Chemical and taxonomic diversity of marine Gram-negative bacteria
Nicholas Lorig-Roach

Notes

**Formatting of taxonomic rankings:** the author has adopted the convention in which only the genus and species names of an organism are italicized. All other taxonomic ranking are capitalized.

**Reference and compound numbering:** for the author’s convenience, compounds and their associated references share the same number and were labeled arbitrarily rather than in order of appearance. This breach of convention made data organization much simpler and simplifies finding the references associated with a given compound.
# Table of Contents

Abstract 4

Introduction 4

**Background**

**Bacterial and chemical diversity of marine, non-cyano Gram-negative bacteria** 4

- Table 1. Gram-negative phyla. 5
- Compounds from Aquificae, Verrucomicrobia, and Chloroflexi 6
  - Figure 1. *Sole marine natural products from...* 6
- Figure 2. *Phylogenetic tree of cultured marine Gram-negative bacteria with known secondary metabolism* 7
- Compounds from Bacteroidetes 8
  - Figure 3. *Representatives of known natural products from...* 8
- Compounds from Deltaproteobacteria—Myxococcales 9
  - Figure 4. *Secondary metabolites of marine Myxococcales* 9
- Compounds from Alphaproteobacteria 9
  - Figure 5. *Selected examples of chemical scaffolds from...* 10
- Compounds from Gammaproteobacteria 11
  - Vibronales – *Photobacterium* 11
    - Figure 6. *Secondary metabolites from Photobacterium.* 11
  - Other Vibronales 11
    - Figure 7. *Natural products from marine Vibrio species.* 12
  - Alteromondales 12
    - Figure 8. *Selection of marine Alteromondales metabolites.* 13
  - Pseudomondales and Oceanospirillales 14
    - Figure 9. *Pseudomondales scaffolds.* 14
    - Figure 10. *Compounds isolated from Oceanospirillales.* 14
- Ecological abundance – learning from the metagenome 15
  - Polyketide synthases, non-ribosomal peptide synthases and other biosynthetic machinery 15
    - Figure 11. *Examples of PKS natural products* 16
    - Figure 12. *Schematic of PKS synthesis and FAS crystal structure* 16
- Biosynthetic potential of bacteria 17
  - Figure 13. *Correlation of bacterial genome size and number of genes* 17
  - Table 2. antiSMASH analyses of a selection of bacterial genomes 18

References 27

Supporting information S1
Introduction

Background
Small molecule natural products have a long history of utility for humans, providing a multitude of therapeutics for a variety of indications. In fact, over half of all FDA approved drugs released between 1981 and 2010 are either natural products or synthetic derivatives of natural products. One of the greatest issues with natural products research is limited supply of source material and questionable ethics (and efficacy) of large-scale sample collections. Bacteria are an ideal solution to this because they can be cultured in the laboratory and genetically manipulated with relative ease. Some bacterial taxa are relatively well examined, particularly Actinomycetes and Cyanobacteria, but subsets of marine bacteria have not been explored exhaustively for their biosynthetic potential despite evidence that exploration would be fruitful. Many of the natural products produced by marine invertebrates are believed to be synthesized by symbiotic or commensal microorganism populations, and there are numerous examples of structural parallels and even identical chemical entities discovered from both macro- and microorganisms. Of particular interest are the Myxococcales, a biosynthetically prolific order with few described marine strains. The myxobacteria have numerous unique properties: unusual gliding motility, fruiting body formation, predatory behavior, and the current record for largest bacterial genome. In addition to the myxobacteria, marine Gram-negative bacteria in general (other than Cyanobacteria) are relatively unexplored, with only about a hundred structural motifs known.

Bacterial and chemical diversity of marine, non-cyano Gram-negative bacteria
Bacteria that stain Gram-negative come from diverse and deeply diverging phyla adapted to a multitude of environments; they can be extremophiles, chemotrophs, symbionts, pathogens, or free living. The number of species known from each phyla ranges from single digits (e.g. Caldiserica) to well over 1,000 (proteobacteria). Some phyla have no known marine representatives (e.g. Elusimicrobia, termite gut symbionts) while the enormous Proteobacteria clade is ubiquitous in both marine and terrestrial environments. Due to differences in abundance, ecological niche, and ease of culturability, relatively few Gram-negative phyla from the marine environment have known secondary metabolites that the author is aware of. Namely Aquificae, Bacteroidetes, Chlamydiae/Verrucomicrobia, Chloroflexi, and the Proteobacteria are known to produce natural products (table 1). However, among these few bacterial orders with known compounds are two strains that produce molecules that have advanced to clinical trials for the treatment of cancer: Tistrella mobilis, producer of the didemnins, and a relative of Vibrio mediterranei that produces the kahalalides. Both of these compounds were originally isolated from a macroorganism; the didemnins were first described as products of a tunicate while the kahalalides were found in an herbivorous mollusk.
Table 1. Gram-negative phyla and their constituent classes and orders. Bolded orders have at least one marine representative with known secondary metabolism. Underlined orders were isolated as part of this work. Grey phyla have no marine representatives with known chemistry. This table was compiled using information from Bergey's manual, the List of Prokaryotes with Standing Nomenclature (LPSN), and NCBI Taxonomy in addition to references 1-95.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Class (Orders)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aquificae</td>
<td>Aquificae (Aquificales, <strong>Desulfobacteriales</strong>)</td>
</tr>
<tr>
<td>Armatimonadetes</td>
<td>Armatimonadetes (Armatimonadales); Chthonomonadetes (Chthonomonadales);</td>
</tr>
<tr>
<td></td>
<td><strong>Fimbrimonas (Fimbrimonas)</strong></td>
</tr>
<tr>
<td>Bacteroidetes/Chlorobi</td>
<td>Bacteroidetes (Bacterioidia, <strong>Cytophagia, Flavobacteria, Sphingobacteria</strong>);</td>
</tr>
<tr>
<td></td>
<td>Chlorobi (Chlorobia); <strong>Ignavibacteriaceae (Ignavibacteria)</strong></td>
</tr>
<tr>
<td>Caldiserica</td>
<td>Caldisericia (Caldisericales)</td>
</tr>
<tr>
<td>Chlamydiae/Verrucomicrobia</td>
<td>Chlamydiae (Chlamydiia); Lentisphaerae (Lentiisphaeria, Oligosphaeria);</td>
</tr>
<tr>
<td></td>
<td><strong>Verrucomicrobia (Opitutae, Spartobacteria, Verrucomicrobia)</strong></td>
</tr>
<tr>
<td>Chloroflexi</td>
<td>Anaerolineae (Anaerolineae); Ardenticatenia (Ardenticatenales); Caldilineae</td>
</tr>
<tr>
<td></td>
<td>(Caldilineae); Chloroflexia (Chloroflexales, <strong>Herpetosipionales</strong>, Kallotenuales);</td>
</tr>
<tr>
<td></td>
<td>Dehalococcoidia (Dehalococcoidales, Dehalogenimonas); Ktedonobacteria</td>
</tr>
<tr>
<td></td>
<td>(Ktedonobacterales, Thermogemmatisporales); Thermoflexia (Thermoflexales);</td>
</tr>
<tr>
<td></td>
<td>Thermomicrobia (Sphaerobacteriae, Thermomicrobiales)</td>
</tr>
<tr>
<td>Chrysiogenetes</td>
<td>Chrysiogenetes (Chrysiogenales)</td>
</tr>
<tr>
<td>Deferribacteres</td>
<td>Deferribacteres (Deferribacterales)</td>
</tr>
<tr>
<td>Deinococcus-Thermus</td>
<td>Deinococci (Deinococcales, <strong>Thermales</strong>)</td>
</tr>
<tr>
<td>Dictyoglomi</td>
<td>Dictyoglophia (Dictyoglomales)</td>
</tr>
<tr>
<td>Elusimicrobia</td>
<td>Elusimicrobia (Elusibicrobiales); <strong>Endomicrobia</strong></td>
</tr>
<tr>
<td>Fibrobacteres/ Acidobacteria</td>
<td>Acidobacteria (Acidobacteriia, Holophagae, Solibacteres); Fibrobacteres (Chitinivirionia, Fibrobacteria); <strong>Marinimicrobia</strong></td>
</tr>
<tr>
<td>Fusobacteria</td>
<td>Fusobacteriaceae (Fusobacteriales)</td>
</tr>
<tr>
<td>Gemmatimonadetes</td>
<td>Gemmatimonadetes (Gemmatimonadales)</td>
</tr>
<tr>
<td>Nitrospinae</td>
<td>Nitrospina (Nitrospinales)</td>
</tr>
<tr>
<td>Nitrospira</td>
<td>Nitrospira (Nitrospirales)</td>
</tr>
<tr>
<td>Planctomycetes</td>
<td><strong>Phycisphaerae (Phycisphaerales)</strong>; Planctomycetia (Planctomycetales)</td>
</tr>
<tr>
<td><strong>Proteobacteria</strong></td>
<td>α (Caulobacteriales, Kiloniellales, Kopriromonadales, Kordiimonadales,</td>
</tr>
<tr>
<td></td>
<td>Magnetococcales, Parvularculales, **Rhizobiales, Rhodobacteriales,</td>
</tr>
<tr>
<td></td>
<td><strong>Rhodospirillales</strong>, Rhodothalassiales, Rickettsiales, Sneathielloides,</td>
</tr>
<tr>
<td></td>
<td><strong>Sphingomonadales</strong>); β (Burkholderiales Ferritrophicales, Ferrovailes, Gallionellales, Hydrogenophilales, Methylophilales, Neisseriales, Nitrosonomadales, Procabacteriales, Rhodocyclales); γ (Acidithiobacillales, Aeromonadales, <strong>Alteromonadales</strong>, Cardiobacteriales, Chromatiales, Enterobacteriales, Legionellae, Methylcoccales, <strong>Oceanospirillales</strong>, Orbales, Pasteurellales, <strong>Pseudomonadales</strong>, Salmonella, Cholera, Thioplovibacteriales, <strong>Vibionales</strong>, Xanthomonadales); δ (Bdellovibriionales, Desulfarculales, Desulfovibacterales, Desulfurellales, Desulfuromonadales, <strong>Myxococcales</strong>, Syntrophobacterales); ε (Campylobacterales, Nautiliales)</td>
</tr>
<tr>
<td>Spirochaetes</td>
<td>Spirochaetia (Spirochaetales)</td>
</tr>
<tr>
<td>Synergistetes</td>
<td>Synergistia (Synergistales)</td>
</tr>
<tr>
<td>Thermodesulfobacteria</td>
<td>Thermodesulfobacteriaceae (Thermodesulfobacteriales)</td>
</tr>
<tr>
<td>Thermotagae</td>
<td>Thermotagae (Thermotagales)</td>
</tr>
</tbody>
</table>
The phylogenetic relationships of secondary metabolite producing strains of marine Gram-negative bacteria were inferred using the neighbor joining method and are shown in figure 2. When possible, the true 16S rRNA sequence was used, unfortunately in many papers bacteria are inadequately identified or the sequencing results are not shared publicly so a type strain best matching the strain reported was used as a placeholder. In this phylogeny, families grouped together well, but there are some unexpected branching patterns: the Alpha-, Beta-, and Gammaproteobacteria appear more closely related to Bacteroidetes than the Deltaproteobacteria, which cluster with phyla generally considered to have diverged long ago in bacterial evolution (Aquificae, Chlamydia/Verrucomicrobia, and the Chloroflexi). This may be in part due the use of the Firmicutes (phylum that includes the genus *Bacillus*) as the out-group—when the Archaea are used to root all bacterial phyla in either a 16S or 23S phylogeny, the Firmicutes diverge at an intermediate branch between the Proteobacteria/Bacteroidetes and the apparently more ancient Chloroflexi and Aquificae phyla. Even with this caveat, the divergent clustering of the Deltaproteobacteria from other Proteobacteria is not satisfactorily explained. Close inspection of the maximum likelihood tree produced by the ‘All-Species Living Tree’ project shows some aberrant clustering of Acidobacteria representatives within Proteobacteria, but otherwise the Proteobacteria classes form a single clade. This indicates that the phylogeny in figure 2 may need additional sequences from the under-represented classes to improve the overall tree topology.

Within bacterial phyla known to produce natural products, the order Gammaproteobacteria is by far the most represented with 59 or more distinct compounds known. Alphaproteobacteria follow with at least 16 structures, then the orders of the phylum Bacteroidetes with nine, the Deltaproteobacteria with six, and the remaining phyla have limited if any secondary metabolites known from marine strains. Possible explanations for this disparity could include a potential abundance of isolated Gammaproteobacteria (i.e. a high percentage of bacteria in researchers’ samples are Gammaproteobacteria), ease of culturability with current methods, or higher biosynthetic output. The author will attempt to address these possibilities in the coming presentation of structural diversity versus taxonomic diversity, metagenomic studies from the literature, and original analyses of bacterial genomes.

**Compounds from Aquificae, Chloroflexi, and Verrucomicrobia**

Perhaps the most impressive feat amongst cultured marine Gram-negatives is the successful isolation and cultivation of *Thermovibrio ammonificans*, a hydrothermal vent associated bacterium sampled from a 2500m depth, and subsequent discovery of a 3-4µM inducer of apoptosis (11). Despite the extreme depth of its natural home, the researchers were able to grow the bacterium at a mere 2 atm (vs 250 atm at collection site) and 75°C. Siphonazole (81) is a member of a new structural class isolated from a member of the Chloroflexi phylum; the isolation of the bacterium was not well described, but the large scale culture was in a simple medium with casein digest and glucose as the primary carbon sources. A new member of the Verrucomicrobia phylum was isolated from a sponge homogenate (*Halichondria okdai*) plated on agar containing only an antifungal (cycloheximide) and antibiotic (rifampicin); from this bacterium a novel carotenoid (34) was obtained.

![ammonifin C (11)](image1)

*Thermovibrio ammonificans*

![diapolyycenedioic acid xyllosyl ester A (34)](image2)

*Rubritalea squalificiens MBIC08254*

![siphonazole (81)](image3)

*Herpetosiphon sp. 060*

**Figure 1.** Sole marine natural products reported from the phyla Aquificae, Chloroflexi, and Verrucomicrobia
Figure 2. Neighbor-joining tree of marine Gram-negative bacteria with known secondary metabolites and strains isolated as part of this work (red and blue). The class of bacteria is labeled and the corresponding branch colorized. Numbers in parentheses preceding strains are entries with a single asterisk are reference strains from NCBI’s RefSeq database. Strains without an asterisk are used when only genus was reported. Entries with two asterisks are placeholder IDs refer to compounds produced (structures are drawn in the upcoming sections). Strains without an asterisk correspond to real 16S rRNA sequence from a natural products manuscript. Entries with a single asterisk are reference strains from NCBI’s RefSeq with homology to a reported producing strain. Strains with two asterisks are placeholders used when only genus was reported.

De sulfurobacteriales
Verrucomicrobiae
Chloroflexia

β-proteobacteria

Cytophagia

Flavobacteria

Sphingobacteria

α-proteobacteria

γ-proteobacteria
Compounds from Bacteroidetes

There are relatively few secondary metabolites known from marine Bacteroidetes, but among the known products there is appreciable structural diversity. The single known scaffold from the Sphingobacteria class is a putative diterpene without reported activity (60); the strain was obtained from the ATCC and grown at 40L scale in a peptone and yeast extract based medium. Two separate but closely related *Rapidithrix* species isolated from near-shore silt in Japan and seaweed in Thailand respectively produced an antibiotic ariakemicin (17) with 0.8µM activity against *S. aureus* and a 5µM acetylcholine esterase inhibitor marinoquinoline (51). The two strains were grown in relatively distinct media and harvested uniquely, 17 from ethanol extract of freeze dried fermentation in Zobell Marine Broth (casein digest and glucose as primary carbon source) and 51 from the methanol extract of XAD-16 resin from culture in SK media (casein digest, beef extract, yeast extract, glycerol, and pyruvate). An acylglycine compound (57) that inhibited N-type calcium channel at ~2µM was isolated from a *Cytophaga* sp. from seawater in Japan that was grown in a casein digest, yeast extract, glycerol and succinate based media. From the Flavobacteriia class, a number of nitro substituted aromatic compounds (18) with mild antibiotic activities were isolated from the ethyl acetate extract of a fermentation of an arctic bacterium grown in casein digest / yeast extract media, some of the chlorinated compounds are new chemical entities. Also of note is the inducer of algal morphogenesis, thalussin (87), isolated from the supernatant of a culture of a bacterium in Zobell Marine Broth passed through an HP-20 column. The producing strain is mis-identified as *Cytophaga* sp.; by the 16S sequence provided by the researchers, the bacterium is certainly a relative of Flavobacteriia. Details regarding the growth and isolation of the remaining structures are available in SI table S10.

**Sphingobacteria**

Neoverrucosan-5B,9B,18B-triol (60)
*Saprospira grandis* ATCC 23116

**Cytophagia**

ariakemicin A (17)
*Rapidithrix* sp. HC35

cyclic polysulfides (32a-32b)
*Cytophaga* sp. BIO137 and BIO 138

N-(3-acyloxy)acylglycine (57)
*Cytophaga* sp. SANK 71996

**Flavobacteria**

flavoristamid A (39)
*Flavobacterium* sp.

myxol (56)
*Flavobacterium* sp. P99-3

N,N-diacyl-tryptamine (58)
*Roseivirga echnicomitas* KMM 6058T

thallusin (87)
*Cytophaga* sp. MBIC 04683

**Figure 3.** Representatives of known natural products from marine Bacteroidetes.
Compounds from Deltaproteobacteria--Myxococcales

The properties of secondary metabolites from marine Myxococcales are well reviewed elsewhere. Most noteworthy is the actin filament stabilizer miuraenamide A (54), which resembles the potent sponge-derived compound jasplakinolide, and the structurally unique salimabromide (78). All of the marine myxobacteria strains were reportedly isolated from sediment with the exception of the haliangin (40) producer which was isolated from seaweed. These environmental samples were processed in the traditional manner of Myxococcales isolation: plated on nutrient deficient agar streaked or spotted with live E. coli bait. Also of note is that adsorbent resins (either XAD-16 or SP207) were used in every case to extract metabolites from culture broth.

![Figure 4. Secondary metabolites of marine Myxococcales.](image)

Compounds from Alphaproteobacteria

The Alphaproteobacteria produce several structurally interesting molecules as well as the first marine natural product to be evaluated clinically, didemnin B (35). Didemnin B, originally isolated from a tunicate, is a potent anti-viral and anti-tumor agent that was determined to be a product of a member of the Rhodospirilalles class by two independent research groups. However, the gene cluster responsible for didemnin B biosynthesis resides on a large plasmid that may be present in a number of other species. Two research groups successfully isolated didemnin B using either XAD-7 resin or ethyl acetate, however the yields using XAD-7 were much higher (3.2 mg/L vs 0.2 mg/L) with slight differences in growth media. Neither group defined their bacterial isolation methods well. Other compounds of note include: B-90063, a structurally unique inhibitor of endothelin converting enzyme; agrochelin A (6), a potent cytotoxic agent (50-300 nM) from a marine agrobacterium isolated from the homogenate of a tunicate that was serially diluted and plated on Zobell Marine Broth agar plates; a cholic acid derivative (30), from a bacterium isolated from ground ascidian siphons diluted in seawater and plated on a starch based agar plate; erythrazole (37) and erythrolic acid (38) low micromolar cytotoxic compounds produced by an Erythrobacter sp. obtained from brackish sediment added to sterile water that was centrifuged and supernatant subsequently plated on humic acid media; and thalassospiramide (86) an inhibitor of IL-5. Information about the remaining compounds can be found in SI table S3.
Figure 5. Selected examples of chemical scaffolds from marine Alphaproteobacteria.
Compounds from Gammaproteobacteria

Compounds from Gammaproteobacteria—Vibrionales, Photobacterium

Natural products from marine Vibrionales have been reviewed by a number of authors. The majority of secondary metabolites from the Vibrionales class are from the genus Vibrio, and the remaining metabolites are produced by Photobacterium species. Among the Photobacterium scaffolds are: the pyrrothine antibiotic holomycin (42); the thiazolidine/dehydro thiazole siderophore that was poorly named as piscibactin (95) despite only a two carbon difference from the known yersiniabactin that was also isolated; the phenazine pigment phototeridine (71); and several cyclic depsipeptides (46, 61, 82, 90). The isolation of these bacteria was poorly described, but the author's experience with this fast-growing genus suggests many methodologies would be effective. Each was grown in similar media, generally containing glucose and casein digest in some form of artificial seawater. Most commonly the fermentations were processed by centrifugation followed by extraction of either the supernatant or the cell pellet. The activity of these compounds is largely not notable. Holomycin is a relatively broad spectrum antibiotic, but is not unique to Gammaproteobacteria—the same compound is produced by Actinomycetales. The depsipeptides ngercheumicin (61) and solonamide (82) as well as the di-ether (2) are implicated in quorum sensing, while the kailuin (46) class was reported to have cytotoxic and anti-dinoflagellate properties.

Figure 6. Secondary metabolites from Photobacterium.

Compounds from Gammaproteobacteria—Other Vibrionales

As with the Photobacterium genus, in most literature reports the isolation of Vibrio species was not well described beyond the source of the environmental sample. The most significant compound isolated from Vibrio is kahalalide F, a potent anti-tumor compound in clinical trials for use against solid tumors. Kahalalide F was originally isolated from a mollusk, Elysia rufescens, that feeds on algae. To cultivate the producing bacterium, researchers ground a piece of E. rufescens, diluted in sterile seawater, and plated onto a variety of agar plate types. The Vibrio sp. was grown in Zobell Marine Broth and harvested by centrifugation followed by lyophilization of the supernatant and extraction with ethanol. The boron containing AI-2 (7) is unique but was not isolated—instead it was co-crystallized with the luxP protein in a structural biology effort. Halogenated diphenyl ethers (24) are known pollutants of human origin, yet were controversially shown to also be produced by microorganisms. The bulk of Vibrio species were grown large scale in a medium based on casein digest and yeast extract, sometimes containing glucose or another sugar, or starch. Extraction was typically performed via centrifugation followed by ethanol or acetone extraction of the pellet, or alternatively
ethyl acetate or butanol extraction of the supernatant. In some cases supernatants were extracted with an adsorbent resin. See SI table S5 for more information.

Figure 7. Natural products from marine Vibrio species.

Compounds from Gammaproteobacteria--Alteromonales

The Alteromonales class, which includes the genera Alteromonas and Pseudoalteromonas, produces a large portion of the known Gammaproteobacteria secondary metabolites. A number of these compounds have notable cytotoxicity: alteramide A (8) is a relative of the sponge compounds cylindramide, discodermide, and geodin with 0.2-10 µM IC50 vs. a number of murine and human tumor cell lines; the pseudoalteromones and prodigiosin also exhibited low µM IC50 against cancer cells; and MC21-A (53), koromicin (47), pseudomonic acid (75) and thiomarinol (88) had demonstrated antimicrobial properties. Many of the Alteromonales species were associated with macroorganisms including sponges, algae/seaweed, and other invertebrates. Tetrodotoxin, a potent neurotoxin infamously found in the pufferfish is reportedly produced by a number of Gammaproteobacteria including Pseudoalteromonas species. However, experimental evidence is often weak, lacking mass spectrometry data or NMR validation and instead relying on retention time comparisons to a standard, and occasionally bioassay. Both techniques certainly contribute valuable evidence but are not conclusive alone. Alteromonales species were commonly cultivated in Zobell Marine Broth, and both ethyl acetate and centrifugation based extraction methods were popular. One strain
listed here, I-L-33, that produced pentabromopseudulalin (68a) and other brominated pyrroles (68b,c) was originally reported as *Chromobacterium* sp., a member of the Betaproteobacteria.\(^8\) Later analysis cast some doubt on I-L-33's classification, suggesting the bacteria resembled Altermonadales species, though the strain may truly have been a Betaproteobacteria.\(^9\) Additional details about the Altermonadales are available in SI table S6 and previously published reviews.\(^9, 10\)

**Figure 8.** Selection of marine Altermonadales metabolites.
Compounds from Gammaproteobacteria—Pseudomondales and Oceanospirillales

Andrimid (13) and a poorly named analog, moiramide (55), are two antibiotic molecules produced by many bacteria including Pseudomonas members. Tambjamine (83) also is questionably named, apparently an analog of another ubiquitous metabolite, prodigiosin. Massetolide (52) had an ~8 µM MIC vs M. tuberculosis and was isolated from two Pseudomonas strains, one from a seaweed and another from a tubeworm. In both cases the natural product 52 was extracted from large-scale culture on agar. The activity of the remaining compounds was limited or not well defined. Where specified, isolation methods involved dilution of an environmental sample in sterile (sea)water followed by plating. Extraction procedures are comparable to those of previous sections.

**Figure 9. Pseudomondales scaffolds.**

**Figure 10. Compounds isolated from Oceanospirillales.**
Ecological abundance -- learning from the metagenome

Advances in genome sequencing technologies and development of new techniques in microbial ecology have afforded a better view of diversity and relative abundance of bacteria in the oceans.\textsuperscript{110, 111, 112, 113} In brief, libraries of 16S ribosomal RNA genes amplified by polymerase chain reaction (PCR) from environmental samples using promiscuous primers can be cloned into \textit{E. coli} or other hosts, separated, and sequenced. Alternatively, genomic DNA of all bacteria in a sample can be isolated and this ‘metagenome’ sequenced, assembled, and sorted into its constituent members. Both techniques have biases and limitations: the PCR based method is biased toward bacteria whose 16S rRNA has higher affinity for the primers used, while the metagenomic method may not have adequate sequence coverage to accurately represent rare strains.\textsuperscript{110} Acknowledging these limitations, these microbial ecology studies have demonstrated that microbial communities from different ocean niches (e.g. surface water, shallow sediment, deep ocean, aphotic sediment, macro-organism associated) are varied and stratified.\textsuperscript{112, 113}

Although there is significant variation between environments, members of the phylum Proteobacteria are often the dominant members of a community, with control oscillating between Alpha-, Beta-, and Gammaproteobacteria. While their presence is common, Delta- and Epsilonproteobacteria are much less abundant than the other Proteobacteria orders (however, Epsilonproteobacteria are sometimes the dominant clade near hydrothermal deposits).\textsuperscript{110, 112, 113} Depending on the sample, Actinobacteria, Acidobacteria, Bacteroidetes, Chlamydia/Verrucomicrobia, Chloroflexi, and Firmicutes may have significant representation. These studies in part support the idea that ecological abundance has caused the apparent over-representation of secondary metabolites from marine Gammaproteobacteria, however amenability to laboratory culture and raw biosynthetic potential remain to be addressed. In fact, assuming equal biosynthetic potential (a poor assumption), the previously cited references indicate many more Alpha- and Betaproteobacteria producers ought to be known from surface water, shallow sediments, and some marine organisms. This could mean Alpha- and Betaproteobacteria are not cultured easily with current methods, or simply do not produce as many metabolites. In either case, development of bacterial isolation methods that select against the ubiquitous members of the Proteobacteria phyla as part of a natural products pipeline seems prudent.

Polyketide synthases, non-ribosomal peptide synthases, and other biosynthetic machinery

Small molecule natural products are commonly produced by several classes of biosynthetic gene families: terpene synthases, nonribosomal peptide synthases (NRPS), and polyketide synthases (PKS).\textsuperscript{97, 114, 115} The products of these large multi-protein synthases are also commonly modified by tailoring enzymes or other metabolic proteins. One method of identifying potentially productive bacterial isolates is through analysis of their genomes and the biosynthetic gene clusters within that encode secondary metabolite synthases.

Polyketide synthases are a family of multi-module enzymes that synthesize products typically with malonyl-CoA units.\textsuperscript{115} These polyketides can be both primary and secondary metabolites, and are found in many organisms including bacteria, plants, fungi, and humans.\textsuperscript{116} Polyketide synthases involved in primary metabolite production are the most ubiquitous, as they produce fatty acids from simpler two-carbon molecules.\textsuperscript{116} However, PKSs have been adapted to produce more complex secondary metabolites as well (figure 11, numerous examples in the preceding figures). These secondary metabolites have the most promise in terms of potential therapeutics or research tools because they all putatively serve some function for the producing organism.
Examples of polyketide natural products from Myxococcales. A: epothilone core structure from Sorangium cellulosum, a tubulin inhibitor, derivatives are currently in clinical trials for cancer. B: myxothiazol, PK from Myxococcus spp., a coenzyme Q inhibitor in mitochondrial cytochrome bc1 complex, known metabolite of strain 088A04A. C: myxovirescin A, putative inhibitor of type II signal peptidase, metabolite of strain M123SB105A.

As is implied by the diversity of compounds produced by PKSs, there is significant diversity in the structure of the PKS proteins. All PKSs do share certain features: modules work together to perform the synthesis of the polyketide product. All functional PKSs contain acyltransferase (AT), acyl carrier protein (ACP), and ketosynthase (KS) modules. Each module has a specific role in the synthetic scheme, such as catalyzing reactions or acting as a scaffold to hold the growing polyketide. Figure 2 shows a schematic of the step-by-step nature of polyketide synthesis and a stylized crystal structure of a fatty acid synthase, highlighting its modularity and complexity.

The fact that all polyketide synthases must have KS and AT genes means that probing for highly conserved regions of these modules can determine the presence or absence of a PKS. Ideally, this analysis would be performed in silico on a fully sequenced genome (see next section), but it is also possible to use PCR based methods using primers that amplify the conserved regions of the biosynthetic modules. However, potential for amplification of non-functional PKS genes as well as off-target genes is possible. This is the basis for part of the experiments performed in this work (see results, 'Probing for PKS genes').

Non-ribosomal peptide synthases operate in a similar fashion to polyketide synthases and are often hybridized with PKS modules. The principal difference is that NRPS modules can use the entire pool of amino acids as starter units which are added to the growing scaffold via ATP dependent condensation reactions. Non-ribosomal peptide synthases also use different carrier proteins (peptidyl or aryl carrier proteins) and alternate catalytic enzymes that introduce additional functionality to the starter units.

Terpene synthases draw from yet another pool of starter units, all of which are monomers or polymers of isopentenyl diphosphate (IPP) or dimethylallyl diphosphate (DMAPP). Terpene synthases are often composed from a core set of protein domains and are in that sense modular, but not in the assembly-line fashion of PKS or NRPS enzymes. Importantly, polyketide, non-ribosomal peptide, terpene, and other synthases share sufficient homology that biosynthetic gene clusters from evolutionarily distant species can be discovered using genomic search tools like BLASTp and antiSMASH.
Biosynthetic potential of bacteria

Although the evolutionary origin of biosynthetic genes are often ubiquitous enzyme families first involved in primary metabolism (e.g. PKS from fatty acid synthases), not all bacteria have equal amounts of genes used for secondary metabolism.\textsuperscript{116} Among bacteria with complete genomes in NCBI’s repository (approx 3000 bacteria, with significant bias towards some groups and species), the average genome size is near 4 million bases (Mb) and ranges from under 1 Mb (a Tenericutes sp., NC_013511) all the way to 14.8 Mb (Sorangium cellulosum So0157-2, NC_021658, a member of Myxococcales in the phylum Proteobacteria).\textsuperscript{100} The number of genes in bacterial genomes is linearly correlated to their size (figure 13), but the percentage of those genomes with homology to biosynthetic gene clusters as detected by antiSMASH, varies from 0% to over 30% (Table 2). antiSMASH is a tool developed to find secondary metabolite gene clusters in genomic sequence using statistical analyses of the query sequence versus a database of known biosynthetic proteins and their constituent domains.\textsuperscript{118}

A selection of genomes from different bacterial phyla was analyzed with antiSMASH, using the most promiscuous settings available (i.e. all of the protein types, including putative biosynthetic proteins, were searched for). Summaries of these analyses are shown in Table 2, and the relevant accession numbers are available in SI table S11. Within each phylum analyzed, the largest genome within the phylum was selected unless a relative of a known marine species had been sequenced. Every available Myxococcales genome was analyzed while only a limited set of representatives from other phyla was selected, which may not accurately reflect that clade’s biosynthetic potential. The Actinomycetes, Firmicutes, and Cyanobacteria are included as benchmarks, and it is clear from even the limited examples in Table 2 why these bacteria are the subject of intense natural product isolation efforts: 10%-30% of their genomes are dedicated to a diverse secondary metabolism with numerous large gene clusters. The majority of the remaining strains analyzed had genomes with less than 10% homology to biosynthetic genes and few had more than 2 large (>50kb) gene clusters, if any at all. The exception to this are the Myxococcales, which have as many or more gene clusters than the Gram-positive organisms, and a significant number of additional large clusters. The majority of biosynthetic genes were either NRPS or hybrid synthases including non-ribosomal peptide synthase modules (the ‘other’ category includes numerous gene types).

There are a number of issues not addressed by the metrics in Table 2. One is the uniqueness of the gene clusters, it is hypothesized that many clusters originate from gene duplication events followed by differentiation. Thus a genome may harbor numerous closely related clusters in addition to the possibility of related bacteria sharing common clusters. Additionally, strains that appear non-productive relative to biosynthetically talented bacteria like the Actinobacteria and the Myxococcales may nonetheless harbor a unique gene cluster and thus produce a unique natural product.

[Figure 13. Plot of bacterial genome size vs predicted number of genes.\textsuperscript{119}]

\textsuperscript{116}\textsuperscript{118}\textsuperscript{119}
Table 2. Overview of antiSMASH analyses of bacterial genomes with representatives from well-sequenced phyla. Accession numbers are available in SI table S11.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Organism</th>
<th>Genome size</th>
<th>Identified gene clusters</th>
<th>Total Gene Clusters</th>
<th>Total Clusters (incl. putative)</th>
<th>Number of gene clusters &gt;50kb</th>
<th>% genome biosynth. genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actin.</td>
<td><em>Salmonella typhimurium</em> CNB-440</td>
<td>5.2 Mb</td>
<td>6 / 5 / 2 / 5</td>
<td>18</td>
<td>37</td>
<td>9</td>
<td>30.3</td>
</tr>
<tr>
<td>Actin.</td>
<td><em>Streptomyces coelicolor</em> A3</td>
<td>8.7 Mb</td>
<td>4 / 7 / 4 / 10</td>
<td>25</td>
<td>54</td>
<td>6</td>
<td>18.4</td>
</tr>
<tr>
<td>Firmi.</td>
<td><em>Pseudomonas aeruginosa</em> K02</td>
<td>8.8 Mb</td>
<td>6 / 2 / 2 / 3</td>
<td>15</td>
<td>34</td>
<td>9</td>
<td>13.9</td>
</tr>
<tr>
<td>Firmi.</td>
<td><em>Clostridium butyricum</em> NCIMB 8052</td>
<td>6.0 Mb</td>
<td>1 / 0 / 0 / 1</td>
<td>2</td>
<td>14</td>
<td>1</td>
<td>4.1</td>
</tr>
<tr>
<td>Cyan.</td>
<td><em>Nostoc punctiforme</em> PCC 73102</td>
<td>9.1 Mb</td>
<td>7 / 4 / 4 / 11</td>
<td>26</td>
<td>47</td>
<td>11</td>
<td>17.2</td>
</tr>
<tr>
<td>Cyan.</td>
<td><em>Microcystis aeruginosa</em> NIES-843</td>
<td>5.8 Mb</td>
<td>4 / 1 / 2 / 2</td>
<td>11</td>
<td>20</td>
<td>7</td>
<td>10.8</td>
</tr>
<tr>
<td>Act.</td>
<td><em>Granulicella mallensis</em> MP5ACTX8</td>
<td>6.2 Mb</td>
<td>3 / 2 / 3 / 4</td>
<td>12</td>
<td>28</td>
<td>4</td>
<td>12.4</td>
</tr>
<tr>
<td>Act.</td>
<td><em>Thermotoga maritima</em> HB-1</td>
<td>1.8 Mb</td>
<td>0 / 0 / 0 / 0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1.8</td>
</tr>
<tr>
<td>Act.</td>
<td><em>Bacillus halodurans</em> P 36-108</td>
<td>4.0 Mb</td>
<td>0 / 0 / 0 / 0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>3.2</td>
</tr>
<tr>
<td>Act.</td>
<td><em>Marinivibrio frunsei</em> DSM 4126</td>
<td>4.5 Mb</td>
<td>0 / 1 / 2 / 0</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>4.2</td>
</tr>
<tr>
<td>Act.</td>
<td><em>Cellulophaga lytica</em> DSM 7489</td>
<td>3.8 Mb</td>
<td>0 / 0 / 1 / 0</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>4.4</td>
</tr>
<tr>
<td>Act.</td>
<td><em>Zobellia galactanivorans</em></td>
<td>5.5 Mb</td>
<td>0 / 2 / 2 / 0</td>
<td>4</td>
<td>21</td>
<td>1</td>
<td>7.1</td>
</tr>
<tr>
<td>Act.</td>
<td><em>Chitinophaga pinensis</em> DSM 2588</td>
<td>9.1 Mb</td>
<td>3 / 2 / 6 / 6</td>
<td>17</td>
<td>38</td>
<td>5</td>
<td>13.4</td>
</tr>
<tr>
<td>Ch.</td>
<td><em>Chlorobacterium thiosulfatophilum</em> ATCC 35110</td>
<td>3.3 Mb</td>
<td>0 / 0 / 1 / 1</td>
<td>2</td>
<td>13</td>
<td>0</td>
<td>5.0</td>
</tr>
<tr>
<td>Ch.</td>
<td><em>Herpetosiphon aurantiacus</em> DSM 785</td>
<td>6.8 Mb</td>
<td>8 / 2 / 2 / 2</td>
<td>14</td>
<td>29</td>
<td>6</td>
<td>15.5</td>
</tr>
<tr>
<td>Pl.</td>
<td><em>Pirulilla salteri</em> DSM 6068</td>
<td>6.2 Mb</td>
<td>0 / 0 / 3 / 2</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>4.4</td>
</tr>
<tr>
<td>Pl.</td>
<td><em>Agrobacterium radiobacter</em> K84</td>
<td>7.3 Mb</td>
<td>2 / 1 / 1 / 4</td>
<td>8</td>
<td>33</td>
<td>1</td>
<td>8.5</td>
</tr>
<tr>
<td>Pl.</td>
<td><em>Pseudonocardia gelatinis</em> DSM 26640</td>
<td>4.5 Mb</td>
<td>0 / 1 / 0 / 3</td>
<td>4</td>
<td>17</td>
<td>1</td>
<td>5.8</td>
</tr>
<tr>
<td>Pl.</td>
<td><em>Tistriola mobilis</em> KA081020-065</td>
<td>6.5 Mb</td>
<td>4 / 1 / 1 / 12</td>
<td>18</td>
<td>45</td>
<td>2</td>
<td>14.7</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Burkholderia xenovorans</em> LB400</td>
<td>9.7 Mb</td>
<td>1 / 0 / 3 / 5</td>
<td>9</td>
<td>49</td>
<td>1</td>
<td>10.0</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Chromobacterium violaceum</em> ATCC 12472</td>
<td>4.8 Mb</td>
<td>2 / 0 / 1 / 4</td>
<td>7</td>
<td>18</td>
<td>1</td>
<td>8.3</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Pseudoalteromonas haloplanktis</em> TAC125</td>
<td>3.9 Mb</td>
<td>0 / 0 / 0 / 1</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>3.6</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Habalia chrysophila</em> KCTC 2396</td>
<td>7.2 Mb</td>
<td>7 / 0 / 0 / 4</td>
<td>11</td>
<td>24</td>
<td>4</td>
<td>9.6</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Pseudomonas fluorescens</em> F113</td>
<td>6.8 Mb</td>
<td>3 / 1 / 0 / 5</td>
<td>9</td>
<td>33</td>
<td>2</td>
<td>11.1</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Photobacterium profundum</em> SS9</td>
<td>6.4 Mb</td>
<td>2 / 1 / 0 / 2</td>
<td>5</td>
<td>14</td>
<td>3</td>
<td>6.0</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Vibrio paraoxybacillus</em> RMID 2210633</td>
<td>5.2 Mb</td>
<td>0 / 0 / 0 / 3</td>
<td>3</td>
<td>15</td>
<td>0</td>
<td>3.6</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Vibrio campbellii</em> ATCC BAA-1116</td>
<td>6.1 Mb</td>
<td>2 / 0 / 1 / 2</td>
<td>4</td>
<td>11</td>
<td>1</td>
<td>3.8</td>
</tr>
<tr>
<td>Prot.</td>
<td><em>Xanthomonas campestris</em> 85-10</td>
<td>5.4 Mb</td>
<td>1 / 0 / 0 / 2</td>
<td>3</td>
<td>17</td>
<td>0</td>
<td>4.7</td>
</tr>
<tr>
<td>Spi.</td>
<td><em>Spiorochaeta sponarinoha</em> DSM 11293</td>
<td>4.7 Mb</td>
<td>0 / 0 / 0 / 2</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>3.1</td>
</tr>
</tbody>
</table>

*The Actinobacteria, Firmicutes, and Cyanobacteria are included for comparative purposes. Phyla with fewer than 10 fully sequenced genomes were omitted from this analysis. *Class shown for proteobacteria; abbreviations: Ac., Acidobacteria; Aqu., Aquificae; Ch., Chlorobi; Cx., Chloroflexi; Pl., Planctomycetes; Sp., Spirochaetes. *Hybrid gene clusters incl. NRPS were counted in the NRPS category, hybrids with PKS but without NRPS were counted as PKS, etc. *Other* includes all other named gene clusters detected by antiSMASH. *Entries with 5 or more clusters > 50kb in size are bold. *Entries with >10% of genome homologous to biosynthetic gene clusters are bold.
References

Please disregard the formatting inconsistencies of this section.


4. See 1.


77. See 26.


79. See 36.


C.


100. Among all fully assembled prokaryotic genomes in NCBI's genome database as of November 2014. (https://www.ncbi.nlm.nih.gov/genome/browse/)
   B: Vol. 2A, p58.


119. This figure was made by downloading search results from NCBI's Genome browser (see ref 100) for all complete and assembled bacterial genomes. The tab-separated values file was opened in Excel, and plotting genome size vs. 'genes.'


