

## Article

# Cryptic Diversity of Black Band Disease Cyanobacteria in *Siderastrea siderea* Corals Revealed by Chemical Ecology and Comparative Genome-Resolved Metagenomics

Julie L. Meyer <sup>1,\*</sup>, Sarath P. Gunasekera <sup>2</sup>, Anya L. Brown <sup>3,4</sup>, Yousong Ding <sup>5</sup>, Stephanie Miller <sup>1</sup>, Max Teplitski <sup>1,2</sup> and Valerie J. Paul <sup>2</sup>

<sup>1</sup> Department of Soil, Water, and Ecosystem Sciences, University of Florida, Gainesville, FL 32610, USA

<sup>2</sup> Smithsonian Marine Station, Ft. Pierce, FL 34949, USA

<sup>3</sup> School of Natural Resources and Environment, University of Florida, Gainesville, FL 32603, USA

<sup>4</sup> Department of Evolution and Ecology & Bodega Marine Lab, University of California, Bodega Bay, CA 94923, USA

<sup>5</sup> Department of Medicinal Chemistry & Center for Natural Products, Drug Discovery and Development, University of Florida, Gainesville, FL 32603, USA

\* Correspondence: juliemeyer@ufl.edu

**Abstract:** Black band disease is a globally distributed and easily recognizable coral disease. Despite years of study, the etiology of this coral disease, which impacts dozens of stony coral species, is not completely understood. Although black band disease mats are predominantly composed of the cyanobacterial species *Roseofilum reptotaenium*, other filamentous cyanobacterial strains and bacterial heterotrophs are readily detected. Through chemical ecology and metagenomic sequencing, we uncovered cryptic strains of *Roseofilum* species from *Siderastrea siderea* corals that differ from those on other corals in the Caribbean and Pacific. Isolation of metabolites from *Siderastrea*-derived *Roseofilum* revealed the prevalence of unique forms of luokeyolides, distinct from previously characterized *Roseofilum reptotaenium* strains. In addition, comparative genomics of *Roseofilum* strains showed that only *Siderastrea*-based *Roseofilum* strains have the genetic capacity to produce lasso peptides, a family of compounds with diverse biological activity. All nine *Roseofilum* strains examined here shared the genetic capacity to produce luokeyolides and malyngamides, suggesting these compounds support the ecology of this genus. Similar biosynthetic gene clusters are not found in other cyanobacterial genera associated with black band disease, which may suggest that luokeyolides and malyngamides contribute to disease etiology through yet unknown mechanisms.

**Keywords:** *Roseofilum*; luokeyolide; malyngamide; lasso peptide; metagenome-assembled genomes; biosynthetic gene cluster



**Citation:** Meyer, J.L.; Gunasekera, S.P.; Brown, A.L.; Ding, Y.; Miller, S.; Teplitski, M.; Paul, V.J. Cryptic Diversity of Black Band Disease Cyanobacteria in *Siderastrea siderea* Corals Revealed by Chemical Ecology and Comparative Genome-Resolved Metagenomics. *Mar. Drugs* **2023**, *21*, 76. <https://doi.org/10.3390/md21020076>

Academic Editors: Rodrigo Costa and Tina Keller-Costa

Received: 9 December 2022

Revised: 13 January 2023

Accepted: 19 January 2023

Published: 22 January 2023



**Copyright:** © 2023 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/>).

## 1. Introduction

Breakthroughs in sequencing technologies over the last few decades have shed light on the extensive genetic diversity of microbial life and its tremendous wealth of biosynthetic gene clusters. Cyanobacteria, especially filamentous types, have proven to be a rich source of secondary metabolites, including antimicrobial and bioactive natural products [1–3]. In some cases, these unique products are toxic and can be produced at levels high enough to be detrimental to humans, pets, and wildlife such as during harmful algal blooms of the cyanobacteria *Microcystis*. Recent metagenomic and metatranscriptomic sequencing of *Microcystis* blooms revealed the presence not only of co-occurring toxigenic and non-toxigenic strains, but also strains that harbored partial gene clusters for microcystin that were abundant and expressed during specific successional phases of the bloom [4]. Thus, the genomic revolution is providing new avenues to explore the functional and ecological roles of cryptic diversity within cyanobacteria.

Black band disease (BBD) is arguably the longest-studied coral disease, as it was first identified in the scientific literature in the 1970s [5] and documented in artwork as early as the 1890s [6]. Yet, we still do not fully understand the etiology of this destructive and globally distributed coral disease. The engineer of BBD is the filamentous cyanobacterium *Roseofilum reptotaenium* [7], which forms a dense, polymicrobial mat under which anoxic and sulfidic conditions smother and kill coral tissue [8]. While *Roseofilum* is part of the normal microflora of corals, where it can be found at low levels even in corals unaffected by BBD [9], it is unknown what triggers *Roseofilum* to form BBD mats. However, the natural products formed by *Roseofilum* may play a role in manipulating the microbial communities on the coral surface through quorum sensing [9] or other means.

While investigating natural products from BBD mats in corals, we uncovered a pair of novel compounds related to previously described loekeyolides A and B [10]. Strikingly, these compounds were detected only in BBD cyanobacterial mats collected from the massive starlet coral, *Siderastrea siderea*. Herein, we characterized the cyanobacteria associated with BBD in *S. siderea* corals through chemical ecology, 16S rRNA sequencing, and genome-resolved metagenomics to determine the differences between these cyanobacteria and previously characterized strains of *Roseofilum*.

## 2. Results

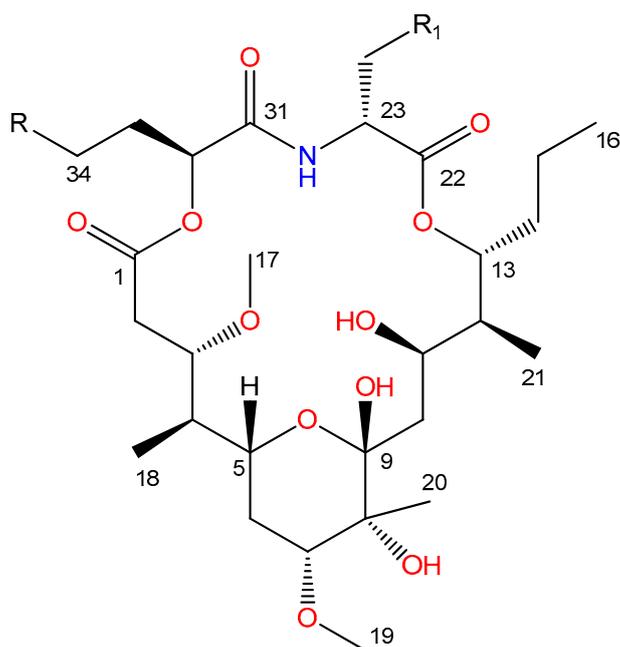
### 2.1. Collection of *S. siderea*-Associated Black Band Disease Cyanobacterial Mats

BBD cyanobacterial mats were sampled from *S. siderea* corals when found during SCUBA diving expeditions in Belize and Florida from 2014 to 2018 (Table S2). These samples were divided among analyses for characterization of major secondary metabolites, characterization of bacterial community composition, and genome-resolved metagenomics. In addition, a non-axenic, cyanobacterial enrichment culture was grown from a BBD mat on a *S. siderea* coral in Florida. The predominant cyanobacterium was likely a *Geitlerinema* strain, as this was the only cyanobacterial genome retrieved from the culture (Table S2).

### 2.2. Isolation and Characterization of Novel Loekeyolides

Previously, we reported two related macrocyclic metabolites, loekeyolide A and loekeyolide B (Figure 1), isolated from the lipophilic extracts of black band disease mats collected from *Montastraea cavernosa*, *Orbicella annularis*, *Orbicella faveolata*, *Pseudodiploria strigosa*, and *Goniopora fruticosa*, and from cultured *Roseofilum reptotaenium* [10]. Loekeyolide A is a 20-membered macrocyclic compound formed by a 16-carbon polyketide chain, 2-deamino-2-hydroxymethionine and D-leucine, and loekeyolide B is its auto-oxidation product at the 2-deamino-2-hydroxymethionine moiety. Interestingly, liquid chromatography-mass spectrometry (LC-MS) analysis of the BBD extracts from a collection of several small samples from the coral *S. siderea* growing in Southwater Caye, Belize, collected in July 2014 indicated the absence of the peaks at  $m/z$  686  $[M + Na]^+$  for loekeyolide A and at  $m/z$  702  $[M + Na]^+$  for loekeyolide B. Instead, we observed peaks at  $m/z$  720  $[M + Na]^+$  for loekeyolide C and 736  $[M + Na]^+$  for loekeyolide D, two new related compounds separated by 16 mass units (Figure 1 and S1) as reported for loekeyolides A and B. Similar results were observed from the lipophilic extract of an August 2018 BBD collection from *S. siderea* growing in Curlew Cay, Belize (Figure S2) and the lipophilic extract of a July 2018 BBD collection from *S. siderea* from Fort Lauderdale, Florida (Figure S3). This mass spectral information prompted us to conduct further chemical investigation of these small samples.

The lipophilic extracts of each batch of freeze-dried sample collections were subjected to reversed phase column chromatography followed by reversed phase high performance liquid chromatography (HPLC) using MeOH-20% water to give loekeyolide D. Although the related compound loekeyolide C was detected in low-resolution electrospray ionization mass spectrometry (LRESIMS) traces, the isolation procedures auto-oxidized it completely to the stable loekeyolide D, and loekeyolide C was not isolated for other spectral studies.



Lookeyolide A (**1**), R = S-CH<sub>3</sub>, R<sub>1</sub> = CH(CH<sub>3</sub>)<sub>2</sub>  
 Lookeyolide B (**2**), R = S(O)-CH<sub>3</sub>, R<sub>1</sub> = CH(CH<sub>3</sub>)<sub>2</sub>  
 Lookeyolide C (**3**), R = S-CH<sub>3</sub>, R<sub>1</sub> = Ph  
 Lookeyolide D (**4**), R = S(O)-CH<sub>3</sub>, R<sub>1</sub> = Ph

**Figure 1.** Chemical structures of previously isolated lookeyolides A and B and the new lookeyolides C and D described in this study.

Lookeyolide D (**4**) was obtained as a white powder;  $[\alpha]^{25}_D 6.0^\circ$  (*c* 0.22, MeOH); UV (MeOH)  $\lambda_{\max}$  (log  $\epsilon$ ) 225 (3.77);  $^1\text{H}$  NMR,  $^{13}\text{C}$  NMR, double quantum filtered correlated spectroscopy (DQF COSY), heteronuclear multiple bond correlation (HMBC), and nuclear Overhauser effect spectroscopy (NOESY) data, see Table 1; high resolution electrospray ionization mass spectrometry (HRESI)/atmospheric-pressure chemical ionization mass spectrometry (APCIMS)  $m/z$  736.3362  $[\text{M} + \text{Na}]^+$  (calcd for C<sub>35</sub>H<sub>55</sub>NO<sub>12</sub>SNa, 702.33337) supported the molecular formula of C<sub>35</sub>H<sub>55</sub>NO<sub>12</sub>S. The  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectral data were indicative of one  $\alpha$ -hydroxy acid, one  $\alpha$ -amino acid, and one highly substituted 16-carbon polyketide chain in the molecule (Table 1). Complete NMR analysis of  $^1\text{H}$ ,  $^{13}\text{C}$ , DQF COSY, edited heteronuclear single quantum coherence spectroscopy (HSQC) and HMBC data for C-22 to C-30 identified this fragment as phenylalanine. The fragment C-31 to C-34 indicated doubling of  $^1\text{H}$  signals as overlapped multiplets for H<sub>2</sub>-34 ( $\delta_{\text{H}}$  2.58, 2.46) and the split methyl signal H<sub>3</sub>-35 ( $\delta_{\text{H}}$  2.508, 2.501). The expected doubling of the  $^{13}\text{C}$  signals for C-31 to C-34 as in lookeyolide B (**2**) were not clearly visible due to weak spectrum. This information, together with the presence of a sulfur atom in the molecular formula, suggested the presence of a hydroxymethionine sulfoxide residue in the molecule as in the related lookeyolide B. The COSY spectrum indicated coupling of hydroxymethine H-32 ( $\delta_{\text{H}}$  4.86,  $\delta_{\text{C}}$  73.6) to methylenes H<sub>2</sub>-33 ( $\delta_{\text{H}}$  1.83,  $\delta_{\text{C}}$  25.4) and then, in turn, to H<sub>2</sub>-34 ( $\delta_{\text{H}}$  2.58, 2.46,  $\delta_{\text{C}}$  48.9). The HMBC spectrum showed correlation of methylene H<sub>2</sub>-34 to methyl carbon signal C-35 ( $\delta_{\text{C}}$  37.3) and the methyl split-singlet-S-35-H<sub>3</sub> ( $\delta_{\text{H}}$  2.508/2.501), in turn, to C-34, ( $\delta_{\text{C}}$  48.9). These data confirmed the presence of a hydroxymethionine sulfoxide residue in the molecule. Following the interpretation of DQF COSY, edited HSQC and  $^{13}\text{C}$  experiments, the remaining  $^1\text{H}$  signals were assignable to two partial structures C-2 to C-7, C-10 to C-16, two O-Me groups (H-17,  $\delta_{\text{H}}$  3.32; H-19,  $\delta_{\text{H}}$  3.38), five hydroxyl methines (H-3,  $\delta_{\text{H}}$  4.34; H-5,  $\delta_{\text{H}}$  3.60; H-7,  $\delta_{\text{H}}$  3.44; H-11,  $\delta_{\text{H}}$  4.70; H-13,  $\delta_{\text{H}}$  4.99), four methyl groups (C-16,  $\delta_{\text{H}}$  0.93,  $\delta_{\text{C}}$  14.3; C-18,  $\delta_{\text{H}}$  0.81,  $\delta_{\text{C}}$  9.7; C-20,  $\delta_{\text{H}}$  1.25,  $\delta_{\text{C}}$  19.6; C-21,  $\delta_{\text{H}}$  0.98,  $\delta_{\text{C}}$  11.9),

two quaternary carbons (C-8,  $\delta_C$  74.7; C-9,  $\delta_C$  102.0) and one ester carbonyl group (C-1,  $\delta_C$  173.4) (Table 1). HMBC correlations from H-3 ( $\delta_H$  4.34) and H<sub>2</sub>-2 ( $\delta_H$  2.52 and 2.36) to C-1 ( $\delta_C$  173.4) connected the remaining ester carbonyl. The weak HMBC data did not show correlation from H<sub>3</sub>-17 ( $\delta_H$  3.32) to C-3 ( $\delta_C$  77.6) and H<sub>3</sub>-19 ( $\delta_H$  3.38) to C-7 ( $\delta_C$  79.5) but the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift values, similar to those reported for **2** and the NOE data observed between H-4 and H-17 and H-20 and H-19, connected the two OMe groups to the C-3 and C-7 positions. HMBC correlations from H<sub>3</sub>-18 methyl ( $\delta_H$  0.81) to C-3 ( $\delta_C$  77.6) and to C-5 ( $\delta_C$  70.8) connected this methyl group to C-4 position. These data connected the methyl group, two OMe groups and the carbonyl group to the C-2 to C-7 partial structure. HMBC correlations indicated H<sub>3</sub>-20 ( $\delta_H$  1.25) to C-7 ( $\delta_C$  80.0), C-8 ( $\delta_C$  74.7), and C-9 ( $\delta_C$  102.0), H<sub>2</sub>-6 ( $\delta_H$  2.08, 1.42) to C-5 ( $\delta_C$  70.8) and C-7 ( $\delta_C$  80.0), H-10a ( $\delta_H$  1.90) to C-11 ( $\delta_C$  66.7), H-10b ( $\delta_H$  1.69) to C-9 ( $\delta_C$  102.0). This information connected the two partial structures C-2 to C-7 and C-10 to C-16 and thus established the planar structure for the substituted 16-carbon polyketide chain. Comparison of  $^{13}\text{C}$  chemical shift values for C-5 ( $\delta_C$  70.8) and C-9 ( $\delta_C$  102.0) with that of loekeyolides A and B, and the degrees of unsaturation calculated from the molecular formula suggested the presence of a pyrano ring system within the 16-carbon polyketide acid chain. HMBC correlation between the H-13 of 16-carbon polyketide chain and C-22 carbonyl of phenylalanine, and H-32 ( $\delta_H$  4.86) of 2-deamino-2-hydroxymethionine and C-1 ( $\delta_C$  173.4) carbonyl of 16-carbon polyketide chain were not observed due to paucity of material available for the experiment. The macrocyclic ring system was established in relation to loekeyolides (**1**, **2**) and thus satisfied the molecular formula. These data establish the planar macrocyclic structure for loekeyolide D (**4**). The 2D NOE data and coupling constants for the chiral centers (Table 1) of **4** observed in CD<sub>3</sub>OD were compared with the reported data for **1** in CD<sub>3</sub>OD and the data were comparable. The key NOE data observed between H-4 and H-17, H-5 and H-18, H-7 and H-20, H-10a and H-20, H-10a and H-21, H-13 and H-21, and H-11 and H-14 and the comparable coupling constants observed for chiral centers H-3 (10.3, 4.1 Hz), H-5 (11.0, 11.0, 2.0 Hz), H-7 (12.0, 4.8 Hz), 11-H (11.0, 1.7, 1.0 Hz), H-12 (7.0, 1.0, 1.0 Hz) and H-13 (9.6, 6.1, 1.0 Hz) suggested the same relative stereochemistry for the chiral centers in substituted 16-carbon polyketide moiety in **4**.

Next, the absolute configuration of the amino acid phenylalanine and the hydroxymethionine sulfoxide unit was determined. Compound **4** (0.1 mg) was dissolved in EtOH (0.1 mL) and treated with an excess of fresh Raney-Ni (2400) as a slurry in H<sub>2</sub>O (0.1 mL) and refluxed for 0.5 h. The product was filtered and concentrated to give a white solid. Desulfurization with Raney-Ni gave *des*-thiomethylloekeyolide D. The desulfurized *des*-thiomethyl product [HRESI/APCIMS  $m/z$  674.3539 [M + Na]<sup>+</sup> (calcd for C<sub>34</sub>H<sub>53</sub>NO<sub>11</sub>Na, 674.3511)] was suspended in 6 N HCl (0.1 mL) and heated at 115 °C for 14 h in a sealed tube. The hydrolysate was concentrated to dryness. The residue was reconstituted in 0.1 mL of H<sub>2</sub>O and analyzed by enantioselective HPLC, comparing the retention times with those of authentic standards [Phenomenex Chirex (D) Penicillamine, 4.6 mm × 250 mm, 5 μm]; detection at 254 nm. Using the solvent mixture of 2.0 mM CuSO<sub>4</sub>-MeCN (90:10), with a flow rate of 1.0 mL/min, the retention times ( $t_R$  min) for authentic standards were L-Phe (70.4) and D-Phe (73.3). The  $t_R$  min of the amino acid in the hydrolysate under the same condition was 73.3, indicating the presence of D-Phe in the hydrolysate. The stereochemistry of the  $\alpha$ -hydroxy acid was determined using a different chiral column for the HPLC analysis [CHIRALPAK MA (+) (4.6 mm × 50 mm), Diacel Chemical industries, Ltd.; solvent, 2.0 mM CuSO<sub>4</sub>-MeCN (95:5); flow rate, 1.0 mL/min; detection at 254 nm]. The  $t_R$  min for authentic standards were *R*-Hba (9.3) and *S*-Hba (13.3). The retention time of the  $\alpha$ -hydroxy acid in the hydrolysate under these conditions was (13.3), indicating the presence of *S*-Hba in the hydrolysate. These data established the absolute configurations for the phenylalanine and hydroxy acid in loekeyolide D (**4**). Loekeyolide D (**4**) is the auto-oxidized product of the 2-deamino-2-hydroxy methionine in loekeyolide C (**3**), and therefore loekeyolide C (**3**) and loekeyolide D (**4**) have the same chiral centers, thus establishing their structures. HRMS data and stereochemical studies are described in Supplementary Materials.

**Table 1.** NMR spectroscopic data for looekeyolide D (4) in CD<sub>3</sub>OD (600 MHz).

Position	$\delta_C$ Mult.	$\delta_H$ (J in Hz)	COSY <sup>a</sup>	HMBC	NOESY <sup>b</sup>
1	173.4, C			2a	
2a	33.6, CH <sub>2</sub>	2.52, d (12.0)	2b,		2b, 3, 5
2b		2.36, dd (12.0, 10.3)	2a, 3		2a, 3
3	77.6, CH	4.34, dd (10.3, 4.1)	2b, 4	2, 18	2a, 2b, 4, 11
4	38.7, CH	2.21, m	3, 5, 18	6b, 18	3, 17, 18
5	70.8, CH	3.60, ddd (11.0, 11.0, 2.0)	4, 6a, 6b	6b, 18	2a, 6a, 18
6a	31.4, CH <sub>2</sub>	2.08, ddd (12.1, 11.0, 4.8)	5, 6b, 7		5, 6b, 7, 18
6b		1.42, ddd (12.1, 12.0, 1.8)	5, 6a,		6a
7	80.0, CH	3.44, dd (12.0, 4.8)	6a, 6b	6a, 6b, 20	6a, 10a, 20
8	74.7, C			6a, 6b, 20	
9	102.0, C			10b, 20	
10a	37.9, CH <sub>2</sub>	1.90, ddd (12.3, 11.0)	10b, 11		7, 10b, 11, 20, 21
10b		1.69, dd (12.3, 2.7)	10a, 11		10a, 11, 12
11	66.7, CH	4.70, ddd (11.0, 1.7, 1.0)	10a, 10b, 12	10a, 12, 21	3, 10ab, 14a, 14b
12	42.6, CH	1.52, ddq (7.1, 1.0, 1.0)	11, 13, 21	21	10b, 13, 21
13	81.4, CH	4.99, ddd (9.6, 6.1, 1.0)	12, 14a, 14b	21	12, 21
14a	35.4, CH <sub>2</sub>	2.15, m	13, 14b, 15a, 15b	12, 16	14b, 15b
14b		1.59, m	13, 14a, 15a, 15b	14a	
15a	20.3, CH <sub>2</sub>	1.36, m	14a, 14b, 15b, 16	16	15b, 16
15b		1.29, m	14a, 14b, 15a, 16		15a, 16
16	14.3, CH <sub>3</sub>	0.93, t (7.5)	15a, 15b		15a, 15b
17	57.4, OCH <sub>3</sub>	3.32, s			4
18	9.7, CH <sub>3</sub>	0.81, d (6.9)	4		4, 5, 6a
19	57.7, OCH <sub>3</sub>	3.38, s			20
20	19.6, CH <sub>3</sub>	1.25, s		7	7, 10a, 19
21	11.9, CH <sub>3</sub>	0.98, d (6.8)	12		12, 13, 10a
22	175.1, C				
23	53.6, CH	4.81, dd (10.8, 4.2)	24a, 24b	24a, 24b	
24a	36.0, CH <sub>2</sub>	3.39, m	23, 24b	26, 30	24b
24b		2.82, m	23, 24a		24a
25	138.9, C			27, 29	
26	130.0, CH	7.23, d (6.8)	27	28	27
27	129.7, CH	7.27, t (6.8)	26, 28		26, 28
28	127.9, CH	7.19, t (6.8)	27, 29	26, 30	27, 29
29	129.7, CH	7.27, t (6.8)	28, 30		28, 30
30	130.0, CH	7.23, d (6.8)	29	28	29
31	171.7, C				
32	73.6, CH	4.86, dd (8.9, 4.0)	33	33	33a, 33b
33	25.4, CH <sub>2</sub>	1.83, m	32, 34a, 34b	32, 34	32, 34
34a	48.9, CH <sub>2</sub>	2.58, m	33, 34b	33, 35	34b
34b		2.46, m	33, 34a		34a
35	37.3, CH <sub>3</sub>	2.508, 2.501, s		34	

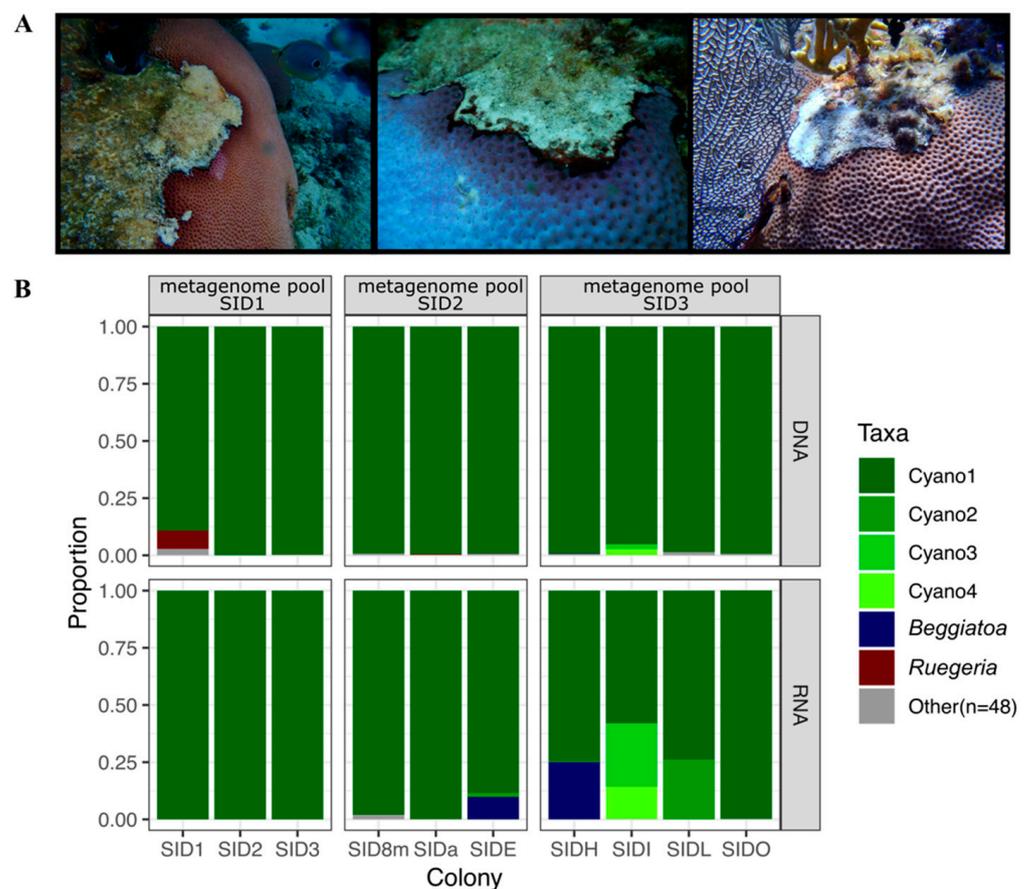
<sup>a</sup> COSY and NOESY correlations are from proton(s) stated to the indicated protons. <sup>b</sup> HMBC correlations are from proton(s) stated to the indicated carbons.

### 2.3. Bacterial Composition of Belize Samples

Bacterial community composition was characterized for both the RNA and DNA fractions of ten Belize *S. siderea* colonies with black band disease (BBD). A total of 1,815,043 16S rRNA gene amplicon sequencing reads passed quality-filtering, with an average of 90,752 reads per sample (min 2973, max 295,333) (Table S1). Only 54 amplicon sequence variants (ASVs) were detected in the twenty libraries (10 RNA, 10 DNA), with just six prevalent ASVs (Table 2, Figure 2B).

**Table 2.** Six predominant Amplicon Sequence Variants (ASVs) in V6 amplicon libraries from black band disease (BBD) cyanobacterial mats from *S. siderea* corals in Belize. For each ASV, the SILVA classification is provided, as well as the closest BLAST match in GenBank for comparison. Accession numbers in bold indicate sequences originating from previous BBD studies.

ASV Name	SILVA Classification	Closest BLAST Match (% Similarity for V6 Region)
Cyano 1	<i>Roseofilum</i> AO1-A	<b>EF123646</b> (100%)
Cyano 2	<i>Geitlerinema</i> PCC-7105	<b>EF110974</b> (100%), <b>EF372580</b> (100%)
Cyano 3	<i>Hormoscilla</i> SI04-45	KY697267 (100%)
Cyano 4	<i>Hormoscilla</i> SI04-45	KY697265 (100%)
<i>Beggiatoa</i>	unclassified genus of Bacteria	<b>KM924160</b> (96.7%)
<i>Ruegeria</i>	<i>Ruegeria</i>	MT484146 (100%)



**Figure 2.** Characterization of black band disease cyanobacterial mats from *Siderastrea siderea* corals. (A) Representative photographs of *S. siderea* colonies with black band disease cyanobacterial mats in Belize and Florida. (B) Proportions of the V6 region of the 16S rRNA gene amplicon sequence variants in the DNA and RNA fractions of ten cyanobacterial mats from *S. siderea* colonies with black band disease. The predominant amplicon sequence variant called “Cyano1” belongs to the genus *Roseofilum*. Samples were subsequently pooled to ensure sufficient DNA for metagenomic sequencing, as indicated by the clustering labels on the top of the figure.

The most prevalent ASV, hereafter referred to as Cyano1, ranged from 58 to 100% relative abundance per sample and was classified as the cyanobacterial genus *Roseofilum*. This was the only ASV detected in the RNA fraction of SID1, the sample with the highest number of sequencing reads (295,333). Five additional prevalent ASVs had relative abundances of less than 30% per sample. These included three additional cyanobacterial ASVs, one ASV classified as *Ruegeria*, and one ASV that was classified only as Bacteria with the

SILVA database but was identified as *Beggiatoa* through a BLASTN search (Table 2). While the Cyano1 ASV was classified as “*Roseofilum* AO1-A” with SILVA, BLASTN searches of these sequences revealed that they share only 95% sequence similarity in the 60-bp V6 region of the 16S rRNA gene with the *Roseofilum* strain AO1-A (KU579397), isolated from the Great Barrier Reef [11], or with *Roseofilum* strain Cy1 (KP689103), isolated from the Florida Reef Tract [12]. Instead, the Cyano1 (*Roseofilum*) ASV was an exact match to several clone library sequences (EF123634, EF123639, EF123644, EF123645, EF123646) that were previously detected in BBD cyanobacterial mats from Caribbean *S. siderea* corals [13].

The Cyano2 (*Geitlerinema*) ASV was an exact match to several clone library sequences (EF110974, EF154084, DQ644020, DQ680351, DQ151461) that were previously detected in BBD cyanobacterial mats from Caribbean *S. siderea* corals [14,15] and to sequences originating from Caribbean stromatolites (EU917946, EU917928, EU917912, EU917848, EU917831, EU917822, EU917819, EU917807) [16]. Two additional cyanobacterial ASVs, Cyano 3 and Cyano 4, were identified as *Hormosilla* (Table 2). The *Beggiatoa* ASV shared 96.7% sequence similarity with *Beggiatoa* isolated from BBD on *Montipora capitata* corals in Hawaii [17]. *Ruegeria* strains have also previously been detected in BBD studies [18], including BBD cyanobacterial mats from *S. siderea* corals [19].

#### 2.4. Metagenome-Assembled Genomes of Cyanobacteria from Belize and Florida

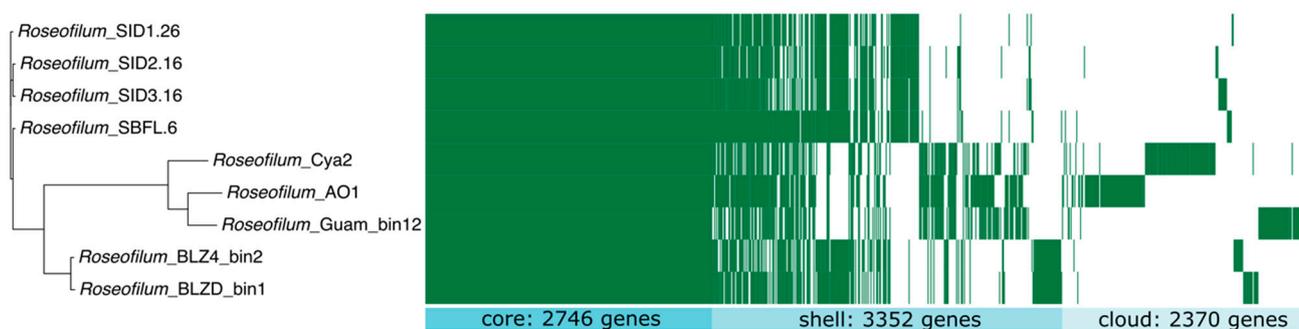
Four of the Belize *S. siderea* colonies (SIDH, SIDI, SIDL, and SIDO) produced detectable levels of loekeyolide C/D by LC-MS, while six of the colonies (SID1, SID2, SID3, SID8m, SIDa, and SIDE) did not. However, the same predominant cyanobacterial ASV (Cyano 1) was found in both DNA and RNA fractions of BBD from all Belize *S. siderea* colonies. The variability in detection of loekeyolides is likely due to the small sample sizes collected and small amounts of chemical extracts. To confirm the genetic potential for the biosynthesis of loekeyolides, we compared metagenome-assembled genomes (MAGs) of cyanobacteria from pooled metagenomes of producer or non-producer samples (as described in the methods) from Belize as well as producer samples from Florida.

Quality-filtered metagenomic sequencing reads ranged from roughly 13 million to 33 million per metagenomic library (Table S2). Seven cyanobacterial MAGs with >90% completeness and <6% contamination were retrieved from all six metagenome libraries (Table S3). *Roseofilum* MAGs were retrieved from all three Belize metagenomes and one Florida metagenome, encompassing both producers and non-producers of loekeyolide C. In addition, cyanobacterial MAGs that do not belong to the genus *Roseofilum* were retrieved from one Belize metagenome and two Florida metagenomes (Table S3). Among the four *Roseofilum* MAGs from *S. siderea* corals, the average nucleotide identity (ANI) of shared genes was >99%, while the ANI of the *Roseofilum* MAGs compared to non-*Roseofilum* cyanobacteria from *S. siderea* corals was too close to the detection limit for accuracy (<75%), suggesting they belong to different genera (Table S4). *Roseofilum* MAGs from *S. siderea* corals had >98% ANI with *Roseofilum* MAGs retrieved from other Caribbean coral species and had >94% ANI with *Roseofilum* MAGs retrieved from Pacific coral species (Table S4). Of the non-*Roseofilum* MAGs from *S. siderea* corals, SID2\_20 and SBC9 had >99% ANI of shared genes with each other and with the *Geitlerinema* BBD 1991 MAG from Caribbean *Montastraea cavernosa* [20,21]. All three strains, SID2\_20, SBC9, and BBD 1991 were classified as *Geitlerinema* species by GTDBtk. The non-*Roseofilum* MAG SBLK1 was too close to the detection limit for accuracy (<75%) from SID2\_20, SBC9, and *Geitlerinema* BBD 1991. SBLK1 was classified to the cyanobacterial family *Spirulinaceae* by GTDB-Tk. The 16S rRNA gene was not detected in the *Spirulinaceae* bacterium SBLK1 for further taxonomic identification. The presence of three distinct cyanobacterial genera within the order *Oscillatoriales* in the MAGs was consistent with the presence of three cyanobacterial genera (*Roseofilum*, *Geitlerinema*, and *Hormosilla*) in the 16S rRNA amplicon libraries, although not an exact match for all genera. As the *Roseofilum* strains were the most predominant cyanobacteria within BBD mats from Belize and previous studies from both

the Caribbean and Pacific [9,22,23], we focused the pangenome analysis primarily on the *Roseofilum* genomes.

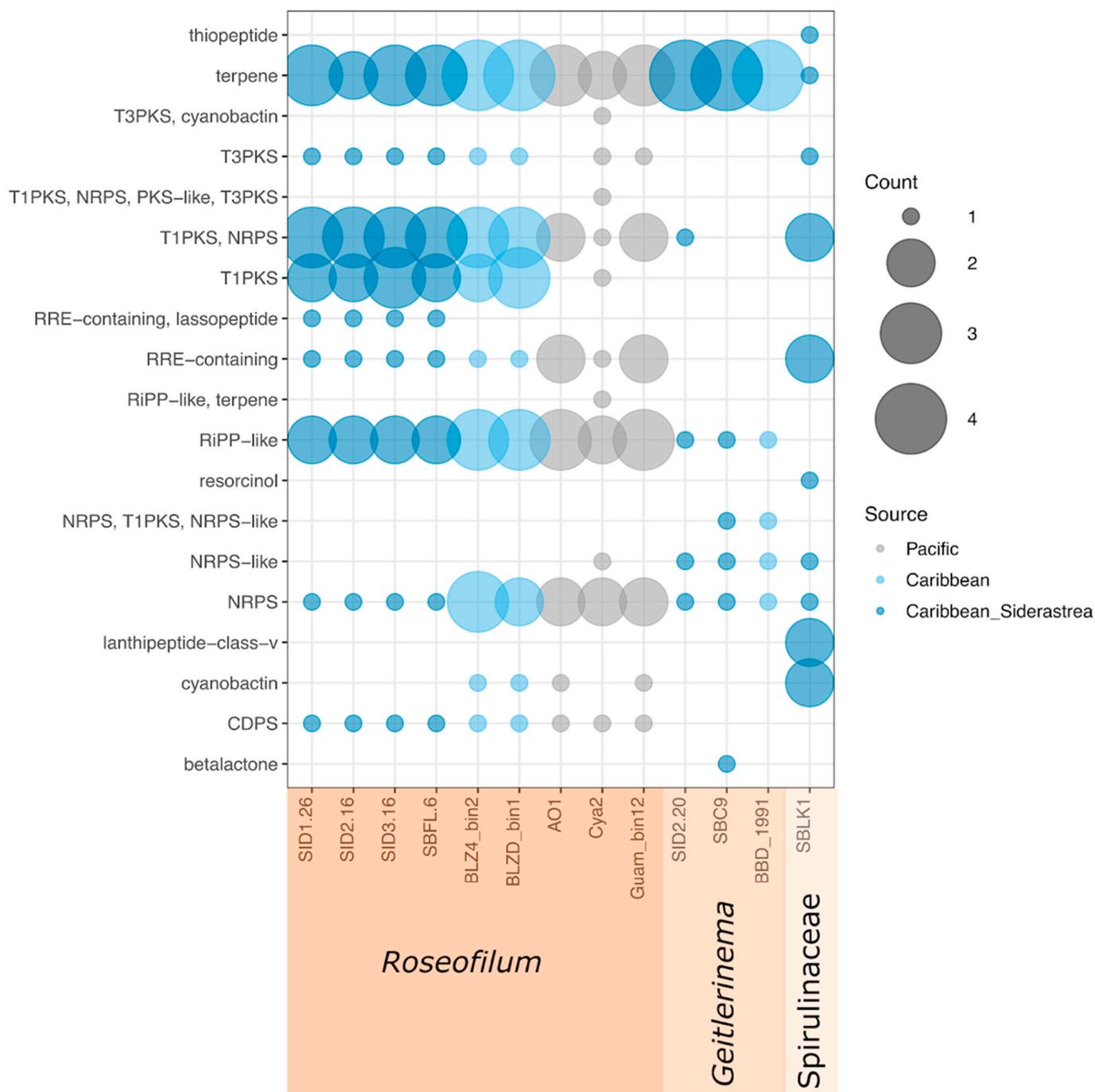
### 2.5. Comparative Genomics of Black Band Disease-Associated Cyanobacteria

Comparative genomics of nine *Roseofilum* MAGs included four *S. siderea*-associated *Roseofilum* MAGs from this study and five *Roseofilum* MAGs from other coral species [12,22,23]. All nine MAGs passed the quality threshold of >90% completeness and <6% contamination (Table S3). Pangenome analysis identified 2746 core genes found in all nine *Roseofilum* genomes, 3352 shell genes in two to eight *Roseofilum* genomes, and 2370 cloud genes found in only one *Roseofilum* genome (Figure 3). Three distinct clusters of genomes were detected: Caribbean *Roseofilum* from four *Siderastrea* corals, three Pacific *Roseofilum* strains, and Caribbean *Roseofilum* from two other boulder corals. A total of 294 genes were found in all 4 *Roseofilum* genomes from *S. siderea* but not in any other *Roseofilum* genomes. Of these, 225 (77%) were annotated as hypothetical proteins, while only 69 (23%) had functional annotations.



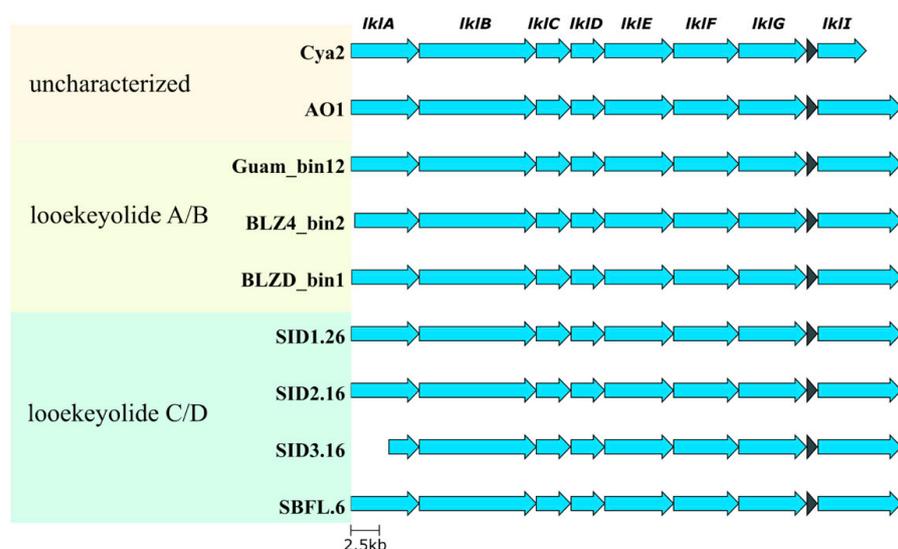
**Figure 3.** Pangenome comparison of nine *Roseofilum* metagenome-assembled genomes from this study and previously published works. Clustering of genomes is based on the alignment of 2746 core genes present in all nine strains. Shell genes were present in two to eight genomes and cloud genes were present in only one genome.

Each of the nine *Roseofilum* MAGs had 14 to 19 biosynthetic gene clusters identified by antiSMASH, including multiple clusters for terpenes, ribosomally synthesized and post-translationally modified (RiPP)-like clusters as well as RiPP recognition elements (RRE), Type I polyketide synthases (T1PKSs), nonribosomal peptide synthetases (NRPSs), and hybrid T1PKS/NRPS clusters (Figure 4). Each of the nine *Roseofilum* MAGs had one biosynthetic gene cluster for tRNA-dependent cyclodipeptide synthase (CDPS), which has been more commonly found in the genomes of *Actinobacteria*, *Firmicutes*, and *Proteobacteria* [24,25]. The *Spirulinaceae* MAG also had 14 detectable biosynthetic gene clusters, including gene clusters for antimicrobial lanthipeptides, thiopeptide, and cyanobactins, as well as resorcinol (Figure 4). In contrast to the *Roseofilum* and *Spirulinaceae* MAGs, only eight or nine biosynthetic gene clusters per genome were detected in *Geitlerinema* MAGs.



**Figure 4.** Comparison of biosynthetic gene clusters in thirteen cyanobacterial metagenome-assembled genomes from this study and previously published works. Counts indicate the number of biosynthetic gene clusters rather than individual genes. Lookeyolide and malynamide clusters are counted as “TIPKS/NRPS”.

The putative biosynthetic gene clusters for lookeyolides are classified as hybrid T1PKS/NRPS clusters and complete biosynthetic gene clusters were retrieved from all nine *Roseofilum* MAGs (Figure 5), meaning both the producer and non-producer samples had the genetic potential to make lookeyolides, although the level of biosynthesis of the lookeyolides may vary among samples. In addition, two Pacific strains [22,23] appear to have the genetic capacity to produce lookeyolides, but their natural products have not been elucidated.



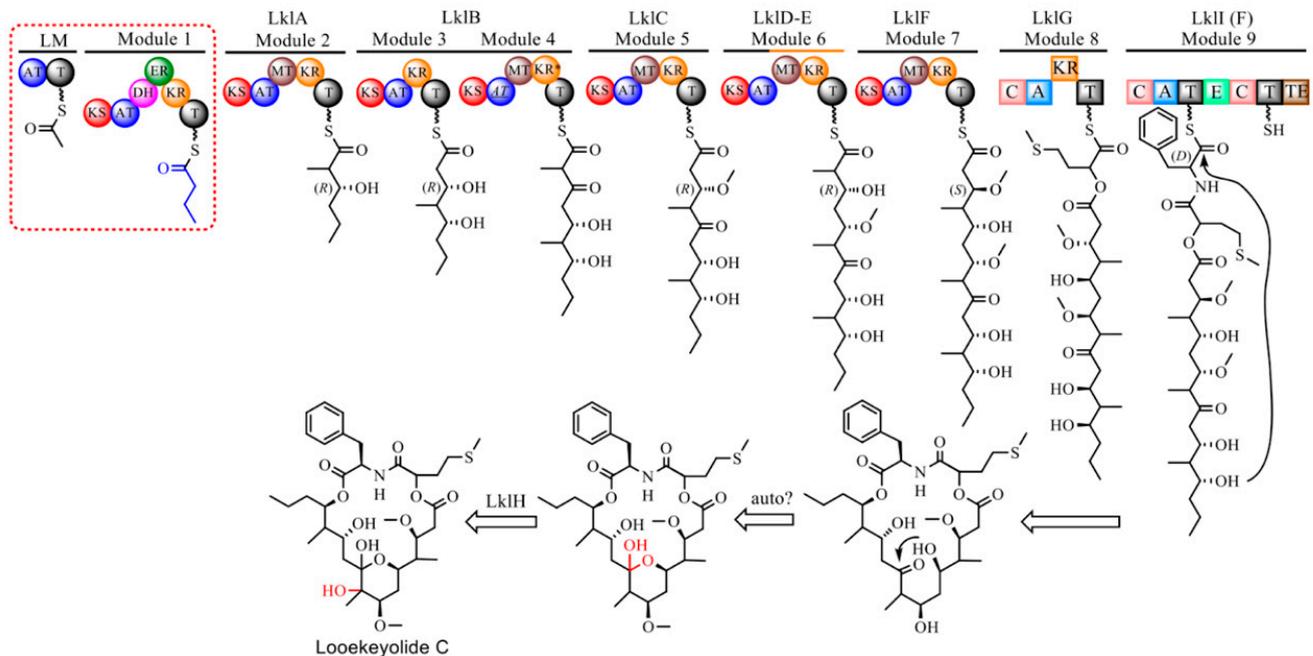
**Figure 5.** Looekeyolide biosynthetic gene clusters among nine *Roseofilum* metagenome-assembled genomes from this study and previously published works. Natural products have not been characterized from two Pacific *Roseofilum* strains (Cya2, AO1). Looekeyolides A/B have been detected from both Caribbean and Pacific strains (Guam\_bin12, BLZ4\_bin2, BLZD4\_bin1). Looekeyolides C/D have only been detected in Caribbean *Roseofilum* from *S. siderea* corals (SID1.26, SID2.16, SID3.16, SBFL.6).

A putative biosynthetic pathway for looekeyolide C/D is proposed (Figure 6), with high similarity to the pathway for looekeyolide A/B previously described [10] in *Roseofilum* MAGs from corals other than *S. siderea*. The adenylation domain of LkII in *Roseofilum* MAGs from *S. siderea* that produce looekeyolide C/D had specificity for L-phenylalanine, while *Roseofilum* MAGs from corals other than *S. siderea* that produce looekeyolide A/B had specificity for L-leucine (Table S5, Figure S11). Most of the looekeyolide biosynthetic genes in Caribbean *Roseofilum* from multiple coral species (*Orbicella annularis*, *Pseudodiploria strigosa*, *Montastraea cavernosa*) were 97% to 99% similar to the genes in Caribbean *Roseofilum* from *S. siderea* except for LkII which was 92% similar due to the low identity (46%) of their A domains (Figure S12). Hybrid T1PKS/NRPS biosynthetic gene clusters predicted to produce looekeyolides were not detected in *Geitlerinema* or *Spirulinaceae* MAGs.

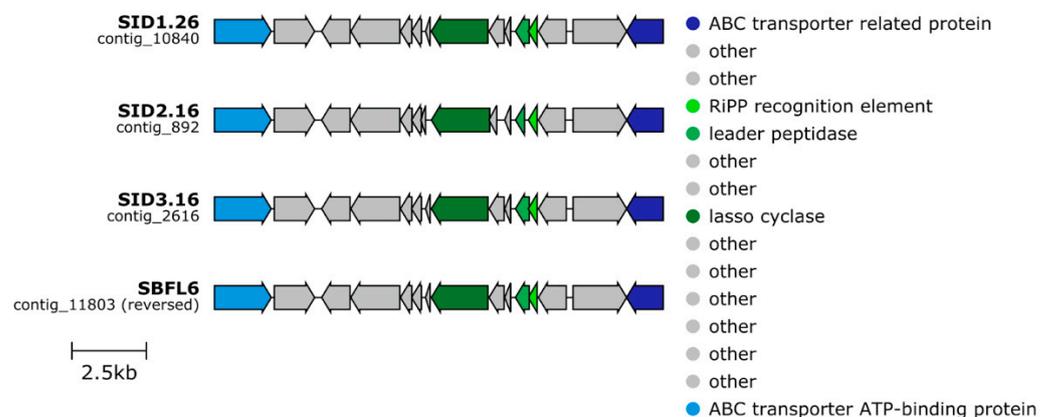
Each of the nine *Roseofilum* MAGs had hybrid T1PKS/NRPS clusters annotated as malyngamides including malyngamide C acetate and malyngamide I (Figure S13). Malyngamides are small amides, many of which have lyngbic acid as a carboxylic acid side chain. Both malyngamides and lyngbic acid from Caribbean filamentous cyanobacteria, including *Roseofilum*, have previously been shown to interfere with bacterial quorum sensing [9,26]. Most of the nine *Roseofilum* MAGs had multiple putative malyngamide biosynthetic gene clusters, and no clear patterns were observed that corresponded to differences in gene clusters among coral hosts or geographic locations (Figure S13). Hybrid T1PKS/NRPS biosynthetic gene clusters predicted to produce malyngamides were not detected in *Geitlerinema* or *Spirulinaceae* MAGs.

Analysis of biosynthetic gene clusters also revealed that the *Roseofilum* MAGs from *S. siderea* corals had one type of biosynthetic gene cluster that was not found in the other *Roseofilum* strains. All four *Roseofilum* MAGs from *S. siderea* corals had a lasso peptide biosynthetic gene cluster that encoded a 98 aa stand-alone RiPP recognition element (RRE), a 135 aa lasso peptide transglutaminase homolog (leader peptidase, capB), and a 633 aa lasso peptide asparagine synthase homolog (lasso cyclase, capC). These biosynthetic genes were flanked on each side by genes for ABC-transporter related genes (Figure 7). While the amino acid sequences in the RiPP recognition element and the capB leader peptidase were identical in all four *Roseofilum* MAGs from *S. siderea* corals, the amino acid sequences for the capC lasso cyclase in SID1.26 and SBFL6 differed by 4 amino acids from SID2.16 and SID3.16. A search with blastp of the *S. siderea*-associated *Roseofilum* lasso cyclase amino

acid sequence (from SID2.16) showed low similarity ( $\leq 67\%$  similarity) to homologues in other cyanobacterial genomes.



**Figure 6.** Proposed biosynthetic pathway for lookeyolide C in Caribbean *Roseofilum* metagenome-assembled genomes from *S. siderea* corals.



**Figure 7.** Lasso peptide biosynthetic gene clusters detected only in Caribbean *Roseofilum* metagenome-assembled genomes from *S. siderea* corals.

### 3. Discussion

*Roseofilum reptotaenium*, the cyanobacterial engineer of black band disease (BBD) in corals [7], is found in tropical coral reefs around the world and impacts at least 72 coral species [27]. Here, we uncovered cryptic diversity among *Roseofilum* strains through both chemical and genomic analyses. The sequence-based threshold of 95% ANI has been proposed as the delineation of bacterial species [28–31]. Using this metric, all six *Roseofilum* strains from the Caribbean are the same cyanobacterial species regardless of the host coral species ( $>98\%$  ANI), while the three Pacific *Roseofilum* strains were very close to this threshold (94.25–94.68% ANI) and thus, potentially represent a separate species. However, using ANI for comparison only reveals the similarity among shared genes and does not capture differences in gene content, i.e., when genes are present in one strain and absent in another. *Roseofilum* strains on *S. siderea* corals were both chemically and genetically

distinct from other strains in the Caribbean despite belonging to the same cyanobacterial species. This difference was consistent across sites in Belize and Florida and through time, as samples were collected in 2014, 2015, and 2018. Of note, surveys using only 16S rRNA amplicons would not be able to distinguish among these distinct strains of *Roseofilum*, thus highlighting the utility of metagenomic sequencing in uncovering the functional differences among visually similar filamentous cyanobacteria in reef ecosystems.

Over 2000 metabolites have been described from *Cyanobacteria* [32]. Some of the ecological roles of these natural products include grazing deterrents, allelopathy, iron scavenging, UV protection, and signaling [33]. Comparative analysis of nine *Roseofilum* genomes showed that each of the *Roseofilum* genomes had multiple terpene, ribosomal and non-ribosomal peptide, and polyketide biosynthetic gene clusters. Each of these classes of cyanobacterial natural products includes potential antibacterial or antiviral compounds [34–36]. In fact, every type of biosynthetic cluster detected in these BBD-associated cyanobacterial genomes, regardless of genus, includes natural products that exhibit antimicrobial properties. These antimicrobial agents may play a role in the progression of BBD by allowing the cyanobacteria to outcompete other coral-associated microorganisms that would normally suppress pathogen growth.

All *Roseofilum* genomes examined here had hybrid peptide/polyketide biosynthetic gene clusters proposed to encode for the cyclic depsipeptide loekeyolides, the lipopeptide malyngamides, and a tRNA-dependent cyclodipeptide that were not found in four non-*Roseofilum* BBD-associated cyanobacterial genomes. Our previous work demonstrated that loekeyolides from *Roseofilum* under laboratory conditions do not alter growth and biofilm formation by marine bacteria, do not act as siderophores, and do not impact photosynthetic performance of the coral [10]. The oxygen sensitive loekeyolide A reduces hydrogen peroxide levels, suggesting a role in combating reactive oxygen species on the coral surface [10]. Malyngamides from filamentous cyanobacteria have demonstrated both cytotoxic and anticancer properties [37] and antibacterial properties against Gram positive pathogens [38]. Malyngamide C and lyngbic acid have also demonstrated quorum-sensing inhibition in marine bacteria [9,26]. In addition, the tRNA-dependent cyclodipeptides have variously shown antibacterial, antifungal, antiviral, and antitumor properties [39]. In contrast to loekeyolides and malyngamides, only the *S. sidera*-associated *Roseofilum* genomes contained biosynthetic gene clusters for lasso peptides. Lasso peptides are underexplored in Cyanobacteria [40]. Characterized lasso peptides have demonstrated a variety of activities including antimicrobial properties, and the unique lasso structure imparts heat and chemical resistance [41].

The biosynthesis of malyngamides and lasso peptides has been well characterized [42,43], setting the stage for their heterologous production and bioactivity investigation. In addition to the proposed biosynthetic pathway for loekeyolide C presented here, we recently proposed a pathway for loekeyolide A [10]. With a cultivated strain of *Roseofilum* that produces loekeyolide A [10] and the genome sequences for multiple, unique *Roseofilum* strains, we are poised for future studies to uncover the bioactivity of these natural products and their potential use for novel applications.

Collectively, *Roseofilum* genomes associated with BBD from locations in the Caribbean and the Pacific share a wide assortment of peptide and polyketide natural products that may have bioactive properties. The exact roles of loekeyolides, malyngamides, and other secondary metabolites are not known, but the conserved nature of these compounds implies they play an important role in the ecology of these cyanobacteria and may also contribute to disease etiology through manipulation of the microbial communities around them.

## 4. Materials and Methods

### 4.1. Sample Collection and Enrichment Culturing

Black band disease (BBD) cyanobacterial mats were collected from *Siderastrea sidera* corals in Belize and Florida by aspiration with a needleless syringe for both chemical analysis and extraction of nucleic acids. BBD mats from several colonies of *Siderastrea*

were combined for bulk analysis in three batches: one from South Water Caye, Belize in July 2014, one from Curlew Cay, Belize in August 2018, and one from Fort Lauderdale, Florida in July 2018. For microbiome analysis, relatively thin BBD mats (Figure 1) from ten colonies of *S. siderea* were sampled while SCUBA diving in September 2015 at Carrie Bow Cay, Curlew Cay, or South Water Channel near the Smithsonian Carrie Bow Cay Field Station in Belize. One additional *S. siderea* coral exhibiting BBD was sampled at Looe Key in the Florida Keys National Marine Sanctuary in July 2017. Finally, a BBD mat was collected from a *S. siderea* coral offshore from Ft. Lauderdale, FL in July 2018. A non-axenic, cyanobacterial enrichment culture of the BBD mat from Ft. Lauderdale, FL was grown in artificial seawater amended with Cyanobacterial BG-11 media (ATCC medium 616) as previously described [12].

#### 4.2. Characterization of Major Secondary Metabolites

Bulk cyanobacterial mats of the 2014 collection were freeze-dried and extracted repeatedly with MeOH. Similarly, the 2018 collection was freeze-dried and extracted with 50% EtOAc-50% MeOH saturated with helium gas. The extracts were chromatographed on a column of C<sub>18</sub> (3 g) using a MeOH-H<sub>2</sub>O step gradient system to give five sub-fractions. The sub-fraction 3 (0.002 g), eluted with 80% MeOH-20% H<sub>2</sub>O was further separated by reversed-phase HPLC (semi-prep 250 mm × 10 mm, 5 μm, RP-18, flow 3.0 mL/min) using 80% MeOH-20% H<sub>2</sub>O to give 0.6 mg of looekeyolide D ( $t_R = 10.3$  min, yield, 0.03% dry wt) (July 2014 batch) and 0.3 mg of looekeyolide D ( $t_R = 10.3$  min, yield, 0.06% dry wt) (August 2018 batch). Looekeyolide C was not isolated and assumed to be completely oxidized during the isolation process.

Optical rotations were recorded on a Jasco P2000 polarimeter. UV spectrophotometric data was acquired on a Shimadzu PharmaSpec UV-visible spectrophotometer. NMR data were collected on a JEOL ECA-600 spectrometer operating at 600.17 MHz for <sup>1</sup>H and 150.9 MHz for <sup>13</sup>C. <sup>1</sup>H NMR chemical shifts (referenced to residual CD<sub>3</sub>OD at δ 3.30) were assigned using a combination of data from 2D DQF COSY and multiplicity-edited HSQC experiments. The edited-HSQC experiment was optimized for  $J_{CH} = 140$  Hz and the HMBC experiment was optimized for  ${}^{2/3}J_{CH} = 8$  Hz. <sup>13</sup>C NMR chemical shifts (referenced to CD<sub>3</sub>OD observed at δ 49.0) were assigned on the basis of multiplicity-edited HSQC experiments. Low resolution liquid chromatography mass spectrometry (LRLC-MS) was performed on a Thermo Scientific (Waltham, MA, USA) LTQ LC-MS ESI instrument connected to a Grace Vydac Reversed-phase column (C18, 218TP, 5 μ, 100 mm × 2.1 mm) using a mixture of 0.1% HCOOH in water (A) and 0.1% HCOOH in CH<sub>3</sub>CN (B) at a rate of 0.2 mL/min. The gradient system used was 90% A to 0% A in 15 min followed by 100% B for the next 10 min. HRMS data was obtained using an Agilent 6210 LC-TOF mass spectrometer equipped with an APCI/ESI multimode ion source detector at the Mass Spectrometer Facility at the University of California, Riverside, California. Varian BondElut octadecyl (C<sub>18</sub>) was used for column chromatography. All solvents used were of HPLC grade (Fisher Scientific).

#### 4.3. Nucleic Acids Extraction

Seawater was decanted from the self-clumping BBD mats and mats were either frozen immediately (Ft. Lauderdale sample) or preserved (Belize and Looe Key samples) with 5–10 volumes of RNAlater (Qiagen, Germantown, MD, USA) before freezing at −80 °C. DNA and RNA were co-extracted with an Allprep DNA/RNA mini kit (Qiagen, Germantown, MD, USA) from the Belize samples. RNA was treated with DNase I (New England Biolabs, Ipswich, MA, USA), concentrated with RNA Clean & Concentrate (Zymo Research, Irvine, CA, USA), and cDNA was synthesized with a VILO Superscript cDNA Synthesis kit (Invitrogen, Carlsbad, CA, USA). DNA was extracted from the two Florida field samples and from the cyanobacterial enrichment culture with a Dneasy Powersoil DNA extraction kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions with the addition of 12 μL of proteinase K (New England Biolabs, Ipswich, MA, USA) to the

bead tube with solution C1 and incubated 30 min at room temperature before the bead beating step.

#### 4.4. V6 Amplicon Libraries of Belize Samples

The V6 region of bacterial 16S rRNA genes were amplified from both DNA and cDNA of the Belize samples with previously published primers [44] using previously described methods [9]. Briefly, the V6 region was amplified in triplicate with Phusion High-Fidelity Polymerase (New England Biolabs, Ipswich, MA, USA). Triplicate PCR amplifications were pooled for each sample, cleaned with a MinElute kit (Qiagen, Germantown, MD, USA), and quantified by NanoDrop (ThermoScientific, NanoDrop Products, Wilmington, DE, USA). Two hundred nanograms of each cleaned amplicon library was submitted to the Interdisciplinary Center for Biotechnology Research at the University of Florida (RRID:SCR\_019152) where the libraries were size selected for fragments from 200 to 240 bp with a 2% agarose PippinPrep cassette and cleaned again to remove agarose. Sequencing was performed on an Illumina MiSeq with a 150-bp paired-end protocol, using single indexing. Sequencing reads were parsed by Illumina index at the sequencing center and further parsed by the inline barcode using with the command-line options of FASTX-toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/) (accessed on 30 July 2018)). Primers and adaptors were removed using cutadapt v. 2.8 [45] and sickle v. 1.33 [46]. Parsed, quality-filtered amplicon sequencing reads are publicly available through NCBI's Sequence Read Archive under the Bioproject ID PRJNA645365. Quality-filtered paired reads were merged and amplicon sequence variants were determined from de-replicated sequences using taxonomic assignment from the SILVA small subunit ribosomal RNA database v. 132 database [47] with DADA2 v. 1.10.1 [48]. Sequences classified as mitochondria or chloroplast were removed from further analysis. Prevalent sequences that were unclassified were searched against NCBI's non-redundant nucleotide collection with BLASTn [49]. Bacterial community analysis was completed with phyloseq v. 1.26.1 [50] and plotted with ggplot2 v. 3.1.1 [51].

#### 4.5. Metagenomic Library Preparation

A total of six metagenomic libraries were prepared. To ensure enough DNA for library preparation, extracted DNA from Belize samples were pooled as follows. Samples SID1, SID2, and SID3 were pooled for metagenome library "SID1", samples SID8m, SIDa, and SIDE were pooled for metagenome library "SID2", and samples SIDH, SIDI, SIDL, and SIDO, all known producers of looekeyolide C/D, were pooled for metagenome library "SID3". The three pooled DNA samples from Belize were sent to the University of Maryland Institute for Bioscience and Biotechnology Research where metagenomic libraries were prepared with a TruSeq DNA Sample Preparation Kit (Illumina, San Diego, CA, USA) and sequenced on an Illumina HiSeq with a 100-bp paired-end protocol. Metagenomic libraries for the three Florida samples were prepared with a Nextera DNA Flex kit (Illumina, San Diego, CA, USA) and sequenced on an Illumina NextSeq500 at the University of Florida Interdisciplinary Center for Biotechnology Research with a 150-bp paired-end protocol.

#### 4.6. Metagenomic Analysis

Quality-filtering and removal of sequencing adaptors of the 100-bp sequencing reads of the Belize samples was performed with cutadapt v. 2.8 [45] and sickle v. 1.33 [46] with a removal of all reads with Ns, a minimum quality score of 30, and a minimum length of 100 bp. Quality-filtering of the 150-bp sequencing reads of the Florida samples was performed with the Minoche [52] filtering pipeline in illumina-utils v. 2.3 [44] and sequencing adaptors were removed with cutadapt v. 2.8 and sickle v. 1.33. Quality-filtered and adaptor-free metagenomic sequencing reads are publicly available through NCBI's Sequence Read Archive under the Bioproject ID PRJNA647383. Metagenomic libraries for Belize samples were assembled with MetaSPAdes v. 3.12 [53] and metagenomic libraries for Florida samples were assembled with MEGAHIT v. 1.1.4 [54,55]. Unassembled quality-

filtered sequencing reads were mapped to the metagenomic assemblies with bowtie2 v. 2.3.5.1 [56] and sorted with SAMtools v. 1.10 [57].

Metagenome-assembled genomes (MAGs) were retrieved by binning of contigs with MetaBAT v. 2.13 [58]. Cyanobacterial MAGs from this study as well as our previously published BBD cyanobacterial MAGs [12] are publicly available through NCBI's Sequence Read Archive under the Bioproject ID PRJNA647383. Genome quality was assessed with the Microbial Genomes Atlas (MiGA) online [59]. Taxonomic classification of cyanobacterial MAGs was performed with GTDB-Tk v. 2.1.0 and database version R207\_v2 using default settings [60,61]. The average nucleotide identity of shared genes was assessed pairwise with the Average Nucleotide Identity calculator from the enveomics toolbox [62]. The genomes of closely related strains of *Roseofilum*, including four strains from this study and five previously published strains [12,22,23], were annotated with Prokka v. 1.12 [63] and comparative genomic content was analyzed with Roary v. 3.12.0 [64]. An approximately-maximum-likelihood phylogenetic tree of the nine *Roseofilum* genomes was created from the alignment of core genes with FastTree v. 2.1.7 [65] and plotted with Phadango v. 1.3.0 [66]. Biosynthetic gene clusters were identified with the online antiSMASH database bacterial version 6 [67] and with PRISM4 v. 4.4.5 [68]. Biosynthetic gene clusters were visualized with clinker v. 0.0.21 [69] and edited with inkscape v. 1.1.0 [70–72].

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/md21020076/s1>, Figure S1: LRLC-MS data showing the presence of looekeyolides C and D in the MeOH extract of 2014 Belize collection; Figure S2: LRLC-MS data showing the presence of looekeyolides C and D in the EtOAc-MeOH (1:1) extract of 2018 Belize collection; Figure S3: LRLC-MS data showing the presence of looekeyolides C and D in the EtOAc-MeOH (1:1) extract of 2018 Fort Lauderdale, Florida collection; Figure S4: LRLC-MS data showing the presence of looekeyolides C and D in fraction 3 of the 2014 Belize collection; Figure S5: <sup>1</sup>H NMR (600 MHz, CD<sub>3</sub>OD) spectrum of looekeyolide D (4); Figure S6: <sup>13</sup>C NMR (151 MHz, CD<sub>3</sub>OD) spectrum of looekeyolide D (4); Figure S7: DQF-COSY NMR (600 MHz, CD<sub>3</sub>OD) spectrum of looekeyolide D (4); Figure S8: HSQC NMR (600 MHz, CD<sub>3</sub>OD) spectrum of looekeyolide D (4); Figure S9: HMBC NMR (600 MHz, CD<sub>3</sub>OD) spectrum of looekeyolide D (4); Figure S10: 2D-NOESY NMR (600 MHz, CD<sub>3</sub>OD) spectrum of looekeyolide D (4); Figure S11: Alignment of A domains of LkII and LkII(F) by ClustalW; Figure S12: Comparison of gene similarities in the biosynthetic gene clusters for looekeyolide A/B and looekeyolide C/D. Figure S13: Biosynthetic gene clusters for malyngamides in *Roseofilum* MAGs; Table S1: Metadata and sequencing read metrics for V6 amplicon libraries from Black Band Disease cyanobacterial mats from *Siderastrea siderea* corals in Belize; Table S2: Metadata and sequencing read metrics for metagenomic libraries from Black Band Disease cyanobacterial mats from *Siderastrea siderea* corals in Belize and Florida; Table S3: Quality metrics of metagenome-assembled genomes (MAGs) of (A) non-*Roseofilum* cyanobacteria from *Siderastrea siderea*, (B) *Roseofilum* from *Siderastrea siderea*, and (C) previously published *Roseofilum* strains; Table S4: (A) Pairwise Average Nucleotide Identity (ANI) of shared genes among cyanobacterial MAGs from Black Band Disease on *Siderastrea siderea* corals. Values of 75% or below are too close to the detection limit for confident assessments. Values above this threshold are highlighted in green, (B) Pairwise Average Nucleotide Identity (ANI) of shared genes among *Roseofilum* MAGs from Black Band Disease on multiple coral species. Table S5: Specificity codes for adenylation (A) domains of two NRPSs for the biosynthesis of Lk-A/B and Lk-C/D.

**Author Contributions:** Conceptualization, M.T., V.J.P. and J.L.M.; formal analysis, J.L.M., S.P.G., A.L.B., S.M. and Y.D.; investigation, J.L.M., S.P.G., M.T. and V.J.P.; resources, M.T. and V.J.P.; data curation, J.L.M.; writing—original draft preparation, J.L.M. and S.P.G.; writing—review and editing, V.J.P., M.T., A.L.B. and Y.D.; visualization, J.L.M. and S.P.G.; funding acquisition, M.T., V.J.P. and Y.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the George E. Burch Fellowship in Theoretic Medicine and Affiliated Theoretic Science from the Smithsonian Institution (M.T.), Mote Marine Laboratory Protect Our Reefs grants POR 2012-1, POR 2013-2, POR 2014-10, and POR 2015-1 (M.T., V.J.P.), and NIH R35 grant GM128742 (Y.D.).

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

**Data Availability Statement:** Metagenomic sequencing reads and metagenome-assembled genomes are available in GenBank under Bioproject PRJNA647383. Amplicon sequencing reads are available in GenBank under Bioproject PRJNA645365.

**Acknowledgments:** We are grateful to Florida Atlantic University and its Harbor Branch Oceanographic Institute for access to their spectrometer facilities. This is contribution # 1070 of the Smithsonian Caribbean Coral Reef Ecosystem and contribution # 1186 of the Smithsonian Marine Station.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Leão, T.; Wang, M.; Moss, N.; da Silva, R.; Sanders, J.; Nurk, S.; Gurevich, A.; Humphrey, G.; Reher, R.; Zhu, Q.; et al. A Multi-Omics Characterization of the Natural Product Potential of Tropical Filamentous Marine Cyanobacteria. *Mar. Drugs* **2021**, *19*, 20. [[CrossRef](#)] [[PubMed](#)]
2. Harvey, A.L.; Edrada-Ebel, R.; Quinn, R.J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat. Rev. Drug Discov.* **2015**, *14*, 111–129. [[CrossRef](#)] [[PubMed](#)]
3. Engene, N.; Gunasekera, S.P.; Gerwick, W.H.; Paul, V.J. Phylogenetic Inferences Reveal a Large Extent of Novel Biodiversity in Chemically Rich Tropical Marine Cyanobacteria. *Appl. Environ. Microbiol.* **2013**, *79*, 1882–1888. [[CrossRef](#)] [[PubMed](#)]
4. Yancey, C.E.; Smith, D.J.; Den Uyl, P.A.; Mohamed, O.G.; Yu, F.; Ruberg, S.A.; Chaffin, J.D.; Goodwin, K.D.; Tripathi, A.; Sherman, D.H.; et al. Metagenomic and Metatranscriptomic Insights into Population Diversity of Microcystis Blooms: Spatial and Temporal Dynamics of Mcy Genotypes, Including a Partial Operon That Can Be Abundant and Expressed. *Appl. Environ. Microbiol.* **2022**, *88*, e0246421. [[CrossRef](#)] [[PubMed](#)]
5. Antonius, A. New Observations on Coral Destruction in Reefs. In Proceedings of the Tenth Meeting of the Association of Island Marine Laboratories of the Caribbean; University of Puerto Rico Mayaguez, Mayagüez, Puerto Rico, 4–7 September 1973; Volume 10.
6. Roff, G. Earliest Record of a Coral Disease from the Indo-Pacific? *Coral Reefs* **2016**, *35*, 457. [[CrossRef](#)]
7. Casamatta, D.; Stanić, D.; Gantar, M.; Richardson, L.L. Characterization of *Roseofilum reptotaenium* (Oscillatoriales, Cyanobacteria) gen. et sp. nov. isolated from Caribbean black band disease. *Phycologia* **2012**, *51*, 489–499. [[CrossRef](#)]
8. Carlton, R.G.; Richardson, L.L. Oxygen and Sulfide Dynamics in a Horizontally Migrating Cyanobacterial Mat: Black Band Disease of Corals. *FEMS Microbiol. Ecol.* **1995**, *18*, 155–162. [[CrossRef](#)]
9. Meyer, J.L.; Gunasekera, S.P.; Scott, R.M.; Paul, V.J.; Teplitski, M. Microbiome Shifts and the Inhibition of Quorum Sensing by Black Band Disease Cyanobacteria. *ISME J.* **2016**, *10*, 1204–1216. [[CrossRef](#)]
10. Gunasekera, S.P.; Meyer, J.L.; Ding, Y.; Abboud, K.A.; Luo, D.; Campbell, J.E.; Angerhofer, A.; Goodsell, J.L.; Raymundo, L.J.; Liu, J.; et al. Chemical and Metagenomic Studies of the Lethal Black Band Disease of Corals Reveal Two Broadly Distributed, Redox-Sensitive Mixed Polyketide/Peptide Macrocycles. *J. Nat. Prod.* **2019**, *82*, 111–121. [[CrossRef](#)]
11. Buerger, P.; Wood-Charlson, E.M.; Weynberg, K.D.; Willis, B.L.; van Oppen, M.J.H. CRISPR-Cas Defense System and Potential Prophages in Cyanobacteria Associated with the Coral Black Band Disease. *Front. Microbiol.* **2016**, *7*, 2077. [[CrossRef](#)]
12. Meyer, J.L.; Paul, V.J.; Raymundo, L.J.; Teplitski, M. Comparative Metagenomics of the Polymicrobial Black Band Disease of Corals. *Front. Microbiol.* **2017**, *8*, 618. [[CrossRef](#)] [[PubMed](#)]
13. Sekar, R.; Kaczmarek, L.T.; Richardson, L.L. Microbial Community Composition of Black Band Disease on the Coral Host *Siderastrea siderea* from Three Regions of the Wider Caribbean. *Mar. Ecol. Prog. Ser.* **2008**, *362*, 85–98. [[CrossRef](#)]
14. Richardson, L.L.; Sekar, R.; Myers, J.L.; Gantar, M.; Voss, J.D.; Kaczmarek, L.; Remily, E.R.; Boyer, G.L.; Zimba, P.V. The presence of the cyanobacterial toxin microcystin in black band disease of corals. *FEMS Micro. Lett.* **2007**, *2*, 182–187. [[CrossRef](#)] [[PubMed](#)]
15. Voss, J.D.; Mills, D.K.; Myers, J.L.; Remily, E.R.; Richardson, L.L. Black Band Disease Microbial Community Variation on Corals in Three Regions of the Wider Caribbean. *Microb. Ecol.* **2007**, *54*, 730–739. [[CrossRef](#)]
16. Havermann, S.A.; Foster, J.S. Comparative characterization of the microbial diversities of an artificial microbialite model and a natural stromatolite. *Appl. Environ. Micro.* **2008**, *74*, 7410–7421. [[CrossRef](#)]
17. Aeby, G.S.; Work, T.M.; Runyon, C.M.; Shore-Maggio, A.; Ushijima, B.; Videau, P.; Beurmann, S.; Callahan, S.M. First Record of Black Band Disease in the Hawaiian Archipelago: Response, Outbreak Status, Virulence, and a Method of Treatment. *PLoS ONE* **2015**, *10*, e0120853. [[CrossRef](#)]
18. Miller, A.W.; Richardson, L.L. A Meta-Analysis of 16S rRNA Gene Clone Libraries from the Polymicrobial Black Band Disease of Corals. *FEMS Microbiol. Ecol.* **2011**, *75*, 231–241. [[CrossRef](#)]
19. Sekar, R.; Mills, D.K.; Remily, E.R.; Voss, J.D.; Richardson, L.L. Microbial Communities in the Surface Mucopolysaccharide Layer and the Black Band Microbial Mat of Black Band-Diseased *Siderastrea siderea*. *Appl. Environ. Microbiol.* **2006**, *72*, 5963–5973. [[CrossRef](#)]
20. Den Uyl, P.A.; Richardson, L.L.; Jain, S.; Dick, G.J. Unraveling the Physiological Roles of the Cyanobacterium *Geitlerinema* sp. BBD and Other Black Band Disease Community Members through Genomic Analysis of a Mixed Culture. *PLoS ONE* **2016**, *11*, e0157953. [[CrossRef](#)]

21. Myers, J.L.; Sekar, R.; Richardson, L.L. Molecular Detection and Ecological Significance of the Cyanobacterial Genera *Geitlerinema* and *Leptolyngbya* in Black Band Disease of Corals. *Appl. Environ. Microbiol.* **2007**, *73*, 5173–5182. [[CrossRef](#)]
22. Sato, Y.; Civiello, M.; Bell, S.C.; Willis, B.L.; Bourne, D.G. Integrated Approach to Understanding the Onset and Pathogenesis of Black Band Disease in Corals: Integrated Approaches to Understand BBD Aetiology. *Environ. Microbiol.* **2016**, *18*, 752–765. [[CrossRef](#)]
23. Buerger, P.; Alvarez-Roa, C.; Weynberg, K.D.; Baekelandt, S.; van Oppen, M.J.H. Genetic, Morphological and Growth Characterisation of a New *Roseofilum* Strain (Oscillatoriales, Cyanobacteria) Associated with Coral Black Band Disease. *PeerJ* **2016**, *4*, e2110. [[CrossRef](#)] [[PubMed](#)]
24. Gondry, M.; Jacques, I.B.; Thai, R.; Babin, M.; Canu, N.; Seguin, J.; Belin, P.; Pernodet, J.-L.; Moutiez, M. A Comprehensive Overview of the Cyclodipeptide Synthase Family Enriched with the Characterization of 32 New Enzymes. *Front. Microbiol.* **2018**, *9*, 46. [[CrossRef](#)]
25. Yao, T.; Liu, J.; Liu, Z.; Li, T.; Li, H.; Che, Q.; Zhu, T.; Li, D.; Gu, Q.; Li, W. Genome Mining of Cyclodipeptide Synthases Unravels Unusual tRNA-Dependent Diketopiperazine-Terpene Biosynthetic Machinery. *Nat. Commun.* **2018**, *9*, 4091. [[CrossRef](#)] [[PubMed](#)]
26. Kwan, J.C.; Teplitski, M.; Gunasekera, S.P.; Paul, V.J.; Luesch, H. Isolation and Biological Evaluation of 8-Epi-Malyngamide C from the Floridian Marine Cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2010**, *73*, 463–466. [[CrossRef](#)] [[PubMed](#)]
27. Morais, J.; Cardoso, A.P.L.R.; Santos, B.A. A Global Synthesis of the Current Knowledge on the Taxonomic and Geographic Distribution of Major Coral Diseases. *Environ. Adv.* **2022**, *8*, 100231. [[CrossRef](#)]
28. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *Nat. Commun.* **2018**, *9*, 5114. [[CrossRef](#)]
29. Konstantinidis, K.T.; Tiedje, J.M. Genomic Insights That Advance the Species Definition for Prokaryotes. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2567–2572. [[CrossRef](#)]
30. Olm, M.R.; Crits-Christoph, A.; Diamond, S.; Lavy, A.; Matheus Carnevali, P.B.; Banfield, J.F. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *MSystems* **2020**, *5*, e00731-19. [[CrossRef](#)]
31. Goris, J.; Konstantinidis, K.T.; Klappenbach, J.A.; Coenye, T.; Vandamme, P.; Tiedje, J.M. DNA-DNA Hybridization Values and Their Relationship to Whole-Genome Sequence Similarities. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 81–91. [[CrossRef](#)]
32. Jones, M.R.; Pinto, E.; Torres, M.A.; Dörr, F.; Mazur-Marzec, H.; Szubert, K.; Tartaglione, L.; Dell’Aversano, C.; Miles, C.O.; Beach, D.G.; et al. CyanoMetDB, a Comprehensive Public Database of Secondary Metabolites from Cyanobacteria. *Water Res.* **2021**, *196*, 117017. [[CrossRef](#)] [[PubMed](#)]
33. Leão, P.N.; Engene, N.; Antunes, A.; Gerwick, W.H.; Vasconcelos, V. The Chemical Ecology of Cyanobacteria. *Nat. Prod. Rep.* **2012**, *29*, 372–391. [[CrossRef](#)] [[PubMed](#)]
34. Kar, J.; Ramrao, D.P.; Zomuanangi, R.; Lalbiaktluangi, C.; Singh, S.M.; Joshi, N.C.; Kumar, A.; Kaushalendra; Mehta, S.; Yadav, M.K.; et al. Revisiting the Role of Cyanobacteria-Derived Metabolites as Antimicrobial Agent: A 21st Century Perspective. *Front. Microbiol.* **2022**, *13*, 1034471. [[CrossRef](#)] [[PubMed](#)]
35. Carpine, R.; Sieber, S. Antibacterial and Antiviral Metabolites from Cyanobacteria: Their Application and Their Impact on Human Health. *Curr. Res. Biotechnol.* **2021**, *3*, 65–81. [[CrossRef](#)]
36. Rojas, V.; Rivas, L.; Cárdenas, C.; Guzmán, F. Cyanobacteria and Eukaryotic Microalgae as Emerging Sources of Antibacterial Peptides. *Molecules* **2020**, *25*, 5804. [[CrossRef](#)]
37. Curren, E.; Leaw, C.P.; Lim, P.T.; Leong, S.C.Y. The Toxic Cosmopolitan Cyanobacteria *Moorena producens*: Insights into Distribution, Ecophysiology and Toxicity. *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 78178–78206. [[CrossRef](#)]
38. Dussault, D.; Vu, K.D.; Vansach, T.; Horgen, F.D.; Lacroix, M. Antimicrobial Effects of Marine Algal Extracts and Cyanobacterial Pure Compounds against Five Foodborne Pathogens. *Food Chem.* **2016**, *199*, 114–118. [[CrossRef](#)]
39. Gondry, M.; Sauguët, L.; Belin, P.; Thai, R.; Amouroux, R.; Tellier, C.; Tuphile, K.; Jacquet, M.; Braud, S.; Courçon, M.; et al. Cyclodipeptide Synthases Are a Family of tRNA-Dependent Peptide Bond-Forming Enzymes. *Nat. Chem. Biol.* **2009**, *5*, 414–420. [[CrossRef](#)]
40. Tietz, J.I.; Schwalen, C.J.; Patel, P.S.; Maxson, T.; Blair, P.M.; Tai, H.-C.; Zakai, U.I.; Mitchell, D.A. A New Genome-Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape. *Nat. Chem. Biol.* **2017**, *13*, 470–478. [[CrossRef](#)]
41. Wang, M.; Fage, C.D.; He, Y.; Mi, J.; Yang, Y.; Li, F.; An, X.; Fan, H.; Song, L.; Zhu, S.; et al. Recent Advances and Perspectives on Expanding the Chemical Diversity of Lasso Peptides. *Front. Bioeng. Biotechnol.* **2021**, *9*, 741364. [[CrossRef](#)]
42. Moss, N.A.; Leão, T.; Rankin, M.R.; McCullough, T.M.; Qu, P.; Korobeynikov, A.; Smith, J.L.; Gerwick, L.; Gerwick, W.H. Ketoreductase domain dysfunction expands chemodiversity: Malyngamide biosynthesis in the cyanobacterium *Okeania hirsuta*. *ACS Chem. Biol.* **2018**, *13*, 3385–3395. [[CrossRef](#)] [[PubMed](#)]
43. Si, Y.; Kretsch, A.M.; Daigh, L.M.; Burk, M.J.; Mitchell, D.A. Cell-free biosynthesis to evaluate lasso peptide formation and enzyme-substrate tolerance. *J. Am. Chem. Soc.* **2021**, *15*, 5917–5927. [[CrossRef](#)] [[PubMed](#)]
44. Eren, A.M.; Vineis, J.H.; Morrison, H.G.; Sogin, M.L. A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology. *PLoS ONE* **2013**, *8*, e66643. [[CrossRef](#)]
45. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* **2011**, *17*, 10–12. [[CrossRef](#)]
46. Joshi, N. *Sickle: A Sliding-Window, Adaptive, Quality—Based Trimming Tool for FastQ Files*; UC Davis Bioinformatics Core: Davis, CA, USA, 2011.

47. Yilmaz, P.; Parfrey, L.W.; Yarza, P.; Gerken, J.; Pruesse, E.; Quast, C.; Schweer, T.; Peplies, J.; Ludwig, W.; Glöckner, F.O. The SILVA and “All-Species Living Tree Project (LTP)” Taxonomic Frameworks. *Nucleic Acids Res.* **2014**, *42*, D643–D648. [[CrossRef](#)] [[PubMed](#)]
48. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)]
49. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
50. McMurdie, P.J.; Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [[CrossRef](#)]
51. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2016; ISBN 9783319242774.
52. Minoche, A.E.; Dohm, J.C.; Himmelbauer, H. Evaluation of Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer Systems. *Genome Biol.* **2011**, *12*, R112. [[CrossRef](#)]
53. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res.* **2017**, *27*, 824–834. [[CrossRef](#)]
54. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. *Bioinformatics* **2015**, *31*, 1674–1676. [[CrossRef](#)] [[PubMed](#)]
55. Li, D.; Luo, R.; Liu, C.-M.; Leung, C.-M.; Ting, H.-F.; Sadakane, K.; Yamashita, H.; Lam, T.-W. MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler Driven by Advanced Methodologies and Community Practices. *Methods* **2016**, *102*, 3–11. [[CrossRef](#)] [[PubMed](#)]
56. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
57. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
58. Kang, D.D.; Froula, J.; Egan, R.; Wang, Z. MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities. *PeerJ* **2015**, *3*, e1165. [[CrossRef](#)]
59. Rodríguez-R, L.M.; Gunturu, S.; Harvey, W.T.; Rosselló-Mora, R.; Tiedje, J.M.; Cole, J.R.; Konstantinidis, K.T. The Microbial Genomes Atlas (MiGA) Webserver: Taxonomic and Gene Diversity Analysis of Archaea and Bacteria at the Whole Genome Level. *Nucleic Acids Res.* **2018**, *46*, W282–W288. [[CrossRef](#)]
60. Chaumeil, P.-A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database. *Bioinformatics* **2019**, *36*, 1925–1927. [[CrossRef](#)]
61. Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.-A.; Hugenholtz, P. A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life. *Nat. Biotechnol.* **2018**, *36*, 996–1004. [[CrossRef](#)]
62. Rodríguez-R, L.M.; Konstantinidis, K.T. *The Enveomics Collection: A Toolbox for Specialized Analyses of Microbial Genomes and Metagenomes*; PeerJ Preprints: London, UK, 2016.
63. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
64. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)] [[PubMed](#)]
65. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)]
66. Hadfield, J.; Croucher, N.J.; Goater, R.J.; Abudahab, K.; Aanensen, D.M.; Harris, S.R. Phandango: An Interactive Viewer for Bacterial Population Genomics. *Bioinformatics* **2018**, *34*, 292–293. [[CrossRef](#)]
67. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [[CrossRef](#)] [[PubMed](#)]
68. Skinnider, M.A.; Johnston, C.W.; Gunabalasingam, M.; Merwin, N.J.; Kieliszek, A.M.; MacLellan, R.J.; Li, H.; Ranieri, M.R.M.; Webster, A.L.H.; Cao, M.P.T.; et al. Comprehensive Prediction of Secondary Metabolite Structure and Biological Activity from Microbial Genome Sequences. *Nat. Commun.* **2020**, *11*, 6058. [[CrossRef](#)]
69. Gilchrist, C.L.M.; Chooi, Y.-H. Clinker & Clustermap.js: Automatic Generation of Gene Cluster Comparison Figures. *Bioinformatics* **2021**, *37*, 2473–2475. [[CrossRef](#)]
70. Inkscape. Inkscape Project. 2020. Available online: <https://inkscape.org> (accessed on 22 December 2022).
71. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
72. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.