Family-wide Structural Characterization and Genomic Comparisons Decode the Diversity-oriented Biosynthesis of Thalassospiramides by Marine Proteobacteria*

Received for publication, September 7, 2016, and in revised form, November 8, 2016. Published, JBC Papers in Press, November 14, 2016, DOI 10.1074/jbc.M116.756858

Weipeng Zhang†‡, Liang Lu†1, Qiliang Lai†, Beika Zhu†, Zhongrui Li†, Ying Xu‡, Zongze Shao‡, Karl Herrup‡, Bradley S. Moore‡, Avena C. Ross*‡, and Pei-Yuan Qian†‡

From the †Division of Life Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong, the ‡Third Institute of Oceanography, State Oceanic Administration, Xiamen 361000, China, the §School of Life Science, Shenzhen University, Shenzhen, Guangdong Province 518060, China, the ¶Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92037, and the **Department of Chemistry, Queen’s University, Kingston, Ontario K7L 3N6, Canada

Edited by Gerald W. Hart

The thalassospiramide lipopeptides have great potential for therapeutic applications; however, their structural and functional diversity and biosynthesis are poorly understood. Here, by cultivating 130 Rhodospirillaceae strains sampled from oceans worldwide, we discovered 21 new thalassospiramide analogues and demonstrated their neuroprotective effects. To investigate the diversity of biosynthetic gene cluster (BGC) architectures, we sequenced the draft genomes of 28 Rhodospirillaceae strains. Our family-wide genomic analysis revealed three types of dysfunctional BGCs and four functional BGCs whose architectures correspond to four production patterns. This correlation allowed us to reassess the “diversity-oriented biosynthesis” proposed for the microbial production of thalassospiramides, which involves iteration of several key modules. Preliminary evolutionary investigation suggested that the functional BGCs could have arisen through module/domain loss, whereas the dysfunctional BGCs arose through horizontal gene transfer. Further comparative genomics indicated that thalassospiramide production is likely to be attendant on particular genes/pathways for amino acid metabolism, signaling transduction, and compound efflux. Our findings provide a systematic understanding of thalassospiramide production and new insights into the underlying mechanism.

Microbial natural products are an excellent source of bioactive molecules (1). Structurally complex and diverse, these compounds have applications in many fields, including medicine, agriculture, cosmetics, and food-production/preservation (2). The thalassospiramides are a family of lipopeptides produced by marine α-proteobacteria belonging to the Rhodospirillaceae family. Fenical and co-workers (3) first characterized thalassospiramides A and B in 2007 from Thalassospira sp. CNJ-328, and then, in 2013, we reported an additional 14 thalassospiramides from Thalassospira and Tistrella strains. We found that many of these compounds display a high inhibitive potency (nanomolar levels) of human calpain protease (4), a calcium-dependent enzyme family implicated in neurological disorders (5), muscular dystrophies, cortical cataracts (6), and cancer (7). For example, inflammatory cytokines that lead to neuronal network disruption, increase intracellular calcium, and increase calpain protease activity are a possible trigger for Alzheimer’s disease (8). The potent activity of thalassospiramides against calpain protease may allow them to act as neuroprotective agents and result in their therapeutic application for Alzheimer’s therapy.

The thalassospiramides are hybrid non-ribosomal peptide-polypeptide molecules. Unlike ribosomal peptides that are encoded directly in the genome and translated by the ribosome (9), NRPSs are custom-produced by a dedicated megasynthetase that acts as both the catalyst and template, incorporating a single amino acid at each enzyme module in a predictable linear fashion. In 2013, we proposed that the thalassospiramide molecules are made by two related hybrid polypeptide synthase-non ribosomal peptide synthetase/NRPS pathways in Thalassospira and Tistrella strains via a highly non-canonical functioning of the megasynthetase using multimodular skipping and iteration, internal supplementation, and nonlinear substrate channeling (4). Thus, the thalassospiramides are produced by “diversity-oriented biosynthesis,” whereby the megasynthetase generates a library of molecules rather than a single refined structure (10). In light of the promising bioactivity and unusual proposed bio-

*This work was supported by grants from the China Ocean Mineral Resources Research and Development Association (DY135 to P. Y. Q.), the Natural Sciences and Engineering Research Council of Canada (NSERC-DG to A. C. R.), and the National Institutes of Health (GM97509 to B. S. M.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

† This article contains supplemental Data S1 and S2, Figs. S1–S6, and Tables S1–S6. All sequenced genomes have been deposited in the NCBI database and can be accessed through the numbers listed in supplemental Table S3.

‡ Both authors contributed equally to this work.

† To whom correspondence may be addressed: Dept. of Chemistry, Queen’s University, Kingston, ON K7L 3N6, Canada. E-mail: avena.ross@queensu.ca.

§ To whom correspondence may be addressed: Div. of Life Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. E-mail: boqianpy@ust.hk.

© 2016 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.
The Diversity-oriented Biosynthesis of Thalassospiramides

Results

New Thalassospiramides and Their Activities—To study the breadth of thalassospiramide production, 76 Thalassospira, 29 Tistrella, 16 Oceanibaculum, two Fodinicurvata, and seven unclassified Rhodospirillaceae strains from marine environments worldwide were obtained, leading to the isolation of 21 new thalassospiramides. Fig. 1 shows the sampling locations of Thalassospira and Tistrella. Fig. 2A shows the chemical structures of 37 previously reported and newly discovered thalassospiramides. The thalassospiramides are divided into two structural groups, “A-like” and “B-like,” distinguished by the identity of the N-terminal amino acid residue. A-like molecules contain a standard amino acid residue, whereas B-like molecules have a statine-like amino acid. The thalassospiramides are further subdivided into subgroups A through J. The majority of new thalassospiramides belong to the previously described subclasses A-F, which differ in the length of the peptide chain; however, seven molecules were placed in the new subgroups H, I, and J. Thalassospiramides H, H1, H2, and H3 are truncated variants much like the C subclass and feature repeating serine or valine motif. Thalassospiramides I and J are related to the A subgroup and possess an additional serine or valine residue, respectively. As the thalassospiramides share a 12-membered ring structure and members are amenable to “b-y” mode sequencing by tandem mass spectrometry (MS2) analysis (4), screening for new analogues and chemical structure determination of all thalassospiramides were first accomplished using high-resolution MS and MS2. Comparison between the m/z of observed MS2 fragments with diagnostic thalassospiramide fragments, such as m/z 390.19 ([M+H]+ ring), 489.26 ([M+H]+ ring+Val) or ring-modified signals m/z 414.20 ([M+H]+ ring+CH3), 503.26 ([M+H]+ ring+CH3+Val) allowed identification of all building blocks and their sequence (complete MS2 data are provided in supplemental Data S1). Additionally, we characterized at least one representative member of each new subclass thoroughly using multidimensional nuclear magnetic resonance (NMR) (supplemental Table S1).

To discover new calpain inhibitors, 12 of the thalassospiramides identified in this paper were subjected to a calpain inhibition assay, exhibiting a range of potency spanning several orders of magnitude (Fig. 2B). Thalassospiramide C2, the smallest molecule, displayed the best inhibitory activity (IC50 = 1.6 nM) of the entire family. The result also hints that the molecular size of these molecules may play a role in the calpain protease inhibitory action. In light of their potency as calpain inhibitors, we examined whether thalassospiramides were also neuropro-
tective agents of potential use in the treatment of conditions such as Alzheimer’s disease. Thalassospiramides A4, A6, A8, C, D1, H, H1, and F1 with a range of calpain inhibitory activities (A4 IC\(_{50}\) = 12.5 nM, A6 IC\(_{50}\) = 37.4 nM, A8 IC\(_{50}\) = 53.8 nM, C IC\(_{50}\) = 3.4 nM, D1 IC\(_{50}\) = 34.7 nM, H (H1) IC\(_{50}\) = 36.8 nM, and F1 IC\(_{50}\) = 41.3 nM) were chosen for testing via a neuroprotec-
tive assay (4, 23). Cultured primary mouse cortical neurons were exposed to inflammatory stress by adding conditioned medium from \(\beta\)-amyloid-stimulated THP1 cells, which are a human monocytic cell line. In the negative control, adding the conditioned medium increased neuronal cell cycling by 5-fold, indicating a neurotoxic response. Adding thalassospiramides A4, H, or H1 significantly reduced this neurotoxic response (Fig. 2C), indicating that thalassospiramides may have potent neuro-
The Diversity-oriented Biosynthesis of Thalassospiramides

Experiments were performed in triplicate. Statistics were performed using Student’s t-test (*, p < 0.05; **, p < 0.01; ***, p < 0.005).
DSM21159 (cluster 3), *Thalassospira lucentensis* 1A00383 (cluster 5), and *Fodinicurvata fenggangensis* DSM21160 (cluster 6) were searched against homologues in other strains using the Basic Local Alignment Search Tool (BLAST) to confirm their relative phylogenetic distances. For example, based on the hit scores of RpsC protein pairs, DSM21159 was closer to *O. pacificum* 1A02656 than to *T. mesophila* 1A00756; by contrast, hit scores of the 2C domain sequences indicated a higher similarity between DSM21159 and 1A00756 than between DSM21159 and 1A02656 (Fig. 4, **B–D**).

**Genome-wide Functions Associated with Thalassospiramide Production**—Linking HGT to functional loss of BGCs suggested that particular metabolic functions beyond the BGCs may influence thalassospiramide production. Comparative genomics were performed between 18 genomes of thalassospiramide-producing bacteria (*Thalassospira*, *Tistrella*, and *Oceanibaculum*) and 14 genomes of non-producing bacteria (*Thalassospira* and *Tistrella*), which in total revealed 108 significantly changed Kyoto Encyclopedia of Genes and Genomes (KEGG) genes (Student’s *t* test, *p* < 0.001; supplemental Data S2). A large proportion of these genes were classified under the categories of amino acid metabolism (24 genes), signal transduction (18 genes), and compound transport (14 genes) (Fig. 5). Remarkably, based on genes within these three categories, the thalassospiramide-producing and -non-producing bacteria formed two distinct groups. Two genes for glutamate biosynthesis via arginine, succinylarginine dihydrolase (*astB* and K01484), and succinyl glutamic semialdehyde dehydrogenase (*astE*, K06447), were present in nearly all genomes of thalassospiramide-producing bacteria but absent in the genomes of all non-producing bacteria. Similarly, a gene involved in L-tyrosine production, prephenate dehydrogenase (K00210), was only found in thalassospiramide-producing bacteria.

In addition to amino acid metabolic pathways, the producing and non-producing bacteria could also be differentiated by a number of genes for signal transduction. For example the *pho*
family regulator (K07657) was highly enriched in the genomes of thalassospiramide-producing bacteria (Fig. 5). The thalassospiramide-producing bacteria were also equipped with a number of specific transporters for compound transport, such as the mac family transporters (K13888 and K05685), which involve efflux of macrolide antibiotics (14), and the major facilitator superfamily (MFS) proteins (K08225 and K08151), which are involved in symport, antiport, and uniport of various substrates, including antibiotics (15). Pathway reconstruction confirmed the presence of specific amino acid metabolism functions for the thalassospiramide-producing bacteria, which could be exemplified by comparison of two closely related strains, 1A00753 and 1A00383 (supplemental Fig. S4). Moreover, the alignment of genomes of these two strains suggested that particular genes involved in amino acid metabolism, carbohydrate transport, and signal transduction are lost from strain 1A00383 (supplemental Fig. S5).

Because the abovementioned genes (Fig. 5) show significant differences between the thalassospiramide producing and non-producing groups, we conjectured that these genes co-occur with the BGCs across all bacterial genomes. Co-occurrence analysis of all genes in the 32 genomes using a Spearman’s correlation coefficient of >0.8 and statistically significant p value of <0.01 revealed a strong correlation between thalassospiramides BGCs and 16 genes, including succinyl arginine dihydrolase (K01484), succinyl glutamic semialdehyde
dehydrogenase (K06447), arginine N-succinyl transferase (K00674), glutamate synthase (NADPH/NADH) large chain (K02023), ammonium transporter, amf family (K03320), rsbV, anti-sigma B factor antagonist (K04749), arginine:pyruvate transaminase (K08688), creatinase (K11741), quaternary ammonium compound-resistance protein, sugE (K11070), spermidine/putrescine transport system permease protein (K11004), ATP-binding cassette, subfamily B, bacterial Hyd/CarB (K00305), sarcosine oxidase, subunit gamma (K01567), acetylpyruvate hydrolase (K00797), spermidine synthase (K00605), aminomethyltransferase (K11475), gntR family transcriptional regulator (K02500), cyclase (K02501), glutamine amidotransferase (K08714), voltage-gated sodium channel (K11717), sufS, cytochrome desulphurase / selenocysteine lyase (K07785), MFS transporter, NRE family, putative nickel resistance protein (K03413), two-component system, chemotaxis family regulator, CheY (K05606), methylmalonyl-CoA/ethylmalonyl-CoA epimerase (K11624), two-component system, narR family, response regulator, ydfl (K02164), nitric oxide reductase, norE (K02448), nitric oxide reductase, norD (K04748), nitric oxide reductase, norQ (K02305), nitric oxide reductase, norC (K06447), succinylglutamic semialdehyde dehydrogenase (K04184), succinyllarginine dihydrolase (K11534), deoR family transcriptional regulator, deoxoyribose repressor (K05346), deoxyribonucleoside repressor, deoR (K05225), MFS transporter, enterobactin (siderophore) exporter (K01745), histidine ammonia-lyase (K08151), MFS transporter, DHA1 family, tetracycline resistance protein (K06073), marR family transcriptional regulator, regulator for hemolysin (K06673), arginine N-succinyltransferase (K00210), prephenate dehydrogenase (K07665), copper resistance phosphate regulon response regulator, cusR (K07657), phosphate regulon response regulator, phoB (K09471), gamma-glutamylputrescine oxidase (K11250), leucine efflux protein (K03499), trk system potassium uptake protein (K03282), large conductance mechanosensitive channel (K02471), putative ATP-binding cassette transporter (K07775), two-component system, ompR family, response regulator, resD (K01914), CRP/FNR family regulator, cyclic AMP receptor protein (K01546), K+ -transporting ATPase ATPase A chain (K01547), K+ -transporting ATPase C chain (K01548), monosaccharide-transporting ATPase (K10820), and large conductance mechanosensitive channel (K03282) (supplemental Fig. S6).

Notably, all of these genes are absent in DSM21159, 1A00383, and DSM21160, the strains with clusters 3, 5, and 6, respectively.

Discussion

This study describes a remarkable diversity-oriented biosynthetic system for the production of 37 thalassospiramides...
within the Rhodospirillaceae family. We discovered new thalassospiramides and demonstrated that the thalassospiramides have neuroprotective activities (Fig. 2) and are worth exploring in the context of human Alzheimer’s disease and other neurodegenerative diseases. The identification of several extremely promising calpain protease inhibitors and neuroprotective agents from the thalassospiramides, where small structural changes have large effects on molecule potency, emphasizes the importance of the complex diversity-oriented biosynthesis.

Our family-wide genome sequencing effort revealed seven BGCs wherein clusters 2, 3, 4, 5, and 6 are newly identified in this report. The pathways differ at biosynthetic modules 1a, 1b, 4, and 5, with newly identified clusters lacking certain functional domains in comparison with cluster 1 (Fig. 3). On a cursory analysis of the pathways, clusters 2–6 appear to lie along a continuum of “completeness” between cluster 1 and 7. Regarding the functional BGCs 1, 2, 4, and 7, the more complete the cluster is, the more diverse the molecules it produces, motivating us to reassess the biosynthetic mechanism based on the compound structures, production patterns, and gene cluster architectures.

Several of the new thalassospiramides, including C2 and F1, are simply additional examples of previously described biosynthetic mechanisms involving in trans amino acid supplementation and adenylation domain promiscuity, respectively. However, the majority of the new molecules result from new instances of non-canonical biosynthesis, thereby expanding our understanding of this biosynthetic system. The first example is that of the new thalassospiramides A6, A7, A8, B3, B4, B5, F2, and F3. They are made by cluster 1-containing Thalassospira strains and have structures very similar to known thalassospiramides, with the addition of a methyl group on the valine nitrogen of the core cyclic peptide. However, based upon protein BLAST analysis, there is no methyltransferase domain in module 5, indicating that this methyl group might result from the in trans activity of an external transferase. Second, clusters 1, 2, and 4 produce thalassospiramide A9, in which the presence of a phenylalanine residue in the core cyclic portion of the peptide suggests that the A domain in module 6 can activate both Tyr and Phe in a similar manner to module 1a. Third, the new A-like thalassospiramide groups H, I, and J are synthesized by cluster 4. A single module stuttering is proposed for the production of thalassospiramides H, H1, H2, H3, and J, where either module 1a or module 2 is used in a repeated fashion. Finally and most intriguingly, the thalassospiramides I and I1 can be produced by the iterative use of modules 1 through 4. This observation is particularly interesting as cluster 4 also produces thalassospiramide A through iterative use of modules 2–4.

A summary of the putative biosynthetic mechanisms of gene clusters 1, 2, 4, and 7, including module skipping and stuttering, multimodular iteration, in trans amino acid activation, and substrate channeling, is represented in Fig. 6. The array of unusual mechanisms used by this family of homologous pathways to produce molecular diversity is without parallel.

The proposed biosynthetic mechanism cannot explain why the dysfunctional clusters 3, 5, and 6 do not produce thalassospiramides while they appear to contain the necessary modules; however, our evolutionary analysis comparing the seven BGCs has revealed a potential explanation. We propose that, being the most complete and common, cluster 1 is likely to be the ancestral pathway inherited vertically by the Thalassospira and that clusters 2, 4, and 7 could have arisen through module/domain losses (Figs. 3 and 4), whereas clusters 3, 5, and 6 have arisen through HGT events (Fig. 4). Therefore, successful thalassospiramide production might require additional metabolic pathways that are deficient in the host strains of clusters 3, 5, and 6. This hypothesis is consistent with the deduced biosynthetic mechanisms that suggest that in trans activity of an
unknown external transferase is involved. To test this hypothesis further, we performed comparative genomics.

Comparison between the genomes of thalassospiramide-producing and -non-producing bacteria identified KEGG genes for amino acid metabolism (such as genes for glutamine and tyrosine biosynthesis), signal transduction (such as the pho family), and compound efflux (such as the MFS transporters) (Fig. 5). Although there is no evidence that glutamate is directly involved in thalassospiramide biosynthesis, glutamate metabolism is at the center of several amino acid metabolic pathways, and it affects the biosynthesis of a variety of bacterium-sourced natural products (16, 17). Enhancement of the tyrosine biosynthetic pathway can increase the production of salidrosides in engineered Escherichia coli (18). The biosynthetic steps to form the thalassospirimides directly involve tyrosine, suggesting potential linkages between tyrosine metabolic pathways and thalassospiramide biosynthesis. Genome mining efforts indicate that secondary metabolites in microbes are under tight regulation, silencing many biosynthetic gene clusters under laboratory conditions (19, 20). The pho family regulators control antibiotic biosynthesis in Streptomyces (21). Thus, the presence or absence of particular regulators in thalassospiramide-producing bacteria suggests their potential roles in controlling thalassospiramide production. In Streptomyces, modifying the MFS transporters can increase the production of secondary metabolites by up to 10-fold (21). In addition, a remarkable degree of evolutionary relatedness has been demonstrated between the transporters and biosynthetic enzymes (22). Currently, the transporters for thalassospiramide secretion have not yet been confirmed experimentally; however, the non-random distribution of transporters in the thalassospiramide-producing and non-producing bacterium genomes indicates their potential impact.

To conclude, the thalassospiramide production patterns and functional BGC architectures are consistent within the Rhodospirillaceae family, allowing a reassessment of the biosynthetic mechanisms. Particular genes/pathways for amino acid metabolism, signal transduction, and molecule transport may have important functions driving thalassospiramide production, emphasizing that molecule production via diversity-oriented biosynthesis is also impacted by genes/pathways beyond the BGCs. This study is an excellent example of the application of integrative biochemical and genomic approaches in discovery of genes and mechanisms within natural product biosynthesis.

**Experimental Procedures**

*Fermentation, Secondary Metabolite Extraction, and LC/MS Analysis—*Information on the 130 strains used for compound production examination is listed in supplemental Table S6. Each bacterial strain was cultured on solid media (yeast extract, 2 g liter⁻¹; agar, 16 g liter⁻¹; sea salt, 35 g liter⁻¹; 37 °C) for 5–7 days, and then a single colony was inoculated into a 50-ml liquid culture of GYTP medium (glucose, 10 g liter⁻¹; yeast extract, 2 g liter⁻¹; tryptone, 2.5 g liter⁻¹; peptone, 2.5 g liter⁻¹; sea salt, 35 g liter⁻¹) and grown with shaking for 3 days at 28 °C. The cultures were extracted twice with an equal volume of ethyl acetate. The organic layers were combined and dried, and the solvent was removed in vacuo. Extracts were analyzed by reverse-phase ultra-performance liquid chromatography-mass spectrometry (Waters 1.7-μm BEH C₁₈ column, 0.25 ml min⁻¹, gradient from 5% to 95% CH₃CN with 0.1% formic acid, 30 min).

*Isolation and Characterization of Thalassospirimides—*Bacterial strains were cultured in GYTP medium (40 × 1 liter) at 28 °C for 4 days. The cultures were extracted twice with an equal volume of ethyl acetate. The organic layers were combined and dried, and the solvent was removed in vacuo. The extracts were subjected to flash C₁₈ chromatography, and elution with 30%, 60%, 90%, and 100% aqueous methanol yielded four fractions. New thalassospirimides were purified from these fractions by reverse-phase HPLC (Phenomenex Luna 5-μm C₁₈, 250 × 100 mm, 100 Å, 3.0 ml min⁻¹) with an isocratic method of 55% CH₃CN in water. Electrospary ionization TOF high-resolution MS analysis was conducted using a microTOF mass spectrometer (Bruker Daltonics GmbH) in positive ion-scanning mode with a mass range of 100–2000 Da and a voltage of 4.5 kV. MS² analysis was performed using an LTQ Velos dual pressure ion trap mass spectrometer (Thermo Fisher Scientific). 1D NMR and 2D NMR spectra were obtained using a 500-MHz Varian Inova spectrometer. All samples were dissolved in methanol-d₄, and chemical shifts were reported in parts per million relative to TMS.

*Calpain Inhibitory Activity Assay—*As described previously (4, 23), the calpain inhibitory activity assay was performed using the calpain activity assay kit (BioVision) with Z-Leu-Leu-Tyr-fluoromethylketone used as a positive control. Briefly, the compounds were dissolved in DMSO and diluted in methanol to a series of concentrations. The assays were performed in 96-well plates, initiated by mixing 85 μl of extraction buffer with 0.1 units HCAN1 (BioVision). Subsequently, 10 μl of reaction buffer, 1 μl of test compound, and 5 μl of calpain substrate (Ac-LLY-AFC) were added to each well. The plate was then subjected to incubation at 37 °C for 1 h. The test samples were then recorded under an excitation wavelength of 400 nm and emission wavelength of 505 nm. Samples without compound were defined as a negative control. Three replicates were performed to generate an average value.

* Cultures of Mouse Primary Neurons—Embryonic cortical neurons were harvested from C57BL/6 E16.5 mouse embryos by established methods. On day 16.5 of gestation, pregnant mice were sacrificed and their embryos removed. Embryonic cerebral cortices were dissected into PBS containing glucose, minced with forceps, and digested with 0.25% trypsin-EDTA (Sigma-Aldrich) for 10 min at 37 °C. After digestion, tissues were rinsed in DMEM (Thermo Fisher Scientific) with 10% FBS (Sigma-Aldrich) to inactivate the trypsin, rinsed a second time in Neurobasal medium (Gibco), and triturated to produce a single-cell suspension. The cells were then plated onto poly-L-lysine-coated glass coverslips at a final density of 45,000 cells/well. The cells were then grown in Neurobasal medium supplemented with B27 (Gibco), 2 mM GlutaMAX (Gibco), 100 units ml⁻¹ penicillin (Gibco), and 100 μg ml⁻¹ streptomycin (Gibco). Cultures were maintained at 37 °C in 5% CO₂. Compound treatments were not initiated until 14 days in vitro.

*Neuroprotective Assay by Mouse Primary Neurons—*Human THP-1 cells (an immortalized human monocyte cell line, ATCC), were grown at 37 °C in a 5% CO₂ atmosphere in RPMI 1640 media supplemented with 10% heat-inactivated FBS, 100 units ml⁻¹ penicillin, 100 μg ml⁻¹ streptomycin, 100 μM 2-ME, and 50 μM 2-mercaptoethanol. The cultures were treated with 10 μM verapamil (a sodium channel blocker) or 10 μM nilutamide (a GABAA receptor antagonist) for 24 h. Cell viability was assessed using the CellTiter-Blue assay (Promega).
1640 culture medium (Gibco) containing 10% heat-inactivated FBS, 5 × 10⁻⁵ M β-mercaptoethanol (Sigma-Aldrich), 1% sodium pyruvate (Sigma-Aldrich), 100 units/ml penicillin (Gibco), and 100 μg ml⁻¹ streptomycin (Gibco). To create the THP-1 conditioned medium, 4 μl of a 2 mM fibrillar Aβ₂₅-₃₅ (Tocris) suspension was added to 24-well plates and allowed to air-dry. THP-1 cells were subsequently plated on the Aβ-coated wells at a density of 20,000/well in Neurobasal medium. After 48 h, the medium was harvested and centrifuged to remove cells and other debris and then used immediately to treat primary neurons as described previously (24).

To test the neuroprotective behavior of the thalassospiramides, the compounds were assayed at 1 μg ml⁻¹. After 24-h pretreatment, half of the culture medium in each well was removed and replaced with conditioned medium diluted 1:3 in Neurobasal medium to reach a final concentration of 12.5%. Cell cycle activity in the neuronal cultures was monitored by the incorporation of 10 μM 5-ethyl-2'-deoxyuridine. After 24 h, the cells were rinsed in buffer and fixed with 4% paraformaldehyde in 0.1 M PBS for 20 min at room temperature followed by repeated rinsing in PBS.

The 5-ethyl-2'-deoxyuridine (EdU) staining was performed using Click-iT chemistry according to the instructions of the manufacturer (Life Technologies). For immunostaining, nonspecific binding staining was blocked with 1% BSA (Sigma-Aldrich) and 0.1% Tween 20. Cells were then incubated overnight at 4°C with chicken anti-microtubule-associated protein 2 (MAP2) antibody (Abcam) diluted 1:5000 in blocking buffer Phosphate Buffered Saline Tween-20 (PBST). Cells were subsequently rinsed in PBS and then incubated with secondary antibody (A488-labeled donkey anti-chicken antibody, 1:500 in PBST) for 1 h at room temperature. Cells were counterstained with DAPI (Sigma-Aldrich) for 5 min. The fluorescently labeled cultures were mounted in Hydromount and analyzed on a fluorescence microscope (Olympus DP80). All assays were run in triplicate.

**DNA Extraction, 16S rRNA Gene Amplification, and Sequencing**—DNA extraction was performed as described previously (25). Briefly, the cultured bacterial cells were pelleted by centrifugation at 4000 × g for 10 min and then lysed with lysozyme, proteinase K, and 10% SDS. Total DNA was extracted using the AllPrep DNA/RNA mini kit (Qiagen) following the instructions of the manufacturer. PCR amplification was carried out using Phusion DNA polymerase (New England Biolabs) and the 16S rRNA gene primer pair 8F/1492R in a thermal cycler (Bio-Rad) following these steps: initial denaturation at 98°C for 30 s, 26 cycles of 98°C for 10 s, 60°C for 10 s and 72°C for 15 s, and a final extension at 72°C for 5 min. The 16S rRNA genes were sequenced in the Sanger sequencing platform before being annotated online at the NCBI website.

**Genome Sequencing, Assembly, and Analysis**—Genomic sequencing and analysis were performed using methods described in our previous study (26). The draft genomes were sequenced using the Illumina Hi-seq 2000 platform (Shanghai South Gene Technology Co.). Assembly was performed using SPAdes Genome Assembler 3.6.1 on our local server (27). The specified K values 21, 31, 41, 51, 61, 71, and 81 were used under the “careful” and “pe” options. The optional output after assembly was selected based on N₅₀ and N₉₀ of the contiguous sequences. The completeness and contamination of the genomes were estimated using CheckM (28). Gene prediction was performed with Glimmer v3.0 (29), and gene annotation was done by searching the KEGG (30) database. Adenylation domain specificity was predicted using Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) (31) and NRPS predictor 2 (32), and the condensation and ketosynthase domain functions were analyzed with NaP-DoS (33). Non-random co-occurrence patterns of KEGG genomes and the thalassospiramide BGCs were constructed according to methods described in a previous study (34). The statistical analyses were performed in the R studio environment (R 2.13) using the vegan, igraph, and hmisc packages. The cutoff for network generation was a Spearman’s correlation coefficient of >0.8 and a statistically significant p value of <0.01. The results were visualized in Gephi (version 0.8.2).

**Phylogenetic Analyses**—Single gene alignment of nearly full-length (>1300 bp) 16S rRNA genes was conducted in Molecular Evolutionary Genetics Analysis (MEGA) version 6.0 (35) using the Muscle algorithm with the following parameters: gap open penalty of -50, unweighted pair group method with arithmetic mean clustering method, and minimum diagonal length of 24. Gblocks (36) analysis was used to eliminate less informative sites in the alignments. The construction of maximum likelihood trees of 16S rRNA genes was conducted using MEGA version 6.0 with the Tamura-Nei model and the nearest neighbor interchange method with 1000 bootstrap replicates. The construction of maximum parsimony trees of 16S rRNA genes was performed using the subtree-pruning-regrafting method with 1000 bootstrap replicates.

The biosynthetic domains 2C, 6C, and 6A of all predicted thalassospiramide clusters were aligned in MEGA version 6.0 using the Muscle algorithm with the following parameters: gap open penalty of -2.9, gap extension penalty of 0, a hydrophobicity multiplier 1.2, unweighted pair group method with arithmetic mean clustering method, and minimum diagonal length of 24. Alignments of the functional gene trees were subjected to Gblocks analysis to remove less informative sites. Maximum likelihood trees were built using the Jones-Taylor-Thornton model and the nearest neighbor interchange method with 1000 bootstrap replicates. Maximum parsimony trees were built using the site rate ranking (SRR) method with 1000 bootstrap replicates. Automated Phylogenomic Inference Application (AMPHORA) (37) was used to predict 31 conserved single-copy genes (tsf, smpB, rpsS, rpsM, rpsK, rpsJ, rpsL, rpsE, rpsC, rpsB, rpoB, rpmA, rplT, rplS, rplP, rplN, rplM, rplL, rplK, rplF, rplE, rplD, rplC, rplB, rplA, pyrG, pgk, nusA, infC, frr, and dnaG). Subsequently, 28 genes (rplL, rplT, and rpmA could not be detected in all genomes) were used to construct the tree. The same parameters used in the construction of the functional gene trees were used.

**Statistical Analyses**—Significant differences between the effects of the tested compounds and the negative controls in calpain inhibitory and neuroprotective assays were assessed using Student’s t tests. Genes with significantly changed numbers in the 18 genomes of thalassospiramide-producing bacteria and the 14 genomes of non-producing bacteria were also identified using Student’s t test (p < 0.001).
The Diversity-oriented Biosynthesis of Thalassospiramides

Author Contributions—P. Y. Q. designed the experiments. W. Z., L. L., and A. C. R. carried out most data analyses. Z. S. and Q. L. provided the samples. Z. L., Y. X., B. Z., and K. H. performed the experiments. L. L., W. Z., A. R., B. S. M., and P. Y. Q. wrote the manuscript.

Acknowledgments—We thank Dr. Nadine Ziemert for discussions regarding the phylogenetic approaches, Prof. William Fenical for providing the strain Thalassospira sp. CNJ-328, and Pui Shuen Wong for performing the MS analysis.

References

1. Newman, D. J., and Cragg, G. M. (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J. Nat. Prod. 75, 311–335

2. Cotter, P. D., Hill, C., and Ross, R. P. (2005) Bacteriocins: developing innate immunity for food. Nat. Rev. Microbiol. 3, 777–788

3. Oh, D. C., Strangman, W. K., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2007) Thalassospiramides A and B, immunosuppressive peptides from the marine bacterium Thalassospira sp. Org. Lett. 9, 1525–1528

4. Ross, A. C., Xu, Y., Lu, L., Kersten, R. D., Shao, Z., Al-Suwailem, A. M., Dorrestein, P. C., Qian, P. Y., and Moore, B. S. (2013) Biosynthetic multi-tasking facilitates thalassospiramide structural diversity in marine bacteria. J. Am. Chem. Soc. 135, 1155–1162

5. Saatman, K. E., Creed, J., and Raghupathi, R. (2010) Calpains as a therapeutic target in traumatic brain injury. Neurotherapeutics 7, 31–42

6. Morton, J. D., Lee, H. Y., McDermott, J. D., Robertson, L. J., Bickerstaffe, R., Jones, M. A., Coxon, J. M., and Abell, A. D. (2013) Amacrocyclic calpain inhibitor slows the development of inherited cortical cataracts in a sheep model. Invest. Ophthalmol. Vis. Sci. 54, 389–395

7. Leloup, L., and Wells, A. (2011) Calpains as potential anti-cancer targets. Expert Opin. Ther. Targets 15, 309–323

8. Herrup, K., and Yang, Y. (2007) Cell cycle regulation in the postmitotic neuron: oxymoron or new biology? Nat. Rev. Neurosci. 8, 368–378

9. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Juli, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., et al. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. 30, 108–160

10. Meiser, P., Weissman, K. J., Bode, H. B., Krug, D., Dorrestein, P. C., Xu, Y., Moore, B. S., and Qian, P. Y. (2015) Mechanism of action of thalassospiramides, a new class of calpain inhibitors. Sci. Rep. 5, 8873

11. Wu, Q., Combs, C., Cannady, S. B., Geldmacher, D. S., and Herrup, K. (2000) β-Amyloid activated microglia induce cell cycling and cell death in cultured cortical neurons. Neurobiol. Aging 21, 797–806

12. Zhang, W., Wang, Y., Tian, R. M., Bougouffa, S., Yang, B., Cao, H. L., Zhang, G., Wang, Y. H., Xu, W., Batang, Z., Al-Suwailem, A., Zhang, X. X., and Qian, P. Y. (2014) Species sorting during biofilm assembly by artificial substrates deployed in a cold seep system. Sci. Rep. 4, 6064

13. Zhang, W., Wang, Y., Bougouffa, S., Tian, R., Cao, H., Li, Y., Cai, L., Wong, Y. H., Zhang, G., Zhou, G., Zhang, X., Bajic, V. B., Al-Suwailem, A., and Qian, P. Y. (2015) Synchronized dynamics of bacterial niche-specific functions during biofilm development in a cold seep brine pool. Environ. Microbiol. 17, 4089–4104

14. Bankewich, A., Nurk, S., Antipov, D., Gurevich, A. V., Dvorkin, M., Kulikov, A. S., Lesin, V. N., Nikolenko, S. I., Pham, S., Prijibelski, A. D., Pshyk, A. V., Sirokin, A. V., Yavahhi, N., Tesler, G., Aleksyev, M. A., and Pevzner, P. A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477

15. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055

16. Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23, 673–679

17. Kang, H., and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30

18. Weber, T., Binn, K., Duderla, S., Krug, D., Kim, H. U., Brucoceri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlbelen, W., Breiting, R., Takano, E., and Medema, M. H. (2015) antiSMASH 3.0: a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237–W243

19. Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPS predictor 2-a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39, W362–W367

20. Ziemert, N., Podell, S., Cote, J. H., Allen, E., and Jensen, P. R. (2015) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS ONE 7, e43064

21. Freiich, S., Zarecki, R., Elam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R., and Ruppin, E. (2011) Competitive and cooperative metabolic interactions in bacterial communities. Nat. Commun. 2, 589

22. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729

23. Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552

24. Wu, M., and Eisen, J. A. (2008) A simple, fast, and accurate method of phylogenomic inference. Genome Biol. 9, 151