

## Article

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

Shang-Tse Ho<sup>1,2,†</sup>, Ying-Ning Ho<sup>3,4,†</sup>, Chih Lin<sup>1</sup>, Wei-Chen Hsu<sup>1</sup> , Han-Jung Lee<sup>1</sup>, Chia-Chi Peng<sup>1</sup>, Han-Tan Cheng<sup>1</sup> and Yu-Liang Yang<sup>1,\*</sup> 

- <sup>1</sup> Agricultural Biotechnology Research Center, Academia Sinica, Taipei 11529, Taiwan; stho@mail.ncyu.edu.tw (S.-T.H.); chih@macrogen-europe.com (C.L.); ny450130@gate.sinica.edu.tw (W.-C.H.); millettaipei@tamu.edu (H.-J.L.); Chia-Chi.Peng@hki-jena.de (C.-C.P.); r04623008@ntu.edu.tw (H.-T.C.)
- <sup>2</sup> Department of Wood Based Materials and Design, College of Agriculture, National Chiayi University, Chiayi 60004, Taiwan
- <sup>3</sup> Institute of Marine Biology, College of Life Science, National Taiwan Ocean University, Keelung 20224, Taiwan; ynho@mail.ntou.edu.tw
- <sup>4</sup> Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung 20224, Taiwan
- \* Correspondence: ylyang@gate.sinica.edu.tw; Tel.: +886-2-2787-2089
- † These authors contributed equally to this work.



**Citation:** Ho, S.-T.; Ho, Y.-N.; Lin, C.; Hsu, W.-C.; Lee, H.-J.; Peng, C.-C.; Cheng, H.-T.; Yang, Y.-L. Integrated Omics Strategy Reveals Cyclic Lipopeptides Empedopeptins from *Massilia* sp. YMA4 and Their Biosynthetic Pathway. *Mar. Drugs* **2021**, *19*, 209. <https://doi.org/10.3390/md19040209>

Academic Editor: Ipek Kurtboke

Received: 11 March 2021

Accepted: 7 April 2021

Published: 9 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Empedopeptins—eight amino acid cyclic lipopeptides—are calcium-dependent antibiotics that act against Gram-positive bacteria such as *Staphylococcus aureus* by inhibiting cell wall biosynthesis. However, to date, the biosynthetic mechanism of the empedopeptins has not been well identified. Through comparative genomics and metabolomics analysis, we identified empedopeptin and its new analogs from a marine bacterium, *Massilia* sp. YMA4. We then unveiled the empedopeptin biosynthetic gene cluster. The core nonribosomal peptide gene null-mutant strains ( $\Delta empC$ ,  $\Delta empD$ , and  $\Delta empE$ ) could not produce empedopeptin, while dioxygenase gene null-mutant strains ( $\Delta empA$  and  $\Delta empB$ ) produced several unique empedopeptin analogs. However, the antibiotic activity of  $\Delta empA$  and  $\Delta empB$  was significantly reduced compared with the wild-type, demonstrating that the hydroxylated amino acid residues of empedopeptin and its analogs are important to their antibiotic activity. Furthermore, we found seven bacterial strains that could produce empedopeptin-like cyclic lipopeptides using a genome mining approach. In summary, this study demonstrated that an integrated omics strategy can facilitate the discovery of potential bioactive metabolites from microbial sources without further isolation and purification.

**Keywords:** *Massilia*; lipopeptides; biosynthesis; genome mining; metabolomics

## 1. Introduction

The production of secondary metabolites is regulated by several genes that are responsible for modifying their chemical structure, transporting substrates and products, and other specific regulatory functions. These genes, named secondary metabolite biosynthesis gene clusters (BGCs), are contiguously aligned in the genome [1]. Advances in computational science and bioinformatic analysis now provide the possibility of dissecting secondary metabolite biosynthetic pathways from microorganisms [2]. Comparative genomics analysis and BGC mining have been applied to the natural product studies to discover new sources of bioactive natural products [3]. In addition, several mass spectrometry based metabolomics analysis approaches, such as Global Natural Products Social Molecular Networking (GNPS), have been utilized to monitor the production of specific molecules [4]. Molecular networking groups similar molecules as a cluster by comparing vector correlations between tandem mass fragment ions of molecules [5]. Thus, integrated

BGC mining and molecular networking analysis are well suited for dereplication and identification of specific metabolites efficiently and even proposes new analog structures.

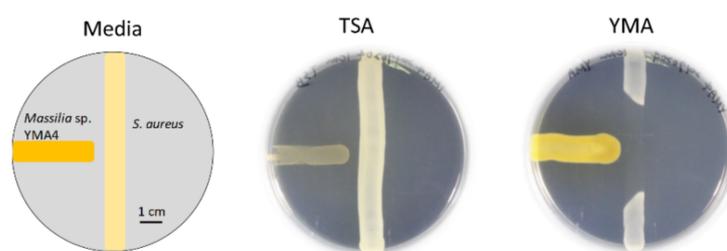
The biodiversity of marine microorganisms is exceptionally high and contributes to the vast chemical diversity seen among their related metabolites [6]. The chemical diversity of metabolites might play vital roles in their bioactivities. One of the significant applications of natural products is antibiotics. As reported previously, approximately 80 natural product-related antibiotics were approved by the FDA from 1981 to 2014 [7]. However, as is now well-known, existing antibiotics cannot effectively treat resistant strains of bacterial infections. Therefore, seeking new antibiotics from natural products is a priority. *Staphylococcus aureus* is a commensal bacterium in humans that colonizes in approximately 30% of the human population. It is also a major human pathogen and the leading cause of bacteremia, infective endocarditis, and other infections in the clinic [8]. Therefore, screening natural resources for effective agents against *S. aureus* infection is a potentially worthwhile challenge.

In this study, we first applied comparative genomics analysis and GNPS to reveal the biosynthetic gene cluster and new analogs of empedopeptin derived from the marine-derived bacterium *Massilia* sp. YMA4. We further confirmed empedopeptins as the main antibiotic utilized by *Massilia* sp. YMA4 against *S. aureus*, using an insertional mutation approach of the biosynthetic genes. We then utilized genome mining to explore new bacterial sources of empedopeptins and related cyclic lipopeptides. Such an integrated approach holds promise for discovering new natural sources of antibiotics and their production.

## 2. Results and Discussion

### 2.1. Antibiotic Activity of *Massilia* sp. YMA4

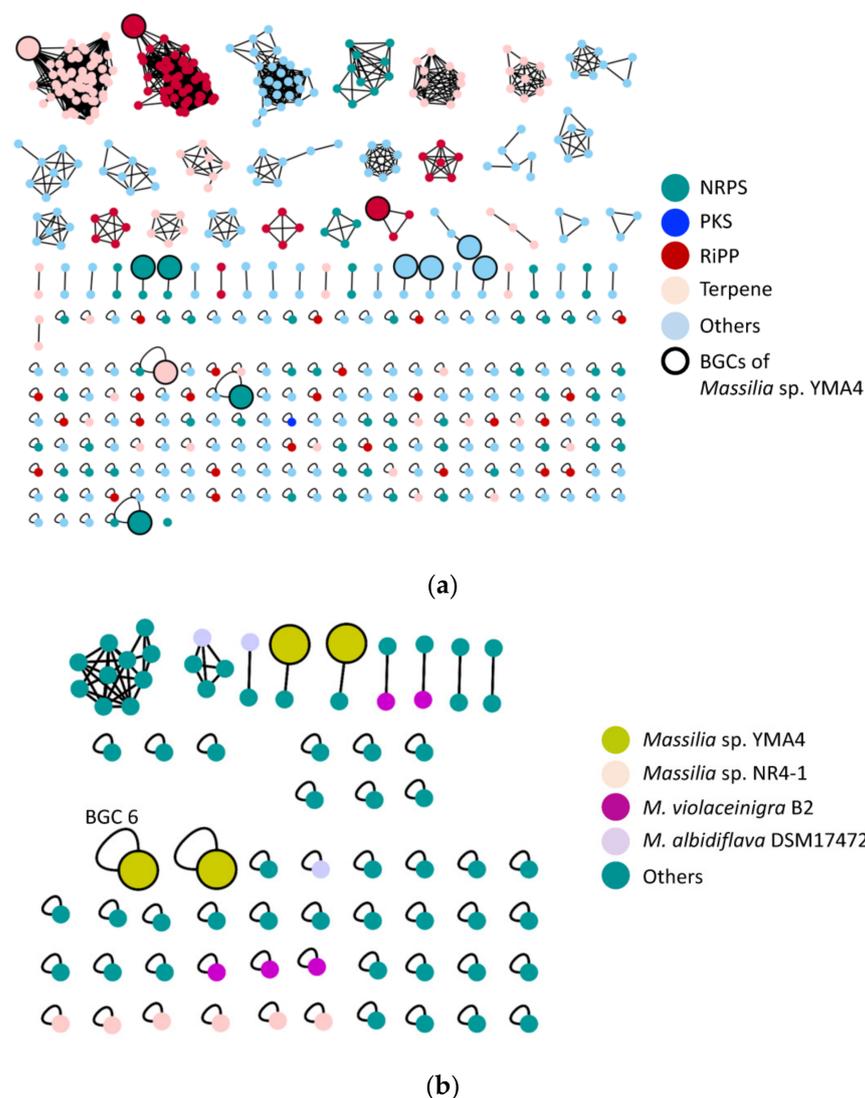
*Massilia*, a Gram-negative bacteria genus, was first found in clinical blood samples in 1998 [9] and was subsequently isolated from various environmental samples [10]. Some strains of the *Massilia* genus, such as *Massilia* sp. BS-1 and *Massilia* sp. NR 4-1 were reported to be antibiotic producers [11,12]. However, knowledge about the potential natural antibiotics related to nonribosomal peptides in the *Massilia* remains scarce. For the screening of new antibiotics, a strain named *Massilia* sp. YMA4 was isolated from a sediment sample collected in the open ocean off Lamay island in Taiwan. The antagonist assay revealed that *Massilia* sp. YMA4 effectively inhibited the growth of *S. aureus* ATCC 29213 only in yeast malt agar (YMA), but not in tryptone soy agar (TSA) culture conditions (Figure 1). To further understand the phylogenetic relationships of *Massilia* sp. YMA4 to other sequenced *Massilia* strains, the whole genome sequences of 69 available *Massilia* strains deposited in the NCBI database (2020/02) were collected and analyzed using the maximum likelihood method. *Massilia* sp. YMA4 was clustered with *M. armeniaca* ZMN-3 (Figure S1). Interestingly, *Massilia* sp. YMA4 and *M. armeniaca* ZMN-3 were close to *Empedobacter halobium* ATCC 31962 according to BLAST analysis using partial 16S rRNA gene [13]. *E. halobium* ATCC 31962 has also been reported to have antimicrobial potential against Gram-positive bacteria [14].



**Figure 1.** Antagonistic assays of *Massilia* sp. YMA4 versus *S. aureus* in yeast malt agar (YMA) and tryptone soy agar (TSA) culture conditions.

## 2.2. The Biosynthetic Potential of Secondary Metabolites of the *Massilia* Strains and the Discovery of the Empedopeptin Biosynthesis Gene Cluster

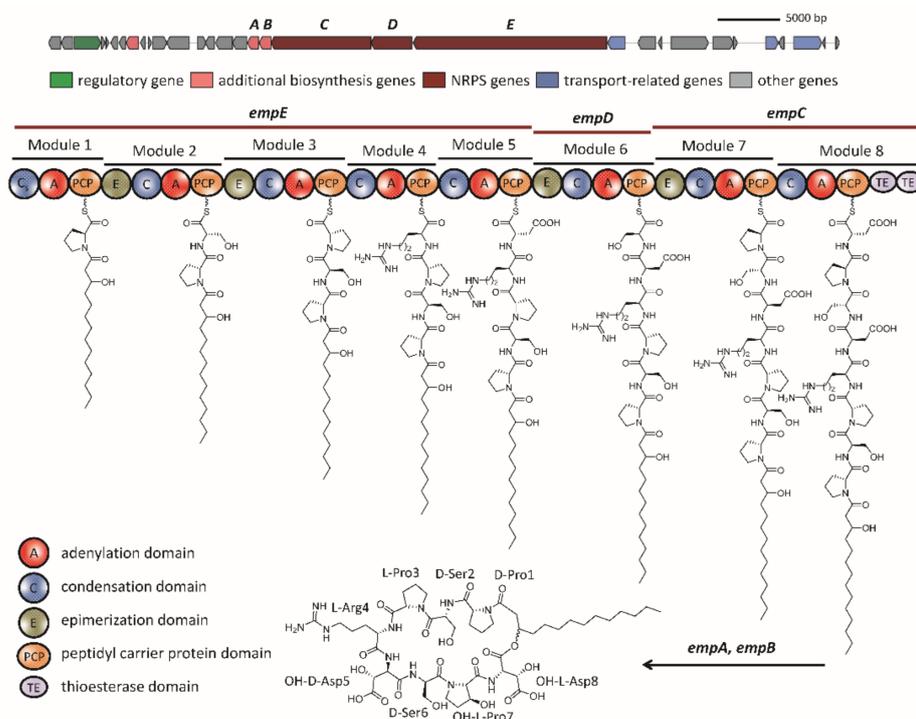
A computational analysis of the secondary metabolite biosynthetic capacity of 72 *Massilia* strains was conducted using antiSMASH 5.1, MiBIG, and BiG-SCAPE [15,16]. The results revealed a high diversity of BGCs in the *Massilia* (Figure 2a). In 72 *Massilia* strains, 490 BGCs were identified, and the homologous BGCs were further grouped into 233 gene cluster families (GCF). Interestingly, 75 nonribosomal peptide BGCs were predicted (NRPS BGC). However, only four *Massilia* strains, *Massilia* sp. NR 4-1 (6 NRPS BGCs), *M. violaceinigra* B2 (5 NRPS BGCs), *Massilia* sp. YMA4 (4 NRPS BGCs), and *M. albidiflava* DSM17472 (3 NRPS BGCs) contained more than three NRPS BGCs (Figure 2b).



**Figure 2.** Sequence similarity network of biosynthetic gene clusters (BGCs) from 72 *Massilia* strains analyzed by antiSMASH. (a) Networks of all the predicted secondary metabolite BGCs. (b) Networks of nonribosomal peptides BGCs. Each node represents one BGC, and the similar BGCs were linked together as one cluster. The single-node indicates unique BGC in the network. All networks were generated by BiG-SCAPE and illustrated with Cytoscape. BGC 6 is the empedopeptin BGC. NRPS: non-ribosomal peptide synthetase; PKS: polyketide synthase; RiPP: ribosomally synthesized and post-translationally modified peptides.

Twelve putative BGCs were predicted in the *Massilia* sp. YMA4 genome, including four NRPS BGCs (Table S1). Among them, BGC 6 was comprised of three NRPS-related

genes, *empC* (two modules, 8.2 kb), *empD* (one module, 3.3 kb), and *empE* (five modules, 15.9 kb); and two additional genes, *empA* (dioxygenase, 0.9 kb) and *empB* (dioxygenase, 0.9 kb) (Figure 3). The modules of *empC*, *empD*, and *empE* genes were comprised of one condensation (C), one adenylation (A), and one peptidyl carrier protein (PCP) domain. In addition, two thioesterase (TE) domains were located in *empC* (Figure 3).



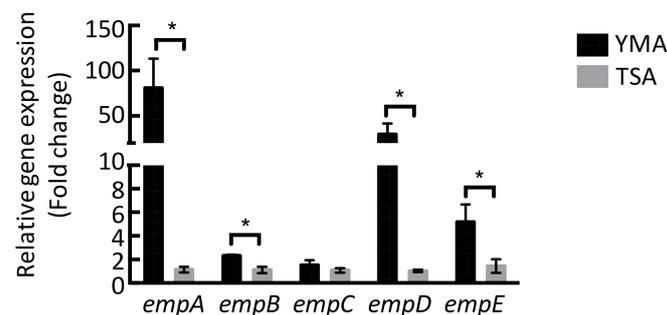
**Figure 3.** In silico analysis of empedopeptin biosynthetic gene cluster and flanking regions of *Massilia* sp. YMA4. The empedopeptin biosynthetic gene cluster was predicted by using antiSMASH5.1. Scale bar = 5 kb.

The A domain in each module was further analyzed using antiSMASH 5.1 and NRPSpredictor2 [15,17], which suggests that the amino acid sequence of the BGC 6 product is: Pro1-Ser2-Pro3-Arg4-Asp5-Ser6-Pro7-Asp8. Additionally, a lipo-initiation (CStarter) domain in module 1 of *empE* was revealed by antiSMASH 5.1 and NaPDos [18]. These results demonstrated that this domain is similar to the CStarter of syringomycin, indicating that this domain can catalyze the acylation of the first amino acid [19]. The combination of the polar peptide and fatty acid tail is a key feature of cyclic lipopeptides that is responsible for their amphiphilic properties [20]. The amphiphilic properties of lipopeptides have great potential for pharmaceutical use and can also deliver drugs by forming active pharmaceutical ingredients as liposomes [21]. The remaining C domains were further analyzed. Among these C domains, modules 2, 3, 6, and 7 were predicted as combined C/E (epimerization) domains, while the other C domains are the conventional  $^L C_L$  domains. The E domain contributes to the stereochemistry of amino acid residues in the NRPS assembly line. It changes the L form amino acid to the D form amino acid ( $^L C_D$ ) [22]. Therefore, Pro1, Ser2, Asp5, and Ser6 are in the D configuration, while Pro3, Arg4, Pro7, and Asp8 are in the L configuration. Interestingly, most NRPS BGCs have one TE domain in the last module, but our results revealed two TE domains in BGC 6. The two TE domains in the last module have also been found in some cyclic lipopeptides BGCs, such as orfamide A, viscosin, and massetolide produced by *P. fluorescens* strains [23]. There are two types of TE domains, type I and II, which exhibit different functions. The type I TE domain can catalyze intramolecular cyclization, while the type II TE domain functions and supports the NRPS assembly line [24]. Taken together, the results of the bioinformatic analysis indi-

cated that the structure produced by the BGC 6 is 3OH-FA-D-Pro1/D-Ser2/L-Pro3/L-Arg4/D-Asp5/D-Ser6/L-Pro7/L-Asp8, which is the same as the core structure of empedopeptin (Figure 3).

### 2.3. Molecular Networking of Empedopeptin and Its Analogs from *Massilia* sp. YMA4 Wild-Type and Mutant Strains

We evaluated the potential empedopeptin biosynthetic-related gene expression under various culture conditions (Figure 4). The results were consistent with the antagonistic assay results, suggesting that the TSA medium could not activate the empedopeptin biosynthetic-related gene expression and thus was unable to inhibit *S. aureus* ATCC 29213 growth (Figure 1).

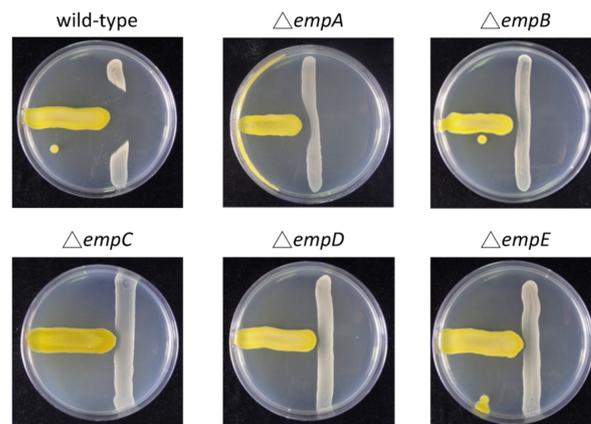


**Figure 4.** Relative gene expression of empedopeptin biosynthetic-related genes in different culture conditions (mean  $\pm$  SD,  $n = 5$ ). The statistical analysis was conducted using the student's t-test. \*  $p < 0.05$ .

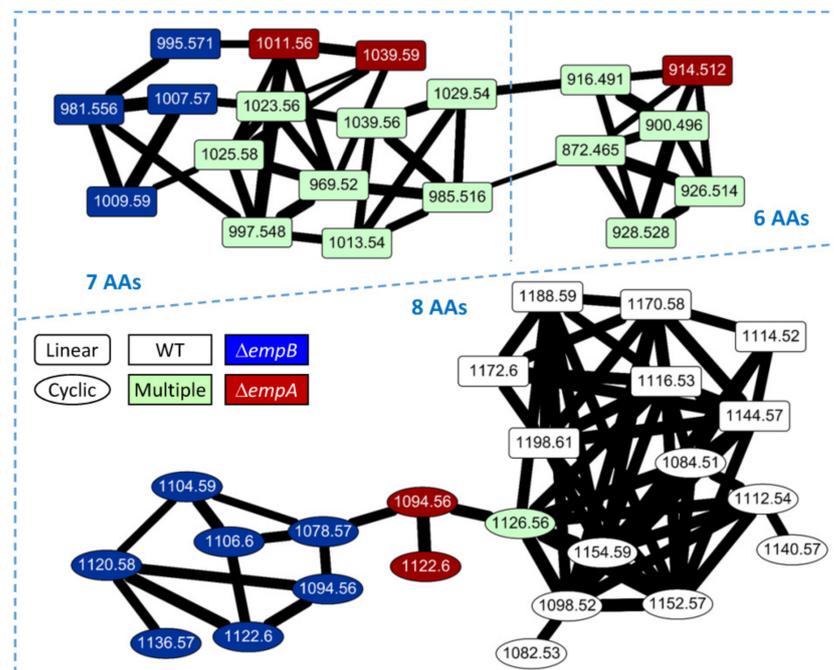
We, therefore, extracted the metabolites from the *Massilia* sp. YMA4 for LC-MS/MS analysis. One significant signal of empedopeptin at  $m/z$  1126.5630 was observed in the extract, suggesting that this compound might act as the main contributor to the antibiotic activity of *Massilia* sp. YMA4 to *S. aureus* ATCC 29213 (Figure S2). The structure of empedopeptin was further confirmed by LC-MS/MS that a base peak at  $m/z$  1126.5630  $[M + H]^+$ , consistent with a molecular formula of  $C_{49}H_{79}N_{11}O_{19}$  and an index of hydrogen deficiency (IHD) of 16. The fragments of b/y ions of empedopeptin are listed in Figure S2. Fragments of b ions in order were, 977.5298 (b7), 864.4842 (b6), 777.4520 (b5), 646.4296 (b4), and 393.2738 (b2), while the fragments of y ion were 918.3830 (y8), 821.3257 (y7), 734.2978 (y6), and 637.2446 (y5). Additionally, we found that the Asp5, Ser7, and Asp8 contain a hydroxyl group.

To further understand the biosynthetic pathway of empedopeptin, we constructed the *empA*, *empB*, *empC*, *empD*, and *empE* null-mutant strains. The mutant strains,  $\Delta empC$ ,  $\Delta empD$ , and  $\Delta empE$ , could not inhibit *S. aureus* ATCC 29213 growth, and  $\Delta empA$  and  $\Delta empB$  strains showed significantly reduced inhibition effects compared to wild-type (Figure 5). We then conducted a comparative metabolomic analysis (molecular networking) of the wild-type and null-mutant strains to elucidate the structures and biosynthetic pathways of empedopeptin and its analogs. We identified 44 empedopeptin analogs, including 17 cyclic lipopeptides and 27 linear lipopeptides (Figure 6). As shown in Figure 6, we observed the empedopeptin analogs only from wild-type,  $\Delta empA$ , and  $\Delta empB$  strains (dioxygenase genes), suggesting that the  $\Delta empC$ ,  $\Delta empD$ , and  $\Delta empE$  strains (core NRPS genes) were unable to produce empedopeptins. Additionally, we found several unique analogs that only appear in the molecular networking analysis of  $\Delta empA$  and  $\Delta empB$  strains. In the lipopeptides containing eight amino acids,  $m/z$  1094.56 and 1122.6 were found in both  $\Delta empA$  and  $\Delta empB$  strains, while  $m/z$  1078.57, 1104.59, 1106.6, 1120.58, and 1136.57 were observed only in the  $\Delta empB$  strain. Comparing the structure of  $m/z$  1094.56 and 1122.6 from  $\Delta empA$  and  $\Delta empB$  strains showed that they comprise the identical amino acid sequences as empedopeptin, but have hydroxyl modification in different amino acids and different chain lengths of 3-hydroxyl fatty acid as well. Asp5, Pro7, and Asp8 residues were

all hydroxylated in empedopeptin. We found cyclic lipopeptides containing OH-Asp5 together with unmodified Pro7 and Asp8 from  $\Delta empA$  strain, while cyclic lipopeptides with OH-Asp8 and even no hydroxylated residues were only identified from  $\Delta empB$  strain (Table S2). These results demonstrated that the dioxygenase genes, *empA* and *empB*, located at the end of the empedopeptin biosynthetic assembly line, functioned in hydroxylation and played an important role in the bioactivity of cyclic lipopeptides. In addition to the cyclic lipopeptides, we also identified several linear lipopeptides, which might be the biosynthetic intermediates of empedopeptins (Figure 6). The structure information of these empedopeptin analogs is summarized in Tables S2 and S3.



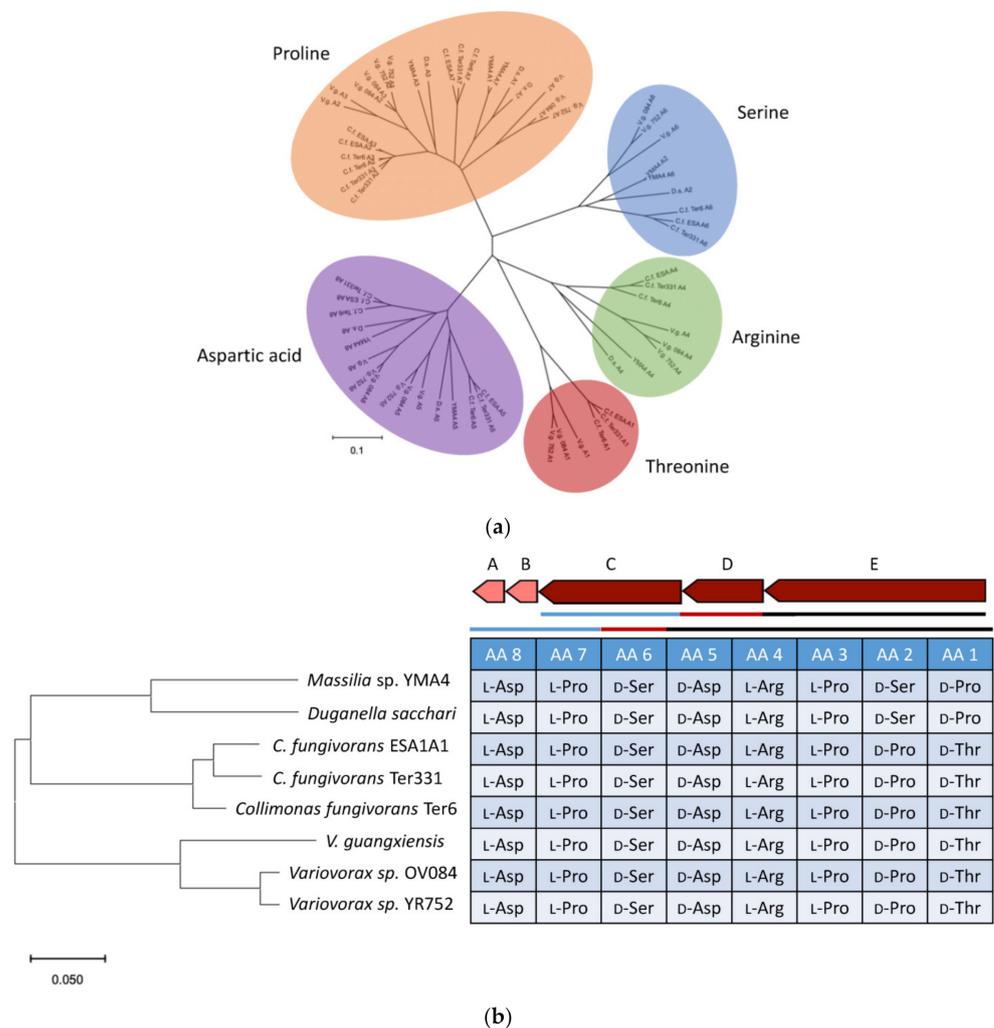
**Figure 5.** Antagonistic assays of the wild-type and core empedopeptin biosynthetic gene null-mutant strains ( $\Delta empA-E$ ) of *Massilia* sp. YMA4 versus *S. aureus* on YMA.



**Figure 6.** Molecular networking of empedopeptin analogs from *Massilia* sp. YMA4 wild-type (WT) and null-mutant strains. 6 AAs: 6 amino acids lipopeptides; 7 AAs: 7 amino acids lipopeptides; and 8 AAs: 8 amino acids lipopeptides. Each node represents one lipopeptide tandem mass spectrum, and the width of the edge indicates the tandem mass spectral similarity of neighboring nodes. The shape of the node represents linear or cyclic lipopeptide. The color of the node represents the source of the lipopeptide. Multiple represents the analog was detected from more than one strain.

#### 2.4. Genome Mining for Discovering Other Empedopeptin-Like Compound Producing Bacteria

Through the MultiGeneBlast analysis, we found that several microorganisms may produce cyclic lipopeptides similar to empedopeptins because similar NRPS BGCs were observed. The adenylation domain plays a role in selecting and recruiting specific amino acids in the NRPS biosynthetic pathway [25]. Therefore, to further propose the structures of cyclic lipopeptides, all the adenylation domains in the empedopeptin-related BGCs were analyzed phylogenetically (Figure 7a). Five types of amino acids, arginine, aspartic acid, proline, serine, and threonine, were involved in those BGCs. As shown in Figure 7b, *Duganella sacchari* has the same NRPS BGC for producing empedopeptin-like compounds. However, *Collimonas* and *Variovorax* strains were proposed to be the producers of tripropeptin-like compounds [26].



**Figure 7.** Phylogenetic analyses of adenylation (A) domains and biosynthetic gene clusters (BGCs) of empedopeptin-like cyclic lipopeptides. (a) Phylogenetic analysis of A domains in BGCs of empedopeptin-like cyclic lipopeptides. (b) Phylogenetic analysis of BGCs of empedopeptin-like cyclic lipopeptides and their proposed amino acid sequences. AA1-AA5 from gene E; AA6 from gene D; AA7-AA8 from gene C.

In the present study, we aimed to explore the antibiotic substances from the marine bacterium *Massilia* sp. YMA4 by integrating comparative genomics and metabolomics analysis. Our results revealed that the primary active substance of *Massilia* sp. YMA4 is a nonribosomal peptide, empedopeptin. We found that *Massilia* sp. YMA4 only inhibited *S. aureus* ATCC 29213 under certain culture conditions, suggesting that the empedopeptin

biosynthetic gene cluster might be a silent gene cluster activated by changing culture conditions. This result was consistent with a previous study, demonstrating that microbial secondary metabolite synthesis is influenced by changing the type and concentration of nutrients in the culture media [27].

Empedopeptin is a calcium-dependent antibiotic that acts against Gram-positive bacteria by inhibiting cell wall biosynthesis [28]. Certain calcium-dependent antibiotics contain a conserved Asp-X-Asp-Gly motif that is thought to facilitate calcium binding [29]. This study confirmed that two dioxygenase genes, *empA* and *empB*, contributed to the hydroxylation of the amino acid residues in empedopeptin. The  $\Delta empA$  and  $\Delta empB$  strains showed significantly reduced inhibition effects compared to the wild-type, which implies the hydroxylated amino acid residues are important to antibiotic activity. We speculated that the two dioxygenases worked synergistically to hydroxylate the amino acid residues of empedopeptin. However, it is challenging to validate the hydroxylation selectivity of these two dioxygenases to specific amino acid residues or peptide sequences.

Some bacteria were found to possess cyclic lipopeptide BGCs that are similar to empedopeptin. However, the amino acid residues of those cyclic lipopeptides were different. As shown in Figure 7b, the first and second residues of cyclic lipopeptides of *Collimonas* and *Variovorax* strains encode threonine and proline, respectively, while the same positions in *Massilia* sp. YMA4 and *D. sacchari* encode proline and serine, respectively. Surprisingly, we did not find the empedopeptin-related BGC from empedopeptin-producing bacteria *E. haloabium* ATCC 31962 through MultiGeneBlast analysis [13,30]. Therefore, we speculate that the whole genome of *E. haloabium* ATCC 31962 might not be well sequenced.

### 3. Materials and Methods

#### 3.1. Bacterial Culture Conditions and Antagonistic Assay

*Massilia* sp. YMA4 was isolated from a marine sediment core collected by research vessel Ocean Researcher No. 3, offshore of Lamay island, Pingtung County, Taiwan, on 9 November, 2013 (OR3-1727). A voucher specimen (BCRC 81003) was deposited in the Bioresource Collection and Research Center (BCRC), Food Industry Research and Development Institute, Hsinchu, Taiwan. The 16S rRNA gene sequencing (Accession: KX444135) was affiliated with the genus *Massilia*. *Massilia* sp. YMA4 was cultivated in yeast malt broth (YMB) consisting of 3 g/L yeast extract; 3 g/L malt extract; 10 g/L dextrose; 5 g/L peptone at 30 °C, with orbital shaking (120 rpm) for 24 h. The barcoding gene, 16S rRNA sequence, was used to identify the strain as a *Massilia* sp. *S. aureus* ATCC 29213 was purchased from BCRC. For the antagonistic assay, *Massilia* sp. YMA4 was grown on tryptone soy broth (TSB) or YMB overnight at 30 °C. When the O.D. value of bacterial cultures reached 2.0, the bacteria were then inoculated to TSA and YMA plates at 30 °C. After two days of incubation, *S. aureus* ATCC 29213 (grown in LB, O.D. = 2.0) were then transferred to the same agar plates to perform the confrontation test at 30 °C.

#### 3.2. DNA Sample Preparation, Whole-Genome Sequencing, Assembly, and Annotation

Genomic DNA of *Massilia* sp. YMA4 was extracted from the cultures grown at 30 °C in YMB medium using a genomic DNA purification kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. The genomic DNA (total 20 µg) was sequenced by the PacBio RS II system (Pacific Biosciences, Menlo Park, CA, USA). A 10-kb SMRTbell library was generated using a DNA Template Prep Kit 2.0 (10–20 Kb; Pacific Biosciences), following the manufacturer's instructions. Sequencing was performed with a PacBio RS II sequencer using one SMRT cell and P6-C4 chemistry at 360 min movie length (Pacific Biosciences, Menlo Park, CA, USA). Single-molecule real-time reads (159,840 filtered subreads and a mean length of 11,850 bp) were de novo assembled using the Hierarchical Genome Assembly Process (HGAP) workflow in the SMRT analysis software version 3 (Pacific Biosciences Inc., Menlo Park, CA, USA) [31]. Genome annotation was performed via Rapid Annotations using Subsystems Technology (RAST) server version 2.0 and blast2go [32,33].

The whole-genome sequence of *Massilia* sp. YMA4 was uploaded to the Genbank database (GenBank assembly accession: GCA\_003293715.1).

### 3.3. Genome Mining and Bioinformatic Analysis

The whole-genome sequence of *Massilia* sp. YMA4 was analyzed by antiSMASH 5.1 for the genome mining of possible secondary metabolite biosynthetic gene clusters [15]. A phylogenetic tree based on assembled genome sequencing data of genus *Massilia* from NCBI, including *Massilia* sp. YMA4 was performed using reference alignment-based phylogenetic builder (REALPHY) via bowtie2 [34]. The phylogeny was estimated on the “polymorphisms\_move.phy” file produced by REALPHY via RAxML with the general time reversible (GTR) nucleotide substitution model and GAMMA model of rate heterogeneity [35]. Finally, RAxML was carried out with 1000 alternative runs on distinct starting trees (N) and a search for the best-scoring ML tree with 1000 replications. The phylogenetic tree was generated by using iTOL [36]. All BGCs associated with the genus *Massilia*, a total of 490 BGCs containing 233 gene cluster families, were downloaded and used for the sequence similarity network. The BIG-SCAPE-CORASON pipeline was utilized locally to analyze the 490 BGCs downloaded from the antiSMASH database (March 2019) [15,16]. The singleton parameter in the sequence similarity network was the BGCs with distances lower than the default cutoff distance of 0.3. Generated sequencing similarity network files separated by BiG-SCAPE class were combined for visualization using Cytoscape version 3.7.1 [37].

Each domain of empedopeptin BGC was predicted using antiSMASH 5.1 or NRP-Spredicator2 [17]. The screening of empedopeptin-related biosynthetic genes from various microorganisms was performed using MultiGeneBlast [30]. For the MultiGeneBlast analysis, a local MultigeneBlast database was applied in this study, built with 101,223 bacterial genomic sequences from the GenBank assembly database of NCBI (download date: March 2019) and processed by MultigeneBlast under a Linux-based system. The sequences from *empA* to *empE* were the query template. The whole-genome sequences of *C. fungivorans* ESAIA1 (GenBank assembly accession: GCA\_003610095.1), *C. fungivorans* Ter331 (GenBank assembly accession: GCA\_000221045.1), *C. fungivorans* Ter6 (GenBank assembly accession: GCA\_001584145.1), *D. sacchari* (GenBank assembly accession: GCA\_900143065.1), *E. haloabium* ATCC 31962 (GenBank assembly accession: GCA\_008011715.1), *V. guangxiensis* (GenBank assembly accession: GCA\_003952165.1), *Variovorax* sp. OV084 (GenBank assembly accession: GCA\_900111625.1), and *Variovorax* sp. YR752 (GenBank assembly accession: GCA\_900215425.1) were downloaded from the GenBank database (NCBI). Phylogenetic analysis sequences of empedopeptin-related biosynthetic genes of selected microbial strains were then analyzed to compare the similarity of empedopeptin-related biosynthetic genes between each strain (Table S4). The phylogenetic analysis of empedopeptin-related genes was constructed with Molecular Evolutionary Genetics Analysis version X (MEGAX) [38]. Predicted BGCs among different genomes were analyzed using a modified Pfam domain similarity metric implemented in BigScape for the similarity network of the BGCs [39,40]. A cutoff of 0.75 was used for the analysis [40].

### 3.4. Construction of Empedopeptin Biosynthetic Gene-Null Mutant Strains

To make insertion mutations of empedopeptin biosynthetic genes *empA*, *empB*, *empC*, *empD*, and *empE* in *Massilia* sp. YMA4, pCM184- $\Delta$ *empA*, pCM184- $\Delta$ *empB*, pCM184- $\Delta$ *empC*, pCM184- $\Delta$ *empD*, and pCM184- $\Delta$ *empE* were constructed. The schematic diagram of the insertion mutant is shown in Figure S3. The pCM184 was purchased from Addgene (MA, USA). The regions for homologous recombination in *Massilia* sp. YMA4 were amplified by primer pairs, which are listed in Tables S5 and S6. The PCR products were digested with KpnI and SacI enzymes and cloned into the KpnI/SacI site of the digested pCM184, followed by transformation into the *E. coli* S17-1  $\lambda$  pir. The *E. coli* S17-1 was applied to transform pCM184- $\Delta$ *empA*, pCM184- $\Delta$ *empB*, pCM184- $\Delta$ *empC*, pCM184- $\Delta$ *empD*, and pCM184- $\Delta$ *empE* individually into *Massilia* sp. YMA4 in this study. Both *Massilia* sp. YMA4

and *E. coli* S17-1 were cultured overnight in YMB at 30 °C and in LB at 37 °C, respectively. Both cultures were then washed by YMB twice, and the pellets were suspended by YMB again. The O.D.600 values of the donor (*E. coli* S17-1) and recipient cells (*Massilia* sp. YMA4) were 2.0, then the donor and recipient cells were mixed at 1:1 (*v:v*) ratio. The mixtures were spotted on YMA for overnight inoculation. The colonies were scraped up and spread on YMA with tetracycline (2 µg/mL) and kanamycin (50 µg/mL), and then the conjugants were picked and checked by PCR after 48 h. The conjugants could grow if the plasmid was integrated into the *Massilia* sp. YMA4 genomic DNA. The insertion sites were confirmed by PCR (Figure S4).

### 3.5. Quantitative Reverse Transcription PCR (qRT-PCR)

To evaluate *emp* biosynthetic gene expressions of *Massilia* sp. YMA4 on TSA and YMA culture conditions, the *empA*, *empB*, *empC*, *empD*, and *empE* were analyzed by qRT-PCR. Firstly, *Massilia* sp. YMA4 was grown on TSB or YMB overnight at 30 °C. When the O.D.600 values of bacterial cultures reached 2.0, the 100 µL of bacterial cultures were then spread on TSA and YMA plates at 30 °C. After two days, the bacterial cells on TSA and YMA were scratched and collected in 1.5-mL microcentrifuge tubes. Total RNA of bacterial cells was extracted using the TRIZOL (Ambion, Waltham, MA, USA) reagents following the manufacturer's instructions. The resulting RNA samples were dissolved in 20 µL RNase-free water and then qualified using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The cDNA samples were reverse transcribed from RNA templates using the ToolsQuant II Fast RT Kit (BioTools, New Taipei City, Taiwan) according to the manufacturer's instructions. In brief, 8 µL of total RNA template (2 µg) was mixed with 2 µL of 5× gDNA Eraser and then incubated at 42 °C for 3 min to remove genomic DNA. The gDNA removal reaction mixtures were mixed with 1 µL of RT Enzyme Mix, 2 µL of RT Primer Mix, 2 µL of 10× Fast RT Buffer, and 5 µL RNase-free water and incubated at 42 °C for 15 min. Terminate the reactions at 95 °C for 3 min. The synthesized cDNA samples were stored at −20 °C until analyzed.

The primer sets used for the qRT-PCR were listed in Table S5. The qPCR reaction consisted of 10 µL of SYBR Green Master Mix (BioTools, New Taipei City, Taiwan), 0.4 µL of each of the forward and reverse primers (100 µM), 2 µL of cDNA template, and 7.2 µL of ddH<sub>2</sub>O for a total volume of 20 µL. The reactions were further amplified by the following thermocycling steps: 95 °C for 5 min; 39 cycles of 95 °C for 10 s (denaturation), 60 °C for 5 s (annealing), 72 °C for 30 s (extension), and then conducted melting curve analysis (95 °C for 15 s, 60 °C for 5 s, then increased 0.5 °C/s up to 95 °C). The RT-qPCR analysis was performed in CFX96™ Real-Time PCR Detection System with C1000™ thermal cycler (Bio-Rad, Hercules, CA, USA) with five replicates. The results were analyzed using the CFX Manager software, the delta CT (cycle of threshold) was detected to determine the relative expression level of genes. All target genes were further normalized to the bacterial 16S gene (Internal control).

### 3.6. Comparative Metabolomics and Molecular Networking Analysis

For comparative metabolomics, *Massilia* sp. YMA4 and mutants were cultured on YMA until colony formation. The single colony of *Massilia* sp. YMA4 was picked and transferred to 3 mL YMB for overnight incubation. For the mutants, the single colony was transferred to 3 mL YMB with 0.1% tetracycline. After overnight incubation, 0.5 mL of *Massilia* sp. YMA4 or mutant cultures were transferred to 500 mL YMB w/wo 500 µg tetracycline for three days. The resulting culture broth was extracted by EtOAc. The remaining broth was subsequently extracted with BuOH to obtain BuOH extracts. The BuOH extracts of each strain (10 mg/mL) were dissolved in MeOH. Therefore, 10 µL of extracts were injected and separated by a C18 column (ACQUITY UPLC BEH C18, 1.7 µm, 2.1 × 100 mm, Waters, Milford, MA, USA) with the following gradient: 0–8 min at 5–99.5% of B (A: ACN: H<sub>2</sub>O = 2: 98 plus 0.1% formic acid; B: ACN plus 0.1% formic acid), 8–10 min at 99.5% of B, 10–10.2 min at 99.5–5% of B, 10.2–12 min at 5% of B. Flow

rate was set at 0.4 mL/min. The mass data were acquired using UPLC-ESIMS (Thermo Scientific Orbitrap Elite Mass Spectrometer), and the mass range was set up as  $m/z$  100–1500. The mass data were acquired in profile mode for molecular networking analysis, positive mode ion detection between  $m/z$  100–1500 with 30,000 resolution. The top four intense ions from each full mass scan were selected for collision-induced dissociation (CID) fragmentation. For CID, isolation width was 2 Da, and the selected ions were fragmented with normalized collision energy 30.0, activation Q 0.250, activation time 10.0, and 15,000 resolution. The mass data (.RAW files) from Xcalibur were converted to mzML file format using ProteoWizard software and subjected to GNPS (<https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>, accessed on 20 March 2019) to generate molecular networking, and the data was visualized in Cytoscape. The LC-MS/MS data of the BuOH extracts of *Massilia* sp. YMA4 and its mutant strains (MSV000086803) are publicly available via MassIVE (<https://massive.ucsd.edu>, accessed on 3 February 2021).

#### 4. Conclusions

In summary, we explored the cyclic lipopeptide, empedopeptin, and its analogs from *Massilia* sp. YMA4 using an integrated omics approach. To elucidate the biosynthetic mechanism of empedopeptin, we constructed five empedopeptin biosynthetic gene null-mutants of *Massilia* sp. YMA4. The antibiotic activities of *Massilia* sp. YMA4 mutants were decreased in comparison with the wild-type, demonstrating that empedopeptins were the primary antibiotic substances of *Massilia* sp. YMA4. In the comparative metabolomics analysis of wild-type and null-mutant strains, we successfully identified 44 empedopeptin analogs, including 17 cyclic lipopeptides and 27 linear lipopeptides. Through MultiGeneBlast analysis, we were also able to survey the similar BGCs of empedopeptin from other microbes, including *D. sacchari*, three *Collimonas*, and three *Variovorax* strains. Our findings illustrated that this integrated omics strategy is suitable to identify and develop potential bioactive metabolites from microbial sources without further isolation and purification.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/1660-397/19/4/209/s1>, Figure S1: Phylogenetic tree of genus *Massilia*, Figure S2: Mass spectrum and fragments information of empedopeptin, Figure S3: Schematic diagram of insertion mutant for *Massilia* sp. YMA4, Figure S4: PCR check of mutants. Table S1: Secondary metabolite gene clusters identified in *Massilia* sp. YMA4, Table S2: Structure information of cyclic lipopeptides, Table S3: Structure information of linear lipopeptides, Table S4: Genetic context and features of the BGC 6 in *Massilia* sp. YMA4 and its related BGCs in various microbes, Table S5: Oligonucleotides primer sets used in this study, Table S6: Primer sets for empedopeptin biosynthetic genes of *Massilia* sp. YMA4.

**Author Contributions:** Conceptualization, Y.-L.Y., Y.-N.H., and S.-T.H.; methodology, Y.-L.Y., Y.-N.H., and C.L.; formal analysis, C.-C.P., Y.-N.H., S.-T.H., and Y.-L.Y.; validation, W.-C.H. and C.L.; investigation, W.-C.H., C.L., H.-T.C., H.-J.L., Y.-N.H., and S.-T.H.; writing—original draft preparation, Y.-N.H. and S.-T.H.; writing—review and editing, S.-T.H., H.-J.L., and Y.-L.Y.; funding acquisition, Y.-L.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology of Taiwan (MOST 104-2320-B-001-019-MY2).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article and Supplementary Materials.

**Acknowledgments:** We thank Chao-Jen Shih (BCRC) for isolation and identification of *Massilia* sp. YMA4; Ning Lu (ABRC) for preliminary data collection. The materials and methods for the construction of biosynthetic gene-null mutant strains were generously provided by Nai-Chun Lin's Lab (National Taiwan University, Taiwan). The mass spectrometry analysis was supported by the Metabolomics Core Facility, Instrument Service Center at Academia Sinica.

**Conflicts of Interest:** The authors declare that there are no conflict of interest.

## References

1. Hoffmeister, D.; Keller, N.P. Natural products of filamentous fungi: Enzymes, genes, and their regulation. *Nat. Prod. Rep.* **2007**, *24*, 393–416. [[CrossRef](#)] [[PubMed](#)]
2. Tran, P.N.; Yen, M.R.; Chiang, C.Y.; Lin, H.C.; Chen, P.Y. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 3277–3287. [[CrossRef](#)] [[PubMed](#)]
3. Atencio, L.A.; Boya, P.C.; Martin, H.C.; Mejia, L.C.; Dorrestein, P.C.; Gutierrez, M. Genome mining, microbial interactions, and molecular networking reveals new dibromoalterochromides from strains of *Pseudoalteromonas* of Coiba National Park-Panama. *Mar. Drugs* **2020**, *18*, 456. [[CrossRef](#)] [[PubMed](#)]
4. Koyama, N.; Tomoda, H. MS network-based screening for new antibiotics discovery. *J. Antibiot.* **2019**, *72*, 54–56. [[CrossRef](#)]
5. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapon, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [[CrossRef](#)]
6. Wiese, J.; Imhoff, J.F. Marine bacteria and fungi as promising source for new antibiotics. *Drug Dev. Res.* **2019**, *80*, 24–27. [[CrossRef](#)]
7. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629–661. [[CrossRef](#)]
8. Tong, S.Y.; Davis, J.S.; Eichenberger, E.; Holland, T.L.; Fowler, V.G., Jr. *Staphylococcus aureus* infections: Epidemiology, pathophysiology, clinical manifestations, and management. *Clin. Microbiol. Rev.* **2015**, *28*, 603–661. [[CrossRef](#)]
9. La Scola, B.; Birtles, R.J.; Mallet, M.N.; Raoult, D. *Massilia timonae* gen. nov., sp. nov., isolated from blood of an immunocompromised patient with cerebellar lesions. *J. Clin. Microbiol.* **1998**, *36*, 2847–2852. [[CrossRef](#)]
10. Ofek, M.; Hadar, Y.; Minz, D. Ecology of root colonizing *Massilia* (Oxalobacteraceae). *PLoS ONE* **2012**, *7*, e40117. [[CrossRef](#)]
11. Agematu, H.; Suzuki, K.; Tsuya, H. *Massilia* sp. BS-1, a novel violacein-producing bacterium isolated from soil. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 2008–2010. [[CrossRef](#)]
12. Myeong, N.R.; Seong, H.J.; Kim, H.J.; Sul, W.J. Complete genome sequence of antibiotic and anticancer agent violacein producing *Massilia* sp. strain NR 4-1. *J. Biotechnol.* **2016**, *223*, 36–37. [[CrossRef](#)]
13. Miess, H.; Arlt, P.; Apel, A.K.; Weber, T.; Nieselt, K.; Hanssen, F.; Czimmel, S.; Nahnsen, S.; Gross, H. The draft whole-genome sequence of the antibiotic producer *Empedobacter haloabium* ATCC 31962 provides indications for its taxonomic reclassification. *Microbiol. Resour. Announc.* **2019**, *8*, 45. [[CrossRef](#)]
14. Konishi, M.; Sugawara, K.; Hanada, M.; Tomita, K.; Tomatsu, K.; Miyaki, T.; Kawaguchi, H.; Buck, R.E.; More, C.; Rossomano, V.Z. Empedopeptin (BMY-28117), a new depsipeptide antibiotic. I. Production, isolation and properties. *J. Antibiot.* **1984**, *37*, 949–957. [[CrossRef](#)]
15. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, *47*, W81–W87. [[CrossRef](#)]
16. Navarro-Munoz, J.C.; Selem-Mojica, N.; Mullowney, M.W.; Kautsar, S.A.; Tryon, J.H.; Parkinson, E.I.; De Los Santos, E.L.C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **2020**, *16*, 60–68. [[CrossRef](#)]
17. Rottig, M.; Medema, M.H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **2011**, *39*, W362–W367. [[CrossRef](#)]
18. Ziemert, N.; Podell, S.; Penn, K.; Badger, J.H.; Allen, E.; Jensen, P.R. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **2012**, *7*, e34064.
19. Girard, L.; Hofte, M.; Mot, R. Lipopeptide families at the interface between pathogenic and beneficial *Pseudomonas*-plant interactions. *Crit. Rev. Microbiol.* **2020**, *46*, 397–419. [[CrossRef](#)]
20. Gross, H.; Loper, J.E. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat. Prod. Rep.* **2009**, *26*, 1408–1446. [[CrossRef](#)]
21. Hutchinson, J.A.; Burholt, S.; Hamley, I.W. Peptide hormones and lipopeptides: From self-assembly to therapeutic applications. *J. Pept. Sci.* **2017**, *23*, 82–94. [[CrossRef](#)] [[PubMed](#)]
22. Sussmuth, R.D.; Mainz, A. Nonribosomal peptide synthesis—principles and prospects. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 3770–3821. [[CrossRef](#)] [[PubMed](#)]
23. Loper, J.E.; Hassan, K.A.; Mavrodi, D.V.; Davis, E.W., 2nd; Lim, C.K.; Shaffer, B.T.; Elbourne, L.D.; Stockwell, V.O.; Hartney, S.L.; Breakwell, K.; et al. Comparative genomics of plant-associated *Pseudomonas* spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet.* **2012**, *8*, e1002784. [[CrossRef](#)] [[PubMed](#)]
24. Jahanshah, G.; Yan, Q.; Gerhardt, H.; Pataj, Z.; Lammerhofer, M.; Pianet, I.; Josten, M.; Sahl, H.G.; Silby, M.W.; Loper, J.E.; et al. Discovery of the cyclic lipopeptide gacamide A by genome mining and repair of the defective GacA regulator in *Pseudomonas fluorescens* Pf0-1. *J. Nat. Prod.* **2019**, *82*, 301–308. [[CrossRef](#)]
25. Strieker, M.; Tanovic, A.; Marahiel, M.A. Nonribosomal peptide synthetases: Structures and dynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 234–240. [[CrossRef](#)]
26. Medema, M.H.; Paalvast, Y.; Nguyen, D.D.; Melnik, A.; Dorrestein, P.C.; Takano, E.; Breitling, R. Pep2Path: Automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput. Biol.* **2014**, *10*, e1003822. [[CrossRef](#)]
27. Sanchez, S.; Chavez, A.; Forero, A.; Garcia-Huante, Y.; Romero, A.; Sanchez, M.; Rocha, D.; Sanchez, B.; Avalos, M.; Guzman-Trampe, S.; et al. Carbon source regulation of antibiotic production. *J. Antibiot.* **2010**, *63*, 442–459. [[CrossRef](#)]

28. Muller, A.; Munch, D.; Schmidt, Y.; Reder-Christ, K.; Schiffer, G.; Bendas, G.; Gross, H.; Sahl, H.G.; Schneider, T.; Brotz-Oesterhelt, H. Lipodepsipeptide empedopeptin inhibits cell wall biosynthesis through Ca<sup>2+</sup>-dependent complex formation with peptidoglycan precursors. *J. Biol. Chem.* **2012**, *287*, 20270–20280. [[CrossRef](#)]
29. Hover, B.M.; Kim, S.H.; Katz, M.; Charlop-Powers, Z.; Owen, J.G.; Ternei, M.A.; Maniko, J.; Estrela, A.B.; Molina, H.; Park, S.; et al. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol.* **2018**, *3*, 415–422. [[CrossRef](#)]
30. Medema, M.H.; Takano, E.; Breitling, R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **2013**, *30*, 1218–1223. [[CrossRef](#)]
31. Chin, C.-S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563. [[CrossRef](#)]
32. Conesa, A.; Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant. Genomics* **2008**, *2008*, 619832. [[CrossRef](#)]
33. Overbeek, R.; Olson, R.; Pusch, G.D.; Olsen, G.J.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Parrello, B.; Shukla, M. The SEED and the rapid rnotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* **2013**, *42*, D206–D214. [[CrossRef](#)]
34. Bertels, F.; Silander, O.K.; Pachkov, M.; Rainey, P.B.; van Nimwegen, E. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol. Biol. Evol.* **2014**, *31*, 1077–1088. [[CrossRef](#)]
35. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [[CrossRef](#)]
36. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [[CrossRef](#)]
37. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
38. Hall, B.G. Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **2013**, *30*, 1229–1235. [[CrossRef](#)]
39. Cenicerros, A.; Dijkhuizen, L.; Petrusma, M.; Medema, M.H. Genome-based exploration of the specialized metabolic capacities of the genus *Rhodococcus*. *BMC Genom.* **2017**, *18*, 1–16. [[CrossRef](#)]
40. Yeong, M. *BiG-SCAPE: Exploring Biosynthetic Diversity through Gene Cluster Similarity Networks*; Wageningen University: Wageningen, The Netherlands, 2016.