UNIVERSITY OF CALIFORNIA SAN DIEGO

Blurring the Lines Between Natural and Synthetic: The Biosynthetic Chemistry of Marine Actinomycete Bacteria

A dissertation submitted in partial satisfaction of the

Requirements for the Degree Doctor of Philosophy

in

Biomedical Sciences

by

Charles Bradford Larson

Committee in charge:

Professor Bradley S. Moore, Chair
Professor Victor Nizet, Co-Chair
Professor Michael Burkart
Professor Tracy Handel
Professor Susan Taylor

2019

The dissertation of Charles Bradford Larson is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2019

Dedication

In loving memory of William Larson

Table of Contents

Signature Page..... iii

Dedication..... iv

Table of Contents..... v

List of Figures..... viii

List of Tables..... xi

Acknowledgements..... xii

Vita..... xiv

Abstract of the Dissertation xvi

Chapter 1: Introduction to the Dissertation.....1

1.1 Natural Products as Traditional Medicines……..……...…...1

1.2 Microbial Natural Products……..…….....3

1.3 Technological Advancements in Analytical Chemistry During the Golden Age……..... 5

1.4 Marine Natural Products…….. 8

1.5 Actinomycetes Contribution to the Natural Products Pharmacopeia……..... 11

1.6 Natural Products Biosynthesis: Connecting Genes to Molecules…….. 13

1.7 DNA Sequencing and Natural Products Biosynthesis…….. 15

1.8 High Throughput and Whole Genome Sequencing Reveals Huge Biosynthetic Potential…… 18

1.9 Heterologous Expression of Natural Products…….. 19

1.10 Opportunities in Modern Natural Products Discovery and Overview of Dissertation Chapters..... 20

1.11 References for Chapter 1..... 24

Chapter 2: Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols…….....34

2.1 Introduction to Chapter 2..... 34

2.2 Chapter 2 Introduction References..... 37

2.3 Reprint of ": Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols"..... 39

Chapter 3: PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete……..... 64

3.1 Introduction to Chapter 3..... 64

3.2 Chapter 3 Introduction References..... 66

3.3 Reprint of "PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete".....67

Chapter 4: Marine Actinobacteria *Salinispora pacifica* Biosynthetic Gene Cluster Produces Phosphonate Degradation Products of Glyphosate…….....99

4.1 Abstract.....99

4.2 Introduction.....100

4.2.1 Phosphorus in Nature and the Marine Phosphorus Cycle.....101

4.2.2 Anthropogenic Phosphonate Contamination in the Environment.....104

4.3 Results.....106

4.3.1 Identification of Salinispora PepM Sequences.....106

4.3.2 Heterologous Expression of S. pacifica Phosphonate BGC.....107

4.3.3 Purification of Compounds 1-3.....108

4.3.4 Cloning and Expression of Phosphonate N-acetyltransferases....117

4.3.5 Distribution of PepM (SppJ) Homologs in the Marine Environment.....121

4.4 Discussion.....122

4.5 Methods/Materials.....125

4.5.1 Genome Mining.....125

4.5.2 Analysis of the distribution of PepM homologs.....125

4.5.3 Cloning and Plasmid Assembly.....125

4.5.4 Heterologous Expression.....127

4.5.5 Purification of Phosphonates Produced in Heterologous Host.....127

4.5.6 NMR and MS/MS Analysis.....128
4.5.7 Cloning and Purification of GNATs from Phosphonate Cluster.....128
4.6 References for Chapter 4.....130

List of Figures

Figure 1.1: Bioactive chemicals from traditional plant medicines.....3

Figure 1.2: Drugs from the Golden Age of natural products discovery.....5

Figure 1.3: Marine Natural Products.....12

Figure 2.1: Molecular network of all generated extracts.....40

Figure 2.2: Effects of genus and species on molecular diversity-network analysis.....41

Figure 2.3: Effects of additional attributes on molecular diversity-network analysis.....43

Figure 2.4: Time-dependent changes in natural product distribution in *Salinispora arenicola* CNH-877, grown in ISP2.....45

Supplementary Figure S2.1: Molecular Network of All *Streptomyces* Extracts Color Coded By Culture Medium.....53

Supplementary Figure S2.2: Molecular Network of All Acquired Extracts Color Coded By Genus.....54

Supplementary Figure S2.3: Molecular Network of All *Salinispora* Extracts Color Coded By Species.....55

Supplementary Figure S2.4: Illustration of the major drivers of chemical speciation within the *Salinispora* data.....56

Supplementary Figure S2.5: Molecular Network of All Generated Extracts Color Coded by Extraction Solvent.....57

Supplementary Figure S2.6: Molecular Network of *Salinispora* Extracts from Different Media Phases.....58

Supplementary Figure S2.7: Comparison of MS/MS spectra of arenicolide A (upper spectrum, m/z: 827.489,) and a novel, formally hydrated arenicolide analogue (lower spectrum, m/z: 845.501) from the ethyl acetate extract of *Salinispora pacifica* CNT-138.....59

Supplementary Figure S2.8: Molecular Network of All Acquired Extracts Color Coded By Strain Location.....60

Supplementary Figure S2.9: Molecular Network of *Salinispora arenicola* CNH877 strains grown in four different liquid media and extracted at three different time points.....61

Figure 3.1: A selection of anthracyclines approved for treatment of cancer.....64

Figure 3.2: Capture vector assembly.....68

Figure 3.3: Gene map of biosynthetic gene clusters associated with the biosynthesis of cosmomycins C and D.....68

Figure 3.4: Cosmomycin cluster identified by molecular networking.....69

Supplementary Figure S3.1: PCR screening of multiple yeast colonies.....73

Supplementary Figure S3.2: PCR screening of individual yeast colonies.....74
Supplementary Figure S3.3: Sequencing of integrated cosmomycin pathway.....75
Supplementary Figure S3.4 PCR amplification of integrated cosmomycin pathway.....76
Supplementary Figure S3.5: MS2 fragmentation analysis of cosmomycin C (1).....77
Supplementary Figure S3.6: MS2 fragmentation analysis of cosmomycin D (2).....78
Supplementary Figure S3.7: MS1 spectrum of cosmomycin analog 3.....79
Supplementary Figure S3.8: MS2 fragmentation analysis of cosmomycin analog 3.....80
Supplementary Figure S3.9: MS1 spectrum of cosmomycin analog A.....81
Supplementary Figure S3.10: MS2 fragmentation analysis of cosmomycin analog A....82
Supplementary Figure S3.11: MS2 fragmentation analysis of cosmomycin analog B....83
Supplementary Figure S3.12: ^{1}H NMR spectrum of cosmomycin C (1)..... 84
Supplementary Figure S3.13: ^{1}H NMR spectrum of A..... 85
Supplementary Figure S3.14: Overlaid 1H NMR spectrum of 1 and A.....86
Supplementary Figure S3.15: HSQC spectrum of 1.....87
Supplementary Figure S3.16: HSQC NMR spectrum of A.....88
Supplementary Figure S3.17: COSY NMR spectrum of 1.....89
Supplementary Figure S3.18: Molecular network of cosmomycin native producers and heterologous host.....90
Supplementary Figure S3.19: *Streptomyces* sp. CNT-302 molecular network.....91
Figure 4.1: Depiction of a *Salinispora* colony, the *S. pacifica* BGC and its small molecule products.....100
Figure 4.2: Biosynthetic scheme for P-C bond installation.....104
Figure 4.3: Depiction of phosphonate BGCs.....107
Figure 4.4: Crude preparation of heterologous host fermentation.....109
Figure 4.5: ^{1}H NMR spectra of both heterologous host and *S. coelicolor* M1152 extracts purified by chromatography.....111
Figure 4.6: Broad band decoupling of the NMR spectrum of the enriched fraction of heterologous host extracts reveals the influence of a phosphorus nucleus on three peaks.....113
Figure 4.7: ^{1}H/^{31}P two-dimensional NMR spectrum (HMBC) of enriched preparation of heterologous host extracts.....114
Figure 4.8: ^{1}H NMR analysis of HILIC purified products of fermentation.....115

Figure 4.9: MS/MS analysis of HILIC purified metabolites.....116

Figure 4.10: ^{31}P NMR analysis of heterologous host extracts spiked with synthetic phosphonate standards.....117

Figure 4.11: MS analysis of GNAT enzyme assays with AMPA substrate.....119

Figure 4.12: MS analysis of GNAT enzyme assays with 2-AEP substrate.....120

Figure 4.13: Sequence Similarity Network of *S. pacifica* PepM homologs found in marine genomes and metagenomes.....122

Figure 4.14: Putative biosynthetic pathway for the production of na-AMPA.....124

Supplementary Figure S4.1: Gel electrophoresis analysis of phosphonate BGC amplicons prior to assembly.....135

Supplementary Figure S4.2: Gel digest of assembled *S. pacifica* BGC in the pCAP03 plasmid.....136

Supplementary Figure S4.3: HRMS data for m/z=110, including molecular formula calculation.....137

Supplementary Figure S4.4: HRMS data for m/z=125, including molecular formula calculation.....138

Supplementary Figure S4.5: HRMS data for m/z=152, including molecular formula calculation.....139

Supplementary Figure S4.6: ^{1}H NMR analysis of spiking experiments with naAMPA, AMPA, and 2-HEP synthetic standards.....140

Supplementary Figure S4.7: NMR Analysis of Crude Extract.....141

Supplementary Figure S4.8: Protein gel (10% acrylamide) of HisTrap purified GNAT enzymes.....142

Supplementary Figure S4.9: Amplification of small regions of the *S. pacifica* phosphonate BGC from genomic DNA isolated from the heterologous host.....143

List of Tables

Supplementary Table S2.1: Overview of the bacterial strains used in this study.....51

Supplementary Table S2.2: Overview of the compounds detected in this study.....52

Supplementary Table S2.3: Information about detected staurosporine analogues in the ISP2 time course network.....62

Supplementary Table S3.1: ^{1}H NMR chemical shift assignments of compounds 1 and 4 and the related cosmomycin analogs obelmycin C4 and A447C/D.....92

Supplementary Table S3.2: Comparison of *Streptomyces* sp. CNT-302 and *S. olindensis* BGCs.....95

Supplementary Table S4.1: Genes in the *Salinispora pacifica* phosphonate BGC and their closest homolog determined by NCBI BLAST.....144

Supplementary Table S4.2: Multiple Sequence Alignment of *S. pacifica* phosphonate BGC GNATs and *B. licheniformis* GNAT.....146

Supplementary Table S4.3: *S. pacifica* PepM homologs in marine genomes and metagenomes in the MAR DB database.....147

Supplementary Table S4.4: Color coded legend for sequence similarity network.....201

Supplementary Table S4.5: Primers used for the amplification of the *S. pacifica* phosphonate BGC and specific GNATs from the cluster.....228

Acknowledgements

Many thanks to my mentor Brad Moore, who offered me the opportunity to perform my doctoral studies in his laboratory and provided the support and mentorship necessary to complete my studies. His kind but rigorous and thorough instruction and mentorship has shaped me into a competent natural products chemist. I would also like to thank my Co-Advisor, Victor Nizet, who offered to liaison with the BMS department which allowed me to work at the beautiful laboratory at the Scripps Institute of Oceanography I have called home for the last several years. Additionally, I would like to thank the wonderful natural products community at SIO. Members of the Jensen, Gerwick, Hughes, Fenical, and Allen laboratories provided an environment that encouraged curiosity, collaboration, and fun. Many thanks to the post-doctoral researchers who aided my progress through the years, including Dr. Yuta Kudo, Dr. Shaun McKinnie, Dr. Hanna Luhavaya, Dr. Jonathan Chekan, Dr. Ellis O'Neill, Dr. Peter Jordan, and Dr. Max Cruseman. I would also like to thank my family for their loving encouragement and support throughout my journey in science.

Chapter 2, in full, is a reprint of materials as it appears in "Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols" in *Journal of Natural Products*, 2016, Max Crusman, Ellis C. O'neill, Charles B. Larson, Alexey V. Melnick, Dimitrios J. Floros, Ricardo R. da Silva, Paul R. Jensen, Pieter C. Dorrestein, Bradley S. Moore. The dissertation author was a secondary investigator and author of this paper.

Chapter 3, in full, is a reprint of materials as it appears in "PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin

Biosynthetic Gene Cluster in an Ocean Streptomycete" in *Journal of Natural Products*, 2017, Charles Bradford Larson, Max Crusman, Bradley S. Moore. The dissertation author was the primary investigator and author of this paper.

Vita

2012-2019 Ph.D., University of California, San Diego

2006-2010 B.S., University of Michigan

Publications

2019 O'Neill E.C., Schorn M., Larson C.B., Millán-Aguiñaga N. Targeted antibiotic discovery through biosynthesis-associated resistance determinants: target directed genome mining. *Crit Rev Microbiol*. 2019 Apr 15:1-23.

2017 Larson C.B., Crüseman M., and Moore B.S. PCR-independent Method of Transformation Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete. *J Nat Prod*. 2017, **80**, 1200-1204.

Crüsemann M., O'Neill E.C., Larson C.B., Melnik A.V., Floros D.J., da Silva R.R., Jensen P.R., Dorrestein P.C., Moore B.S. Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols. *J Nat Prod*. 2017, **80**, 588-597.

2016 Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein

PC, Bandeira N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking Nature Biotech. 2016, **34**, 828-837.

2011 Zhu J.*, Larson C.B.*, Ramaker M.A., Quandt K., Wendte J.M., Ku K.P., Chen F., Jourdian G.W., Vemulapalli R., Schurig G.G., He Y. Characterization of Recombinant *B. abortus* Strain RB51SOD towards Understanding the Uncorrelated Innate and Adaptive Immune Responses Induced by RB51SOD Compared to Its Parent Vaccine Strain RB51 (* co-first authors) *Front. Cell. Inf. Microbio*. 2011, **1**, 1-10.

2008 Xiang Z., Todd T., Ku K.P., Kovacic B.L., Larson C.B., Chen F., Hodges A.P., Tian Y., Olenzek E.A., Zhao B., Colby L.A., Rush H.G., Gilsdorf J.R., Jourdian G.W., He Y. VIOLIN: Vaccine Investigation and Online Information Network *Nucleic Acids Res*. 2008, **36**, D923-928.

Abstract of the Dissertation

Blurring the Lines Between Natural and Synthetic: The Biosynthetic Chemistry of Marine Actinomycete Bacteria

by

Charles Bradford Larson

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2019

Professor Bradley S. Moore, Chair

Professor Victor Nizet, Co-Chair

Nature is the source of an incredible diversity of complex chemistry. The wide variety of chemical scaffolds produced as secondary metabolites have been the source

of many useful medicinal therapies for the treatment of human disease, in addition to providing useful tools for probing biology. The biosynthetic machinery that produces these molecules is encoded in the genomes of organisms, and recent technological advancements in bioinformatics, DNA sequencing, and gene cloning and capture techniques offer researchers new opportunities for natural products discovery. When combined with modern analytical chemistry techniques, these methods are a powerful engine for compound discovery. The marine environment represents a unique well of biodiversity to investigate with this new methodology, with a unique evolutionary history ancient than terrestrial habitats. Chapter 2 of this dissertation describes the in-depth analysis of a collection of 146 marine actinomycetes by mass spectrometry and molecular networking to investigate their biosynthetic capacity. These bacteria were cultured in a variety of media, and their extracts were used to generate approximately 1.8 million mass spectra. Analysis of this large data set was accomplished by the Global Natural Product Social Molecular Networking platform, which allowed for the identification of known metabolites and revealed information about the production of novel metabolites. Chapter 3 of this dissertation presents a more focused study on a single biosynthetic gene cluster from an ocean streptomycete, which produces a suite of glycosylated anthracyclines. In addition to revealing the biosynthetic source of the cosmomycins and describing novel analogs of these compounds, this study describes a novel PCR-independent gene capture method utilizing the recent advances in gene synthesis technology. In Chapter 4 of this dissertation, a phosphonate biosynthetic gene cluster from another marine actinomycete, *Salinispora pacifica*, was investigated for small molecule production. Analysis of this gene cluster revealed the production of glyphosate degradation products, previously assumed to be anthropogenic pollutants.

Bacterial production of these molecules has important implications for environmental study and monitoring of phosphonate pollutants.

Chapter 1: Introduction to the Dissertation

1.1 Natural Products as Traditional Medicines

Natural products encompass, in the broadest sense, any material produced by biology. This includes biotic materials like wood, materials such as bioplastics or cornstarch, bodily fluids, and other natural materials that were once found in living organisms. This rather broad definition can be winnowed down to include only organic compounds synthesized by a living organism. This definition could refer to primary metabolites, molecules such as nucleic acids, amino acids, sugars, and fatty acids that are essential for the survival of an organism. For our purposes, the most restrictive definition of 'natural product' will be used, including only secondary or specialized metabolites. These are organic molecules not essential to survival, but rather compounds that increase the competitiveness of the organism within its environment. Secondary metabolites generally have an extrinsic function that mainly affects other organisms outside of the producer. These molecules have many applications, including: functioning as compounds essential for intercell communication or quorum sensing; for reproduction; for cellular development such as sporulation; and for antagonism in the case of antibiotics and toxins.[1]

Natural products have formed the bedrock of medicinal therapies for thousands of years. The first recorded use of plant compounds as medicines comes from Mesopotamian cuneiform tablets, which describe the use of oils of cedar trees (*Cedrus* species) and cypress trees (*Cupressus sempervirens*), licorice (*Glycyrrhiza glabra*), myrrh from *Commiphora* species, and juice of the poppy flower *Papaver somniferum*.[2] Ancient Egyptian pharmaceutical records dating from 1500 BCE describe the use of 700

drugs and their formulations such as pills, ointments, and infusions.[3] Traditional Chinese medical records can be dated to 1100 BCE, and they represent yet another large body of knowledge describing treatments for human disease using natural preparations.[4] Indeed, the use of traditional plant medicines has been extensively documented in cultures from all over the world.[5–9]

In many cases, these traditional medicinal practices and plants have led to the discovery of important pharmaceutical compounds that are still relevant today (Figure 1). Investigations performed by pharmacist Freidrich Serturner on the flowering plant *Papaver somniferum* yielded a suite of alkaloids including the commercially important morphine in 1803.[10] Later this natural product would serve as the precursor to the more potent diacetylmorphine (heroin) and the painkiller codeine, both of which remain clinical drugs. The willow bark preparations used in a variety of folk medicine traditions yielded their active ingredient, salicin, which would serve as the natural product scaffold for the salicylate medicines including acetylsalicylate (aspirin).[11] The antimalarial drug quinine, approved for clinical use by the FDA in 2004, was isolated from the bark of *Cinchona succirubra*. This plant had been used for centuries for the treatment of malaria, fever, indigestion, mouth and throat diseases, and cancer.[12] Another antimalarial success story involves the plant *Artemisia annua*, used for millennia in traditional Chinese medicine for the treatment of fevers. This plant is the source of the potent antimalarial drug artemisinin and its derivatives, effective in the treatment of drug resistant malaria strains.[13]

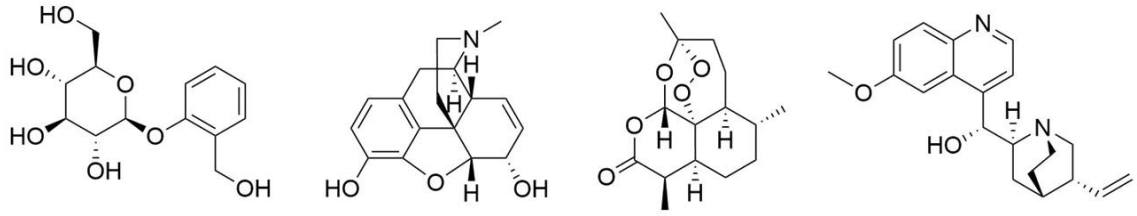


Figure 1.1: Bioactive chemicals from traditional plant medicines. The metabolites depicted in this figure were isolated from traditional medicine sources from a variety of cultures. Despite their primitive origins, these metabolites are still employed as clinical therapies today.

1.2 Microbial Natural Products

The beginning of the modern age of natural products can be dated to 1929, when Alexander Fleming reported his discovery of the antibiotic penicillin.[14] Rather than investigating traditional plant medicines, Fleming serendipitously observed antibacterial biological activity and followed that activity to its small molecule source. The humble 'mold juice' from the fungus *Penicillium notatum* would prove to be the world's first effective antibiotic. His work began a new paradigm of research, investigating microbial sources for bioactive molecules.

Due to the success of penicillin, many scientific research efforts were focused on the isolation of natural products from microbial sources. The scientific and financial foundations supporting such programs would be initiated before and during the Second World War.[15] In the post-war period, this trend led to the birth of the modern pharmaceutical industry, which focused efforts on the research of new molecules with antibiotic activity.[16] These efforts proved fruitful, and during this "Golden Age of

Antibiotics" a staggering number of novel bioactive metabolites were discovered in microbes.[17] During this period, roughly between 1940 and 1970, the majority of the antibiotic chemical scaffolds currently in clinical use were discovered.[18] Some of these molecules' intricate chemical structures are depicted in Figure 2.

The powerful compounds discovered during this time include: antibacterial agents, such as the penicillins, cephalosporins, aminoglycosides, and tetracyclines; immunosuppressive agents, such as the cyclosporins and rapamycin; cholesterol-lowering agents, such as mevastatin and lovastatin; and anthelminthics and antiparasitic drugs, such as the ivermectins.[19] Antitumor antibiotics discovered during this time are still important cancer chemotherapeutic agents, which include members of the anthracycline, bleomycin, actinomycin, mitomycin, and aureolic acid families.[20] Clinically useful agents from these families are the daunomycin-related agents, daunomycin itself, doxorubicin, idarubicin, and epirubicin; the glycopeptidic bleomycins A2 and B2 (Blenoxane®); peptolides such as d-actinomycin; the mitosanes; and the glycosylated anthracenone, mithramycin. These metabolites were isolated from various *Streptomyces* species, in addition two other clinically active agents, streptozocin and deoxycoformycin.[21] Calicheamicin, possibly the most potent antitumor compound to be approved for clinical use, initially proved too toxic for the clinic, in spite of its exquisite sub-picomolar activity.[22] The compound has since been linked to a specific monoclonal antibody directed against chronic myelogenous leukemia, and recently has been approved for clinical use. In this case, the compound is carried to the site where needed and released *in situ*, thereby reducing the systemic toxicity to tolerable levels without a loss in efficacy.[23]

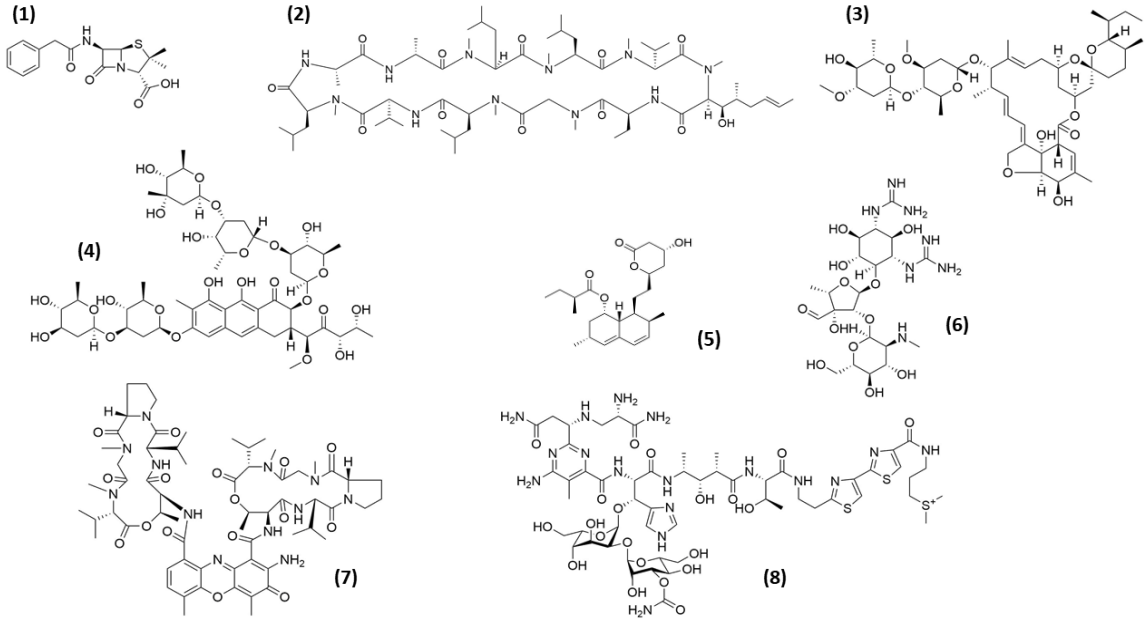


Figure 1.2: Drugs from the Golden Age of natural products discovery

(1) Penicillin G, an antibiotic used for the treatment of a wide range of bacterial infections; **(2)** Cyclosporine, an immunosuppressant used to treat organ transplant rejection, psoriasis, and severe rheumatoid arthritis; **(3)** Ivermectin, an antiparasitic used to treat head lice, scabies, river blindness (onchocerciasis), strongyloidiasis, trichuriasis, and lymphatic filariasis; **(4)** mithramycin A, an antineoplastic antibiotic used for the treatment of testicular cancer, chronic myeloid leukemia **(5)** lovastatin, a statin used for the treatment of dyslipidemia and the prevention of cardiovascular disease **(6)** streptomycin, an aminoglycoside antibiotic used for the treatment of bacterial infections including tuberculosis **(7)** Actinomycin D, a chemotherapeutic agent used for the treatment of various cancers; **(8)** bleomycin, a peptidic anti-cancer drug

1.3 Technological Advancements in Analytical Chemistry During the Golden Age

It is important to note that this flurry of antibiotics discovery was facilitated not just by the hard work of individual researchers, but was also a result of rapidly improving

analytical chemistry techniques. Predating antibiotic research by many decades, the first reproducible and accurate quantitative analytical chemistry instrument was developed in 1830 by Justus Liebig. Known as the Kaliapparat, it could precisely measure the carbon, oxygen, and hydrogen content of samples.[24] Liebig would concurrently perfect the combustion analysis and flame analysis methods developed by Joseph Louis Gay-Lussac.[25] Quantitative analysis methods would continue to improve throughout the 1800s, including the field of spectroscopy. Although the spectrum of light was first characterized by Isaac Newton in 1704, it wasn't until 1860 that Gustav Kirchoff and chemist Robert Bunsen would use light spectra to identify different elements. Their work would spawn the field of analytical spectroscopy, which would eventually lead to the development of infrared and ultraviolet spectroscopy. The first fully automated infrared spectrometer would be introduced in 1937, and the technology would continue maturing throughout the 20th century with advancements like detectors covering the spectrum from UV to visible light and diode arrays able to simultaneously scan the full spectrum of wavelengths in seconds.[26] This fast, non-destructive analysis would become standard in natural products laboratories and facilitate the discovery of thousands of new metabolites.

Mass spectrometry, an analytical technique that ionizes chemical species and sorts the ions based on their mass-to-charge ratio, would become another essential tool for the natural products researcher. The first experiments that would lead to the modern mass spectrometer were performed in the late 1800's. Gas filled discharge tubes, electrodes encased in a gas-filled and insulated envelope, would be used to study positive and negative "rays." These positive and negative "rays" could be deflected by magnetic fields, and it was this insight that would lead to the discovery of chemical

isotopes in 1913 when J.J. Thompson used the magnetic deflection of positive rays to describe the two isotopes of neon, neon-20 and neon-22.[27] Thompson and his colleague would go on to develop the first mass spectrometer based on this experiment, soon describing 212 naturally occurring isotopes with their instrument. Although the mass spectrometer soon found a home in physics labs for analyzing and separating isotopes, it wasn't until the 1960's that new ionization techniques for analysis of larger non-elemental molecules would be developed. These new methods, such as chemical ionization and field-desorption ionization, would open mass spectrometry to analytical chemistry applications. Further refinements of the technology like electrospray ionization, Fourier-transform ion cyclotron resonance, matrix-assisted laser desorption ionization, and the ion trap would continue to improve the scope and usefulness of the technique.[28] These improvements would enable the chemical formula determination of unknown metabolites, in addition to fragmenting molecules to aid in structural determination. Useful in analysis of both pure samples and complex mixtures, mass spectrometry remains one of the most powerful analytical techniques for elucidating the structure of unknown natural products.

Nuclear magnetic resonance spectroscopy (NMR), one of the few applications of quantum mechanical theory, is another experimental method that would become crucial to natural products research. The technology relies on the detection of radio frequencies generated by the spin of atomic nuclei when they are subjected to a strong magnetic field. The first measurement of the magnetic moments of atomic nuclei would take place in 1938,[29] but it would take many years for scientists to observe the changes in atomic nucleus Larmor frequencies due to the chemical bonding state of the atoms. The analysis of these chemical shifts and spin-spin coupling led to the idea of using nuclear

magnetic resonance to analyze and identify materials,[30] and the first NMR machine would be constructed in 1952.[31] Early iterations of NMR machines suffered from poor signal-to-noise ratios, but improvements such as signal averaging, strong superconducting magnets, Fourier-Transform operations, and non-uniform sampling would deliver the sensitive and accurate instruments used today. Additionally, the advancement of 2-dimensional NMR would greatly improve the utility of NMR experiments by providing data about the connectivity and spatial arrangements between nuclei.[32] Modern NMR techniques are irreplaceable and essential to the identification and structural elucidation of novel natural products.

1.4 Marine Natural Products

Terrestrial environments had proven to be a rich source of important NPs throughout the 'Golden Age' from the 1940s through the 1970's. However, as the field matured, problems began to arise with the traditional discovery pipeline. The standard methods involving the isolation of bacterial strains followed by fermentation and bioassay guided fractionation resulted in fewer novel carbon scaffolds and research efforts instead produced many "re-discovered" compounds. This would eventually lead to pharmaceutical companies abandoning nearly all of their natural products research efforts by the early 2000's.[33]

While terrestrial natural product discoveries were slowing down, new avenues of biochemical diversity were being opened during this period. Oceans cover over 70% of the earth's surface, and they represent a tremendous source of biodiversity for the discovery of bioactive metabolites. At the present time, 26 of the 28 major animal phyla

are found in aquatic environments, with eight being exclusively aquatic and mostly marine.[34] Prior to the invention of the self-contained underwater breathing apparatus (SCUBA) in the 1940s, marine organism collection was performed by free diving or snorkeling and had a very limited range. Subsequently, the popularization of SCUBA allowed sample collection to depths of 10-120 feet on a routine basis. Because of this revolutionary technology, the marine environment has been increasingly explored as a source of novel bioactive agents. Deep water, inaccessible to scuba divers, can still be probed by dredging or trawling, both of which are non-selective and can cause environmental damage. These challenges can be overcome by use of manned submarines or remotely operated vehicles, but these methods suffer from an extremely high cost of operation.

As these new technologies were deployed, novel bioactive metabolites would begin to be discovered. Initially, isolation efforts focused on large or conspicuous marine species such as sponges and brown algae. These early investigations would uncover a modified nucleoside in the Caribbean sponge *Cryptotethia crypta*, where β-D-arabinofuranose replaced the usual 2-deoxyribose ring of deoxythymine.[35] Biochemical studies would reveal that cytosine arabinoside, also known as cytarabine or Ara-C, was a potent disrupter of DNA replication and led to cellular toxicity.[36] Used clinically to induce remission of acute myelocytic leukemia, perhaps a more important contribution from this metabolite would be highlighting nucleoside therapy as a valuable therapeutic strategy. This motif would later be used to great effect in antiviral chemotherapy, with nucleosides treating diseases like hepatitis B and human immunodeficiency virus.[37,38]

Not long after the discovery of this interesting nucleoside chemistry, extracts of the tunicate *Ecteinascidia turbinate* were found to contain potent anti-cancer activity.

Although it would take nearly 30 years to elucidate the structure of the active compound ecteinascidin 743,[39] it would become another marine natural product success story and is now approved for the treatment of soft tissue sarcomas and relapsed ovarian cancer. Another anticancer marine discovery would come in 1986, when the polyether metabolite halichondrin A from the sponge *Halichondria okadai* was described.[40] The compound possessed exquisite cancer cell toxicity through an antitubulin mechanism, but differed from other antitubulin agents by virtue of its unique binding site on β-tubulin.[41] Medicinal chemistry investigation would reveal that the macrocycle was responsible for the bioactivity, and modifications to the original carbon scaffold would give rise to the clinical drug eribulin.[42] One year later in 1987, investigation of the Indian Ocean sea hare *Dolabella auricularia* would produce another antitubulin agent, dolastatin 10.[43] Although clinical trials of the compound itself were unsuccessful, the dolastatin analog monomethyl auristatin E would prove to be a successful treatment for Hodgkin's lymphoma when conjugated with an anti-CD30 antibody.[44] Another clinical drug to emerge from the marine environment would be ziconotide, a peptide toxin from the fish-hunting marine mollusk, *Conus magus*.[45] Used to treat severe pain, the ω-conotoxin peptide functions through binding as an antagonist to N-type voltage-gated calcium channels. Unlike many treatments for chronic pain, ziconotide has the additional benefit of not inducing tolerance.

All told, there are 13 marine derived molecules used as clinical drugs,[46] but the marine environment has also produced valuable tools for molecular biology and biochemistry. Kainic acid, isolated from the red alga *Digenia simplex*,[47] was an essential neurobiological tool used to define the kainite receptor within the mammalian nervous system.[48] Saxitoxin, isolated from Alaskan butter clams,[49] is an extremely potent toxin

(LD_{50} in humans is 5.7 μg/kg) and was crucial tool for describing sodium channel function. Another potent sodium channel blocker, tetrodotoxin, was isolated from the puffer fish[50] and also used to probe sodium channel function. Additionally, tetrodotoxin was investigated as a pain reliever for oncology patients.[51] The diverse and complex structures of these metabolites are shown in Figure 3. These examples highlight the marine environment as not only source of novel bioactive metabolites, but also as 'first-in-class' inspirations for new kinds of chemotherapies and tools for probing molecular biology.

1.5 Actinomycetes Contribution to the Natural Products Pharmacopeia

One particularly impressive group of microbes proved to be an extremely rich source of novel therapies, the phylum actinobacteria. One of the largest taxonomic units among the major lineages currently recognized within the Bacteria domain,[52] actinobacteria are filamentous and Gram-positive. Their genomes contain a high guanine-plus-cytosine (G+C) content as well.[53] These fungus-like bacteria grow by a combination of tip extension and branching of hyphae, their name derived from the Greek words for ray (aktis or aktin) and fungi (mukēs). In the past, actinomycetes were considered transitional forms between fungi and bacteria because, like filamentous fungi, many *Actinobacteria* are mycelial and reproduce via sporulation. However, when investigated closely it is clear that these microorganisms are indeed bacteria. Their genetic material is organized in a prokaryotic nucleoid, and their cell walls are made of peptidoglycan rather than chitin.[54] Actinobacteria can be found in nearly any environment on earth, and have many different lifestyles. Some genera are soil or aquatic bacteria (*Streptomyces*, *Micromonospora*, *Rhodococcus*, and *Salinispora* species), others live as plant symbionts (*Frankia*), some are plant or animal pathogens

(*Corynebacterium*, *Mycobacterium*, or *Nocardia* species), and some make up a portion of the human microbiome as gastrointestinal commensals (*Bifidobacterium* spp.).[55]

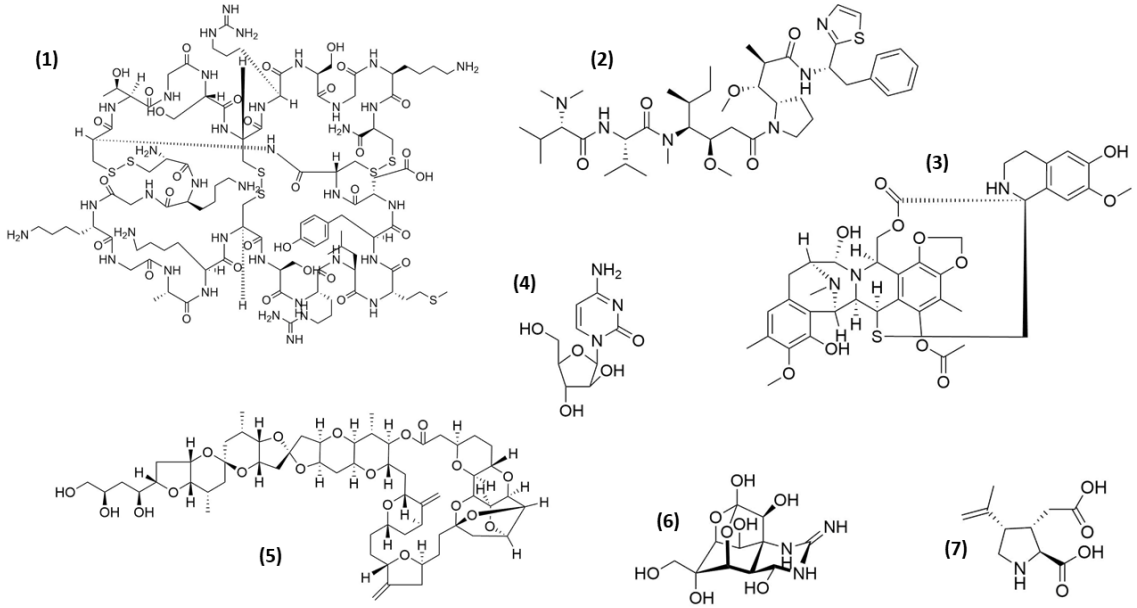


Figure 1.3: Marine Natural Products. Isolated from a variety of kingdoms, these natural products represent the incredible diversity of metabolites found in the marine environment. **(1)** Ziconotide, **(2)** Dolastatin 10, **(3)** ET-743 or Trabectedin, **(4)** Ara-C, **(5)** Halichondrin B, **(6)** Tetrodotoxin, and **(7)** Kainic acid

Actinobacteria have an extraordinary ability to make diverse natural products. Actinomycetes produce approximately 30% of all known 60,000–80,000 microbial natural products, and approximately 40% of all known 33,500 bioactive microbial metabolites.[56] They are the source of over 70% of practically used drugs isolated from microbial sources (drugs used for humans, animals, or agriculture).[57] Because of this proven track record, they have been the focus of many isolation efforts and biochemical investigations.

1.6 Natural Products Biosynthesis: Connecting Genes to Molecules

At the time of the first antibiotic discovery, little was known about the molecular mechanics of heritability. In 1858, Charles Darwin and Alfred Wallace were the first to postulate a 'hereditary material' which mediated the inheritance of certain traits from one generation to the next.[58] Shortly after, this idea would gain traction as Gregor Mendel published the results of his investigations into the laws of hereditary transmission in pea plants.[59] Although the basic properties of DNA would be described by Friedrich Miescher in 1869, scientists initially believed that it was too simple to convey hereditary information.[60] While investigating virulent and non-virulent strains of *Streptococcus pneumoniae* in 1928, bacteriologist Frederick Griffith discovered a "transforming principle" that was able transform the nonvirulent strain of the bacteria into a virulent strain. However, it was still believed that the transforming agent was likely a protein.[61] Decades later, immunologist Oswald Avery had set out to ultimately identify the transforming material. Along with Colin MacLeod and Maclyn McCarty, he published results in 1944 which repeated Griffith's experiments and showed that the active fraction of transforming principle consisted solely of DNA. However, it wasn't until years later in that DNA would be confirmed as the sole mediator of heritability in the eyes of the scientific community. In 1952, Alfred Hershey and Martha Chase were able to definitively show that DNA was the hereditary material using viral DNA labeled with radioactive phosphorous and viral protein labeled with radioactive sulfur.[62] This discovery, in concert with the determination of DNA's molecular structure,[63] would herald the dawn of a new age in molecular biology and genetics.

Because of their massive impact on the field of medicine, there was great interest in determining how natural products are created within organisms. The new discipline of

genetics was quick to investigate the problem, and by the mid 1950's several groups had begun to investigate the biosynthetically talented genus *Streptomyces*.[64] One of these dedicated investigators was David Hopwood, whose pioneering work would lay the foundations of *Streptomyces* genetics and eventually natural products biosynthesis. His painstaking work on genetic linkage maps for *S. coelicolor A3(2)* continued throughout the 1960's, and would facilitate the first studies connecting genes to molecules in *Streptomyces*.[65] In 1975, these efforts would begin to bear fruit when Hopwood and colleagues were able to show that the biosynthetic genes responsible for making methylenomycin A and the genes responsible for resistance to the antibiotic were present on an single plasmid, SCP1.[66–68] These studies represent the first genetic localization of antibiotic biosynthesis genes, but would prove somewhat misleading as most biosynthetic genes are found on the chromosomes of bacteria rather than plasmids. Hopwood's group would also show that the biosynthetic genes for actinorhodin existed as closely linked chromosomal genes.[69] This "clustering" of biosynthesis genes in a bacterium's genome would give rise to the idea of "biosynthetic gene clusters" (BGCs), a useful paradigm that still stands today.[70]

Although many advances were made in the study of biosynthesis using mutagenesis and genetic crossing experiments, the field would greatly accelerate with the invention of effective gene cloning methods, DNA sequencing, and the polymerase chain-reaction (PCR). First developed in *Escherichia coli*, recombinant DNA technology allows researchers to join pieces of DNA from multiple sources into a single molecule. Restriction enzymes were the first tools used to cut and re-join DNA at specific sequences.[71] In 1970, it was demonstrated that special treatment of *E. coli* induced uptake of DNA.[72] This technique was used to transform sensitive *E. coli* strains with

genes, conferring antibiotic resistance using a plasmid.[73] One year later, in 1973, the first recombinant DNA was introduced by transformation into *E. coli*, launching the field of synthetic biology. In 1974, this exciting DNA manipulation technology was adapted for use in *Streptomyces* by Okanashi and others,[74–76] allowing for the genetic manipulation of the most prolific antibiotic producers known to man.

Aided by new genetic tools for *Streptomyces* such as antibiotic-selectable plasmids,[77,78] researchers were able to start probing biosynthetic gene clusters. The first step was taken in 1983, when a chromosomal O-methyltransferase from *S. coelicolor* was shown to be essential for the production of undecylprodigiosin.[79] Soon after, biosynthesis genes for the production of methylenomycin A and candicidin were also described.[80,81] These initial investigations would set the stage for elucidating the entire biosynthetic pathway for actinorhodin, a feat accomplished in 1984 when a plasmid containing the entire BGC was cloned and used to transform the non-producing *Streptomyces parvulus*, which subsequently produced the blue pigmented antibiotic.[82] These foundational studies would encourage researchers to eventually connect thousands of BGCs to their small molecule products.[83]

1.7 DNA Sequencing and Natural Products Biosynthesis

The earliest sequencing methods were developed for proteins. In 1950, Pehr Edman published a paper demonstrating a label-cleavage method for protein sequencing which was later termed “Edman degradation”.[84] Concurrently, Fred Sanger was developing his own labelling and separation method which culminated in the sequencing of insulin.[85–88] For this work, Sanger was eventually awarded the 1958 Nobel Prize for Chemistry. Sanger would continue his sequencing work, but shifted his focus to

RNA, demonstrating the first version of electrophoretic sequencing as we know it today.[89,90] Despite these successes, a reliable DNA sequencing method would remain elusive. It wasn't until 1968 that the first 12 bases of DNA were sequenced using DNA polymerase and primer extensions,[91] with another method using an RNA intermediate delivering 24 deciphered bases soon after.[92] In 1977, Sanger would publish his seminal paper describing a DNA sequencing method based on dideoxy-nucleotide triphosphates which, with several key improvements, would become the gold standard of sequencing methodology used to this day.[93]

The so-called "Sanger sequencing" method, together with advancements in automated equipment to streamline and automate the sequencing protocol, would yield many advancements in the field of natural products discovery and biosynthesis. Although the structures of various natural products suggested they shared similar biosynthetic machinery, this was difficult to confirm without sequence information. Isotope labeled feeding studies had determined that polyketides were biosynthesized from condensed acetate units, but it took genome sequencing to reveal the fatty acid synthase-like type II polyketide synthase (PKS) which produced molecules like granaticin and tetracenomycin C.[94,95] These BGCs consist of two ketosynthase subunits (KSα and KSβ), an acyl carrier protein that anchors the growing polyketide chain, and various tailoring enzymes such as oxygenases, glycosyl and methyl-transferases. Further sequencing studies would reveal type I polyketide synthases or PKSs, first in the erythromycin producer *Saccharopolyspora erythraea*.[96] Differing from the compact type II "aromatic" PKSs, these enzymes are large, multi-modular proteins consisting of all the necessary domains for constructing the polyketide backbone.[96] Subsequent

investigations would also reveal type III PKS enzymes, homologous to the plant enzymes responsible for chalcone and stilbene synthesis in higher plants.[97,98]

Non-ribosomal peptides, distinguished from ribosomal peptides by the presence of non-proteinogenic amino acids; carboxy-acids; heterocyclic rings; modified amino acids and fatty acids,[99] were determined to be biosynthesized by unknown enzymatic machinery in 1968. An elegant experiment performed in the Lipmann laboratory employing cell extracts of the producer strains showed that synthesis of these peptides was possible even in the presence of RNases or inhibitors of the ribosomal machinery.[100] It would take another twenty years for the first NRPS gene to be sequenced,[101] but by 1999 enough of these enzymes had been cloned to launch the first tool predicting the amino acid specificity of NRPS modules.[102] Although much was still unknown, these studies hinted at the vast impact that gene sequencing would have had on understanding, predicting, and manipulating natural product biosynthesis.

The first bacterial antibiotic producer to have its entire genome sequenced was the laboratory workhorse *S. coelicolor* A3(2) in 2002.[103] Known then to produce only 4 natural products,[67,104–106] bioinformatic analysis revealed over 20 predicted biosynthetic gene clusters. Although the gene clusters for actinorhodin, Calcium-dependent antibiotic, prodigiosin, and the whiE grey spore pigment had been identified previously, the presence of at least eighteen uncharacterized gene clusters containing markers of secondary metabolism was surprising. These results were corroborated when the genome of the industrially important *Streptomyces avermitilis*, which produces the antiparasitic avermectins, was sequenced. Revealed to contain 38 biosynthetic gene clusters,[107] researchers began to see the large untapped potential of cryptic and silent gene clusters that were not expressed under normal laboratory conditions.

1.8 High Throughput and Whole Genome Sequencing Reveals Huge Biosynthetic Potential

Formally launched in 1990, the ambitious Human Genome Project would be the largest collaborative research project ever undertaken. Initially, the sequencing work would be accomplished by technology based on capillary electrophoresis of individual fluorescently labelled Sanger sequencing reaction products. Because of the high demand this enormous sequencing project had for fast and accurate sequence information, biotech companies soon developed new Sanger-based platforms that were able to increase the sequencing efficiency several fold.[108] However, these improvements would pale in comparison to the quantum leap in sequencing technology represented by the introduction of next-generation sequencing (NGS). Platforms like Roche's 454 pyrosequencer, Illumina's HiSeq, and Pacific Biosciences RS platform would increase the sequencing output of a genome sequencer over a billion-fold. Moving at a staggering rate of improvement between 2004-2010, DNA sequencing capacity doubled every five months.[109]

Although not explicitly developed to sequence bacterial genomes, NGS would launch a new era in natural products biosynthesis and discovery by facilitating fast and inexpensive whole genome sequencing of bacteria. An avalanche of sequence data has followed, and useful bioinformatic programs have been developed to identify and characterize BGCs from this sequence data. These include generalized algorithms such as antiSMASH[110,111] to predict a wide variety of pathways including PKS, NRPS, and other pathways, and more specific prediction software for analyzing specific types of BGCs like NRPSPredictor,[112] Natural Product Domain Seeker (NaPDoS),[113] BAGEL,[114] or PKSIIIexplorer.[115] These programs have help to identify an astonishing number of

cryptic gene clusters which represent ‘microbial dark matter’, chemical diversity that is inaccessible to traditional culture and fermentation methods.

1.9 Heterologous Expression of Natural Products

Naturally occurring antibiotics are produced by fermentation, an ancient technique that can be traced back almost 8000 years, initially utilized for beverages and food production. Beer is one of the world's oldest beverages, produced from barley by fermentation, possibly dating back to the sixth millennium BC and recorded in the written history of ancient Egypt and Mesopotamia.[116] Another old fermentation method is used to initiate the koji process, which takes advantage of *Aspergillus oryzae* to ferment rice grains. During the past 4000 years, *Penicillium roqueforti* has been utilized for cheese production, and for the past 3000 years soy sauce in Asia and bread in Egypt have represented other examples of traditional fermentations.[117]

The genome sequencing revolution and bioinformatic software has since revealed a colossal number of cryptic biosynthetic gene clusters, inaccessible to researchers using traditional fermentation methods. Scientists have since been looking for tools to probe these secretive BGCs. Although approaches such as new culturing methods have seen some success,[118,119] one exceptional tool in the hands of natural products researchers is the cloning and heterologous expression of cryptic BGCs. In this case, individual genes or whole clusters contained on plasmids or bacterial artificial chromosomes (BACs) are introduced to suitable laboratory host bacteria, which then produce the gene products and ultimately small molecule metabolites of cryptic gene clusters. The first example of this technique for natural product BGC elucidation was the heterologous expression of actinorhodin.[120] Early pathway expression techniques relied

on random capture of bacterial DNA and subsequent screening of hundreds or thousands of clones for the biosynthetic genes of interest. This time consuming library-based methodology would later be eclipsed by the direct capture method developed by Larionov and colleagues,[121,122] later adapted for bacterial natural product investigation by Sean Brady.[123] This modernized capture methodology has elucidated a wide variety of metabolites and their associated BGCs, both from culturable[124–126] and unculturable sources.[127]

1.10 Opportunities in Modern Natural Products Discovery and Overview of Dissertation Chapters

There has been incredible advancement in genome sequencing, bioinformatics, analytical chemistry, and molecular biology in the past decade. In order to leverage these powerful techniques for discovery, researchers have turned to the vast biodiversity of the marine environment. The ocean presents itself as a deep well of interesting and novel metabolites, with practically untouched fauna and flora, hosting over 80% of life on earth.[128] Because of their proven track record of producing bioactive natural products and unique evolutionary environment compared to the majority of laboratory studied actinomycetes, newly isolated obligate-marine actinomycetes are an attractive area to focus research efforts on. This thesis presents several facets of the same effort to synthesize modern molecular biology, analytical chemistry, and bioinformatics to uncover new and interesting natural product chemistry in actinomycetes.

The first study presented here in Chapter 2 was an effort to utilize sensitive, state-of-the-art mass spectrometry machines available at both the Scripps Institution of

Oceanography and the UCSD Skaggs School of Pharmacy and Pharmaceutical Sciences to investigate the products of many different strains of ocean bacteria across multiple fermentation and extraction conditions. Although this fermentation/extraction/analysis methodology has been used by many investigators previously, the availability of genome sequence information for all bacterial strains in a large culture library presented the possibility of connecting many molecules, both known and unknown, to their BGCs in the genome. Because of the huge amount of data contained in both the genome sequence and mass spectrometry analysis, this ambitious project would require the use of computer algorithms to analyze the data generated therein. This hypothesis independent "big data" path to natural products discovery differs significantly from the traditional experimental approach, which generally trains a researcher's focus on one or a few strains of bacteria for exhaustive investigation. Despite these differences, the investigators felt it was important to use the new powerful tools that had recently become available to natural products researchers. These tools include bioinformatic algorithms to parse the wealth of sequence data generated from the available culture library, in addition to newly developed molecular networking software pioneered by our collaborators in the Dorrestein lab to organize and make sense of thousands of mass spectrometry runs. Without these technological developments, such a study would be impossible. Leveraging this technology to investigate the metabolites of hundreds of ocean bacteria was an important undertaking, helping to highlight the promise of such an approach while at the same time revealing new information about natural product expression.

The later studies included in this thesis represent a more focused approach to natural products discovery. Specific BGCs were targeted for manipulation and capture

based on algorithmic analysis of genome information. In one case, a glycosylated anthracycline BGC was chosen after a survey of nearly two dozen ocean *Streptomyces* genomes. Many anthracyclines are powerful anti-cancer drugs, making this BGC more attractive than the average gene cluster. However, the most interesting portion of the cluster was the three distinct glycosyltransferases, which putatively decorated the anthracycline core with an unknown number and type of sugar moieties. These sugars can have profound effects on the biological activity of a metabolite, in addition to offering the opportunity for engineering the cluster itself by adding, subtracting, or modifying the glycosyltransferases. Additionally, a sensitive mass spectrometry technique developed in the Moore lab known as glycogenomics can quickly identify glycosylated metabolites in complex crude mixtures. This study reported in Chapter 3 was an effort to use and improve this glycogenomic technique while identifying new glycosylated anthracyclines and shed light on the biosynthetic machinery that produces these metabolites.

The final study presented here in Chapter 4 focused on another interesting and commercially important class of metabolites, phosphonates. These C-P bonded metabolites have a staggeringly wide range of bioactivities, and as of 2009 approximately 15% of all phosphonate natural products were commercialized. These metabolites have a wide range of uses including antibiotics, pesticides, and anti-parasitics.[129] These facts make phosphonate BGCs attractive targets for investigation, along with their relatively small size for ease of cloning and manipulation. New methods for cloning, based on generating high concentrations of PCR fragments for assembly rather than direct-capture from genomic material, were employed to assemble and express a putative phosphonate BGC from the marine bacterium *Salinispora pacifica*. Because of their extreme hydrophilicity, purification of these compounds proved difficult,

but by using advanced NMR techniques such as broad-band decoupling and two-dimensional NMR in concert with mass spectrometry it was possible to identify the metabolites produced by this BGC.

These studies represent different facets of natural products research, utilizing new technologies only recently available to individual investigators to cover new ground and advance the field. These include advanced bioinformatic analysis of genomic and metagenomic information, the development and implementation of novel gene cloning or gene capture methods, and modern analytical chemistry techniques used to elucidate the biosynthetic chemistry of ocean bacteria. These studies illuminate the research area of their specific focus, but also serve as examples of how to use new powerful tools for innovative research.

References

1. O'Brien, J. & Wright, G. D. An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* **22**, 552–558 (2011).

2. Cragg, G. M. & Newman, D. J. Biodiversity: A continuing source of novel drug leads. *Pure Appl. Chem.* **77**, 7–24 (2005).

3. Hallmann-Mikołajczak, A. [Ebers Papyrus. The book of medical knowledge of the 16th century B.C. Egyptians]. *Arch. Hist. i Filoz. Med.* **67**, 5–14 (2004).

4. Harper, D. J. *Early Chinese medical literature : the Mawangdui medical manuscripts*. (Kegan Paul International, 1998).

5. Gupta, M. P., CYTED (Organization). Subprograma de Química Fina Farmacéutica. & Secretaría Ejecutiva Permanente del Convenio Andrés Bello. *270 plantas medicinales iberoamericanas*. (Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo, Subprograma de Química Fina Farmacéutica, 1995).

6. Schultes, R. E. & Raffauf, R. F. *The healing forest: medicinal and toxic plants of the Northwest Amazonia. The healing forest: medicinal and toxic plants of the Northwest Amazonia.* (Dioscorides Press, 1990).

7. Ayensu, E. S. *Medicinal plants of the West Indies. Medicinal plants of the West Indies.* (Reference Publications, Inc., 1981).

8. Kapoor, L. D. *Handbook of ayurvedic medicinal plants*. (CRC Press, 2001).

9. Iwu, M. M. *Handbook of African medicinal plants*. (CRC Press, 1993).

10. Krishnamurti, C. & Rao, S. C. The isolation of morphine by Serturner. *Indian J. Anaesth.* **60**, 861–862 (2016).

11. Wood, J. N. From plant extract to molecular panacea: a commentary on Stone (1763) 'An account of the success of the bark of the willow in the cure of the agues'. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**, (2015).

12. DerMarderosian, A. & Beutler, J. A. *The review of natural products : the most complete source of natural product information.* (Facts & Comparisons, 2012).

13. Faurant, C. From bark to weed: the history of artemisinin. *Parasite* **18**, 215–8 (2011).

14. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *Br. J. Exp. Pathol.* **10**, 226 (1929).

15. Bernardini, S., Tiezzi, A., Laghezza Masci, V. & Ovidi, E. Natural products for human health: an historical overview of the drug discovery approaches. *Nat. Prod. Res.* **32**, 1926–1950 (2018).

16. Baker, D. D., Chu, M., Oza, U. & Rajgarhia, V. The value of natural products to

future pharmaceutical discovery. *Nat. Prod. Rep.* **24**, 1225 (2007).

17. Mann, J. *Murder, magic, and medicine*. (Oxford University Press, 1992).

18. Demain, A. L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J. Antibiot. (Tokyo).* **62**, 5–16 (2009).

19. Abraham, D. J., Rotella, D. P. & Burger, A. *Burger's Medicinal Chemistry, Drug Discovery and Development.* (Wiley, 2010).

20. Newman, D. J. & Cragg, G. M. *Drug Discovery, Therapeutics, and Preventive Medicine*. (Humana Press, 2005).

21. Foye, W. O. *Cancer chemotherapeutic agents*. (American Chemical Society, 1995).

22. Lee, M. D., Ellestad, G. A. & Borders, D. B. Calicheamicins: discovery, structure, chemistry, and interaction with DNA. *Acc. Chem. Res.* **24**, 235–243 (1991).

23. Peters, C. & Brown, S. Antibody-drug conjugates as novel anti-cancer chemotherapeutics. *Biosci. Rep.* **35**, (2015).

24. Liebig, J. Ueber einen neuen Apparat zur Analyse organischer Körper, und über die Zusammensetzung einiger organischen Substanzen. *Ann. der Phys. und Chemie* **97**, 1–43 (1831).

25. Holmes, F. L. Elementary Analysis and the Origins of Physiological Chemistry. *Isis* **54**, 50–81 (1963).

26. Thomas, N. C. The early history of spectroscopy. *J. Chem. Educ.* **68**, 631 (1991).

27. Thomson, J. J. Rays of Positive Electricity. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **89**, 1–20 (1913).

28. Griffiths, J. A Brief History of Mass Spectrometry. *Anal. Chem.* **80**, 5678–5683 (2008).

29. Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. A New Method of Measuring Nuclear Magnetic Moment. *Phys. Rev.* **53**, 318–318 (1938).

30. Robert H. Varian. Method and means for correlating nuclear properties of atoms and magnetic fields. (1948).

31. Lee, W. D., Drazen, J. M., Sharp, P. A. & Langer, R. S. *From X-rays to DNA : how engineering drives biology*. (2014).

32. Keifer, P. A. NMR spectroscopy in drug discovery: tools for combinatorial chemistry, natural products, and metabolism research. *Prog. Drug Res.* **55**, 137–211 (2000).

33. Dias, D. A., Urban, S., Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**, 303–336 (2012).

34. Iwu, M. M. & Wootton, J. C. *Ethnomedicine and drug discovery*. (Elsevier, 2002).

35. BERGMANN, W., WATKINS, J. C. & STEMPIEN, M. F. Contributions to the Study of Marine Products. XLV. Sponge Nucleic Acids [*,1]. *J. Org. Chem.* **22**, 1308–1313 (1957).

36. Brunton, L. *Goodman & Gilman: Las bases farmacológicas de la terapéutica*. (McGraw Hill, 2012).

37. Quan, D. J. & Peters, M. G. Antiviral therapy: nucleotide and nucleoside analogs. *Clin. Liver Dis.* **8**, 371–85 (2004).

38. Lehmann, C., Malin, J., Suárez, I. & Fätkenheuer, G. Moderne HIV-Therapie. *Internist (Berl).* (2019). doi:10.1007/s00108-019-0564-0

39. Rinehart, KL, Holt TG, Fregeau NL, Stroh JG, Kiefer PA, Sun F. Ecteinascidins 729, 743, 745, 759A, 759B, and 770: potent antitumor agents from the Caribbean tunicate Ecteinascidia turbinata. *J. Org. Chem.* **55**, 4512–4515 (1990).

40. Hickford, S. J. H., Blunt, J. W. & Munro, M. H. G. Antitumour polyether macrolides: Four new halichondrins from the New Zealand deep-water marine sponge Lissodendoryx sp. *Bioorg. Med. Chem.* **17**, 2199–2203 (2009).

41. Jordan, M. A. The primary antimitotic mechanism of action of the synthetic halichondrin E7389 is suppression of microtubule growth. *Mol. Cancer Ther.* **4**, 1086–1095 (2005).

42. Yeung, B. K. Natural product drug discovery: the successful optimization of ISP-1 and halichondrin B. *Curr. Opin. Chem. Biol.* **15**, 523–528 (2011).

43. Pettit, G. R. Kamano, Y., Herald, C.L.,Tuinman, A.A., Boettner, F.E., Kizu, H., Schmidt, J.M., Baczynskyj, L., Tomer, K.B., Bontems, R.J. The isolation and structure of a remarkable marine animal antineoplastic constituent: dolastatin 10. *J. Am. Chem. Soc.* **109**, 6883–6885 (1987).

44. Katz, J., Janik, J. E. & Younes, A. Brentuximab Vedotin (SGN-35). *Clin. Cancer Res.* **17**, 6428–6436 (2011).

45. Teichert, R. W. & Olivera, B. M. Natural products and ion channel pharmacology. *Future Med. Chem.* **2**, 731–744 (2010).

46. Gerwick, W. H. & Moore, B. S. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.* **19**, 85–98 (2012).

47. Moloney, M. G. Excitatory amino acids. *Nat. Prod. Rep.* **15**, 205–19 (1998).

48. Bunch, L. & Krogsgaard-Larsen, P. Subtype selective kainic acid receptor agonists: Discovery and approaches to rational design. *Med. Res. Rev.* **29**, 3–28 (2009).

49. Schantz, E. J., Ghazarossian VE, Schnoes HK, Strong FM, Springer JP, Pezzanite JO, Clardy J. Structure of saxitoxin. *J. Am. Chem. Soc.* **97**, 1238–1239 (1975).

50. Woodward, R. B. The structure of tetrodotoxin. *Pure Appl. Chem.* **9**, 49–74

(1964).

51. Hagen NA, Lapointe B, Ong-Lam M, Dubuc B, Walde D, Gagnon B, Love R, Goel R, Hawley P, Ngoc AH, du Souich P.. A multicentre open-label safety and efficacy study of tetrodotoxin for cancer pain. *Curr. Oncol.* **18**, e109-16 (2011).

52. Ludwig, W., Euzéby, J., Schumann, P., Busse, H.J., Trujillo, M.E., Kämpfer, P., Whitman, W.B. Road map of the phylum Actinobacteria. in *Bergey's Manual® of Systematic Bacteriology* 1–28 (Springer New York, 2012). doi:10.1007/978-0-387-68233-4_1

53. Macagnan, D., Romeiro, R. da S., de Souza, J. T. & Pomella, A. W. V. Isolation of actinomycetes and endospore-forming bacteria from the cacao pod surface and their antagonistic activity against the witches' broom and black pod pathogens. *Phytoparasitica* **34**, 122–132 (2006).

54. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Meier-Kolthoff JP, Klenk HP, Clément C, Ouhdouch Y, van Wezel GP. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **80**, 1–43 (2016).

55. Lewin GR, Carlos C, Chevrette MG, Horn HA, McDonald BR, Stankey RJ, Fox BG, Currie CR. Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annu. Rev. Microbiol.* **70**, 235–254 (2016).

56. Niu, G. Genomics-Driven Natural Product Discovery in Actinomycetes. *Trends Biotechnol.* **36**, 238–241 (2018).

57. Raja, A. & Prabakaran, P. Actinomycetes and Drugs-An Overview. *Am. J. Drug Discov. Dev.* **1**, 75–84 (2011).

58. Darwin, C. & Wallace, A. On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *J. Proc. Linn. Soc. London. Zool.* **3**, 45–62 (1858).

59. Abbott, S. & Fairbanks, D. J. Experiments on Plant Hybrids by Gregor Mendel. *Genetics* **204**, 407–422 (2016).

60. Dahm, R. Friedrich Miescher and the discovery of DNA. *Dev. Biol.* **278**, 274–288 (2005).

61. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond).* **27**, 113–59 (1928).

62. HERSHEY, A. D. & CHASE, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* **36**, 39–56 (1952).

63. WATSON, J. D. & CRICK, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–8 (1953).

64. Hopwood, D. A. Forty years of genetics with Streptomyces: from in vivo through in vitro to in silico. *Microbiology* **145**, 2183–2202 (1999).

65. Hopwood, D. A. LINKAGE AND THE MECHANISM OF RECOMBINATION IN

STREPTOMYCES COELICOLOR*. *Ann. N. Y. Acad. Sci.* **81**, 887–898 (2006).

66. Kirby, R. & Hopwood, D. A. Genetic determination of methylenomycin synthesis by the SCP1 plasmid of Streptomyces coelicolor A3(2). *J. Gen. Microbiol.* **98**, 239–52 (1977).

67. WRIGHT, L. F. & HOPWOOD, D. A. Identification of the Antibiotic Determined by the SCP1 Plasmid of Streptomyces coelicolor A3 (2). *J. Gen. Microbiol.* **95**, 96–106 (1976).

68. KIRBY, R., WRIGHT, L. F. & HOPWOOD, D. A. Plasmid-determined antibiotic synthesis and resistance in Streptomyces coelicolor. *Nature* **254**, 265–267 (1975).

69. Wright, L. F. & Hopwood, D. A. Actinorhodin is a chromosomally-determined antibiotic in Streptomyces coelicolar A3(2). *J. Gen. Microbiol.* **96**, 289–97 (1976).

70. RUDD, B. A. M. & HOPWOOD, D. A. Genetics of Actinorhodin Biosynthesis by Streptomyces coelicolor A3(2). *J. Gen. Microbiol.* **114**, 35–43 (1979).

71. Smith, H. O. & Wilcox, K. W. A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *J. Mol. Biol.* **51**, 379–91 (1970).

72. Mandel, M. & Higa, A. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**, 159–62 (1970).

73. Cohen, S. N., Chang, A. C. & Hsu, L. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 2110–4 (1972).

74. Okanishi, M., Suzuki, K. & Umezawa, H. Formation and reversion of Streptomycete protoplasts: cultural condition and morphological study. *J. Gen. Microbiol.* **80**, 389–400 (1974).

75. Bibb, M. J., Ward, J. M. & Hopwood, D. A. Transformation of plasmid DNA into Streptomyces at high frequency. *Nature* **274**, 398–400 (1978).

76. BALTZ, R. H. & MATSUSHIMA, P. Protoplast Fusion in Streptomyces: Conditions for Efficient Genetic Recombination and Cell Regeneration. *Microbiology* **127**, 137–146 (1981).

77. Thompson, C. J., Ward, J. M. & Hopwood, D. A. DNA cloning in Streptomyces: resistance genes from antibiotic-producing species. *Nature* **286**, 525–7 (1980).

78. Bibb, M. J. & Cohen, S. N. Gene expression in Streptomyces: construction and application of promoter-probe plasmid vectors in Streptomyces lividans. *Mol. Gen. Genet.* **187**, 265–77 (1982).

79. Feitelson, J. S. & Hopwood, D. A. Cloning of a Streptomyces gene for an O-methyltransferase involved in antibiotic biosynthesis. *Mol. Gen. Genet.* **190**, 394–8 (1983).

80. Chater, K. F. & Bruton, C. J. Mutational cloning in Streptomyces and the isolation of antibiotic production genes. *Gene* **26**, 67–78 (1983).

81. Gil, J. A. & Hopwood, D. A. Cloning and expression of a p-aminobenzoic acid synthetase gene of the candicidin-producing Streptomyces griseus. *Gene* **25**, 119–32 (1983).

82. Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **309**, 462–4

83. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJ, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann HW, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus EP, Barona-Gómez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P, Cheng YQ, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein PC, Doyle S, Driessen AJ, Ehling-Schulz M, Entian KD, Fischbach MA, Gerwick L, Gerwick WH, Gross H, Gust B, Hertweck C, Höfte M, Jensen SE, Ju J, Katz L, Kaysser L, Klassen JL, Keller NP, Kormanec J, Kuipers OP, Kuzuyama T, Kyrpides NC, Kwon HJ, Lautru S, Lavigne R, Lee CY, Linquan B, Liu X, Liu W, Luzhetskyy A, Mahmud T, Mast Y, Méndez C, Metsä-Ketelä M, Micklefield J, Mitchell DA, Moore BS, Moreira LM, Müller R, Neilan BA, Nett M, Nielsen J, O'Gara F, Oikawa H, Osbourn A, Osburne MS, Ostash B, Payne SM, Pernodet JL, Petricek M, Piel J, Ploux O, Raaijmakers JM, Salas JA, Schmitt EK, Scott B, Seipke RF, Shen B, Sherman DH, Sivonen K, Smanski MJ, Sosio M, Stegmann E, Süssmuth RD, Tahlan K, Thomas CM, Tang Y, Truman AW, Viaud M, Walton JD, Walsh CT, Weber T, van Wezel GP, Wilkinson B, Willey JM, Wohlleben W, Wright GD, Ziemert N, Zhang C, Zotchev SB, Breitling R, Takano E, Glöckner FO.. Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).

84. EDMAN, P. A method for the determination of amino acid sequence in peptides. *Arch. Biochem.* **22**, 475 (1949).

85. SANGER, F. & THOMPSON, E. O. P. The amino-acid sequence in the glycyl chain of insulin. II. The investigation of peptides from enzymic hydrolysates. *Biochem. J.* **53**, 366–74 (1953).

86. SANGER, F. & THOMPSON, E. O. P. The amino-acid sequence in the glycyl chain of insulin. *Biochem. J.* **52**, iii (1952).

87. SANGER, F. & TUPPY, H. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochem. J.* **49**, 481–90 (1951).

88. SANGER, F. & TUPPY, H. The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochem. J.* **49**, 463–81 (1951).

89. Sanger, F., Brownlee, G. G. & Barrell, B. G. A two-dimensional fractionation procedure for radioactive nucleotides. *J. Mol. Biol.* **13**, 373–98 (1965).

90. Brownlee, G. G., Sanger, F. & Barrell, B. G. Nucleotide sequence of 5S-ribosomal RNA from Escherichia coli. *Nature* **215**, 735–6 (1967).

91. Wu, R. & Kaiser, A. D. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J. Mol. Biol.* **35**, 523–37 (1968).

92. Gilbert, W. & Maxam, A. The nucleotide sequence of the lac operator. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3581–4 (1973).

93. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–7 (1977).

94. Sherman DH, Malpartida F, Bibb MJ, Kieser HM, Bibb MJ, Hopwood DA. Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of Streptomyces violaceoruber Tü22. *EMBO J.* **8**, 2717–2725 (1989).

95. Bibb, M. J., Biró, S., Motamedi, H., Collins, J. F. & Hutchinson, C. R. Analysis of the nucleotide sequence of the Streptomyces glaucescens tcmI genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *EMBO J.* **8**, 2727–36 (1989).

96. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **348**, 176–178 (1990).

97. Funa N, Ohnishi Y, Fujii I, Shibuya M, Ebizuka Y, Horinouchi S. A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897–899 (1999).

98. Bangera, M. G. & Thomashow, L. S. Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from Pseudomonas fluorescens Q2-87. *J. Bacteriol.* **181**, 3155–63 (1999).

99. Finking, R. & Marahiel, M. A. Biosynthesis of Nonribosomal Peptides. *Annu. Rev. Microbiol.* **58**, 453–488 (2004).

100. Gevers, W., Kleinkauf, H. & Lipmann, F. The activation of amino acids for biosynthesis of gramicidin S. *Proc. Natl. Acad. Sci. U. S. A.* **60**, 269–76 (1968).

101. Weckermann, R., Fürbass, R. & Marahiel, M. A. Complete nucleotide sequence of the tycA gene coding the tyrocidine synthetase 1 from Bacillus brevis. *Nucleic Acids Res.* **16**, 11841 (1988).

102. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–83 (1997).

103. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth

S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417**, 141–147 (2002).

104. Tsao, S. W., Rudd, B. A., He, X. G., Chang, C. J. & Floss, H. G. Identification of a red pigment from Streptomyces coelicolor A3(2) as a mixture of prodigiosin derivatives. *J. Antibiot. (Tokyo).* **38**, 128–31 (1985).

105. RUDD, B. A. M. & HOPWOOD, D. A. A Pigmented Mycelial Antibiotic in Streptomyces coelicolor: Control by a Chromosomal Gene Cluster. *Microbiology* **119**, 333–340 (1980).

106. Lakey, J. H., Lea, E. J. A., Rudd, B. A. M., Wright, H. M. & Hopwood, D. A. A New Channel-forming Antibiotic from Streptomyces coelicolor A3(2) Which Requires Calcium for its Activity. *Microbiology* **129**, 3565–3573 (1983).

107. Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M. Genome sequence of an industrial microorganism Streptomyces avermitilis: Deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci.* **98**, 12215–12220 (2001).

108. Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470**, 198–203 (2011).

109. Stein, L. D. The case for cloud computing in genome informatics. *Genome Biol.* **11**, 207 (2010).

110. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339-46 (2011).

111. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).

112. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799–5808 (2005).

113. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **7**, e34064 (2012).

114. de Jong, A., van Hijum, S. A. F. T., Bijlsma, J. J. E., Kok, J. & Kuipers, O. P. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res.* **34**, W273-9 (2006).

115. Vijayan, M., Chandrika, S. K. & Vasudevan, S. E. PKSIIIexplorer: TSVM approach for predicting Type III polyketide synthase proteins. *Bioinformation* **6**, 125–7 (2011).

116. Gallone B, Mertens S, Gordon JL, Maere S, Verstrepen KJ, Steensels J. Origins, evolution, domestication and diversity of Saccharomyces beer yeasts. *Curr. Opin. Biotechnol.* **49**, 148–155 (2018).

117. Holker, U., Hofer, M. & Lenz, J. Biotechnological advantages of laboratory-scale solid-state fermentation with fungi. *Appl. Microbiol. Biotechnol.* **64**, 175–186 (2004).

118. Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, Kanigan T, Lewis K, Epstein SS.. Use of ichip for high-throughput in situ cultivation of 'uncultivable' microbial species. *Appl. Environ. Microbiol.* **76**, 2445–50 (2010).

119. Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schäberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C, Lewis K. A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).

120. Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. 1984. *Biotechnology* **24**, 342–3 (1992).

121. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 491–6 (1996).

122. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. *Nat. Protoc.* **3**, 371–377 (2008).

123. Kallifidas, D. & Brady, S. F. Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. *Methods Enzymol.* **517**, 225–39 (2012).

124. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1957–62 (2014).

125. Bonet, B., Teufel, R., Crüsemann, M., Ziemert, N. & Moore, B. S. Direct Capture and Heterologous Expression of Salinispora Natural Product Genes for the Biosynthesis of Enterocin. *J. Nat. Prod.* **78**, 539–542 (2015).

126. Ross, A. C., Gulland, L. E. S., Dorrestein, P. C. & Moore, B. S. Targeted Capture and Heterologous Expression of the Pseudoalteromonas Alterochromide Gene Cluster in Escherichia coli Represents a Promising Natural Product Exploratory Platform. *ACS Synth. Biol.* **4**, 414–420 (2015).

127. Katz, M., Hover, B. M. & Brady, S. F. Culture-independent discovery of natural

products from soil metagenomes. *J. Ind. Microbiol. Biotechnol.* **43**, 129–141 (2016).

128. Malve, H. Exploring the ocean for new drug developments: Marine pharmacology. *J. Pharm. Bioallied Sci.* **8**, 83–91 (2016).

129. Metcalf, W. W. & van der Donk, W. A. Biosynthesis of phosphonic and phosphinic acid natural products. *Annu. Rev. Biochem.* **78**, 65–94 (2009).

Chapter 2: Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols

2.1 Chapter Introduction

The modern mass spectrometer can be dated to 1910, when J.J. Thomson and Francis Aston used an instrument of their own design to describe the molecular weight of several small nuclei and molecules, including hydrogen, helium, and carbon dioxide.[1] Later, in 1919, an improved mass spectrometer would be used determine the masses of single atomic nuclei.[2] The decades that followed brought many more advancements in the technology including the addition of liquid[3] or gas chromatographic separation[4] before ionization, high-resolution MS able to determine the elemental composition of molecules,[5,6] tandem mass spectrometers able to fragment molecules into smaller constitutive pieces,[7,8] and electrospray ionization.[9] Mass spectrometry has become an extremely powerful tool for structural elucidation, sensitive enough to identify ions in the zeptomole (10^{-21} M) concentration range.[10]

Another important advance in the field of mass spectrometry was the compilation of spectral databases. Databases containing reference mass spectra were being compiled as early as the 1960s[11,12], and proved useful for identifying known compounds in complex mixtures and also for structural elucidation of unknown molecules. However, comparing experimental spectra to a library of reference spectra manually is both laborious and time consuming. The GNPS program recently developed in the Dorrestein lab aims to improve this workflow by avoiding manual comparison and instead algorithmically comparing MS^2 spectra in a sample group to both reference spectra and other spectra within (standards or deposited by other labs).[13] The result of GNPS

analysis is a network of nodes representing specific masses/molecules and edges representing the relatedness between these nodes based on their fragmentation pattern. This allows users to analyze massive data sets and compare large numbers of extracts both to each other and to a curated spectral library of known compounds. The technique has been successfully applied to a variety of analytical problems including natural product dereplication,[14] the identification of unknown and known molecules in bacterial extracts,[15,16] and metabolomics including the analysis of chemistry of the human microbiome.[17]

The phyla Actinobacteria has long been known as a producer of natural products. In fact, 70% of known microbial secondary metabolites were isolated from actinomycetes, including 50% of known antibiotics.[18] Traditionally isolated in the terrestrial environment, obligate marine actinomycetes were first isolated and characterized in the lab of Dr. William Fenical.[19,20] This genus quickly revealed its first bioactive natural products, the salinisporamides, which proved to be powerful anticancer molecules.[21] Indeed, salinisporomide A is now in phase III trials under the trade name Marizomib.[22] Because of this early success and because terrestrial actinomycetes are such a treasure trove of bioactive molecules, these newly discovered marine species deserve special attention when searching for novel drug candidates.

In this study we apply the latest mass spectrometry techniques and GNPS program to investigate 146 species of marine actinomycetes. These state-of-the art techniques allowed us to analyze and compare approximately 1.8 million mass spectra from over 600 samples representing different extraction techniques, incubation times, and media for the various strains. Such large-scale analysis and comparison of chemistry of the strains was plausible due to automated algorithm of GNPS tool. The

results revealed 15 groups of known metabolites in addition to identifying the most prolific producers of unknown metabolites. This information can be used to prioritize the investigation of wild-type strains that produce multiple unknown compounds, in addition to aiding in the connection of genes to molecules.

2.2 Chapter 2 Introduction References

1. Thomson, G. P. & Feather, N. J. J. Thomson and the Cavendish Laboratory in his Day. *Phys. Today* **18**, 65–65 (1965).

2. Squires, G. Francis Aston and the mass spectrograph. *J. Chem. Soc. Dalt. Trans.* **0**, 3893–3900 (1998).

3. Baldwin, M. A. & McLafferty, F. W. Liquid chromatography-mass spectrometry interface-I: The direct introduction of liquid solutions into a chemical ionization mass spectrometer. *Org. Mass Spectrom.* **7**, 1111–1112 (1973).

4. Ryhage, R. Use of a Mass Spectrometer as a Detector and Analyzer for Effluent Emerging from High Temperature Gas Liquid Chromatography Columns. *Anal. Chem.* **36**, 759–764 (1964).

5. Beynon, J. H. (John H. Mass spectrometry and its applications to organic chemistry. (1960).

6. BEYNON, J. H. Qualitative Analysis of Organic Compounds by Mass Spectrometry. *Nature* **174**, 735–737 (1954).

7. McLafferty, F. W., Bente, P.F.I., Kornfield, R., Tsai, S., Howe, I. Metastable ion characteristics. XXIV. Application of collisional activation spectra to the elucidation of organic ion structures. *J. Am. Chem. Soc.* **95**, 3886–3892 (1973).

8. Cheng, M. T., Kruppa, G. H., McLafferty, F. W. & Cooper, D. A. Structural information from tandem mass spectrometry for China White and related fentanyl derivatives. *Anal. Chem.* **54**, 2204–2207 (1982).

9. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).

10. Belov, M. E., Gorshkov, M. V, Udseth, H. R., Anderson, G. A. & Smith, R. D. Zeptomole-sensitivity electrospray ionization--Fourier transform ion cyclotron resonance mass spectrometry of proteins. *Anal. Chem.* **72**, 2271–9 (2000).

11. Gohlke RS. *Uncertified Dow Mass Spectral Data*. (Dow Chem. Co., 1963).

12. Abrahamsson S, Stenhagen E, M. F. *Atlas of Mass Spectral Data*. *Atlas of Mass Spectral Data* (Wileyy, 1969).

13. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel

F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).

14. Mohimani H, Gurevich A, Mikheenko A, Garg N, Nothias LF, Ninomiya A, Takada K, Dorrestein PC, Pevzner PA. Dereplication of peptidic natural products through database search of mass spectra. *Nat. Chem. Biol.* **13**, 30–37 (2017).

15. Wu C, van der Heul HU, Melnik AV, Lübben J, Dorrestein PC, Minnaard AJ, Choi YH, van Wezel GP. Lugdunomycin, an Angucycline-Derived Molecule with Unprecedented Chemical Architecture. *Angew. Chemie Int. Ed.* **58**, 2809–2814 (2019).

16. Hoffmann M, Auerbach D, Panter F, Hoffmann T, Dorrestein PC, Müller R. Homospermidine Lipids: A Compound Class Specifically Formed during Fruiting Body Formation of Myxococcus xanthus DK1622. *ACS Chem. Biol.* **13**, 273–280 (2018).

17. Quinn RA, Navas-Molina JA, Hyde ER, Song SJ, Vázquez-Baeza Y, Humphrey G, Gaffney J, Minich JJ, Melnik AV, Herschend J, DeReus J, Durant A, Dutton RJ, Khosroheidari M, Green C, da Silva R, Dorrestein PC, Knight R. From Sample to Multi-Omics Conclusions in under 48 Hours. *mSystems* **1**, (2016).

18. Subramani, R. & Aalbersberg, W. Marine actinomycetes: An ongoing source of novel bioactive metabolites. *Microbiol. Res.* **167**, 571–580 (2012).

19. Maldonado LA, Fenical W, Jensen PR, Kauffman CA, Mincer TJ, Ward AC, Bull AT, Goodfellow M. Salinispora arenicola gen. nov., sp. nov. and Salinispora tropica sp. nov., obligate marine actinomycetes belonging to the family Micromonosporaceae. *Int. J. Syst. Evol. Microbiol.* **55**, 1759–1766 (2005).

20. Jensen, P. R. & Mafnas, C. Biogeography of the marine actinomycete Salinispora. *Environ. Microbiol.* **8**, 1881–1888 (2006).

21. Buchanan GO, Williams PG, Feling RH, Kauffman CA, Jensen PR, Fenical W. Sporolides A and B: Structurally Unprecedented Halogenated Macrolides from the Marine Actinomycete Salinispora tropica. *Org. Lett.* **7**, 2731–2734 (2005).

22. A Phase III Trial of With Marizomib in Patients With Newly Diagnosed Glioblastoma - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT03345095. (Accessed: 24th February 2019)

2.3 Reprint of ": Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols"

JOURNAL OF NATURAL PRODUCTS

Article

pubs.acs.org/jnp

Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols

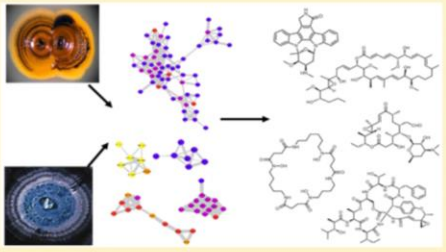
Max Crüsemann,[†,⊥,Δ] Ellis C. O'Neill,[†,∥,Δ] Charles B. Larson,[†] Alexey V. Melnik,[‡] Dimitrios J. Floros,[‡] Ricardo R. da Silva,[‡,§] Paul R. Jensen,[†] Pieter C. Dorrestein,[*,†,‡] and Bradley S. Moore[*,†,‡]

[†]Center for Marine Biotechnology & Biomedicine, Scripps Institution of Oceanography, and [‡]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, United States

[§]Research Support Center in Natural and Synthetic Products, Department of Physics and Chemistry, Faculty of Pharmaceutical Sciences, University of São Paulo, Ribeirão Preto, 14040-903, Brazil

Supporting Information

ABSTRACT: In order to expedite the rapid and efficient discovery and isolation of novel specialized metabolites, while minimizing the waste of resources on rediscovery of known compounds, it is crucial to develop efficient approaches for strain prioritization, rapid dereplication, and the assessment of favored cultivation and extraction conditions. Herein we interrogated bacterial strains by systematically evaluating cultivation and extraction parameters with LC-MS/MS analysis and subsequent dereplication through the Global Natural Product Social Molecular Networking (GNPS) platform. The developed method is fast, requiring minimal time and sample material, and is compatible with high-throughput extract analysis, thereby streamlining strain prioritization and evaluation of culturing parameters. With this approach, we analyzed 146 marine *Salinispora* and *Streptomyces* strains that were grown and extracted using multiple different protocols. In total, 603 samples were analyzed, generating approximately 1.8 million mass spectra. We constructed a comprehensive molecular network and identified 15 molecular families of diverse natural products and their analogues. The size and breadth of this network shows statistically supported trends in molecular diversity when comparing growth and extraction conditions. The network provides an extensive survey of the biosynthetic capacity of the strain collection and a method to compare strains based on the variety and novelty of their metabolites. This approach allows us to quickly identify patterns in metabolite production that can be linked to taxonomy, culture conditions, and extraction methods, as well as informing the most valuable growth and extraction conditions.



Nearly half of all small-molecule drugs approved for use in humans are derived from natural products.[1] The ability to sequence bacterial genomes at constantly decreasing costs and time has dramatically changed the field of natural products discovery research over the past decade. With a growing number of genomes sequenced, comparative genomics and novel bioinformatics approaches have been used to analyze and classify biosynthetic gene clusters (BGCs) on a larger scale.[2] It has been commonly observed that many organisms contain far more BGCs than characterized natural products. One approach to overcome this gap and further characterize natural product diversity is to culture and extract the microbes in many different ways. This approach has been named OSMAC (one strain, many compounds) by Zeeck and co-workers.[3] It was first employed in the early 2000s and has led to the isolation of large numbers of novel metabolites by systematically altering cultivation parameters.[4]

Natural products chemists frequently face the challenge of rediscovery of known compounds. Several mass-spectrometry-based metabolomics workflows have been developed to ameliorate this high rediscovery rate, referred to as "dereplication".[5] However, many of these approaches solely use MS1 data, thus identifying compounds only by mass and chromatographic and spectroscopic properties, and are not able to determine structural relationships between the metabolites.

Molecular networking is a recently introduced concept for the analysis of mass spectrometric fragmentation data and assessment of structural similarities between measured metabolites. The molecular networking concept enables the visualization of large data sets and the grouping of fragmented ions into clusters, using an algorithm to compare the similarity of the fragmentation spectra.[6] In a natural product molecular network, these clusters represent molecular families (MFs) putatively synthesized by gene cluster families (GCFs).[7] Molecular networking is a powerful approach that has advanced

Special Issue: Special Issue in Honor of Phil Crews

Received: August 4, 2016
Published: November 11, 2016

DOI: 10.1021/acs.jnatprod.6b00722
J. Nat. Prod. 2017, 80, 588−597

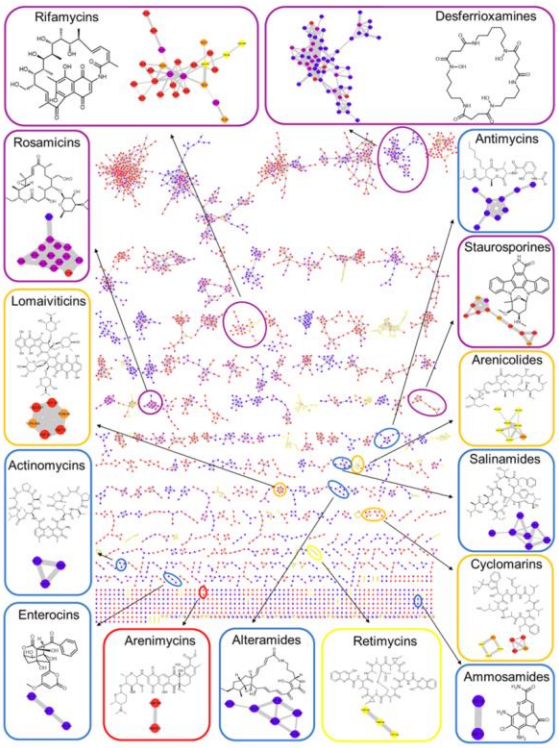


Figure 1. Molecular network of all generated extracts. Blue nodes represent ions detected only from *Streptomyces*; red nodes represent ions detected only from *Salinispora* strains grown on agar. The yellow nodes represent ions detected from *Salinispora* only in liquid medium,[19] while the orange nodes represent ions detected in both liquid and solid media. Purple nodes represent ions detected in both *Streptomyces* and *Salinispora*. Molecular families that include standards from the GNPS library are highlighted in the network (displaying the structure of the most abundant analogue) and color-coded according to their source. GNPS IDs of the standards can be found in Table S2. Only clusters containing at least two nodes are shown.

several natural-product-related research projects involving dereplication and quantification,[8] discovery,[9] biosynthesis,[10] and chemical ecology.[11] It has also been integrated as a central component of the Global Natural Products Social (GNPS) molecular networking platform, where dereplication is performed against a large, community-acquired reference library of spectra.[12] Molecular networking further allows for the screening of large numbers of strains for metabolic assessment.[7]

Creating networks with large numbers of closely related strains provides opportunities to identify new molecular families and investigate differences when growth and extraction conditions are changed. In this study, we screened 146 marine *Salinispora* and *Streptomyces* strains using HPLC-MS/MS, molecular networking, and the GNPS platform. We aimed to systematically explore the culturing and extraction of these strains to gain insight into the distribution of known and unknown metabolites and the effects of different growth and extraction protocols on the compounds detected. Analysis of the networks showed that varying conditions such as culture medium, extraction solvent, and time impact the networks. Furthermore, this study highlights species- and genus-specific metabolite production on a larger scale and allows for the prioritization of strains and optimized conditions for future MS-guided natural product discovery projects.

RESULTS AND DISCUSSION

Cultivation, Extraction, and Generation of Molecular Networks. The objective of this study was to apply large-scale molecular networking to a related group of sequenced bacteria to comprehensively interrogate the effects of growth media and extraction methods on the production and recovery of specialized metabolites respectively and to prioritize strains that produce novel molecular families for further study. We selected marine actinomycete bacteria, as they are known to be prolific producers of secondary metabolites.[13] First, we established an effective small-scale extraction method for the HPLC-MS/MS-based screening and analysis. Extraction was carried out sequentially with three solvents of increasing polarity (EtOAc, *n*-butanol, and MeOH). To evaluate the

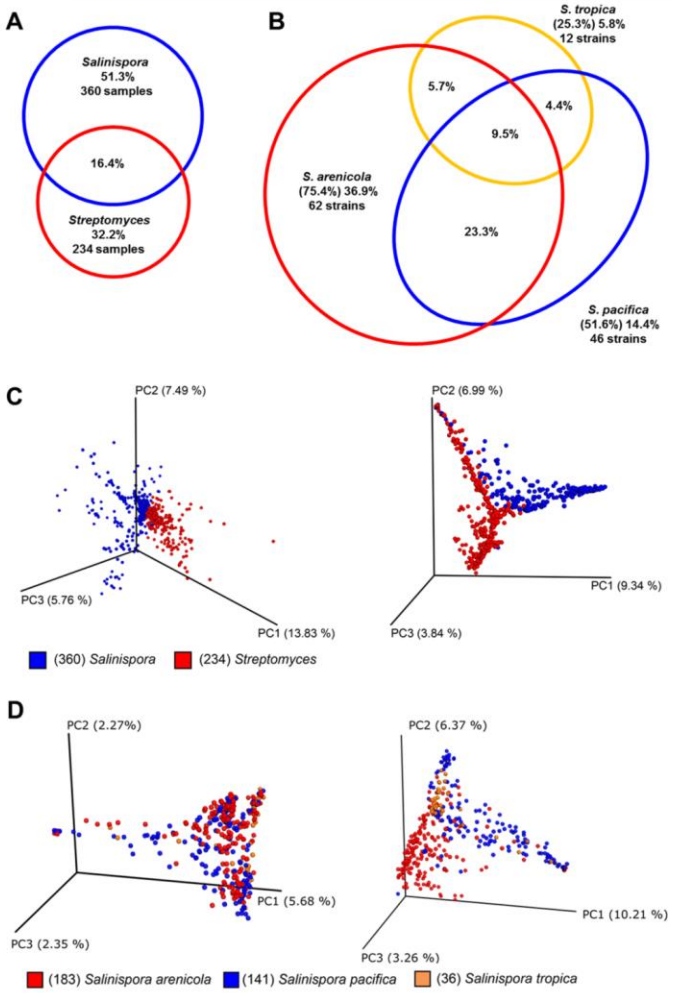


Figure 2. Effects of genus and species on molecular diversity-network analysis. (A) Venn diagram for *Salinispora* and *Streptomyces* specific nodes in the networks. (B) Venn diagram displaying *Salinispora* species node distribution. Percentages are shown for each sector, with the total percentage for each species in parentheses. (C) PCoA plots of *Salinispora* (blue) and *Streptomyces* (red) samples separated using Gower (left) and random forest classifier (right). (D) *Salinispora* samples were reanalyzed with unsupervised (left) and supervised (right) random forest based on *Salinispora* species.

relationships between bacterial species and chemical diversity, 120 *Salinispora* strains were cultivated on A1 agar (Table S1). The obligate marine actinomycete genus *Salinispora* consists of three closely related species, *arenicola*, *pacifica*, and *tropica*,[14,15] that produce a wide range of bioactive secondary metabolites.[16] Recent studies have shown that certain secondary metabolites are consistently produced by individual species,[15,17] which has been supported at the genomic level based on BGC distributions.[2c,18] We additionally selected 26 marine *Streptomyces* strains and, in order to evaluate the effects of media composition on metabolite production, grew them on three different media, A1, MS, and R5 agar (Table S1). Complete genomes are available for all strains to facilitate future research programs. In total, 603 samples were analyzed, generating approximately 1.8 million mass spectra that were processed with the GNPS molecular networking workflow.[12]

A comprehensive network was generated for spectra with a minimum of four fragment ions and by merging all identical

spectra into individual consensus nodes. Only nodes that had at least two identical spectra were displayed. After removal of nodes associated with the solvent controls, the molecular network consisted of 5526 nodes connected with 7396 edges; 54.6% of the nodes were organized into a total of 472 molecular families, comprising two or more nodes each. The remainder of the MS/MS spectra are sufficiently unique that they did not form any connections to other spectra. Additionally, previously published data from 35 *Salinispora* strains[19] grown in liquid culture were incorporated from the publically available GNPS-MassIVE database,[12] allowing for comparisons between liquid and solid cultivation conditions. For comparisons between growth media, only the samples obtained from *Streptomyces* were networked with each other (Figure S1). It is important to note that the number of nodes in the network does not correspond exactly to the number of metabolites, as different adducts or different charge states of the same chemical species can generate different nodes. Rather, molecular networking provides an overview of the different chemistries detected by mass spectrometry.

Analysis of Known Molecular Families in the Molecular Network. We identified 15 molecular families that contained spectra that matched known compounds in the network based on the curated GNPS natural products library (Figure 1, Table S2). In our analysis, we applied a mass exclusion threshold of 400 Da to limit our detection to large metabolites. In doing so we excluded well-known *Salinispora* molecules such as the saliniketals and salinisporamides, although we did identify some small molecules that formed oligomers in the gas phase of the mass spectrometer, such as ammosamide B ($[3M + Na]^+$: 896.14 Da). Several of these known compounds, such as the enterocins,[20] were identified in strains that were not previously known as producers. A large number of putative new analogues of known compounds were also identified in the network. For example there are three analogues of salinamide that do not correspond to any library variants, one of which, based on the parent mass, likely corresponds to salinamide F.[21]

Four molecular families were produced by both *Salinispora* and *Streptomyces*. The desferrioxamine family, a group of hydroxamate siderophores,[22] includes over 50 congeners, including acylated derivatives that have only been detected from *Streptomyces* strains before.[11b,23] The staurosporine molecular family[24] consists of a total of 11 members, mainly produced by *Salinispora* strains, but hydroxystaurosporine was also found in *Streptomyces* CNQ-149. This molecular family is produced by a large portion of the *Salinispora* strains: staurosporine was detected from a total of 61 strains, 56 *S. arenicola* and 5 *S. pacifica*, while the gene cluster is present in all 62 *S. arenicola* and 16 *S. pacifica* strains.[25] The rifamycin molecular family consists of 25 members mostly detected in *S. arenicola* strains;[26] however, rifamycin W was also produced by one *Streptomyces* strain. The rosamicin family, a group of glycosylated polyketides, is produced by five *Salinispora* strains and one *Streptomyces* strain. Having initially detected this family in this data set, we recently isolated and characterized three novel rosamicins and their biosynthetic byproducts salinipyrones and pacificanones from *S. pacifica* CNS-237.[27]

Analysis of Network by Genus and Species and Principal Coordinates Analysis (PCoA) Visualization of the Overall Chemical Diversity. As might be expected for bacteria in different families, there was little overlap in the parent ions detected (16.4%) in the *Salinispora* and *Streptomyces* extracts (Figures 2A, S2). Those shared between the two genera include lipids but also natural products, including the desferrioxamines and the rosamicins, as described above. *Salinispora* and *Streptomyces* extracts in general have similar molecular diversities, averaging 10.4 and 11.5 different nodes per sample, respectively. However, the larger number of *Salinispora* extracts examined accounts for the relatively high percentage of the total (51.3%) that is specific to this genus. Within the closely related and well-defined *Salinispora* strains, it is possible to analyze the distribution of extracted metabolites by species (Figures 2B, S3). As described above, *Salinispora* is known to produce species-specific metabolites.[17] In this work, the production of known molecules follows a similar pattern to that in previous studies showing rifamycins as a strong marker for *S. arenicola*, while lomaiviticins are produced only by *S. pacifica* and *S. tropica*. Staurosporines are produced by *S. arenicola* and *S. pacifica*, while desferrioxamines are produced by all three species (Table S2). The wide distribution of desferrioxamines and staurosporines is reflected in the corresponding gene cluster patterns, where, of the 120 strains, 92 and 78, respectively, possess these gene clusters.

When the whole *Salinispora* network is analyzed for metabolite production by species, it is apparent that more than half of the total nodes (57.6%) were found in only one of the three species. This observation clearly shows that there can be great differences in secondary metabolism even among very closely related species. Less than 10% of the nodes are produced by all three species, suggesting secondary metabolism is more a species-defining trait than a genus characteristic.

In addition, network consensus nodes in each sample were subjected to multivariate analyses. Intrasample distances were determined using both the Gower distance metric and the random forest classifier and visualized using PCoA dimensional reduction. The PCoA analysis showed that *Salinispora* and *Streptomyces* samples occupy mutually exclusive areas of this chemical space outside of a shared core (Figure 2C). Since the unsupervised Gower PCoA analysis (Figure 2C, left) did not show a clear grouping pattern between most metadata labels, we turned to the random forest classifier (Figure 2C, right).[28] With the PCoA approach, one can use the random forest algorithm's ability to classify samples into specified (supervised) or unspecified (unsupervised) groups as the basis of a dissimilarity metric retrieved from proximity matrices.[29] We applied this technique to the 360 *Salinispora*-derived samples, classifying on the basis of species (Figure 2D). The random forest classifier was able to differentiate the *Salinispora* species with an accuracy of 87%, showing that the metabolic information captured by mass spectrometry provides a consistent fingerprint of each species. Interestingly, the top drivers for this species-specific PCoA separation were analogues of the previously discussed bioactive alkaloid staurosporine, while the influence of media components for this analysis could be excluded (Figure S4). The PCoA analysis thus supports the observations from the molecular network and helps in building a global and comprehensive metabolic picture for the genus *Salinispora*.

Impact of Additional Attributes on the Network. *Strain.* For the prioritization of strains for chemical analysis, a direct comparison of their metabolic profiles is beneficial. Within the network, each strain contributes to a certain number of nodes, thus giving a direct measure of extracted molecular diversity. Because the strains from each genus were grown and extracted under the same conditions, it was possible to compare

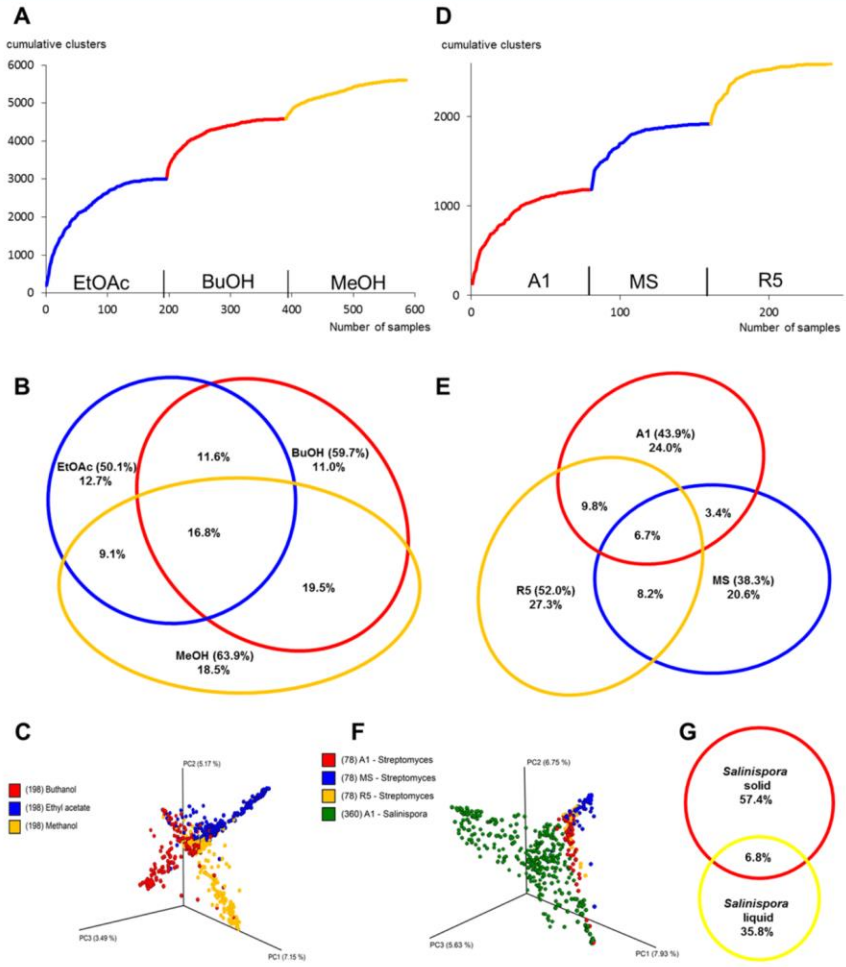


Figure 3. Effects of additional attributes on molecular diversity-network analysis. Cumulative consensus curves for added unique spectra by each additional solvent (A) and Venn diagram for node distributions in the network for each solvent (B). (C) Supervised random forest analysis of all samples classified by solvent. Cumulative consensus curves (D) and Venn diagram (E) for each medium (only from the *Streptomyces* extracts). Percentages are shown for each sector, with the total percentage for each treatment in parentheses. (F) Supervised random forest analysis of all samples classified by growth medium. (G) Comparison of liquid and solid extraction for 30 *Salinispora* strains (solvent: EtOAc).

them in the network. One example of a chemically rich strain is *Streptomyces* sp. CNQ-329, which contributes to 451 nodes in the network (Table S1). This number can be further broken down into nodes by medium, revealing that medium R5 gives the most diversity with 339 nodes. Looking at extraction solvents, from the R5 samples, MeOH and *n*-butanol (BuOH) provide the most chemical diversity, with 269 and 252 nodes per sample, respectively. From the *Salinispora* strains, *S. arenicola* CNT-798 yielded the highest chemical diversity, contributing to 288 nodes in total. In this case, each solvent extracts a similar amount of molecular diversity (239, 245, and 261 nodes from BuOH, EtOAc, and MeOH, respectively). It is important to note that the individual *Salinispora* strains were grown on just one medium, giving rise to the smaller total number of nodes compared to *Streptomyces* strains. Conversely, some strains yielded very little chemical diversity, with the approach used. *Salinispora pacifica* CNY-703, for example, contributed to only six nodes in the network. These results help

to prioritize strains with higher chemical diversity and also to define the culture and extraction conditions that give the highest metabolite yields.

Solvent and Medium. When the network was sorted by solvent, a comparable number of nodes were shown to be extracted by each of the three solvents (Figures 3A,B, S5). This can be visualized with an accumulation curve describing the number of unique clusters added by each additional sample, incorporating all of those extracted with EtOAc, before adding those from BuOH and then MeOH, reflecting the order they were used in the extraction (Figure 3A). The inflection upon addition of spectra from samples extracted with a new solvent indicates an influx of new clusters. A Venn diagram of the nodes originating from each solvent shows that almost half (42%) of the nodes were extracted by just one solvent, and, of the three solvents, MeOH yielded the highest number of unique nodes (Figure 3B). A total of 57 molecular families were extracted by a single solvent (EtOAc = 12, BuOH = 20, and MeOH = 25), which may have been missed in the network by the exclusion of any of these solvents. Several of the known compounds were extracted by only one solvent, including salinamide E (BuOH), antimycin A1 (MeOH), and arenimycin A (EtOAc). Three analogues of the lomaiviticin family, including lomaiviticin A, were extracted by only EtOAc, and two compounds of the cyclomarin family by MeOH only. When random forest dissimilarities are visualized in PCoA space, the distinctions caused by solvent differences spread samples in distinct directions (Figure 3C). The three solvents are likely able to capture a common core metabolome, but also allow for capturing solvent-specific metabolites. These results clearly demonstrate that using three extraction solvents, instead of one, greatly enhances the molecular diversity that can be detected by mass spectrometry.

To gain insight into medium-dependent metabolomics, all *Streptomyces* strains were grown on three different media (A1, MS, and R5). Extracts from these cultures were networked together and then analyzed for extracted nodes by culture medium (Figures 3D,E, S1). In this case, the generated accumulation curve shows a similar trend to that with the solvent extraction analysis for all samples, showing a rapid increase in molecular diversity with each change of medium (Figure 3D). Analysis of nodes in the network by medium reveals that over 70% of the nodes were extracted from just one of the three media (Figure 3E).

This observation corroborates observations from the OSMAC method,[3] that culture medium is a key factor in secondary metabolite biosynthesis. We identified a total of 89 clusters in the *Streptomyces* network that were produced on just one medium (35 on A1, 29 on MS, and 25 on R5), including some of the detected standards. As we only evaluated metabolites with a molecular weight over 400 Da, we do not anticipate inclusion of byproducts of core metabolism. To provide some examples, most of the rosamicin molecular family was produced only on A1, which was also necessary for production of five of the seven known salinamides, including salinamides A and E. Additionally, many of the detected desferrioxamine family analogues were produced only on medium R5, as were the entire alteramide and antimycin molecular families. All samples were classified with supervised random forest by the media information (Figure 3F), and differences are clearly seen to spread samples in distinct directions in the PCoA space. Thus, this analysis rapidly visualizes how much the metabolic repertoire is dependent on medium composition.

Solid versus Liquid Media. Previously, 35 *Salinispora* strains were grown in liquid A1 medium, extracted and analyzed in a similar way to this project.[19] When the comparable data from this previous work is networked with the same 30 sequenced strains from solid A1 media, we observed less than 7% overlap of extracted metabolites (Figures 3G, S6). Interestingly, most of the metabolic overlap belongs to known molecular families that could be dereplicated by comparison to standards in the GNPS database. The cyclomarins are represented in two adducts in the network, the sodiated form and the dehydrated and protonated form (Figure 1). These adducts display significantly different fragmentation patterns, and thus they form two distinct molecular families. Analysis of both molecular families shows that only cyclomarin A was extracted from both solid and liquid cultures. We observed that some cyclomarin analogues were extracted from only the liquid or the solid cultures, thereby clearly demonstrating the culture-dependent variability in production of compounds of the same class. In the case of the arenicolide molecules, we detected only an unprecedented hydrated analogue of arenicolide A on the solid growth medium (Figure S7), while six arenicolide analogues were produced in liquid medium.

Taken together, the comparison between growth on solid and liquid media for 30 *Salinispora* strains shows production of almost entirely different chemistry. The observed metabolic differences of liquid versus solid media suggest that a network with liquid culturing data of all 146 strains would look significantly different and may help to capture a broader map of the metabolic potential of these bacteria.

Location. The network was also queried for molecular families produced by several strains from one collection location or locations relatively close to each other (Figure S8). One example is a molecular cluster found to be extracted from two strains collected from Guam, *Salinispora pacifica* CNQ-768 and *Streptomyces* sp. CNQ-865 with a parent mass of m/z 878.152. This observation implies that the corresponding gene cluster, which we have yet to identify, is shared between these two geographically related strains. Although other strains may also have the cluster, it was apparently not expressed under these experimental conditions. A second cluster, consisting of 16 nodes, is derived from 10 strains of *S. pacifica* and *S. arenicola* isolated from expeditions to Hawaii and Fiji in the central Pacific. A third cluster, consisting of four nodes, is produced by four Fijian *S. pacifica* strains. One of the largest molecular families consisted of 152 nodes, in which 137 were extracted from strains collected from Hawaii. The remaining 15 nodes were derived from *Salinispora* and *Streptomyces* strains isolated from the Pacific (Fiji, Guam, Palmyra, San Diego, Channel Islands, and the Sea of Cortez). Thus, it appears that the biosynthetic genes responsible for these metabolites are locally restricted to strains in our collection isolated from the Pacific Ocean.

Culture Time. Another dimension that can be added to the analyses is the length of culturing before extraction. This is particularly valuable for experiments seeking to determine the best time point for preparative isolation of molecules or to observe formation and changes of compound patterns over time.[6b] To gain insight into temporal changes in natural product production, we grew *S. arenicola* CNH-877 in four different liquid media and extracted at three different time points: 14, 21, and 28 days postinoculation (Figure S9). We

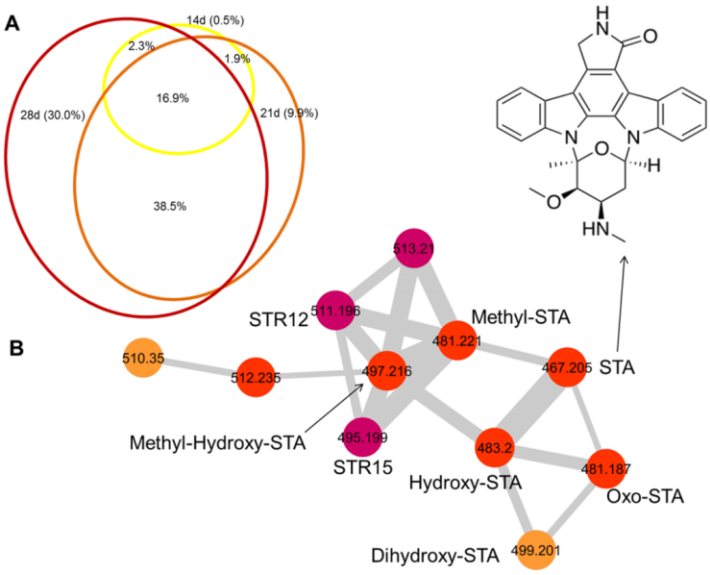


Figure 4. Time-dependent changes in natural product distribution in *Salinispora arenicola* CNH-877, grown in ISP2. (A) Venn diagram representing node distributions in the molecular network at three time points (14, 21, and 28 days). (B) The staurosporine (STA) molecular family in the network. Nodes represent masses (m/z), and edge thickness corresponds to cosine score between the nodes. Highlighted in red are masses that are present in samples taken at 14, 21, and 28 days. Orange nodes represent masses present only after 21 and 28 days; violet nodes are present only after 28 days.

observed that the number of extracted ions steadily increased over time (Figure 4A) and that after 28 days there is far higher molecular diversity than at 14 or 21 days. The temporal changes in metabolite production and distribution can be exemplified with the staurosporine molecular family group (Figure 4B, Table S3).[24] The staurosporine (STA) molecular family in the ISP2 network consists of 11 nodes, eight of which can be connected to known STA analogues by exact mass. Analysis of the nodes reveals a steady number of spectral counts for STA and oxo-STA across the different samples, while there is an increase in spectra corresponding to hydroxy-STA, methyl-STA, and methyl-hydroxy-STA. Dihydroxy-STA was detected after 21 days and with an increase in spectral counts in the last time point. The production of minor analogues, whose masses were previously reported from a *Saccharothrix* strain, was detected only after 28 days.[30] These results illustrate the biosynthetic changes and intramolecular conversions within a family of related molecules over time.[6b,31]

■ CONCLUSIONS

Natural products discovery and structure elucidation is a time-consuming and sometimes inefficient process fraught with the rediscovery of known compounds. To advance natural product research, it is thus crucial to develop rational and effective strategies for the discovery of novel natural products entities and scaffolds. Emerging concepts such as genome mining and MS-guided metabolomics have accelerated this process in recent years. We believe that one efficient strategy in a rational, state-of-the-art drug discovery program is the quick assessment of the metabolic capacities of natural product producers under various lab conditions, coupled with correlation of genomic and metabolic data for accelerated discovery and dereplication processes. In this study, we generated a comprehensive picture of the molecular diversity from 146 actinomycete strains from the marine environment. The selection of strains with sequenced genomes will help in the utilization of this data in future studies. To maximize molecular diversity in an efficient manner, we developed a simple culturing and extraction protocol and evaluated the variables that influence metabolite identification.

In previous studies on the molecular diversity of *Salinispora*, much smaller numbers of strains were grown on just one medium and extracted under just one condition.[19,32] Thus, the data in this study, generated from 120 *Salinispora* strains with three extraction solvents, and the visualization in a molecular network give a more comprehensive and detailed picture of the *Salinispora* metabolome. The species-specific production of many known and unknown metabolites is well reflected in the network and clearly visualized by supervised random forest analysis. To produce an even larger picture with two "talented" genera, 26 *Streptomyces* strains were grown and extracted in the same way as the *Salinispora* strains, but on three media instead of one. In total, 15 structurally diverse molecular families could be annotated as known compound classes in the network, including numerous as yet undescribed congeners. The size of the network and diversity of the samples allowed us to observe

how attributes, such as growth and extraction conditions, affect chemical diversity. It also allowed us to quickly compare strains and prioritize chemically rich isolates for more detailed profiling, as well as informing the most valuable growth and extraction conditions.

In light of the OSMAC approach, the changes of the metabolomes can be rationalized, as the different media represent different environments that the bacteria are exposed to, requiring them to alter their behavior. The culturing in liquid versus solid media, comparable to environments on surfaces versus in suspension, greatly influences the production of specialized metabolites. To our knowledge, there has not been a systematic investigation of the effect of culturing and extraction parameters on a larger number of strains with mass spectrometric tools.[33] Here, the expanded molecular diversity that is added to the network by each additional treatment (medium, agar, solvent) shows clearly how much molecular diversity can be missed when just one medium, solvent, or time point is used to assess the metabolic capacity. Parameters such as time of extraction and solvent are of great importance for the extracted metabolite spectrum and should always be kept in mind when creating a natural products isolation workflow. With molecular networking, optimization of culturing and extraction parameters can now be assessed quickly and implemented early into the discovery workflow. The results of this study encourage further applications of the OSMAC approach by natural product chemists, and this workflow can be applied to microbes across all three domains of life (Eukaryotes, Prokaryotes, and Archaea). To conclude, this network provides an extensive survey of the biosynthetic capacity of this strain collection, and, with the GNPS database continuing to expand, this will provide a living data set to inform future rational and automated natural product discovery efforts in the genera *Salinispora* and *Streptomyces*.

■ EXPERIMENTAL SECTION

Culturing and Extraction. *Salinispora* strains were cultured from frozen stock cultures on 10 mL of A1 agar (six-well plates, 9 cm^2). Ten mg/mL phenol red was added to the medium to indicate the beginning of stationary phase when the color of the medium shifted from yellow to red,[34] at which point they were extracted. *Streptomyces* strains were cultured in 24-well plates on A1, MS, and R5 agar for 7 days before extraction. For the extraction, a plug of agar and cell lawn was removed and crushed with a glass pipet. First, agar and cells were washed with 500 μL of H_2O in an ultrasonic bath (30 min) to remove salts. Then, agar and cells were extracted subsequently with 500 μL of EtOAc, *n*-BuOH, and MeOH each (ultrasonic bath, 5 min). All *Streptomyces* samples were extracted by vortexing for 30 s with each solvent. After each extraction the solvent was evaporated, and the residue was redissolved in 1 mL of MeOH and filtered through a 0.2 μm membrane into HPLC vials. Solvent blanks were generated by extracting media using the same protocols. For this study, all strains were grown and analyzed once. For the time course experiment, *S. arenicola* CNH-877, CNY-011, CNS-690, and CNS-694 were grown in either liquid A1, ISP2, MB, or production medium (1% soytone, 1% soluble starch, 1% maltose) supplemented with Instant Ocean sea salt. A 1 mL amount of the cultures was extracted after 7, 14, 21, and 28 days with 1 mL of EtOAc and BuOH, and the solvent treated as above.

HPLC-MS. Samples were analyzed using an Agilent 6530 Accurate-Mass Q-TOF spectrometer coupled to an Agilent 1260 LC system. A Phenomenex Luna C18 HPLC column (2.6 mm, 150 × 4.6 mm) was used under the following LC conditions with 0.1% TFA: 1−5 min (10% MeCN in H_2O), 5−26 min (10−100% MeCN), 26−28 min (100% MeCN). The divert valve was set to waste for the first 4 min. Q-TOF MS settings during the LC gradient were as follows: positive ion mode mass range 300−1700 m/z, static exclusion 300−400 m/z, MS scan rate 1/s, MS/MS scan rate 3/s, fixed collision energy 20 keV; source gas temperature 300 °C, gas flow 11 L/min, nebulizer 45 psig, scan source parameters: VCap 3000, fragmentor 100, skimmer 1 65, octopoleRFPeak 750. The MS was autotuned using Agilent tuning solution in positive mode before each measurement. MS data were analyzed with MassHunter software (Agilent).

Molecular Networking and Data Analysis. All MS/MS data were converted from Agilent MassHunter data files (.d) to mzXML file format using the software Trans-Proteomic pipeline (Institute for Systems Biology).[35] The data were transferred onto the GNPS server (gnps.ucsd.edu), and molecular networking was performed using the GNPS data analysis workflow using the spectral clustering algorithm.[6a] Sample attributes were linked to the data (146 strains, 2 genera, 3 species, 3 media, 16 locations, 3 solvents). Different parameters (cosine, minimum matched peaks) were evaluated to determine the best networking conditions. Finally, a cosine of 0.5 and a minimum number of matched peaks of 4 were chosen for further analyses. The chosen parameters include mass tolerance for fragment peaks (0.5 Da), parent mass tolerance (2.0 Da), a minimum cluster size of 2, and a maximum cluster size of 250. These settings yielded the highest number of connected nodes with no standards having clustering with other standards. To facilitate network analysis, all nodes that contained ions that were present in the media controls were subtracted from the networks. The spectral networks were imported into Cytoscape 3.1.0[36] and visualized using the force-directed layout. Nodes represent parent masses, and edge thickness corresponds to cosine score. Group and attributes files and cumulative consensus curves were generated according to the GNPS documentations (https://bix-lab.ucsd.edu/display/Public/GNPS+Documentation+Page). To generate cumulative consensus curves, the network was rerun using the same parameters with input files being allocated to the spectrum file groups based on attribute. The data are publically accessible as MassIVE data sets MSV000078836 and MSV000078839. Ellipsoid area-proportional Venn diagrams were generated with the tool eulerAPE v3 (http://www.eulerdiagrams.org/eulerAPE).[37] Bioinformatic genome, gene cluster, and domain analysis was performed using the tools antiSMASH 3.0 (antismash.secondarymetabolites.org)[38] and NaPDoS (napdos.ucsd.edu).[39]

Statistical Analysis. The intensity of the precursor ions of MS/MS clusters was exported through the "Create Cluster Buckets" option on GNPS (gnps.ucsd.edu) data analysis Advanced Output Options. The table was used to perform unsupervised and supervised analysis using R statistical environment[40] and Qiime bioinformatics pipeline.[41] The unsupervised analysis consisted of calculating Gower distance with the R package VEGAN[42] and using the distance matrix to perform PCoA using Qiime and visualized using EMPeror.[43] The supervised analysis consisted of training classifiers for different partitions of the data (e.g., classifying samples according to solvent extraction, growth media, or species labels based on whole metabolomics profile). The random forest classifier was used through the RandomForest package.[29] The model accuracies were calculated by subsampling the data in training and test data sets with the package Caret.[44] The random forest sample proximity values were used to calculate sample to sample dissimilarities and repeat the PCoA analysis for classification.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.6b00722.

Tables S1−S3, Figures S1−S9 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: pdorrestein@ucsd.edu.
*E-mail: bsmoore@ucsd.edu.

ORCID

Bradley S. Moore: 0000-0002-4652-1253

Present Addresses

[⊥]Institute of Pharmaceutical Biology, University of Bonn, 53115 Bonn, Germany.

[∥]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, Oxfordshire, OX1 3RB, UK.

Author Contributions

[Δ]M. Crüsemann and E. C. O'Neill contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the NIH (R01-GM085770 to B.S.M. and P.R.J. and R01-GM097509 to B.S.M. and P.C.D.), the São Paulo Research Foundation (FAPESP-2015/03348-3 to R.R.d.S.), and a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft (CR 464-1 to M.C.).

■ DEDICATION

Dedicated to Professor Phil Crews, of the University of California, Santa Cruz, for his pioneering work on bioactive natural products.

■ REFERENCES

(1) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012**, *75*, 311−335.

(2) (a) Doroghazi, J. R.; Metcalf, W. W.*BMC Genomics***2013**, *14*, 61110.1186/1471-2164-14-611. (b) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J.; Birren, B. W.; Takano, E.; Sali, A.; Linington, R. G.; Fischbach, M. A. *Cell* **2014**, *158*, 412−421. (c) Ziemert, N.; Lechner, A.; Wietz, M.; Millán-Aguiñaga, N.; Chavarria, K. L.; Jensen, P. R. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E1130−E1139. (d) Hadjithomas, M.; Chen, I.-M. A.; Chu, K.; Ratner, A.; Palaniappan, K.; Szeto, E.; Huang, J.; Reddy, T. B. K.; Cimermančič, P.; Fischbach, M. A.; Ivanova, N. N.; Markowitz, V. M.; Kyrpides, N. C.; Pati, A. *mBio* **2015**, *6*, e00932.

(3) Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A. *ChemBioChem* **2002**, *3*, 619−627.

(4) (a) Yuan, C.; Guo, Y. H.; Wang, H. Y.; Ma, X. J.; Jiang, T.; Zhao, J. L.; Zou, Z. M.; Ding, G. *Sci. Rep.* **2016**, *6*, 19350. (b) Hewage, R. T.; Aree, T.; Mahidol, C.; Ruchirawat, S.; Kittakoop, P. *Phytochemistry* **2014**, *108*, 87−94. (c) Wang, Q. X.; Bao, L.; Yang, X. L.; Guo, H.; Ren, B.; Guo, L. D.; Song, F. H.; Wang, W. Z.; Liu, H. W.; Zhang, L. X. *Fitoterapia* **2013**, *85*, 8−13.

(5) (a) Doroghazi, J. R.; Albright, J. C.; Goering, A. W.; Ju, K. S.; Haines, R. R.; Tchalukov, K. A.; Labeda, D. P.; Kelleher, N. L.; Metcalf, W. W. *Nat. Chem. Biol.* **2014**, *10*, 963−968. (b) Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. *Anal. Chem.* **2014**, *86*, 10780−10788. (c) Krug, D.; Müller, R. *Nat. Prod. Rep.* **2014**, *31*, 768−783. (d) Wu, C.; Kim, H. K.; van Wezel, G. P.; Choi, Y. H. *Drug Discovery Today: Technol.* **2015**, *13*, 11−17. (e) Gaudencio, S. P.; Pereira, F. *Nat. Prod. Rep.* **2015**, *32*, 779−810.

(6) (a) Guthals, A.; Watrous, J. D.; Dorrestein, P. C.; Bandeira, N. *Mol. BioSyst.* **2012**, *8*, 2535−2544. (b) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1743−1752.

(7) Nguyen, D. D.; Wu, C.-H.; Moree, W. J.; Lamsa, A.; Medema, M. H.; Zhao, X.; Gavilan, R. G.; Aparicio, M.; Atencio, L.; Jackson, C.; Ballesteros, J.; Sanchez, J.; Watrous, J. D.; Phelan, V. V.; van de Wiel, C.; Kersten, R. D.; Mehnaz, S.; De Mot, R.; Shank, E. A.; Charusanti, P.; Nagarajan, H.; Duggan, B. M.; Moore, B. S.; Bandeira, N.; Palsson, B. Ø.; Pogliano, K.; Gutiérrez, M.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2611−E2620.

(8) (a) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. *J. Nat. Prod.* **2013**, *76*, 1686−1699. (b) Winnikoff, J. R.; Glukhov, E.; Watrous, J.; Dorrestein, P. C.; Gerwick, W. H. *J. Antibiot.* **2014**, *67*, 105−112.

(9) (a) Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.; Duggan, B. M.; Di Marzo, V.; Sherman, D. H.; Dorrestein, P. C.; Gerwick, L.; Gerwick, W. H. *J. Nat. Prod.* **2015**, *78*, 1671−1682. (b) Liaw, C.-C.; Chen, P.-C.; Shih, C.-J.; Tseng, S.-P.; Lai, Y.-M.; Hsu, C.-H.; Dorrestein, P. C.; Yang, Y.-L. *Sci. Rep.* **2015**, *5*, 12856. (c) Henke, M. T.; Soukup, A. A.; Goering, A. W.; McClure, R. A.; Thomson, R. J.; Keller, N. P.; Kelleher, N. L. *ACS Chem. Biol.* **2016**, *11*, 2117.

(10) (a) Schorn, M.; Zettler, J.; Noel, J. P.; Dorrestein, P. C.; Moore, B. S.; Kaysser, L. *ACS Chem. Biol.* **2013**, *9*, 301−309. (b) Crone, W. J. K.; Vior, N. M.; Santos-Aberturas, J.; Schmitz, L. G.; Leeper, F. J.; Truman, A. W.*Angew. Chem., Int. Ed.***2016**, *55*, 963910.1002/anie.201604304. (c) Wu, C.; Medema, M. H.; Lakamp, R. M.; Zhang, L.; Dorrestein, P. C.; Choi, Y. H.; van Wezel, G. P. *ACS Chem. Biol.* **2016**, *11*, 478−490.

(11) (a) Moree, W. J.; Phelan, V. V.; Wu, C.-H.; Bandeira, N.; Cornett, D. S.; Duggan, B. M.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13811−13816. (b) Traxler, M. F.; Watrous, J. D.; Alexandrov, T.; Dorrestein, P. C.; Kolter, R.*mBio***2013**, *4*, e00459-1310.1128/mBio.00459-13. (c) Briand, E.; Bormans, M.; Gugger, M.; Dorrestein, P. C.; Gerwick, W. H. *Environ. Microbiol.* **2016**, *11*, 384−400.

(12) Wang, M.; Carver, J.; Phelan, V.; Sanchez, L.; Garg, N.; Peng, Y.; Nguyen, D.; Watrous, J.; Kapono, C.; Luzzatto-Knaan, T.; et al. *Nat. Biotechnol.* **2016**, *34*, 828−837.

(13) Nett, M.; Ikeda, H.; Moore, B. S. *Nat. Prod. Rep.* **2009**, *26*, 1362−1384.

(14) (a) Maldonado, L. A.; Fenical, W.; Jensen, P. R.; Kauffman, C. A.; Mincer, T. J.; Ward, A. C.; Bull, A. T.; Goodfellow, M. *Int. J. Syst. Evol. Microbiol.* **2005**, *55*, 1759−1766. (b) Ahmed, L.; Jensen, P. R.; Freel, K. C.; Brown, R.; Jones, A. L.; Kim, B. Y.; Goodfellow, M. *Antonie van Leeuwenhoek* **2013**, *103*, 1069−1078.

(15) Freel, K. C.; Millán-Aguiñaga, N.; Jensen, P. R. *Appl. Environ. Microbiol.* **2013**, *79*, 5997−6005.

(16) Jensen, P. R.; Moore, B. S.; Fenical, W. *Nat. Prod. Rep.* **2015**, *32*, 738−751.

(17) Jensen, P. R.; Williams, P. G.; Oh, D.-C.; Zeigler, L.; Fenical, W. *Appl. Environ. Microbiol.* **2007**, *73*, 1146−1152.

(18) (a) Udwary, D. W.; Zeigler, L.; Asolkar, R. N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P. R.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10376−10381. (b) Penn, K.; Jenkins, C.; Nett, M.; Udwary, D. W.; Gontang, E. A.; McGlinchey, R. P.; Foster, B.; Lapidus, A.; Podell, S.; Allen, E. E.; Moore, B. S.; Jensen, P. R. *ISME J.* **2009**, *3*, 1193−1203.

(19) Duncan, K. R.; Crüsemann, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, B. S.; Dorrestein, P. C.; Jensen, P. R. *Chem. Biol.* **2015**, *22*, 460−471.

(20) Piel, J.; Hertweck, C.; Shipley, P. R.; Hunt, D. M.; Newman, M. S.; Moore, B. S. *Chem. Biol.* **2000**, *7*, 943−955.

(21) Ray, L.; Yamanaka, K.; Moore, B. S. *Angew. Chem., Int. Ed.* **2016**, *55*, 364−367.

(22) (a) Challis, G. L. *ChemBioChem* **2005**, *6*, 601−611. (b) Roberts, A. A.; Schultz, A. W.; Kersten, R. D.; Dorrestein, P. C.; Moore, B. S. *FEMS Microbiol. Lett.* **2012**, *335*, 95−103.

(23) Sidebottom, A. M.; Johnson, A. R.; Karty, J. A.; Trader, D. J.; Carlson, E. E. *ACS Chem. Biol.* **2013**, *8*, 2009−2016.

(24) Park, B. S.; Abdel-Azeem, A. Z.; Al-Sanea, M. M.; Yoo, K. H.; Tae, J. S.; Lee, S. H. *Curr. Med. Chem.* **2013**, *20*, 3872−3902.

(25) Freel, K. C.; Nam, S.-J.; Fenical, W.; Jensen, P. R. *Appl. Environ. Microbiol.* **2011**, *77*, 7261−7270.

(26) Wilson, M. C.; Gulder, T. A. M.; Mahmud, T.; Moore, B. S. *J. Am. Chem. Soc.* **2010**, *132*, 12757−12765.
(27) Awakawa, T.; Crüsemann, M.; Munguia, J.; Ziemert, N.; Nizet, V.; Fenical, W.; Moore, B. S. *ChemBioChem* **2015**, *16*, 1443−1447.
(28) Ramette, A. *FEMS Microbiol. Ecol.* **2007**, *62*, 142−160.
(29) Liaw, A.; Wiener, M. *R News* **2002**, *2*, 18−22.
(30) Barrabee, E. B.; Horan, A. C.; Gentile, F. A.; Patel, M. G. Indolocarbazoles from *Saccharothrix aerocolonigenes copiosa* subsp. nov SCC 1951 ATCC 53856, USPTO. United States, Schering Corporation, 1997.
(31) Esquenazi, E.; Jones, A. C.; Byrum, T.; Dorrestein, P. C.; Gerwick, W. H. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5226−5231.
(32) Bose, U.; Hewavitharana, A. K.; Vidgen, M. E.; Ng, Y. K.; Shaw, P. N.; Fuerst, J. A.; Hodson, M. P. *PLoS One* **2014**, *9*, e91488.
(33) Reen, F.; Romano, S.; Dobson, A.; Gara, F. *Mar. Drugs* **2015**, *13*, 4754−4783.
(34) Wolfe, A. J. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 12−50.
(35) (a) Keller, A.; Eng, J.; Zhang, N.; Li, X. j.; Aebersold, R.*Mol. Syst. Biol.***2005**, *1*,.E110.1038/msb4100024 (b) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Farrah, T.; Lam, H.; Tasman, N.; Sun, Z.; Nilsson, E.; Pratt, B.; Prazen, B.; Eng, J. K.; Martin, D. B.; Nesvizhskii, A. I.; Aebersold, R. *Proteomics* **2010**, *10*, 1150−1159.
(36) Cline, M. S.; Smoot, M.; Cerami, E.; Kuchinsky, A.; Landys, N.; Workman, C.; Christmas, R.; Avila-Campilo, I.; Creech, M.; Gross, B.; Hanspers, K.; Isserlin, R.; Kelley, R.; Killcoyne, S.; Lotia, S.; Maere, S.; Morris, J.; Ono, K.; Pavlovic, V.; Pico, A. R.; Vailaya, A.; Wang, P. L.; Adler, A.; Conklin, B. R.; Hood, L.; Kuiper, M.; Sander, C.; Schmulevich, I.; Schwikowski, B.; Warner, G. J.; Ideker, T.; Bader, G. D. *Nat. Protoc.* **2007**, *2*, 2366−2382.
(37) Micallef, L.; Rodgers, P. *PLoS One* **2014**, *9*, e101717.
(38) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Muller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43*, W237−243.
(39) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. *PLoS One* **2012**, *7*, e34064.
(40) R_Core_Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. In 2013.
(41) Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Peña, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. *Nat. Methods* **2010**, *7*, 335−336.
(42) Dixon, P. *Journal of Vegetation Science* **2003**, *14*, 927−930.
(43) Vazquez-Baeza, Y.; Pirrung, M.; Gonzalez, A.; Knight, R. *GigaScience* **2013**, *2*, 16.
(44) Kuhn, M. *J. Statistical Software* **2008**, *28*, 26.

SUPPORTING INFORMATION FOR

Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains based on Growth and Extraction Protocols

Max Crüsemann,[†,⊥,△] Ellis C. O'Neill,[†,∥,△] Charles B. Larson,[†] Alexey V. Melnik,[‡] Dimitrios J. Floros,[‡] Ricardo R. da Silva,[‡,§] Paul R. Jensen,[†] Pieter C. Dorrestein,[†,‡] and Bradley S. Moore[†,‡*]*

[†]Center for Marine Biotechnology & Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA

[‡]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA

[§]Research Support Center in Natural and Synthetic Products, Department of Physics and Chemistry, Faculty of Pharmaceutical Sciences, University of São Paulo, Ribeirão Preto, 14040-903, Brazil

Contents **Page**

Table S1: 3

Table S2: 4

Figure S1: 5

Figure S2: 6

Figure S3: 7

Figure S4: 8

Figure S5: 9

Figure S6: 10

Figure S7: 11

Figure S8: 12

Figure S9: 13

Table S3: 14

References: 15

Table S1: Overview of the bacterial strains used in this study.

Species	Strain	Sampling Location	IMG JGI submission ID	Number of nodes in network
Salinispora tropica	CNB440	Bahamas	369723 (NCBI Taxon ID)	236
	CNB476	Bahamas	11057	142
	CNB536	Bahamas	11058	63
	CNH898	Bahamas	9582	185
	CNH699	Bahamas	11506	167
	CNB187	Bahamas	9669	230
	CNS416	Bahamas	10982	102
	CNT250	Bahamas	20026	156
	CNT261	Bahamas	14345	14
	CNY012	Bahamas	20025	238
	CNY678	Cozumel	35127	222
	CNY681	Yucatan	35126	29
Salinispora arenicola	CNB456	Bahamas	11056	163
	CNB527	Bahamas	9581	113
	CNH643	Bahamas	34967	27
	CNH646	Bahamas	9688	244
	CNH713	Red Sea	37236	74
	CNH718	Red Sea	35123	8
	CNH877	Bahamas	11857	111
	CNH905	Bahamas	9690	221
	CNH941	Sea of Cortez	9700	270
	CNH962	Sea of Cortez	11938	20
	CNH963	Sea of Cortez	14289	18
	CNH964	Sea of Cortez	9614	232
	CNH994	Sea of Cortez	35122	93
	CNP105	Sea of Cortez	11185	127
	CNP150	Sea of Cortez	11186	97
	CNQ748	Guam	9687	92
	CNQ884	Guam	34969	57
	CNR107	Guam	11939	189
	CNR425	Guam	15667	169
	CNR921	Palau	9715	184
	CNS051	Palau	11187	95
	CNS206	Palau	391037 (NCBI Taxon ID)	210
	CNS243	Palau	14327	84
	CNS296	Palau	36410	34
	CNS299	Palau	14344	48
	CNS325	Palau	37242	30
	CNS342	Fiji	37324	26
	CNS673	Fiji	11949	210
	CNS744	Fiji	11188	8
	CNS820	Fiji	36417	20
	CNS848	Fiji	38495	141
	CNS991 (DSM 45549)	Fiji	10070	192
	CNT005	Fiji	10949	99
	CNT088	Fiji	24413	68
	CNT798	Hawaii	9693	288
	CNT799	Hawaii	14925	182
	CNT800	Hawaii	9575	179
	CNT849	Hawaii	11184	129
	CNT850	Hawaii	9637	264
	CNT857	Hawaii	9616	195
	CNT859	Hawaii	11088	165
	CNX481	Palmyra	11189	85
	CNX482	Palmyra	9639	109
	CNX536	Palmyra	9695	71
	CNX814	Palmyra	10970	115
	CNX891	Palmyra	9694	189
	CNY011	Bahamas	10971	104
	CNY230	Fiji	35133	28
	CNY231	Fiji	11233	279
	CNY234	Fiji	11940	83
	CNY237	Fiji	10981	72
	CNY244	Fiji	35128	17
	CNY256	Fiji	11234	105
	CNY260	Fiji	11235	161
	CNY280	Fiji	10972	102
	CNY281	Fiji	35121	54
	CNY282	Fiji	12935	113
	CNY486	Puerto Vallarta	35125	82
	CNY679	Yucatan	35131	34
	CNY685	Banco Chinchorro	35945	69
	CNY690	Banco Chinchorro	35129	15
	CNY694	Yucatan	35132	17
Salinispora pacifica	CNH732	Red Sea	37235	22
	CNQ768	Guam	10973	217
	CNR114	Guam	9685	83
	CNR510	Guam	37241	15
	CNR894	Palau	9701	168
	CNR909	Palau	34968	8
	CNR942	Palau	11236	74
	CNS055	Palau	11237	75
	CNS103	Palau	9618	6
	CNS143	Palau	24476	112
	CNS237	Palau	14784	173
	CNS801	Fiji	34966	53
	CNS860	Fiji	11238	8
	CNS860 (DSM 45544)	Fiji	11030 (DSM45543)	115
	CNS960 (DSM 45544)	Fiji	10584 (DSM45544)	162
	CNS996	Fiji	10975	184
	CNT001	Fiji	9691	110
	CNT003	Fiji	9615	30
	CNT029	Fiji	9694	171
	CNT045	Fiji	10976	188
	CNT084	Fiji	9714	8
	CNT124	Fiji	10977	102
	CNT131	Fiji	9712	135
	CNT133 (DSM 45546)	Fiji	10313 (DSM45546)	71
	CNT136 (DSM 45547)	Fiji	10287 (DSM45547)	193
	CNT146 (DSM 45548)	Fiji	10602 (DSM45548)	178
	CNT150 (DSM 45549)	Fiji	10678 (DSM45549)	8
	CNT403	Fiji	34964	17
	CNT569	Fiji	9613	191
	CNT584	Fiji	10978	10
	CNT603	Fiji	9692	141
	CNT609	Fiji	10979	134
	CNT796	Hawaii	9689	61
	CNT851	Hawaii	10980	98
	CNT854	Hawaii	9676	9
	CNT855	Hawaii	9617	10
	CNY302	Sea of Cortez	15468	134
	CNY239	Fiji	14461	122
	CNY330	Sea of Cortez	11508	136
	CNY331	Sea of Cortez	11509	85
	CNY363	Sea of Cortez	35581	10
	CNY498	Puerto Vallarta	35667	18
	CNY646	Red Sea/Sponge	35578	28
	CNY666	Madeiras	35579	9
	CNY673	Madeiras	35580	7
	CNY703	Madeiras	35561	6
Streptomyces sp.	CNB091	Florida Keys	11141	262
	CNH099	San Diego	9717	174
	CNH189	San Diego	9655	405
	CNH287	San Diego	14288	182
	CNP092	Sea of Cortez	15294	196
	CNQ149	San Diego	not published	218
	CNQ329	San Diego	15683	451
	CNQ525	San Diego	35130	304
	CNQ766	Guam	10983	200
	CNQ865	Guam	14300	390
	CNR698	Bahamas	9744	302
	CNS335	Fiji	10984	327
	CNS606	Fiji	15295	286
	CNS615	Fiji	11142	309
	CNS654	Fiji	35124	336
	CNT302	Fiji	11025	181
	CNT318	Fiji	14387	286
	CNT360	Fiji	14389	254
	CNT371	Fiji	9716	319
	CNT372	Fiji	10985	203
	CNX435	Channel Islands	34950	291
	CNY228	Fiji	11019	265
	CNY243	Fiji	11239	83
	TAA040	Channel Islands	14404	146
	TAA204	Channel Islands	14405	204
	TAA486	Channel Islands	14866	222

Table S2: Overview of the compounds detected in this study

compound name	molecular family	gene cluster	analogues	molecular formula	m/z measured	m/z theor	adduct	GNPS score	GNPS Library ID	genus	species	strains detected	solvent	media	solid/liquid
actinomycin D	actinomycins	acm	3	$C_{62}H_{86}N_{12}O_{16}$	1255.64	1255.64	$[M+H]^+$	0.87	CCMSLIB00000068673	*Streptomyces*		CNS654	EA,BU,ME	MS,RS	solid
ammosamide B	ammosamides	amm	1(trimer+trimer)	$C_{17}H_{13}N_5O_3Cl$	896.14	896.14	$[3M-3Aq]^+$	0.87	CCMSLIB00000223962	*Streptomyces*		CNS613	EA,ME	A1,RS	solid
antimycin A1	antimycins	ant	9	$C_{28}H_{40}N_2O_9$	549.28	549.282	$[M+H]^+$	0.76	CCMSLIB00000079705	*Streptomyces*		CNQ329,CNY228	EA,BU,ME	R5	solid
arenicolide A	arenicolides	PKS28	7	$C_{28}H_{42}O_{11}$	827.49	827.492	$[M+Na]^+$	0.91	CCMSLIB00000001478	*Salinispora*	*arenicola,pacifica*	CNQ748,CNT138	EA	A1	liquid
areninycin A	areninycins	arn	2	$C_{31}H_{43}NO_{10}$	668.20	668.197	$[M+H]^+$	0.8	CCMSLIB00000081200	*Salinispora*	*arenicola*	CNB527	EA,BU	A1	solid
areninycin B	areninycins	arn	2	$C_{31}H_{43}NO_{11}$	809.31	809.312	$[M+H]^+$	0.83	CCMSLIB00000081201	*Salinispora*	*arenicola*	CNB527	EA,BU,ME	A1	solid
atteramide B	atteramides		6	$C_{28}H_{45}N_5O_5$	497.36	495.278	$[M+H]^+$	0.82	CCMSLIB00000072256	*Streptomyces*		CNQ329,CNS654	EA,BU,ME	R5	solid
cyclomarin A	cyclomarins	cym	6	$C_{46}H_{63}N_7O_{11}$	1025.61	1025.608	$[M-H_2O+H]^+$	0.74	CCMSLIB00000001554	*Salinispora*	*arenicola*	CNS205	EA,BU,ME	A1	solid/liquid
cyclomarin D	cyclomarins	cym	6	$C_{46}H_{63}N_7O_{10}$	1035.59	1035.59	$[M+Na]^+$	0.92	CCMSLIB00000001618	*Salinispora*	*arenicola*	CNS205	EA	A1	liquid
acyldeferoxamine-C13 (pramiceferrisoxamine)	deferoxamines	dez	>50	$C_{37}H_{71}N_7O_9$	729.55	729.548	$[M+H]^+$	0.66	CCMSLIB00000072053	*Salinispora*	*arenicola*	CNS991,CNY231	BU	A1	solid
acyldeferoxamine-C15	deferoxamines	dez	>50	$C_{39}H_{75}N_7O_9$	757.58	757.579	$[M+H]^+$	0.66	CCMSLIB00000072056	*Salinispora*	*arenicola*	CNS991,CNY231	BU	A1	solid
deferoxamine B	deferoxamines	dez	>50	$C_{25}H_{61}N_7O_8$	561.36	561.36	$[M+H]^+$	0.85	CCMSLIB00000072336	*Salinispora, Streptomyces*	*arenicola, tropica*	CNB440,CNH643,CNH713,CNS325,CNS673, CNS744,CNS820,CNS991,CNY578,CNT302,CNY236,CNT372,CNR688,CNS615,CNQ329,CNH189,CNQ766,CNP092,CNQ525	EA,BU,ME	A1,MS,RS	solid
deferoxamine E	deferoxamines	dez	>50	$C_{27}H_{63}N_7O_{10}$	601.36	601.36,623.33	$[M+H]^+$ $[M+Na]^+$	0.96	CCMSLIB00000001621, CCMSLIB00000001622	*Salinispora, Streptomyces*	*arenicola,pacifica*	CNH664,CNS673,CNS991,CNS996,CNT798,CNT869,CNY678,CNT302,TAA204,CNY228,TAA48 6,CNT318,CNR698,CNS615,CNQ329,CNH189, CNT360,CNS654	EA,BU,ME	A1,MS,RS	solid
desmethylenylnocardamine	deferoxamines	dez	>50	$C_{26}H_{59}N_7O_8$	586.5	587.339	$[M+H]^+$	0.72	CCMSLIB00000079904	*Salinispora, Streptomyces*	*arenicola*	CNH664,CNS673,CNS991,TAA204,CNY228,CNT372,TAA486,CNT318,CNQ329,CNH189,CNT360,CNS654	EA,BU,ME	A1,MS,RS	solid
enterocin	enterocins	enc	3	$C_{16}H_{12}O_{12}$	445.11	445.113	$[M+H]^+$	0.77	CCMSLIB00000001445	*Streptomyces*		CNT318	EA,BU,ME	A1,RS	solid
enterocin-Bdesoxy	enterocins	enc	3	$C_{16}H_{12}O_8$	429.12	429.118	$[M+H]^+$	0.86	CCMSLIB00000001844	*Streptomyces*		CNT318	EA,BU,ME	ME,RS	solid
tomaviticin A	tomaviticins	tom	9	$C_{58}H_{74}N_2O_{20}$	683.27	683.268	$[M+H]^{2+}$	0.92	CCMSLIB00000081198	*Salinispora*	*pacifica, tropica*	CNH898,CNT148,CNT350,CNT569	EA	A1	solid
tomaviticin C	tomaviticins	tom	9	$C_{58}H_{74}N_2O_{20}$	670.27	670.274	$[M+H]^{2+}$	0.94	CCMSLIB00000081196	*Salinispora*	*pacifica, tropica*	CNB440,CNB476,CNB536,CNH898,CNS095,CNT148,CNT350,CNT569	EA,BU,ME	A1	solid/liquid
tomaviticin D	tomaviticins	tom	9	$C_{58}H_{74}N_2O_{19}$	677.41	677.281	$[M+H]^{2+}$	0.94	CCMSLIB00000081197	*Salinispora*	*pacifica, tropica*	CNB476,CNB536,CNH898,CNT148,CNT569	EA,BU,ME	A1	solid
netropcin A	netropcins	ntn	2	$C_{68}H_{94}N_{16}O_{13}Br_2$	1171.42	1171.422	$[M+H]^+$	0.86	CCMSLIB00000223995	*Salinispora*	*arenicola*	CNT005	EA	A1	liquid
rifamycin S	rifamycins	rif	28	$C_{37}H_{45}NO_{12}$	696.28	696.302	$[M+H]^+$	0.76	CCMSLIB00000001633	*Salinispora*	*arenicola*	CNT005,CNX814	EA	A1	solid/liquid
rifamycin W	rifamycins	rif	28	$C_{35}H_{43}NO_{11}$	656.31	656.306	$[M+H]^+$	0.73	CCMSLIB00000001635	*Salinispora, Streptomyces*	*arenicola*	CNH646,CNH941,CNS205,CNS991,CNT005,CNT799,CNT850,CNX891,CNY234,CNY260,CNQ149,CNP193,CNT005,CNT798,CNT849,CNX814,CNQ149	EA,BU,ME	A1, MS	solid/liquid
rosamicin	rosamicins	spr	15	$C_{31}H_{51}NO_9$	582.36	582.364	$[M+H]^+$	0.85	CCMSLIB00000223877	*Salinispora, Streptomyces*	*pacifica*	CNS237,CNT609	BU	A1	solid
21-hydroxyrosamicin	rosamicins	spr	15	$C_{31}H_{51}NO_{10}$	598.35	598.359	$[M+H]^+$	0.93	CCMSLIB00000513126	*Salinispora, Streptomyces*	*pacifica*	CNS237,CNT609,CNT360	EA,BU,ME	A1	solid
19-dihydro-14-hydroxyrosamicin	rosamicins	spr	15	$C_{31}H_{51}NO_{10}$	600.43	600.374	$[M+H]^+$	0.86	CCMSLIB00000513128	*Salinispora, Streptomyces*	*pacifica*	CNS237,CNT609,CNT360	EA,BU,ME	A1	solid
19-dihydro-21-hydroxyrosamicin	rosamicins	spr	15	$C_{31}H_{51}NO_{10}$	600.38	600.374	$[M+H]^+$	0.84	CCMSLIB00000513129	*Salinispora, Streptomyces*	*pacifica*	CNS237,CNT609,CNT360	EA,BU,ME	A1	solid
14-hydroxyrosamicin	rosamicins	spr	15	$C_{31}H_{51}NO_{10}$	598.36	598.359	$[M+H]^+$	0.92	CCMSLIB00000513127	*Salinispora, Streptomyces*	*pacifica*	CNS237,CNT584,CNT133,CNT609,CNY646,CNT369	EA,BU,ME	A1	solid
salinamide A	salinamides	sln	7	$C_{54}H_{81}N_7O_{14}$	1020.49	1020.493	$[M+H]^+$	0.81	CCMSLIB00000223953	*Streptomyces*		CNB591	EA,BU,ME	A1,MS	solid
salinamide B	salinamides	sln	7	$C_{54}H_{80}N_7O_{13}Cl$	1056.47	1056.47	$[M+H]^+$	0.67	CCMSLIB00000223954	*Streptomyces*		CNB591	EA,BU,ME	A1	solid
salinamide E	salinamides	sln	7	$C_{49}H_{74}N_6O_{12}$	955.45	955.45	$[M+H]^+$	0.79	CCMSLIB00000223958	*Streptomyces*		CNH287	BU	A1	solid
desmethylsalinamide E	salinamides	sln	7	$C_{48}H_{72}N_6O_{12}$	841.44	841.436	$[M+H]^+$	0.76	CCMSLIB00000223960	*Streptomyces*		CNH287,CNB591	EA,BU	A1,MS	solid
staurosporine	staurosporines	sta	11	$C_{28}H_{26}N_4O_3$	467.20	467.208	$[M+H]^+$	0.97	CCMSLIB00000001615	*Salinispora*	*arenicola, pacifica,*	CNB458,CNB527,CNH643,CNH646,CNH713,CNH718,CNH877,CNH905,CNH941,CNH962,CNH963,CNH984,CNH996,CNP105,CNP193,CNQ748, CNQ884,CNR167,CNR425,CNR921,CNS051,CNS095,CNS343,CNS396,CNS399,CNS325,CNS342,CNS673,CNS744,CNS820,CNS848,CNS863 ,CNS991,CNS996,CNT005,CNT029,CNT138,CNT403,CNT798,CNT799,CNT800,CNT849,CNT850,CNX682,CNX814,CNX891,CNY011,CNY230,CNY231,CNY234,CNY237,CNY244,CNY256,CNY260,CNY280,CNY281,CNY282,CNY486,CNY679,CNY694 ,CNY694	EA,BU,ME	A1	solid/liquid
staurosporine-7-OH	staurosporines	sta	11	$C_{28}H_{26}N_4O_4$	483.20	483.203	$[M+H]^+$	0.88	CCMSLIB00000001672	*Salinispora, Streptomyces*	*arenicola, pacifica,*	CNB458,CNB527,CNH643,CNH646,CNH713,CNH718,CNH877,CNH905,CNH941,CNH962,CNH963,CNH984,CNH996,CNP105,CNP193,CNQ748, CNQ884,CNR167,CNR425,CNR921,CNS051,CNS095,CNS343,CNS396,CNS399,CNS325,CNS342,CNS673,CNS744,CNS820,CNS848,CNS863 ,CNS991,CNS996,CNT005,CNT124,CNT138,CNT584,CNT798,CNT799,CNT849,CNT850,CNT857,CNX682,CNX814,CNX891,CNY011,CNY202,CNY230,CNY231,CNY234,CNY237,CNY244,CNY256,CNY260,CNY280,CNY282,CNY486,CNY679,CNY694 ,CNQ149	EA,BU,ME	A1,MS	solid/liquid

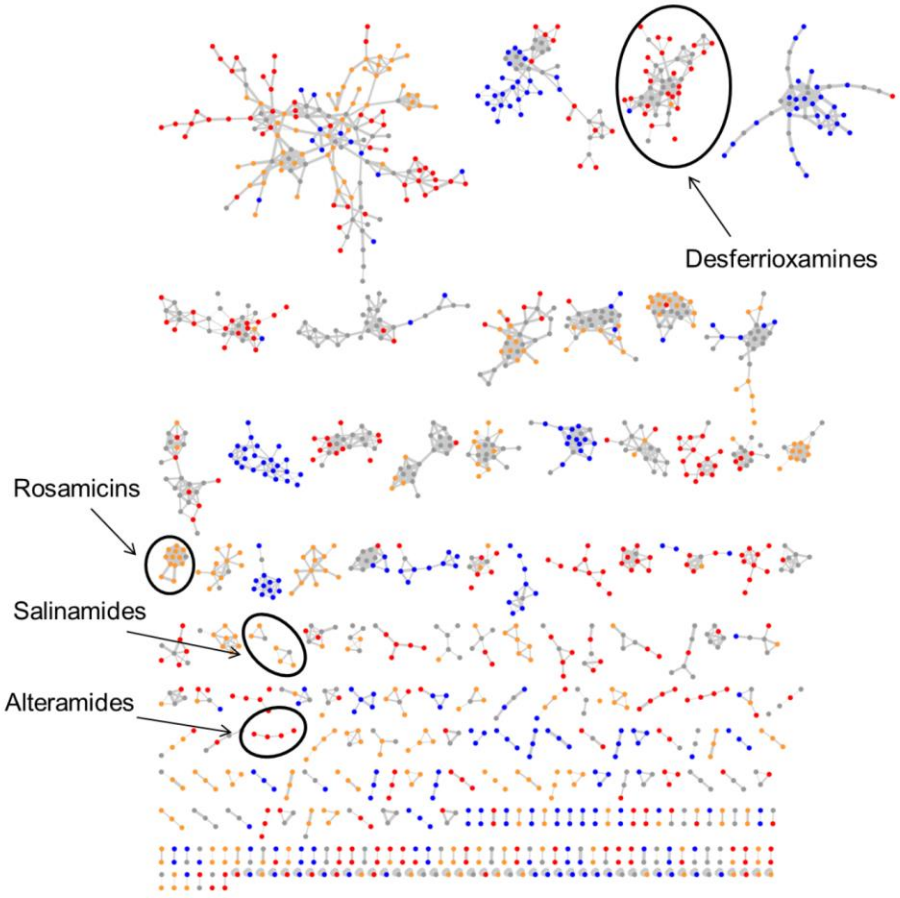


Figure S1: Molecular Network of All *Streptomyces* Extracts Color Coded By Culture Medium. Blue nodes represent ions that were only detected in MS extracts, orange nodes those only found in A1 and the red nodes represent ions only detected from R5 extracts. Only the clusters containing at least two nodes are shown. See also Figure 2E.

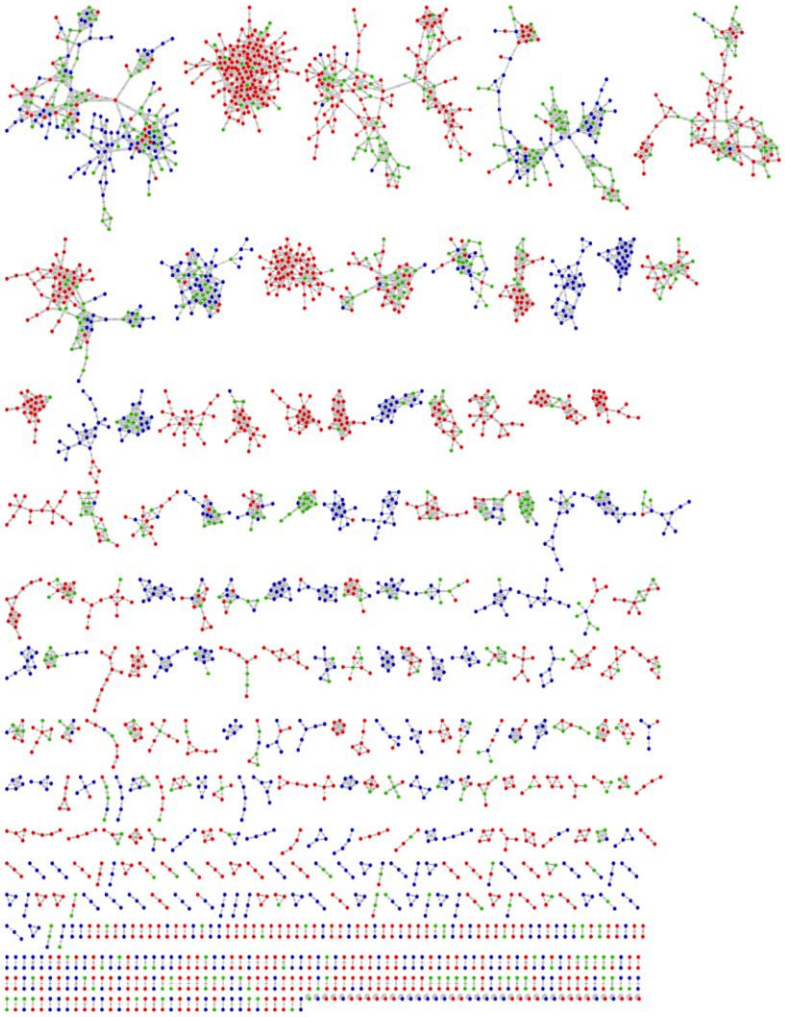


Figure S2: Molecular Network of All Acquired Extracts Color Coded By Genus. Blue nodes represent ions that were only detected from *Salinispora* extracts, red nodes those only found from *Streptomyces* extracts and the green nodes represent ions that are shared by the genera. Only the clusters containing at least two nodes are shown. See also Figure 2A.

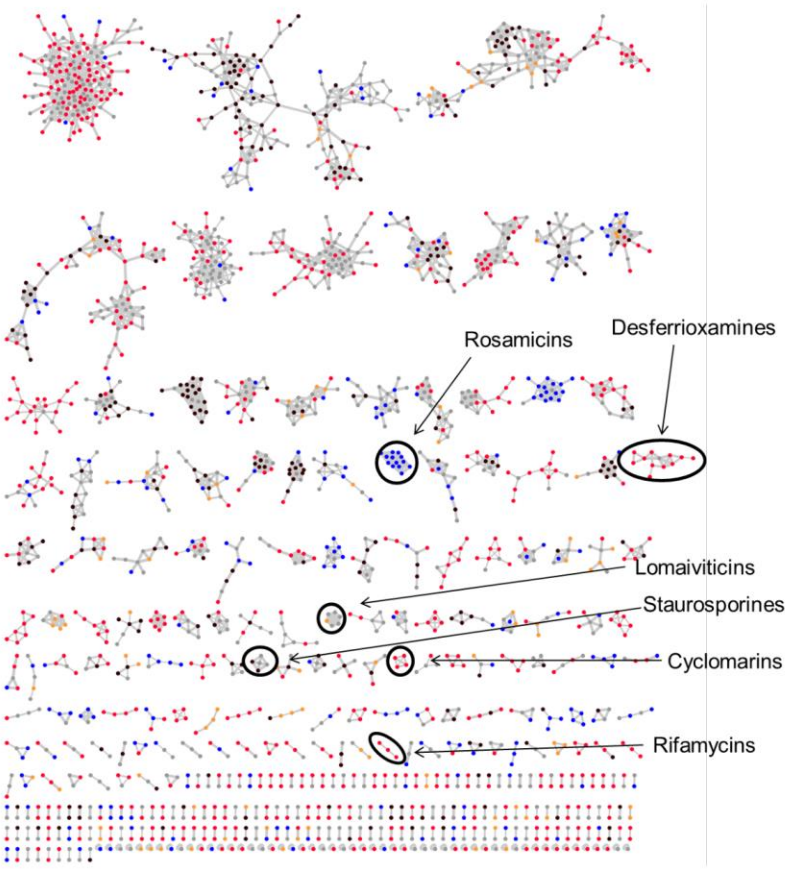


Figure S3: Molecular Network of All *Salinispora* Extracts Color Coded By Species. Blue nodes represent ions that were only detected from *S. pacifica* extracts, orange nodes those only found from *S. tropica* extracts and the red nodes represent ions only detected from *S. arenicola* extracts. Black nodes represent ions that found in all three species. Only the clusters containing at least two nodes are shown. See also Figure 2B.

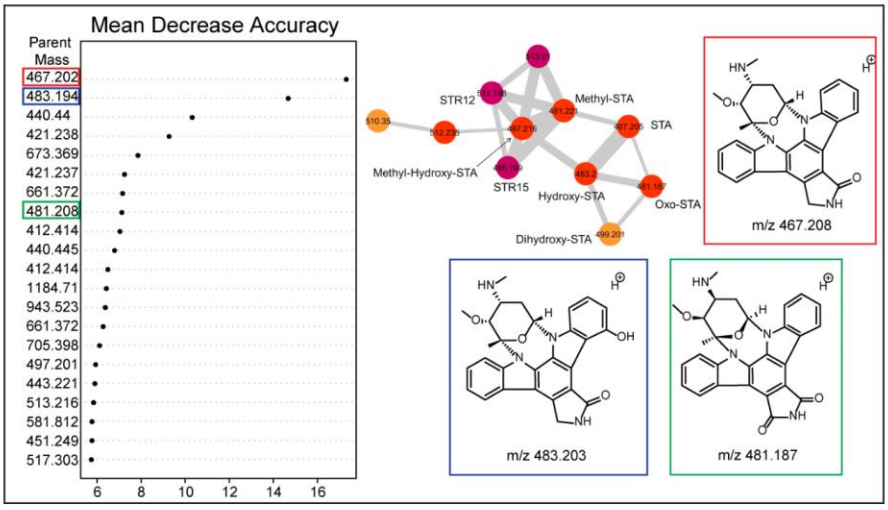


Figure S4: Illustration of the major drivers of chemical speciation within the *Salinispora* data. Depicted are the structures of three major examples of this analysis, staurosporine and its hydroxylated and oxidized analogs.

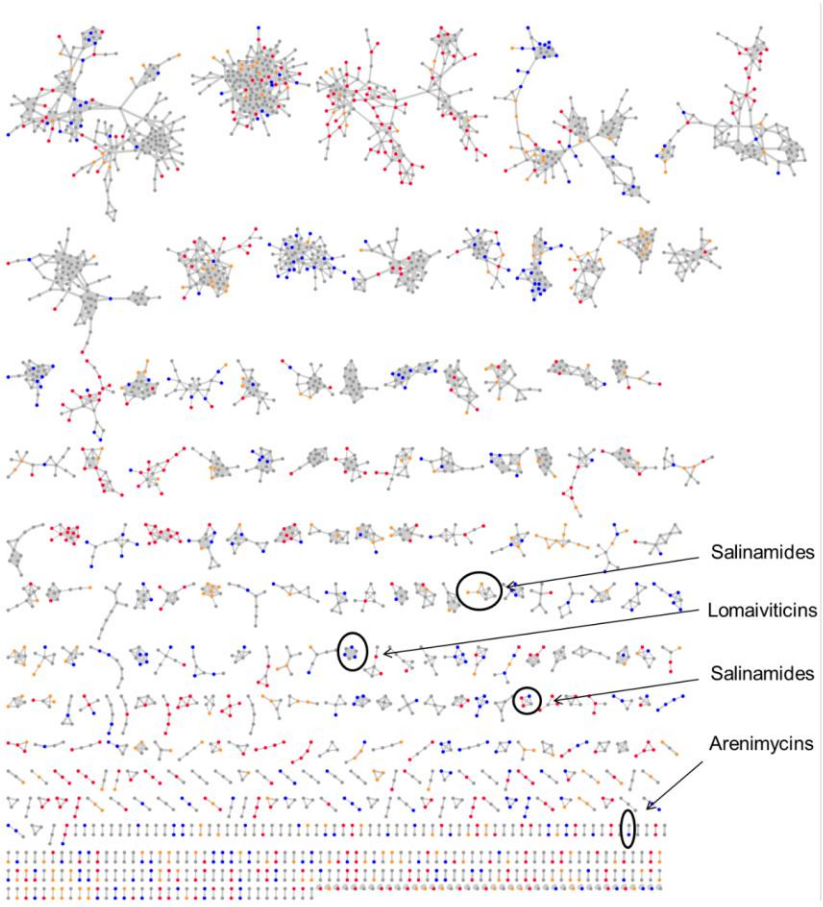


Figure S5: Molecular Network of All Generated Extracts Color Coded by Extraction Solvent. Blue nodes represent ions that were only detected in the ethyl acetate extracts, orange nodes those only found in butanol and the red nodes represent ions only detected from methanol extracts. Only the clusters containing at least two nodes are shown. See also Figure 3 A and B.

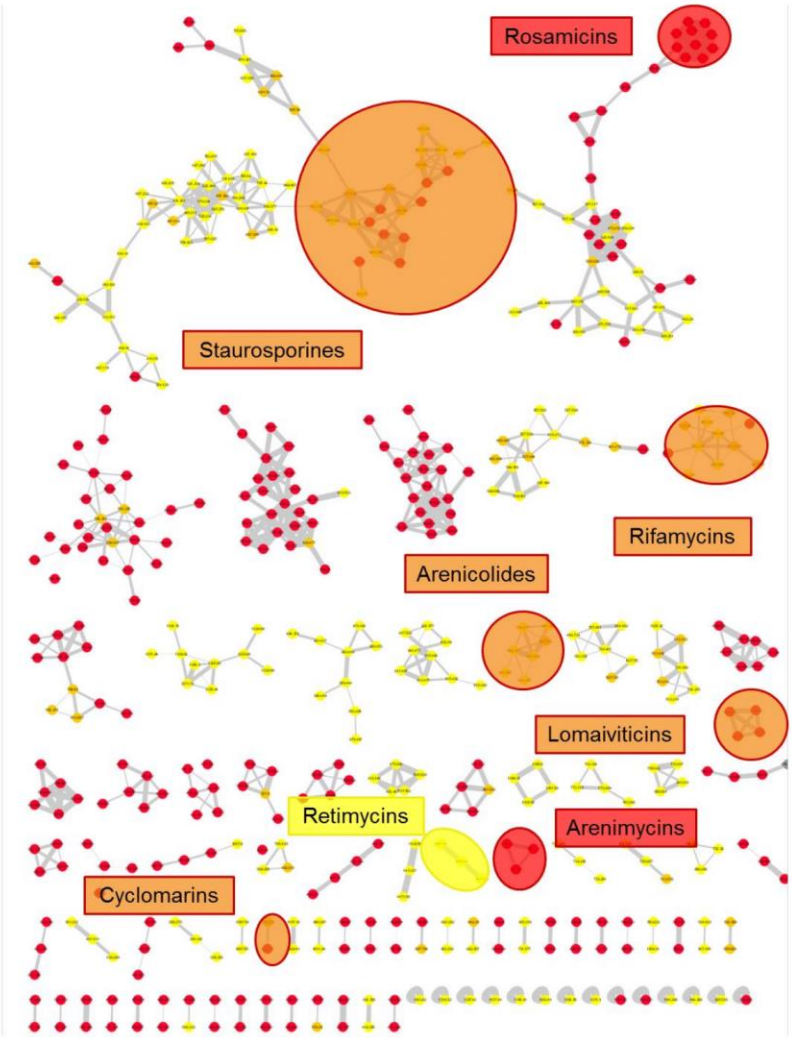


Figure S6: Molecular Network of *Salinispora* Extracts from Different Media Phases. Extracts from previous work (Duncan, Crüsemann et al. 2015) (30 *Salinispora* strains, liquid A1, solvent: ethyl acetate) and ethyl acetate extracts of the same 30 strains, grown on solid A1, extracted with ethyl acetate, were networked together. Yellow nodes represent ions extracted from liquid, red nodes from solid medium and orange nodes represent ions extracted from both approaches.

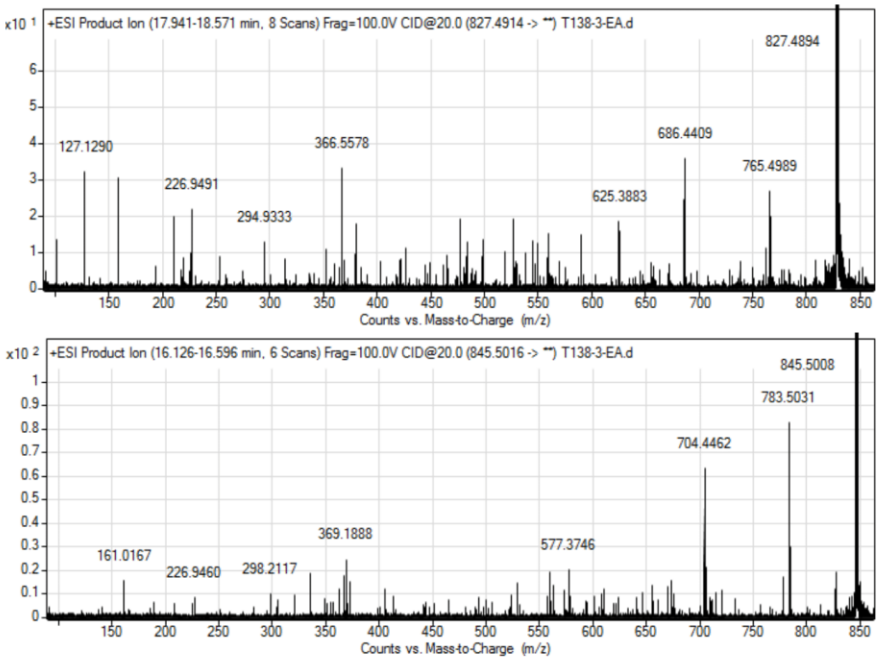


Figure S7: Comparison of MS/MS spectra of arenicolide A (upper spectrum, *m/z:* 827.489,) and a novel, formally hydrated arenicolide analogue (lower spectrum, m/z: 845.501) from the ethyl acetate extract of *Salinispora pacifica* CNT-138.

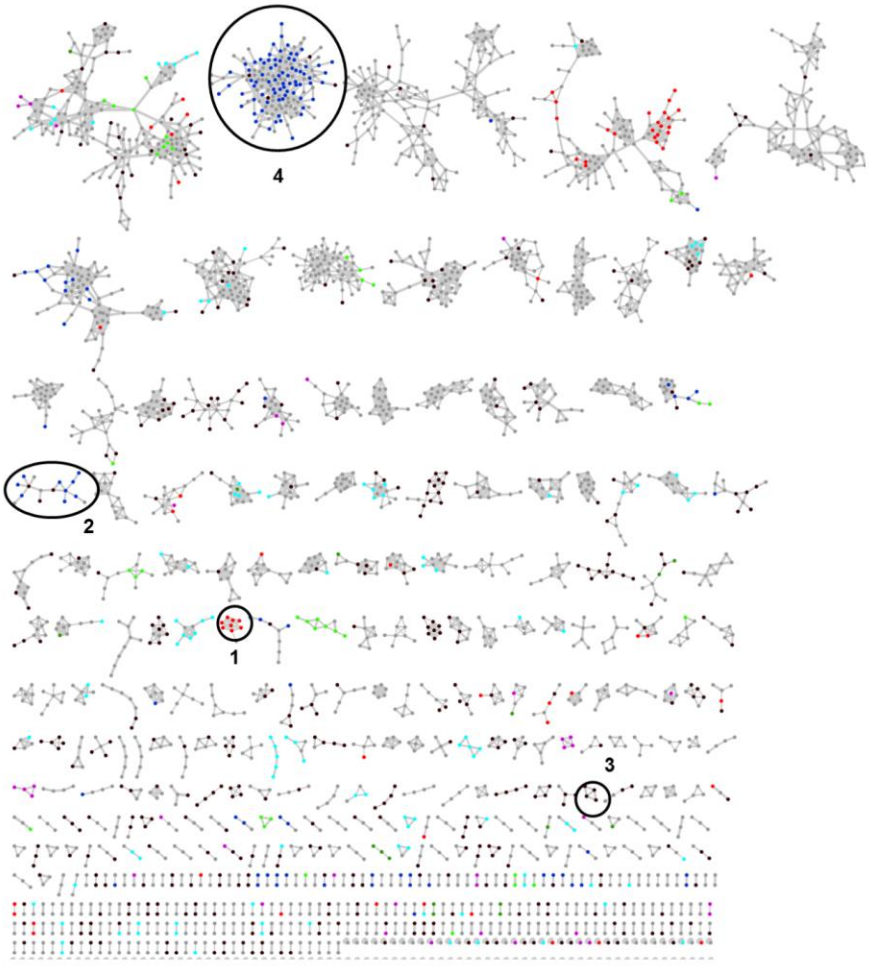


Figure S8: Molecular Network of All Acquired Extracts Color Coded By Strain Location. Ions that were only extracted from one location are color-coded as follows Bahamas: yellow, Fiji: black, Guam: red, Hawaii: blue, Palau: purple, Palmyra: light green, Red Sea: dark green, San Diego: light blue. Highlighted are clusters 1, 2, 3 and 4 in the order mentioned in the main text.

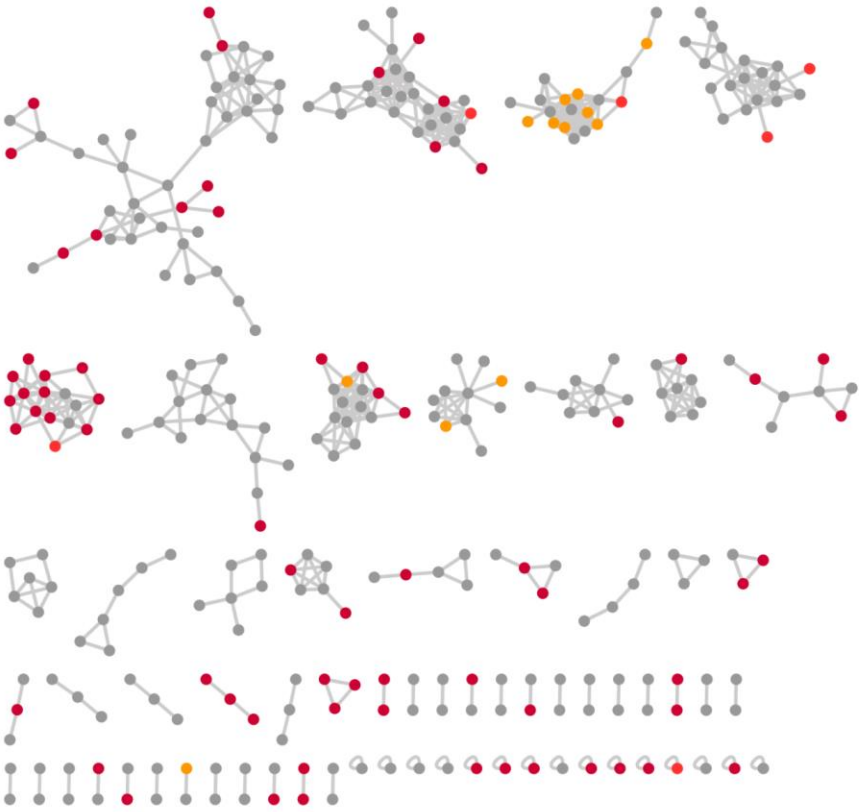


Figure S9: Molecular Network of *Salinispora arenicola* CNH877 strains grown in four different liquid media and extracted at three different time points. Ions that were only extracted after one time point are color-coded as follows: 14 days orange, 21 days red, 28 days dark red (see Figure 4).

Table S3: Information about detected staurosporine analogues in the ISP2 time course network (see Figure 4) (Barrabee, Horan et al. 1997, Park, Abdel-Azeem et al. 2013).

m/z (Da) measured	m/z (Da) theor	error (ppm)	exact mass	molecular formula	spectral counts after 14 d	21 d	28 d	name	characterized
467.205	467.208	-6.4	466.2	$C_{28}H_{26}N_4O_3$	3	4	5	staurosporine	
481.187	481.188	-2.1	480.18	$C_{28}H_{24}N_4O_4$	1	1	1	oxo-STA	RK-1409 (7-oxo-STA)
481.221	481.224	-6.2	480.216	$C_{29}H_{28}N_4O_3$	2	6	5	Methyl-STA	N-methyl-STA
483.2	483.203	-6.2	482.195	$C_{28}H_{26}N_4O_4$	6	7	9	hydroxy-STA	7,11 and 15-hydroxy-STA
495.199	495.203	-7.7	494.195	$C_{29}H_{26}N_4O_4$	0	0	3		*Saccharothrix* STR15
497.216	497.218	-4.0	496.211	$C_{29}H_{28}N_4O_4$	6	7	12	methyl-hydroxy-STA	TAN-999 (10-methoxy-STA),4-N-methyl-5-hydroxy-STA
499.201	499.198	6.0	498.19	$C_{28}H_{26}N_4O_5$	0	3	10	dihydroxy-STA	3,11-dihydroxy-STA
511.196	511.198	-3.9	510.19	$C_{29}H_{26}N_4O_5$	0	0	5		*Saccharothrix* STR12

References

Barrabee, E. B.; Horan, A. C.; Gentile, F. A.; Patel, M. G. **1997**, Indolocarbazoles from *Saccharothrix aerocolonigenes copiosa* subsp. nov SCC 1951 ATCC 53856. USPTO. United States, Schering Corporation.

Duncan, K. R.; Crüsemann, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, B.S.; Dorrestein, P.C.; Jensen, P.R. *Chem. Biol.* **2015,** 22, 460-471.

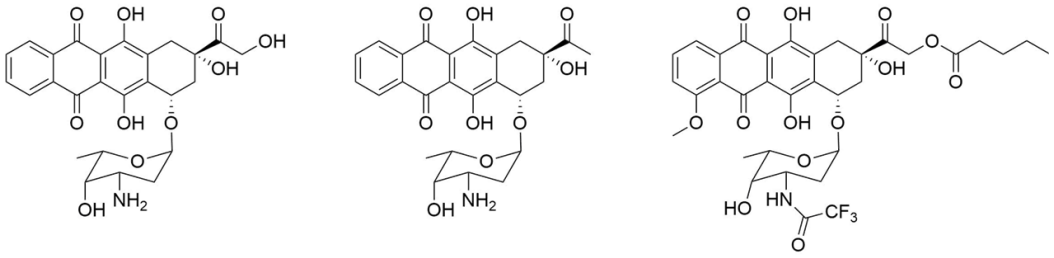
Park, B. S.; Abdel-Azeem, A. Z.; Al-Sanea, M. M.; Yoo, K. H.; Tae, J. S.; Lee, S. H. *Curr. Med. Chem.* **2013,** *20*, 3872-3902.

Chapter 2 section 2.3, in full, is a reprint of materials as it appears in “Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols” in *Journal of Natural Products*, 2016, Max Crusman, Ellis C. O’neill, Charles B. Larson, Alexey V. Melnick, Dimitrios J. Floros, Ricardo R. da Silva, Paul R. Jensen, Pieter C. Dorrestein, Bradley S. Moore. The dissertation author was a secondary author of this paper. M.C., E.C.O., and C.B.L. designed research; M.C., E.C.O., and C.B.L. performed extractions; M.C., E.C.O., A.V.M., D.J.F., R.R.S., and C.B.L. performed mass spectrometry experiments; M.C. and E.C.O. generated molecular networks; C.B.L., M.C., and E.C.O. and analyzed data; P.R.J. provided bacterial strains, and all authors contributed to writing the manuscript.

Chapter 3: PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete

3.1 Introduction to Chapter 3

Cancer is a major worldwide health concern, with an economic cost estimated to be over $1 trillion annually[1]. It is the leading cause of death worldwide, claiming an estimated 9.6 million lives in 2018 and accounting for approximately one in six deaths globally[2]. Since the discovery of daunorubicin in 1950s and its clinical trials in the following decade[3], anthracyclines have been valuable tools for studying and treating cancer. Still being used in the clinic under the trade name daunomycin, along with the structurally similar natural product doxorubicin,[4] these anthracyclines are on the WHO's list of essential medicines.[5] More recently, the anthracyclines epirubicin[6], idarubicin[7], and valrubicin[8] have also been FDA approved for various cancers[9].



Doxorubicin (Adriamycin® or Rubex®) | Daunorubicin (Cerubidine®) | Valrubicin (Valstar®)

Figure 3.1: A selection of anthracyclines approved for treatment of cancer.

Anthracycline natural products such as daunomycin and doxorubicin are produced by type-II polyketide synthase (PKS) biosynthetic gene clusters (BGCs), which condense malonyl-CoA into the polyketide chains forming the backbone of the four-ringed aromatic structure. Several tailoring enzymes subsequently cyclize, hydroxylate, glycosylate and reduce this core molecule.[10] These modifications can have a profound effect on the bioactivity of these molecules.[11] Of particular importance are the sugars attached to the core of the molecule. Changes to these moieties can dramatically affect their activity, for example, abolishing activity, enhancing efficacy in drug resistant cancers, in addition to changes to cardiotoxicity and other side effects.[12,13]

In addition to contributing to the bioactivity of anthracyclines, presence of sugar moieties in the structure can facilitate detection of these molecules using LC-MS/MS (MS^2) analysis and using glycogenomics to link these molecules to their BGC. Glycogenomics is a genome mining approach in which characteristic mass fragments of glycosyl moieties of sugar-containing natural products are used to predict required for the biosynthesis glycosylation genes that become a “search hook” for the corresponding BGC.[14,15]

Bioinformatic analysis of the draft genome of marine actinomycete strain CNT-302 revealed a BGC that contained three glycosyltransferase-encoding genes, type II PKS genes and tailoring genes for redox chemistry. Such gene combination suggested that the product of the cluster is a glycosylated anthracycline. Considering potent anticancer properties of this class of molecules, it became a target for PCR-independent direct capture method and subsequent heterologous expression. Additionally, this cluster presented an opportunity to use glycogenomics to link the captured genes to the gene cluster product(s).

3.2 References for Chapter 3 Introduction

1. Stewart, B. W., Wild, C., International Agency for Research on Cancer & World Health Organization. *World cancer report 2014*.

2. WHO | Cancer. *WHO* (2019).

3. DIMARCO A, GAETANI M, OREZZI P, SCARPINATO BM, SILVESTRINI R, SOLDATI M, DASDIA T, VALENTINI L. 'DAUNOMYCIN', A NEW ANTIBIOTIC OF THE RHODOMYCIN GROUP. *Nature* **201**, 706–7 (1964).

4. Weiss, R. B. The anthracyclines: will we ever find a better doxorubicin? *Semin. Oncol.* **19**, 670–86 (1992).

5. WHO | Essential medicines. *WHO* (2016).

6. Bonfante, V., Bonadonna, G., Villani, F. & Martini, A. Preliminary clinical experience with 4-epidoxorubicin in advanced human neoplasia. *Recent Results Cancer Res.* **74**, 192–9 (1980).

7. Kaplan, S., Martini, A., Varini, M., Togni, P. & Cavalli, F. Phase I trial of 4-demethoxydaunorubicin with single i.v. doses. *Eur. J. Cancer Clin. Oncol.* **18**, 1303–6 (1982).

8. Onrust, S. V. & Lamb, H. M. Valrubicin. *Drugs Aging* **15**, 69–75 (1999).

9. McGowan JV, Chung R, Maulik A, Piotrowska I, Walker JM, Yellon DM. Anthracycline Chemotherapy and Cardiotoxicity. *Cardiovasc. drugs Ther.* **31**, 63–75 (2017).

10. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162–190 (2007).

11. Shaul P, Frenkel M, Goldstein EB, Mittelman L, Grunwald A, Ebenstein Y, Tsarfaty I, Fridman M. The structure of anthracycline derivatives determines their subcellular localization and cytotoxic activity. *ACS Med. Chem. Lett.* **4**, 323–8 (2013).

12. Weymouth-Wilson, A. C. The role of carbohydrates in biologically active natural products. *Nat. Prod. Rep.* **14**, 99 (1997).

13. Booser, D. J. & Hortobagyi, G. N. Anthracycline antibiotics in cancer therapy. Focus on drug resistance. *Drugs* **47**, 223–58 (1994).

14. Kersten RD, Ziemert N, Gonzalez DJ, Duggan BM, Nizet V, Dorrestein PC, Moore BS. Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl. Acad. Sci.* **110**, E4407–E4416 (2013).

15. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B.. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).

3.3 Reprint of "PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete"

JOURNAL OF NATURAL PRODUCTS

Note

pubs.acs.org/jnp

PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete

Charles B. Larson,[†,‡] Max Crüsemann,[†,⊥] and Bradley S. Moore*[,†,§]

[†]Center for Marine Biotechnology & Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92037, United States

[‡]Biomedical Sciences Graduate Program, University of California at San Diego, La Jolla, California 92093, United States

[§]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California 92093, United States

Supporting Information

ABSTRACT: The transformation-associated recombination cloning methodology facilitates the genomic capture and heterologous expression of natural product biosynthetic gene clusters (BGCs). We have streamlined this procedure by introduction of synthetic DNA gene blocks for the efficient capture of BGCs. We show the successful capture and expression of the aromatic polyketide antitumor agent cosmomycin from streptomycete bacteria and the discovery of new cosmomycin analogues by mass spectral molecular networking.

With the advent of rapid and inexpensive genome sequencing methods, the process of natural product discovery is advancing to incorporate biosynthetic logic.[1] Traditional discovery efforts have focused on the collection, culturing, and subsequent analysis of organismal extracts to detect pharmacological activities harbored in bacteria, plants, or animals in a "grind and find" approach.[2] However, this manner of investigation can often overlook compounds produced in minuscule amounts or in unique environmental conditions that are difficult to replicate in the laboratory. Some of the most well-studied bacteria have been shown to produce only a subset of natural product chemicals even though their genomes frequently encode the biosynthetic machinery to synthesize many additional compounds of often unknown structure and function.[3] Recent advances in synthetic biology have allowed investigators to avoid this traditional discovery approach of culturing and extracting wild-type strains by directly manipulating the genetic sequence that encodes components of biosynthetic pathways.[4] Using this genetic information, entire pathways can be targeted within the native producer or removed in their entirety and expressed in a heterologous host.[5] Utilizing these techniques, many pathways of interest have been elucidated that are intractable to traditional natural product discovery approaches.[6]

In order to clone large biosynthetic gene clusters (BGCs) that are regularly 40−60 kilobases in size, transformation-associated recombination (TAR) has emerged as a powerful method to selectively capture and incorporate genes encoding complete biosynthetic pathways into multihost plasmids. This method was originally described by Kouprina and Larionov for cloning selective genomic loci from human DNA[7] and later adapted by Brady to capture BGCs from environmental DNA (eDNA) libraries, including the BGC producing the MRSA-active antibiotic tetarimycin A.[8] Our laboratory developed TAR vectors and methods to capture and express an assortment of BGCs directly from gDNA from taxonomically diverse microbes.[6,9,10] In addition to direct cloning from libraries or genomes, TAR has been applied as a DNA assembly technique[11] for the creation of a fully synthetic genome and has facilitated the refactoring of BGCs to support natural product production.[12]

Despite its successes, challenges are often experienced during the experimentally involved TAR cloning process. Vector assembly in particular is a tedious, multistep process that relies heavily on PCR. Complications can arise during this assembly process, especially with high GC or repetitive sequences that are common among actinomycete bacteria known to harbor large numbers of BGCs. Herein we report a modified approach employing a fully synthetic "capture arm," thus eliminating the need for traditional PCR amplification during the vector assembly process. This new methodology allows for a significant decrease in the duration of the cloning process and opens the door for higher-throughput applications. To demonstrate the feasibility of this optimized procedure, we set out to capture and heterologously express the cosmomycin BGC from a marine streptomycete bacterium.

Received: December 5, 2016
Published: March 23, 2017

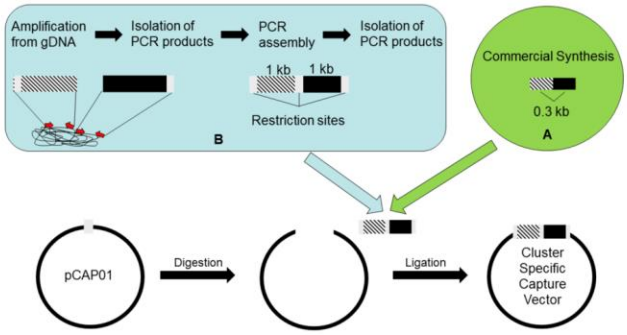


Figure 1. Capture vector assembly. (A) Cluster-specific capture arms, one kilobase in size, are amplified from genomic DNA and subsequently assembled using PCR. (B) Cluster-specific capture arms (360 bp) are synthesized commercially as a single 750 bp dsDNA gene block including restriction sites for insertion and linearization.

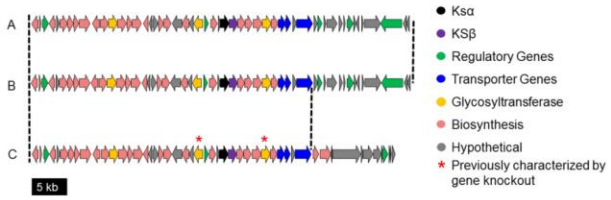


Figure 2. Gene map of biosynthetic gene clusters associated with the biosynthesis of cosmomycins C and D. The cosmomycin BGCs from (A) *Streptomyces* sp. CNS-615, (B) *Streptomyces* sp. CNT-302, and (C) *Streptomyces olindensis* are >90% similar on the amino acid sequence level. A gene-by-gene comparison can be found in Table S2. Homology of the *Streptomyces* sp. CNT-302 and the *S. olindensis* BGCs is restricted to the first 40 kb of the captured sequence (denoted by dotted lines), suggesting the last 14 kb of genomic material captured in this study is not necessary for cosmomycin production. Genes marked with a "*" have been interrogated by gene knockout experiments in previous studies.

We previously reported the utilization of TAR-based direct cloning and expression of BGCs with the taromycin series of cyclic lipopeptide antibiotics.[6] Briefly, our original protocol with the *Saccharomyces cerevisiae*/*Escherichia coli* shuttle-actinobacterial chromosome integrative capture vector pCAP01 required a pair of one kilobase (kb) regions, homologous to the two flanking regions of the BGC of interest, to be amplified from genomic DNA (Figure 1A). These fragments were then assembled, inserted into the pCAP01 capture vector, and linearized prior to use in the capture/transformation steps. The time-consuming process of the multiple PCR reactions and subsequent ligations necessary to obtain the assembled capture vector demanded development of a more efficient protocol.

We sought to streamline the original protocol by replacing all vector assembly steps by a simple digestion/ligation reaction. To do so, we designed a synthetic DNA fragment containing shortened capture arms and the appropriate restriction sites to facilitate both its insertion into pCAP01 and plasmid linearization (Figure 1B). We selected an orphan 54 kb BGC encoding a type II PKS from *Streptomyces* sp. CNT-302 (Figure 2) as a target pathway. The type II PKS shares 94% and 90% identity with the cosmomycin[13,14] KS and CLF subunits, respectively. In addition, the pathway contains 32 genes encoding putative tailoring enzymes and three genes encoding glycosyltransferases indicative of a highly glycosylated end product. Proposed functions and homologies for each gene product in the pathway are shown in Table S2. Initially investigated for its extensive glycosylation biochemistry involving three separate glycosyltransferases, we analyzed CNT-302 extracts by glycogenomic mass spectral analysis[15] to no avail. Although this BGC appeared nonfunctional in the native host under the growth conditions tested, we explored its heterologous expression using the modified TAR capturing method.

We employed capture arms of 360 base pairs (bps), about a third of the length previously used,[6] although capture arms as small as 60 bps have been shown to be effective.[16] The size of the capture arms was determined by the maximum size of commercial oligonucleotide synthesis at the time and included three restriction sites flanking the two homology capture arms for ligation and linearization. After insertion of the 750 bp synthetic double-stranded DNA into pCAP01 to form plasmid pCAP01-COM, we transformed yeast spheroplasts with digested *Streptomyces* sp. CNT-302 gDNA and linear pCAP01-COM on selective media. Yeast clones were screened by PCR, and subsequent analysis via restriction digestion confirmed the successful capture of the type II PKS BGC in plasmid pCAP01-COS in three of 200 transformants screened (Figures S1 and S2). After the isolation of pCAP01-COS, we confirmed the vector by restriction digest and transformed it into an *E. coli* ET12567 shuttle, which integrated the vector into the genome of *S. coelicolor* M512[17] by conjugation. To

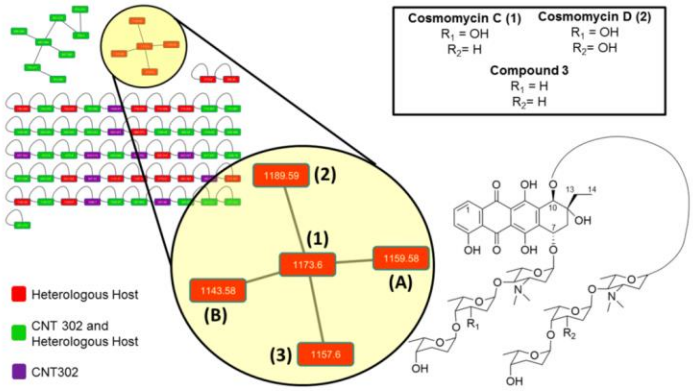


Figure 3. Cosmomycin cluster identified by molecular networking. Nodes matching the masses for cosmomycins C (**1**) and D (**2**) network with nodes matching values of cosmomycin derivatives lacking methyl groups and/or oxygen. These include deoxycosmomycin C (**3**), desmethylcosmomycin C (**A**), and desmethyl-deoxycosmomycin C (**B**).

confirm the stable integration of pCAP01-COS without unintended rearrangements, several regions of the integrated plasmid were PCR amplified and Sanger sequenced (Figures S3 and S4). In order to determine if glycosylated natural products were produced by this type II PKS BGC, we cultured the host *S. coelicolor* M512/pCAP01-COS and analyzed its products by mass spectral (MS) molecular networking.[18]

Spectral networks, originally developed for proteomics, have recently been adopted as a tool for general MS data analysis.[19] The foundation of a molecular network is the comparison of MS/MS spectra for molecular ions captured by mass spectrometry. Molecular families are then created based on the similarity of the analyzed spectra. Analysis of the network, which included extracts of M512 as a control, M512/pCAP01-COS, and CNT-302 immediately revealed a five-molecule family in *S. coelicolor* M512/pCAP01-COS absent in the "native" producer, including the previously described antitumor molecules cosmomycin C (**1**) (HRESIMS *m/z* 1173.5964, calcd *m/z* for $C_{60}H_{88}N_2O_{21}$ ($[M + H]^+$) 1173.5952), cosmomycin D (**2**) (HRESIMS *m/z* 1189.5889, calcd *m/z* for $C_{60}H_{88}N_2O_{22}$ ($[M + H]^+$), 1189.5901), and deoxycosmomycin C (**3**) (HRESIMS *m/z* 1157.6007, calcd *m/z* for $C_{60}H_{88}N_2O_{20}$ ($[M + H]^+$) 1157.6003) (Figure 3).[20] These three molecules were identified by mass fragmentation patterns and isotope distributions in comparison to those previously reported (Figures S5–S8).[21] Additionally, two new cosmomycin analogues were identified in *S. coelicolor* M512/pCAP01-COS, which appear to be missing methyl and/or hydroxy groups on the inner amino sugar rhodosamine or aglycone when compared to cosmomycin C (**1**). These include compound **A** (HRESIMS *m/z* 1159.5793 (calcd *m/z* for $C_{59}H_{86}N_2O_{21}$ ($[M + H]^+$) 1159.5796) and compound **B** (HRESIMS *m/z* 1143.5846, calcd *m/z* for $C_{59}H_{86}N_2O_{20}$ ($[M + H]^+$) 1143.5847), which have masses consistent with desmethylcosmomycin C and desmethyl-deoxycosmomycin C, respectively (Figures S9–S11).

The two major cosmomycin species **1** and **A** were further purified and analyzed by ^{1}H and HSQC NMR (Figures S12–S17). Because of the small amount of **B** produced in culture, further structural elucidation was not pursued. The chemical shifts observed for both **1** and **A** closely matched literature values for cosmomycin analogues A447 C and D, obelmycin C, and cosmomycin D (Table S1). While **1** could be unequivocally established as cosmomycin C, spectroscopic analysis of the closely related **A** suggested that it was a new analogue. The C-14 triplet methyl peak at 1.09 ppm in **1** is conspicuously absent in the 1D NMR and HSQC spectra of compound **A** (Figures S14–S16). Additionally, we observed a new peak in the HSQC spectrum of **A** at 1.47 ppm correlating to a carbon with a chemical shift of 26.8 ppm, matching expected values of the terminal methyl at C-13. Thus, these data suggest that the canonical propionate starter unit of the cosmomycin type II PKS is replaced with acetate in the heterologous system, thereby shortening the carbon chain by one carbon in **A**.

After the identification of the cosmomycin series in *S. coelicolor* M512/pCAP01-COS, genome analysis revealed a homologous BGC in *Streptomyces* sp. CNS-615 (Figure 2). Following analysis of culture extracts from *Streptomyces* sp. CNS-615 and subsequent molecular networking against M512/pCAP01-COS and *Streptomyces* sp. CNT-302, *Streptomyces* sp. CNS-615 was similarly identified as a producer of **1**–**3** but not of any of the putatively acetate-primed desmethyl species **A** or **B** (Figure S18).

Identification of the cosmomycin series in culture extracts of *S. coelicolor* M512/pCAP01-COS demonstrated that the type II PKS BGC captured via TAR is sufficient for the production of the cosmomycin natural products. These results confirm the findings by Padilla and co-workers, who demonstrated the necessity of two glycosyltransferases in the biosynthesis of cosmomycin D from *Streptomyces olindensis*, which each share approximately 95% identity with the glycosyltransferases captured within the pathway from CNT-302.[13]

During the course of this study, the genome sequence of *S. olindensis* was published,[14] which included a type II PKS BGC with an average amino acid identity of 90% to the *cos* cluster captured in this study (Figure 1, Table S2). Knowledge of the final product has allowed us to confirm complete absence of production by the native "producer" *Streptomyces* sp. CNT-

302 under a range of media conditions (Figure S19), yet expression was achieved in *S. coelicolor* M512/pCAP01-COS without the need for genetic manipulation of the pathway. This observation illustrates the potential for using TAR capturing and heterologous expression to unleash otherwise silent gene clusters, while also demonstrating the synergy of combining heterologous expression and molecular networking to connect genes to molecules and reveal new members of molecular families.

The TAR cloning protocol has been successful at capturing BGCs directly from genomic[6] and metagenomic[8] DNA samples for the efficient characterization of natural products and the processes involved in their assembly. However, the staggering number of available bacterial genomes and orphan biosynthetic clusters therein suggests the need for a high-throughput methodology to investigate novel biosynthetic chemistry. The lengthy TAR protocol requiring individual optimization and troubleshooting is unsuitable for high-throughput technologies. Here, we present a streamlined version of the protocol, avoiding PCR assembly of each custom capture vector, instead utilizing synthetic DNA inserted with a simple ligation. Although the synthetic capture arms used in this study cost approximately $1/bp, at the time of publication the price has fallen to $0.15/bp, and the time savings compared to PCR-based assembly of capture vectors is estimated to be 1 week for a single investigator. We have demonstrated that this new method is as effective as the original approach, showing similar efficiency of transformation in addition to effectively capturing and expressing the cosmomycin biosynthetic pathway. Since this work was performed, the pCAP system has undergone further improvements such as the shortening of the synthetic insert length to 144 bp by including 60 bp capture arms, the addition of regions homologous to the vector backbone allowing for one-step assembly, and the insertion of a 5-fluoroorotic acid-mediated negative selection mechanism, which dramatically increases efficiency from ∼1−2% to over 50%.[22] These recent advancements allow us to entertain the idea of a high-throughput TAR protocol, which could potentially automate a large number of capture experiments through rapid, PCR-independent assembly of capture vectors and subsequent parallel transformations with prepared gDNA. Furthermore, whole pathway capture allows for the simple manipulation of biosynthetic machinery. Through manipulating biosynthetic genes, novel analogues and compounds can be engineered and produced from known clusters.[23] In the case of the cosmomycin series, new glycosylation patterns are envisioned through glycorandomization engineering processes.[24] Considering the promising antitumor activity of the known cosmomycin series and the importance of the carbohydrate moiety for bioactivity,[25] this approach could yield powerful analogues with even greater efficacy. With this engineering approach to modifying biosynthesis, combined with the powerful TAR pathway cloning, there is a large area of chemical space waiting to be explored.

■ EXPERIMENTAL SECTION

General Experimental Procedures. NMR spectra were collected at 298 K using a Bruker Avance III 600 MHz spectrometer fitted with a 5 mm TCI cryoprobe. The HSQC spectra were acquired using 12.5% nonuniform sampling and reconstructed using SMILE and NMRPipe. Spectra were recorded in $CDCl_3$ at 25 °C, and chemical shifts are given on the δ scale referenced to residual chloroform (δ_H 7.26) (Figures S12−17, Table S1). For mass spectrometry analysis, an Agilent 1290 liquid chromatography system coupled to an Agilent 6530 Q-TOF was used. Liquid chromatography fractionation was carried out on both an Agilent Prepstar and an Agilent 1260 Infinity LC system. Electroporations were carried out on an Eppendorf Electroporator 2510.

TAR Cloning Procedure. The *cos* BGC was identified in the genome of *Streptomyces* sp. CNT-302 (GenBank accession ARIM00000000.1) using the antiSMASH software suite,[26] and homologous capture arms were designed with 360 base pairs of homology to the boundaries of the cluster. The capture arms were synthesized as a DNA geneblock (Integrated DNA Technologies) with a *Bam*HI restriction site between them. Capture arms were inserted into the SpeI and XhoI restriction sites in the pCAP01 (AddGene accession number 59981) backbone using digestion and ligation to form plasmid pCAP01-COM. Successful assembly was confirmed by restriction analysis.

A 1 μg amount of pCAP01-COM was linearized using the *Bam*HI restriction site included in the synthetic insert sequence, and the linear vector was gel purified using a Qiagen gel extraction kit following the standard manufacturer's protocol. From this point onward, a slightly modified version of the original TAR protocol[10] was followed and can be found deposited with the pCAP01 sequence on the Addgene database. Briefly, yeast spheroplasts (200 μL) were transformed with linear vector (200 ng) and gDNA (1 μg) from CNT-302 previously digested with ClaI. Transformed spheroplasts were plated on selective agar and incubated for 4 days at 30 °C. Colonies were PCR screened for regions of the target biosynthetic cluster, and plasmids from positive hits (1.5% of colonies analyzed) were isolated using the Zymoprep Yeast Plasmid Miniprep I kit (Zymo Research) and transformed into *E. coli* Top10 (Life Technologies) by electroporation (Figures S1 and S2). The correct capture of the biosynthetic cluster into pCAP01-COM was confirmed by restriction analysis, and the resulting plasmid was named pCAP01-COS.

The plasmid pCAP01-COS was transformed into *S. coelicolor* M512 using triparental conjugation[27] with *E. coli* ET12567 cells containing either pUB307 conjugation plasmid or the pCAP01-COS vector containing the captured cluster. Three sequential rounds of selection were performed on selective media (MS agar + nalidixic acid (100 μg/mL) + kanamycin (50 μg/mL) + chloramphenicol (35 μg/mL)) to ensure the plasmid was integrated and maintained. Integration was confirmed by PCR amplification and Sanger sequencing (Figures S3 and S4).

Heterologous Expression of the Pathway and Identification of the Products. Spores of *S. coelicolor* M512/pCAP01-COS were used to streak plates of R5 production media. After 1 week, plugs of the agar plate (1 cm in diameter) were removed, washed with 5 mL of H_2O, and extracted with 5 mL of EtOAc. During the course of this study, a second marine streptomycete (*Streptomyces* sp. CNS-615, GenBank accession AQPE00000000.1) was found to contain the same target type II PKS BGC. In addition to M512 and M512/pCAP01-COS, the two marine *Streptomyces* sp. strains containing the S4 kb pathway (CNT-302 and CNS-615) were also grown and extracted in the same manner. Production of secondary metabolites was confirmed by HPLC-HR-ESI-MSMS analysis, carried out on an Agilent 1290 liquid chromatography system coupled to an Agilent 6530 Q-TOF (200−2000 m/z, 20 keV) (Figures S5−S11). Prepared samples were dissolved in MeOH and injected on a C18 column (Phenomenex Luna 5 μm C18(2) 100A 100 × 4.6 mm) with an initial mobile phase of 10% MeCN at a flow rate of 0.7 mL/min. Over the course of 30 min, the MeCN concentration was linearly increased to 100%. The collected data were subjected to the GNPS molecular networking workflow[19] and analyzed as described previously (Cosine Score = 0.7).[18] Samples were then grouped and visualized using Cytoscape (Figures S18 and S19).

Purification of Compounds 1 and A. Liquid R5 media was inoculated with *S. coelicolor* M512/pCAP01-COS and grown for 10 days. One liter of culture supernatant was extracted three times with 500 mL of EtOAc and subsequently dried. The extract was then loaded on a preparative HPLC column (Agilent Pursuit XRs 5 μm C18, 100 × 21.2 mm) and subjected to a standard reversed-phase gradient from

5% to 100% MeCN in 100 mM ammonium acetate buffer over 60 min at a flow rate of 15 mL/min. All subsequent HPLC purifications utilized the same 100 mM ammonium acetate buffer. Individual fractions collected from the preparative column containing cosmomycin family members (eluted at ~40% MeCN) were then subjected to another round of HPLC purification (Phenomenex Luna 5 μm C8 100A, 250 × 10 mm). Cosmomycin family members were eluted on an isocratic 40% MeCN concentration, and individual peaks were collected, dried, and subjected to a final round of C18 purification (Phenomenex Kinetex 5 μm XB-C18 100A, 250 × 4.6 mm) using the same conditions as the previous C8 column purification. The major individual peaks were collected and dried on a lyophilizer for NMR analysis. Two liters of extracted culture supernatant yielded 2.5 mg of **1** and 1.5 mg of **A**.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.6b01121.

HRMS and MS/MS data for compounds **1**−**3**, **A**, and **B**; ^{1}H and 2D NMR of **1** and **A**; analysis of plasmid pCAP01-COS, molecular network of organisms containing the cosmomycin BGC (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bsmoore@ucsd.edu.

ORCID

Bradley S. Moore: 0000-0002-4652-1253

Present Address

⊥Institute of Pharmaceutical Biology, University of Bonn, 53115 Bonn, Germany.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank P. Jensen and W. Fenical (UC San Diego) for providing *Streptomyces* strains CNT-302 and CNS-615, M. Bibb (John Innes Centre, UK) for providing *S. coelicolor* M512, Y. Su at the UCSD Molecular Mass Spectrometry Facility for mass spectrometry assistance, B. Duggan for NMR spectroscopy assistance, and E. O'Neill, X. Tang, S. McKinnie, Y. Kudo, and J. Li (UC San Diego) for valuable discussions. This work was supported by U.S. National Institutes of Health grant R01-GM085770 and a DFG postdoctoral fellowship CR464-1 to M.C. The genome sequences were provided by the Jensen lab and obtained by the U.S. Department of Energy Joint Genome Institute with support by the Office of Science, U.S. Department of Energy, under Contract No. DE-AC02-05CH11231.

■ REFERENCES

(1) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. *Nat. Rev. Drug Discovery* **2015**, *14*, 111−129.

(2) Cragg, G. M.; Newman, D. J. *Biochim. Biophys. Acta, Gen. Subj.* **2013**, *1830*, 3670−3695.

(3) Nett, M.; Ikeda, H.; Moore, B. S. *Nat. Prod. Rep.* **2009**, *26*, 1362−1384.

(4) Medema, M. H.; Breitling, R.; Takano, E. *Synthetic Biology in Streptomyces Bacteria*; Academic Press, 2011; Vol. *497*, pp 485−502.

(5) Ongley, S. E.; Bian, X.; Neilan, B. A.; Muller, R. *Nat. Prod. Rep.* **2013**, *30*, 1121−38.

(6) Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1957−1962.

(7) Kouprina, N.; Larionov, V. *Nat. Protoc.* **2008**, *3*, 371−7.

(8) Kallifidas, D.; Kang, H. S.; Brady, S. F. *J. Am. Chem. Soc.* **2012**, *134*, 19552−19555.

(9) Ross, A. C.; Gulland, L. E.; Dorrestein, P. C.; Moore, B. S. *ACS Synth. Biol.* **2014**, *4*, 414−420.

(10) Li, Y.; Li, Z.; Yamanaka, K.; Xu, Y.; Zhang, W.; Vlamakis, H.; Kolter, R.; Moore, B. S.; Qian, P. Y. *Sci. Rep.* **2015**, *5*, 9383.

(11) Chao, R.; Yuan, Y.; Zhao, H. *FEMS Yeast Res.* **2014**, *15*, 1−9.

(12) Luo, Y.; Li, B. Z.; Liu, D.; Zhang, L.; Chen, Y.; Jia, B.; Zeng, B. X.; Zhao, H.; Yuan, Y. J. *Chem. Soc. Rev.* **2015**, *44*, 5265−5290.

(13) Garrido, L. M.; Lombo, F.; Baig, I.; Nur, E. A. M.; Furlan, R. L.; Borda, C. C.; Brana, A.; Mendez, C.; Salas, J. A.; Rohr, J.; Padilla, G. *Appl. Microbiol. Biotechnol.* **2006**, *73*, 122−131.

(14) Rojas, J. D.; Starcevic, A.; Baranasic, D.; Ferreira-Torres, M. A.; Contreras, C. A.; Garrido, L. M.; Araujo, W. L.; de Souza, R. F.; Zucko, J.; Hranueli, D.; Long, P. F.; Cullum, J.; Padilla, G. *Genome Announc* **2014**, *2*, 1−2.

(15) Kersten, R. D.; Ziemert, N.; Gonzalez, D. J.; Duggan, B. M.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E4407−4416.

(16) Noskov, V. N.; Koriabine, M.; Solomon, G.; Randolph, M.; Barrett, J. C.; Leem, S.-H.; Stubbs, L.; Kouprina, N.; Larionov, V. *Nucleic Acids Res.* **2001**, *29*, e32.

(17) Floriano, B.; Bibb, M. *Mol. Microbiol.* **1996**, *21*, 385−96.

(18) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. *J. Nat. Prod.* **2013**, *76*, 1686−1699.

(19) Wang,; et al. *Nat. Biotechnol.* **2016**, *34*, 828−837.

(20) Hirayama, K.; Akashi, S.; Ando, T.; Horino, I.; Etoh, Y.; Morioka, H.; Shibai, H.; Murai, A. *Biomed. Environ. Mass Spectrom.* **1987**, *14*, 305−312.

(21) Kelso, C.; Rojas, J. D.; Furlan, R. L.; Padilla, G.; Beck, J. L. *Eur. Mass Spectrom.* **2009**, *15*, 73−81.

(22) Tang, X.; Li, J.; Millan-Aguinaga, N.; Zhang, J. J.; O'Neill, E. C.; Ugalde, J. A.; Jensen, P. R.; Mantovani, S. M.; Moore, B. S. *ACS Chem. Biol.* **2015**, *10*, 2841−2849.

(23) Metsä-Ketelä, M.; Palmu, K.; Kunnari, T.; Ylihonko, K.; Mäntsälä, P. *Antimicrob. Agents Chemother.* **2003**, *47*, 1291−1296.

(24) Langenhan, J. M.; Griffith, B. R.; Thorson, J. S. *J. Nat. Prod.* **2005**, *68*, 1696−1711.

(25) Furlan, R. L.; Watt, S. J.; Garrido, L. M.; Amarante-Mendes, G. P.; Nur-e-alam, M.; Rohr, J.; Braña, A.; Mendez, C.; Salas, J. A.; Sheil, M. M.; Beck, J. L.; Padilla, G. *J. Antibiot.* **2004**, *57*, 647−654.

(26) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Muller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43*, W237−243.

(27) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*; John Innes Foundation: Norwich, 2000; Vol. *1*, pp 250−251.

A PCR-independent Method of Transformation Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete

Charles B. Larson, Max Crüsemann, and Bradley S. Moore

Supporting Information

Table of Contents

SI2: **Figure S1** PCR screening of multiple yeast colonies

SI3: **Figure S2** PCR screening of individual yeast colonies

SI4: **Figure S3** PCR amplification of integrated cosmomycin pathway

SI5: **Figure S4** Sequencing of integrated cosmomycin pathway

SI6: **Figure S5** MS2 fragmentation analysis of cosmomycin C **(1)**

SI7: **Figure S6** MS2 fragmentation analysis of cosmomycin D **(2)**

SI8: **Figure S7** MS1 spectrum of cosmomycin analog **3**

SI9: **Figure S8** MS2 fragmentation analysis of cosmomycin analog **3**

SI10: **Figure S9** MS1 spectrum of cosmomycin analog **A**

SI11: **Figure S10** MS2 fragmentation analysis of cosmomycin analog **A**

SI12: **Figure S11** MS2 fragmentation analysis of cosmomycin analog **B**

SI13: **Figure S12** ^{1}H NMR spectrum of cosmomycin C (**1**)

SI14: **Figure S13** ^{1}H NMR spectrum of **A**

SI15: **Figure S14** Overlaid ^{1}H NMR spectra of **1** and **A**

SI16: **Figure S15** HSQC NMR spectrum of **1**

SI17: **Figure S16** HSQC NMR spectrum of **A**

SI18: **Figure S17** COSY NMR spectrum of **1**

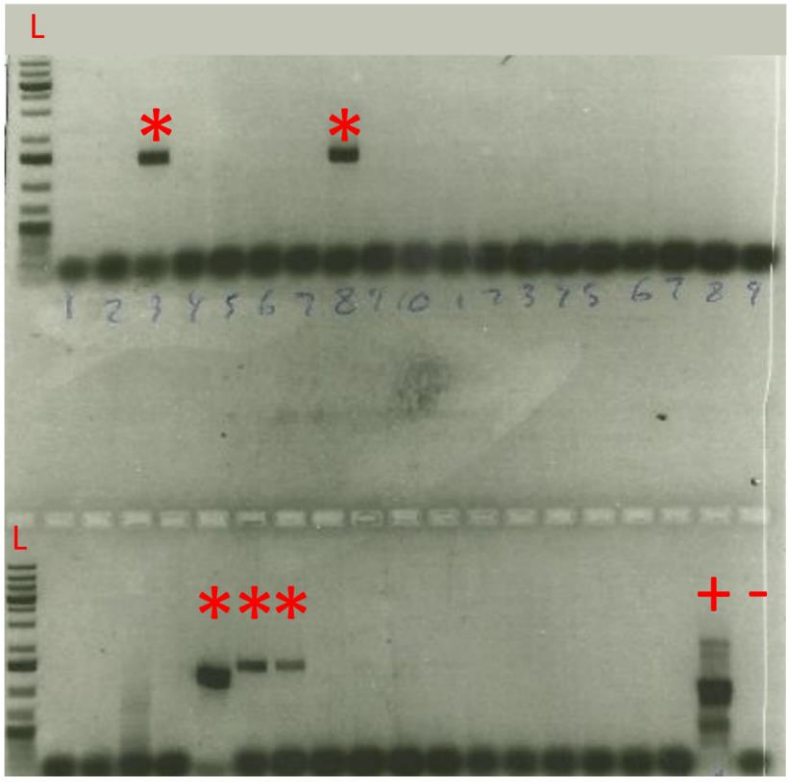
SI19: **Figure S18** Molecular network of cosmomycin native producers and heterologous host.

SI20: **Figure S19** *Streptomyces* sp. CNT-302 molecular network

SI21-23: **Table S1** ^{1}H NMR chemical shift assignments of compound **1**, **A**, and related cosmomycin analogs

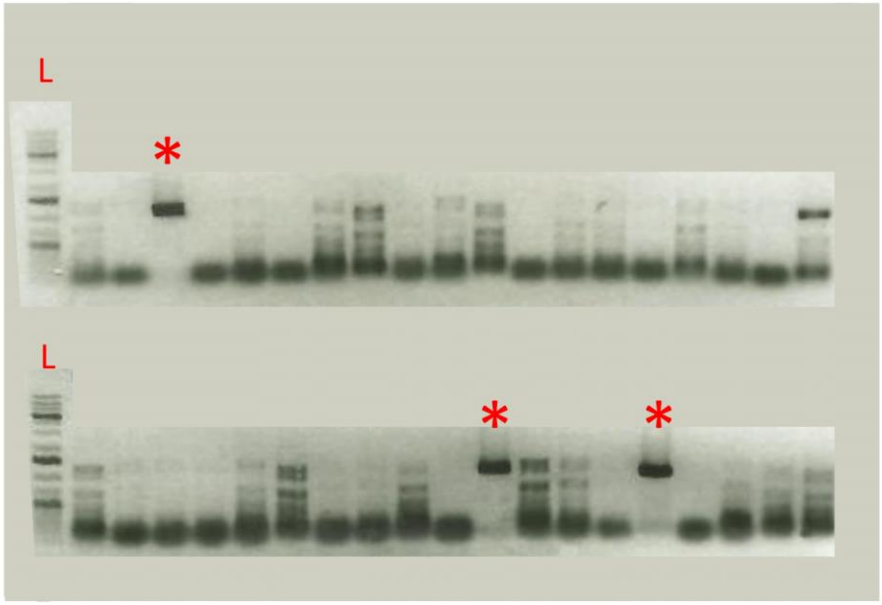
SI24-26: **Table S2** Comparison of *Streptomyces* sp. CNT-302 and *S. olindensis* BGCs

Figure S1. PCR screening of multiple yeast colonies



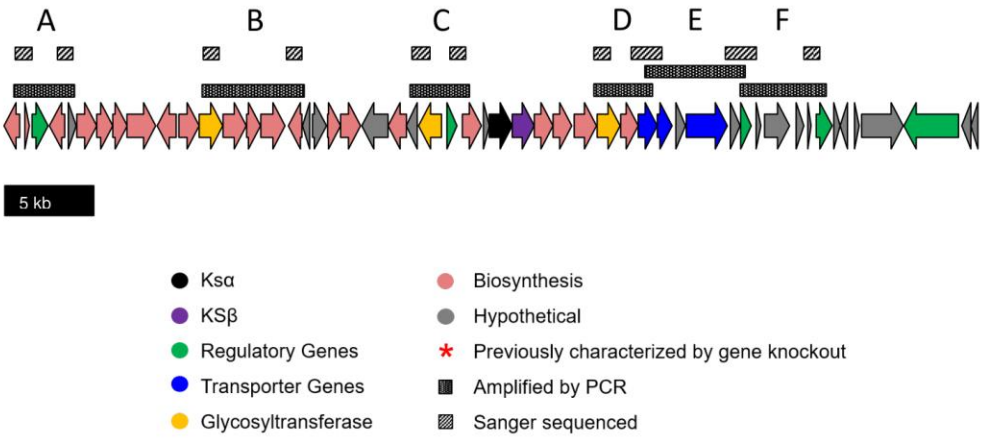
PCR amplification of a 1400 bp fragment of the target BGC performed on groups of 10 yeast clones. Positive amplification of target region indicated by "*". Positive and negative controls with and without CNT-302 gDNA shown as "+" and "–", respectively. Lanes loaded with 1 kb-plus DNA ladder (ThermoFisher) labeled with "L".

Figure S2. PCR screening of individual yeast colonies



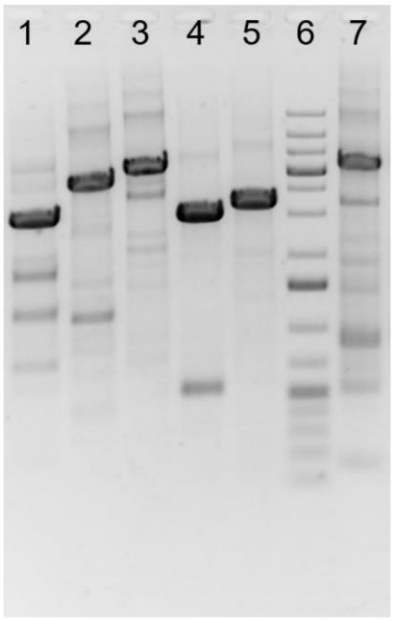
PCR amplification of a 1400 bp fragment of the target BGC performed on individual yeast clones. Lanes loaded with 1 kb-plus DNA ladder labeled with "L". Positive hits indicated by "*". A total of three clones showed clean amplification, with another clone showing an amplification band with lower molecular weight DNA also in the lane. A total of 200 colonies were initially investigated, calculated transformation efficiency was 1.5%.

Figure S4. Sequencing of integrated cosmomycin pathway



In order to further confirm proper assembly and integration of plasmid pCAP01-COS, regions A-F depicted above were amplified and Sanger sequenced. Results confirmed the sequence of the integrated pathway was correct for all amplicons. A total of 10 kb of the pathway was covered by the sequencing effort.

Figure S3. PCR amplification of integrated cosmomycin pathway

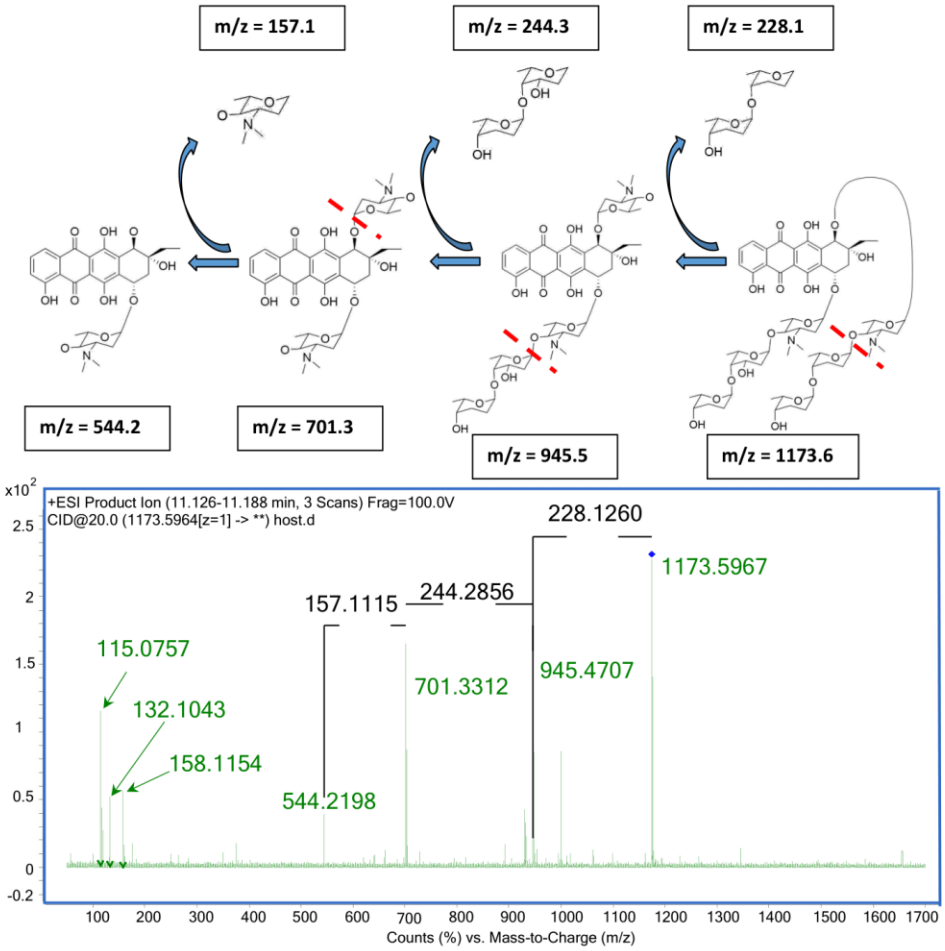


After integrating pCAP01-COS into the genome of *S. coelicolor* M512, various regions were amplified by PCR to confirm their size, proper assembly, and integration.

Expected Size (bp)

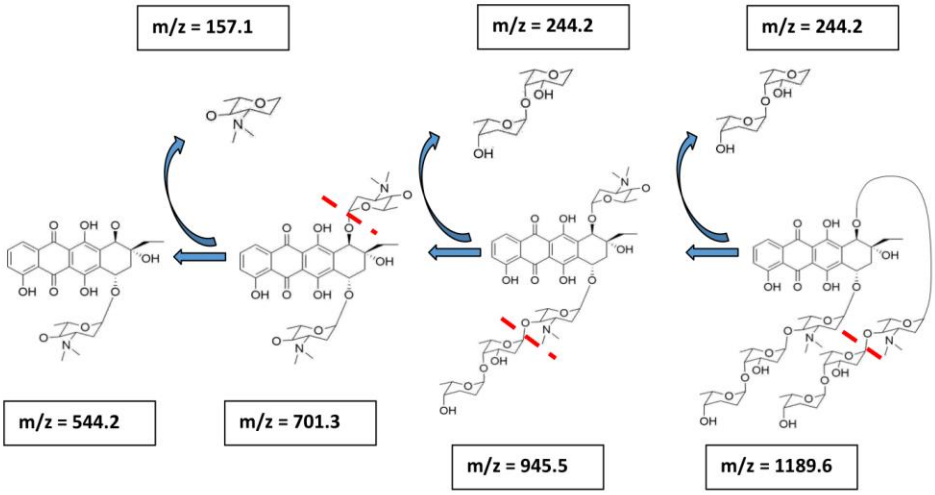
1: **3191** (Region D)
2: **4639** (Region F)
3: **5602** (Region E)
4: **3040** (Region C)
5: **3419** (Region A)
6: 1kb Plus Ladder: 20000, 10000, 7000, **5000**, 4000, 3000, 2000, **1500**, 1000, 700, **500**, 400, 300, 200
7: **5555** (Region B)

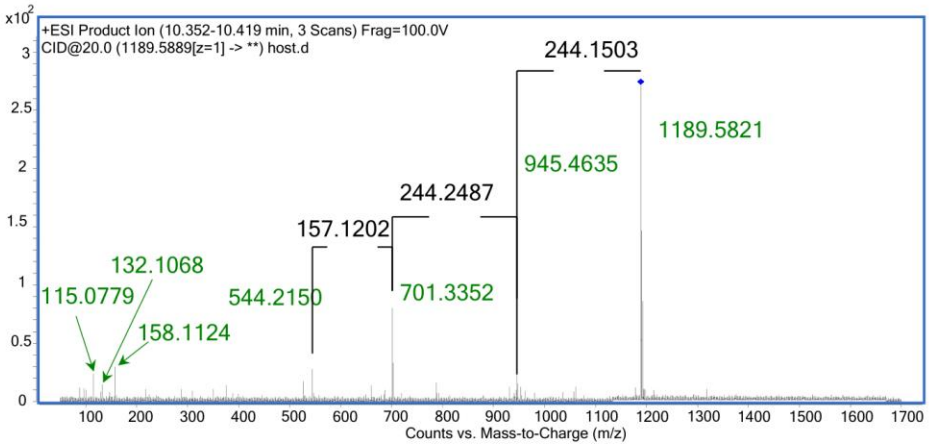
Figure S5. MS2 fragmentation analysis of cosmomycin C **(1)**



Fragmentation of parent ion 1173.59 shows losses corresponding to sugar moieties of cosmomycin C. The initial loss of 228 mass units correspond to the B ions of two rhodinose moieties, the two outermost sugars on one of the trisaccharide chains. The second loss of 244 corresponds to rhodinose and 2-deoxyfucose lost from the second trisaccharide chain. The last shift seen represents the loss of rhodosamine, with a B ion mass of 157. All three sugars can be seen in the lower mass range of the spectrum.

Figure S6. MS2 fragmentation analysis of cosmomycin D **(2)**





Fragmentation of parent ion 1189.589 shows losses corresponding to sugar moieties of cosmomycin D. The initial loss of 244 mass units correspond to the B ions of rhodinose and L-2-deoxy-fucose, the two outermost sugars on both the trisaccharide chains. The second loss of 244 corresponds to the same two sugars lost from the second trisaccharide chain. The last shift seen represents the loss of rhodosamine, with a B ion mass of 157. All three sugars can be seen in the lower mass range of the spectrum.

Figure S7. MS1 spectrum of cosmomycin analog **3**

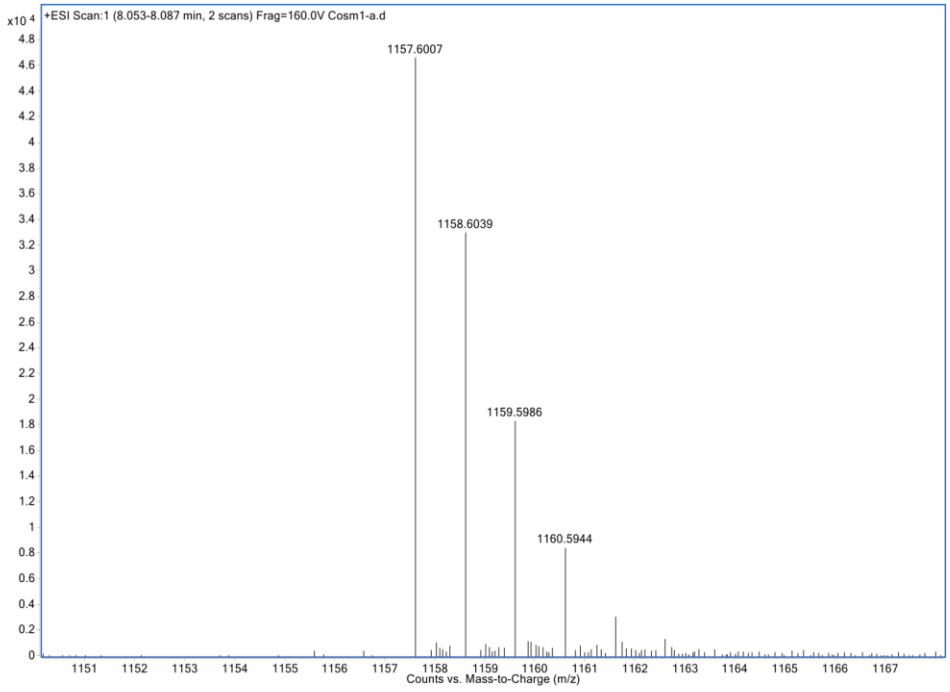
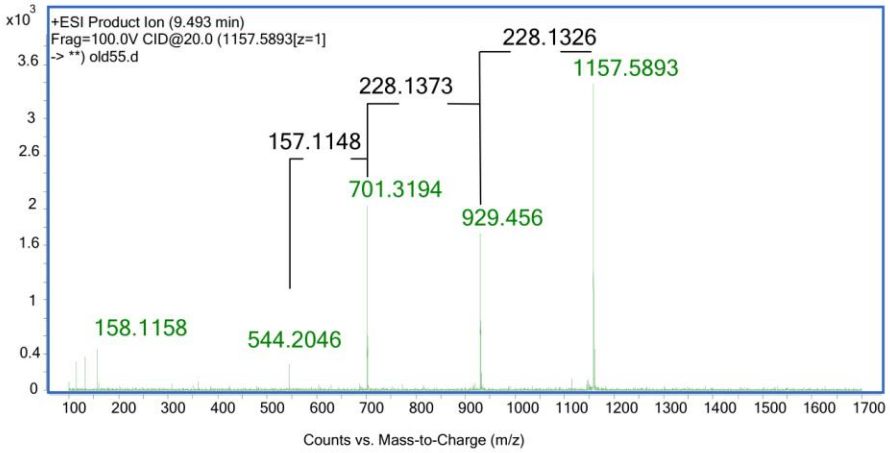


Figure S8. MS2 fragmentation analysis of cosmomycin analog **3**



Fragmentation of parent ion 1157.5893 shows losses corresponding to sugar moieties of the previously reported compound A447 C, also reported by Kelso et al[1]. The initial loss of 228 mass units corresponds to the B ions of two rhodinose moieties, the two outermost sugars on one of the trisaccharide chains. The second loss of 228 corresponds to the same moieties lost from the second trisaccharide chain. The final loss of 157 corresponds to a rhodosamine B ion.

Figure S9. MS1 spectrum of cosmomycin analog **A**

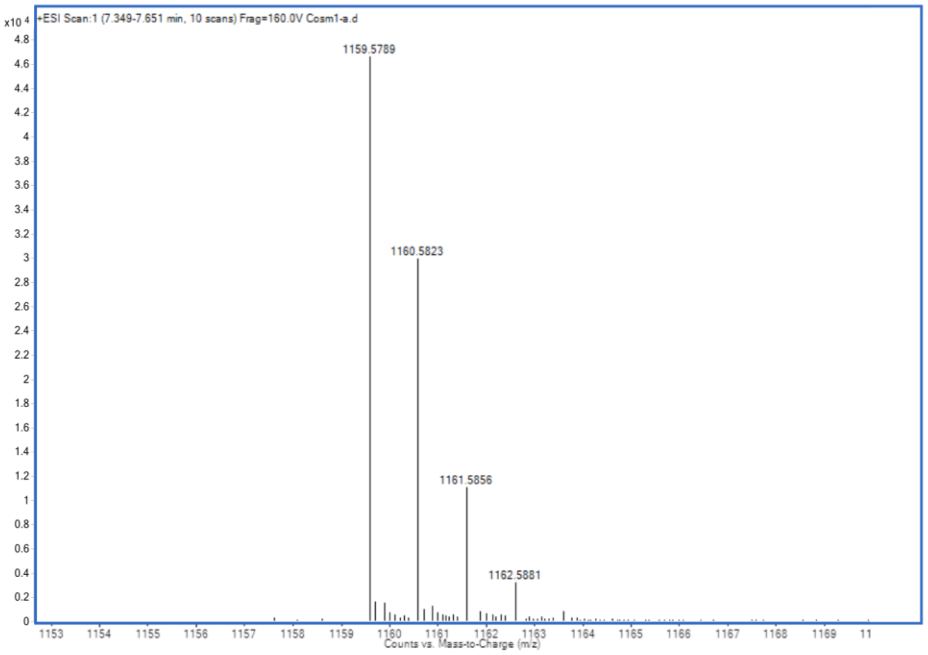
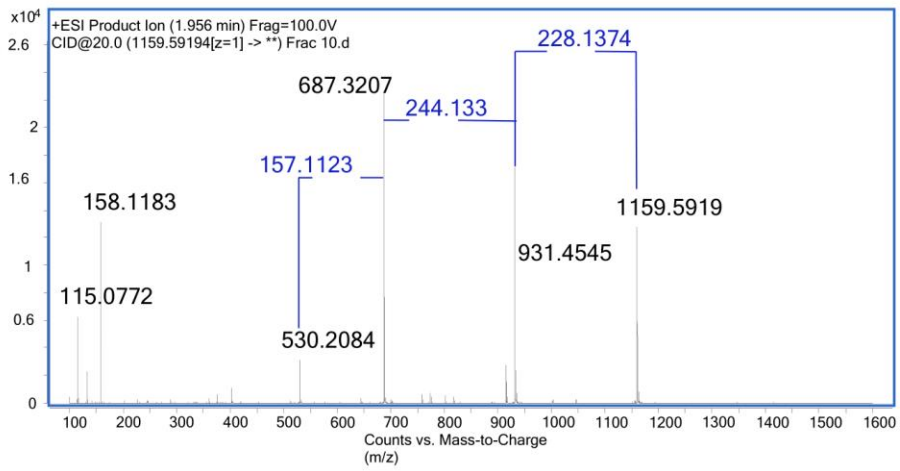
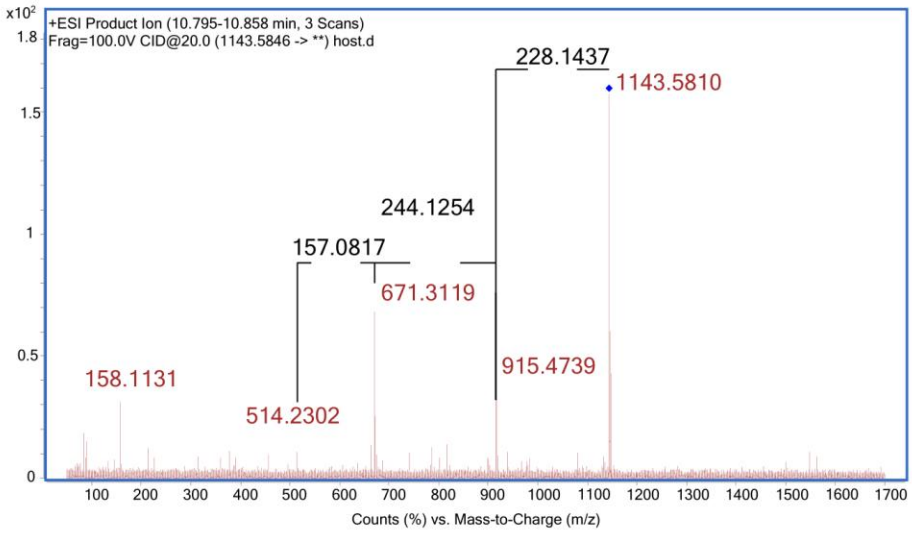


Figure S10. MS2 fragmentation analysis of cosmomycin analog **A**



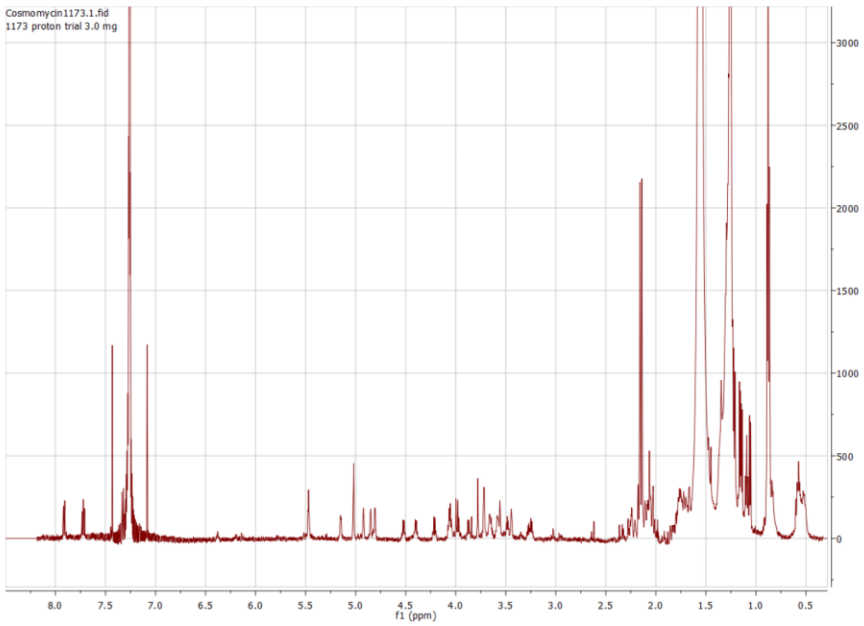
Fragmentation of parent ion 1159.5919 shows an initial loss of 228 mass units corresponding to the B ions of two rhodinose moieties, the two outermost sugars on one of the trisaccharide chains. The second loss of 244 corresponds to rhodinose and 2-deoxyfucose lost from the second trisaccharide chain. Although the sugar masses fragmented here are identical to the sugars lost from cosmomycin C, the mass of the rest of the molecule is 14 Daltons lighter than cosmomycin C at each stage of fragmentation. This suggests the lack of a methylene group the aglycone.

Figure S11. MS2 fragmentation analysis of cosmomycin analog **B**



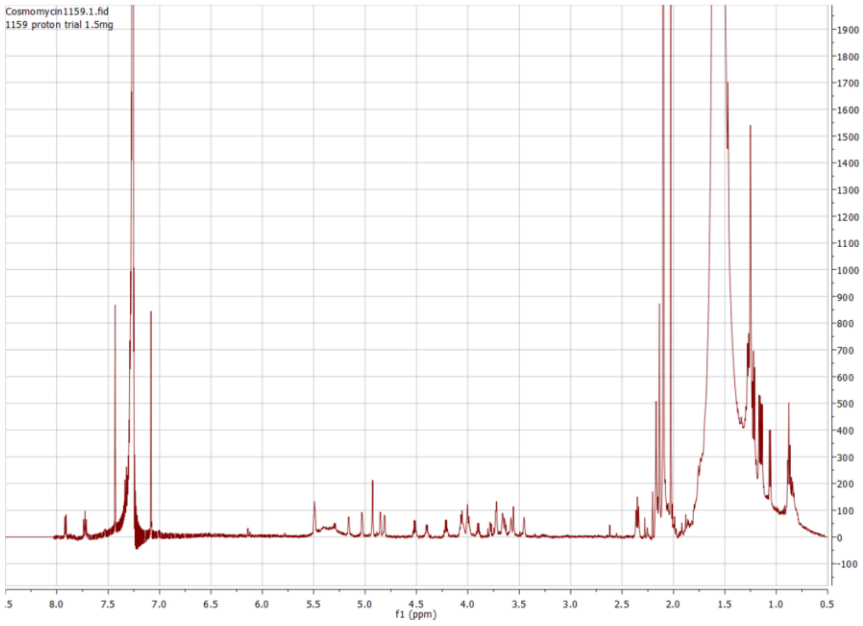
The initial loss of 228 mass units corresponds to the B ions of two rhodinose residues, the two outermost sugar units on one of the trisaccharide chains. The second loss of 244 corresponds to rhodinose and 2-deoxyfucose lost from the second trisaccharide chain. The mass shifts of the sugar moieties fragmented here match the first two fragments of cosmomycin C, indicating the same glycosylation pattern on the four outermost sugars. The mass of the unfragmented molecule, as well as the mass of the first three fragments, show a mass shift of -30 Daltons compared to cosmomycin C. This mass shift is consistent with the loss of an oxygen and a methylene group, suggesting this species is desoxy, desmethyl-cosmomycin C (compound **B**).

Figure S12. ^{1}H NMR spectrum of cosmomycin C (**1**)



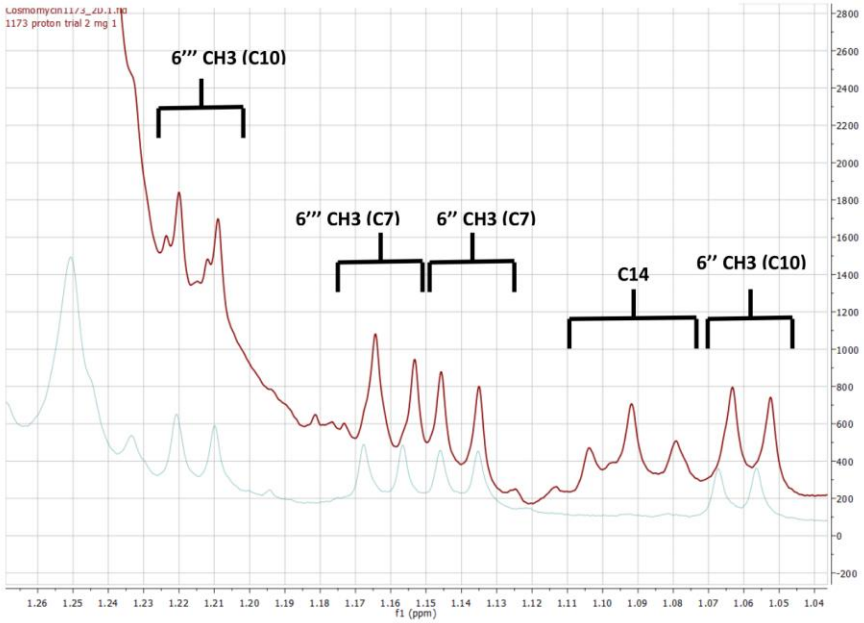
NMR Spectrum of cosmomycin C shows three distinct aromatic protons at 7.92 ppm, 7.72 ppm, and 7.33 ppm. Sugar protons can be seen from 3.5 to 5.5 ppm, and the N-methyl protons from the inner rhodosamines can be seen at 2.17 and 2.14.

Figure S13. ^{1}H NMR spectrum of **A**



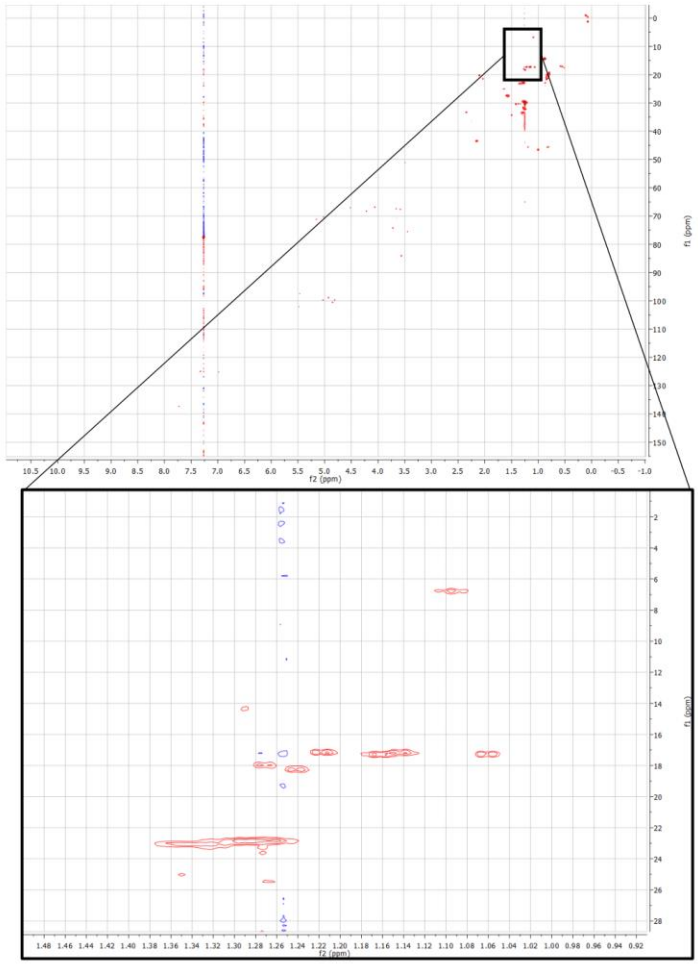
NMR Spectrum of compound **4** shows three distinct aromatic protons at 7.92 ppm , 7.72 ppm, and 7.33 ppm. Sugar protons can be seen from 3.5 to 5.5 ppm, and the N-methyl protons from the inner rhodosamines can be seen at 2.17 and 2.14.

Figure S14. Overlaid ^{1}H NMR spectrum of **1** and **A**



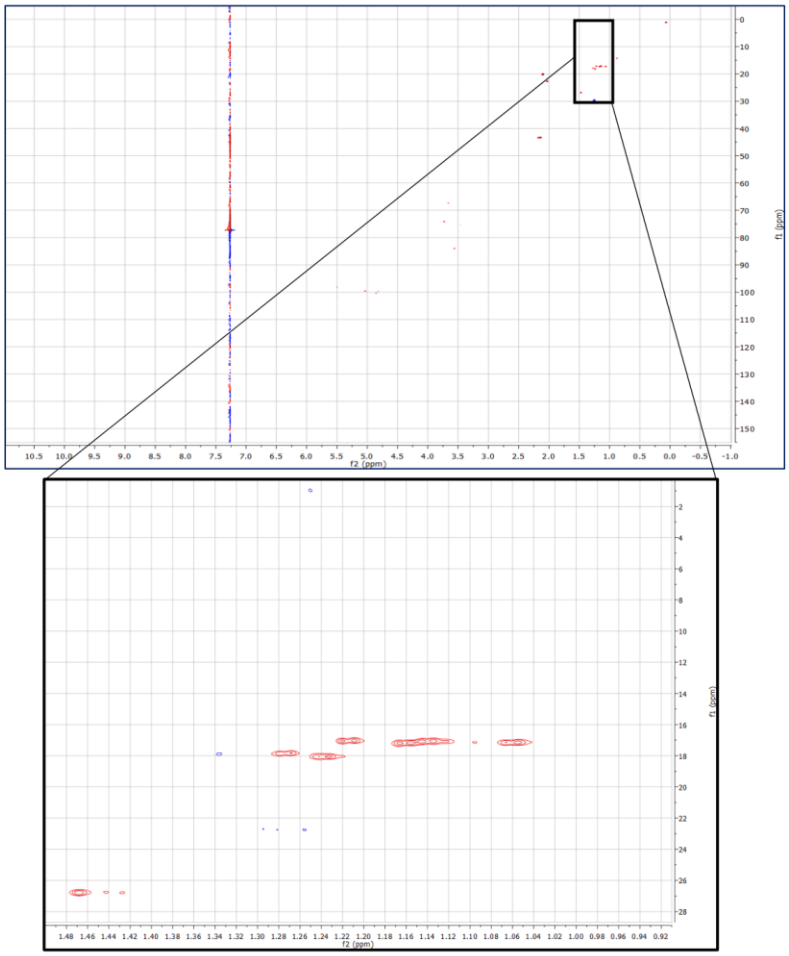
Overlaid ^{1}H NMR spectra of compounds **1** (red signal) and **A** (blue signal) show the loss of a triplet signal at 1.09 ppm in compound **A**. This signal corresponds to the C-14 methyl group of the aglycone.[4,5]

Figure S15. HSQC spectrum of **1**



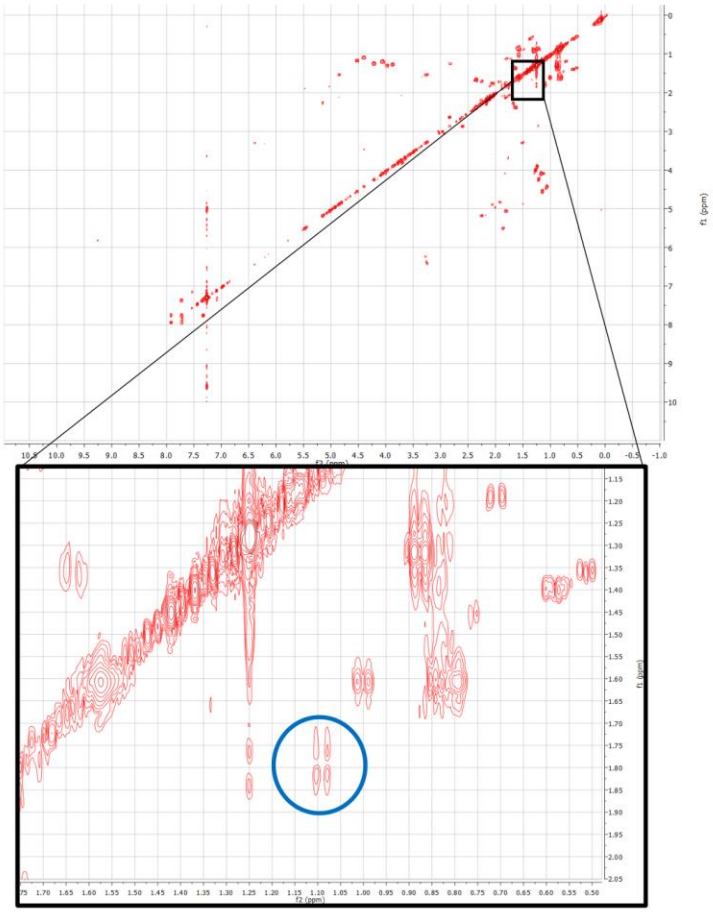
In the HSQC spectrum of compound **1** was acquired using 12.5% non-uniform sampling and reconstructed using SMILE[2] and NMRPipe[3] the triplet at 1.09 can clearly be seen correlating to carbon 14 of the aglycone which has a diagnostic chemical shift of 6.7 ppm. Also in the expanded region, the methyl protons attached to carbon six of all six sugars can be seen correlating to carbons with chemical shifts from ~17 to 18.3 ppm, matching literature values.[4-8]

Figure S16. HSQC NMR spectrum of **A**



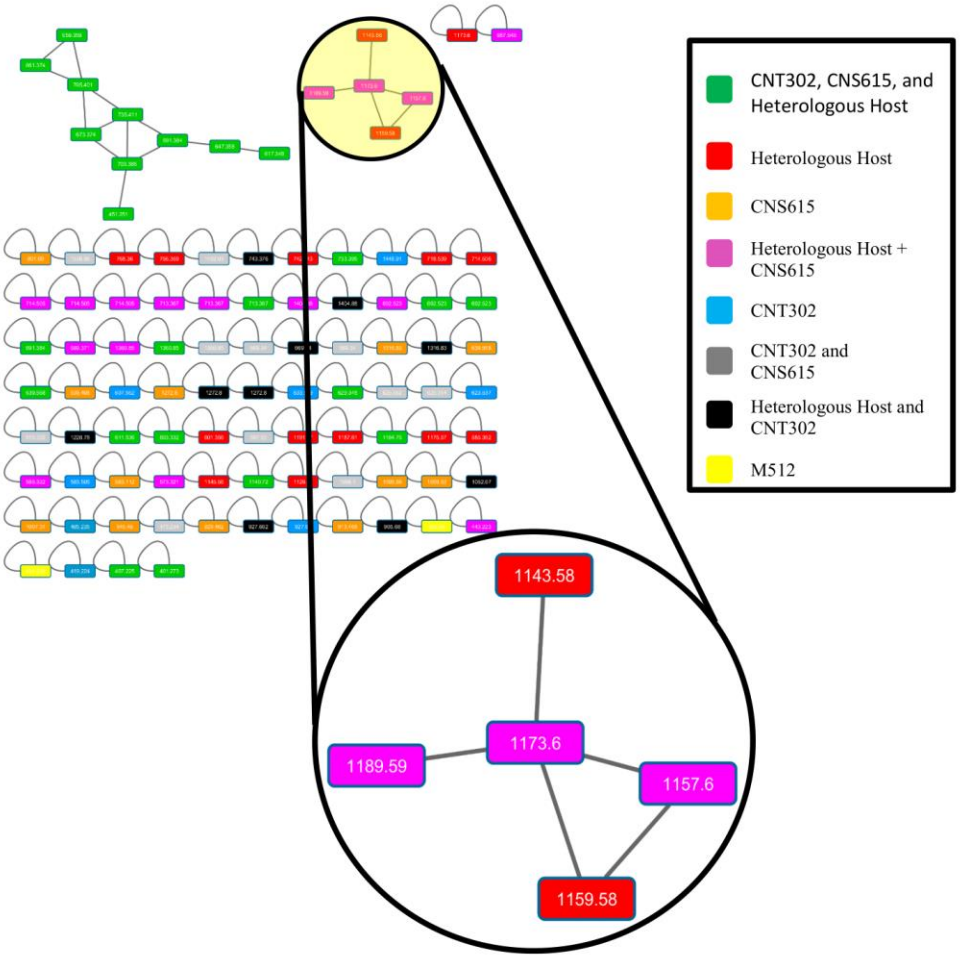
In the HSQC spectrum of compound **A**, the triplet at 1.09 ppm corresponding to carbon 14 is missing, and an additional singlet peak at 1.47 ppm which correlates to a carbon with a chemical shift of 26.8 ppm can be clearly seen. This data suggests the new peak represents carbon 13, which is now a terminal methyl group rather than a methylene. Such a structure is consistent with the canonical propionate PKS starter unit of cosmomycin being replaced by a malonate starter unit in the heterologous host.

Figure S17. COSY NMR spectrum of **1**



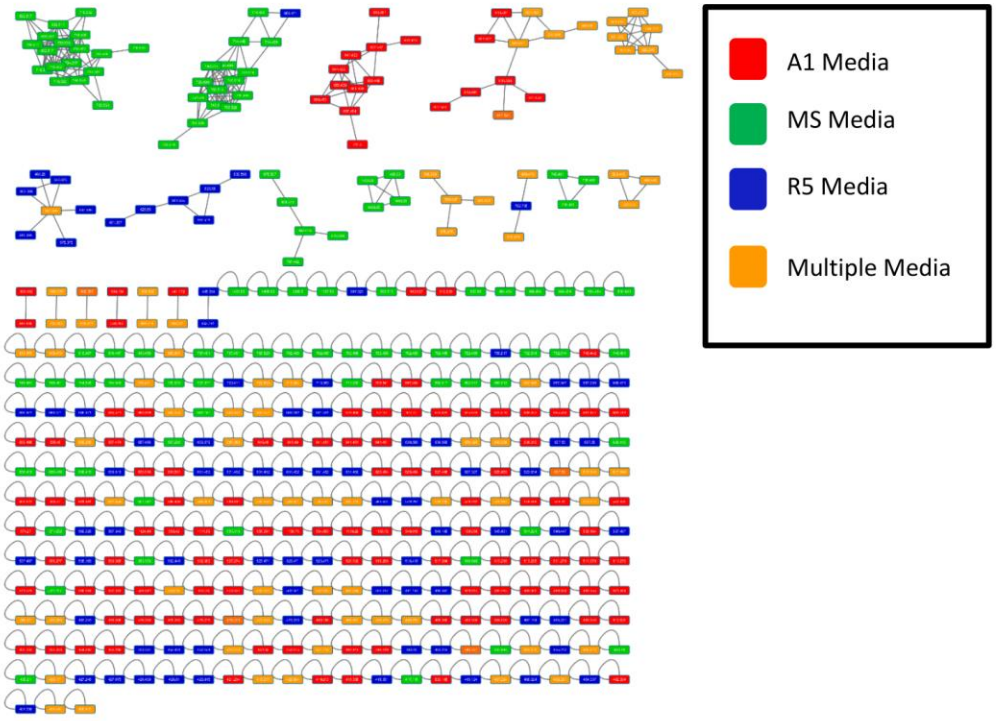
COSY correlation between carbon 14 protons observed at 1.09 ppm and two protons with chemical shifts of 1.76 and 1.82 (circled). These shifts are consistent with literature values for protons 13a and 13b.[6-8]

Figure S18. Molecular network of cosmomycin native producers and heterologous host



Molecular network generated using the GNPS algorithm.[9,10] Red nodes are found only in the heterologous host, while orange nodes are found only in CNS-615. Nodes colored purple are found in both strains. CNT-302 extracts were included in the networking, but no compounds related to cosmomycin were present. In addition to our own extracts, publicly available data on the strains was also included in the network (MassIVE accession: MSV000078836)[11] and visualization was performed by cytoscape.[12] The cosmomycin molecular family is highlighted and expanded.

Figure S19. *Streptomyces* sp. CNT-302 molecular network

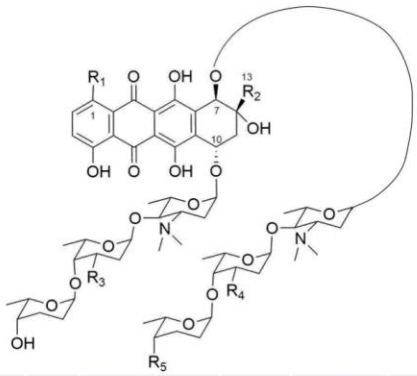


Networking results from CNT-302 butanol, methanol, and ethyl acetate extractions obtained from archived MassIVE data set MSV000078836. No masses matching cosmomycin C, cosmomycin D, or other cosmomycin family members were detected.

Table S1. [1]H NMR chemical shift assignments of compounds **1** and **4** and the related cosmomycin analogs obelmycin C[4] and A447C/D[5]

	Obelmycin C	A447 C	A447 D	Cosmomycin C (1)	Compound A
R1	OH	H	H	H	H
R2	CH2CH3	CH2CH3	CH2CH3	CH2CH3	CH3*
R3	OH	H	H	OH	OH
R4	OH	H	H	H	H
R5	OH	OH	=O	OH	OH

'*' denotes a putative assignment not confirmed by full NMR characterization.



	Obelmycin C	A447D	A447C	Cosmomycin C (**1**)	(**A**)
Aglycone					
1 H	-	dd, J = 7.2, 1.1 Hz 7.84	d, J = 7.4 Hz 7.89	d, J = 7.2 Hz 7.91	d, J = 7.2 Hz 7.91
2H	d, J=10 Hz 7.31	dd, J = 7.4, 7.2 Hz 7.65	dd, J = 8.3, 7.4 Hz 7.7	t, J= 8.4, 7.8 Hz 7.72	t, J= 8.4, 7.8 Hz 7.72
3H	d, J=10 Hz 7.29	dd, J = 7.2, 1.1 Hz 7.25	dd, J = 8.3 Hz 7.3	d, J = 8.4 Hz 7.33	d, J = 8.4 Hz 7.33
7H	5.15 br d	5.15 m	5.14 m	5.15 br d 3 Hz	5.16 br d, J = 3 Hz
8 Ha	2.2	NR	NR	NR	NR
8 Hb	2.2	NR	NR	NR	NR
10 H	5.02 s	5.02 s	5.01 s	br s (2 protons) 5.02	4.92 s (2 protons)
13 Ha	NR	NR	NR	NR	NR

13 Hb	NR		NR		NR		NR		NR	
14 CH3	1.1	t, J= 7 Hz	1.1	t, J = 7.4 Hz	1.1	t, J = 7.5 Hz	1.09	t, J=7.2 Hz		
C7 Trisaccharide										
1' H	5.48	d, J = 3.5 Hz	5.48	dd, J = 3.4 Hz	5.5	d, J = 3.6 Hz	5.47	m (2 protons)	5.49	m (2 protons)
2' H	NR		NR		NR		NR		NR	
3' H	NR		NR		NR		NR		NR	
4' H	3.73	br s	3.73	br s	3.78	br s	3.78	s	3.78	m
5' H	4.06	q, J = 7 Hz	3.99	q, J = 6.4 Hz	4	q, J = 6.8 Hz	4.06	q, J= 6.6 Hz	4.06	q, J = 6.6 Hz
6'-CH3	1.27	d, J= 7 Hz	1.28	dd, J = 6.4 Hz	1.27	d, J = 6.8 Hz	NR		NR	
3'-N(CH3)2	2.16	s	2.15	s	2.16	s	2.14	s (6 protons)	2.14	s (6 protons)
1" H	5.03		5.04	br s	4.95	br s	5.02	br s, 2 protons	5.03	br d, J = 3 Hz
2" H	NR		NR		NR		NR		NR	
3" H	3.98	br d	NR		NR		3.99	d, J= 10.2 Hz	4	br m
4" H	3.56	br s	4.06	m	3.47	br s	3.48	m	3.56	s
5" H	4.53	q, J = 7 Hz	4.52	q, J = 6.3 Hz	4.44	q, J = 6.7 Hz	4.52	q, J= 6.6 Hz	4.51	q, J = 6.6 Hz
6" CH3	1.14	q, J = 7 Hz	1.15	d, J = 6.3 Hz	1.07	d, J = 6.7 Hz	1.14	d, J=6.6 Hz	1.14	d, J = 6.6 Hz
1'" H	4.85	br s	4.85	br s	4.82	d, J = 3 Hz	4.85	s	4.85	br s
2'" H	NR		NR		NR		NR		NR	
3'" H	NR		NR		NR		NR		NR	
4'" H	3.66	br s	3.66	br s	3.58	br s	3.65	m	3.65	m
5'" H	4.06	q, J = 7 Hz	4.22	q, J = 6.7 Hz	4.06	q, J = 6.8 Hz	4.06	q, J= 6.6 Hz	4.06	q, J = 6.6 Hz
6'" CH3	1.16	q, J = 7 Hz	1.22	d, J = 6.7 Hz	1.15	d, J = 6.8 Hz	1.16	d, J=6.6 Hz	1.16	d, J = 6.6 Hz
C10 Trisaccharide										
1' H	5.44	d, J = 3.5 Hz	5.46	d, J = 3.5 Hz	5.46	d, J = 3.3 Hz	5.47	m (2 protons)	5.49	m (2 protons)
2' H	NR		NR		NR		NR		NR	
3' H	NR		NR		NR		NR		NR	
4' H	3.73	br s	3.73	br s	3.72	br s	3.72	br s (2 proton)	3.72	br s (2 protons)
5' H	3.88	q, J = 7 Hz	3.88	q, J = 6.7 Hz	3.87	q, J = 6.6 Hz	3.87	q, J= 6.6 Hz	3.9	q, J=6.6 Hz
6' CH3	1.24	d, J = 7 Hz	1.24	d, J = 6.7 Hz	1.23	d, J = 6.6 Hz	NR		NR	
3'-N(CH3)2	2.16	s	2.17	s	2.16	s	2.16	s (6 protons)	2.17	s (6 protons)
1" H	4.93	br s	4.92	br s	4.92	br s	4.92	s (2 protons)	4.92	s (2 protons)
2"	NR		NR		NR		NR		NR	
3" CH3	1.7, 2.0		NR		NR		NR		NR	
4" H	3.45	br s	3.55	br s	3.44	br s	3.45	s	3.45	s
5" H	4.4	q, J = 7 Hz	4.44	q, J = 6.3 Hz	4.39	q, J = 6.3 Hz	4.4	q, J= 6.6 Hz	4.4	q, J = 6.6 Hz

6" CH3	1.07	d, J = 7 Hz	1.09	d, J = 6.3 Hz	1.06	d, J = 6.3 Hz	1.06	d, J=6.6 Hz	1.06	d, J = 6.6 Hz
1'" H	4.81	d, J = 3.5 Hz	5.03	t, J = 5.3 Hz	4.8	d, J = 3.2 Hz	4.8	doublet J= 3.6 Hz	4.81	d, J = 3.6 Hz
2'" H	NR		NR		NR		NR		NR	
3'" H	NR		2.51	ddd, J = 16, 6.7, 5.3 Hz	NR		NR			
			2.42	ddd, J = 16, 9.5, 5.3 Hz						
4'" H	3.58	br s			3.58	br s	3.57	m	3.58	s
5'" H	4.22	q, J = 7 Hz	4.33	q, J = 6.6 Hz	4.07	q, J = 6.8 Hz	4.21	q, J= 6.6 Hz	4.21	q, J = 6.6 Hz
6'" CH3	1.22	d, J = 7 Hz	1.27	d, J = 6.6 Hz	1.17	d, J = 6.6 Hz	1.21	d, J=6.6 Hz	1.21	d, J = 6.6 Hz

All spectra were recorded in $CDCl_3$ and spectra for this study were referenced to residual chloroform (δH 7.26).

Table S2. Comparison of *Streptomyces* sp. CNT-302 and *S. olindensis* BGCs

Gene no. on CNT-302 ctg13	Gene product in CNT-302 (smcog)	Homologue in *S. olindensis*	Identity	Accession no.
91	glucose-1-phosphate thymidylyltransferase	glucose-1-phosphate thymidylyltransferase	97%	KDN80037.1
92	acyl carrier protein	actinorhodin polyketide synthase ACP	88%	WP_031119660.1
93	SARP family transcriptional regulator	transcriptional regulator	95%	KDN80039.1
94	methyltransferase	methyltransferase	92%	WP_037757689.1
95	unannotated	ester cyclase	95%	KDN80041.1
96	O-methyltransferase	methyltransferase	88%	KDN80042.1
97	alpha/beta hydrolase fold protein	short chain dehydrogenase	39%	WP_037760089.1
98	methyltransferase	SAM-dependent methyltransferase	89%	KDN80044.1
99	mono-oxygenase, FAD binding	FAD-dependent oxidoreductase	94%	WP_037757693.1
100	oxidoreductase	oxidoreductase	91%	WP_037757695.1
101	CYP450	hypothetical protein	74%	WP_037757696.1
102	glycosyl transferase	glycosyl transferase	94%	WP_037757699.1
103	aminotransferase	lipopolysaccharide biosynthesis protein Rfbh	95%	WP_037757701.1
104	short chain dehydrogenase	ketoacyl reductase	93%	WP_037757702.1
105	cyclase/dehydrase	actinorhodin polyketide cyclase	90%	WP_037757704.1
106	short chain dehydrogenase	ketoreductase	50%	WP_037758759.1
107	hypothetical	hydroxylacyl-CoA dehydrogenase	86%	WP_051648278.1

108	hypothetical	cyclase	94%	WP_037757865.1
109	methyltransferase	SAM-dependent methyltransferase	53%	KDN80044.1
110	aminotransferase	daunorubicin biosynthesis sensory transduction protein DnrJ	95%	WP_037757705.1
111	hypothetical	NDP-hexose 2,3-dehydratase	88%	WP_051648283.1
112	NAD-dependent epimerase/dehydratase	NAD-dependent epimerase	82%	KDN80058.1
113	hypothetical	dTDP-4-dehydrorhamnose 3,5-epimerase	89%	WP_037757707.1
114	glycosyl transferase	CosK	95%	ABC00725.1
115	PadR transcriptional regulator	CosS	95%	ABC00736.1
116	ornithine cyclodeaminase	hypothetical protein	81%	WP_037757710.1
117	hypothetical	CosX	91%	ABC00737.1
118	Beta-Ketoacyl Synthase (T2pks)	Beta-ACP synthase	94%	WP_037757712.1
119	Beta-Ketoacyl Synthase (clf)	Beta-ketoacyl synthase	90%	WP_037757715.1
120	3-oxoacyl ACP synthase	3-oxoacyl ACP synthase	91%	WP_037757720.1
121	malonyl-CoA ACP transacylase	CosF	83%	ABC00730.1
122	cytochrome CYP450	cytochrome CYP450	83%	WP_037757725.1
123	glycosyl transferase	glycosyl transferase family 28	94%	WP_037757728.1
124	NAD dependent epimerase/dehydratase	dTDP-glucose 4,6-dehydratase	94%	WP_037757730.1
125	ABC transporter ATP binding protein	CosI	94%	ABC00731.1
126	ABC-2 transporter	multidrug ABC transporter permease	97%	WP_037757733.1
127	hypothetical	glutathione peroxidase	87%	WP_037757871.1

128	ABC transporter ATP binding protein	daunorubicin resistance protein DrrC	94%	KDN80074.1
129	hypothetical	hypothetical protein DF19_01985	77%	KDN77779.1
130	PadR transcriptional regulator	transcriptional regulator	76%	KDN77778.1
131	hypothetical	membrane protein	51%	WP_031115629.1
132	hypothetical	ribonuclease BN	67%	KDN74392.1
133	hypothetical	membrane protein	86%	KDN73505.1
134	hypothetical	hypothetical protein	42%	WP_037764046.1
135	extracytoplasmic-function sigma-70 factor	RNA polymerase sigma factor	64%	KDN79303.1
136	hypothetical	hypothetical protein DF19_30470	47%	KDN79723.1
137	hypothetical	oxidoreductase	60%	KDN78332.1
138	hypothetical	hypothetical protein DF19_13800	31%	KDN77257.1
139	hypothetical	antitermination regulator	40%	WP_037761378.1
140	Sensor histidine kinase	histidine kinase	62%	WP_037761461.1
141	hypothetical	ATPase	43%	KDN79305.1
142	hypothetical	metal ABC transporter substrate-binding protein	52%	KDN79304.1

REFERENCES

(1) Kelso, C.; Rojas, J. D.; Furlan, R. L.; Padilla, G.; Beck, J. L. *Eur J Mass Spectrom* **2009,** *15*, 73-81.
(2) Johdo, O.; Watanabe, Y.; Ishikura, T.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T. *J Antibiot (Tokyo)* **1991,** *44*, 1130-1140
(3) Shimosaka, A.; Hayakawa, Y.; Nakagawa, M.; Furihata, K.; Seto, H.; Otake, N *Antibiot (Tokyo)* **1987,** *40*, 116-121.
(4) J. Ying, F. Delaglio, D.A. Torchia, and A. Bax. *J Biomol NMR*, in press **2016.**
(5) F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Baxz. *J. Biomol. NMR*, **1995**, *6*, 277-293.
(6) Johdo, O.; Watanabe, Y.; Ishikura, T.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T. *J Antibiot (Tokyo)* **1991**, *44*, 1121-1129.
(7) Johdo, O.; Tone, H.; Okamoto, R.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T., Anthracycline metabolites from Streptomyces violaceus A262. V. New anthracycline alldimycin A: a minor component isolated from obelmycin beer. *J Antibiot (Tokyo)* **1991**, *44*, 1160-1164.
(8) Johdo, O.; Tone, H.; Okamoto, R.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T. *J Antibiot (Tokyo)* **1991**, *44*, 1155-1159.
(9) Bandeira, N.; Tsur, D.; Frank, A.; Pevzner, P. A. *Proc Natl Acad Sci U S A* **2007,** *104*, 6140-6145.
(10) Nguyen, D. D.; Wu, C. H.; Moree, W. J.; Lamsa, A.; Medema, M. H.; Zhao, X.; Gavilan, R. G.; Aparicio, M.; Atencio, L.; Jackson, C.; Ballesteros, J.; Sanchez, J.; Watrous, J. D.; Phelan, V. V.; van de Wiel, C.; Kersten, R. D.; Mehnaz, S.; De Mot, R.; Shank, E. A.; Charusanti, P.; Nagarajan, H.; Duggan, B. M.; Moore, B. S.; Bandeira, N.; Palsson, B. O.; Pogliano, K.; Gutierrez, M.; Dorrestein, P. C. *Proc Natl Acad Sci U S A* **2013,** *110*, E2611-2620.
(11) Crüsesmann, M.; O'Neill, E.; Larson, C.; Melnik, A.; Floros, D.; da Silva, R.; Jensen, P.; Dorrestein, P.; Moore, BS. *Journal of Natural Products* **2016**, epub November 11, 2016.
(12) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res* **2003**, *13*, 2498-2504.

Chapter 3 section 3.3, in full, is a reprint of materials as it appears in "PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete" in *Journal of Natural Products*, 2017, Charles Bradford Larson, Max Crusman, Bradley S. Moore. The dissertation author was the primary investigator and author of this paper. C.B.L and B.S.M. designed research; C.B.L. performed bioinformatic, genetic, mass spectrometry and NMR experiments; C.B.L. and M.C. generated molecular networks; C.B.L., M.C. and B.S.M. analyzed data and wrote the manuscript.

Chapter 4: Marine Actinobacteria *Salinispora pacifica* Biosynthetic Gene Cluster Produces Phosphonate Degradation Products of Glyphosate

4.1 Chapter 4 Abstract:

Phosphonates are a diverse family of chemicals, both man-made and natural, containing a phosphorous-carbon bond which have a wide range of functions and biological activities. Examples of commercially important natural product phosphonates include the antibiotic fosfomycin used to treat urinary tract infections, the antimalarial drug FR-900098 and bialaphos, a potent herbicide. Additionally, the most widely used herbicide on the planet is a synthetic phosphonate, glyphosate. In this chapter, I describe the discovery and characterization of an orphan phosphonate biosynthetic gene cluster (BGC) from the bacterium *Salinispora pacifica* strain CNS-865. Upon introduction to a heterologous host, this BGC was found to produce a suite of small molecular weight phosphonate compounds including aminomethylphosphonate (AMPA), 2-hydroxyethyl-phosphonate (2HEP), and *N*-acetyl-aminomethylphosphonate (na-AMPA). Because two of these compounds, AMPA and na-AMPA, are used as environmental markers of glyphosate contamination, this study may have implications for the environmental study of the most widely used herbicide on the planet.

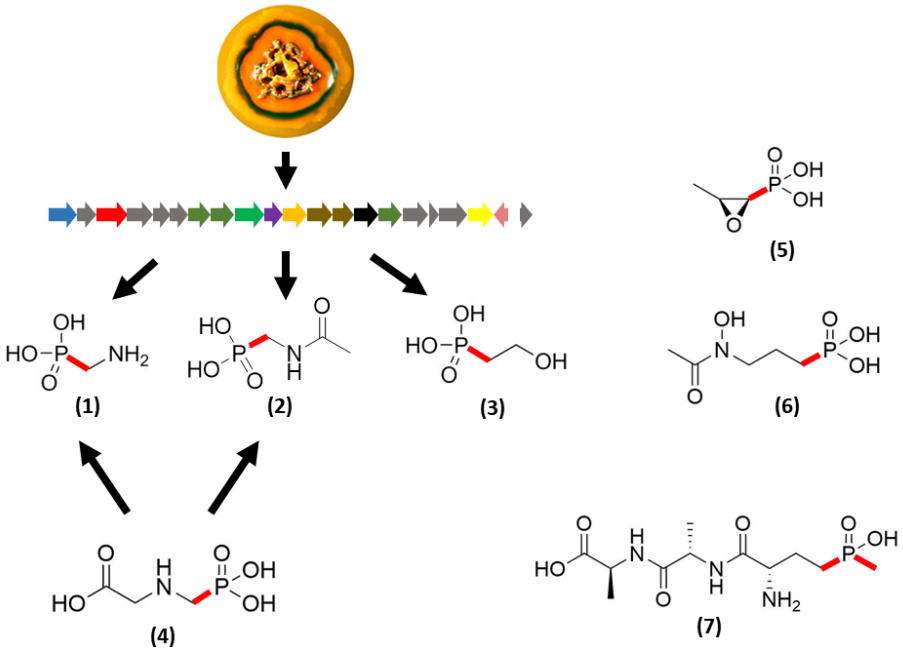


Figure 4.1: Depiction of a *Salinispora* colony, the *S. pacifica* BGC and its small molecule products na-AMPA (**1**), AMPA (**2**), 2-HEP (**3**). Glyphosate (**4**) has been demonstrated to decompose into **1** and **2** in the environment. Also shown are the bioactive phosphonates fosfsomycin (**5**), FR-900098 (**6**), and bialaphos (**7**). Phosphonate bonds are highlighted in red.

4.2 Introduction

Phosphonates are molecules containing a carbon-phosphorous bond, with the general chemical formula C−PO(OH)$_2$ or C−PO(OR)$_2$ where R is an alkyl or aryl functional group.[1] These compounds form a diverse family of molecules with many useful activities, including antibiotics (e. g., fosfomycin and fosmidomycin), immunosuppressants and bone disorders treatments (e. g., alendronate and zolendronate), antivirals (tenofovir), herbicides (glyphosate and phosphinothricin), chelators (methylenediphosphonic acid), deadly nerve agents (sarin and VX),

compounds for water treatment (aminotrimethylene phosphonic acid and hydroxy-phosphonoacetic acid), and carriers for radionuclides used in cancer therapy (e. g., phosphonate functionalized polymers like *N*,*N*′,*N*′-trimethylenephosphonate-polyethyleneimine). This variety of potent activities is, in many cases, a result of the stable C-P bond in these molecules mimicking the phosphate ester bond common in many biological substrates. Some examples of phosphonate molecules can be seen in Figure 1.

4.2.1 Phosphorus in Nature and the Marine Phosphorus Cycle

Phosphorus is essential to the structure and function of all life forms, forming the backbone of both DNA, RNA, and the standard energy currency of cells, ATP. As such, phosphorous is an obligate requirement for the growth of all organisms. In marine environments, phosphorous can control the species distribution and ecosystem structure,[2–4] and has been demonstrated to be the limiting nutrient in certain areas such as the eastern Mediterranean Sea,[5,6] the Sargasso Sea,[7] and areas with large freshwater or agricultural runoff inputs.[8,9] It is estimated that phytoplankton consume 1.5-2.5 gigatons of phosphorous annually.[10] The pool of available phosphorus in the oceans consists of both dissolved and particulate forms, which can be inorganic (such as orthophosphate, pyrophosphate, polyphosphate) or organic (such as P-esters, P-diesters, and phosphonates).[11] Phosphonates are a major component of high molecular-weight dissolved organic phosphorus (DOP) as determined by solid-state ^{31}P NMR.[12] Phosphonates are especially stable molecules and do not degrade as easily as other forms of organic phosphorus, and they form a large component of marine

sediments (around 25% of total organic P).[13] When compared to the proportion of phosphonates in living organisms and sinking particulates (1% and 3%, respectively), these data suggests that phosphonates are a major sink of P in the marine environment.[14]

Although initially described as synthetic molecules, phosphonates have been recognized as natural products for over 60 years.[15] Both marine and aquatic environments have been a rich source of these interesting metabolites, with examples like 2-aminoethylphosphonic acid (2AEP) found in amoeba plasma membranes,[16] complex phosphonoglycocerebrosides found in the sea slug *Aplysia kurodai,*[17,18] *N*-methylated phosphonoglycocerebrosides found in sea snails,[19] and glycoproteins in the sea anemone *Metridium dianthus* decorated with derivatives of 2AEP. Eggs of the freshwater snail *Helisoma* also contain 2AEP as a component of phosphonoglycoproteins,[20] and similarly the marine snail *Volvarina rubella* contains 2AEP and its N-methylated derivative as components of polysaccharides. Recently, a mucin isolated from the jellyfish *Aurelia aurita* was shown to contain *O*-glycosylated peptide with a phospohonate moiety.[21] Marine sponges have also been found to incorporate phosphonate moieties into polyketides.[22]

Since the advent of late-generation sequencing technology and the large amount of data generated from whole genome and metagenomic sequence projects, genome mining studies have revealed new sources of phosphonates in the marine environment. Using genomic data, methylphosphonic acid was identified in the exopolysaccharides of the marine bacteria *Nitrosopumilus maritimus*. Although the native function of these metabolites remains unknown, these molecules have been shown to be a major source of methane production when acted upon by C-P lyases.[23] Genome mining also revealed

the biosynthesis genes for 2AEP in the reef building coral *Acropora digitifera*, although production has not been confirmed by chemical analysis.[24]

The entry for phosphonate biosynthesis from primary metabolism is the enzyme phosphoenolpyruvate mutase (PepM). The PepM enzyme catalyzes the reaction that forms the carbon-phosphorous bond in phosphonopyruvate from the substrate phosphoenolpyruvate. However, the equilibrium for this reaction favors phosphoenolpyruvate by a factor of at least 500.[25] In order for the biosynthesis of phosphonates to proceed, this unfavorable equilibrium must be overcome. Nature has achieved this by coupling P-C bond formation with a variety of irreversible reactions, including decarboxylation, transamination, and aldol reactions. The most common strategy is decarboxylation, accomplished by phosphonopyruvate decarboxylase (PpD), forming the phosphonate building block phosphonoacetaldehyde. This molecule is used as the starting material to biosynthesize a variety of natural phosphonate, including dehydrophos, fosfomycin, the rhizocticins, and the plumbemycins.[26] A depiction of this reaction scheme is shown in Figure 2A. It is worth noting that *Streptomyces hygroscopicus*, *Streptomyces viridochromogenes*, and *Kitasatospora phosalacinea* harbor an exception to this PepM paradigm, utilizing a cobalamin-dependent methyl transferase to install a methyl group to a phosphinate nucleophile forming a P-C bond (Figure 2B).[27,28]

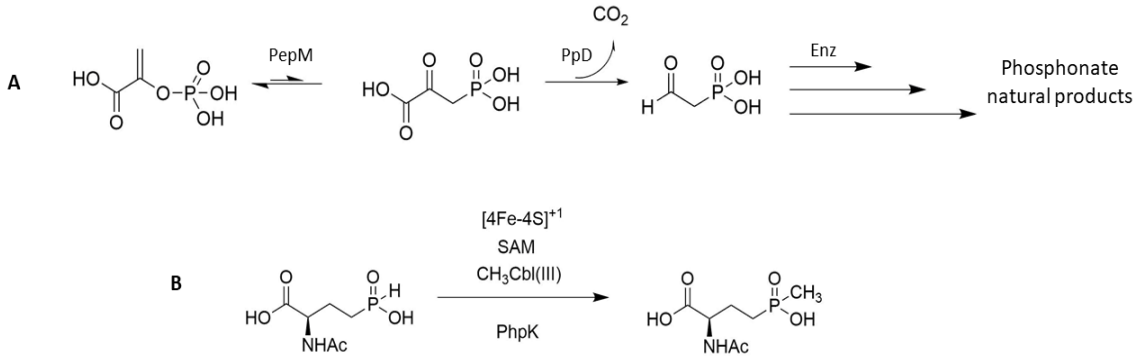


Figure 4.2: Biosynthetic scheme for P-C bond installation. (**A**) Canonical PepM/PpD P-C bond formation (**B**) Cobalamin-dependent methyl transferase P-C bond formation

4.2.2 Anthropogenic Phosphonate Contamination in the Environment

One of the most popular and widely used phosphonates is glyphosate (*N*-(phosphonomethyl)glycine). It was first synthesized and tested for herbicide use in 1970 and entered the market in 1974 as a non-selective herbicide.[29] Glyphosate inhibits 5-enolpyruvyl-shikimate-3-phosphate synthase, a key enzyme within the shikimate pathway necessary for the synthesis of aromatic amino acids.[30] Since its introduction, glyphosate has become one of the world's most common herbicides with over 270 million pounds used in the US alone in 2012.[31] Because of its widespread use, there has been interest in studying its persistence and degradation in the environment.

Aminomethylphosphonic acid (AMPA) is the main metabolite of glyphosate degradation by both microbial and abiotic means. Acetylation of AMPA has also been demonstrated to occur in some soil microorganisms, and may be necessary to liberate P from the molecule for use in primary metabolism (Figure 1).[32,33] Phosphonate

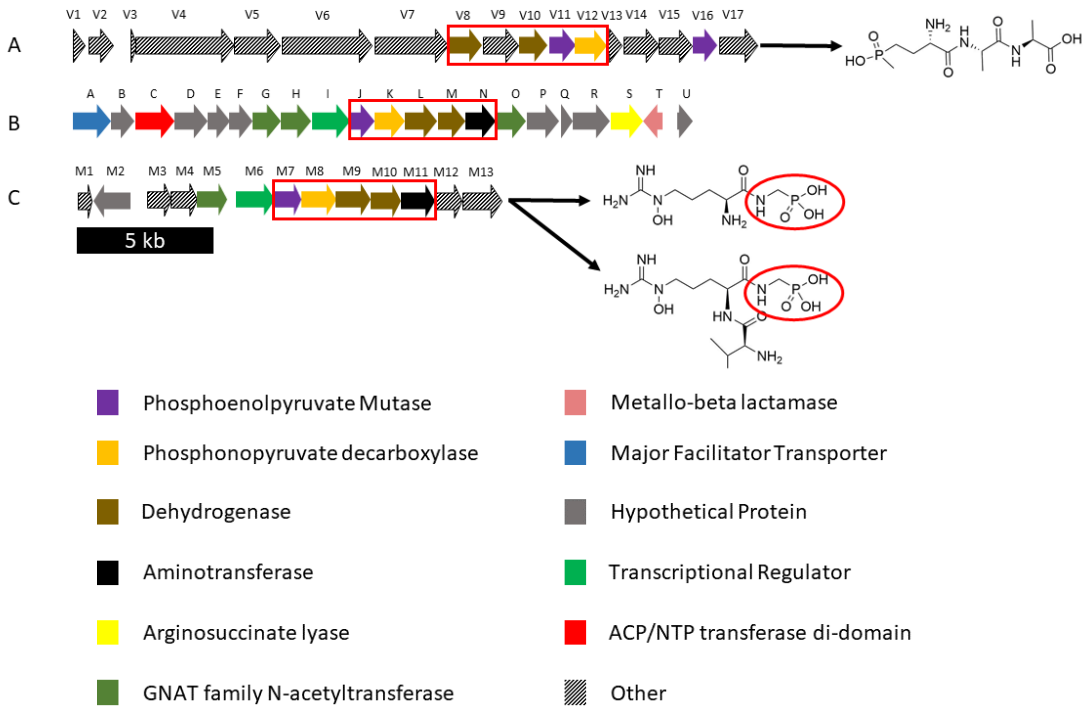
detergents also liberate AMPA when degraded, and it has been frequently detected in surface water, sediment and groundwater.[34–36] Several studies revealed that AMPA slightly effects human erythrocytes *in vitro*[37] but can cause DNA and chromosomal damage in fish.[38] Various studies have reported that glyphosate and AMPA are persistent in both terrestrial and aquatic environments.[39,40] The source of environmental AMPA has been assumed to be entirely anthropogenic until this time, as it has never been reported as a product of microbial fermentation or demonstrated to be the biosynthetic end-product of a BGC.

Therefore, because of their interesting bioactivities, notable contribution to the marine phosphorus cycle, and roles in environmental pollution monitoring, phosphonate BGCs are of great interest to the scientific community. *Salinispora* species are obligate marine actinomycetes with a unique lineage due to their evolution in the marine environment.[41] The genus is the source of many bioactive metabolites, including the anti-cancer drug salinisporamide, currently in phase II clinical trials for the treatment of multiple myeloma and phase III trials in patients with glioblastoma.[42,43] However, phosphonates have never before been reported from this actinomycete genus despite them being prevalent in terrestrial actinomycete bacteria. This report describes the search for phosphonate BGCs within sequenced *Salinispora* genomes, the heterologous expression of one such BGC, and the elucidation of the metabolites produced upon BGC heterologous expression.

4.3 Results

4.3.1 Identification of *Salinispora* PepM Sequences

To identify potential phosphonate BGCs, 119 *Salinispora* genomes from the Jensen strain collection[44] were queried for phosphoenolpyruvate mutase genes using the IMG genome browser (https://img.jgi.doe.gov/). This search returned two strains with homologous *pepM* genes, *S. arenicola* CNS-205 and *S. pacifica* CNS-863. Both strains were grown on various phosphorus-replete media, however no phosphonate chemistry was detected by ^{31}P NMR analysis or MS/MS networking. Because only *S. pacifica* CNS-863 possessed the canonical *pepM*/*ppD* pair, its phosphonate BGC was selected for gene cloning, heterologous expression and chemical analysis. Further bioinformatic investigation of this 24 kb phosphonate BGC revealed its homology to two clusters found in *Streptomyces monomycini* and *Streptomyces viridichromogenes* responsible for producing argolaphos A and B and phosphonithricin tripeptide (PTT), respectively (Figure 3). All of these strains contain the enzymes necessary for the biosynthesis of AMPA, with AMPA related enzymes highlighted in red boxes. Although PTT does not contain an AMPA moiety, various knockout mutants of this strain have been shown to accumulate AMPA in the culture media.[45] Notable differences in the clusters include the presence of an ATP-grasp ligase in the *S. monomycini* BGC, and non-ribosomal peptide synthetases in addition to a second *pepM* copy in the *S. viridichromogenes* BGC. A detailed description of each protein in the *S. pacifica* BGC can be found in Supplemental Table S4.1.



D

S. pacifica enzyme	*S. viridichromogenes* enzyme	*S. monomycini* enzyme	Enzyme Function
sppG	N/A	M5 (63%)	GCN5 family acetyltransferase
sppI	N/A	M6 (59%)	XRE family transcriptional regulator
sppJ	V11 (40%)	M7 (46%)	phosphoenolpyruvate mutase
sppK	V12 (42%)	M8 (59%)	phosphonopyruvate decarboxylase
sppL	V8 (32%)	M9 (51%)	alcohol dehydrogenase
sppM	V10 (53%)	M10 (71%)	3-phosphoglycerate dehydrogenase
sppN	V9 (26%)	M11 (55%)	aspartate aminotransferase

Figure 4.3: Depiction of phosphonate BGCs. **(A)** *S. viridichromogenes* phosphonate BGC responsible for producing phosphinothricin tripeptide. **(B)** *Salinispora pacifica* phosphonate BGC. **(C)** *S. monomycini* phosphonate BGC responsible for producing argolaphos A and B. AMPA moieties are circled in red. **(D)** Summary of the protein homologs shared between the BGCs (% amino acid identity).

4.3.2 Heterologous Expression of *S. pacifica* Phosphonate BGC

To investigate the chemistry and enzymology of this uncharacterized *Salinispora* phosphonate BGC, it was amplified by PCR in 6 overlapping fragments to obtain large

amounts of genetic material, ensuring efficient assembly of the BGC. The fragments were assembled by *Saccharomyces cerevisiae* into the self-replicating plasmid pCAP03 (Supplemental Figures S4.1 and S4.2). This plasmid was then integrated into the genome of *Streptomyces coelicolor* M1152. Fermentation of the heterologous host in both solid and liquid media yielded three peaks in the ^{31}P NMR spectrum with chemical shifts consistent with phosphonates (Figure 4). It should be noted, however, that production was not consistent in liquid media, and thus agar plates were used for secondary metabolite expression for the rest of this study.

4.3.3 Purification of Compounds 1-3

Purification of the three compounds of interest, compounds **1**–**3**, was challenging due to the chemical properties of small phosphonate molecules. Lacking UV chromophores, these compounds are invisible to the spectrophotometers generally used to detect secondary metabolites during initial purification procedures. Additionally, their extreme hydrophilicity complicated standard chromatography methods to separate these compounds from the abundance of salts, organic phosphates, and other polar molecules present as the main fermentation components in the crude extracts. Although phosphonates are detectible by mass spectrometry, the usual separation methods used in LC/MS systems fail to retain these molecules, and the high salt content crude extracts will generate noisy spectra with fidelity too low for untargeted analysis in addition to damaging the detectors in sensitive modern machines.

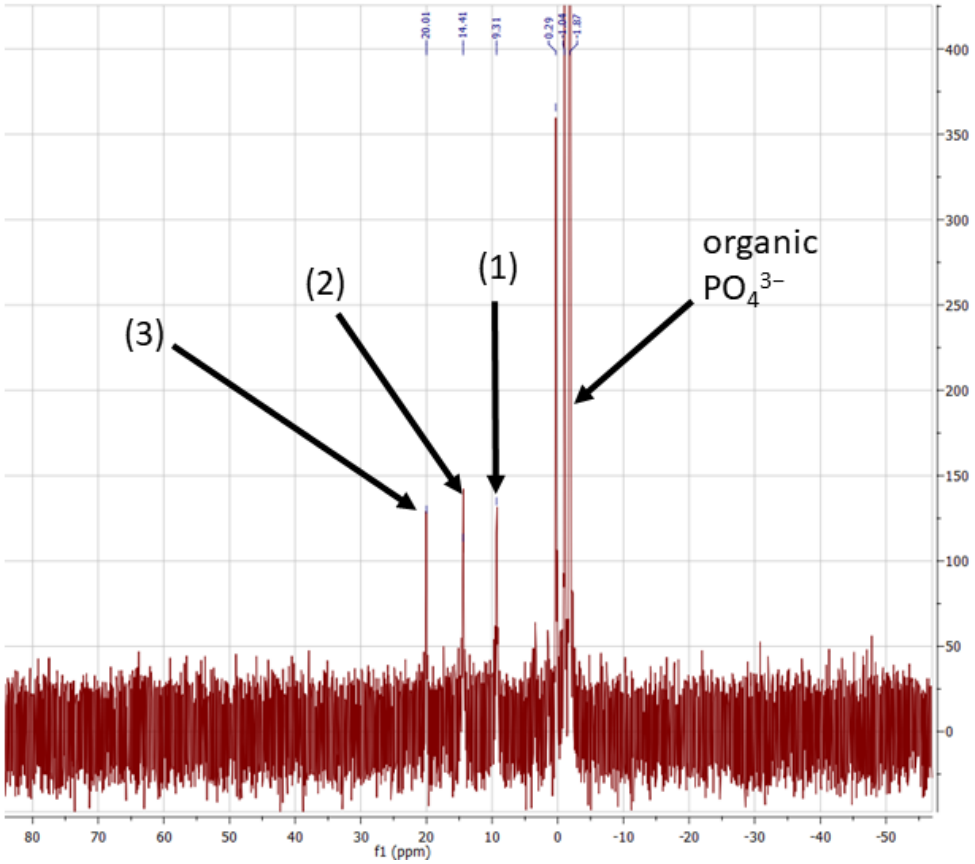


Figure 4.4: Crude preparation of heterologous host fermentation. Analysis was accomplished by ^{31}P NMR in D_2O. Three distinct chemical shifts can be seen at 9.3 ppm (compound **1**), 14.4 ppm (compound **2**), and 20.0 ppm (compound **3**). This region of the spectrum (chemical shifts above 7 ppm) is diagnostic for phosphonate phosphorus nuclei bonded to a carbon. Peak labels correspond with compounds **1**-**3** shown in Figure 4.1.

Because of the strong negative charge present in the phosphonate moiety, ion-exchange chromatography was determined to be the best method to de-salt crude extracts and enrich for our target metabolites. Several resins were investigated for their ability to retain the molecules responsible for the ^{31}P signals observed in the crude extracts, but only AG-1x8 (-OH form) was able to bind the phosphonates produced by

the captured BGC. This anion-exchange step was able to remove the majority of salts and other positively charged or neutral polar contaminants in the crude extract. In order to remove any lingering hydrophobic contaminants in the preparation, the ^{31}P NMR positive fractions from the AG1x8 column were applied to a column of C18 resin, and the H_2O wash fraction was collected and dried. These successive rounds of ion exchange and standard reversed phase chromatography yielded an enriched, semi-pure mixture. Analysis of this semi-pure mixture by ^{1}H NMR revealed two peaks with a large *J*-coupling constant, another characteristic of phosphonate molecules (Figure 4.5). None of these peaks were present in the extracts of the negative control in which cultures of the empty host *S. coelicolor* M1152 processed in the same manner. The presence of a phosphorus in compounds associated with these two doublets was further confirmed by a broad band decoupling experiment, collapsing the peaks to singlets (Figure 4.6). Additionally, this experiment revealed that a multiplet at 3.6 ppm was also being split by a phosphorus nucleus and likely represented the third molecule identified in the crude ^{31}P spectrum. These hydrogen peaks were subsequently connected to their respective ^{31}P NMR signals using a 2D $^{31}P/^{1}H$ NMR experiment (Figure 4.7).

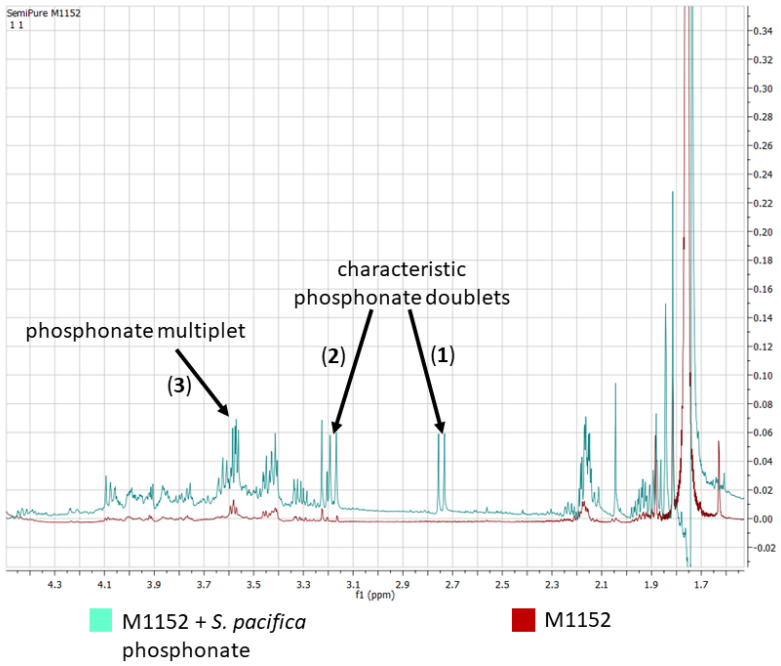


Figure 4.5: ^{1}H NMR spectra of both heterologous host and *S. coelicolor* M1152 extracts purified by chromatography. The doublets at 2.75 ppm (compound **1**) and 3.2 ppm (compound **2**) both exhibit the large coupling constant associated with protons split by a nearby phosphorus nucleus. Additionally, there is a multiplet at 3.55 ppm (compound **3**) in the heterologous host extract. None of these features are present in the cultures of the empty host M1152.

Although the semi-pure preparations allowed for the identification of ^{1}H and ^{31}P NMR signals produced by the phosphonate BGC metabolites, this information was insufficient to elucidate the chemical structure of these molecules. Analysis of this preparation by MS/MS experiments was still too noisy to positively identify the masses responsible for the diagnostic NMR signals, and it was determined further chromatographic purification was necessary. Hydrophilic Interaction Liquid Chromatography (HILIC) was an attractive option to separate impurities from our molecules of interest. Despite its long equilibration times (> 10 times the column volume of buffer is required to equilibrate) and the scant amount of our material that could be

injected due to the sensitive nature of the column, this method proved successful in purifying the target compounds. The process yielded trace amounts of a mixture of the three target phosphonates free from obvious contaminants when analyzed by ^{1}H NMR (Figure 4.8). This pure mixture was then subjected to MS/MS analysis to reveal mass/charge (*m/z*) ratios of 111, 153, and 125, with all of these target masses showing characteristic fragmentation patterns for phosphonate molecules (Figure 4.9). High-resolution MS was able to confirm the molecular weight and chemical formula of these compounds (Supplemental Figures S4.3-S4.5). The molecular formulas determined from the high-resolution MS suggested these metabolites may be the known compounds aminomethylphosphonate (AMPA) (compound **1**), *N*-acetyl aminomethylphosphonate (na-AMPA) (compound **2**), and 2-hydroxyethylphosphonate (2HEP) (compound **3**), respectively. Synthetic standards of these three molecules were obtained and used to confirm the identity of the heterologous host fermentation products by spiking extracts with these standards followed by NMR analysis (Figure 4.10, Supplemental Figure S4.6). Although only trace amounts of the pure phosphonate compounds were purified, NMR analysis show a production of approximately 4 mg/L of solid media (Supplemental Figure S4.7). Both AMPA and 2HEP (compounds **1** and **3**) have been isolated from bacterial mutants blocked for phosphonate biosynthesis,[45–47] however na-AMPA (compound **2**) has only been found in environmental samples and assumed to be the product of anthropogenic phosphonate degradation by microbes.[48]

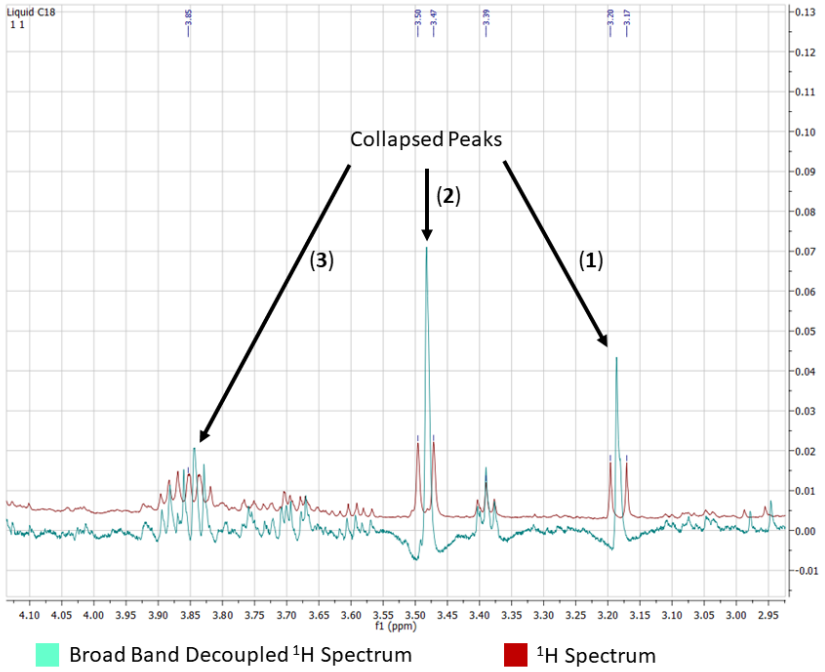


Figure 4.6: Broad band decoupling of the NMR spectrum of the enriched fraction of heterologous host extracts reveals the influence of a phosphorus nucleus on three peaks.

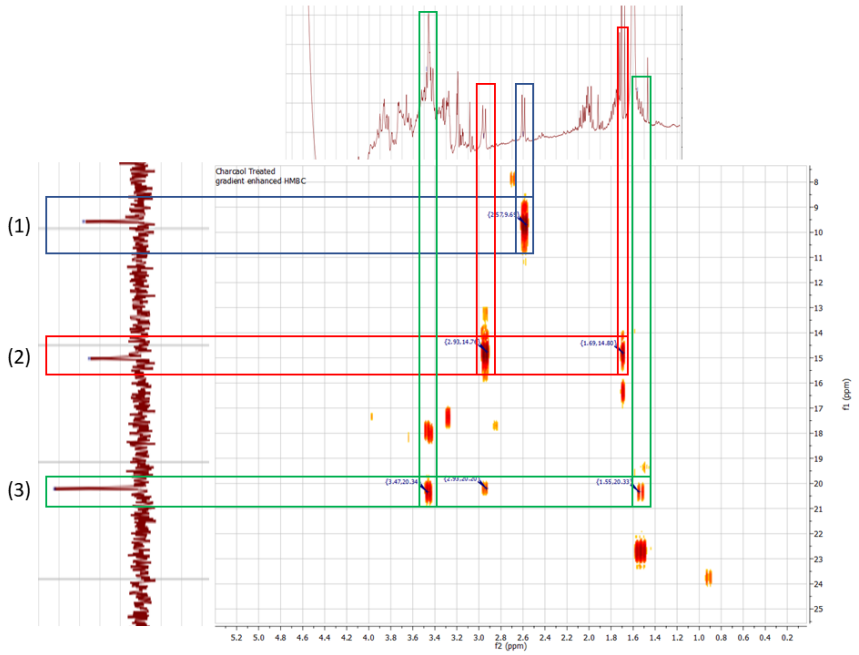


Figure 4.7: $^{1}H/^{31}P$ two-dimensional NMR spectrum (HMBC) of enriched preparation of heterologous host extracts. Correlations can be seen between the ^{31}P chemical shift at 9.6 ppm and the ^{1}H doublet at 2.5 ppm (compound **1**); the ^{31}P shift at 14.7 ppm and the ^{1}H shifts at 2.93 ppm and 1.69 ppm (compound **2**); and the ^{31}P shift at 20 ppm and ^{1}H shifts at 3.47 ppm and 1.55 ppm (compound **3**).

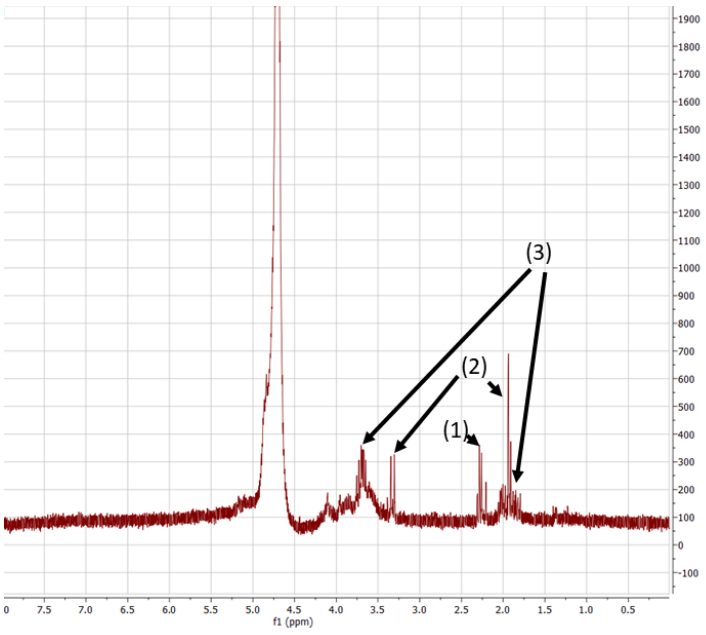


Figure 4.8: ^{1}H NMR analysis of HILIC purified products of fermentation. Based on the previous HMBC experiment, the doublet (compound **1**) correlates to the ^{31}P shift at 9.8 ppm; the doublet and singlet (compound **2**) correlates to the ^{31}P shift at 14.7 ppm; and the multiplets correlate to the ^{31}P shift at 20 ppm (compound **3**).

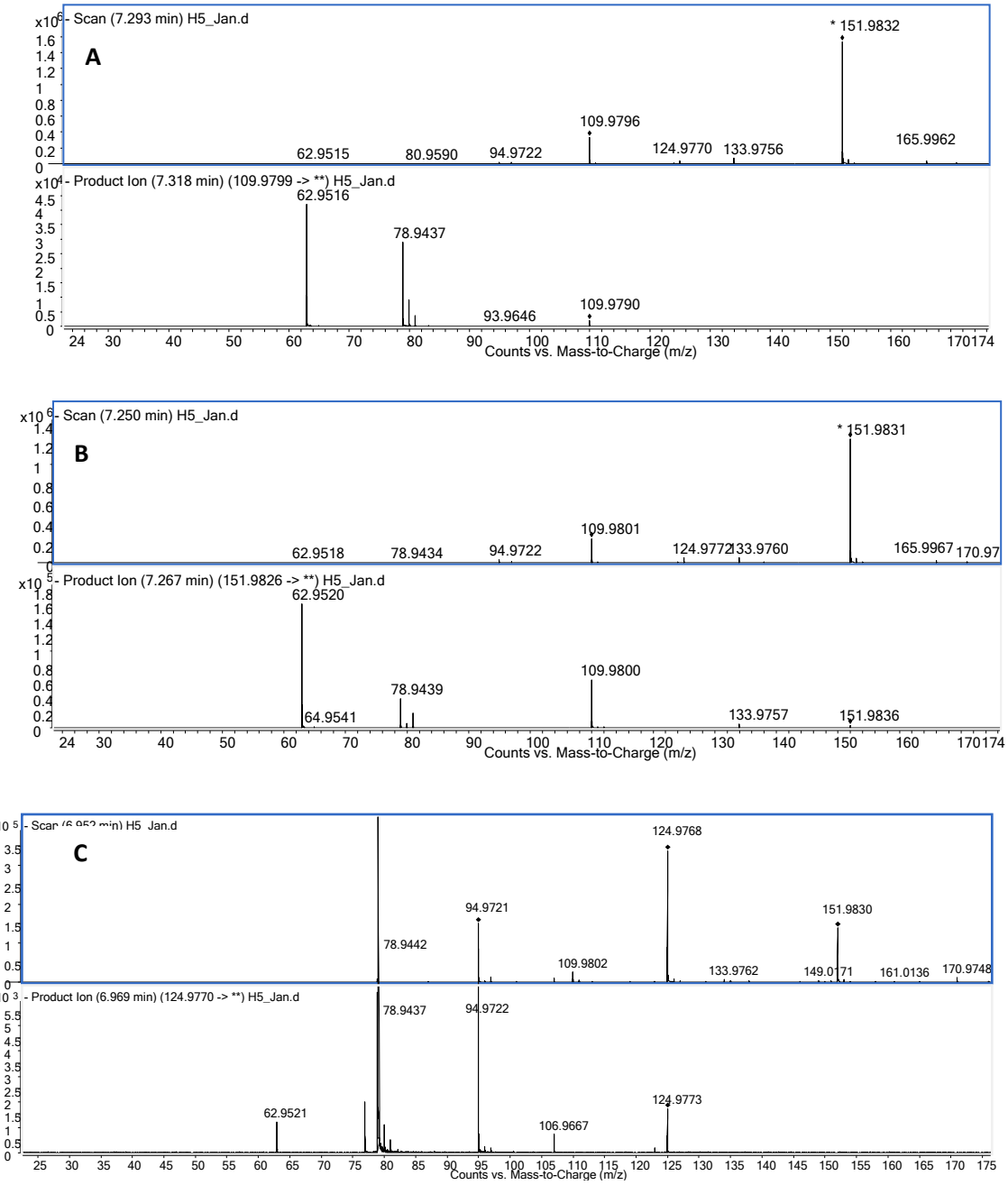


Figure 4.9: MS/MS analysis of HILIC purified metabolites. Characteristic fragments of phosphonate molecules can be seen at *m/z* =63, and 81 in all of the putative phosphonate metabolites. (**A**) MS/MS analysis of compound **1**. (**B**) MS/MS analysis of compound **2**. (**C**) MS/MS analysis of compound **3**.

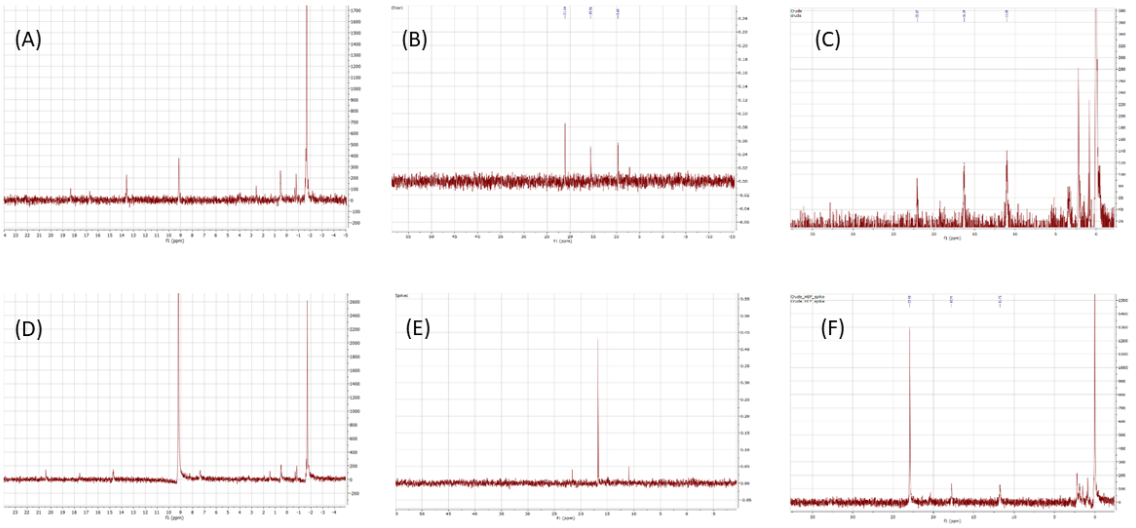


Figure 4.10: ^{31}P NMR analysis of heterologous host extracts spiked with synthetic phosphonate standards. (A-C) Heterologous host extracts, without synthetic standard added. (D) Extract spiked with synthetic AMPA (compound **1**) (E) Extract spiked with synthetic na-AMPA (**2**) (F) Extract spiked with synthetic 2HEP (compound **3**).

4.3.4 Cloning and Expression of Phosphonate *N*-acetyltransferases

One method for inactivating glyphosate is *N*-acetylation, as *N*-acetylglyphosate is non-herbicididal. There is currently only a single known *N*-acetyltransferase isolated from *Bacillus licheniformis* capable of performing this reaction.[49] This activity is of great commercial interest, as it represents an alternative strategy for glyphosate tolerance compared to modifying the enzymatic target of the herbicide, enolpyruvylshikimate-3-phosphate synthase. The *B. licheniformis* glyphosate acetyltransferase (GAT) is a 146 AA protein weighing 17 kDa, and shares between 30 and 64% AA identity to its closest

homologs in a generally uncharacterized family of N-acetyltransferases containing a conserved domain (Acetyltransf_1: pfam00583).[50] The native enzyme is an extremely poor catalyst for the acetylation reaction, with a k_{cat} of 5.3 min^{-1} and $Km_{,GPJ}$ of 1.3 mM. Despite exhaustive studies on the substrate specificity of the enzyme, its native substrate remains unknown.[51]

Interestingly, the *S. pacifica* phosphonate BGC contains three genes that code for putative GCN5-family acyltransferase type of enzymes (GNATs). Two of these enzymes, SppG and SppH, contain conserved domains (Acetyltransf_6: pfam13480), and a third enzyme SppO shares between 33 and 49% AA identity with other enzymes containing the same pfam13480 domain. Although these enzymes are found in the same phosphonate BGC, they share little homology amongst themselves or with GAT. A multiple sequence alignment shows that none of the enzymes share more than 29.3% AA identity (Supplemental Table S4.2), and there are no biochemically characterized enzymes with significant homology when analyzed by NCBI BLAST.

In order to explore the activity and the substrate specificity of these GNATs, the genes (*sppG*, *sppH*, and *sppO*) were cloned, heterologously expressed in *E. coli,* and purified (Supplemental Figure S4.8). Each enzyme was incubated overnight with the nitrogen containing phosphonates AMPA, 2-AEP, and glyphosate to probe the activity of these GNATs. All three enzymes were able to acetylate AMPA and 2-AEP (Figures 4.11 and 4.12), although none of the enzymes consumed all of the substrate *in vitro*. Furthermore, none of the cloned enzymes were able to acetylate glyphosate (data not shown).

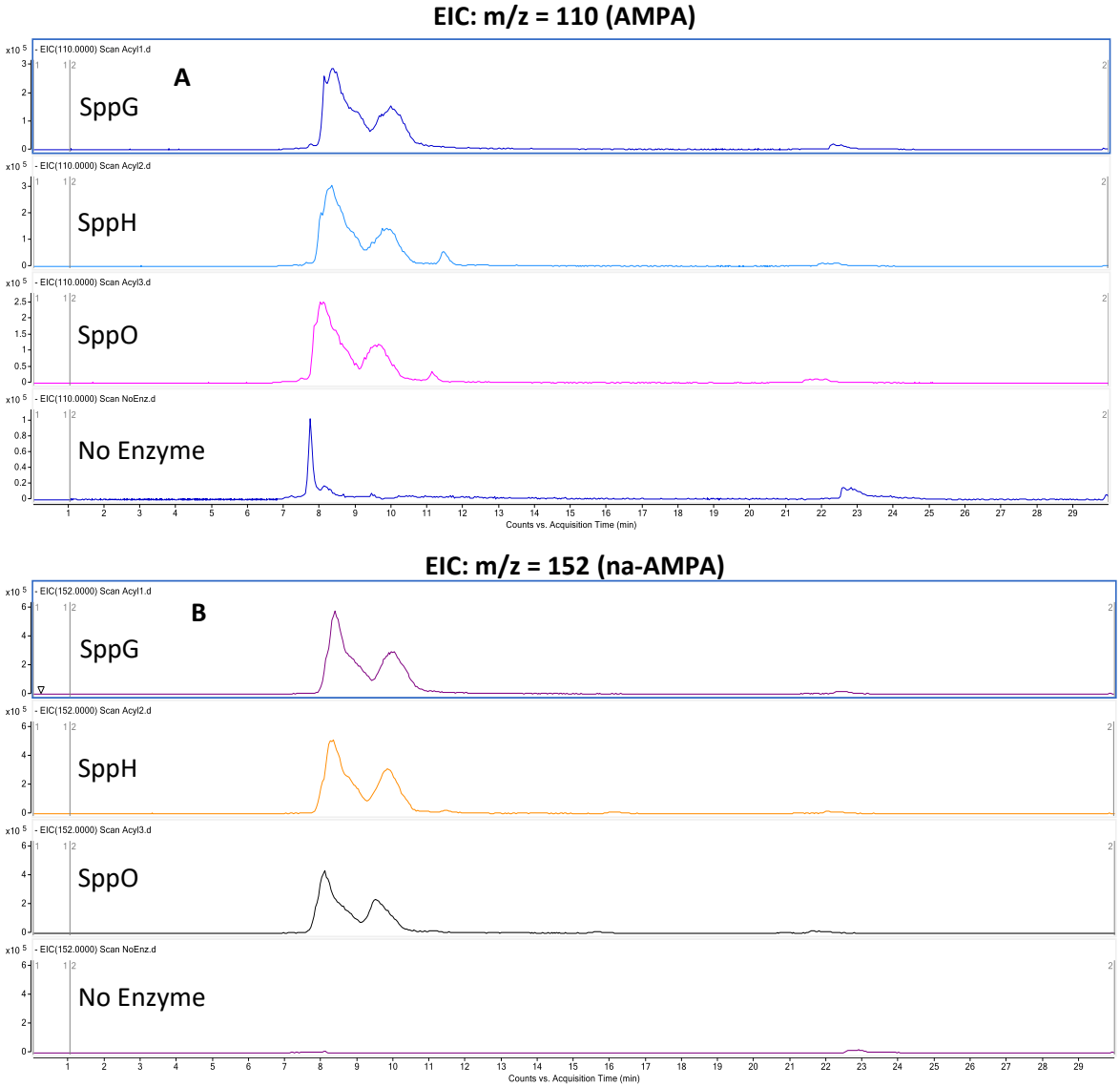


Figure 4.11: MS analysis of GNAT enzyme assays with AMPA substrate. (A) Extracted ion chromatograms for mass of substrate AMPA (*m/z* = 110) for all reactions including incubations with purified SppG, SppH, or SppO and reaction mixture without enzyme. (B) Extracted ion chromatograms for mass of putative enzyme product na-AMPA (*m/z* = 152) for all reactions including the purified enzymes and reaction mixture without enzyme.

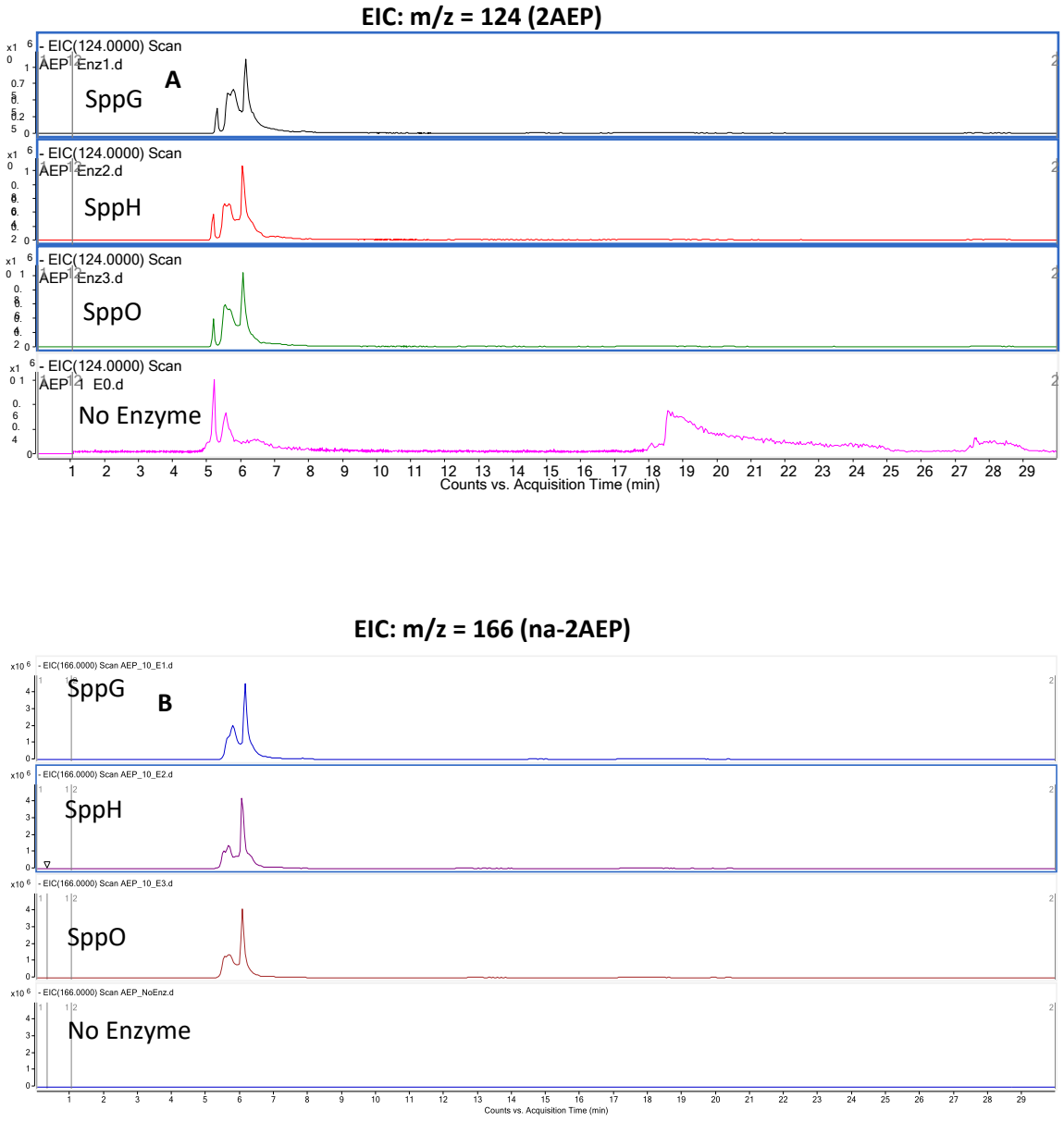


Figure 4.12: MS analysis of GNAT enzyme assays with 2-AEP substrate. (A) Extracted ion chromatograms for mass of substrate 2-AEP (m/z = 124) for all reactions including incubations with SppG, SppH, or SppO, and reaction mixture without enzyme. (B) Extracted ion chromatograms for mass of putative enzyme product na-AMPA (m/z = 166) for all reactions including purified enzymes and reaction mixture without enzyme.

4.3.5 Distribution of PepM (SppJ) Homologs in the Marine Environment

Finally, the distribution of PepM homologs in the marine environment was investigated. The SppJ sequence was used to query both marine metagenomes and the fully sequenced genomes of marine organisms in the MARDB database. Analysis by NCBI BLAST revealed over 500 homologs of the *S. pacifica* PepM with an expect value below 1.3×10^{-38} in the database, suggesting that this pathway and its homologs are widespread and could have a major impact on the ocean phosphorus cycle (Figure 4.13 and Supplemental Tables S4.3 and S4.4). In order to investigate the relationship between SppJ and its homologs found in the marine environment, the protein sequences were used to create a sequence similarity network using the Enzyme Similarity Tool (https://efi.igb.illinois.edu/). This algorithm generates a visual representation of the homologs found in the MARDB database using an all-versus-all comparison of the protein sequences.

This network analysis revealed that the *S. pacifica* PepM clusters closely with 50 other enzymes found in the marine environment (colored light blue in Figure 4.13), in addition to sharing some characteristics with 414 other sequences connected in the network (other nodes excluding red, black and gray). A legend detailing the proteins in each node can be found in Supplemental Tables S4.3 and S4.4. One might expect that the most closely related enzymes would be found in other *Salinispora* or actinomycete strains, but surprisingly the homologs are spread across an extremely diverse group of bacteria and archaea. These include sequences from the genera Clostridiales, Chlorflexi, Nitrospina, Streptomyces, Verrucosispora, Thaumarchaeota, Dehalococcoidaceae, Nitrosoarchaeum, Reichenbachiella, and Terasakiella. The

diversity of the PepM sources lends credence to the hypothesis that the biosynthesis of AMPA is common in the marine ecosystem.

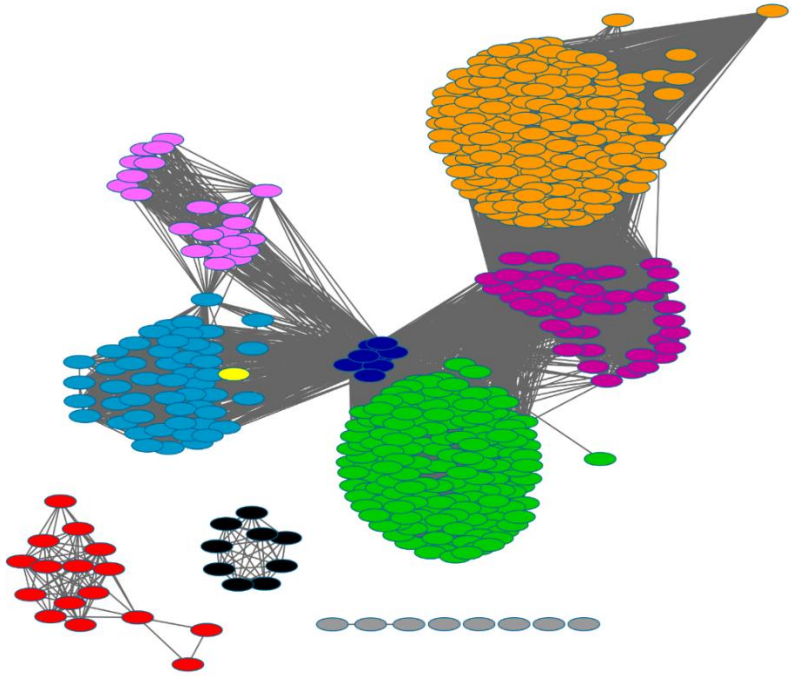


Figure 4.13: Sequence Similarity Network of *S. pacifica* PepM homologs found in marine genomes and metagenomes. Nodes represent SppJ homologs, and edges connecting these nodes represent homology shared between nodes. The node for the *S. Pacifica* PepM (SppJ) is highlighted in yellow.

4.4 Discussion

This study represents the first report of na-AMPA as a natural product, as this metabolite has only been recognized previously as a synthetic compound or as a catabolic breakdown product of herbicidal glyphosate. Additionally, the captured *S. pacifica* phosphonate cluster produces 2HEP and AMPA in similar titers to na-AMPA. Although both of these metabolites have been shown to be intermediates in

phosphonate biosynthesis by knocking out enzymes in phosphonate BGCs such as the phosphinothricin pathway from *S. viridichromogenes*,[45] this is the first report of these metabolites as apparent products of a biosynthetic pathway.

Comparisons to similar clusters in *Streptomyces monomycini* and *Streptomyces viridichromogenes* suggest that the genes *sppJ-sppL* are responsible for the formation of 2HEP (compound **3**). As is the case in *S. monomycini*, an unknown enzyme then generates hydroxymethyl phosphonate from 2HEP. This is accomplished by the V9 enzyme, a dioxygenase, in *S. viridichromogenes*, which has been swapped for an aminotransferase in both the *S. pacifica* and *S. monomycini* BGC. Although no close homologs of V9 are found in the *S. pacifica* genome, this reaction likely proceeds through one of the 23 other dioxygenases in the genome. The genes *sppM* and *sppN* then transform hydroxymethyl phosphonate to AMPA (compound **1**). Finally, from both bioinformatic comparison and *in vitro* experiments, it appears that the acetylation of AMPA is accomplished by any one of the *N*-acetyltransferase gense (*sppG*, *sppH*, or *sppO*) to yield na-AMPA (compound **2**). A putative biosynthetic pathway for these isolated metabolites is shown in Figure 4.14.

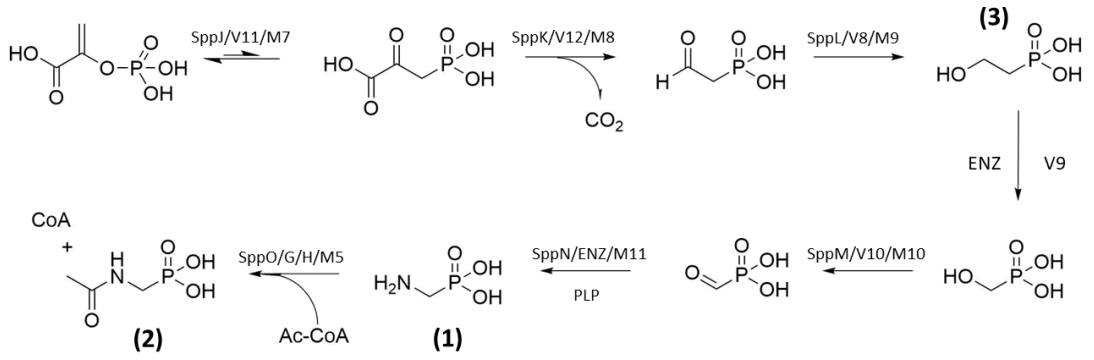


Figure 4.14: Putative biosynthetic pathway for the production of na-AMPA.

Although the biological function of these metabolites remains unknown and whether a more complex phosphonate is produced by *S. pacifica* CNS-865, the discovery of phosphonate producing enzyme machinery in a sediment dwelling ocean bacterium has potential wide reaching implications for environmental studies in the marine environment. There is great concern about terrestrial, aquatic, and marine contamination of the environment due to the large amounts of phosphonates used as pesticides agriculturally, in cities, and in residential areas.[36,52–54] Currently, AMPA is often used as a proxy for glyphosate and phosphonate detergent pollution. However, this study's biochemical and bioinformatic analysis suggests that in the marine environment, at least some of the AMPA detected is produced naturally in the ocean biosphere. The wide-spread presence of homologs of the *S. pacifica* PepM suggest that AMPA production may be a common occurrence. These results call into question the validity of using environmental AMPA as a proxy for glyphosate and detergent contamination, and also suggest that bacteria have evolved to not only naturally biosynthesize phosphonates like AMPA but also to degrade them.

Additionally, these results shed light on the role of phosphonate biosynthesis in the marine phosphorus cycle. Phosphonates represent a large fraction of both particulate and dissolved organic phosphorus in the marine environment and are stable molecules that can persist for long periods of time. This study has revealed at least one source of phosphonate production is a sedimentary actinomycete. Because phosphorus is often the limiting nutrient in marine ecosystems and specialized enzymes are required to liberate P from phosphonates, sequestration of phosphorus into phosphonates likely effects the local environment where these bacteria live and reproduce.

4.5 Methods/Materials

4.5.1 Genome Mining

The genomes of 119 *Salinispora* strains available for culture were mined using PepM as a genetic hook. Homologs were identified in the Jensen lab strain collection using both the IGM genome browser and NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). This search returned two hits, *Salinispora arenicola* CNS-205 (Genbank accession CP000850.1) and *Salinispora pacifica* CNS-865 (GenBank accession ARGJ00000000). Further analysis was performed by the antiSMASH 3.0 algorithm.[55] Multiple sequence alignments were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

4.5.2 Analysis of the distribution of PepM homologs

The protein sequence of the *S. pacifica* PepM enzyme was used to search the Marine Metagenomics Portal (https://mmp.sfb.uit.no/) for homologs found in marine metagenomes and in complete and incomplete genome sequences. The MAR BLAST algorithm was used to search the database, and results were returned in a table for further analysis. Sequence similarity networks were then generated from these results using the Enzyme Function Initiative's Enzyme Similarity Tool (https://efi.igb.illinois.edu/).

4.5.3 Cloning and Plasmid Assembly

After identification of the phosphonate BGC in *S. pacifica* CNS-863, the cluster was cloned using PCR assembler method. The cluster was amplified in six 4 kb pieces

using PrimeSTAR HS DNA polymerase (Takara) according to the manufacturer's guidelines. The fragments covering the beginning and the end of the cluster were designed to include 40 bp homology sequences to the pCAP03 capture vector (AddGene accession number 69862). The amplified fragments had 250-585 bp sequence overlaps to allow the assembly by homology recombination. Each amplified fragment was analyzed by gel electrophoresis, purified, and confirmed by Sanger sequencing (Supplemental Figure S4.1). A table of the primers used can be found in Supplemental Table S4.5. From this point onward, a slightly modified version of the original TAR protocol[56] was followed and can be found deposited with the pCAP03 sequence on the Addgene database.[57] Rather than capturing the BGC directly from gDNA, purified amplicons and linear plasmid were used to assemble the full BGC in yeast. Yeast spheroplasts (200 µL) were transformed with linear vector (200 ng) and amplified fragments (50 ng each). Transformed spheroplasts were plated on selective agar and incubated for 4 days at 30 °C. Colonies were PCR screened for regions of the target biosynthetic cluster, and plasmid from positive hits were transformed into *E. coli* Top10 (Life Technologies) by electroporation. The correct assembly of the BGC into pCAP03 was confirmed by restriction analysis, and the resulting plasmid was named pCAP03-phos (Supplemental Figure S4.2).

The plasmid pCAP03-phos was transformed into *E. coli* ET12567 cells for further conjugation into *S. coelicolor* M1152 using tri-parental mating. *E. coli* ET12567 carrying the pUB307 conjugation plasmid was used as a helper strain. After exconjugates appeared, three sequential rounds of growth selection were performed on selective media (MS agar + nalidixic acid (100 µg/mL) + kanamycin (50 µg/mL) to ensure the

plasmid was stably integrated and maintained in the heterologous host genome as determined by PCR (Supplemental Figure S4.9).

4.5.4 Heterologous Expression

Spores of the negative control, blank host *S. coelicolor* M1152, and *S. coelicolor* M1152/pCAP03-phos were used to streak plates of International Streptomyces Project Medium #4 (ISP4). After 10 days, plates were frozen at -20 °C overnight and subsequently defrosted. The resulting 'juice', separated from the polymerized agar matrix, was frozen and lyophilized to yield a powdered crude containing both media components and the products of fermentation. This crude powder was dissolved in 600 uL of D_2O at a concentration of 0.5 g/mL, and particulate was removed by centrifugation (5 min 10,000 x g). The resulting soluble fraction of the crude was subjected to ^{31}P NMR analysis.

4.5.5 Purification of Phosphonates Produced in Heterologous Host

Lyophilized crude preparations were dissolved in dH_2O (0.5 g/mL), centrifuged to remove particulate (10 min at 15,000 x g), and applied to a gravity-fed column of AG1x8 resin (-OH formulation) (Biorad). The column was washed with three column volumes of dH_2O to remove residual salts, and phosphonates were eluted with 0.2M ammonium acetate (pH = 7.2). The eluant was frozen and dried by lyophilization, during which time ammonium acetate was also removed. This enriched fraction was re-dissolved in a minimal dH_2O and applied to a gravity-fed column of C-18 resin (PolygoPrep 100A C-18 silica gel 25-40uM, Anspec;). The wash fraction (100% H_2O) was collected and dried by lyophilization. This semi-purified material was then subjected to hydrophilic interaction liquid chromatography (HILIC) (Phenomonex Luna 5u HILIC 200A 250x10mm). Initial

conditions for the column were 10% buffer A (5mM ammonium formate in H_2O pH=6.8) and 90% buffer B (90% MeCN, 5mM ammonium formate pH=7.0) with a flow rate of 1.5 mL/min. These initial conditions were held for 15 minutes after sample injection, and subsequently a gradient from 10% A to 50% A was run over 30 minutes. Fractions were collected every 5 minutes and analyzed by ^{31}P NMR for diagnostic phosphonate chemical shifts. The positive fractions were pooled and used for ^{1}H NMR and MS/MS analysis.

4.5.6 NMR and MS/MS Analysis

Production of secondary metabolites was confirmed by HPLC-HR-ESI-MSMS analysis, carried out on an Agilent 1290 liquid chromatography system coupled to an Agilent 6530 Q-TOF (200−2000m/z, 20 keV). Prepared samples were dissolved in H_2O and injected on a C18 column (Phenomenex Luna 5µm C18(2) 100A 100x4.6 mm) with an initial mobile phase of 5% MeCN at a flow rate of 0.7 mL/min. Initial conditions were held for 10 min, and the MeCN concentration was linearly increased to 100% over 20 min afterwards. NMR spectra were collected at 298K using a JEOL ECA 500 MHz spectrometer fitted with a 5mm TCI cryoprobe.

4.5.7 Cloning and Purification of GNATs from Phosphonate Cluster

Three GNATs from the *S. pacifica* phosphonate BGC were amplified by PCR using primers that included 22 bp of homology sequences to allow cloning of the DNA fragments into the pET28a vector using Gibson assembly. Primers used for amplification can be found in Supplemental Table S4.5. After assembly, the reaction mixture was transformed into *E. coli* DH10B using electroporation and plated on LB agar

plates containing kanamycin. After overnight incubation, colonies were picked and grown overnight in liquid LB media with kanamycin. Plasmids were isolated using Qiagen's miniprep kit, and successful insertion of the GNAT sequences was confirmed by restriction analysis and sequencing. Expression plasmids were then transformed into *E. coli* BL21 for protein production. Overnight cultures (5 mL) were used to inoculate 1 L flasks of Terrific Broth and the cultures were grown until OD_{600}=0.6, at which time the flasks were cooled to 18 °C and subsequently induced with 1 mM IPTG to final concentration and incubated overnight. The cells were then harvested by centrifugation (10,000 x g for 20 min) and lysed by sonication. Proteins were isolated by Ni^{2+}-affinity chromatography using a HisTrap column (GE Life Sciences). Protein fractions were confirmed by running polyacrylamide gel electrophoresis (Supplemental Figure S4.8).

4.6 Chapter 4 References

1. Svara, J., Weferling, N. & Hofmann, T. Phosphorus Compounds, Organic. in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH Verlag GmbH & Co. KGaA, 2006). doi:10.1002/14356007.a19_545.pub2

2. Smith, S. V. Phosphorus versus nitrogen limitation in the marine environment1. *Limnol. Oceanogr.* **29**, 1149–1160 (1984).

3. Hansell, D. A. & Carlson, C. A. *Biogeochemistry of marine dissolved organic matter*. (Academic Press, 2002).

4. Benitez-Nelson, C. R. The biogeochemical cycling of phosphorus in marine systems. *Earth-Science Rev.* **51**, 109–135 (2000).

5. Krom, M. D., Herut, B. & Mantoura, R. F. C. Nutrient budget for the Eastern Mediterranean: Implications for phosphorus limitation. *Limnol. Oceanogr.* **49**, 1582–1592 (2004).

6. Thingstad TF, Krom MD, Mantoura RF, Flaten GA, Groom S, Herut B, Kress N, Law CS, Pasternak A, Pitta P, Psarra S, Rassoulzadegan F, Tanaka T, Tselepides A, Wassmann P, Woodward EM, Riser CW, Zodiatis G, Zohary T. Nature of Phosphorus Limitation in the Ultraoligotrophic Eastern Mediterranean. *Science (80-.).* **309**, 1068–1071 (2005).

7. Cotner, J., Ammerman, J., Peele, E. & Bentzen, E. Phosphorus-limited bacterioplankton growth in the Sargasso Sea. *Aquat. Microb. Ecol.* **13**, 141–149 (1997).

8. T. R. Fisher, A. B. Gustafson, K. Sellner, R. Lacouture, L. W. Haas, R. L. Wetzel, R. Magnien, D. Everitt, B. Michaels, R. Karrh. Spatial and temporal variation of resource limitation in Chesapeake Bay. *Mar. Biol.* **133**, 763–778 (1999).

9. Fisher, T., Peele, E., Ammerman, J. & Harding, L. Nutrient limitation of phytoplankton in Chesapeake Bay. *Mar. Ecol. Prog. Ser.* **82**, 51–63 (1992).

10. Baturin, G. N. Phosphorus Cycle in the Ocean. *Lithol. Miner. Resour.* **38**, 101–119 (2003).

11. Paytan, A. & McLaughlin, K. The oceanic phosphorus cycle. *Chem. Rev.* **107**, 563–76 (2007).

12. Kolowith, L. C., Ingall, E. D. & Benner, R. Composition and cycling of marine organic phosphorus. *Limnol. Oceanogr.* **46**, 309–320 (2001).

13. Ingall, E. D., Schroeder, P. A. & Berner, R. A. The nature of organic phosphorus in marine sediments: New insights from 31P NMR. *Geochim. Cosmochim. Acta* **54**, 2617–2620 (1990).

14. Paytan, A., Cade-Menun, B. J., McLaughlin, K. M. & Faul, K. L. Selective phosphorus regeneration of sinking marine particles : evidence from 31 P-NMR. (2003).

15. Kittredge, J. S. & Roberts, E. A carbon-phosphorus bond in nature. *Science* **164**, 37–42 (1969).

16. Korn, E. D., Dearborn, D. G., Fales, H. M. & Sokoloski, E. A. Phosphonoglycan. A major polysaccharide constituent of the amoeba plasma membrane contains 2-aminoethylphosphonic acid and 1-hydroxy-2-aminoethylphosphonic acid. *J. Biol. Chem.* **248**, 2257–9 (1973).

17. Yamada S, Araki S, Abe S, Kon K, Ando S, Satake M. Structural analysis of a novel triphosphonoglycosphingolipid from the egg of the sea hare, Aplysia kurodai. *J. Biochem.* **117**, 794–9 (1995).

18. Araki S, Abe S, Ando S, Kon K, Fujiwara N, Satake M. Structure of phosphonoglycosphingolipid containing pyruvylated galactose in nerve fibers of Aplysia kurodai. *J. Biol. Chem.* **264**, 19922–7 (1989).

19. Matsubara, T. & Hayashi, A. Structural studies on glycolipids of shellfish. V. Gala-6 series glycosphingolipids of the marine snail, Chlorostoma argyrostoma turbinatum. *J. Biochem.* **99**, 1401–8 (1986).

20. Miceli, M. V, Henderson, T. O. & Myers, T. C. 2-aminoethylphosphonic Acid metabolism during embryonic development of the planorbid snail helisoma. *Science* **209**, 1245–7 (1980).

21. Urai M, Nakamura T, Uzawa J, Baba T, Taniguchi K, Seki H, Ushida K.. Structural analysis of O-glycans of mucin from jellyfish (Aurelia aurita) containing 2-aminoethylphosphonate. *Carbohydr. Res.* **344**, 2182–7 (2009).

22. Kim H, Chin J, Choi H, Baek K, Lee TG, Park SE, Wang W, Hahn D, Yang I, Lee J, Mun B, Ekins M, Nam SJ, Kang H. Phosphoiodyns A and B, unique phosphorus-containing iodinated polyacetylenes from a Korean sponge Placospongia sp. *Org. Lett.* **15**, 100–3 (2013).

23. Metcalf WW, Griffin BM, Cicchillo RM, Gao J, Janga SC, Cooke HA, Circello BT, Evans BS, Martens-Habbena W, Stahl DA, van der Donk WA. Synthesis of Methylphosphonic Acid by Marine Microbes: A Source for Methane in the Aerobic Ocean. *Science (80-.).* **337**, 1104–1107 (2012).

24. Shoguchi, E., Tanaka, M., Takeuchi, T., Shinzato, C. & Satoh, N. Probing a Coral Genome for Components of the Photoprotective Scytonemin Biosynthetic Pathway and the 2-Aminoethylphosphonate Pathway. *Mar. Drugs* **11**, 559–570 (2013).

25. Bowman, E., McQueney, M., Barry, R. J. & Dunaway-Mariano, D. Catalysis and thermodynamics of the phosphoenolpyruvate/phosphonopyruvate rearrangement. Entry into the phosphonate class of naturally occurring organophosphorus compounds. *J. Am. Chem. Soc.* **110**, 5575–5576 (1988).

26. Horsman, G. P. & Zechel, D. L. Phosphonate Biochemistry. *Chem. Rev.* **117**, 5704–5783 (2017).

27. Zhang, Q., van der Donk, W. A. & Liu, W. Radical-Mediated Enzymatic

Methylation: A Tale of Two SAMS. *Acc. Chem. Res.* **45**, 555–564 (2012).

28. Werner WJ, Allen KD, Hu K, Helms GL, Chen BS, Wang SC. In Vitro Phosphinate Methylation by PhpK from Kitasatospora phosalacinea. *Biochemistry* **50**, 8986–8988 (2011).

29. Duke, S. O. The history and current status of glyphosate. *Pest Manag. Sci.* **74**, 1027–1034 (2018).

30. Steinrücken, H. C. & Amrhein, N. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. *Biochem. Biophys. Res. Commun.* **94**, 1207–12 (1980).

31. Epa, Ocspp & Opp. *US EPA - Pesticides Industry Sales and Usage 2008 - 2012*. (2008).

32. la Cecilia, D. & Maggi, F. Analysis of glyphosate degradation in a soil microcosm. *Environ. Pollut.* **233**, 201–207 (2018).

33. Hove-Jensen, B., Zechel, D. L. & Jochimsen, B. Utilization of glyphosate as phosphate source: biochemistry and genetics of bacterial carbon-phosphorus lyase. *Microbiol. Mol. Biol. Rev.* **78**, 176–97 (2014).

34. Botta F, Lavison G, Couturier G, Alliot F, Moreau-Guigon E, Fauchon N, Guery B, Chevreuil M, Blanchoud H. Transfer of glyphosate and its degradate AMPA to surface waters through urban sewerage systems. *Chemosphere* **77**, 133–139 (2009).

35. Grandcoin, A., Piel, S. & Baurès, E. AminoMethylPhosphonic acid (AMPA) in natural waters: Its sources, behavior and environmental fate. *Water Res.* **117**, 187–197 (2017).

36. Van Stempvoort DR, Spoelstra J, Senger ND, Brown SJ, Post R, Struger J. Glyphosate residues in rural groundwater, Nottawasaga River Watershed, Ontario, Canada. *Pest Manag. Sci.* **72**, 1862–1872 (2016).

37. Kwiatkowska, M., Huras, B. & Bukowska, B. The effect of metabolites and impurities of glyphosate on human erythrocytes (in vitro). *Pestic. Biochem. Physiol.* **109**, 34–43 (2014).

38. Guilherme, S., Santos, M. A., Gaivão, I. & Pacheco, M. DNA and chromosomal damage induced in fish (Anguilla anguilla L.) by aminomethylphosphonic acid (AMPA)—the major environmental breakdown product of glyphosate. *Environ. Sci. Pollut. Res.* **21**, 8730–8739 (2014).

39. Scribner, E. A., Battaglin, W. A., Gilliom, R. J. & Meyer, M. T. *Concentrations of Glyphosate, Its Degradation Product, Aminomethylphosphonic Acid, and Glufosinate in Ground- and Surface-Water, Rainfall, and Soil Samples Collected in the United States, 2001-06*. *Scientific Investigations Report* (2007). doi:10.3133/SIR20075122

40. Battaglin, W. A., Meyer, M. T., Kuivila, K. M. & Dietze, J. E. Glyphosate and Its Degradation Product AMPA Occur Frequently and Widely in U.S. Soils, Surface

Water, Groundwater, and Precipitation. *JAWRA J. Am. Water Resour. Assoc.* **50**, 275–290 (2014).

41. Jensen, P. R., Moore, B. S. & Fenical, W. The marine actinomycete genus Salinispora: a model organism for secondary metabolite discovery. *Nat. Prod. Rep.* **32**, 738–51 (2015).

42. Phase 2 Clinical Trial of NPI-0052 in Patients With Relapsed or Relapsed/Refractory Multiple Myeloma. (2017). Available at: https://clinicaltrials.gov/ct2/show/NCT00461045. (Accessed: 23rd April 2019)

43. A Phase III Trial of With Marizomib in Patients With Newly Diagnosed Glioblastoma - Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT03345095. (Accessed: 23rd April 2019)

44. Ziemert N, Lechner A, Wietz M, Millán-Aguiñaga N, Chavarria KL, Jensen PR.. Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. *Proc. Natl. Acad. Sci.* **111**, E1130–E1139 (2014).

45. Blodgett JA, Thomas PM, Li G, Velasquez JE, van der Donk WA, Kelleher NL, Metcalf WW. Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. *Nat. Chem. Biol.* **3**, 480–485 (2007).

46. Imai S, Seto H, Sasaki T, Tsuruoka T, Ogawa H, Satoh A, Inouye S, Niida T, Otake N. Studies on the biosynthesis of bialaphos (SF-1293). 4. Production of phosphonic acid derivatives, 2-hydroxyethylphosphonic acid, hydroxymethylphosphonic acid and phosphonoformic acid by blocked mutants of Streptomyces hygroscopicus SF-1293 and their roles in the biosynthesis of bialaphos. *J. Antibiot. (Tokyo).* **37**, 1505–8 (1984).

47. Woodyer, R. D., Li, G., Zhao, H. & van der Donk, W. A. New insight into the mechanism of methyl transfer during the biosynthesis of fosfomycin. *Chem. Commun. (Camb).* 359–61 (2007). doi:10.1039/b614678c

48. Drzyzga, D., Forlani, G., Vermander, J., Kafarski, P. & Lipok, J. Biodegradation of the aminopolyphosphonate DTPMP by the cyanobacterium Anabaena variabilis proceeds via a C-P lyase-independent pathway. *Environ. Microbiol.* **19**, 1065–1076 (2017).

49. Castle LA, Siehl DL, Gorton R, Patten PA, Chen YH, Bertain S, Cho HJ, Duck N, Wong J, Liu D, Lassner MW. Discovery and directed evolution of a glyphosate tolerance gene. *Science* **304**, 1151–4 (2004).

50. Siehl, D. L., Castle, L. A., Gorton, R. & Keenan, R. J. The molecular basis of glyphosate resistance by an optimized microbial acetyltransferase. *J. Biol. Chem.* **282**, 11446–55 (2007).

51. Siehl DL, Castle LA, Gorton R, Chen YH, Bertain S, Cho HJ, Keenan R, Liu D, Lassner MW. Evolution of a microbial acetyltransferase for modification of glyphosate: a novel tolerance strategy. *Pest Manag. Sci.* **61**, 235–240 (2005).

52. Mercurio, P., Flores, F., Mueller, J. F., Carter, S. & Negri, A. P. Glyphosate

persistence in seawater. *Mar. Pollut. Bull.* **85**, 385–390 (2014).

53. Bai, S. H. & Ogbourne, S. M. Glyphosate: environmental contamination, toxicity and potential risks to human health via food contamination. *Environ. Sci. Pollut. Res.* **23**, 18988–19001 (2016).

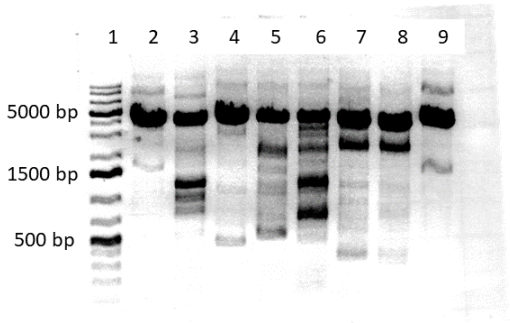
54. Kissane, Z. & Shephard, J. M. The rise of glyphosate and new opportunities for biosentinel early-warning studies. *Conserv. Biol.* **31**, 1293–1300 (2017).

55. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015).

56. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 491–6 (1996).

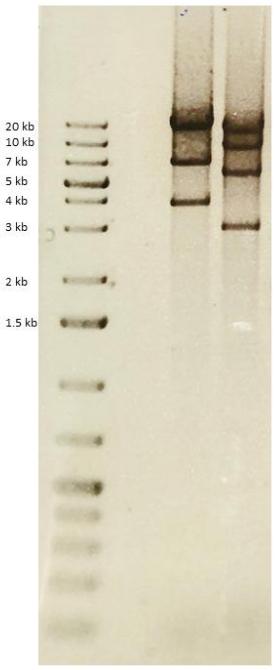
57. Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS. Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem. Biol.* **10**, 2841–2849 (2015).

4.7 Chapter 4 Supplementary Information



Lane 1: GeneRuler 1kb Plus DNA Ladder
Lane 2: Fragment 1 without pCAP homology (expected size 4166 bp)
Lane 3: Fragment 2 (expected size 4466 bp)
Lane 4: Fragment 3 (expected size 4417 bp)
Lane 5: Fragment 4 (expected size 4444 bp)
Lane 6: Fragment 5 (expected size 4400 bp)
Lane 7: Fragment 6 including pCAP homology (expected size 4268 bp)
Lane 8: Fragment 6 without pCAP homology (expected size 4188 bp)
Lane 9: Fragment 1 including pCAP homology (expected size 4246 bp)

Supplementary Figure S4.1: Gel electrophoresis analysis of phosphonate BGC amplicons prior to assembly.

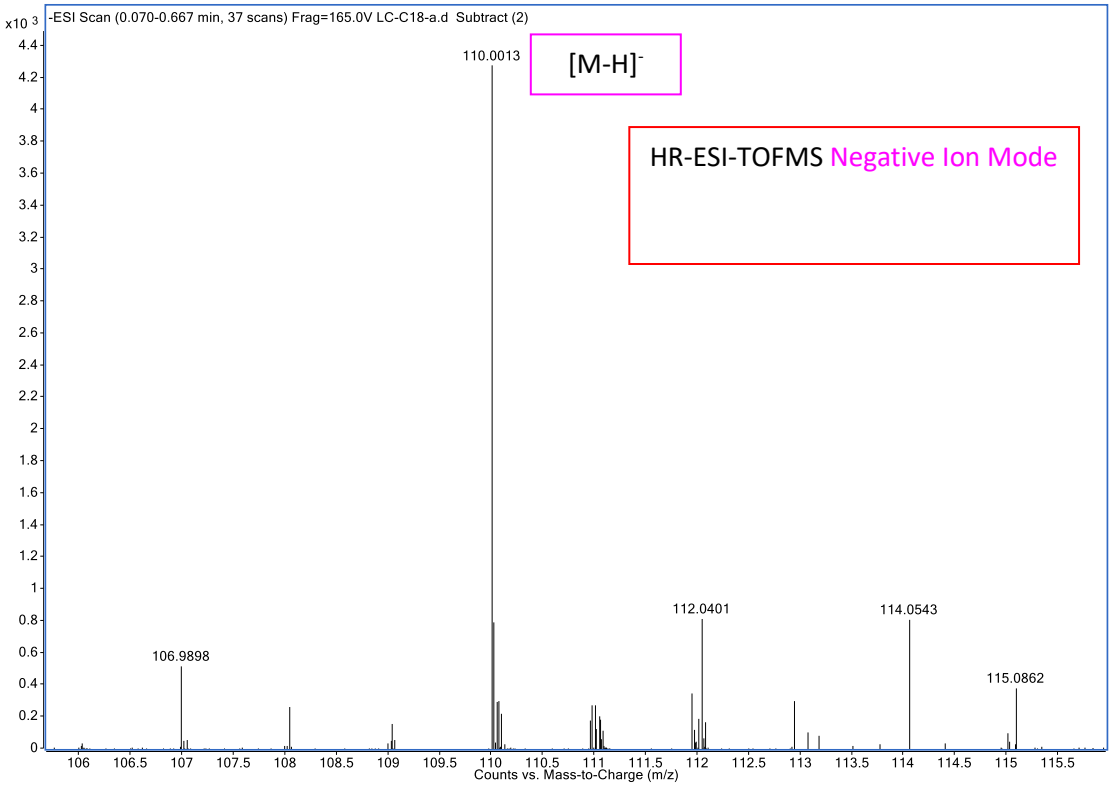


Lane 1: GeneRuler 1kb Plus DNA Ladder

Lane 2: pCAP03-phos plasmid Digest with ScaI;
Expected Bands (in bp): 23730, 7029, 3849

Lane 3: pCAP03-phos plasmid Digest with SacI;
Expected Bands (in bp): 16435, 9435, 5788, 3000

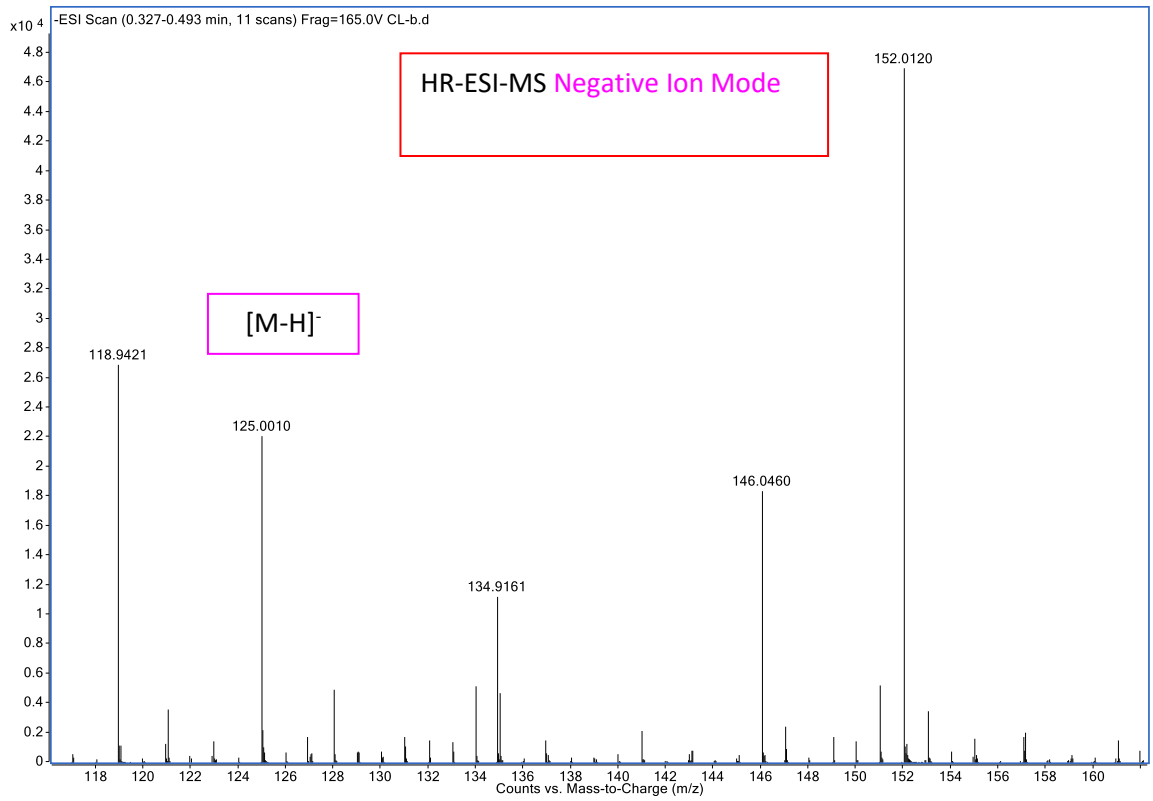
Supplementary Figure S4.2: Gel digest of assembled *S. pacifica* BGC in the pCAP03 plasmid.



Search Results: Sample LC-C18

Mass Measured	Theo. Mass	Delta (ppm)	Composition
110.0013	110.0013	0.0	$[C H_5 N O_3 P]^+$

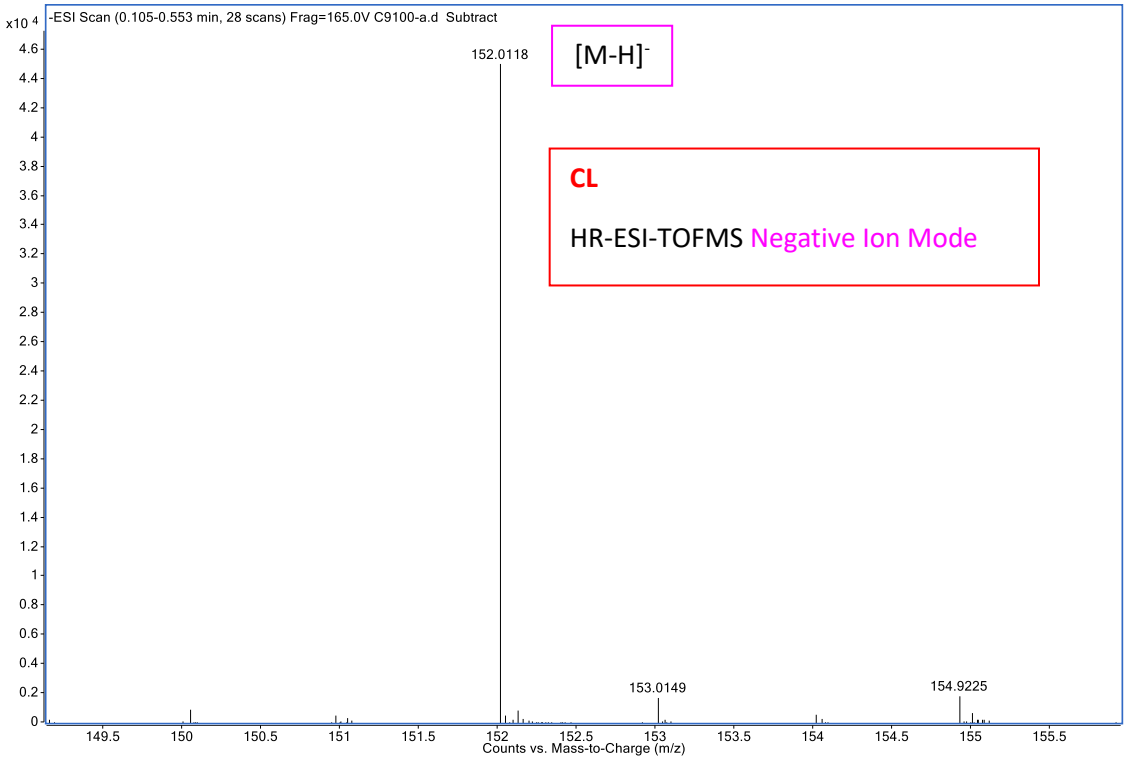
Supplementary Figure S4.3: HRMS data for m/z=110, including molecular formula calculation.



Search Results: **Sample CL**

Mass Measured	Theo. Mass	Delta (ppm)	Composition
125.0010	125.0009	0.8	$[C_2H_6O_4P]^-$

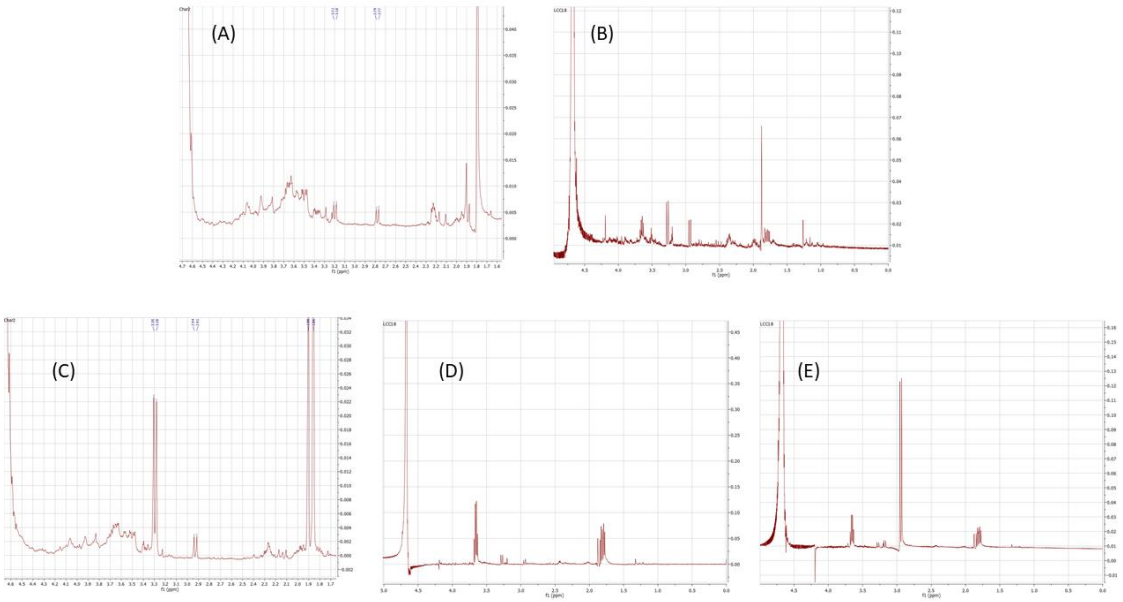
Supplementary Figure S4.4: HRMS data for m/z=125, including molecular formula calculation.



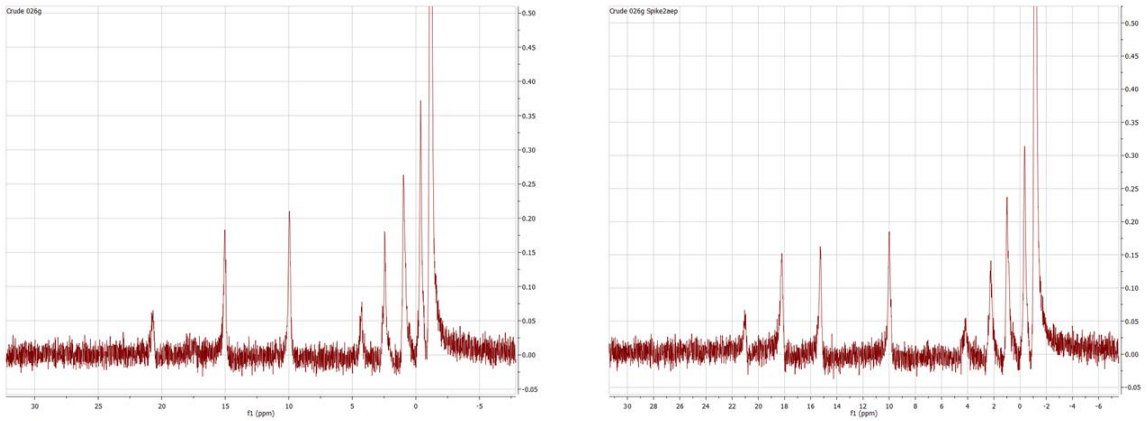
Search Results: Sample CL

Mass Measured	Theo. Mass	Delta (ppm)	Composition
152.0118	152.0118	0.0	$[C_3H_7NO_4P]^-$

Supplementary Figure S4.5: HRMS data for m/z=152, including molecular formula calculation.

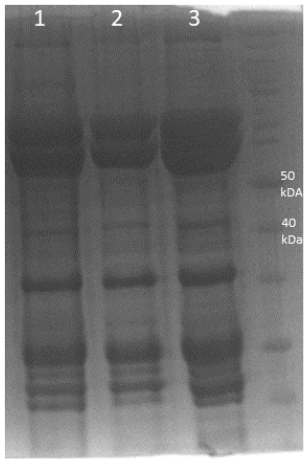


Supplementary Figure S4.6: ^{1}H NMR analysis of spiking experiments with naAMPA, AMPA, and 2-HEP synthetic standards. (A and B) ^{1}H NMR spectrum of enriched preparation; (C) ^{1}H NMR spectrum of enriched preparation shown in A spiked with 1 mg of na-AMPA; (D) ^{1}H NMR spectrum of enriched preparation shown in B spiked with 0.5 mg of 2-HEP standard (E) ^{1}H NMR spectrum of enriched preparation shown in B spiked with 0.5 mg of 2-HEP standard and 1 mg AMPA standard.



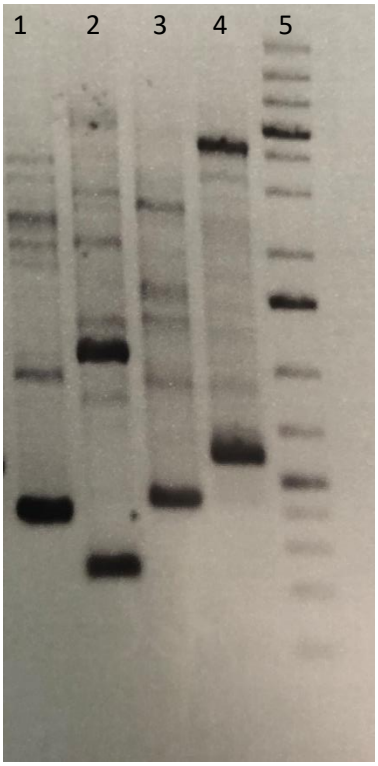
A B

Supplementary Figure S4.7: NMR Analysis of Crude Extract. (A) ^{31}P NMR spectrum of 0.25g crude extract dissolved in 500 uL D_2O (B) ^{31}P NMR analysis of the same crude extract, spiked with 0.5 mg of 2-AEP synthetic standard. This spiking experiment shows the production levels are approximately equal to 1 mg of phosphonate natural products AMPA and na-AMPA per gram of crude extract, with 2-HEP produced in lower titer (approximately 0.4 mg per gram of crude extract). Each 75 mL plate of ISP4 solid media yields 0.29 grams of crude extract, with the calculated yield of phosphonates being 3.9 mg/L and 1.5 mg/L respectively.



Expected Bands:
Lane 1: 37.659 kDa
Lane 2: 40.492 kDa
Lane 3: 40.543 kDa

Supplementary Figure S4.8: Protein gel (10% acrylamide) of HisTrap purified GNAT enzymes. Lane 1 = SppG; Lane 2 = SppH; Lane 3 = SppO.



Lane 1: Primers Fragment 2 Reverse and Fragment 1 Forward

Lane 2: Primers Fragment 3 Forward and Fragment 2 Reverse

Lane 3: Primers Fragment 4 Forward and Fragment 3 Reverse

Lane 4: Primers Fragment 5 Forward and Fragment 4 Reverse

Lane 5: 1 kb Plus DNA Ladder

Supplementary Figure S4.9: Amplification of small regions of the *S. pacifica* phosphonate BGC from genomic DNA isolated from the heterologous host. Expected Bands: Lane 1 = 405 bp; Lane 2 = 248 bp Lane 3 = 443 bp; Lane 4 = 585 bp; Lane 5 = 20 kb, 10 kb, 7kb, 5kb, 4kb, 3kb, 2kb, 1500 bp, 1000 bp, 700 bp, 500 bp, 400 bp, 300 bp, 200 bp, 75 bp.

Supplementary Table S4.1: Genes in the *Salinispora pacifica* phosphonate BGC and their closest homolog determined by NCBI BLAST.

S. pacifica **phosphonate BGC Gene**	**Annotation**	**Closest Homolog**	**Homolog Accession Number**	**% Amino Acid Identity**
sppA	MFS transporter	MFS transporter [Streptomyces griseus]	WP_050513796.1	61%
sppB	nucleotidyl transferase	nucleotidyltransferase domain-containing protein [Streptomyces griseus]	WP_030720118.1	51%
sppC	conserved hypothetical	hypothetical protein [Streptomyces griseus]	WP_030720115	62%
sppD	conserved hypothetical/tRNA synthetase	hypothetical protein [Streptomyces griseus]	WP_050513795.1	57%
sppE	conserved hypothetical	YqcI/YcgG family protein [Streptomyces griseochromogenes]	WP_079146497	60%
sppF	hypothetical/nucleoside diphosphate kinase	hypothetical protein [Streptomyces griseus]	WP_030720106	63%
sppG	GCN5 family acetyltransferase	GNAT family N-acetyltransferase [Streptomyces griseus]	WP_032769332	61%
sppH	GCN5 family acetyltransferase	GNAT family N-acetyltransferase [Streptomyces griseus]	WP_030720101	55%

sppI	XRE family transcriptional regulator	helix-turn-helix transcriptional regulator [Streptomyces griseochromogenes]	WP_067299575.1	62%
sppJ	phosphoenolpyruvate mutase	phosphoenolpyruvate phosphomutase [Streptomyces griseus]	WP_032769330.1	79%
sppK	phosphonopyruvate decarboxylase	phosphonopyruvate decarboxylase [Streptomyces rimosus]	WP_125058098.1	59%
sppL	alcohol dehydrogenase	iron-containing alcohol dehydrogenase [Streptomyces griseofuscus]	WP_085567253	53%
sppM	3-phosphoglycerate dehydrogenase	C-terminal binding protein [Streptomyces griseofuscus]	WP_037661451	72%
sppN	aspartate aminotransferase	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme [Streptomyces rimosus]	WP_125058095	56%
sppO	hypothetical/peptidoglycan related	GNAT family N-acetyltransferase [Nocardiopsis chromatogenes]	WP_017624875	48%
sppP	stearoyl-CoA 9-desaturase	hypothetical protein [Streptomyces griseus]	WP_030720072.1	64%
sppQ	methylaspartate mutase, cobalamine binding	methylaspartate mutase [Streptomyces griseus]	WP_078614503	64%

sppR	methylaspartate mutase, no cobalamine binding	methylaspartate mutase [Streptomyces griseus]	WP_030720070	68%
sppS	ATP-grasp, phosphoribosylglycinamide synthetase	ATP-grasp domain-containing protein [Streptomyces griseus]	WP_030720068	64%
sppT	MBL fold metallohydrolase	MBL fold metallo-hydrolase [Pseudonocardiales bacterium]	PZS25218	59%
sppU	conserved hypothetical, NTP transferase family	nucleotidyltransferase family protein [Micromonospora aurantiaca]	WP_091329982.1	71%

Supplementary Table S4.2: Multiple Sequence Alignment of *S. pacifica* phosphonate BGC GNATs and *B. licheniformis* GNAT

	SspG	GAT	SspH	SspO
SspG	100	8.85	22.19	22.26
GAT	8.85	100	19.12	19.85
SspH	22.19	19.12	100	29.3
SspO	22.26	19.85	29.3	100

Supplementary Table S4.3: *S. pacifica* PepM homologs in marine genomes and metagenomes in the MAR DB database.

Accession # for BLAST Hit	Annotation for Hit	Expect Value for Hit
WP_018735391.1_MMP02441365	hypothetical protein [Salinispora pacifica] [mmp_id=MMP02441365] [mmp_db=mardb]	0
MAR14796.1_MMP07618462	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618462] [mmp_db=mardb]	2.87165E-92
PXF51769.1_MMP08434975	phosphoenolpyruvate mutase [ANME-1 cluster archaeon] [mmp_id=MMP08434975] [mmp_db=mardb]	6.04147E-91
MBL76896.1_MMP07618508	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]	6.3566E-91
ABX12056.1_MMP00000032	putative phosphoenolpyruvate phosphomutase [Nitrosopumilus maritimus SCM1] [mmp_id=MMP00000032] [mmp_db=marref]	1.31302E-90
MMP494431_182150	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	8.01278E-90
ARF65919.1_MMP06323808	phosphoenolpyruvate phosphomutase [Streptomyces violaceoruber] [mmp_id=MMP06323808] [mmp_db=marref]	2.63058E-89

MMP490065_5965	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]	4.94183E-89
PIG51557.1_MMP07626382	phosphoenolpyruvate mutase [Verrucosispora sp. CNZ293] [mmp_id=MMP07626382] [mmp_db=mardb]	5.49309E-89
MMP494431_91392	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	1.7326E-88
MBR73495.1_MMP07618617	phosphoenolpyruvate phosphomutase [Dehalococcoidaceae bacterium] [mmp_id=MMP07618617] [mmp_db=mardb]	6.83327E-88
PXF27507.1_MMP08159281	phosphoenolpyruvate phosphomutase [Thaumarchaeota archaeon] [mmp_id=MMP08159281] [mmp_db=mardb]	6.92148E-88
MBL75307.1_MMP07618508	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]	1.14974E-87
MMP490065_193497	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]	2.27528E-87
MMP494431_249509	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase]	2.77307E-87

	[Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	
WP_049564768.1_MMP03468208	phosphoenolpyruvate phosphomutase [Streptomyces sp. SBT349] [mmp_id=MMP03468208] [mmp_db=mardb]	5.01368E-87
GBH35141.1_MMP00113975	hypothetical protein NZNM25_19320 [Candidatus Nitrosopumilus sp. NM25] [mmp_id=MMP00113975] [mmp_db=mardb]	8.77841E-87
MMP491463_206367	[UniRef50=UniRef50_A0A1G1J1V0,Cluster: Uncharacterized protein] [Interpro=IPR029044,Nucleotide-diphospho-sugar transferases] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	9.08667E-87
MBQ09481.1_MMP07619140	hypothetical protein CMD96_06800 [Gammaproteobacteria bacterium] [mmp_id=MMP07619140] [mmp_db=mardb]	9.18418E-87
MMP490065_229543	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]	1.4156E-86
MAS51187.1_MMP07618457	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618457] [mmp_db=mardb]	1.75549E-86
MMP494431_228096	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain]	1.83272E-86

	[Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	
WP_007403095.1_MMP02471010	phosphoenolpyruvate synthase [Candidatus Nitrosoarchaeum limnia] [mmp_id=MMP02471010] [mmp_db=marref]	1.96452E-86
EGG41199.1_MMP02471010	putative phosphoenolpyruvate phosphomutase [Candidatus Nitrosoarchaeum limnia SFB1] [mmp_id=MMP02471010] [mmp_db=mardb]	1.96452E-86
GCA_002506165.1_00636_MMP06027469	Phosphonopyruvate hydrolase [mmp_id=MMP06027469] [mmp_db=mardb]	2.13886E-86
MMP494431_137904	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	4.35924E-86
MMP494431_37872	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]	6.64614E-86
MBL77230.1_MMP07618508	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]	6.93181E-86
GCA_003235875.1_00167_MMP09287831	Phosphonopyruvate hydrolase [mmp_id=MMP09287831] [mmp_db=mardb]	9.05617E-86
GCA_000230485.1_00568_MMP02954312	Phosphonopyruvate hydrolase [mmp_id=MMP02954312] [mmp_db=mardb]	2.95525E-85

MBL76962.1_MMP07618508	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]	3.32845E-85
MAR36777.1_MMP07618460	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618460] [mmp_db=mardb]	5.46609E-85
MAG08768.1_MMP07620233	phosphoenolpyruvate mutase [Candidatus Woesearchaeota archaeon] [mmp_id=MMP07620233] [mmp_db=mardb]	5.58133E-85
GCA_003229915.1_00709_MMP09240145	Phosphonopyruvate hydrolase [mmp_id=MMP09240145] [mmp_db=mardb]	9.16058E-85
MMP494431_194244	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	9.19547E-85
GCA_000230485.1_01502_MMP02954312	Phosphonopyruvate hydrolase [mmp_id=MMP02954312] [mmp_db=mardb]	1.14002E-84
MBL77037.1_MMP07618508	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]	1.16493E-84
MAF20390.1_MMP07619599	hypothetical protein CMI55_01780 [Parcubacteria group bacterium] [mmp_id=MMP07619599] [mmp_db=mardb]	2.06478E-84
MBR74309.1_MMP07618617	phosphoenolpyruvate phosphomutase [Dehalococcoidaceae bacterium] [mmp_id=MMP07618617] [mmp_db=mardb]	3.52925E-84

MAR68904.1_MMP07619518	phosphoenolpyruvate phosphomutase [Nitrospina sp.] [mmp_id=MMP07619518] [mmp_db=mardb]	5.80503E-84
GCA_002502135.1_00411_MMP06027474	Phosphonopyruvate hydrolase [mmp_id=MMP06027474] [mmp_db=mardb]	8.99961E-84
WP_066506643.1_MMP04316851	phosphoenolpyruvate phosphomutase [Clostridiales bacterium MCWD3] [mmp_id=MMP04316851] [mmp_db=mardb]	5.8019E-83
MMP490065_168674	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]	5.96839E-83
GCA_001674955.1_00721_MMP03840809	Phosphonopyruvate hydrolase [mmp_id=MMP03840809] [mmp_db=mardb]	1.29735E-82
MMP490065_209202	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]	1.58908E-82
MBA30896.1_MMP07618650	phosphoenolpyruvate phosphomutase [Dehalococcoidia bacterium] [mmp_id=MMP07618650] [mmp_db=mardb]	3.6666E-82
MMP490065_23702	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]	5.53324E-82

MMP491463_307773	[UniRef50=UniRef50_A9EXQ4,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	9.80769E-81
GCA_001657385.1_03693_MMP04939327	Phosphonopyruvate hydrolase [mmp_id=MMP04939327] [mmp_db=mardb]	1.41245E-80
MMP494431_197159	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]	1.50947E-80
MMP490065_74875	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]	2.09914E-80
MAG96192.1_MMP07619952	phosphonopyruvate hydrolase [Rhodospirillaceae bacterium] [mmp_id=MMP07619952] [mmp_db=mardb]	4.93569E-80
MMP494431_93748	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]	1.55154E-79
MMP490065_195017	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain]	1.90467E-79

	[Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]	
MMP494431_133827	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	2.17395E-79
MMP490065_237234	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]	5.87316E-79
MAS50419.1_MMP07618457	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618457] [mmp_db=mardb]	6.2201E-79
MBN18910.1_MMP07618503	phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618503] [mmp_db=mardb]	2.31081E-78
GCA_003251795.1_01021_MMP09288020	Phosphonopyruvate hydrolase [mmp_id=MMP09288020] [mmp_db=mardb]	7.54687E-78
MMP494431_85898	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]	7.99216E-78
MBL24275.1_MMP07619992	phosphonopyruvate hydrolase [Rhodospirillaceae bacterium] [mmp_id=MMP07619992] [mmp_db=mardb]	2.25393E-77

MMP494431_1390	[UniRef50=UniRef50_A0A075H9G4,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]	6.56725E-77
RAS38621.1_MMP05720407	phosphoenolpyruvate mutase [Thaumarchaeota archaeon SCGC AC-337_F14] [mmp_id=MMP05720407] [mmp_db=mardb]	9.27298E-77
RAI28397.1_MMP07424250	phosphonopyruvate hydrolase [Rhodobium orientis] [mmp_id=MMP07424250] [mmp_db=mardb]	9.7111E-77
GBE44636.1_MMP00081387	phosphonopyruvate hydrolase [bacterium BMS3Bbin10] [mmp_id=MMP00081387] [mmp_db=mardb]	1.8664E-74
MAM74537.1_MMP07620173	phosphoenolpyruvate phosphomutase [Tistrella sp.] [mmp_id=MMP07620173] [mmp_db=mardb]	1.56806E-73
GCA_003247775.1_00216_MMP09287867	Phosphonopyruvate hydrolase [mmp_id=MMP09287867] [mmp_db=mardb]	2.76597E-72
MBA77101.1_MMP07620174	phosphoenolpyruvate phosphomutase [Tistrella sp.] [mmp_id=MMP07620174] [mmp_db=mardb]	7.20537E-72
MAD38201.1_MMP07620172	phosphoenolpyruvate phosphomutase [Tistrella sp.] [mmp_id=MMP07620172] [mmp_db=mardb]	7.20537E-72
GCA_003228955.1_00049_MMP09240132	Phosphonopyruvate hydrolase [mmp_id=MMP09240132] [mmp_db=mardb]	1.00518E-71
WP_062762299.1_MMP04332635	phosphoenolpyruvate phosphomutase [Tistrella mobilis] [mmp_id=MMP04332635] [mmp_db=mardb]	2.33802E-71
WP_007668809.1_MMP02436095	phosphonopyruvate hydrolase [alpha proteobacterium BAL199]	5.76003E-71

	[mmp_id=MMP02436095] [mmp_db=mardb]	
AWG33697.1_MMP06948887	phosphonopyruvate hydrolase [Alcaligenes aquatilis] [mmp_id=MMP06948887] [mmp_db=mardb]	7.20173E-71
AFK56356.1_MMP02603926	phosphoenolpyruvate phosphomutase (plasmid) [Tistrella mobilis KA081020-065] [mmp_id=MMP02603926] [mmp_db=marref]	1.73797E-70
EDN71450.1_MMP02470514	phosphoenolpyruvate phosphomutase [Beggiatoa sp. PS] [mmp_id=MMP02470514] [mmp_db=mardb]	1.88386E-69
MAO54861.1_MMP07620005	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620005] [mmp_db=mardb]	5.02677E-69
WP_003808778.1_MMP02231287	MULTISPECIES: phosphoenolpyruvate phosphomutase [Bordetella] [mmp_id=MMP02231287] [mmp_db=mardb]	6.08127E-69
WP_003808778.1_MMP02231461	MULTISPECIES: phosphoenolpyruvate phosphomutase [Bordetella] [mmp_id=MMP02231461] [mmp_db=mardb]	6.08127E-69
WP_003808778.1_MMP02231460	MULTISPECIES: phosphoenolpyruvate phosphomutase [Bordetella] [mmp_id=MMP02231460] [mmp_db=mardb]	6.08127E-69
WP_037338324.1_MMP02929445	phosphoenolpyruvate mutase [Salinisphaera hydrothermalis] [mmp_id=MMP02929445] [mmp_db=mardb]	7.10816E-69
GCA_001643485.1_00432_MMP04113178	Phosphonopyruvate hydrolase [mmp_id=MMP04113178] [mmp_db=mardb]	1.69693E-68

GCA_002415625.1_02740_MMP06457039	Phosphonopyruvate hydrolase [mmp_id=MMP06457039] [mmp_db=mardb]	4.90634E-68
GCA_002359255.1_02600_MMP06456280	Phosphonopyruvate hydrolase [mmp_id=MMP06456280] [mmp_db=mardb]	4.90634E-68
MBB91078.1_MMP07619312	phosphoenolpyruvate mutase [Magnetovibrio sp.] [mmp_id=MMP07619312] [mmp_db=mardb]	4.90634E-68
PJN35138.1_MMP07346485	phosphoenolpyruvate phosphomutase [Streptomyces sp. CB02613] [mmp_id=MMP07346485] [mmp_db=mardb]	6.11661E-68
MBL68683.1_MMP07620218	phosphoenolpyruvate mutase [Verrucomicrobiales bacterium] [mmp_id=MMP07620218] [mmp_db=mardb]	6.563E-68
MBL68532.1_MMP07620218	phosphoenolpyruvate mutase [Verrucomicrobiales bacterium] [mmp_id=MMP07620218] [mmp_db=mardb]	6.563E-68
GCA_002433235.1_02961_MMP06454690	Phosphonopyruvate hydrolase [mmp_id=MMP06454690] [mmp_db=mardb]	6.563E-68
WP_071022580.1_MMP05904704	MULTISPECIES: phosphonopyruvate hydrolase [Cupriavidus] [mmp_id=MMP05904704] [mmp_db=marref]	9.42216E-68
WP_036540049.1_MMP02743897	phosphoenolpyruvate mutase [Pseudooceanicola nanhaiensis] [mmp_id=MMP02743897] [mmp_db=mardb]	1.07187E-67
MBH03641.1_MMP07620252	phosphoenolpyruvate mutase [Xanthomonadales bacterium] [mmp_id=MMP07620252] [mmp_db=mardb]	4.62882E-67

GCA_003283105.1_00089_MMP08886030	Phosphonopyruvate hydrolase [mmp_id=MMP08886030] [mmp_db=mardb]	5.38726E-67
GCA_002495315.1_00380_MMP06027030	Phosphonopyruvate hydrolase [mmp_id=MMP06027030] [mmp_db=mardb]	6.9355E-67
GCA_003235555.1_00156_MMP09287802	Phosphonopyruvate hydrolase [mmp_id=MMP09287802] [mmp_db=mardb]	9.97912E-67
MAH21593.1_MMP07620156	phosphoenolpyruvate mutase [Thaumarchaeota archaeon] [mmp_id=MMP07620156] [mmp_db=mardb]	1.65749E-66
OYD24006.1_MMP06648058	phosphonopyruvate hydrolase [Oceanimonas baumannii] [mmp_id=MMP06648058] [mmp_db=mardb]	4.45142E-66
GCA_003209165.1_01310_MMP08886582	Phosphonopyruvate hydrolase [mmp_id=MMP08886582] [mmp_db=mardb]	4.65079E-66
MAA98202.1_MMP07620130	phosphoenolpyruvate mutase [Stappia sp.] [mmp_id=MMP07620130] [mmp_db=mardb]	5.86162E-66
MAN80433.1_MMP07619308	phosphoenolpyruvate mutase [Magnetovibrio sp.] [mmp_id=MMP07619308] [mmp_db=mardb]	7.6484E-66
PTX60241.1_MMP08776942	phosphoenolpyruvate phosphomutase [Melghirimyces profundicolus] [mmp_id=MMP08776942] [mmp_db=mardb]	8.07569E-66
GCA_003251555.1_00343_MMP09288039	Phosphonopyruvate hydrolase [mmp_id=MMP09288039] [mmp_db=mardb]	1.0896E-65
MAI26272.1_MMP07620119	phosphoenolpyruvate mutase [Spirochaeta sp.] [mmp_id=MMP07620119] [mmp_db=mardb]	1.18995E-65

MAS83288.1_MMP07619288	phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619288] [mmp_db=mardb]	1.2768E-65
MBK68764.1_MMP07619298	phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619298] [mmp_db=mardb]	1.33161E-65
SDU33903.1_MMP05428979	phosphoenolpyruvate mutase [Stappia sp. ES.058] [mmp_id=MMP05428979] [mmp_db=marref]	1.36389E-65
PPR37399.1_MMP07286218	Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin4] [mmp_id=MMP07286218] [mmp_db=mardb]	1.68099E-65
MBK20413.1_MMP07619996	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619996] [mmp_db=mardb]	3.67511E-65
WP_070393116.1_MMP05826283	phosphoenolpyruvate mutase [Moorea producens] [mmp_id=MMP05826283] [mmp_db=marref]	4.80691E-65
PPR31149.1_MMP07286220	Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin6] [mmp_id=MMP07286220] [mmp_db=mardb]	6.03543E-65
GCA_002500605.1_00781_MMP06450621	Phosphonopyruvate hydrolase [mmp_id=MMP06450621] [mmp_db=mardb]	7.52064E-65
GCA_003250975.1_00685_MMP09288140	Phosphonopyruvate hydrolase [mmp_id=MMP09288140] [mmp_db=mardb]	9.00062E-65
GCA_003235935.1_01300_MMP09287818	Phosphonopyruvate hydrolase [mmp_id=MMP09287818] [mmp_db=mardb]	9.00062E-65

MAZ79102.1_MMP07619291	phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619291] [mmp_db=mardb]	1.53655E-64
GCA_002471405.1_00998_MMP06455356	Phosphonopyruvate hydrolase [mmp_id=MMP06455356] [mmp_db=mardb]	3.04451E-64
GCA_002433595.1_00639_MMP06451400	Phosphonopyruvate hydrolase [mmp_id=MMP06451400] [mmp_db=mardb]	3.04451E-64
MBD97865.1_MMP07618393	phosphoenolpyruvate mutase [bacterium] [mmp_id=MMP07618393] [mmp_db=mardb]	3.04451E-64
PCJ58663.1_MMP07568849	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07568849] [mmp_db=mardb]	4.01065E-64
WP_083373742.1_MMP05826284	phosphoenolpyruvate mutase [Moorea producens] [mmp_id=MMP05826284] [mmp_db=marref]	6.64107E-64
WP_023477253.1_MMP02469983	phosphoenolpyruvate phosphomutase [Burkholderia cenocepacia] [mmp_id=MMP02469983] [mmp_db=mardb]	6.79445E-64
WP_039357130.1_MMP07498378	MULTISPECIES: phosphoenolpyruvate mutase [Burkholderia] [mmp_id=MMP07498378] [mmp_db=mardb]	7.70262E-64
KHL13034.1_MMP03070120	phosphoenolpyruvate phosphomutase [Mumia flava] [mmp_id=MMP03070120] [mmp_db=mardb]	7.70262E-64
MAF61158.1_MMP07618425	phosphoenolpyruvate mutase [Blastomonas sp.] [mmp_id=MMP07618425] [mmp_db=mardb]	7.74954E-64

MBV31477.1_MMP07620290	phosphoenolpyruvate mutase [Rickettsiales bacterium] [mmp_id=MMP07620290] [mmp_db=mardb]	1.09019E-63
MAI81079.1_MMP07618662	phosphoenolpyruvate mutase [Deltaproteobacteria bacterium] [mmp_id=MMP07618662] [mmp_db=mardb]	2.04046E-63
MMP494431_190222	[UniRef50=UniRef50_A0A0D2W1R5,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=2.7.7.33] [mmp_id=MMP494431] [mmp_db=marcat]	2.56569E-63
MMP494431_190221	[UniRef50=UniRef50_A0A0D2W1R5,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]	2.56569E-63
GCA_003228495.1_01494_MMP09240199	Phosphonopyruvate hydrolase [mmp_id=MMP09240199] [mmp_db=mardb]	2.90463E-63
AUS80769.1_MMP08364581	phosphoenolpyruvate mutase [Actinoalloteichus sp. AHMU CJ021] [mmp_id=MMP08364581] [mmp_db=marref]	4.70282E-63
KHD11437.1_MMP03165106	phosphoenolpyruvate phosphomutase [Candidatus Thiomargarita nelsonii] [mmp_id=MMP03165106] [mmp_db=mardb]	5.34294E-63
KHD06688.1_MMP03165106	phosphoenolpyruvate phosphomutase [Candidatus Thiomargarita nelsonii] [mmp_id=MMP03165106] [mmp_db=mardb]	5.34294E-63
WP_051242647.1_MMP02441158	phosphoenolpyruvate mutase [Stappia stellulata] [mmp_id=MMP02441158] [mmp_db=mardb]	6.78803E-63
WP_005001090.1_MMP02436141	phosphoenolpyruvate mutase [Nitrococcus mobilis] [mmp_id=MMP02436141] [mmp_db=mardb]	8.52667E-63

MBR93142.1_MMP07619802	phosphoenolpyruvate mutase [Proteobacteria bacterium] [mmp_id=MMP07619802] [mmp_db=mardb]	8.73092E-63
GCA_001657395.1_01944_MMP04939355	Phosphonopyruvate hydrolase [mmp_id=MMP04939355] [mmp_db=mardb]	1.03724E-62
MAI82122.1_MMP07620418	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620418] [mmp_db=mardb]	1.24264E-62
MAJ30509.1_MMP07620417	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620417] [mmp_db=mardb]	1.26894E-62
MBK31660.1_MMP07619297	phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619297] [mmp_db=mardb]	1.45186E-62
SEO57998.1_MMP04490248	phosphoenolpyruvate mutase [Salinihabitans flavidus] [mmp_id=MMP04490248] [mmp_db=mardb]	2.56101E-62
GCA_003247415.1_00636_MMP09287891	Phosphonopyruvate hydrolase [mmp_id=MMP09287891] [mmp_db=mardb]	2.61062E-62
MMP491463_299831	[UniRef50=UniRef50_A0A1M3MSG2,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.3.1.157] [mmp_id=MMP491463] [mmp_db=marcat]	3.07956E-62
MMP491463_299830	[UniRef50=UniRef50_A0A1M3MSG2,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	3.07956E-62

ABA87887.1_MMP02598295	phosphoenolpyruvate phosphomutase [Pelobacter carbinolicus DSM 2380] [mmp_id=MMP02598295] [mmp_db=marref]	3.68831E-62
MBS57551.1_MMP07618480	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618480] [mmp_db=mardb]	6.08797E-62
GCA_003235785.1_00267_MMP09287860	Phosphonopyruvate hydrolase [mmp_id=MMP09287860] [mmp_db=mardb]	8.05709E-62
WP_072286117.1_MMP04909702	phosphoenolpyruvate mutase [Pelobacter acetylenicus] [mmp_id=MMP04909702] [mmp_db=marref]	9.09894E-62
GCA_001657395.1_01961_MMP04939355	Phosphonopyruvate hydrolase [mmp_id=MMP04939355] [mmp_db=mardb]	9.20715E-62
WP_042155018.1_MMP02261502	MULTISPECIES: phosphoenolpyruvate mutase [Planktothrix] [mmp_id=MMP02261502] [mmp_db=mardb]	1.15014E-61
WP_042155018.1_MMP02261501	MULTISPECIES: phosphoenolpyruvate mutase [Planktothrix] [mmp_id=MMP02261501] [mmp_db=mardb]	1.15014E-61
MAP83779.1_MMP07620403	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620403] [mmp_db=mardb]	1.21183E-61
GCA_001643555.1_00024_MMP04113182	Phosphonopyruvate hydrolase [mmp_id=MMP04113182] [mmp_db=mardb]	1.33482E-61
MBB42517.1_MMP07619988	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619988] [mmp_db=mardb]	1.47153E-61
MBI32378.1_MMP07619142	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium]	1.57928E-61

	[mmp_id=MMP07619142] [mmp_db=mardb]	
WP_045519159.1_MMP00000021	phosphoenolpyruvate mutase [Bacillus niacini] [mmp_id=MMP00000021] [mmp_db=mardb]	1.66269 E-61
MBC36642.1_MMP07620041	phosphoenolpyruvate mutase [Rickettsiales bacterium] [mmp_id=MMP07620041] [mmp_db=mardb]	2.46485 E-61
RAP34809.1_MMP08965197	phosphoenolpyruvate mutase, partial [Candidatus Marinamargulisbacteria bacterium SCGC AG-439-L15] [mmp_id=MMP08965197] [mmp_db=mardb]	2.75474 E-61
PPR57756.1_MMP07286199	Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha3_Bin6] [mmp_id=MMP07286199] [mmp_db=mardb]	4.65648 E-61
MBQ09441.1_MMP07619140	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619140] [mmp_db=mardb]	4.78168 E-61
RAP31434.1_MMP08965200	phosphoenolpyruvate mutase [Candidatus Marinamargulisbacteria bacterium SCGC AG-343-D04] [mmp_id=MMP08965200] [mmp_db=mardb]	4.82408 E-61
MBS94531.1_MMP07618544	phosphoenolpyruvate mutase [Chromatiales bacterium] [mmp_id=MMP07618544] [mmp_db=mardb]	6.42813 E-61
MBF42503.1_MMP07619141	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619141] [mmp_db=mardb]	7.18648 E-61
WP_083202259.1_MMP05382895	phosphoenolpyruvate mutase [Stappia indica] [mmp_id=MMP05382895] [mmp_db=mardb]	7.25135 E-61

PRZ53072.1_MMP06264457	phosphoenolpyruvate mutase [Paraburkholderia insulsa] [mmp_id=MMP06264457] [mmp_db=mardb]	7.4128E-61
MAD47893.1_MMP07620397	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620397] [mmp_db=mardb]	8.49362E-61
GCA_002470825.1_00459_MMP06453278	Phosphonopyruvate hydrolase [mmp_id=MMP06453278] [mmp_db=mardb]	9.87021E-61
MAS81580.1_MMP07619289	phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619289] [mmp_db=mardb]	9.87021E-61
WP_083206217.1_MMP05382876	phosphoenolpyruvate mutase [Stappia indica] [mmp_id=MMP05382876] [mmp_db=mardb]	1.41843E-60
MBQ26716.1_MMP07619525	phosphoenolpyruvate mutase [Nitrospiraceae bacterium] [mmp_id=MMP07619525] [mmp_db=mardb]	1.42879E-60
MAM67613.1_MMP07619956	phosphoenolpyruvate mutase, partial [Rhodospirillaceae bacterium] [mmp_id=MMP07619956] [mmp_db=mardb]	2.22744E-60
AWB42962.1_MMP08934269	phosphoenolpyruvate mutase [Paenibacillus sp. CAA11] [mmp_id=MMP08934269] [mmp_db=marref]	2.50636E-60
MBV40496.1_MMP07619967	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619967] [mmp_db=mardb]	3.17028E-60
MAG58367.1_MMP07619727	phosphoenolpyruvate mutase [Planctomycetes bacterium] [mmp_id=MMP07619727] [mmp_db=mardb]	3.2859E-60

ABE54447.1_MMP02598300	2,3-dimethylmalate lyase [Shewanella denitrificans OS217] [mmp_id=MMP02598300] [mmp_db=marref]	3.78883E-60
MAF82498.1_MMP07618542	phosphoenolpyruvate mutase [Chromatiales bacterium] [mmp_id=MMP07618542] [mmp_db=mardb]	4.21608E-60
MBE11134.1_MMP07619979	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619979] [mmp_db=mardb]	4.22756E-60
GCA_002469945.1_01135_MMP06457391	Phosphonopyruvate hydrolase [mmp_id=MMP06457391] [mmp_db=mardb]	4.24909E-60
MAF95691.1_MMP07620004	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620004] [mmp_db=mardb]	5.85926E-60
GCA_002500465.1_00032_MMP06455355	Phosphonopyruvate hydrolase [mmp_id=MMP06455355] [mmp_db=mardb]	6.39048E-60
GCA_002390245.1_02979_MMP06457461	Phosphonopyruvate hydrolase [mmp_id=MMP06457461] [mmp_db=mardb]	6.85815E-60
WP_005011554.1_MMP2272521	phosphoenolpyruvate mutase [Nitrospina gracilis] [mmp_id=MMP2272521] [mmp_db=mardb]	8.31446E-60
MBT77452.1_MMP07618543	phosphoenolpyruvate mutase [Chromatiales bacterium] [mmp_id=MMP07618543] [mmp_db=mardb]	8.64637E-60
MBG05926.1_MMP07620003	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620003] [mmp_db=mardb]	9.13398E-60

WP_070991887.1_MMP05792550	phosphoenolpyruvate mutase [Pseudoalteromonas byunsanensis] [mmp_id=MMP05792550] [mmp_db=mardb]	1.12569E-59
WP_102772790.1_MMP04858689	phosphonopyruvate hydrolase [Achromobacter pulmonis] [mmp_id=MMP04858689] [mmp_db=mardb]	1.21318E-59
MBI07088.1_MMP07619998	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619998] [mmp_db=mardb]	1.31446E-59
GCA_002434595.1_01329_MMP06450461	Phosphonopyruvate hydrolase [mmp_id=MMP06450461] [mmp_db=mardb]	1.44381E-59
PPR42533.1_MMP07286216	Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin2] [mmp_id=MMP07286216] [mmp_db=mardb]	1.64025E-59
OEU71675.1_MMP05301632	phosphoenolpyruvate mutase [Desulfuromonadales bacterium C00003068] [mmp_id=MMP05301632] [mmp_db=mardb]	1.88451E-59
MBV28968.1_MMP07619966	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619966] [mmp_db=mardb]	2.21492E-59
MBU70460.1_MMP07618603	phosphoenolpyruvate mutase [Cupriavidus sp.] [mmp_id=MMP07618603] [mmp_db=mardb]	4.49543E-59
MMP492357_300029	[UniRef50=UniRef50_A0A0B5FB86,Cluster: Phosphonopyruvate hydrolase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.60] [mmp_id=MMP492357] [mmp_db=marcat]	5.5522E-59

MMP492357_300028	[UniRef50=UniRef50_A0A0B5FB86,Cluster: Phosphonopyruvate hydrolase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	5.5522E-59
GCA_003233615.1_02490_MMP09239988	Phosphonopyruvate hydrolase [mmp_id=MMP09239988] [mmp_db=mardb]	6.8694E-59
GCA_003230275.1_01538_MMP09240095	Phosphonopyruvate hydrolase [mmp_id=MMP09240095] [mmp_db=mardb]	9.39344E-59
PCH55231.1_MMP07568997	phosphoenolpyruvate mutase [Burkholderiaceae bacterium] [mmp_id=MMP07568997] [mmp_db=mardb]	1.01649E-58
PPR48436.1_MMP07286215	Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin1] [mmp_id=MMP07286215] [mmp_db=mardb]	1.30489E-58
WP_036500424.1_MMP02898118	phosphoenolpyruvate mutase [Nitrosococcus oceani] [mmp_id=MMP02898118] [mmp_db=mardb]	1.51943E-58
ABA57815.1_MMP02598329	2,3-dimethylmalate lyase [Nitrosococcus oceani ATCC 19707] [mmp_id=MMP02598329] [mmp_db=marref]	1.61706E-58
WP_002810069.1_MMP02864944	phosphoenolpyruvate mutase [Nitrosococcus oceani] [mmp_id=MMP02864944] [mmp_db=mardb]	1.61706E-58
WP_002810069.1_MMP02436179	phosphoenolpyruvate mutase [Nitrosococcus oceani] [mmp_id=MMP02436179] [mmp_db=mardb]	1.61706E-58

GCA_002340825.1_03645_MMP06452276	Phosphonopyruvate hydrolase [mmp_id=MMP06452276] [mmp_db=mardb]	1.85454E-58
GCA_002341065.1_04382_MMP06457331	Phosphonopyruvate hydrolase [mmp_id=MMP06457331] [mmp_db=mardb]	1.85454E-58
ADJ28131.1_MMP02598531	phosphoenolpyruvate phosphomutase [Nitrosococcus watsonii C-113] [mmp_id=MMP02598531] [mmp_db=marref]	1.92906E-58
MBV24227.1_MMP07619965	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619965] [mmp_db=mardb]	2.44059E-58
GCA_003233115.1_01686_MMP09240020	Phosphonopyruvate hydrolase [mmp_id=MMP09240020] [mmp_db=mardb]	4.29526E-58
GCA_002453985.1_00693_MMP06451993	Phosphonopyruvate hydrolase [mmp_id=MMP06451993] [mmp_db=mardb]	4.45039E-58
GCA_001542995.1_00717_MMP03777284	Phosphonopyruvate hydrolase [mmp_id=MMP03777284] [mmp_db=mardb]	5.26325E-58
PPR73314.1_MMP07286195	Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha3_Bin2] [mmp_id=MMP07286195] [mmp_db=mardb]	6.10317E-58
ADE14458.1_MMP02598510	phosphoenolpyruvate phosphomutase [Nitrosococcus halophilus Nc 4] [mmp_id=MMP02598510] [mmp_db=marref]	8.29563E-58
MAW55866.1_MMP07620013	phosphoenolpyruvate mutase, partial [Rhodospirillaceae bacterium] [mmp_id=MMP07620013] [mmp_db=mardb]	8.48912E-58

PPR19649.1_MMP07286231	Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha10_Bin2] [mmp_id=MMP07286231] [mmp_db=mardb]	9.99687E-58
MBO44291.1_MMP07619972	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619972] [mmp_db=mardb]	1.44855E-57
MBE03790.1_MMP07619111	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619111] [mmp_db=mardb]	1.77535E-57
GCA_002390205.1_00197_MMP06451823	Phosphonopyruvate hydrolase [mmp_id=MMP06451823] [mmp_db=mardb]	1.80791E-57
MBE88682.1_MMP07620020	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620020] [mmp_db=mardb]	2.01839E-57
GCA_002352385.1_00263_MMP06451510	Phosphonopyruvate hydrolase [mmp_id=MMP06451510] [mmp_db=mardb]	2.1738E-57
AVX06931.1_MMP07838500	phosphoenolpyruvate mutase [Bacillus sp. Y-01] [mmp_id=MMP07838500] [mmp_db=marref]	2.79487E-57
GCA_003247575.1_03863_MMP09287884	Phosphonopyruvate hydrolase [mmp_id=MMP09287884] [mmp_db=mardb]	4.45314E-57
MBH71388.1_MMP07619641	phosphoenolpyruvate mutase [Pelagibacteraceae bacterium] [mmp_id=MMP07619641] [mmp_db=mardb]	7.45183E-57
MAG23290.1_MMP07619953	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619953] [mmp_db=mardb]	8.45201E-57

PCI21724.1_MMP07568950	phosphoenolpyruvate mutase, partial [SAR324 cluster bacterium] [mmp_id=MMP07568950] [mmp_db=mardb]	8.55358E-57
PCN45699.1_MMP06759470	phosphoenolpyruvate mutase [Brevibacillus laterosporus] [mmp_id=MMP06759470] [mmp_db=mardb]	1.62012E-56
MAX17400.1_MMP07619517	phosphoenolpyruvate mutase [Nitrospina sp.] [mmp_id=MMP07619517] [mmp_db=mardb]	1.9081E-56
GAX62849.1_MMP00036669	PEP phosphomutase [Candidatus Scalindua sp. husup-a2] [mmp_id=MMP00036669] [mmp_db=mardb]	2.70165E-56
GCA_003229345.1_00234_MMP09240210	Phosphonopyruvate hydrolase [mmp_id=MMP09240210] [mmp_db=mardb]	4.38246E-56
MBM24054.1_MMP07619122	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619122] [mmp_db=mardb]	9.81665E-56
MAH84731.1_MMP07619309	phosphoenolpyruvate mutase [Magnetovibrio sp.] [mmp_id=MMP07619309] [mmp_db=mardb]	2.82176E-55
GCA_002501105.1_00891_MMP06457369	Phosphonopyruvate hydrolase [mmp_id=MMP06457369] [mmp_db=mardb]	3.39047E-55
KMP11646.1_MMP03376160	phosphoenolpyruvate mutase [Nitrospina sp. SCGC_AAA799_C22] [mmp_id=MMP03376160] [mmp_db=mardb]	2.39979E-54
MAQ83588.1_MMP07619456	phosphoenolpyruvate mutase [Maritimibacter sp.] [mmp_id=MMP07619456] [mmp_db=mardb]	3.12875E-54

MBN08222.1_MMP07619989	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619989] [mmp_db=mardb]	7.11748E-54
MAI10307.1_MMP07620010	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620010] [mmp_db=mardb]	8.27619E-54
MBA30457.1_MMP07618650	phosphoenolpyruvate mutase [Dehalococcoidia bacterium] [mmp_id=MMP07618650] [mmp_db=mardb]	1.02093E-53
MBN18068.1_MMP07618503	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618503] [mmp_db=mardb]	1.45504E-53
GCA_003249475.1_01054_MMP09288137	Phosphonopyruvate hydrolase [mmp_id=MMP09288137] [mmp_db=mardb]	1.50245E-53
MBJ78420.1_MMP07619524	phosphoenolpyruvate mutase [Nitrospinae bacterium] [mmp_id=MMP07619524] [mmp_db=mardb]	2.0392E-53
MBS86051.1_MMP07619227	phosphoenolpyruvate mutase [Candidatus Heimdallarchaeota archaeon] [mmp_id=MMP07619227] [mmp_db=mardb]	2.64585E-53
GCF_000375765.1_00099_MMP02256471	Phosphonopyruvate hydrolase [mmp_id=MMP02256471] [mmp_db=mardb]	3.76556E-53
MBM32497.1_MMP07618509	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618509] [mmp_db=mardb]	3.95559E-53
GCA_003249475.1_00019_MMP09288137	Phosphonopyruvate hydrolase [mmp_id=MMP09288137] [mmp_db=mardb]	4.01287E-53
MAX28946.1_MMP07620167	phosphoenolpyruvate mutase [Thiotrichales bacterium]	4.40846E-53

	[mmp_id=MMP07620167] [mmp_db=mardb]	
GCA_000372225.1_01694_MMP02261272	Phosphonopyruvate hydrolase [mmp_id=MMP02261272] [mmp_db=mardb]	4.96176E-53
MBR71886.1_MMP07619974	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619974] [mmp_db=mardb]	5.27756E-53
MBJ26613.1_MMP07618312	phosphoenolpyruvate mutase [Alphaproteobacteria bacterium] [mmp_id=MMP07618312] [mmp_db=mardb]	5.38733E-53
MAG26910.1_MMP07619591	phosphoenolpyruvate mutase [Candidatus Pacearchaeota archaeon] [mmp_id=MMP07619591] [mmp_db=mardb]	5.81182E-53
GCA_003232235.1_00208_MMP09240083	Phosphonopyruvate hydrolase [mmp_id=MMP09240083] [mmp_db=mardb]	1.22118E-51
MBM01825.1_MMP07618507	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618507] [mmp_db=mardb]	1.2653E-51
MBP63019.1_MMP07619706	hypothetical protein CMJ62_15975 [Planctomycetaceae bacterium] [mmp_id=MMP07619706] [mmp_db=mardb]	1.65974E-51
GBL05756.1_MMP00115722	phosphoenolpyruvate phosphomutase [Glaciecola sp. KUL10] [mmp_id=MMP00115722] [mmp_db=mardb]	2.10018E-51
OEU65941.1_MMP05301624	phosphoenolpyruvate mutase, partial [Desulfobacterales bacterium S5133MH16] [mmp_id=MMP05301624] [mmp_db=mardb]	2.57234E-51

GCA_002434715.1_04403_MMP06452182	Phosphonopyruvate hydrolase [mmp_id=MMP06452182] [mmp_db=mardb]	1.96625E-50
WP_101758417.1_MMP4467369	phosphoenolpyruvate mutase [Oceanicoccus sp. KOV_DT_Chl] [mmp_id=MMP4467369] [mmp_db=mardb]	3.59461E-50
MAG47670.1_MMP07618362	phosphoenolpyruvate mutase [archaeon] [mmp_id=MMP07618362] [mmp_db=mardb]	3.95981E-50
MAV14427.1_MMP07618450	phosphoenolpyruvate mutase, partial [Chloroflexi bacterium] [mmp_id=MMP07618450] [mmp_db=mardb]	4.62182E-50
MBM03222.1_MMP07618511	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618511] [mmp_db=mardb]	4.95555E-50
GCA_001629325.1_00331_MMP04534682	Phosphonopyruvate hydrolase [mmp_id=MMP04534682] [mmp_db=mardb]	1.45484E-49
MAF51975.1_MMP07618439	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618439] [mmp_db=mardb]	1.57426E-49
GCA_003232435.1_01571_MMP09240056	Phosphonopyruvate hydrolase [mmp_id=MMP09240056] [mmp_db=mardb]	1.90713E-49
KFM21051.1_MMP02869637	Phosphoenolpyruvate phosphomutase protein [Marine Group I thaumarchaeote SCGC AAA799-B03] [mmp_id=MMP02869637] [mmp_db=mardb]	2.06684E-49
GCA_003246675.1_01424_MMP09287898	Phosphonopyruvate hydrolase [mmp_id=MMP09287898] [mmp_db=mardb]	2.18189E-49
GCA_003235975.1_00265_MMP09287816	Phosphonopyruvate hydrolase [mmp_id=MMP09287816] [mmp_db=mardb]	2.18189E-49

MMP492357_152197	[UniRef50=UniRef50_A0A0S7X8Q6,Cluster: Uncharacterized protein] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	2.24078E-49
MAU56302.1_MMP07618644	phosphoenolpyruvate mutase [Dehalococcoidia bacterium] [mmp_id=MMP07618644] [mmp_db=mardb]	3.43487E-49
GCA_002420445.1_00259_MMP06455107	Phosphonopyruvate hydrolase [mmp_id=MMP06455107] [mmp_db=mardb]	3.72232E-49
MBM14753.1_MMP07619519	phosphoenolpyruvate mutase [Nitrospina sp.] [mmp_id=MMP07619519] [mmp_db=mardb]	4.39359E-49
MAS50841.1_MMP07618457	hypothetical protein CL712_02820 [Chloroflexi bacterium] [mmp_id=MMP07618457] [mmp_db=mardb]	4.56589E-49
GCA_001577055.1_00552_MMP04423161	Phosphonopyruvate hydrolase [mmp_id=MMP04423161] [mmp_db=mardb]	5.10539E-49
MBO82775.1_MMP07618251	phosphoenolpyruvate mutase [Actinobacteria bacterium] [mmp_id=MMP07618251] [mmp_db=mardb]	6.71757E-49
WP_034225084.1_MMP03004373	phosphoenolpyruvate mutase [Arenimonas donghaensis] [mmp_id=MMP03004373] [mmp_db=mardb]	7.22965E-49
MMP491463_358252	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=2.7.7.14] [mmp_id=MMP491463] [mmp_db=marcat]	7.83241E-49
MMP491463_358251	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR004821,Cytidyltransferase-	7.83241E-49

	like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	
MBE31609.1_MMP07620261	hypothetical protein CMP17_01390 [Rickettsiales bacterium] [mmp_id=MMP07620261] [mmp_db=mardb]	9.68259E-49
MBO39511.1_MMP07620019	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620019] [mmp_db=mardb]	1.04618E-48
GCA_002348795.1_01296_MMP06451904	Phosphonopyruvate hydrolase [mmp_id=MMP06451904] [mmp_db=mardb]	1.04618E-48
GCA_003232445.1_02328_MMP09240055	Phosphonopyruvate hydrolase [mmp_id=MMP09240055] [mmp_db=mardb]	1.44359E-48
GCA_003229735.1_00663_MMP09240163	Phosphonopyruvate hydrolase [mmp_id=MMP09240163] [mmp_db=mardb]	1.54965E-48
GCA_003235465.1_01204_MMP09287807	Phosphonopyruvate hydrolase [mmp_id=MMP09287807] [mmp_db=mardb]	2.05127E-48
GCA_002500965.1_00733_MMP06452478	Phosphonopyruvate hydrolase [mmp_id=MMP06452478] [mmp_db=mardb]	2.61667E-48
MBL75785.1_MMP07618508	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]	3.87439E-48
GCA_003213695.1_00213_MMP08886429	Phosphonopyruvate hydrolase [mmp_id=MMP08886429] [mmp_db=mardb]	4.097E-48
WP_103439162.1_MMP08100003	phosphoenolpyruvate mutase [Arenibacter hampyeongensis] [mmp_id=MMP08100003] [mmp_db=mardb]	4.60144E-48

WP_111976523.1_MMP09469620	phosphoenolpyruvate mutase [Catenovulum sp. RQJ05] [mmp_id=MMP09469620] [mmp_db=mardb]	4.89741E-48
SEA64967.1_MMP05660420	phosphoenolpyruvate mutase [Desulfuromusa kysingii] [mmp_id=MMP05660420] [mmp_db=mardb]	5.07928E-48
ARN73867.1_MMP06075352	phosphoenolpyruvate mutase [Oceanicoccus sagamiensis] [mmp_id=MMP06075352] [mmp_db=marref]	5.96603E-48
MMP492357_270844	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.7.39] [mmp_id=MMP492357] [mmp_db=marcat]	6.27216E-48
MMP492357_270843	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	6.27216E-48
MBO51482.1_MMP07619713	phosphoenolpyruvate mutase [Planctomycetaceae bacterium] [mmp_id=MMP07619713] [mmp_db=mardb]	6.41729E-48
WP_068379690.1_MMP04487161	phosphoenolpyruvate mutase [Paraglaciecola sp. S66] [mmp_id=MMP04487161] [mmp_db=mardb]	7.88192E-48
GCA_002450435.1_00579_MMP06450360	Phosphonopyruvate hydrolase [mmp_id=MMP06450360] [mmp_db=mardb]	8.48761E-48
GCA_003281795.1_00664_MMP08886115	Phosphonopyruvate hydrolase [mmp_id=MMP08886115] [mmp_db=mardb]	9.15359E-48
MBD86332.1_MMP07618488	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618488] [mmp_db=mardb]	9.39594E-48

MMP492357_176301	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]	1.56989E-47
MMP492357_176300	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	1.56989E-47
MBK66302.1_MMP07618514	hypothetical protein CL769_05050 [Chloroflexi bacterium] [mmp_id=MMP07618514] [mmp_db=mardb]	1.91223E-47
MAG39062.1_MMP07620234	phosphoenolpyruvate mutase [Candidatus Woesearchaeota archaeon] [mmp_id=MMP07620234] [mmp_db=mardb]	2.14761E-47
MMP492357_71496	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	2.21138E-47
MMP492357_71495	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=3.11.1.3] [mmp_id=MMP492357] [mmp_db=marcat]	2.21138E-47
MMP492357_155310	[UniRef50=UniRef50_G0ENI0,Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	2.25078E-47
MAX58083.1_MMP07618500	phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618500] [mmp_db=mardb]	2.28063E-47

WP_067553003.1_MMP04332636	phosphoenolpyruvate mutase [Oceanibaculum pacificum] [mmp_id=MMP04332636] [mmp_db=mardb]	2.31942E-47
AWK12922.1_MMP08984914	phosphoenolpyruvate phosphomutase [Streptomyces spongiicola] [mmp_id=MMP08984914] [mmp_db=marref]	2.36302E-47
GCA_003282145.1_00604_MMP08886084	Phosphonopyruvate hydrolase [mmp_id=MMP08886084] [mmp_db=mardb]	2.67009E-47
PKH00836.1_MMP08125768	phosphoenolpyruvate mutase [Paraglaciecola sp. MB-3u-78] [mmp_id=MMP08125768] [mmp_db=mardb]	2.94432E-47
GCA_001626765.1_04224_MMP04534590	Phosphonopyruvate hydrolase [mmp_id=MMP04534590] [mmp_db=mardb]	3.57753E-47
WP_078509320.1_MMP02441002	hypothetical protein [Streptomyces sp. CNT302] [mmp_id=MMP02441002] [mmp_db=mardb]	4.33335E-47
MAG50420.1_MMP07618366	phosphoenolpyruvate mutase [archaeon] [mmp_id=MMP07618366] [mmp_db=mardb]	4.36627E-47
MBN22549.1_MMP07618423	phosphoenolpyruvate mutase [Bdellovibrionaceae bacterium] [mmp_id=MMP07618423] [mmp_db=mardb]	5.57348E-47
PSM55803.1_MMP08493996	phosphoenolpyruvate mutase [Clostridium diolis] [mmp_id=MMP08493996] [mmp_db=mardb]	5.72991E-47
MBL19935.1_MMP07619026	phosphoenolpyruvate mutase [Flavobacteriaceae bacterium] [mmp_id=MMP07619026] [mmp_db=mardb]	6.59415E-47

GCA_002413545.1_01931_MMP06456309	Phosphonopyruvate hydrolase [mmp_id=MMP06456309] [mmp_db=mardb]	6.95611E-47
GCA_003280295.1_01643_MMP08886241	Phosphonopyruvate hydrolase [mmp_id=MMP08886241] [mmp_db=mardb]	7.08148E-47
MMP492357_221512	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]	7.59762E-47
MMP492357_221511	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	7.59762E-47
WP_017212535.1_MMP08107721	phosphoenolpyruvate mutase [Clostridium beijerinckii] [mmp_id=MMP08107721] [mmp_db=mardb]	8.06672E-47
MMP492357_311978	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]	9.29222E-47
MMP492357_311977	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	9.29222E-47
KIE23726.1_MMP03070123	phosphoenolpyruvate phosphomutase [Streptomyces sp. MUSC 125] [mmp_id=MMP03070123] [mmp_db=mardb]	1.03405E-46

MBH60513.1_MMP07618652	hypothetical protein CL907_05055 [Dehalococcoidia bacterium] [mmp_id=MMP07618652] [mmp_db=mardb]	1.1237E-46
MMP492357_283672	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	1.13898E-46
MMP492357_283671	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Priam=3.11.1.3] [mmp_id=MMP492357] [mmp_db=marcat]	1.13898E-46
AVP54233.1_MMP08398331	phosphoenolpyruvate mutase [Clostridium tetani] [mmp_id=MMP08398331] [mmp_db=marref]	1.25336E-46
MMP492357_310525	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]	1.58182E-46
MMP492357_310524	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	1.58182E-46
MBH44942.1_MMP07619132	hypothetical protein CMD88_05755 [Gammaproteobacteria bacterium] [mmp_id=MMP07619132] [mmp_db=mardb]	1.5831E-46
KYK28104.1_MMP04495909	hypothetical protein AYK20_01710 [Thermoplasmatales archaeon SG8-52-1] [mmp_id=MMP04495909] [mmp_db=mardb]	1.68444E-46
PCI19122.1_MMP07568951	phosphoenolpyruvate mutase [SAR202 cluster bacterium]	1.71762E-46

	[mmp_id=MMP07568951] [mmp_db=mardb]	
PPR62234.1_MMP07286202	hypothetical protein CFH10_00887, partial [Alphaproteobacteria bacterium MarineAlpha4_Bin2] [mmp_id=MMP07286202] [mmp_db=mardb]	2.22026E-46
MMP492357_293075	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]	2.32028E-46
MMP492357_293074	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	2.32028E-46
MAG69988.1_MMP07618216	phosphoenolpyruvate mutase [Acidobacteria bacterium] [mmp_id=MMP07618216] [mmp_db=mardb]	3.0607E-46
AVQ37341.1_MMP08397979	phosphoenolpyruvate mutase [Clostridium botulinum] [mmp_id=MMP08397979] [mmp_db=marref]	3.52848E-46
AVQ40888.1_MMP08398059	phosphoenolpyruvate mutase [Clostridium botulinum] [mmp_id=MMP08398059] [mmp_db=marref]	3.52848E-46
MMP492357_208316	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]	4.10955E-46
MMP492357_208315	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain]	4.10955E-46

	[Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	
MMP491463_91825	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=2.7.1.167] [mmp_id=MMP491463] [mmp_db=marcat]	4.47623E-46
MMP491463_91824	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	4.47623E-46
SMF02759.1_MMP02745866	phosphoenolpyruvate mutase [Alteromonadaceae bacterium Bs31] [mmp_id=MMP02745866] [mmp_db=mardb]	6.60293E-46
SNS59778.1_MMP05421640	phosphoenolpyruvate mutase [Ekhidna lutea] [mmp_id=MMP05421640] [mmp_db=mardb]	7.99127E-46
GCA_003233015.1_00740_MMP09240024	Phosphonopyruvate hydrolase [mmp_id=MMP09240024] [mmp_db=mardb]	8.90629E-46
MAV05675.1_MMP07619615	phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619615] [mmp_db=mardb]	1.12496E-45
GCA_003233155.1_00187_MMP09240016	Phosphonopyruvate hydrolase [mmp_id=MMP09240016] [mmp_db=mardb]	1.25273E-45
GCA_001180265.1_00184_MMP3368574	Phosphonopyruvate hydrolase [mmp_id=MMP3368574] [mmp_db=mardb]	1.75676E-45
GCA_003233015.1_00171_MMP09240024	Phosphonopyruvate hydrolase [mmp_id=MMP09240024] [mmp_db=mardb]	2.32869E-45
GCA_001655195.1_02715_MMP03763493	Phosphonopyruvate hydrolase [mmp_id=MMP03763493] [mmp_db=mardb]	2.8545E-45

MBS59812.1_MMP07618351	phosphoenolpyruvate mutase [Anaerolineaceae bacterium] [mmp_id=MMP07618351] [mmp_db=mardb]	3.47227E-45
MMP491463_347519	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=2.7.7.2] [mmp_id=MMP491463] [mmp_db=marcat]	3.8229E-45
MMP491463_347518	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	3.8229E-45
MMP494431_95861	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=2.7.7.39] [mmp_id=MMP494431] [mmp_db=marcat]	3.83896E-45
MMP494431_95860	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]	3.83896E-45
BAN03327.1_MMP00060985	putative phosphoenolpyruvate phosphomutase [Ilumatobacter coccineus YM16-304] [mmp_id=MMP00060985] [mmp_db=marref]	3.89425E-45
MMP492357_352174	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.7.39] [mmp_id=MMP492357] [mmp_db=marcat]	3.96663E-45
MMP492357_352173	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold]	3.96663E-45

	[Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	
GCA_002346435.1_00149_MMP06457216	Phosphonopyruvate hydrolase [mmp_id=MMP06457216] [mmp_db=mardb]	4.40685E-45
GCA_003281605.1_01756_MMP08886141	Phosphonopyruvate hydrolase [mmp_id=MMP08886141] [mmp_db=mardb]	6.79276E-45
MAW74576.1_MMP07619353	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619353] [mmp_db=mardb]	7.30142E-45
MMP490065_154055	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=2.7.7.39] [mmp_id=MMP490065] [mmp_db=marcat]	1.21872E-44
MMP490065_154054	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]	1.21872E-44
MAJ22838.1_MMP07619619	phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619619] [mmp_db=mardb]	1.38295E-44
MMP492357_25364	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	2.20454E-44
MMP492357_25363	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold]	2.20454E-44

	[Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	
GCA_003211015.1_01168_MMP08886617	Phosphonopyruvate hydrolase [mmp_id=MMP08886617] [mmp_db=mardb]	2.21119E-44
MMP494431_144294	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Priam=2.7.7.14] [mmp_id=MMP494431] [mmp_db=marcat]	2.24422E-44
MMP494431_144293	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]	2.24422E-44
MMP492357_273581	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	2.34845E-44
MMP492357_273580	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	2.34845E-44
GCA_003212275.1_00314_MMP08886523	Phosphonopyruvate hydrolase [mmp_id=MMP08886523] [mmp_db=mardb]	2.62143E-44
MBS83214.1_MMP07619109	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619109] [mmp_db=mardb]	2.74107E-44
GCA_002377585.1_00088_MMP06457119	Phosphonopyruvate hydrolase [mmp_id=MMP06457119] [mmp_db=mardb]	3.10447E-44
MAH97793.1_MMP07618756	phosphoenolpyruvate mutase [Euryarchaeota archaeon]	3.11217E-44

	[mmp_id=MMP07618756] [mmp_db=mardb]	
ADV46661.1_MMP00713578	phosphoenolpyruvate phosphomutase [Nitratifractor salsuginis DSM 16511] [mmp_id=MMP00713578] [mmp_db=marref]	5.08519E-44
GCA_003211435.1_00667_MMP08886590	Phosphonopyruvate hydrolase [mmp_id=MMP08886590] [mmp_db=mardb]	5.31076E-44
GCA_002501165.1_00652_MMP06457363	Phosphonopyruvate hydrolase [mmp_id=MMP06457363] [mmp_db=mardb]	5.50444E-44
MBQ34353.1_MMP07619376	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619376] [mmp_db=mardb]	5.54793E-44
GCA_003215185.1_00709_MMP08886380	Phosphonopyruvate hydrolase [mmp_id=MMP08886380] [mmp_db=mardb]	6.43588E-44
KPJ58683.1_MMP03994146	hypothetical protein AMJ42_02810 [Deltaproteobacteria bacterium DG_8] [mmp_id=MMP03994146] [mmp_db=mardb]	7.8991E-44
WP_023475077.1_MMP02469983	phosphoenolpyruvate mutase [Burkholderia cenocepacia] [mmp_id=MMP02469983] [mmp_db=mardb]	1.13878E-43
ODS34541.1_MMP05368398	Phosphonopyruvate hydrolase [Candidatus Scalindua rubra] [mmp_id=MMP05368398] [mmp_db=mardb]	1.17802E-43
WP_029455268.1_MMP02841150	phosphoenolpyruvate mutase [Candidatus Pelagibacter ubique] [mmp_id=MMP02841150] [mmp_db=mardb]	1.38869E-43
MAW17907.1_MMP07619884	phosphoenolpyruvate mutase [Rhodobacteraceae bacterium] [mmp_id=MMP07619884] [mmp_db=mardb]	1.66639E-43

GCF_000438945.1_01691_MMP02440705	Phosphonopyruvate hydrolase [mmp_id=MMP02440705] [mmp_db=mardb]	2.28656E-43
WP_020403838.1_MMP02440412	phosphoenolpyruvate mutase [Gracilimonas tropica] [mmp_id=MMP02440412] [mmp_db=mardb]	2.60265E-43
GCA_002470485.1_00859_MMP06451384	Phosphonopyruvate hydrolase [mmp_id=MMP06451384] [mmp_db=mardb]	2.64297E-43
GCA_002341165.1_01494_MMP06457326	Phosphonopyruvate hydrolase [mmp_id=MMP06457326] [mmp_db=mardb]	3.29226E-43
GCA_003281725.1_00410_MMP08886127	Phosphonopyruvate hydrolase [mmp_id=MMP08886127] [mmp_db=mardb]	3.38051E-43
WP_035250982.1_MMP02742681	phosphoenolpyruvate mutase [Actibacterium atlanticum] [mmp_id=MMP02742681] [mmp_db=mardb]	3.39528E-43
SFG78580.1_MMP05216175	phosphoenolpyruvate phosphomutase [Neptunomonas qingdaonensis] [mmp_id=MMP05216175] [mmp_db=mardb]	3.55169E-43
MAD11214.1_MMP07618948	phosphoenolpyruvate mutase [Flavobacteriaceae bacterium] [mmp_id=MMP07618948] [mmp_db=mardb]	3.58836E-43
GCA_003282985.1_00886_MMP08886037	Phosphonopyruvate hydrolase [mmp_id=MMP08886037] [mmp_db=mardb]	4.89026E-43
GCA_003281565.1_01066_MMP08886145	Phosphonopyruvate hydrolase [mmp_id=MMP08886145] [mmp_db=mardb]	5.21456E-43
MMP492357_327676	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821,Cytidyltransferase-	6.26053E-43

	like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]	
MMP492357_327675	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	6.26053E-43
GCA_003213815.1_01110_MMP08886426	Phosphonopyruvate hydrolase [mmp_id=MMP08886426] [mmp_db=mardb]	7.0933E-43
GCA_002179655.1_03643_MMP06226650	Phosphonopyruvate hydrolase [mmp_id=MMP06226650] [mmp_db=mardb]	1.01197E-42
WP_032520302.1_MMP02769562	phosphoenolpyruvate mutase [Prochlorococcus marinus] [mmp_id=MMP02769562] [mmp_db=mardb]	1.06891E-42
RAI02086.1_MMP09079831	isocitrate lyase [Acuticoccus sp. PTG4-2] [mmp_id=MMP09079831] [mmp_db=mardb]	1.07746E-42
WP_072054505.1_MMP3355991	phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP3355991] [mmp_db=mardb]	1.37163E-42
WP_005419153.1_MMP02470712	phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP02470712] [mmp_db=mardb]	1.70944E-42
WP_005419153.1_MMP04520027	phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP04520027] [mmp_db=mardb]	1.70944E-42
WP_019882558.1_MMP02261247	MULTISPECIES: phosphoenolpyruvate mutase [Methylophilus] [mmp_id=MMP02261247] [mmp_db=mardb]	1.81578E-42
WP_063662137.1_MMP04519656	phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP04519656] [mmp_db=mardb]	1.82041E-42

GCA_002471275.1_04526_MMP06450620	Phosphonopyruvate hydrolase [mmp_id=MMP06450620] [mmp_db=mardb]	1.88842E-42
GCA_002473315.1_00356_MMP06452437	Phosphonopyruvate hydrolase [mmp_id=MMP06452437] [mmp_db=mardb]	1.88842E-42
GCA_002336125.1_02319_MMP06456076	Phosphonopyruvate hydrolase [mmp_id=MMP06456076] [mmp_db=mardb]	1.88842E-42
MBT05342.1_MMP07619976	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619976] [mmp_db=mardb]	2.02893E-42
GCA_002377485.1_00039_MMP06450560	Phosphonopyruvate hydrolase [mmp_id=MMP06450560] [mmp_db=mardb]	2.02893E-42
AAW85744.1_MMP02604303	phosphoenolpyruvate phosphomutase [Vibrio fischeri ES114] [mmp_id=MMP02604303] [mmp_db=marref]	2.22159E-42
GCA_002179675.1_03385_MMP06226649	Phosphonopyruvate hydrolase [mmp_id=MMP06226649] [mmp_db=mardb]	2.22159E-42
WP_011261863.1_MMP05449650	phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP05449650] [mmp_db=mardb]	2.22159E-42
WP_011261863.1_MMP03652539	phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP03652539] [mmp_db=mardb]	2.22159E-42
GCA_003213855.1_00634_MMP08886424	Phosphonopyruvate hydrolase [mmp_id=MMP08886424] [mmp_db=mardb]	2.93001E-42
MBA65651.1_MMP07619399	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619399] [mmp_db=mardb]	4.79928E-42

MMP492357_181017	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	6.08161E-42
MMP492357_181016	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698,Phosphoenolpyruvate phosphomutase, core] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	6.08161E-42
MMP492357_219960	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	6.9536E-42
MMP492357_219959	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	6.9536E-42
GCA_002471885.1_01119_MMP06451807	Phosphonopyruvate hydrolase [mmp_id=MMP06451807] [mmp_db=mardb]	7.07019E-42
GCA_002436185.1_00480_MMP06450768	Phosphonopyruvate hydrolase [mmp_id=MMP06450768] [mmp_db=mardb]	7.07019E-42
MAD52059.1_MMP07619323	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619323] [mmp_db=mardb]	7.07019E-42
GCA_002346665.1_00711_MMP06452558	Phosphonopyruvate hydrolase [mmp_id=MMP06452558] [mmp_db=mardb]	7.78165E-42

GCA_000384875.1_00772_MMP02440973	Phosphonopyruvate hydrolase [mmp_id=MMP02440973] [mmp_db=mardb]	9.3988E-42
MBP80234.1_MMP07618671	phosphoenolpyruvate mutase [Deltaproteobacteria bacterium] [mmp_id=MMP07618671] [mmp_db=mardb]	1.03066E-41
MBL51963.1_MMP07619401	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619401] [mmp_db=mardb]	1.03536E-41
MBN42917.1_MMP07618318	phosphoenolpyruvate mutase [Alphaproteobacteria bacterium] [mmp_id=MMP07618318] [mmp_db=mardb]	1.30476E-41
MBH44265.1_MMP07619132	phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619132] [mmp_db=mardb]	1.33697E-41
MAZ47818.1_MMP07619209	phosphoenolpyruvate mutase [Halobacteriovoraceae bacterium] [mmp_id=MMP07619209] [mmp_db=mardb]	1.45724E-41
MMP492357_378316	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]	1.51295E-41
MMP492357_378315	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	1.51295E-41
WP_015871974.1_MMP02469763	phosphoenolpyruvate mutase [Edwardsiella ictaluri]	1.67696E-41

	[mmp_id=MMP02469763] [mmp_db=mardb]	
WP_015871974.1_MMP03761503	phosphoenolpyruvate mutase [Edwardsiella ictaluri] [mmp_id=MMP03761503] [mmp_db=mardb]	1.67696E-41
KMQ77667.1_MMP03755268	phosphoenolpyruvate phosphomutase [Edwardsiella ictaluri] [mmp_id=MMP03755268] [mmp_db=mardb]	1.67696E-41
MMP492357_185032	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]	1.73628E-41
MMP492357_185031	[UniRef50=UniRef50_V5C6M9,Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	1.73628E-41
MMP492357_160277	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	2.00254E-41
MMP492357_160276	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	2.00254E-41
MAD12485.1_MMP07618948	phosphoenolpyruvate mutase [Flavobacteriaceae bacterium] [mmp_id=MMP07618948] [mmp_db=mardb]	2.32637E-41
MMP491463_239882	[UniRef50=UniRef50_B9JNZ3,Cluster: Phosphoenolpyruvate phosphomutase protein] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain]	2.396E-41

	[Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	
MMP491463_295256	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Priam=1.6.99.3] [mmp_id=MMP491463] [mmp_db=marcat]	2.54538E-41
MMP491463_295255	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	2.54538E-41
MAQ43600.1_MMP07619369	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619369] [mmp_db=mardb]	2.79048E-41
AUV80813.1_MMP08381148	carboxyvinyl-carboxyphosphonate phosphorylmutase [Salinigranum rubrum] [mmp_id=MMP08381148] [mmp_db=marref]	2.83503E-41
GCA_003245675.1_00800_MMP09287973	Phosphonopyruvate hydrolase [mmp_id=MMP09287973] [mmp_db=mardb]	2.87795E-41
MMP491463_208017	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Priam=2.7.7.39] [mmp_id=MMP491463] [mmp_db=marcat]	2.90042E-41
MMP491463_208016	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	2.90042E-41
MBO22452.1_MMP07619970	phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619970] [mmp_db=mardb]	3.04266E-41
WP_046426931.1_MMP03445833	hypothetical protein [Streptomyces malaysiense] [mmp_id=MMP03445833] [mmp_db=mardb]	4.1797E-41

WP_028880465.1_MMP02743896	phosphoenolpyruvate mutase [Terasakiella pusilla] [mmp_id=MMP02743896] [mmp_db=mardb]	4.24156E-41
MBO89973.1_MMP07620258	carboxyvinyl-carboxyphosphonate phosphorylmutase [Rickettsiales bacterium] [mmp_id=MMP07620258] [mmp_db=mardb]	4.37237E-41
MBE09768.1_MMP07619979	carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619979] [mmp_db=mardb]	4.37237E-41
MBE09208.1_MMP07619979	carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619979] [mmp_db=mardb]	4.37237E-41
WP_067344657.1_MMP4029003	phosphoenolpyruvate mutase [Marinomonas spartinae] [mmp_id=MMP4029003] [mmp_db=mardb]	4.46374E-41
MAV76723.1_MMP07619346	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619346] [mmp_db=mardb]	4.53247E-41
GCA_003210075.1_01187_MMP08886458	Phosphonopyruvate hydrolase [mmp_id=MMP08886458] [mmp_db=mardb]	4.64943E-41
GCA_002375965.1_00799_MMP06455818	2,3-dimethylmalate lyase [mmp_id=MMP06455818] [mmp_db=mardb]	6.47714E-41
GCA_002376045.1_01806_MMP06451006	2,3-dimethylmalate lyase [mmp_id=MMP06451006] [mmp_db=mardb]	6.47714E-41
MMP494431_74556	[UniRef50=UniRef50_A0A160V7K5,Cluster: Methylisocitrate lyase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.8.23] [mmp_id=MMP494431] [mmp_db=marcat]	7.58269E-41

GCA_003278125.1_00325_MMP08886282	Phosphonopyruvate hydrolase [mmp_id=MMP08886282] [mmp_db=mardb]	7.78273E-41
GCA_002413545.1_01794_MMP06456309	Carboxyvinyl-carboxyphosphonate phosphorylmutase [mmp_id=MMP06456309] [mmp_db=mardb]	7.82597E-41
MAX59592.1_MMP07618500	carboxyvinyl-carboxyphosphonate phosphorylmutase [Chloroflexi bacterium] [mmp_id=MMP07618500] [mmp_db=mardb]	8.68869E-41
SMD37771.1_MMP04488029	phosphoenolpyruvate mutase [Reichenbachiella faecimaris] [mmp_id=MMP04488029] [mmp_db=mardb]	9.89243E-41
MBQ11747.1_MMP07619689	carboxyvinyl-carboxyphosphonate phosphorylmutase [Planctomyces sp.] [mmp_id=MMP07619689] [mmp_db=mardb]	1.08934E-40
MAD58100.1_MMP07619788	phosphoenolpyruvate mutase [Porticoccus sp.] [mmp_id=MMP07619788] [mmp_db=mardb]	1.57229E-40
MAJ57951.1_MMP07619618	phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619618] [mmp_db=mardb]	1.63643E-40
GCA_003211655.1_00530_MMP08886573	Phosphonopyruvate hydrolase [mmp_id=MMP08886573] [mmp_db=mardb]	3.31679E-40
MBB70843.1_MMP07619294	isocitrate lyase [Legionellales bacterium] [mmp_id=MMP07619294] [mmp_db=mardb]	3.31936E-40
MBO43289.1_MMP07619972	carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619972] [mmp_db=mardb]	3.40433E-40

GCF_000384895.1_00821_MMP02440972	Phosphonopyruvate hydrolase [mmp_id=MMP02440972] [mmp_db=mardb]	4.09323E-40
GCF_000291925.1_00190_MMP02470491	Phosphonopyruvate hydrolase [mmp_id=MMP02470491] [mmp_db=mardb]	4.23314E-40
MBJ62004.1_MMP07618901	hypothetical protein CMB57_02000 [Euryarchaeota archaeon] [mmp_id=MMP07618901] [mmp_db=mardb]	4.9228E-40
MAV76785.1_MMP07619346	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619346] [mmp_db=mardb]	5.06087E-40
MMP492357_275649	[UniRef50=UniRef50_A0A0S4RUB7,Cluster: Methylisocitrate lyase] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]	5.0827E-40
MMP492357_275648	[UniRef50=UniRef50_A0A0S4RUB7,Cluster: Methylisocitrate lyase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	5.0827E-40
GCA_003212695.1_00172_MMP08886495	Phosphonopyruvate hydrolase [mmp_id=MMP08886495] [mmp_db=mardb]	5.45731E-40
AGM40194.1_MMP02603230	isocitrate lyase and phosphorylmutase [Spiribacter salinus M19-40] [mmp_id=MMP02603230] [mmp_db=marref]	5.83795E-40
WP_030276407.1_MMP02645363	phosphoenolpyruvate phosphomutase [Streptomyces sp. NRRL B-24484] [mmp_id=MMP02645363] [mmp_db=mardb]	6.20143E-40
WP_040540712.1_MMP02436226	phosphoenolpyruvate mutase [marine gamma proteobacterium HTCC2148] [mmp_id=MMP02436226] [mmp_db=mardb]	6.29476E-40
PPR72984.1_MMP07286201	2,3-dimethylmalate lyase [Alphaproteobacteria bacterium MarineAlpha4_Bin1]	6.44361E-40

	[mmp_id=MMP07286201] [mmp_db=mardb]	
PPR24431.1_MMP07286232	2,3-dimethylmalate lyase [Alphaproteobacteria bacterium MarineAlpha10_Bin3] [mmp_id=MMP07286232] [mmp_db=mardb]	6.44361E-40
MMP490065_44831	[UniRef50=UniRef50_A0A160V7K5,Cluster: Methylisocitrate lyase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=4.1.3.30] [mmp_id=MMP490065] [mmp_db=marcat]	8.8974E-40
MMP491463_364497	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP491463] [mmp_db=marcat]	1.1129E-39
MMP491463_364496	[UniRef50=UniRef50_B0NP48,Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813,Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	1.1129E-39
MBE90608.1_MMP07620020	carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620020] [mmp_db=mardb]	1.19399E-39
MAJ23737.1_MMP07619619	phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619619] [mmp_db=mardb]	1.94035E-39
MMP491463_310111	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.7.14] [mmp_id=MMP491463] [mmp_db=marcat]	1.99665E-39
MMP491463_310110	[UniRef50=UniRef50_UPI00049790D9,Cluster: phosphoenolpyruvate	1.99665E-39

	phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]	
MMP492357_79768	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]	2.33759E-39
MMP492357_79767	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR004821,Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	2.33759E-39
WP_116450246.1_MMP08954526	oxaloacetate decarboxylase [Blastococcus litoris] [mmp_id=MMP08954526] [mmp_db=mardb]	2.42274E-39
GCA_003212655.1_01295_MMP08886498	Phosphonopyruvate hydrolase [mmp_id=MMP08886498] [mmp_db=mardb]	2.46066E-39
MAJ86251.1_MMP07618432	phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07618432] [mmp_db=mardb]	3.50175E-39
WP_034126960.1_MMP05967369	MULTISPECIES: phosphoenolpyruvate mutase [Pseudomonas] [mmp_id=MMP05967369] [mmp_db=mardb]	4.56278E-39
WP_019818012.1_MMP06140193	MULTISPECIES: phosphoenolpyruvate mutase [Pseudomonas] [mmp_id=MMP06140193] [mmp_db=mardb]	4.6106E-39
GCA_003216835.1_01727_MMP08886316	Phosphonopyruvate hydrolase [mmp_id=MMP08886316] [mmp_db=mardb]	4.65362E-39
PCI17085.1_MMP07568951	carboxyvinyl-carboxyphosphonate phosphorylmutase [SAR202 cluster bacterium] [mmp_id=MMP07568951] [mmp_db=mardb]	5.18631E-39

PKB78155.1_MMP06121366	carboxyvinyl-carboxyphosphonate phosphorylmutase [SAR202 cluster bacterium MP-SInd-SRR3963457-G2] [mmp_id=MMP06121366] [mmp_db=mardb]	5.63727E-39
WP_027983379.1_MMP02440745	carboxyvinyl-carboxyphosphonate phosphorylmutase [delta proteobacterium PSCGC 5342] [mmp_id=MMP02440745] [mmp_db=mardb]	8.17101E-39
MBA94038.1_MMP07619391	phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619391] [mmp_db=mardb]	8.21891E-39
WP_050713289.1_MMP3245995	hypothetical protein [Halomonas sp. R57-5] [mmp_id=MMP3245995] [mmp_db=marref]	1.1716E-38
WP_050713289.1_MMP02744764	hypothetical protein [Halomonas sp. TG39a] [mmp_id=MMP02744764] [mmp_db=mardb]	1.1716E-38
MAZ63906.1_MMP07618640	carboxyvinyl-carboxyphosphonate phosphorylmutase [Dehalococcoidia bacterium] [mmp_id=MMP07618640] [mmp_db=mardb]	1.21866E-38
MMP492357_2904	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]	1.30914E-38
MMP492357_2903	[UniRef50=UniRef50_A0A0G1W6N0,Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR014729,Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]	1.30914E-38

Supplementary Table S4.4: Color coded legend for sequence similarity network (Figure 13 in main text).

Node Color	Description
	WP_018735391.1_MMP02441365 hypothetical protein [Salinispora pacifica] [mmp_id=MMP02441365] [mmp_db=mardb]
	WP_066506643.1_MMP04316851 phosphoenolpyruvate phosphomutase [Clostridiales bacterium MCWD3] [mmp_id=MMP04316851] [mmp_db=mardb] MBL76896.1_MMP07618508 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb] MBL77230.1_MMP07618508 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb] MMP494431_91392 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat] MMP490065_168674 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat] MMP494431_194244 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat] MMP490065_229543 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat] GBH35141.1_MMP00113975 hypothetical protein NZNM25_19320 [Candidatus Nitrosopumilus sp. NM25] [mmp_id=MMP00113975] [mmp_db=mardb] MAR68904.1_MMP07619518 phosphoenolpyruvate phosphomutase [Nitrospina sp.] [mmp_id=MMP07619518] [mmp_db=mardb] ABX12056.1_MMP00000032 putative phosphoenolpyruvate phosphomutase [Nitrosopumilus maritimus SCM1] [mmp_id=MMP00000032] [mmp_db=marref] MMP494431_37872 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat] MMP490065_209202 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat] ARF65919.1_MMP06323808 phosphoenolpyruvate phosphomutase [Streptomyces violaceoruber] [mmp_id=MMP06323808] [mmp_db=marref] PIG51557.1_MMP07626382 phosphoenolpyruvate mutase [Verrucosispora sp. CNZ293] [mmp_id=MMP07626382] [mmp_db=mardb]

MMP494431_85898 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]
MMP494431_1390 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]
GCA_003235875.1_00167_MMP09287831 Phosphonopyruvate hydrolase [mmp_id=MMP09287831] [mmp_db=mardb]
RAS38621.1_MMP05720407 phosphoenolpyruvate mutase [Thaumarchaeota archaeon SCGC AC-337_F14] [mmp_id=MMP05720407] [mmp_db=mardb]
GCA_000230485.1_00568_MMP02954312 Phosphonopyruvate hydrolase [mmp_id=MMP02954312] [mmp_db=mardb]
MBR73495.1_MMP07618617 phosphoenolpyruvate phosphomutase [Dehalococcoidaceae bacterium] [mmp_id=MMP07618617] [mmp_db=mardb]
MAS51187.1_MMP07618457 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618457] [mmp_db=mardb]
MMP490065_5965 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]
MAR14796.1_MMP07618462 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618462] [mmp_db=mardb]
GCA_000230485.1_01502_MMP02954312 Phosphonopyruvate hydrolase [mmp_id=MMP02954312] [mmp_db=mardb]
GCA_001674955.1_00721_MMP03840809 Phosphonopyruvate hydrolase [mmp_id=MMP03840809] [mmp_db=mardb]
MMP494431_228096 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]
MBL77037.1_MMP07618508 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]
MMP494431_197159 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]
MBL75307.1_MMP07618508 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]
WP_007403095.1_MMP02471010 phosphoenolpyruvate synthase [Candidatus Nitrosoarchaeum limnia] [mmp_id=MMP02471010] [mmp_db=marref]
MBA30896.1_MMP07618650 phosphoenolpyruvate phosphomutase [Dehalococcoidia bacterium] [mmp_id=MMP07618650] [mmp_db=mardb]

EGG41199.1_MMP02471010 putative phosphoenolpyruvate phosphomutase [Candidatus Nitrosoarchaeum limnia SFB1] [mmp_id=MMP02471010] [mmp_db=mardb]

PXF27507.1_MMP08159281 phosphoenolpyruvate phosphomutase [Thaumarchaeota archaeon] [mmp_id=MMP08159281] [mmp_db=mardb]
MMP490065_23702 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]
MAR36777.1_MMP07618460 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618460] [mmp_db=mardb]
WP_049564768.1_MMP03468208 phosphoenolpyruvate phosphomutase [Streptomyces sp. SBT349] [mmp_id=MMP03468208] [mmp_db=mardb]
MMP494431_137904 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]
MMP494431_182150 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]
MMP494431_93748 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]
MBR74309.1_MMP07618617 phosphoenolpyruvate phosphomutase [Dehalococcoidaceae bacterium] [mmp_id=MMP07618617] [mmp_db=mardb]
MAS50419.1_MMP07618457 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618457] [mmp_db=mardb]
MMP490065_74875 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]
KFM21051.1_MMP02869637 Phosphoenolpyruvate phosphomutase protein [Marine Group I thaumarchaeote SCGC AAA799-B03] [mmp_id=MMP02869637] [mmp_db=mardb]
MMP490065_237234 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP490065] [mmp_db=marcat]
MBN18910.1_MMP07618503 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618503] [mmp_db=mardb]
MMP490065_195017 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]
MMP494431_133827 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]

MMP494431_249509 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=3.11.1.3] [mmp_id=MMP494431] [mmp_db=marcat]
MBL76962.1_MMP07618508 phosphoenolpyruvate phosphomutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]
MMP490065_193497 [UniRef50=UniRef50_A0A075H9G4Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]
SMD37771.1_MMP04488029 phosphoenolpyruvate mutase [Reichenbachiella faecimaris] [mmp_id=MMP04488029] [mmp_db=mardb]
GCA_003278125.1_00325_MMP08886282 Phosphonopyruvate hydrolase [mmp_id=MMP08886282] [mmp_db=mardb]
MAJ57951.1_MMP07619618 phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619618] [mmp_db=mardb]
WP_028880465.1_MMP02743896 phosphoenolpyruvate mutase [Terasakiella pusilla] [mmp_id=MMP02743896] [mmp_db=mardb]
WP_067344657.1_MMP4029003 phosphoenolpyruvate mutase [Marinomonas spartinae] [mmp_id=MMP4029003] [mmp_db=mardb]
MAV76723.1_MMP07619346 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619346] [mmp_db=mardb]
GCA_003210075.1_01187_MMP08886458 Phosphonopyruvate hydrolase [mmp_id=MMP08886458] [mmp_db=mardb]

MMP491463_295255 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]

MMP491463_239882 [UniRef50=UniRef50_B9JNZ3Cluster: Phosphoenolpyruvate phosphomutase protein] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]
MBO22452.1_MMP07619970 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619970] [mmp_db=mardb]
MAQ43600.1_MMP07619369 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619369] [mmp_db=mardb]

MMP492357_378315 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP491463_208017 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Priam=2.7.7.39] [mmp_id=MMP491463] [mmp_db=marcat]

MMP491463_208016 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]
MAD12485.1_MMP07618948 phosphoenolpyruvate mutase [Flavobacteriaceae bacterium] [mmp_id=MMP07618948] [mmp_db=mardb]

MMP492357_160277 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]
MMP492357_160276 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP491463_295256 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Priam=1.6.99.3] [mmp_id=MMP491463] [mmp_db=marcat]
WP_015871974.1_MMP02469763 phosphoenolpyruvate mutase [Edwardsiella ictaluri] [mmp_id=MMP02469763] [mmp_db=mardb]
MBN42917.1_MMP07618318 phosphoenolpyruvate mutase [Alphaproteobacteria bacterium] [mmp_id=MMP07618318] [mmp_db=mardb]

MBH44265.1_MMP07619132 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619132] [mmp_db=mardb]
GCA_000384875.1_00772_MMP02440973 Phosphonopyruvate hydrolase [mmp_id=MMP02440973] [mmp_db=mardb]

MMP492357_378316 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]
KMQ77667.1_MMP03755268 phosphoenolpyruvate phosphomutase [Edwardsiella ictaluri] [mmp_id=MMP03755268] [mmp_db=mardb]
WP_015871974.1_MMP03761503 phosphoenolpyruvate mutase [Edwardsiella ictaluri] [mmp_id=MMP03761503] [mmp_db=mardb]
MBL51963.1_MMP07619401 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619401] [mmp_db=mardb]
MAD52059.1_MMP07619323 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619323] [mmp_db=mardb]
GCA_002436185.1_00480_MMP06450768 Phosphonopyruvate hydrolase [mmp_id=MMP06450768] [mmp_db=mardb]
GCA_002471885.1_01119_MMP06451807 Phosphonopyruvate hydrolase [mmp_id=MMP06451807] [mmp_db=mardb]

MMP492357_185031 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_185032 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_181017 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_181016 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
MBA65651.1_MMP07619399 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619399] [mmp_db=mardb]
GCA_002346665.1_00711_MMP06452558 Phosphonopyruvate hydrolase [mmp_id=MMP06452558] [mmp_db=mardb]
MBP80234.1_MMP07618671 phosphoenolpyruvate mutase [Deltaproteobacteria bacterium] [mmp_id=MMP07618671] [mmp_db=mardb]
MMP492357_219960 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]
MMP492357_219959 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
AAW85744.1_MMP02604303 phosphoenolpyruvate phosphomutase [Vibrio fischeri ES114] [mmp_id=MMP02604303] [mmp_db=marref]
GCA_002179675.1_03385_MMP06226649 Phosphonopyruvate hydrolase [mmp_id=MMP06226649] [mmp_db=mardb]
GCA_003213855.1_00634_MMP08886424 Phosphonopyruvate hydrolase [mmp_id=MMP08886424] [mmp_db=mardb]
WP_019882558.1_MMP02261247 MULTISPECIES: phosphoenolpyruvate mutase [Methylophilus] [mmp_id=MMP02261247] [mmp_db=mardb]
MBT05342.1_MMP07619976 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619976] [mmp_db=mardb]
GCA_002377485.1_00039_MMP06450560 Phosphonopyruvate hydrolase [mmp_id=MMP06450560] [mmp_db=mardb]
WP_011261863.1_MMP05449650 phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP05449650] [mmp_db=mardb]
WP_011261863.1_MMP03652539 phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP03652539] [mmp_db=mardb]
WP_005419153.1_MMP02470712 phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP02470712] [mmp_db=mardb]
WP_005419153.1_MMP04520027 phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP04520027] [mmp_db=mardb]
GCA_002473315.1_00356_MMP06452437 Phosphonopyruvate hydrolase [mmp_id=MMP06452437] [mmp_db=mardb]
GCA_002471275.1_04526_MMP06450620 Phosphonopyruvate hydrolase [mmp_id=MMP06450620] [mmp_db=mardb]
WP_063662137.1_MMP04519656 phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP04519656] [mmp_db=mardb]
GCA_002336125.1_02319_MMP06456076 Phosphonopyruvate hydrolase [mmp_id=MMP06456076] [mmp_db=mardb]

WP_032520302.1_MMP02769562 phosphoenolpyruvate mutase [Prochlorococcus marinus] [mmp_id=MMP02769562] [mmp_db=mardb]
WP_072054505.1_MMP3355991 phosphoenolpyruvate mutase [Aliivibrio fischeri] [mmp_id=MMP3355991] [mmp_db=mardb]
GCA_002179655.1_03643_MMP06226650 Phosphonopyruvate hydrolase [mmp_id=MMP06226650] [mmp_db=mardb]
GCA_003282985.1_00886_MMP08886037 Phosphonopyruvate hydrolase [mmp_id=MMP08886037] [mmp_db=mardb]
GCA_003281565.1_01066_MMP08886145 Phosphonopyruvate hydrolase [mmp_id=MMP08886145] [mmp_db=mardb]
GCA_003213815.1_01110_MMP08886426 Phosphonopyruvate hydrolase [mmp_id=MMP08886426] [mmp_db=mardb]
GCA_002341165.1_01494_MMP06457326 Phosphonopyruvate hydrolase [mmp_id=MMP06457326] [mmp_db=mardb]
SFG78580.1_MMP05216175 phosphoenolpyruvate phosphomutase [Neptunomonas qingdaonensis] [mmp_id=MMP05216175] [mmp_db=mardb]
MMP492357_327676 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]
MMP492357_327675 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
GCA_002470485.1_00859_MMP06451384 Phosphonopyruvate hydrolase [mmp_id=MMP06451384] [mmp_db=mardb]
MAW17907.1_MMP07619884 phosphoenolpyruvate mutase [Rhodobacteraceae bacterium] [mmp_id=MMP07619884] [mmp_db=mardb]
WP_035250982.1_MMP02742681 phosphoenolpyruvate mutase [Actibacterium atlanticum] [mmp_id=MMP02742681] [mmp_db=mardb]
GCA_003281725.1_00410_MMP08886127 Phosphonopyruvate hydrolase [mmp_id=MMP08886127] [mmp_db=mardb]
WP_023475077.1_MMP02469983 phosphoenolpyruvate mutase [Burkholderia cenocepacia] [mmp_id=MMP02469983] [mmp_db=mardb]
GCF_000438945.1_01691_MMP02440705 Phosphonopyruvate hydrolase [mmp_id=MMP02440705] [mmp_db=mardb]
WP_020403838.1_MMP02440412 phosphoenolpyruvate mutase [Gracilimonas tropica] [mmp_id=MMP02440412] [mmp_db=mardb]
GCA_003215185.1_00709_MMP08886380 Phosphonopyruvate hydrolase [mmp_id=MMP08886380] [mmp_db=mardb]
WP_029455268.1_MMP02841150 phosphoenolpyruvate mutase [Candidatus Pelagibacter ubique] [mmp_id=MMP02841150] [mmp_db=mardb]
ODS34541.1_MMP05368398 Phosphonopyruvate hydrolase [Candidatus Scalindua rubra] [mmp_id=MMP05368398] [mmp_db=mardb]
MBQ34353.1_MMP07619376 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619376] [mmp_db=mardb]
GCA_002501165.1_00652_MMP06457363 Phosphonopyruvate hydrolase [mmp_id=MMP06457363] [mmp_db=mardb]

MAD11214.1_MMP07618948 phosphoenolpyruvate mutase [Flavobacteriaceae bacterium] [mmp_id=MMP07618948] [mmp_db=mardb]

ADV46661.1_MMP00713578 phosphoenolpyruvate phosphomutase [Nitratifractor salsuginis DSM 16511] [mmp_id=MMP00713578] [mmp_db=marref]
GCA_003211435.1_00667_MMP08886590 Phosphonopyruvate hydrolase [mmp_id=MMP08886590] [mmp_db=mardb]
GCA_003212275.1_00314_MMP08886523 Phosphonopyruvate hydrolase [mmp_id=MMP08886523] [mmp_db=mardb]

MBS83214.1_MMP07619109 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619109] [mmp_db=mardb]
GCA_003211015.1_01168_MMP08886617 Phosphonopyruvate hydrolase [mmp_id=MMP08886617] [mmp_db=mardb]

MMP492357_273580 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_273581 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]

MMP494431_144293 [UniRef50=UniRef50_A0A0G1W6N0Cluster: Phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]

MMP494431_144294 [UniRef50=UniRef50_A0A0G1W6N0Cluster: Phosphoenolpyruvate phosphomutase] [Priam=2.7.7.14] [mmp_id=MMP494431] [mmp_db=marcat]
MAH97793.1_MMP07618756 phosphoenolpyruvate mutase [Euryarchaeota archaeon] [mmp_id=MMP07618756] [mmp_db=mardb]

MMP492357_25364 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_25363 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP490065_154054 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=5.4.2.9] [mmp_id=MMP490065] [mmp_db=marcat]
MAJ22838.1_MMP07619619 phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619619] [mmp_db=mardb]
MAW74576.1_MMP07619353 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619353] [mmp_db=mardb]

MMP490065_154055 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=2.7.7.39] [mmp_id=MMP490065] [mmp_db=marcat]

MMP494431_95860 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]
GCA_003281605.1_01756_MMP08886141 Phosphonopyruvate hydrolase [mmp_id=MMP08886141] [mmp_db=mardb]

MMP494431_95861 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=2.7.7.39] [mmp_id=MMP494431] [mmp_db=marcat]

BAN03327.1_MMP00060985 putative phosphoenolpyruvate phosphomutase [Ilumatobacter coccineus YM16-304] [mmp_id=MMP00060985] [mmp_db=marref]

MMP491463_347518 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]

MMP492357_352173 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP491463_347519 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=2.7.7.2] [mmp_id=MMP491463] [mmp_db=marcat]
MAG26910.1_MMP07619591 phosphoenolpyruvate mutase [Candidatus Pacearchaeota archaeon] [mmp_id=MMP07619591] [mmp_db=mardb]
MBS86051.1_MMP07619227 phosphoenolpyruvate mutase [Candidatus Heimdallarchaeota archaeon] [mmp_id=MMP07619227] [mmp_db=mardb]
MAG47670.1_MMP07618362 phosphoenolpyruvate mutase [archaeon] [mmp_id=MMP07618362] [mmp_db=mardb]
WP_101758417.1_MMP4467369 phosphoenolpyruvate mutase [Oceanicoccus sp. KOV_DT_Chl] [mmp_id=MMP4467369] [mmp_db=mardb]
GCA_003232435.1_01571_MMP09240056 Phosphonopyruvate hydrolase [mmp_id=MMP09240056] [mmp_db=mardb]
GBL05756.1_MMP00115722 phosphoenolpyruvate phosphomutase [Glaciecola sp. KUL10] [mmp_id=MMP00115722] [mmp_db=mardb]
WP_034225084.1_MMP03004373 phosphoenolpyruvate mutase [Arenimonas donghaensis] [mmp_id=MMP03004373] [mmp_db=mardb]
MMP491463_358251 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]

MMP491463_358252 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=2.7.7.14] [mmp_id=MMP491463] [mmp_db=marcat]
SEA64967.1_MMP05660420 phosphoenolpyruvate mutase [Desulfuromusa kysingii] [mmp_id=MMP05660420] [mmp_db=mardb]
ARN73867.1_MMP06075352 phosphoenolpyruvate mutase [Oceanicoccus sagamiensis] [mmp_id=MMP06075352] [mmp_db=marref]
WP_111976523.1_MMP09469620 phosphoenolpyruvate mutase [Catenovulum sp. RQJ05] [mmp_id=MMP09469620] [mmp_db=mardb]
MBO51482.1_MMP07619713 phosphoenolpyruvate mutase [Planctomycetaceae bacterium] [mmp_id=MMP07619713] [mmp_db=mardb]
GCA_003229735.1_00663_MMP09240163 Phosphonopyruvate hydrolase [mmp_id=MMP09240163] [mmp_db=mardb]
GCA_003213695.1_00213_MMP08886429 Phosphonopyruvate hydrolase [mmp_id=MMP08886429] [mmp_db=mardb]
GCA_003232445.1_02328_MMP09240055 Phosphonopyruvate hydrolase [mmp_id=MMP09240055] [mmp_db=mardb]
WP_103439162.1_MMP08100003 phosphoenolpyruvate mutase [Arenibacter hampyeongensis] [mmp_id=MMP08100003] [mmp_db=mardb]
GCA_003282145.1_00604_MMP08886084 Phosphonopyruvate hydrolase [mmp_id=MMP08886084] [mmp_db=mardb]

MMP492357_155310 [UniRef50=UniRef50_G0ENI0Cluster: Putative phosphoenolpyruvate phosphomutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
MAG39062.1_MMP07620234 phosphoenolpyruvate mutase [Candidatus Woesearchaeota archaeon] [mmp_id=MMP07620234] [mmp_db=mardb]
WP_067553003.1_MMP04332636 phosphoenolpyruvate mutase [Oceanibaculum pacificum] [mmp_id=MMP04332636] [mmp_db=mardb]

MMP492357_176301 [UniRef50=UniRef50_A0A0G1W6N0Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]
MMP492357_270843 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_176300 [UniRef50=UniRef50_A0A0G1W6N0Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
GCA_003281795.1_00664_MMP08886115 Phosphonopyruvate hydrolase [mmp_id=MMP08886115] [mmp_db=mardb]
WP_068379690.1_MMP04487161 phosphoenolpyruvate mutase [Paraglaciecola sp. S66] [mmp_id=MMP04487161] [mmp_db=mardb]
MMP492357_270844 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.7.39] [mmp_id=MMP492357] [mmp_db=marcat]

WP_017212535.1_MMP08107721 phosphoenolpyruvate mutase [Clostridium beijerinckii] [mmp_id=MMP08107721] [mmp_db=mardb]

MMP492357_283672 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]

WP_009828628.1_MMP02436225 phosphoenolpyruvate mutase [Rhodobacteraceae bacterium HTCC2083] [mmp_id=MMP02436225] [mmp_db=mardb]

MMP492357_283671 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Priam=3.11.1.3] [mmp_id=MMP492357] [mmp_db=marcat]
MBA94038.1_MMP07619391 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619391] [mmp_db=mardb]
MBN22549.1_MMP07618423 phosphoenolpyruvate mutase [Bdellovibrionaceae bacterium] [mmp_id=MMP07618423] [mmp_db=mardb]
WP_019818012.1_MMP06140193 MULTISPECIES: phosphoenolpyruvate mutase [Pseudomonas] [mmp_id=MMP06140193] [mmp_db=mardb]
WP_034126960.1_MMP05967369 MULTISPECIES: phosphoenolpyruvate mutase [Pseudomonas] [mmp_id=MMP05967369] [mmp_db=mardb]
PSM55803.1_MMP08493996 phosphoenolpyruvate mutase [Clostridium diolis] [mmp_id=MMP08493996] [mmp_db=mardb]
MMP491463_91824 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]
MAG50420.1_MMP07618366 phosphoenolpyruvate mutase [archaeon] [mmp_id=MMP07618366] [mmp_db=mardb]
GCA_001626765.1_04224_MMP04534590 Phosphonopyruvate hydrolase [mmp_id=MMP04534590] [mmp_db=mardb]
MBL19935.1_MMP07619026 phosphoenolpyruvate mutase [Flavobacteriaceae bacterium] [mmp_id=MMP07619026] [mmp_db=mardb]
MMP492357_71496 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]
MMP492357_71495 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=3.11.1.3] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_2904 [UniRef50=UniRef50_A0A0G1W6N0Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.1.167] [mmp_id=MMP492357] [mmp_db=marcat]
PKH00836.1_MMP08125768 phosphoenolpyruvate mutase [Paraglaciecola sp. MB-3u-78] [mmp_id=MMP08125768] [mmp_db=mardb]
MMP491463_91825 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=2.7.1.167] [mmp_id=MMP491463] [mmp_db=marcat]

MMP491463_310110 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]
MMP492357_293074 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
MMP491463_310111 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.7.14] [mmp_id=MMP491463] [mmp_db=marcat]
MMP492357_293075 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]

MMP491463_364496 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]

MMP492357_310524 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
GCA_003212695.1_00172_MMP08886495 Phosphonopyruvate hydrolase [mmp_id=MMP08886495] [mmp_db=mardb]

MMP491463_364497 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP491463] [mmp_db=marcat]

MMP492357_310525 [UniRef50=UniRef50_V5C6M9Cluster: Phosphonopyruvate hydrolase PphA] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]
MAJ86251.1_MMP07618432 phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07618432] [mmp_db=mardb]
AVP54233.1_MMP08398331 phosphoenolpyruvate mutase [Clostridium tetani] [mmp_id=MMP08398331] [mmp_db=marref]
GCA_003216835.1_01727_MMP08886316 Phosphonopyruvate hydrolase [mmp_id=MMP08886316] [mmp_db=mardb]
GCA_003280295.1_01643_MMP08886241 Phosphonopyruvate hydrolase [mmp_id=MMP08886241] [mmp_db=mardb]

MMP492357_79768 [UniRef50=UniRef50_A0A0G1W6N0Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_79767 [UniRef50=UniRef50_A0A0G1W6N0Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_311977 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
GCA_003212655.1_01295_MMP08886498 Phosphonopyruvate hydrolase [mmp_id=MMP08886498] [mmp_db=mardb]
MMP492357_311978 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]
MMP492357_221511 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
MAJ23737.1_MMP07619619 phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619619] [mmp_db=mardb]
MMP492357_221512 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR004821Cytidyltransferase-like domain] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]
GCA_001655195.1_02715_MMP03763493 Phosphonopyruvate hydrolase [mmp_id=MMP03763493] [mmp_db=mardb]

MMP492357_275649 [UniRef50=UniRef50_A0A0S4RUB7Cluster: Methylisocitrate lyase] [Priam=2.7.7.14] [mmp_id=MMP492357] [mmp_db=marcat]

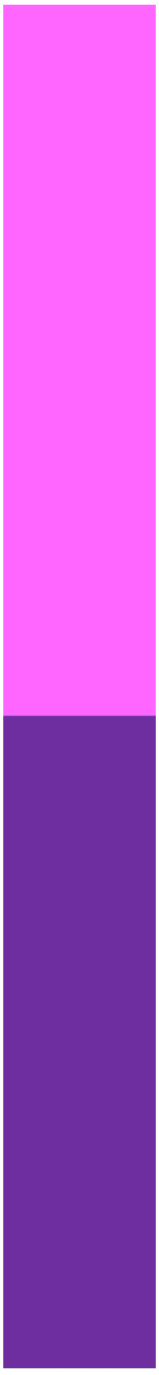
MMP492357_352174 [UniRef50=UniRef50_B0NP48Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR014729Rossmann-like alpha/beta/alpha sandwich fold] [Priam=2.7.7.39] [mmp_id=MMP492357] [mmp_db=marcat]
GCF_000291925.1_00190_MMP02470491 Phosphonopyruvate hydrolase [mmp_id=MMP02470491] [mmp_db=mardb]
GCA_003233015.1_00171_MMP09240024 Phosphonopyruvate hydrolase [mmp_id=MMP09240024] [mmp_db=mardb]
GCA_001180265.1_00184_MMP3368574 Phosphonopyruvate hydrolase [mmp_id=MMP3368574] [mmp_db=mardb]
GCA_003211655.1_00530_MMP08886573 Phosphonopyruvate hydrolase [mmp_id=MMP08886573] [mmp_db=mardb]
MAV05675.1_MMP07619615 phosphoenolpyruvate mutase [Candidatus Pelagibacter sp.] [mmp_id=MMP07619615] [mmp_db=mardb]
GCF_000384895.1_00821_MMP02440972 Phosphonopyruvate hydrolase [mmp_id=MMP02440972] [mmp_db=mardb]
MAD58100.1_MMP07619788 phosphoenolpyruvate mutase [Porticoccus sp.] [mmp_id=MMP07619788] [mmp_db=mardb]
GCA_003233155.1_00187_MMP09240016 Phosphonopyruvate hydrolase [mmp_id=MMP09240016] [mmp_db=mardb]
GCA_003233015.1_00740_MMP09240024 Phosphonopyruvate hydrolase [mmp_id=MMP09240024] [mmp_db=mardb]
SNS59778.1_MMP05421640 phosphoenolpyruvate mutase [Ekhidna lutea] [mmp_id=MMP05421640] [mmp_db=mardb]

SMF02759.1_MMP02745866 phosphoenolpyruvate mutase [Alteromonadaceae bacterium Bs31] [mmp_id=MMP02745866] [mmp_db=mardb]
MMP492357_208315 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
MAV76785.1_MMP07619346 phosphoenolpyruvate mutase [Candidatus Marinimicrobia bacterium] [mmp_id=MMP07619346] [mmp_db=mardb]
MAG69988.1_MMP07618216 phosphoenolpyruvate mutase [Acidobacteria bacterium] [mmp_id=MMP07618216] [mmp_db=mardb]
AVQ40888.1_MMP08398059 phosphoenolpyruvate mutase [Clostridium botulinum] [mmp_id=MMP08398059] [mmp_db=marref]
MMP492357_208316 [UniRef50=UniRef50_UPI00049790D9Cluster: phosphoenolpyruvate phosphomutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.2] [mmp_id=MMP492357] [mmp_db=marcat]
MBJ62004.1_MMP07618901 hypothetical protein CMB57_02000 [Euryarchaeota archaeon] [mmp_id=MMP07618901] [mmp_db=mardb]

WP_040540712.1_MMP02436226 phosphoenolpyruvate mutase [marine gamma proteobacterium HTCC2148] [mmp_id=MMP02436226] [mmp_db=mardb]
AVQ37341.1_MMP08397979 phosphoenolpyruvate mutase [Clostridium botulinum] [mmp_id=MMP08397979] [mmp_db=marref]

MMP492357_275648 [UniRef50=UniRef50_A0A0S4RUB7Cluster: Methylisocitrate lyase] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

GCA_003251795.1_01021_MMP09288020 Phosphonopyruvate hydrolase [mmp_id=MMP09288020] [mmp_db=mardb]
WP_071022580.1_MMP05904704 MULTISPECIES: phosphonopyruvate hydrolase [Cupriavidus] [mmp_id=MMP05904704] [mmp_db=marref]
GBE44636.1_MMP00081387 phosphonopyruvate hydrolase [bacterium BMS3Bbin10] [mmp_id=MMP00081387] [mmp_db=mardb]
WP_102772790.1_MMP04858689 phosphonopyruvate hydrolase [Achromobacter pulmonis] [mmp_id=MMP04858689] [mmp_db=mardb]
MAM74537.1_MMP07620173 phosphoenolpyruvate phosphomutase [Tistrella sp.] [mmp_id=MMP07620173] [mmp_db=mardb]
GCA_003247775.1_00216_MMP09287867 Phosphonopyruvate hydrolase [mmp_id=MMP09287867] [mmp_db=mardb]
MBA77101.1_MMP07620174 phosphoenolpyruvate phosphomutase [Tistrella sp.] [mmp_id=MMP07620174] [mmp_db=mardb]
MBL24275.1_MMP07619992 phosphonopyruvate hydrolase [Rhodospirillaceae bacterium] [mmp_id=MMP07619992] [mmp_db=mardb]
OYD24006.1_MMP06648058 phosphonopyruvate hydrolase [Oceanimonas baumannii] [mmp_id=MMP06648058] [mmp_db=mardb]
RAI28397.1_MMP07424250 phosphonopyruvate hydrolase [Rhodobium orientis] [mmp_id=MMP07424250] [mmp_db=mardb]



AWG33697.1_MMP06948887 phosphonopyruvate hydrolase [Alcaligenes aquatilis] [mmp_id=MMP06948887] [mmp_db=mardb]
GCA_003228955.1_00049_MMP09240132 Phosphonopyruvate hydrolase [mmp_id=MMP09240132] [mmp_db=mardb]
MAD38201.1_MMP07620172 phosphoenolpyruvate phosphomutase [Tistrella sp.] [mmp_id=MMP07620172] [mmp_db=mardb]
GCA_001657385.1_03693_MMP04939327 Phosphonopyruvate hydrolase [mmp_id=MMP04939327] [mmp_db=mardb]
WP_007668809.1_MMP02436095 phosphonopyruvate hydrolase [alpha proteobacterium BAL199] [mmp_id=MMP02436095] [mmp_db=mardb]
AFK56356.1_MMP02603926 phosphoenolpyruvate phosphomutase (plasmid) [Tistrella mobilis KA081020-065] [mmp_id=MMP02603926] [mmp_db=marref]
WP_062762299.1_MMP04332635 phosphoenolpyruvate phosphomutase [Tistrella mobilis] [mmp_id=MMP04332635] [mmp_db=mardb]
GCA_001657395.1_01944_MMP04939355 Phosphonopyruvate hydrolase [mmp_id=MMP04939355] [mmp_db=mardb]
WP_003808778.1_MMP02231461 MULTISPECIES: phosphoenolpyruvate phosphomutase [Bordetella] [mmp_id=MMP02231461] [mmp_db=mardb]
WP_003808778.1_MMP02231287 MULTISPECIES: phosphoenolpyruvate phosphomutase [Bordetella] [mmp_id=MMP02231287] [mmp_db=mardb]
MAG96192.1_MMP07619952 phosphonopyruvate hydrolase [Rhodospirillaceae bacterium] [mmp_id=MMP07619952] [mmp_db=mardb]
WP_003808778.1_MMP02231460 MULTISPECIES: phosphoenolpyruvate phosphomutase [Bordetella] [mmp_id=MMP02231460] [mmp_db=mardb]
MBM32497.1_MMP07618509 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618509] [mmp_db=mardb]
MAX58083.1_MMP07618500 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618500] [mmp_db=mardb]
GCF_000375765.1_00099_MMP02256471 Phosphonopyruvate hydrolase [mmp_id=MMP02256471] [mmp_db=mardb]
MBK66302.1_MMP07618514 hypothetical protein CL769_05050 [Chloroflexi bacterium] [mmp_id=MMP07618514] [mmp_db=mardb]
MBN18068.1_MMP07618503 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618503] [mmp_db=mardb]
MBJ78420.1_MMP07619524 phosphoenolpyruvate mutase [Nitrospinae bacterium] [mmp_id=MMP07619524] [mmp_db=mardb]
WP_005011554.1_MMP2272521 phosphoenolpyruvate mutase [Nitrospina gracilis] [mmp_id=MMP2272521] [mmp_db=mardb]
MBA30457.1_MMP07618650 phosphoenolpyruvate mutase [Dehalococcoidia bacterium] [mmp_id=MMP07618650] [mmp_db=mardb]
MAV14427.1_MMP07618450 phosphoenolpyruvate mutase partial [Chloroflexi bacterium] [mmp_id=MMP07618450] [mmp_db=mardb]
MBH60513.1_MMP07618652 hypothetical protein CL907_05055 [Dehalococcoidia bacterium] [mmp_id=MMP07618652] [mmp_db=mardb]
MBM03222.1_MMP07618511 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618511] [mmp_db=mardb]

MMP492357_152197 [UniRef50=UniRef50_A0A0S7X8Q6Cluster: Uncharacterized protein] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]
MAF51975.1_MMP07618439 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618439] [mmp_db=mardb]
GCA_002413545.1_01931_MMP06456309 Phosphonopyruvate hydrolase [mmp_id=MMP06456309] [mmp_db=mardb]
GCA_000372225.1_01694_MMP02261272 Phosphonopyruvate hydrolase [mmp_id=MMP02261272] [mmp_db=mardb]
MBM01825.1_MMP07618507 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618507] [mmp_db=mardb]
MBP63019.1_MMP07619706 hypothetical protein CMJ62_15975 [Planctomycetaceae bacterium] [mmp_id=MMP07619706] [mmp_db=mardb]
GCA_003232235.1_00208_MMP09240083 Phosphonopyruvate hydrolase [mmp_id=MMP09240083] [mmp_db=mardb]

KPJ58683.1_MMP03994146 hypothetical protein AMJ42_02810 [Deltaproteobacteria bacterium DG_8] [mmp_id=MMP03994146] [mmp_db=mardb]
GCA_003245675.1_00800_MMP09287973 Phosphonopyruvate hydrolase [mmp_id=MMP09287973] [mmp_db=mardb]
GCA_002434715.1_04403_MMP06452182 Phosphonopyruvate hydrolase [mmp_id=MMP06452182] [mmp_db=mardb]

KYK28104.1_MMP04495909 hypothetical protein AYK20_01710 [Thermoplasmatales archaeon SG8-52-1] [mmp_id=MMP04495909] [mmp_db=mardb]
MBE31609.1_MMP07620261 hypothetical protein CMP17_01390 [Rickettsiales bacterium] [mmp_id=MMP07620261] [mmp_db=mardb]
GCA_001542995.1_00717_MMP03777284 Phosphonopyruvate hydrolase [mmp_id=MMP03777284] [mmp_db=mardb]
GCA_001577055.1_00552_MMP04423161 Phosphonopyruvate hydrolase [mmp_id=MMP04423161] [mmp_db=mardb]
MBO82775.1_MMP07618251 phosphoenolpyruvate mutase [Actinobacteria bacterium] [mmp_id=MMP07618251] [mmp_db=mardb]
MBO39511.1_MMP07620019 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620019] [mmp_db=mardb]
GCA_002377585.1_00088_MMP06457119 Phosphonopyruvate hydrolase [mmp_id=MMP06457119] [mmp_db=mardb]
PCI19122.1_MMP07568951 phosphoenolpyruvate mutase [SAR202 cluster bacterium] [mmp_id=MMP07568951] [mmp_db=mardb]
GCA_003235975.1_00265_MMP09287816 Phosphonopyruvate hydrolase [mmp_id=MMP09287816] [mmp_db=mardb]
GCA_003246675.1_01424_MMP09287898 Phosphonopyruvate hydrolase [mmp_id=MMP09287898] [mmp_db=mardb]

MBH44942.1_MMP07619132 hypothetical protein CMD88_05755 [Gammaproteobacteria bacterium] [mmp_id=MMP07619132] [mmp_db=mardb]

MAU56302.1_MMP07618644 phosphoenolpyruvate mutase [Dehalococcoidia bacterium] [mmp_id=MMP07618644] [mmp_db=mardb]
MAS50841.1_MMP07618457 hypothetical protein CL712_02820 [Chloroflexi bacterium] [mmp_id=MMP07618457] [mmp_db=mardb]
MBM14753.1_MMP07619519 phosphoenolpyruvate mutase [Nitrospina sp.] [mmp_id=MMP07619519] [mmp_db=mardb]
MBL75785.1_MMP07618508 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618508] [mmp_db=mardb]
GCA_002450435.1_00579_MMP06450360 Phosphonopyruvate hydrolase [mmp_id=MMP06450360] [mmp_db=mardb]
MBD86332.1_MMP07618488 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618488] [mmp_db=mardb]
KMP11646.1_MMP03376160 phosphoenolpyruvate mutase [Nitrospina sp. SCGC_AAA799_C22] [mmp_id=MMP03376160] [mmp_db=mardb]
GCA_002346435.1_00149_MMP06457216 Phosphonopyruvate hydrolase [mmp_id=MMP06457216] [mmp_db=mardb]
GCA_002348795.1_01296_MMP06451904 Phosphonopyruvate hydrolase [mmp_id=MMP06451904] [mmp_db=mardb]
GCA_003235465.1_01204_MMP09287807 Phosphonopyruvate hydrolase [mmp_id=MMP09287807] [mmp_db=mardb]
MBS59812.1_MMP07618351 phosphoenolpyruvate mutase [Anaerolineaceae bacterium] [mmp_id=MMP07618351] [mmp_db=mardb]
GCA_003229345.1_00234_MMP09240210 Phosphonopyruvate hydrolase [mmp_id=MMP09240210] [mmp_db=mardb]
GCA_002500965.1_00733_MMP06452478 Phosphonopyruvate hydrolase [mmp_id=MMP06452478] [mmp_db=mardb]
MAX17400.1_MMP07619517 phosphoenolpyruvate mutase [Nitrospina sp.] [mmp_id=MMP07619517] [mmp_db=mardb]
MAF20390.1_MMP07619599 hypothetical protein CMI55_01780 [Parcubacteria group bacterium] [mmp_id=MMP07619599] [mmp_db=mardb]

MBQ09481.1_MMP07619140 hypothetical protein CMD96_06800 [Gammaproteobacteria bacterium] [mmp_id=MMP07619140] [mmp_db=mardb]
GCA_003229915.1_00709_MMP09240145 Phosphonopyruvate hydrolase [mmp_id=MMP09240145] [mmp_db=mardb]
GCA_002506165.1_00636_MMP06027469 Phosphonopyruvate hydrolase [mmp_id=MMP06027469] [mmp_db=mardb]
PXF51769.1_MMP08434975 phosphoenolpyruvate mutase [ANME-1 cluster archaeon] [mmp_id=MMP08434975] [mmp_db=mardb]
MMP491463_206367 [UniRef50=UniRef50_A0A1G1J1V0Cluster: Uncharacterized protein] [Interpro=IPR029044Nucleotide-diphospho-sugar transferases] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]
MAG08768.1_MMP07620233 phosphoenolpyruvate mutase [Candidatus Woesearchaeota archaeon] [mmp_id=MMP07620233] [mmp_db=mardb]
GCA_002502135.1_00411_MMP06027474 Phosphonopyruvate hydrolase [mmp_id=MMP06027474] [mmp_db=mardb]

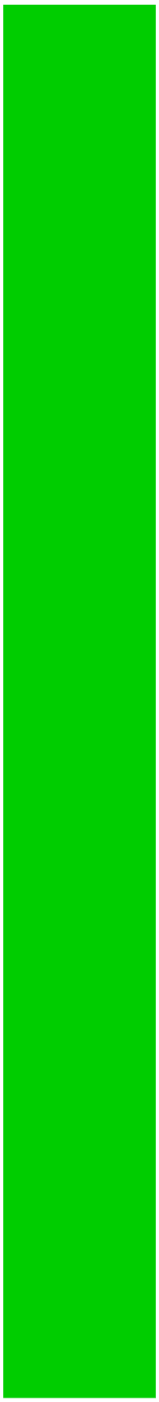
MAF61158.1_MMP07618425 phosphoenolpyruvate mutase [Blastomonas sp.] [mmp_id=MMP07618425] [mmp_db=mardb]
PCJ58663.1_MMP07568849 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07568849] [mmp_db=mardb]
GCA_002471405.1_00998_MMP06455356 Phosphonopyruvate hydrolase [mmp_id=MMP06455356] [mmp_db=mardb]
GCA_002433595.1_00639_MMP06451400 Phosphonopyruvate hydrolase [mmp_id=MMP06451400] [mmp_db=mardb]
MBD97865.1_MMP07618393 phosphoenolpyruvate mutase [bacterium] [mmp_id=MMP07618393] [mmp_db=mardb]
WP_083373742.1_MMP05826284 phosphoenolpyruvate mutase [Moorea producens] [mmp_id=MMP05826284] [mmp_db=marref]
WP_039357130.1_MMP07498378 MULTISPECIES: phosphoenolpyruvate mutase [Burkholderia] [mmp_id=MMP07498378] [mmp_db=mardb]
KHL13034.1_MMP03070120 phosphoenolpyruvate phosphomutase [Mumia flava] [mmp_id=MMP03070120] [mmp_db=mardb]
WP_023477253.1_MMP02469983 phosphoenolpyruvate phosphomutase [Burkholderia cenocepacia] [mmp_id=MMP02469983] [mmp_db=mardb]
MBV31477.1_MMP07620290 phosphoenolpyruvate mutase [Rickettsiales bacterium] [mmp_id=MMP07620290] [mmp_db=mardb]
MAI81079.1_MMP07618662 phosphoenolpyruvate mutase [Deltaproteobacteria bacterium] [mmp_id=MMP07618662] [mmp_db=mardb]

MMP494431_190222 [UniRef50=UniRef50_A0A0D2W1R5Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=2.7.7.33] [mmp_id=MMP494431] [mmp_db=marcat]

MMP494431_190221 [UniRef50=UniRef50_A0A0D2W1R5Cluster: Phosphoenolpyruvate phosphomutase] [Interpro=IPR012698Phosphoenolpyruvate phosphomutase core] [Priam=5.4.2.9] [mmp_id=MMP494431] [mmp_db=marcat]
GCA_003228495.1_01494_MMP09240199 Phosphonopyruvate hydrolase [mmp_id=MMP09240199] [mmp_db=mardb]
AUS80769.1_MMP08364581 phosphoenolpyruvate mutase [Actinoalloteichus sp. AHMU CJ021] [mmp_id=MMP08364581] [mmp_db=marref]

KHD06688.1_MMP03165106 phosphoenolpyruvate phosphomutase [Candidatus Thiomargarita nelsonii] [mmp_id=MMP03165106] [mmp_db=mardb]

KHD11437.1_MMP03165106 phosphoenolpyruvate phosphomutase [Candidatus Thiomargarita nelsonii] [mmp_id=MMP03165106] [mmp_db=mardb]
MBR93142.1_MMP07619802 phosphoenolpyruvate mutase [Proteobacteria bacterium] [mmp_id=MMP07619802] [mmp_db=mardb]
WP_051242647.1_MMP02441158 phosphoenolpyruvate mutase [Stappia stellulata] [mmp_id=MMP02441158] [mmp_db=mardb]
WP_005001090.1_MMP02436141 phosphoenolpyruvate mutase [Nitrococcus mobilis] [mmp_id=MMP02436141] [mmp_db=mardb]



MAJ30509.1_MMP07620417 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620417] [mmp_db=mardb]
SEO57998.1_MMP04490248 phosphoenolpyruvate mutase [Salinihabitans flavidus] [mmp_id=MMP04490248] [mmp_db=mardb]

MAI82122.1_MMP07620418 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620418] [mmp_db=mardb]

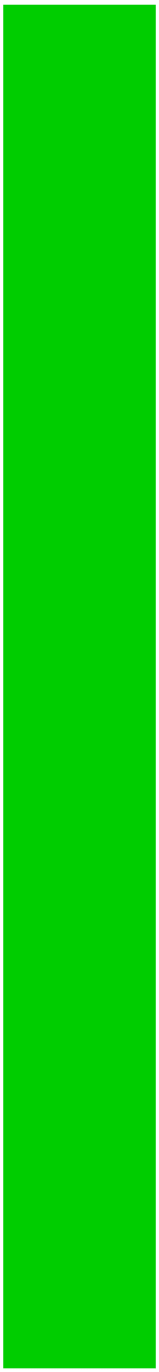
ABA87887.1_MMP02598295 phosphoenolpyruvate phosphomutase [Pelobacter carbinolicus DSM 2380] [mmp_id=MMP02598295] [mmp_db=marref]
MBK31660.1_MMP07619297 phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619297] [mmp_db=mardb]

MMP491463_299831 [UniRef50=UniRef50_A0A1M3MSG2Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.3.1.157] [mmp_id=MMP491463] [mmp_db=marcat]
GCA_003247415.1_00636_MMP09287891 Phosphonopyruvate hydrolase [mmp_id=MMP09287891] [mmp_db=mardb]

MMP491463_299830 [UniRef50=UniRef50_A0A1M3MSG2Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]
WP_072286117.1_MMP04909702 phosphoenolpyruvate mutase [Pelobacter acetylenicus] [mmp_id=MMP04909702] [mmp_db=marref]
GCA_001657395.1_01961_MMP04939355 Phosphonopyruvate hydrolase [mmp_id=MMP04939355] [mmp_db=mardb]
MBS57551.1_MMP07618480 phosphoenolpyruvate mutase [Chloroflexi bacterium] [mmp_id=MMP07618480] [mmp_db=mardb]
GCA_003235785.1_00267_MMP09287860 Phosphonopyruvate hydrolase [mmp_id=MMP09287860] [mmp_db=mardb]
WP_042155018.1_MMP02261502 MULTISPECIES: phosphoenolpyruvate mutase [Planktothrix] [mmp_id=MMP02261502] [mmp_db=mardb]
WP_042155018.1_MMP02261501 MULTISPECIES: phosphoenolpyruvate mutase [Planktothrix] [mmp_id=MMP02261501] [mmp_db=mardb]

MAP83779.1_MMP07620403 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620403] [mmp_db=mardb]
GCA_001643555.1_00024_MMP04113182 Phosphonopyruvate hydrolase [mmp_id=MMP04113182] [mmp_db=mardb]
WP_045519159.1_MMP00000021 phosphoenolpyruvate mutase [Bacillus niacini] [mmp_id=MMP00000021] [mmp_db=mardb]
MBB42517.1_MMP07619988 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619988] [mmp_db=mardb]

MBI32378.1_MMP07619142 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619142] [mmp_db=mardb]

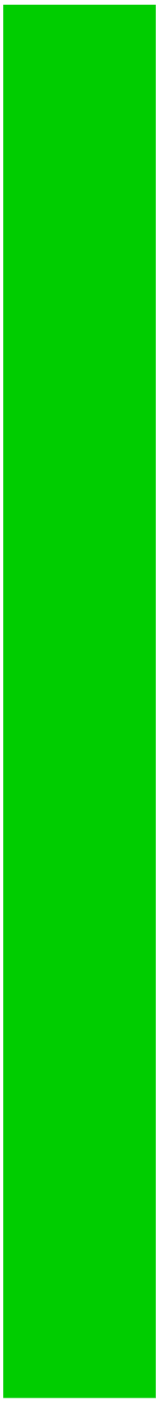


MBQ09441.1_MMP07619140 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619140] [mmp_db=mardb]

PPR57756.1_MMP07286199 Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha3_Bin6] [mmp_id=MMP07286199] [mmp_db=mardb]
MBC36642.1_MMP07620041 phosphoenolpyruvate mutase [Rickettsiales bacterium] [mmp_id=MMP07620041] [mmp_db=mardb]
RAP34809.1_MMP08965197 phosphoenolpyruvate mutase partial [Candidatus Marinamargulisbacteria bacterium SCGC AG-439-L15] [mmp_id=MMP08965197] [mmp_db=mardb]
WP_083202259.1_MMP05382895 phosphoenolpyruvate mutase [Stappia indica] [mmp_id=MMP05382895] [mmp_db=mardb]
PRZ53072.1_MMP06264457 phosphoenolpyruvate mutase [Paraburkholderia insulsa] [mmp_id=MMP06264457] [mmp_db=mardb]
RAP31434.1_MMP08965200 phosphoenolpyruvate mutase [Candidatus Marinamargulisbacteria bacterium SCGC AG-343-D04] [mmp_id=MMP08965200] [mmp_db=mardb]
GCA_002470825.1_00459_MMP06453278 Phosphonopyruvate hydrolase [mmp_id=MMP06453278] [mmp_db=mardb]

MBF42503.1_MMP07619141 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619141] [mmp_db=mardb]

MAD47893.1_MMP07620397 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07620397] [mmp_db=mardb]
MBS94531.1_MMP07618544 phosphoenolpyruvate mutase [Chromatiales bacterium] [mmp_id=MMP07618544] [mmp_db=mardb]
MBQ26716.1_MMP07619525 phosphoenolpyruvate mutase [Nitrospiraceae bacterium] [mmp_id=MMP07619525] [mmp_db=mardb]
MAM67613.1_MMP07619956 phosphoenolpyruvate mutase partial [Rhodospirillaceae bacterium] [mmp_id=MMP07619956] [mmp_db=mardb]
WP_083206217.1_MMP05382876 phosphoenolpyruvate mutase [Stappia indica] [mmp_id=MMP05382876] [mmp_db=mardb]
MAS81580.1_MMP07619289 phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619289] [mmp_db=mardb]
ABE54447.1_MMP02598300 23-dimethylmalate lyase [Shewanella denitrificans OS217] [mmp_id=MMP02598300] [mmp_db=marref]
MBV40496.1_MMP07619967 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619967] [mmp_db=mardb]
AWB42962.1_MMP08934269 phosphoenolpyruvate mutase [Paenibacillus sp. CAA11] [mmp_id=MMP08934269] [mmp_db=marref]
GCA_002469945.1_01135_MMP06457391 Phosphonopyruvate hydrolase [mmp_id=MMP06457391] [mmp_db=mardb]
GCA_002500465.1_00032_MMP06455355 Phosphonopyruvate hydrolase [mmp_id=MMP06455355] [mmp_db=mardb]



MBT77452.1_MMP07618543 phosphoenolpyruvate mutase [Chromatiales bacterium] [mmp_id=MMP07618543] [mmp_db=mardb]
GCA_002390245.1_02979_MMP06457461 Phosphonopyruvate hydrolase [mmp_id=MMP06457461] [mmp_db=mardb]
MAG58367.1_MMP07619727 phosphoenolpyruvate mutase [Planctomycetes bacterium] [mmp_id=MMP07619727] [mmp_db=mardb]
MAF82498.1_MMP07618542 phosphoenolpyruvate mutase [Chromatiales bacterium] [mmp_id=MMP07618542] [mmp_db=mardb]
MBE11134.1_MMP07619979 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619979] [mmp_db=mardb]
MAF95691.1_MMP07620004 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620004] [mmp_db=mardb]
GCA_002434595.1_01329_MMP06450461 Phosphonopyruvate hydrolase [mmp_id=MMP06450461] [mmp_db=mardb]

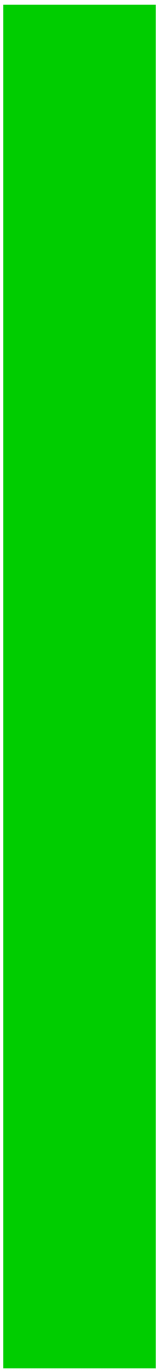
PPR42533.1_MMP07286216 Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin2] [mmp_id=MMP07286216] [mmp_db=mardb]
MBI07088.1_MMP07619998 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619998] [mmp_db=mardb]
MBG05926.1_MMP07620003 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620003] [mmp_db=mardb]

WP_070991887.1_MMP05792550 phosphoenolpyruvate mutase [Pseudoalteromonas byunsanensis] [mmp_id=MMP05792550] [mmp_db=mardb]
MBU70460.1_MMP07618603 phosphoenolpyruvate mutase [Cupriavidus sp.] [mmp_id=MMP07618603] [mmp_db=mardb]
PCH55231.1_MMP07568997 phosphoenolpyruvate mutase [Burkholderiaceae bacterium] [mmp_id=MMP07568997] [mmp_db=mardb]
GCA_003233615.1_02490_MMP09239988 Phosphonopyruvate hydrolase [mmp_id=MMP09239988] [mmp_db=mardb]

OEU71675.1_MMP05301632 phosphoenolpyruvate mutase [Desulfuromonadales bacterium C00003068] [mmp_id=MMP05301632] [mmp_db=mardb]
MBV28968.1_MMP07619966 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619966] [mmp_db=mardb]

MMP492357_300028 [UniRef50=UniRef50_A0A0B5FB86Cluster: Phosphonopyruvate hydrolase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP492357] [mmp_db=marcat]

MMP492357_300029 [UniRef50=UniRef50_A0A0B5FB86Cluster: Phosphonopyruvate hydrolase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.7.60] [mmp_id=MMP492357] [mmp_db=marcat]
WP_002810069.1_MMP02436179 phosphoenolpyruvate mutase [Nitrosococcus oceani] [mmp_id=MMP02436179] [mmp_db=mardb]
WP_002810069.1_MMP02864944 phosphoenolpyruvate mutase [Nitrosococcus oceani] [mmp_id=MMP02864944] [mmp_db=mardb]



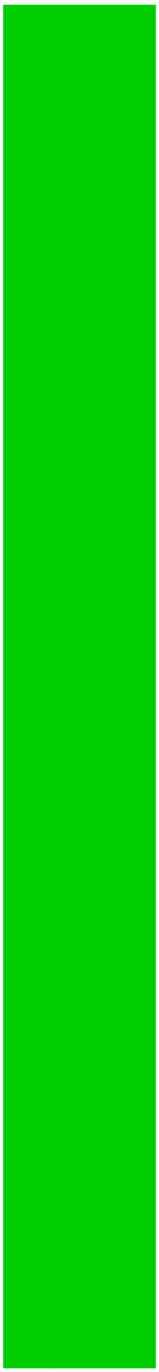
GCA_002340825.1_03645_MMP06452276 Phosphonopyruvate hydrolase [mmp_id=MMP06452276] [mmp_db=mardb]
WP_036500424.1_MMP02898118 phosphoenolpyruvate mutase [Nitrosococcus oceani] [mmp_id=MMP02898118] [mmp_db=mardb]
GCA_003230275.1_01538_MMP09240095 Phosphonopyruvate hydrolase [mmp_id=MMP09240095] [mmp_db=mardb]
ABA57815.1_MMP02598329 23-dimethylmalate lyase [Nitrosococcus oceani ATCC 19707] [mmp_id=MMP02598329] [mmp_db=marref]

PPR48436.1_MMP07286215 Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin1] [mmp_id=MMP07286215] [mmp_db=mardb]

PPR73314.1_MMP07286195 Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha3_Bin2] [mmp_id=MMP07286195] [mmp_db=mardb]
GCA_003233115.1_01686_MMP09240020 Phosphonopyruvate hydrolase [mmp_id=MMP09240020] [mmp_db=mardb]
GCA_002453985.1_00693_MMP06451993 Phosphonopyruvate hydrolase [mmp_id=MMP06451993] [mmp_db=mardb]
MBV24227.1_MMP07619965 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619965] [mmp_db=mardb]
ADJ28131.1_MMP02598531 phosphoenolpyruvate phosphomutase [Nitrosococcus watsonii C-113] [mmp_id=MMP02598531] [mmp_db=marref]
GCA_002341065.1_04382_MMP06457331 Phosphonopyruvate hydrolase [mmp_id=MMP06457331] [mmp_db=mardb]
MBE88682.1_MMP07620020 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620020] [mmp_db=mardb]
GCA_002390205.1_00197_MMP06451823 Phosphonopyruvate hydrolase [mmp_id=MMP06451823] [mmp_db=mardb]

MBE03790.1_MMP07619111 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619111] [mmp_db=mardb]
MBO44291.1_MMP07619972 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619972] [mmp_db=mardb]
MAW55866.1_MMP07620013 phosphoenolpyruvate mutase partial [Rhodospirillaceae bacterium] [mmp_id=MMP07620013] [mmp_db=mardb]
ADE14458.1_MMP02598510 phosphoenolpyruvate phosphomutase [Nitrosococcus halophilus Nc 4] [mmp_id=MMP02598510] [mmp_db=marref]

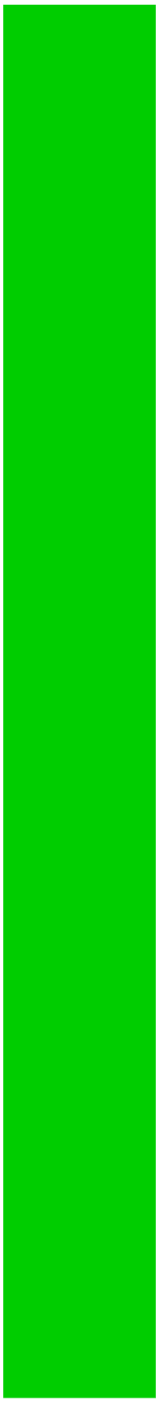
PPR19649.1_MMP07286231 Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha10_Bin2] [mmp_id=MMP07286231] [mmp_db=mardb]
PCN45699.1_MMP06759470 phosphoenolpyruvate mutase [Brevibacillus laterosporus] [mmp_id=MMP06759470] [mmp_db=mardb]
PCI21724.1_MMP07568950 phosphoenolpyruvate mutase partial [SAR324 cluster bacterium] [mmp_id=MMP07568950] [mmp_db=mardb]
MBH71388.1_MMP07619641 phosphoenolpyruvate mutase [Pelagibacteraceae bacterium] [mmp_id=MMP07619641] [mmp_db=mardb]



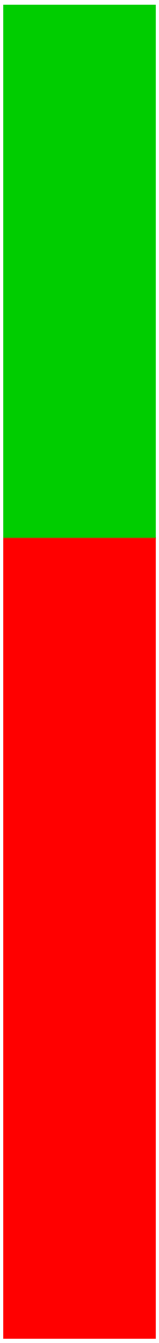
MAG23290.1_MMP07619953 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619953] [mmp_db=mardb]
GCA_003247575.1_03863_MMP09287884 Phosphonopyruvate hydrolase [mmp_id=MMP09287884] [mmp_db=mardb]
AVX06931.1_MMP07838500 phosphoenolpyruvate mutase [Bacillus sp. Y-01] [mmp_id=MMP07838500] [mmp_db=marref]
GCA_002352385.1_00263_MMP06451510 Phosphonopyruvate hydrolase [mmp_id=MMP06451510] [mmp_db=mardb]
MAQ83588.1_MMP07619456 phosphoenolpyruvate mutase [Maritimibacter sp.] [mmp_id=MMP07619456] [mmp_db=mardb]
GCA_002501105.1_00891_MMP06457369 Phosphonopyruvate hydrolase [mmp_id=MMP06457369] [mmp_db=mardb]
MAH84731.1_MMP07619309 phosphoenolpyruvate mutase [Magnetovibrio sp.] [mmp_id=MMP07619309] [mmp_db=mardb]

MBM24054.1_MMP07619122 phosphoenolpyruvate mutase [Gammaproteobacteria bacterium] [mmp_id=MMP07619122] [mmp_db=mardb]
GAX62849.1_MMP00036669 PEP phosphomutase [Candidatus Scalindua sp. husup-a2] [mmp_id=MMP00036669] [mmp_db=mardb]
MAX28946.1_MMP07620167 phosphoenolpyruvate mutase [Thiotrichales bacterium] [mmp_id=MMP07620167] [mmp_db=mardb]
GCA_003249475.1_00019_MMP09288137 Phosphonopyruvate hydrolase [mmp_id=MMP09288137] [mmp_db=mardb]
MBR71886.1_MMP07619974 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619974] [mmp_db=mardb]
MBJ26613.1_MMP07618312 phosphoenolpyruvate mutase [Alphaproteobacteria bacterium] [mmp_id=MMP07618312] [mmp_db=mardb]
MBN08222.1_MMP07619989 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619989] [mmp_db=mardb]
MAI10307.1_MMP07620010 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620010] [mmp_db=mardb]
GCA_003249475.1_01054_MMP09288137 Phosphonopyruvate hydrolase [mmp_id=MMP09288137] [mmp_db=mardb]
MAN80433.1_MMP07619308 phosphoenolpyruvate mutase [Magnetovibrio sp.] [mmp_id=MMP07619308] [mmp_db=mardb]
MAA98202.1_MMP07620130 phosphoenolpyruvate mutase [Stappia sp.] [mmp_id=MMP07620130] [mmp_db=mardb]
PTX60241.1_MMP08776942 phosphoenolpyruvate phosphomutase [Melghirimyces profundicolus] [mmp_id=MMP08776942] [mmp_db=mardb]
GCA_001629325.1_00331_MMP04534682 Phosphonopyruvate hydrolase [mmp_id=MMP04534682] [mmp_db=mardb]

OEU65941.1_MMP05301624 phosphoenolpyruvate mutase partial [Desulfobacterales bacterium S5133MH16] [mmp_id=MMP05301624] [mmp_db=mardb]
MBL68683.1_MMP07620218 phosphoenolpyruvate mutase [Verrucomicrobiales bacterium] [mmp_id=MMP07620218] [mmp_db=mardb]



GCA_002359255.1_02600_MMP06456280 Phosphonopyruvate hydrolase [mmp_id=MMP06456280] [mmp_db=mardb]
MBB91078.1_MMP07619312 phosphoenolpyruvate mutase [Magnetovibrio sp.] [mmp_id=MMP07619312] [mmp_db=mardb]
WP_036540049.1_MMP02743897 phosphoenolpyruvate mutase [Pseudooceanicola nanhaiensis] [mmp_id=MMP02743897] [mmp_db=mardb]
MBL68532.1_MMP07620218 phosphoenolpyruvate mutase [Verrucomicrobiales bacterium] [mmp_id=MMP07620218] [mmp_db=mardb]
GCA_002433235.1_02961_MMP06454690 Phosphonopyruvate hydrolase [mmp_id=MMP06454690] [mmp_db=mardb]
GCA_002495315.1_00380_MMP06027030 Phosphonopyruvate hydrolase [mmp_id=MMP06027030] [mmp_db=mardb]
MBH03641.1_MMP07620252 phosphoenolpyruvate mutase [Xanthomonadales bacterium] [mmp_id=MMP07620252] [mmp_db=mardb]
GCA_003283105.1_00089_MMP08886030 Phosphonopyruvate hydrolase [mmp_id=MMP08886030] [mmp_db=mardb]
GCA_003209165.1_01310_MMP08886582 Phosphonopyruvate hydrolase [mmp_id=MMP08886582] [mmp_db=mardb]
GCA_002420445.1_00259_MMP06455107 Phosphonopyruvate hydrolase [mmp_id=MMP06455107] [mmp_db=mardb]
MAH21593.1_MMP07620156 phosphoenolpyruvate mutase [Thaumarchaeota archaeon] [mmp_id=MMP07620156] [mmp_db=mardb]
GCA_003235555.1_00156_MMP09287802 Phosphonopyruvate hydrolase [mmp_id=MMP09287802] [mmp_db=mardb]
EDN71450.1_MMP02470514 phosphoenolpyruvate phosphomutase [Beggiatoa sp. PS] [mmp_id=MMP02470514] [mmp_db=mardb]
GCA_001643485.1_00432_MMP04113178 Phosphonopyruvate hydrolase [mmp_id=MMP04113178] [mmp_db=mardb]
MAO54861.1_MMP07620005 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620005] [mmp_db=mardb]
GCA_002415625.1_02740_MMP06457039 Phosphonopyruvate hydrolase [mmp_id=MMP06457039] [mmp_db=mardb]
WP_037338324.1_MMP02929445 phosphoenolpyruvate mutase [Salinisphaera hydrothermalis] [mmp_id=MMP02929445] [mmp_db=mardb]
GCA_003250975.1_00685_MMP09288140 Phosphonopyruvate hydrolase [mmp_id=MMP09288140] [mmp_db=mardb]
PPR62234.1_MMP07286202 hypothetical protein CFH10_00887 partial [Alphaproteobacteria bacterium MarineAlpha4_Bin2] [mmp_id=MMP07286202] [mmp_db=mardb]
GCA_003235935.1_01300_MMP09287818 Phosphonopyruvate hydrolase [mmp_id=MMP09287818] [mmp_db=mardb]
MAZ79102.1_MMP07619291 phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619291] [mmp_db=mardb]
MBK20413.1_MMP07619996 phosphoenolpyruvate mutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619996] [mmp_db=mardb]
WP_070393116.1_MMP05826283 phosphoenolpyruvate mutase [Moorea producens] [mmp_id=MMP05826283] [mmp_db=marref]



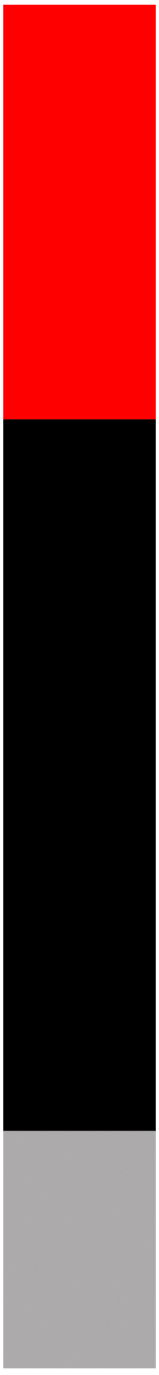
PPR31149.1_MMP07286220 Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin6] [mmp_id=MMP07286220] [mmp_db=mardb]
GCA_002500605.1_00781_MMP06450621 Phosphonopyruvate hydrolase [mmp_id=MMP06450621] [mmp_db=mardb]
MAS83288.1_MMP07619288 phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619288] [mmp_db=mardb]
GCA_003251555.1_00343_MMP09288039 Phosphonopyruvate hydrolase [mmp_id=MMP09288039] [mmp_db=mardb]

PPR37399.1_MMP07286218 Phosphonopyruvate hydrolase [Alphaproteobacteria bacterium MarineAlpha6_Bin4] [mmp_id=MMP07286218] [mmp_db=mardb]
MBK68764.1_MMP07619298 phosphoenolpyruvate mutase [Legionellales bacterium] [mmp_id=MMP07619298] [mmp_db=mardb]
MAI26272.1_MMP07620119 phosphoenolpyruvate mutase [Spirochaeta sp.] [mmp_id=MMP07620119] [mmp_db=mardb]
SDU33903.1_MMP05428979 phosphoenolpyruvate mutase [Stappia sp. ES.058] [mmp_id=MMP05428979] [mmp_db=marref]
GCA_002376045.1_01806_MMP06451006 23-dimethylmalate lyase [mmp_id=MMP06451006] [mmp_db=mardb]
AGM40194.1_MMP02603230 isocitrate lyase and phosphorylmutase [Spiribacter salinus M19-40] [mmp_id=MMP02603230] [mmp_db=marref]

MAX59592.1_MMP07618500 carboxyvinyl-carboxyphosphonate phosphorylmutase [Chloroflexi bacterium] [mmp_id=MMP07618500] [mmp_db=mardb]
MMP494431_74556 [UniRef50=UniRef50_A0A160V7K5Cluster: Methylisocitrate lyase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=2.7.8.23] [mmp_id=MMP494431] [mmp_db=marcat]
GCA_002413545.1_01794_MMP06456309 Carboxyvinyl-carboxyphosphonate phosphorylmutase [mmp_id=MMP06456309] [mmp_db=mardb]

MBQ11747.1_MMP07619689 carboxyvinyl-carboxyphosphonate phosphorylmutase [Planctomyces sp.] [mmp_id=MMP07619689] [mmp_db=mardb]
RAI02086.1_MMP09079831 isocitrate lyase [Acuticoccus sp. PTG4-2] [mmp_id=MMP09079831] [mmp_db=mardb]

AUV80813.1_MMP08381148 carboxyvinyl-carboxyphosphonate phosphorylmutase [Salinigranum rubrum] [mmp_id=MMP08381148] [mmp_db=marref]
WP_116450246.1_MMP08954526 oxaloacetate decarboxylase [Blastococcus litoris] [mmp_id=MMP08954526] [mmp_db=mardb]
GCA_002375965.1_00799_MMP06455818 23-dimethylmalate lyase [mmp_id=MMP06455818] [mmp_db=mardb]
MMP490065_44831 [UniRef50=UniRef50_A0A160V7K5Cluster: Methylisocitrate lyase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=4.1.3.30] [mmp_id=MMP490065] [mmp_db=marcat]



WP_027983379.1_MMP02440745 carboxyvinyl-carboxyphosphonate phosphorylmutase [delta proteobacterium PSCGC 5342] [mmp_id=MMP02440745] [mmp_db=mardb]

PCI17085.1_MMP07568951 carboxyvinyl-carboxyphosphonate phosphorylmutase [SAR202 cluster bacterium] [mmp_id=MMP07568951] [mmp_db=mardb]
MBB70843.1_MMP07619294 isocitrate lyase [Legionellales bacterium] [mmp_id=MMP07619294] [mmp_db=mardb]
PKB78155.1_MMP06121366 carboxyvinyl-carboxyphosphonate phosphorylmutase [SAR202 cluster bacterium MP-SInd-SRR3963457-G2] [mmp_id=MMP06121366] [mmp_db=mardb]

MAZ63906.1_MMP07618640 carboxyvinyl-carboxyphosphonate phosphorylmutase [Dehalococcoidia bacterium] [mmp_id=MMP07618640] [mmp_db=mardb]

MBO89973.1_MMP07620258 carboxyvinyl-carboxyphosphonate phosphorylmutase [Rickettsiales bacterium] [mmp_id=MMP07620258] [mmp_db=mardb]

MBO43289.1_MMP07619972 carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619972] [mmp_db=mardb]

MBE09768.1_MMP07619979 carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619979] [mmp_db=mardb]

MBE09208.1_MMP07619979 carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619979] [mmp_db=mardb]
PPR24431.1_MMP07286232 23-dimethylmalate lyase [Alphaproteobacteria bacterium MarineAlpha10_Bin3] [mmp_id=MMP07286232] [mmp_db=mardb]
PPR72984.1_MMP07286201 23-dimethylmalate lyase [Alphaproteobacteria bacterium MarineAlpha4_Bin1] [mmp_id=MMP07286201] [mmp_db=mardb]

MBE90608.1_MMP07620020 carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07620020] [mmp_db=mardb]

MBL24863.1_MMP07619992 carboxyvinyl-carboxyphosphonate phosphorylmutase [Rhodospirillaceae bacterium] [mmp_id=MMP07619992] [mmp_db=mardb]
PPR64289.1_MMP07286202 23-dimethylmalate lyase [Alphaproteobacteria bacterium MarineAlpha4_Bin2] [mmp_id=MMP07286202] [mmp_db=mardb]
PJN35138.1_MMP07346485 phosphoenolpyruvate phosphomutase [Streptomyces sp. CB02613] [mmp_id=MMP07346485] [mmp_db=mardb]

MMP491463_307773 [UniRef50=UniRef50_A9EXQ4Cluster: Phosphoenolpyruvate mutase] [Interpro=IPR015813Pyruvate/Phosphoenolpyruvate kinase-like domain] [Priam=5.4.2.9] [mmp_id=MMP491463] [mmp_db=marcat]
KIE23726.1_MMP03070123 phosphoenolpyruvate phosphomutase [Streptomyces sp. MUSC 125] [mmp_id=MMP03070123] [mmp_db=mardb]

WP_030276407.1_MMP02645363 phosphoenolpyruvate phosphomutase [Streptomyces sp. NRRL B-24484] [mmp_id=MMP02645363] [mmp_db=mardb]
AWK12922.1_MMP08984914 phosphoenolpyruvate phosphomutase [Streptomyces spongiicola] [mmp_id=MMP08984914] [mmp_db=marref]

MAZ47818.1_MMP07619209 phosphoenolpyruvate mutase [Halobacteriovoraceae bacterium] [mmp_id=MMP07619209] [mmp_db=mardb]
WP_046426931.1_MMP03445833 hypothetical protein [Streptomyces malaysiense] [mmp_id=MMP03445833] [mmp_db=mardb]
WP_078509320.1_MMP02441002 hypothetical protein [Streptomyces sp. CNT302] [mmp_id=MMP02441002] [mmp_db=mardb]

Supplementary Table S4.5: Primers used for the amplification of the *S. pacifica* phosphonate BGC and specific GNATs from the cluster.

Primers for amplification of *S. pacifica* phosphonate BGC	
Fragment 1 Forward (Including pCAP03 homology)	AAAAACTCGGTTTGACGCCTCCCATGGTATAAATAGTGGCCGAGCACACGTTCCGCCTGT
Fragment 1 Reverse	GGACGGGATCCAGACGGCCT
Fragment 2 Forward	CACGGCATCAACGCGGTCCT
Fragment 2 Reverse	CACCCGGATCAGTCGGCAGC
Fragment 3 Forward	CGGCCGATGTCGAGGAGCAC
Fragment 3 Reverse	TCCCTCGCCGATCAGCACGA
Fragment 4 Forward	TCGGATTGTGCACCGGCGTC
Fragment 4 Reverse	GCACACCCGGCTCCATCACC
Fragment 5 Forward	ATCTACCGGGGTGTGCCGCT
Fragment 5 Reverse	TACCGGGAGACCGCCACCTG
Fragment 6 Forward	GGGCCATTGCCGACCTGACC
Fragment 6 Reverse (including pCAP03 homology)	ATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATAGGTCGGGCATCTCGCGCTAC

Primers for amplification of GNATs	
sppG Forward	GCCTGGTGCCGCGCGGCAGCCAATGACCGTGTTCTCCGCCCC
sppG Reverse	GTGGTGGTGCTCGAGTGCGGCCTCAGCCCTCCCTGCGTCGGA
sppH Forward	GCCTGGTGCCGCGCGGCAGCCAGTGGCCGTACGAGTCGTCGA
sppH Reverse	GTGGTGGTGCTCGAGTGCGGCCCTAGGTGCCTTCGGTGAGCA
sppO Forward	GCCTGGTGCCGCGCGGCAGCCAGTGCGGGTCGCCGCGACCTA
sppO Reverse	GTGGTGGTGCTCGAGTGCGGCCTCATCGGGGGCCGCCCACCA