UNIVERSITY OF CALIFORNIA, SAN DIEGO

Genetic and functional diversity of microbial secondary lipid biosynthetic pathways

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Biology

by

Christine Nicole Shulse

Committee in charge:

Eric Allen, Chair
Douglas Bartlett
Paul Jensen
Joseph Pogliano
Kit Pogliano
Emily Troemel

2012

The Dissertation of Christine Nicole Shulse is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

DEDICATION

This dissertation is dedication to my parents, Eric and Cheryl Shulse, and my sister Julie Shulse. Mom and Dad, thank you for your certainty that I could reach this point and your support in getting me there. Julie, thank you for being a great friend as well as a sister. I've enjoyed living in SoCal with you these past several years and I'll miss our weekend visits!

EPIGRAPH

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

T.S. Eliot
Little Gidding

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AA	arachidonic acid (20:4*n*-6)
ACP	acyl carrier protein
AT	acyltransferase
CLF	chain length factor
DHA	docosahexaenoic acid (22:6*n*-3)
DH/I	dehydratase/isomerase
DPA	docosapentaenoic acid (22:5*n*-6)
EPA	eicosapentaenoic acid (20:5*n*-3)
ER	enoyl reductase
FAME	fatty acid methyl ester
FAS	fatty acid synthesis
FAS/PKS	fatty acid synthase/polyketide synthase
GC-MS	gas chromatography-mass spectrometry
GOS	Global Ocean Sampling
KR	ketoacyl reductase
KS	ketoacyl synthase
MAT	malonyl-CoA:ACP transacylase
NMDS	non-parametric multidimensional scaling
PPTase	phosphopantetheinyl transferase
PUFA	polyunsaturated fatty acid

LIST OF FIGURES

Introduction

Chapter 1

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

ACKNOWLEDGEMENTS

I'd like to start by thanking my advisor Eric Allen. It was his intriguing research and obvious enthusiasm for the field that first drew me into environmental microbiology. In the almost six years since that first meeting he has been a great mentor, giving me freedom to explore my own ideas yet somehow keeping me pointed in the right direction. I would also like to thank my other committee members, Doug Bartlett, Paul Jensen, Joe Pogliano, Kit Pogliano, and Emily Troemel, for all their time and great advice on my research over the years. Doug has also opened his office and lab doors to me since my first day as a rotation student in Hubbs Hall, and I thank him for that.

A big thank you to all members of the no longer fledgling Allen Lab, both past and present, especially Brittany Fernandez, Priya Narasingarao, Sheila Podell, and Juan Ugalde. Beyond the Allen Lab, my science and sanity have benefited tremendously from the Hubbs Hall 4th Floor community, especially the Azam and Bartlett Labs. I'm proud to be able to call my fellow graduate students Emiley Eloe, Nellie Shaul, and Kristen Espantman both colleagues and friends, and all three of them have contributed to my growth as a scientist and person.

I would like to acknowledge the financial support of the NIH CMG Training Grant, the NSF Socrates GK-12 Fellowship, and the ARCS Foundation. The Biological Sciences graduate department has provided logistical support throughout my time here, especially Cathy Pugh and Tom Thomp.

Chapter 1, in full, is a reprint of the material as it appears in PLoS ONE, 2011, Shulse, Christine; Allen, Eric. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Environmental Microbiology, 2011, Shulse, Christine; Allen, Eric. The dissertation author was the primary investigator and author of this paper.

VITA

2006	Bachelor of Science, Biology/Spanish, Georgetown University
2006 – 2012	Graduate Student Researcher, Division of Biological Sciences, University of California, San Diego
2007 – 2009	Teaching Assistant, Division of Biological Sciences, University of California, San Diego
2012	Doctor of Philosophy, Biology University of California, San Diego

PUBLICATIONS

Shulse C., and Allen E. (2011) Widespread occurrence of secondary lipid biosynthesis in microbial lineages. **PLoS ONE** 6(5): e20146, doi:10.1371/journal.pone.0020146.

Shulse C., and Allen E. (2011) Diversity and distribution of microbial long-chain fatty acid biosynthetic genes in the marine environment. **Environ Microbiol** 13(3), 684-695.

Eloe E., Shulse C., Fadrosh D., Williamson S., Allen E., and Bartlett D. (2010) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep ocean environment. **Environ Microbiol Reports**, no. doi: 10.1111/j.1758-2229.2010.00223.x

Shulse C., Steiner N.K., and Hurley C.K. (2007) Allelic diversity in KIR2DL4 in a bone marrow transplant population: description of three novel alleles. **Tissue Antigens** 70(2):157-9.

Goodridge J.P., Lathbury L.J., Steiner N.K., Shulse C., Pullikotil P., Seidah N.G., Hurley C.K., Christiansen F.T., and Witt C.S. (2007) Three common alleles of KIR2DL4 (CD158d) encode constitutively expressed, inducible and secreted receptors in NK cells. **Eur J Immunol** 37(1):199-211.

FIELDS OF STUDY

Major Field: Biological Sciences
Studies in Molecular Biology
Assistant Professor Eric Allen

ABSTRACT OF THE DISSERTATION

Genetic and functional diversity of microbial secondary lipid biosynthetic pathways

by

Christine Nicole Shulse

Doctor of Philosophy in Biology

University of California, San Diego, 2012

Professor Eric Allen, Chair

Bacterial production of long-chain omega-3 polyunsaturated fatty acids (PUFAs), such as eicosapentaenoic acid (EPA, 20:5*n*-3) and docosahexaenoic acid (DHA, 22:6*n*-3), is constrained to a narrow subset of marine Gammaproteobacteria. The genes responsible for *de novo* PUFA biosynthesis, designated *pfaEABCD*, encode large, multi-domain enzyme complexes akin to type I iterative fatty acid and polyketide synthases. Likewise, the heterocyst glycolipids produced by nitrogen-fixing cyanobacteria and the phenolic lipids of *Azotobacter vinelandii* are produced by similar complexes. The prevalence of these "secondary lipid synthases," so named because their products have thus far not been shown to be essential to cell growth and

survival under laboratory conditions, in both cultured and uncultured bacteria is completely unknown.

Bioinformatic methods were used to identify homologous type I FAS/PKS gene clusters in diverse microbial lineages representing 10 phyla. Phylogenomic analysis reveals a high degree of functional conservation within distinct biosynthetic pathways. Next, PCR primers targeting the keto-acyl synthase (KS) domain of the *pfaA* gene involved in PUFA biosynthesis were used to construct environmental clone libraries to investigate the potential for microbial secondary lipid synthesis in disparate marine habitats. Of the 446 sequences recovered, 27.6% clustered with KS sequences involved in the synthesis of EPA, DHA and arachadonic acid (AA, 20:4*n*-6). The remaining 72.4% of clones formed environmental-only clades or clustered with KS domains of *pfaA* homologs from organisms producing unidentified products. Lastly, the production of two distinct secondary lipid products at various temperatures was analyzed in six strains of *Shewanella* in order to provide insight into the factors governing secondary lipid synthesis. The current dissertation significantly expands the known genetic and ecological prevalence of microbial processes involved in secondary metabolism and delivers new opportunities to explore the physiological basis and biotechnological value of novel lipid molecules.

Introduction

Fatty acids are the major component of lipids in two of the three domains of life (*Bacteria* and *Eukarya* but not *Archaea*). They are of interest from a variety of perspectives. First and foremost, fatty acids are essential to the cell structurally. The fatty acid carbon chain makes up the hydrophobic "tail" of phospholipids, allowing these lipids to assemble into a bilayer and form the cytoplasmic membrane. Although both *Bacteria* and *Eukarya* produce fatty acids, the enzymes responsible for fatty acid synthesis in the two domains differs enough on a molecular level that two of the most widely used antibacterial agents, isoniazid and triclosan, target fatty acid synthesis in *Bacteria* [1]. Physiologically, the hydrophobic fatty acid portion of the cytoplasmic membrane serves as a permeability barrier, preventing polar and charged molecules from entering or exiting the cell without the aid of a transport protein. From a biotechnological perspective, fatty acids are converted into biodiesel to power cars, used in drug delivery to aid passage across hydrophobic barriers, and the omega-3 fatty acids in particular are used as nutraceuticals.

The canonical Type II fatty acid synthase (FAS), which generally produces 12 to 18 carbon (C12-C18) saturated and monounsaturated fatty acids, has been studied in the model bacterium *Escherichia coli* for over fifty years [2]. In contrast, an additional system of fatty acid synthesis, co-existing with the Type II FAS and synthesizing specialized lipid products (or "secondary lipids"), was first described in cyanobacteria just 15 years ago [3]. Here both our growing body of knowledge about this "secondary lipid synthase" and areas for further research are summarized. Additionally, Figure 1 gives a structural overview of select fatty acids discussed in this introduction.

Early reports of atypical fatty acids in bacteria

The first studies of fatty acid composition in bacteria report mainly C12-C18 saturated and monounsaturated fatty acids, which represent the dominant fatty acids in the vast majority of bacterial species [4,5]. Indeed, in the early 1960s it was generally accepted that bacteria do not produce long-chain polyunsaturated fatty acids (PUFAs) [6]. The 1977 discovery of the highly unsaturated fatty acid eicosapentaenoic acid (EPA; 20:5*n*3) incorporated into the phospholipids of the marine bacterium *Flexibacter polymorphus* overthrew this paradigm [7]. However, *Flexibacter polymorphus* remained a lone oddity for close to a decade, when in 1986 EPA and another long chain omega-3 fatty acid, docosahexaenoic acid (DHA; 22*n*6), were discovered in 9 of 11 isolates from the deep sea [8]. The biosynthetic mechanism of these PUFAs in bacteria remained unknown, although it was hypothesized that EPA, at least, was synthesized via the oxygen-dependent pathway utilized by eukaryotes [7].

As researchers in Australia and North America were rewriting the literature on polyunsaturated fatty acids in bacteria, their colleagues in England were identifying and characterizing novel glycolipids in nitrogen-fixing cyanobacteria [9]. Further studies in the 1970s revealed that the aglycone moiety of these novel lipids was composed of long, hydroxylated alkyl chains typically 26 or 28 carbons in length [10,11]. While the biosynthetic mechanism behind the production of these specialized lipids remained unknown, the critical ecological role of the heterocyst glycolipids in these nitrogen-fixing bacteria was eventually elucidated. The glycolipid layer serves as a barrier to oxygen diffusion across the heterocyst membrane, protecting the oxygen-sensitive nitrogenase enzyme [12].

An additional group of specialized lipid was discovered in the cysts of the soil bacterium *Azotobacter vinelandii* in 1981 [13] and subsequently found to compose the entire membrane in this differentiated cell type [14]. Again, the alkyl chain was found to be unusually long (C21-C23) [13]. However, at this time no connections were drawn between the production of specialized long-chain lipids in such diverse microorganisms as marine bacteria, nitrogen-fixing cyanobacteria, and soil bacteria.

A distinct biosynthetic mechanism for secondary lipids

Although these unique lipid products were initially described in the late 1960s, 70s, and early 80s, no biosynthetic mechanisms for production were discovered until the late 1990s. Researchers first identified a "polyketide sythase (PKS)-like" gene (so named because its predicted protein product has two active sites typical of PKS enzymes, which produce secondary metabolites in bacteria, fungi, and plants) that was involved in the synthesis of heterocyst glycolipids in cyanobacteria [3]. This was follow by the 2001 description of a "polyketide synthase" responsible for the production of EPA and DHA in the bacterium *Shewanella* sp. strain SCRC2738 and the single-celled eukaryote *Schizochytrium* [15]. Finally, at the turn of the 21st century, the phenolic lipids of *Azotobacter vinelandii* were found to be synthesized by a hybrid Type I FAS/Type III PKS [16]. Additionally, phenolic lipids were found to have an important ecological role for *A. vinelandii*, as inactivation of the genes responsible for phenolic lipid production results in cells unable to form desiccation-resistant cysts [17].

The polyketide synthases of bacteria and fungi are of great interest, primarily because they produce useful antibiotics (e.g., reviewed in [18]) and anti-cancer compounds (e.g. [19,20]), thus their enzymology has been studied intensively. Polyketides synthases can be categorized as Types I-III; here I describe Type I polyketide sythases, as those are the most similar to the unique "PKS/FAS" responsible for producing secondary lipids. See Figure 2 for a visual representation of the domain architecture of a Type I polyketide synthase. A minimal base consisting of a ketosynthase domain (KS), acyltransferase domain (AT), and an acyl carrier protein (ACP) is required to extend a growing acyl chain. Other optional domains, such as a ketoreductase (KR), enoylreductase (ER), or a dehydratase (DH) will act to modify this base chain [21]. The number and order of each distinct domain will ultimately determine the final polyketide, resulting in a large diversity of products [22].

Type I polyketide synthases can be further subdivided into modular and iterative PKSs (reviewed in [23]). With modular PKSs, a given "module," consisting of a KS, AT, and an ACP, as well as any optional domains, catalyzes only one cycle of chain extension. If five cycles of chain extension are necessary, five modules must exist on the PKS. For iterative PKSs, a single enzymatic domain will act in multiple successive cycles of chain extension. In this case, if five cycles of chain extension were necessary, a single "module" would act iteratively through all five cycles. The PKS/FAS responsible for secondary lipid production acts similarly to a Type I iterative PKS.

Although PKSs and FASs share an evolutionary origin [24], the two systems have taken quite different directions in bacteria. Most relevant to this body of work,

the Type II FAS seems to be highly conserved throughout the bacterial domain, and homologs of the *f*atty *a*cid *b*iosynthesis, or *fab*, genes consistently produce C12-C18 saturated and monounsaturated fatty acids in bacteria as phylogenetically distant as *E. coli* and *Bacillus subtilis* [25]. Therefore the discovery of fatty acyl chains produced by "PKS-like" systems is especially exciting, in that it opens the door to incredible diversity of fatty acid structure and thus function in environmental bacteria.

Secondary metabolites and genomics

The field of biology was thrust into the genomics era in 1995, when J. Craig Venter and colleagues published the 1.8 Mbp genome of *Haemophilus influenzae* Rd [26]. Subsequent improvements in sequencing technologies have resulted in an explosion of sequencing data, with close to 8000 finished or draft genomes and almost 400 metagenomes available. The vast majority of these are genomes and metagenomes are microbial. The availability of sequencing data has revolutionized many of the natural sciences, including natural product discovery. Natural product gene clusters can be identified from genomes and metagenomes to inform subsequent isolation and characterization efforts [27].

With this background, the dissertation seeks to answer two questions:

1. Everything we know about the extent of secondary lipid synthesis comes from cultured isolates. However, the vast majority of environmental bacteria are not

in culture [28]. What is the distribution and diversity of secondary lipid synthesis in the natural environment, among uncultured bacteria?

2. Many putative natural products identified by genomics, including those of some known secondary lipids, do not appear to be expressed under laboratory conditions (e.g. [29,30]). What factors govern bacterial production of secondary lipids, and can we further understand their biological role?

To address these questions, this dissertation generates and assimilates genomic, genetic, and physiological data. The research makes use of the wealth of sequencing data recently available as well as traditional methods in analytical chemistry and molecular biology.

Chapter 1 is an analysis of all sequenced microbial genomes to identify secondary lipid synthase biosynthetic pathways. Putative secondary lipid synthase pathways were found in 45 genera representing 10 phyla, indicating that this mode of fatty acid synthesis is widespread in the microbial world.

Chapter 2 turns from the culture flask to the high seas to determine the relevance of secondary lipid synthesis for environmental bacteria. This was accomplished via a two-pronged approach, using a PCR-based culture-independent query of the genetic capacity for long-chain fatty acid biosynthesis as well as a query of metagenomic databases. These data demonstrate that the potential for this mode of fatty acid biosynthesis occurs across ocean basins, in such diverse and geographically

disparate habitats as the Puerto Rico Trench, the surface waters off the Scripps Institution of Oceanography Pier, and the pelagic waters of the Indian Ocean.

Chapter 3 returns to the lab to characterize the ecological relevance and trade-offs in the production of two secondary lipid products that make use of a common pathway in the biotechnologically relevant genus *Shewanella*. The results indicate that although the two products share a common pathway, they play distinct roles in the bacterium's adaptation to low temperature.

References

1. Wright HT, Reynolds KA (2007) Antibacterial targets in fatty acid biosynthesis. Curr Opin Microbiol 10: 447-453.

2. Rock CO, Jackowski S (2002) Forty years of bacterial fatty acid synthesis. Biochem Biophys Res Commun 292: 1155-1166.

3. Campbell EL, Cohen MF, Meeks JC (1997) A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. Arch Microbiol 167: 251-258.

4. Oliver JD, Colwell RR (1973) Extractable lipids of gram-negative marine bacteria: fatty-acid composition. Int J Syst Bacteriol 23: 442-458.

5. O'Leary WM (1962) The fatty acids of bacteria. Bacteriological Reviews 26: 421-447.

6. Erwin J, Bloch K (1964) Biosynthesis of unsaturated fatty acids in microorganisms - structures and biosynthetic pathways are compared and related to physiological properties of the organims. Science 143: 1006-1012.

7. Johns RB, Perry GJ (1977) Lipids of the marine bacterium *Flexibacter polymorphus*. Arch Microbiol 114: 267-271.

8. Delong EF, Yayanos AA (1986) Biochemical function and ecological significance of novel bacterial lipids in deep-sea procaryotes. Appl Environ Microbiol 51: 730-737.

9. Nichols BW, Wood BJB (1968) New glycolipid specific to nitrogen-fixing blue-green algae. Nature 217: 767-768.

10. Bryce TA, Welti D, Walsby AE, Nichols BW (1972) Monohexoside derivatives of long-chain polyhydroxy alcohols; a novel class of glycolipid specific to heterocystous alge. Phytochemistry 11: 295-302.

11. Lambein F, Wolk CP (1973) Structural studies on the glycolipids from the envelope of the heterocyst of *Anabaena cylindrica*. Biochemistry 12: 791-798.

12. Murry MA, Wolk CP (1989) Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two layers of the cell envelope. Arch Microbiol 151: 469-474.

13. Su CJ, Reusch RN, Sadoff HL (1981) Isolation and characterization of several unique lipids from *Azotobacter vinelandii* cysts. J Bacteriol 147: 80-90.

14. Reusch RN, Sadoff HL (1983) Novel lipid components of the *Azotobacter vinelandii* cyst membrane. Nature 302: 268-270.

15. Metz JG, Roessler P, Facciotti D, Levering C, Dittrich F, et al. (2001) Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. Science 293: 290-293.

16. Miyanaga A, Funa N, Awakawa T, Horinouchi S (2008) Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. Proc Natl Acad Sci USA 105: 871-876.

17. Funa N, Ozawa H, Hirata A, Horinouchi S (2006) Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in *Azotobacter vinelandii*. Proc Natl Acad Sci USA 103: 6356-6361.

18. Staunton J, Wilkinson B (1997) Biosynthesis of erythromycin and rapamycin. Chemical Reviews 97: 2611-2629.

19. Du LC, Sanchez C, Chen M, Edwards DJ, Shen B (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. Chem Biol 7: 623-642.

20. Chang ZX, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, et al. (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula. J Nat Prod 67: 1356-1367.

21. Hopwood DA (1997) Genetic contributions to understanding polyketide synthases. Chemical Reviews 97: 2465-2497.

22. Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. Chemical Reviews 106: 3468-3496.

23. Staunton J, Weissman KJ (2001) Polyketide biosynthesis: a millennium review. Natural Product Reports 18: 380-416.

24. Jenke-Kodama H, Sandmann A, Muller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol 22: 2027-2039.

25. White SW, Zheng J, Zhang YM, Rock CO (2005) The structural biology of type II fatty acid biosynthesis. Annu Rev Biochem 74: 791-831.

26. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496-512.

27. Challis GL (2008) Genome mining for novel natural product discovery. J Med Chem 51: 2618-2628.

28. Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57: 369-394.

29. Motoigi T, Okuyama H (2011) Fatty acid and hydrocarbon composition in tropical marine Shewanella amazonensis strain SB2B(T). J Basic Microbiol 51: 484-489.

30. Jones AC, Monroe EA, Podell S, Hess WR, Klages S, et al. (2011) Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium Lyngbya majuscula. Proceedings of the National Academy of Sciences of the United States of America 108: 8815-8820.

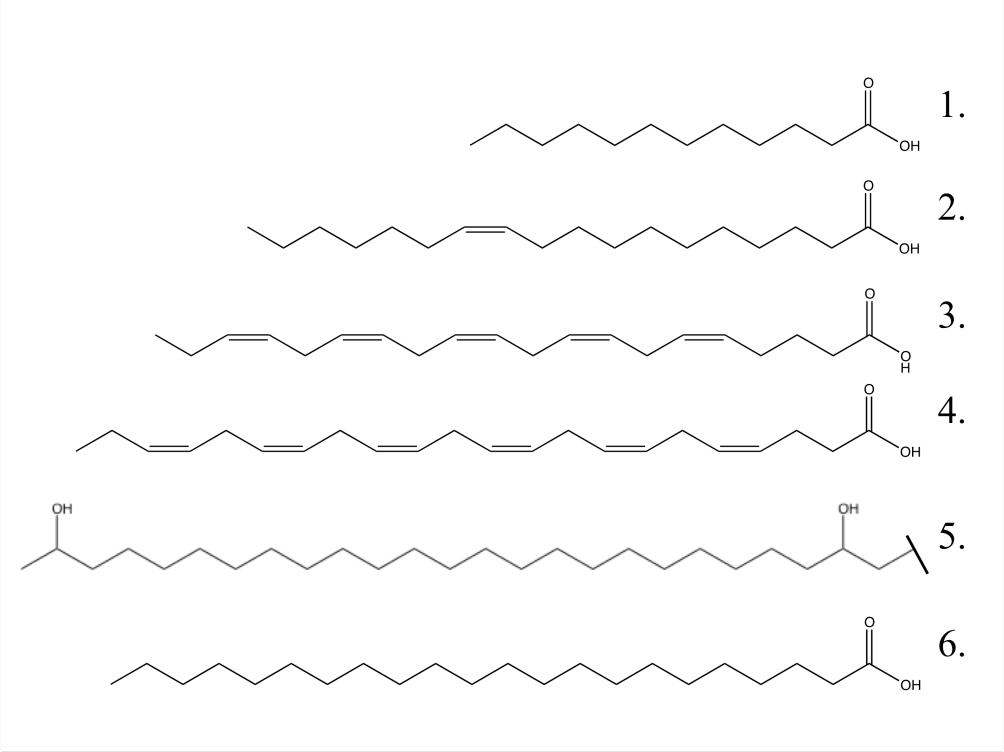

Figure 1. Structures of notable fatty acids/fatty acyl chains. Lauric acid (**1**); cis-vaccenic acid (**2**); eicosapentaenoic acid (**3**); docosahexaenoic acid (**4**); 3,25-hexacosanediol (**5**); behenic acid (**6**).

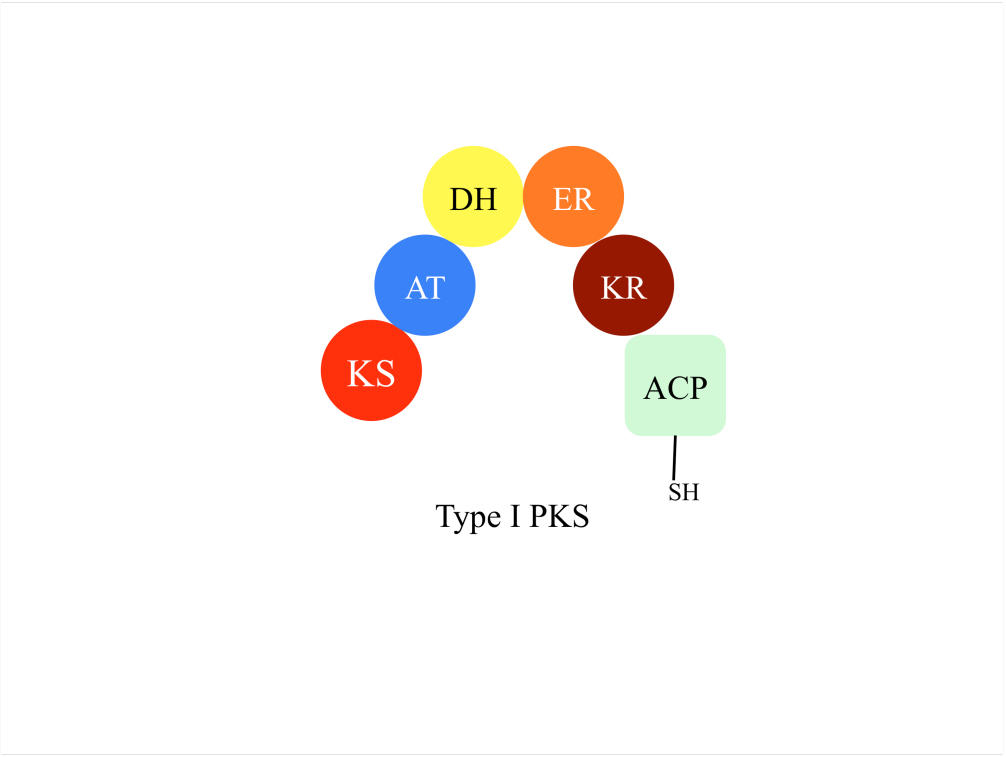

Figure 2. Schematic representation of the domain architecture of a Type I PKS. Adapated from Fischbach and Walsh, 2006.

Chapter 1:

Widespread Occurrence of Secondary Lipid Biosynthesis Potential in Microbial Lineages

OPEN ACCESS Freely available online

PLoS one

Widespread Occurrence of Secondary Lipid Biosynthesis Potential in Microbial Lineages

Christine N. Shulse[1], Eric E. Allen[1,2*]

1 Division of Biological Sciences, University of California San Diego, La Jolla, California, United States of America, 2 Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, United States of America

Abstract

Bacterial production of long-chain omega-3 polyunsaturated fatty acids (PUFAs), such as eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3), is constrained to a narrow subset of marine c-proteobacteria. The genes responsible for de novo bacterial PUFA biosynthesis, designated pfaEABCD, encode large, multi-domain protein complexes akin to type I iterative fatty acid and polyketide synthases, herein referred to as "Pfa synthases". In addition to the archetypal Pfa synthase gene products from marine bacteria, we have identified homologous type I FAS/PKS gene clusters in diverse microbial lineages spanning 45 genera representing 10 phyla, presumed to be involved in long-chain fatty acid biosynthesis. In total, 20 distinct types of gene clusters were identified. Collectively, we propose the designation of "secondary lipids" to describe these biosynthetic pathways and products, a proposition consistent with the "secondary metabolite" vernacular. Phylogenomic analysis reveals a high degree of functional conservation within distinct biosynthetic pathways. Incongruence between secondary lipid synthase functional clades and taxonomic group membership combined with the lack of orthologous gene clusters in closely related strains suggests horizontal gene transfer has contributed to the dissemination of specialized lipid biosynthetic activities across disparate microbial lineages.

Citation: Shulse CN, Allen EE (2011) Widespread Occurrence of Secondary Lipid Biosynthesis Potential in Microbial Lineages. PLoS ONE 6(5): e20146. doi:10.1371/journal.pone.0020146

Editor: Jack Anthony Gilbert, Argonne National Laboratory, United States of America

Received March 29, 2011; Accepted April 19, 2011; Published May 19, 2011

Funding: A gift from the Legler Benbough Foundation: http://www.benbough.org/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: eallen@ucsd.edu

Introduction

Bacteria have evolved the capacity for fatty acid biosynthesis for incorporation into membrane phospholipids in three distinct ways. The most common mechanism is the prototypical type II Fatty Acid Synthase (FAS II), well characterized in *E. coli* [1]. In this system individual enzymatic activities reside on discrete enzyme products, encoded by the **f**atty **a**cid **b**iosynthesis, or *fab*, genes. An alternative pathway, albeit significantly less pervasive in bacterial lineages, is the type I FAS system (FAS I). The canonical pathway found in eukaryotic organisms, FAS I is also found in the Corynebacterineae of the order Actinomycetales [2]. FAS I consists of a large, multifunctional biosynthetic complex containing all enzymatic domains necessary for acyl chain elongation and functional derivatization and is responsible for the production of both membrane phospholipid fatty acyl chains as well as precursor fatty acid molecules for elongation to long-chain mycolic acids in members of the Corynebacteriaceae, Mycobacteriaceae and Nocardiaceae families [3].

A third mechanism of *de novo* fatty acid synthesis coexists with the FAS II in a narrow subset of marine Gammaproteobacteria [4,5]. This pathway consists of a novel iterative FAS/PKS system and is responsible for the production of long-chain omega-3 polyunsaturated fatty acids (PUFAs) such as eicosapentaenoic acid (EPA, 20:5*n*-3) and docosahexaenoic acid (DHA, 22:6*n*-3) [6,7,8]. The genes responsible for bacterial omega-3 PUFA production, designated *pfaA–E*, possess multiple fatty acid biosynthetic enzyme activities as integrated domains within operon-encoded gene products [9].

In addition to polyenoic fatty acyl products, related FAS/PKS gene clusters have been shown to synthesize other specialized long-chain fatty acid products. The C_{26} to C_{32} fatty acid alkyl chains containing hydroxyl or ketone moieties found in the heterocyst glycolipids of filamentous nitrogen-fixing cyanobacteria [10,11] and the C_{22} to C_{26} fatty acids of phenolic lipids comprising the dormant cysts of the gram-negative bacterium *Azotobacter vinelandii* [12] are both synthesized via an analogous iterative type I FAS/PKS mechanism. Their long chain length, typically containing $\geq$20 carbons, distinguishes these fatty acyl products from those produced by FAS II ($\leq$18 carbons). We collectively term these specialized lipid products *"secondary lipids"*, to emphasize the accessory nature of these lipid molecules and distinguish these products from those synthesized via core, or primary, fatty acid biosynthetic mechanisms.

All three of these biosynthesis methods make use of highly conserved enzyme activities encoded on integrated domains (FAS I and Pfa synthases) or autonomous monofunctional gene products (FAS II) to accomplish the cycle of condensation, reduction, dehydration, and reduction necessary to produce a final fatty acid product. The biosynthetic reaction sequence of an elongation cycle includes the following activities: 1) Ketoacyl synthase [KS] catalyzes a condensing function responsible for chain-elongation; 2) Ketoacyl reductase [KR] catalyzes the reduction of the carbonyl group to a hydroxyl group; 3) Dehydratase/Isomerase [DH/I]

catalyzes the dehydration of the b-hydroxyacyl intermediate generated by KR to a *trans*-2-enoyl derivative and the subsequent isomerization from the *trans*-2 to the *cis*-3 configuration; and lastly 4) Enoyl reductase [ER] catalyzes the reduction of double bond generated by DH to complete the chain elongation process. Other essential activities include acyl carrier protein [ACP] function which tethers the growing fatty acyl chain as a thioester as it is acted upon by other enzyme activities and phosphopantetheinyl transferase [PPTase] activity which converts ACP products from the inactive apo- form to the active holo-form via the posttranslational addition of a 4′-phosphopantetheine prosthetic group from acetyl coenzyme A. Acyltransferases [AT] catalyze the general transfer of a nascent acyl substrate from acyl-CoA to ACP for elongation of the fatty acyl chain. Malonyl-CoA:ACP transacylase [MAT] is a type of acyltranferase that specifically catalyzes the transfer of a 2C malonyl moiety from malonyl-CoA to ACP. A final component, specific to PKSs and Pfa synthases, is the chain length factor [CLF] domain shown to determine the ultimate chain length of the reaction product [13].

Evidence for additional widespread capacity for secondary lipid production potential via the Pfa synthase mechanism in environmental samples has recently been reported [14]. In that study, culture-independent molecular surveys were used to identify 13 novel groups based on KS domain homology from disparate marine habitats. Beyond the marine environment, the phylogenetic extent and ecological breadth of secondary lipid biosynthetic potential has not been investigated.

The current capacity for inexpensive, rapid, whole genome sequencing has allowed for broad genome comparisons among diverse microbial lineages. Here, we expand upon previous studies of the distribution and diversity of secondary lipid production potential in the marine environment [14] by analyzing all sequenced microbial genomes for the presence of FAS/PKS gene clusters homologous to those involved in PUFA secondary lipid biosynthesis. We uncover and classify multiple previously unrecognized FAS/PKS gene clusters in diverse bacterial lineages representing varying physiologies and life histories, significantly expanding the palette and pervasiveness of gene products linked to specialized microbial metabolites.

Results and Discussion

In the following sections, we describe the diversity and organization of secondary lipid biosynthetic gene clusters identified in this study, first addressing those with characterized products then progressing to novel clusters with uncharacterized products. Additional analyses are presented to support the definition of secondary lipid synthases and differentiate these gene clusters from those involved in PKS or NRPS products based on PPTase domain conservation. Next, we analyze the genomes of secondary lipid synthase containing organisms for the presence of other lipid biosynthetic activities, including FAS II and *ole* gene functions, two systems whose products interact with *pfa* gene products. Lastly, we analyze the ecology and physiological properties of these organisms to provide insight into possible traits unifying secondary lipid production potential and present evidence showing that horizontal gene transfer has aided in the dissemination of these biosynthetic gene clusters.

Diversity and organization of FAS/PKS gene clusters

The presence of multiple acyl carrier protein (ACP) domains in a single gene product is a distinguishing characteristic of Pfa synthases. It has been shown that an increase in the number of ACPs increases the biosynthetic throughput of PUFA product synthesis [15]. Most FAS/PKS gene clusters retrieved in this study contain multiple tandem ACP domains (**Figure 1**). However, although all gene clusters with tandem ACPs are presumed to produce fatty acyl products, not all fatty acyl-producing FAS/PKS gene clusters contain multiple ACPs (e.g. *ars* gene cluster responsible for the production of alkylresorcinols and alkylpyrones in *Azotobacter vinelandii*). Therefore we did not exclude gene clusters containing a single ACP if the domain content and organization was consistent within a candidate FAS/PKS cluster and phylogenetic analysis of the proximal ketoacyl synthase (KS) domain, harbored within the *pfaA* homolog (**Figure 1**), supported a common evolutionary relationship with validated FAS/PKS pathways.

A total of twenty distinct classes of FAS/PKS gene clusters, designated Types A–T (**Figure 1**), were identified in 86 finished or draft genomes available in GenBank and/or the Joint Genome Institute's Integrated Microbial Genomes (IMG) databases as of February 2011 (2.2% of 3839 genomes analyzed). The division of gene clusters into "Types" was strongly supported by three independent metrics: (i) domain count and organization as analyzed by non-parametric multidimensional scaling (NMDS; **Figure S1**); (ii) phylogenetic analysis of component enzymatic domains in each secondary lipid synthase identified (**Figure 2** and **Figure S2**); and (iii) pathway-product information, if known. Alphabetic ordering of cluster "Types" is based on phylogeny of KS domains (**Figure 2**). All FAS/PKS gene clusters analysed were found to contain at least one KS domain, ACP domain, and ketoacyl reductase (KR) domain, and various combinations of malonyl-CoA:ACP transacylase (MAT), acyltransferase (AT), chain length factor (CLF), dehydratase/isomerase (DH/I), enoyl reductase (ER), and phosphopantetheinyl transferase (PPTase) domains. It is important to investigate both the domain content and organization of each gene cluster as these factors determine the ultimate length and functionality of the chemical product [16]. The NMDS plot (**Figure S1**) is a visual representation of the similarity among all the gene clusters based on these two factors.

Secondary lipid synthase types with characterized products. Type A and Type B produce omega-3 PUFAs and are found primarily in c-proteobacteria of marine origin (**Table 1**). Type A represents the canonical secondary lipid synthesizing gene cluster responsible for eicosapentaenoic acid (EPA, 20:5*n*-3) synthesis and consists of five genes, *pfaABCDE* [9]. Domain order within these genes is highly conserved: *pfaA* [KS-MAT-$ACP_{(4-6)}$-KR], *pfaB* [AT], *pfaC* [KS-CLF-DH_2], *pfaD* [ER], and *pfaE* [PPTase] (**Figure 1**). An exception to this conservation is found in *Pseudoalteromonas* sp. DS-12 where the PPTase domain is incorporated into *pfaC* [17]. Intriguingly, analysis of the DH domains in *Pseudoalteromonas* sp. DS-12 reveals that the first DH domain is phylogenetically more similar to the second DH domain from all other gene clusters harboring two DH domains, and vice versa (**Figure S2**). This implies that at some point in the evolution of this gene cluster a section of *pfaC* and all of *pfaE* were translocated. It is unknown whether the *Pseudoalteromonas* gene cluster retains the ability to produce EPA [17].

The Type A cluster was previously found to be conserved in 15 sequenced *Shewanella* strains [18]. In the present study, nine additional *Shewanella* genomes were analyzed and all were found to contain a coherent Type A gene cluster (**Table 1**). The presence of a TypeA *pfa* gene operon in all genomically characterized members of the Shewanellaceae (n = 24) suggests the genetic potential for EPA production is a defining characteristic of this lineage. The *Shewanella* are a genus of Gammaproteobacteria known for their ability to utilize a wide variety of electron acceptors and have been recovered from diverse environmental sources [19,20,21]. In addition to the 24 analyzed *Shewanella*

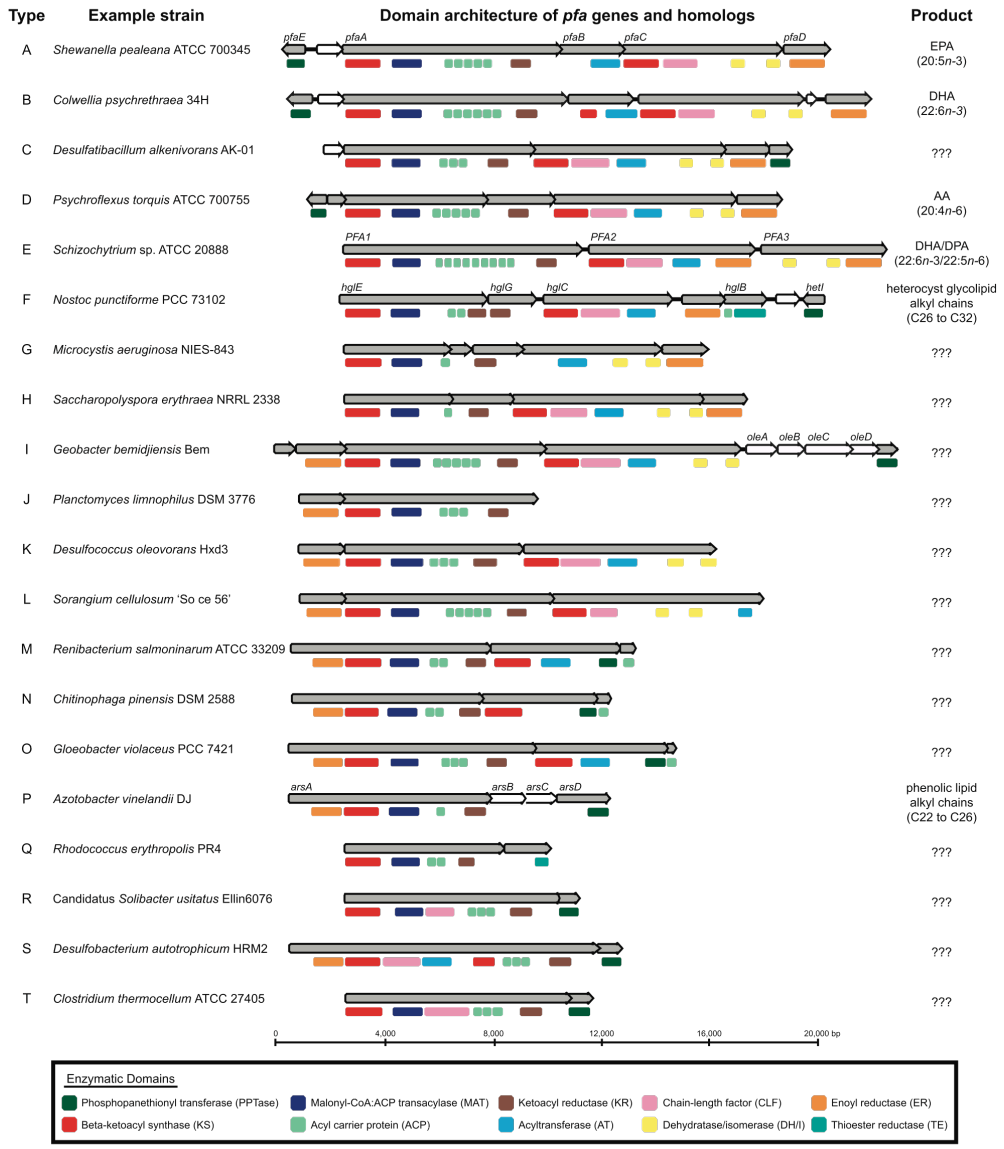

Figure 1. Diversity of *pfa*-like gene clusters. Eighty-six microbial genomes with either the *pfa* gene cluster or a homolog were identified. Type designations are based on KS phylogeny, conserved domain structure and product information, if available.
doi:10.1371/journal.pone.0020146.g001

genomes, the EPA-synthesizing Type A gene cluster is also found in three Vibrio genomes (Vibrio sp. MED222, V. splendidus 12B01, and V. splendidus LGP32), two Photobacterium genomes [9] and one Pseudoalteromonas genome [17].

Type B represents the docosahexaenoic acid (DHA)-producing pfa gene cluster, which differs from the EPA-producing pfa cluster by the insertion of an additional KS domain in pfaB (Figure 1). An active site cysteine was found to be absent in all Type B pfaB

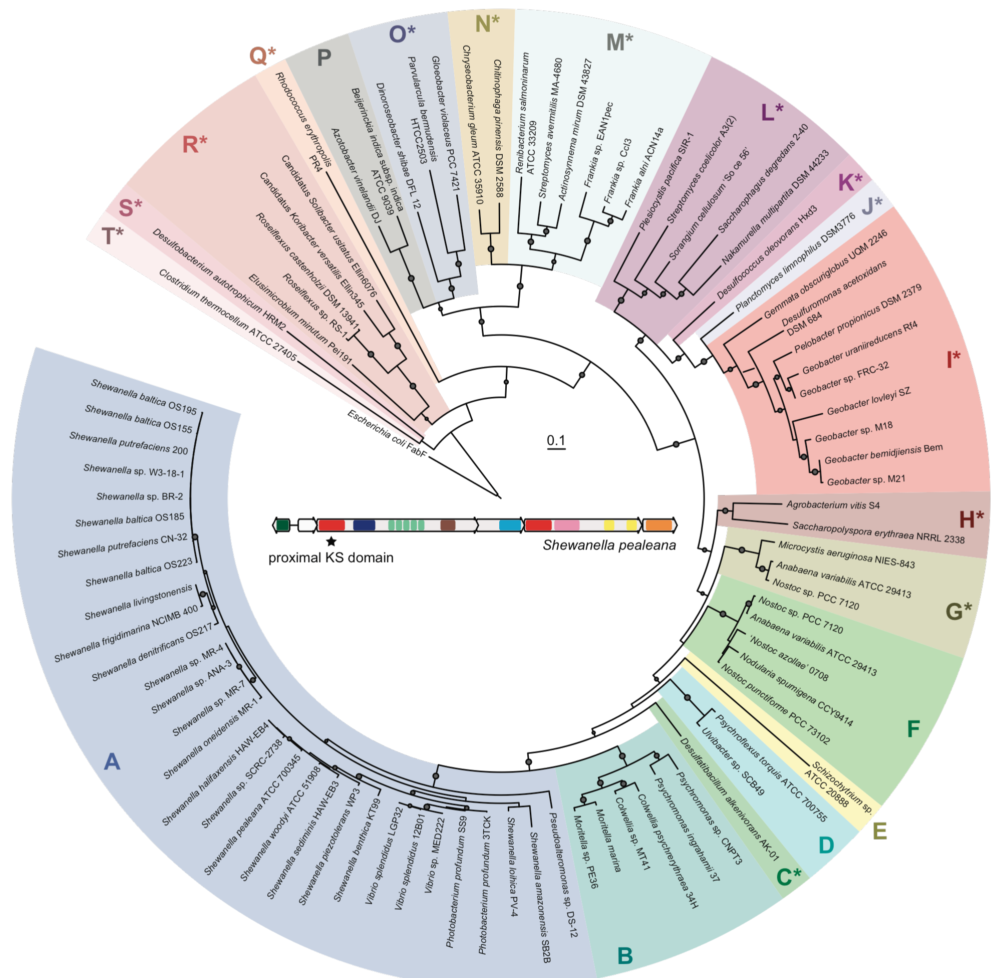

Figure 2. Maximum likelihood phylogenetic tree of proximal keto-acyl synthase protein domains (364 conserved amino acids). Gene cluster Types are colored and given a letter label. Asterisks represent Types first clustered and described in this study. Bootstrap values ≥50% are indicated by dots. The *Escherichia coli* DH10B FabF protein was used as the outgroup.
doi:10.1371/journal.pone.0020146.g002

KS domains examined in the current study. A recent study suggested that this *pfaB* KS domain may play a role in determining the final PUFA end product, as *E. coli* transformed with *pfaA*, *pfaC*, *pfaD* and *pfaE* from an EPA-producer and *pfaB* from a DHA-producer made both EPA and DHA [22]. The Type B gene cluster is found exclusively in marine c-proteobacteria representing three genera: *Colwellia* [23], *Psychromonas* [24], and *Moritella* [8]. Although the Type B gene cluster retains a completely conserved domain structure in the genomes of the *Colwellia*, *Moritella*, and *Psychromonas* and the proximal *pfaA* KSs from these species group together at 74% amino acid sequence identity, *Colwellia* and *Moritella* are capable of producing significant quantities of DHA while the *Psychromonas* that have been investigated either do not produce DHA or produce it only in trace quantities [25,26].

Type D is found in two Bacteriodetes strains (**Table 1**), of which *Psychroflexus torquis* ATCC 700755 is known to produce arachidonic acid (AA, 20:4*n*-6) and EPA [27]. The Type D gene cluster has the same domain content as the Type A EPA-producing gene cluster, however the AT domain is rearranged and domains reside on split or fused gene products – the *pfaA* homolog

Table 1. Membership and description of secondary lipid synthase types with characterized products.

Type	Organism(s)	Unique Characteristics	Product
A	*Photobacterium profundum* SS9 and 3TCK | *Pseudoalteromonas* sp. DS-12 | *Shewanella amazonensis* SB2B, *S. baltica* OS155, OS185, OS195 and OS223 *S. benthica* KT99, *S. denitrificans* OS217, *S. frigidimarina* NCIMB 400, *S. halifaxensis* HAW-EB4, *S. livingstonensis*, *S. loihica* PV-4, *S. oneidensis* MR-1, *S. pealeana* ATCC 700345, *S. piezotolerans* WP3, *S. putrefaciens* 200 and CN-32, *S. sediminis* HAW-EB3, *S.* sp. ANA-3, sp. BR-2, sp. MR-4, sp. MR-7, sp. SCRC-2738, sp. W3-18-1, *S. woodyi* ATCC 51908 | *Vibrio* sp. MED222, *V. splendidus* 12B01 and LGP32	Marine γ-proteobacteria	Eicosapentaenoic acid (EPA, 20:5*n*-3)
B	*Colwellia psychrerythraea* 34H, *C.* sp. MT41 | *Moritella marina*, *Moritella* sp. PE36 | *Psychromonas ingrahamii* 37, *P.* sp. CNPT3	Marine γ-proteobacteria	Docosahexaenoic acid (DHA, 22:6*n*-3)
D	*Psychroflexus torquis* ATCC 700755 | *Ulvibacter* sp. SCB49	Marine Bacteroidetes[1,2]	Arachidonic acid (AA, 20:4*n*-6) and EPA
E	*Schizochytrium* sp. ATCC 20888	Osmoheterotrophic Protist	Docosapentaenoic acid (DPA, 22:5*n*-6) and DHA
F	*Anabaena variabilis* ATCC 29413 | *Nodularia spumigena* CCY9414 | '*Nostoc azollae*' 0708, *Nostoc punctiforme* PCC 73102, *N.* sp. PCC 7120	Nitrogen fixation; Cyanobacteria	Heterocyst glycolipid alkyl chains (e.g. hexacosanediol, C_{26})
P	*Azotobacter vinelandii* DJ | *Beijerinckia indica* subsp. indica ATCC 9039	N_2 fixation; soil bacteria with varying pH requirements[3]	Phenolic lipid alkyl chains (e.g. behenic acid, C_{22})

[1]Bowman JP, McCammon SA, Lewis T, Skerratt JH, Brown JL, et al. (1998) *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson et al. 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiol* 144: 1601–1609.
[2]Pinhassi J, Zweifel UL, Hagstrom A (1997) Dominant marine bacterioplankton species found among colony-forming bacteria. *Appl Environ Microbiol* 63: 3359–3366.
[3]Starkey RL, De PK (1939) A new species of *Azotobacter*. *Soil Sci*ence 47: 329–343.
doi:10.1371/journal.pone.0020146.t001

is split into two genes, designated pfaA1 [KS-AT-ACP_5] and pfaA2 [KR], and pfaB and pfaC are fused into one gene, pfaBC [KS-CLF-AT-DH_2] (Figure 1).

The Type E gene cluster is found in the marine thraustochytrid Schizochytrium sp. ATCC 20888 (order Labyrinthulida) and is involved in the production of DHA and the omega-6 PUFA docosapentaenoic acid (DPA, 22:5n-6) [6]. This gene cluster consists of three genes, designated PFA1, 2, and 3, and contains the same domains found in the five genes of the Type A cluster. Although the Type E gene cluster is clearly homologous to the Type A and B clusters, differences do exist. For example, in contrast to the Type A and B gene clusters, each containing one ER domain, the Type E cluster contains two, one located on PFA2 and the other on PFA3 (Figure 1). Sequence composition analyses indicate that these Schizochytrium ER domains share 89% identity at the amino acid level. Thus one of the ER domains may have resulted from a duplication and translocation event after transfer of an ancestral Pfa synthase gene cluster into the Schizochytrium genome. In addition to the Type E cluster for PUFA synthesis, Schizochytrium sp. ATCC 20888 also contains the canonical eukaryotic PUFA biosynthesis pathway, involving elongase and desaturase enzymes [28]. The presence of a bacterial Pfa synthase in Schizochytrium is significant as it suggests evidence of lateral gene transfer between a bacterium and eukaryote. The integration and retention of a Pfa synthase in the Schizochytrium genome may contribute to the high PUFA content observed in certain labyrinthulids [29,30]. DHA- and DPA-producing thraustochytrids have been isolated from marine environments around the globe [31], although the relative contribution of each pathway to PUFA synthesis has not been determined in these strains.

Type F is found in the genomes of five nitrogen-fixing cyanobacteria. First characterized in the cyanobacterium Nostoc punctiforme strain ATCC 29133, the hgl genes, for heterocyst glycolipid [10], were shown to be involved in the production of the lipid moiety of heterocyst glycolipids. The other four cyanobacteria containing the Type F gene cluster, Nostoc azollae 0708, Anabaena variabilis ATCC 29413, Nodularia spumigena CCY 9414, and Nostoc sp. PCC 7120, also form heterocysts for nitrogen fixation. The Type F gene cluster is remarkably similar to Types A and B. One notable difference is the fusion of some of the domains contained on pfaB and pfaC on a single gene, hglC [KS-CLF-AT]. Additionally, the Type F cluster lacks the DH/I domains typically found on pfaC.

The final Type with a characterized pathway for which the chemical product has been verified is the Type P gene cluster, found in the genomes of two nitrogen-fixing Proteobacteria. This synthase is responsible for the production of the alkyl moiety of phenolic lipids in the cyst-forming Gammaproteobacterium Azotobacter vinelandii [12]. The Type P gene cluster is also found in the Alphaproteobacterium Beijerinckia indica subsp. Indica ATCC 9039, which has not been found to produce cysts or phenolic lipids [32].

Secondary lipid synthase types with uncharacterized products. In addition to the above pathways with characterized products, 14 additional gene clusters homologous to the pfa genes were discovered with putative fatty acyl end products (Table 2). The Type C cluster has been found in only one sequenced bacterial genome, that of the Deltaproteobacterium Desulfatibacillum alkenivorans AK-01, which has not been reported to produce PUFAs. This gene cluster contains a pfaBC fusion [KS-CLF-AT-DH_2] and pfaD [ER] is located at the 3′ end of the gene cluster.

The Type G gene cluster, like the characterized Type F cluster, is found in cyanobacterial genomes. In fact, in the case of Anabaena

Table 2. Membership and description of secondary lipid synthase types with uncharacterized products.

Type	Organism(s)	Unique Characteristics
C	*Desulfatibacillum alkenivorans* AK-01	Isolated from sediments[1]
G	*Anabaena variabilis* ATCC 29413 | *Microcystis aeruginosa* NIES-843 | *Nostoc* sp. PCC 7120	Cyanobacteria
H	*Agrobacterium vitis* S4 | *Saccharopolyspora erythraea* NRRL 2338	Plant pathogen[2]; Soil bacterium[3]
I	*Desulfuromonas acetoxidans* DSM 684 | *Gemmata obscuriglobus* UQM 2246 | *Geobacter bemidjiensis* Bem, *G. lovleyi* SZ, *G.* sp. FRC-32, sp. M18, sp. M21, *G. uraniireducens* Rf4 | *Pelobacter propionicus* DSM 2379	Deltaproteobacteria, except *G. obscuriglobus* (Planctomycete)
J	*Planctomyces limnophilus* DSM 3776	Planctomycete
K	*Desulfococcus oleovorans* Hxd3	Sulfate-reducer[4]
L	*Nakamurella multipartita* DSM 44233 | *Plesiocystis pacifica* SIR-1 | *Saccharophagus degredans* 2-40 | *Sorangium cellulosum* 'So ce 56' | *Streptomyces coelicolor* A3(2), *S. ghanensis* ATCC 14672	Actinobacteria, Deltaproteobacteria, Gammaproteobacteria
M	*Actinosynnema mirum* DSM 43827 | *Frankia alni* ACN14a, *F.* sp. CcI3, *F.* sp. EAN1pec | *Renibacterium salmoninarum* ATCC 33209 | *Streptomyces avermitilis* MA-4680	*Frankia* = nitrogen fixers
N	*Chitinophaga pinensis* DSM 2588 | *Chryseobacterium gleum* ATCC 35910	Bacteroidetes
O	*Dinoroseobacter shibae* DFL 12 | *Gloeobacter violaceus* PCC 7421 | *Parvularcula bermudensis* HTCC2503	*Gloeobacter violaceus* PCC 7421 produces PUFAs (18:2, 18:3); desaturases[5]
Q	*Rhodococcus erythropolis* PR4	Alkane-degrader[6]
R	Candidatus *Koribacter versatilis* Ellin345 | Candidatus *Solibacter usitatus* Ellin6076 | *Elusimicrobium minutum* Pei191 | *Roseiflexus castenholzii* DSM 13941, *R.* sp. RS-1	
S	*Desulfobacterium autotrophicum* HRM2	Sulfate-reducer[7]
T	*Clostridium thermocellum* ATCC 27405	Thermophilic, anaerobic[8]

[1]So CM, Young LY (1999) Isolation and characterization of a sulfate-reducing bacterium that anaerobically degrades alkanes. Applied and Environmental Microbiology 65: 2969–2976.
[2]Ophel K, Kerr A (1990) *Agrobacterium vitis* sp. nov. for Strains of *Agrobacterium* biovar 3 from Grapevines. International Journal of Systematic Bacteriology 40: 236–241.
[3]Labeda DP (1987) Transfer of the Type Strain of *Streptomyces erythraeus* (Waksman 1923) Waksman and Henrici 1948 to the Genus *Saccharopolyspora* Lacy and Goodfellow 1975 as *Saccharopolyspora erythraea* sp. nov., and Designation of a Neotype Strain for *Streptomyces erythraeus*. International Journal of Systematic Bacteriology 37: 19–22.
[4]Aeckersberg F, Bak F, Widdel F (1991) Anaerobic oxidation of saturated hydrocarbons to CO2 by a new type of sulfate-reducing bacterium. Archives of Microbiology 156: 5–14.
[5]Chi XY, Yang QL, Zhao FQ, Qin S, Yang Y, et al. (2008) Comparative Analysis of Fatty Acid Desaturases in Cyanobacterial Genomes. Comparative and Functional Genomics.
[6]KomukaiNakamura S, Sugiura K, Yamauchilnomata Y, Toki H, Venkateswaran K, et al. (1996) Construction of bacterial consortia that degrade Arabian light crude oil. Journal of Fermentation and Bioengineering 82: 570–574.
[7]Brysch K, Schneider C, Fuchs G, Widdel F (1987) Lithoautotrophic growth of sulfate-reducing bacteria, and description of *Desulfobacterium autotrophicum* gen. nov., sp. nov. Archives of Microbiology 148: 264–274.
[8]McBee R (1954) The characteristics of *Clostridium thermocellum*. Journal of Bacteriology 67: 505–506.
doi:10.1371/journal.pone.0020146.t002

variabilis ATCC 29413 and *Nostoc* sp. PCC 7120, the Type F and Type G gene clusters coexist in the same genome. Notably, the Type G gene cluster is always found to have two uncharacterized type I PKS genes directly upstream. The product of the Type G gene cluster may interact with the product of the linked type I PKS genes, producing a hybrid polyketide/fatty acid product. A precedent for the interaction of products from a PKS and a FAS/PKS exists, for example, in the case of phenolic lipid production in *Azotobacter vinelandii* (Type P). In this case, the phenolic functional group is produced by two type III PKS genes, which are flanked on both sides by the Type P FAS/PKS genes responsible for the production of the C_{22-26} fatty acid moiety of phenolic lipids. A similar head/tail functional group production relationship could be occurring between the Type G gene cluster and the upstream PKS genes.

Agrobacterium vitis S4 (class/subphylum Alphaproteobacteria) and *Saccharopolyspora erythraea* NRRL 2338 (phylum Actinobacteria) harbor the Type H gene cluster (**Table 2**). This gene cluster is defined by a split *pfaA* and a fused *pfaBC* homolog (**Figure 1**).

Type I contains a *pfaBC* fusion and is found in multiple genera of Deltaproteobacteria and one Planctomycete (**Table 2**). The Type J gene cluster is only found in one genome, that of *Planctomyces limnophilus*. It consists of just two genes, a *pfaD* homolog [ER] followed by a *pfaA* homolog [KS-MAT-ACP_3-KR] (**Figure 1**). Phylogenetically, the Type J KS domain is closely related to the KS domain from the Type I gene cluster (**Figure 2**). This apparently partial gene cluster could have arisen after a partial transfer of a more complete Type I-like cluster. The Type K gene cluster also has a similar configuration to the Type I cluster (**Figure 1**) and is found in a single Deltaproteobacterium, *Desulfococcus oleovorans* Hxd3 (**Table 2**).

The Type L gene cluster is distinguished by a fused *pfaBC* [KS-CLF-DH_2-AT] homolog, and a *pfaD* [ER] homolog at the 5′ end of the gene cluster. It has been suggested that this gene cluster plays a role in the production of polyunsaturated fatty acids in *Streptomyces coelicolor* A3(2) [33] however PUFA production has not been demonstrated in this strain.

The Type M cluster is conserved in all three finished genomes from strains representing the nitrogen-fixing genus *Frankia* (phylum Actinobacteria) (**Table 2**). This is especially significant given that these genomes range in size from 5.43 to 9.04 Mbp, indicating extensive genome expansion, contraction, and adaptation to specific niches [34].

The Type N cluster is found in the chitin-degrading species Chitinophaga pinensis DSM 2588, which forms spherical ''resting bodies,'' or microcysts, upon aging [35], as well as the clinical isolate Chryseobacterium gleum ATCC 35910. The composition of the C. pinensis DSM 2588 ''resting body'' has not been determined, and it is possible that the product of the Type N plays a structural role in these differentiated cells, a proposal consistent with the structural role of phenolic lipids (Type P) in differentiated cysts of A. vinelandii [36]. However Chryseobacterium gleum ATCC 35910 is not known to produce differentiated cell types. The Type O gene cluster is found in the genomes of two marine Alphaproteobacteria, Dinoroseobacter shibae DFL 12 and Parvularcula bermudensis HTCC2503, and the rock-dwelling cyanobacterium Gloeobacter violaceus PCC 7421. The Type Q gene cluster is found in a single bacterial strain, Rhodococcus erythropolis PR4 (phylum Actinobacteria) (Table 2).

The Type R cluster is dispersed throughout several diverse microbial phyla, including the Acidobacteria (Candidatus Koribacter versatilis Ellin345 and Candidatus Solibacter usitatus Ellin6076), Elusimicrobia (Elusimicrobium minutum Pei191), and Chloroflexi (Roseiflexus castenholzii DSM 13941 and Roseiflexus sp. RS-1). The Type R cluster consists of one pfaA homolog containing the domain architecture KS-MAT-CLF-$ACP_{(3-6)}$-KR followed by a downstream gene encoding phosphopantetheinyl tranferase (PPTase) activity. Although there is strong evidence that this gene cluster shares an evolutionary origin (Figure 2), it is possible that the fatty acyl products are modified to produce different end products in these physiologically diverse and ecologically distinct species. The Type R core gene cluster homologous to the pfa genes is often associated with other genes involved in lipid production. For example, in Korebacter versatilis Ellin345, the pfaA and pfaE homologs are separated by a gene encoding part of the acetyl-CoA carboxylase (ACC) complex, which catalyzes the carboxylation of acetyl-CoA to produce malonyl-CoA for fatty acid biosynthesis. This intervening gene is absent in the remaining four gene clusters comprising Type R.

Types S and T are each found in single representatives (Table 2).

Analysis of phosphopantetheinyl transferase domains

One characteristic distinguishing PKSs from lipid synthases is the molecular sequence of the associated PPTase products responsible for posttranslational modification and activation of ACP domains [37]. To provide additional evidence that we have identified novel secondary lipid synthases rather than PKSs, we investigated PPTase domain diversity. The informative variable motifs are known as P0, P1a, and P1b in PUFA-specific PPTases and 1A, P1a9, and P1b9 in PKS/NRPS-specific PPTases. Secondary lipid synthase Types A, B, C, D, F, I, L, M, N, O, P, R, S, and T most often contain an associated PPTases in direct proximity to the Pfa synthase. Of these, Types A, B, C, D, F, I, and L contain the PUFA-associated P0 motif – defined as (L/V)Rx(L/V)LS in [37] and modified to hRxhLS in this comprehensive study of 69 secondary lipid-associated PPTases, where h = hydrophobic residue (Figure S3). Importantly, the PPTase from the Type P gene cluster, known to be involved in the production of phenolic lipids, does not contain the P0 motif. Consequently, the P0 domain may not be necessary for the production of secondary lipids in general. All secondary lipid Types investigated contain the PUFA-associated P1a motif – defined as K(G/D)KP in [37] and modified to x(G/D)xP in this study – rather than the PKS-related P1a9 motif. Exceptions were the Photobacterium PPTases which were previously shown to be EntD-like PPTases [37]. EntD in E. coli is involved in the production of siderophore enterobactin [38] and is included in the PKS/NRPS-producing PPTase group. Examination of the final domain region revealed that Types A, B, C, D, F, I, and L contain the P1b domain – FNxSH in [37] and modified to (F/S)NxSH in this study – while Types M, N, O, P, R, T, and S contain the P1b9 domain – GSIxH in [37] and modified to hShxH in this study (Figure S3). This bifurcation between Types A–L and Types M–T is also reflected in the KS tree (Figure 2) and could have implications for the type of end product produced from these gene clusters.

Relationship of secondary lipid pathways with FAS II

The relationship between secondary lipid pathways and FAS II is of interest as both pathways draw from the same intracellular pool of precursor molecules for their biosynthetic activities. All bacteria possessing the archetypal Pfa synthase (Type A and B) also possess a complete FAS II and generally produce polyunsaturated fatty acids as a small percentage of their total fatty acids, with saturated and monounsaturated fatty acids comprising the majority of their phospholipids [25]. However, several interesting exceptions exist among the bacteria harboring other putative secondary lipid synthases. The genome of the fish pathogen Renibacterium salmoninarum (Type M) lacks both fabA [DH/I] and fabZ [DH] homologs, at least one of which is necessary to perform the dehydration step involved in the biosynthesis of both saturated and monounsaturated fatty acids [39]. Based on genome analysis and growth capabilities it has been suggested that this strain requires exogenous fatty acids, possibly scavenged from its host, for incorporation into membrane phospholipids [40]. Consequently, the Type M pathway present in R. salmoninarum genome likely represents the only complete fatty acid biosynthesis pathway present in this strain. Likewise, Desulfobacterium autotrophicum HRM2 (Type S) is also missing fabA and fabZ homologs. Unlike the pathogenic R. salmoninarum, the completely oxidizing sulfate-reducing bacterium D. autotrophicum is free-living. Presumably, it could incorporate the fatty acids typically used as carbon sources into its membrane phospholipids. Given that D. autotrophicum harbors an incomplete FAS II, it is possible that the Type S secondary lipid synthase gene cluster is functioning as the ''core'' fatty acid synthase for this organism, thus the main producer of fatty acids for incorporation into phospholipids.

Co-occurrence of secondary lipid synthesis pathways with olefin-producing genes

Several major outstanding questions exist regarding the products synthesized by uncharacterized secondary lipid synthases described here (Types C, G, H, I J, K, L, M N, O, Q, R, S, T). It has previously been demonstrated that pfa gene products are necessary for the production of olefinic hydrocarbons in Shewanella onedensis MR-1 [41]. pfa genes and several secondary lipid synthase types described here co-occur with the ole genes responsible for the head-to-head condensation of fatty acyl products and the formation of olefinic neutral lipids [42], in 44 of the 80 (55%) sequenced genomes examined (Pseudoalteromonas sp. DS-12, Shewanella livingstonensis, Shewanella sp. BR-2, Shewanella sp. SCRC-2738, Moritella marina, and Schizochytrium sp. ATCC 20888 were excluded from this analysis as their genomes have not been sequenced). Furthermore, Geobacter bemidjiensis produces the same olefinic hydrocarbon, hentriacontanonaene ($C_{31}H_{46}$), found in Shewanella strains harboring the pfa gene cluster. Therefore, we suggest that the Type I ''pfa-like'' secondary lipid synthase gene cluster in the Deltaprotoebacteria and Planctomycete strains may provide precursors necessary for olefin biosynthesis. Additionally, it is possible that the Type L cluster contributes to the production

of olefinic hydrocarbons. *Nakamurella multipartita* DSM 44233 and *Plesiocystis pacifica* SIR-1 both contain *ole* genes [42].

These *ole* genes are often linked with the FAS/PKS gene clusters but can also reside elsewhere in the genome. In cases where unlinked *pfa* and *ole* genes exist, the transfer of these genes into the genomes was not coordinated, as bacteria possessing the same "Type" of *pfa* gene cluster often have differing *ole* gene configurations. An interesting question then arises regarding the primacy of these two biosynthetic gene clusters. One possibility is that the *pfa* homologs were originally producing a different product and were co-opted for the production of precursors for olefin biosynthesis over evolutionary time. It is also possible that the *pfa* homologs retain the ability to produce a stand-alone product and only contribute to olefin production under certain environmental conditions. Very little is known regarding the regulation and synthesis of olefin products.

Ecology of microorganisms possessing secondary lipid production potential

The microorganisms harboring the *pfa* genes and homologs belong to diverse phyla throughout the bacterial domain, exhibiting varied ecophysiologies and lifestyle strategies. Examination of this metadata provided important insight into the ecological basis of secondary lipid synthesis across the spectrum of biosynthetic diversity examined. A permutational analysis of variance (PERMOVA) was used to rigorously test the association between gene cluster Type and life history traits, for which there was adequate replication ($n \geq 5$, Types A, B, F, I, L, M, and R) (**Figure S4**). All gene clusters differed from each other in terms of life history traits (PERMANOVA pair-wise tests, $P = \leq 0.004$), with the exception that L did not differ from M or R, and M did not differ from R ($P = \geq 0.05$). Overall allocation success across all seven gene clusters equaled 60.6%, with all genes having a more distinct set of life history traits than expected by chance alone, with the exception of R. Type I was associated with being an obligate anaerobe. Many of the known fatty acid products produced by characterized Types are known to play a role in excluding oxygen from the cell (e.g. heterocyst glycolipids), or have been shown to provide antioxidative functions to the cell [43]. Consistent with these functions, the product of the Type I *pfa* homologs in the Deltaproteobacteria may play a role in protecting these sensitive cells from lethal levels of oxygen in the environment. A complete PERMOVA analysis on all secondary lipid synthase types is presented in **Figure S5**.

Contribution of horizontal gene transfer to the dissemination of secondary lipid synthase pathways

Horizontal gene transfer (HGT) may help explain the patchy distribution of secondary lipid pathways among bacterial and eukaryotic phyla. Specifically, multiple secondary lipid Types may be found in the same phyla, while a single Type may be found across several phyla (**Figure 3**). To examine the possibility of HGT, all species of the same genus having an identified secondary lipid synthase gene cluster with two or more sequenced genomes were analyzed for evidence of HGT. For example, the *pfa* gene cluster (Type A) is present in three strains of *Vibrio* but absent in many closely related *Vibrio* genomes. A comparison of the *pfa* gene region of *Vibrio splendidus* 12B01 and corresponding regions of the genomes of multiple *Vibrio* species is shown in **Figure 4**. The two genomic regions immediately flanking the *pfa* gene cluster show 93% and 87% identity at the nucleotide level, respectively, to their homologs in the Vibrionales bacterium SWAT-3 genome. The *pfa* gene cluster clearly shows a perfect insertion characteristic of a genomic island.

The Type I gene cluster has an especially sporadic distribution among the Deltaproteobacteria. Of the two sequenced *Pelobacter* strains, *Pelobacter propionicus* DSM 2379 harbors a Type I cluster while *Pelobacter carbinolicus* DSM 2380 does not. Likewise, of the nine sequenced *Geobacter*, only six contain the Type I secondary lipid synthase. Furthermore, omega-3 PUFA production in marine Labyrinthulid protists (Eukaryota) of the genus *Schizochytrium* is known to proceed via the same FAS/PKS-related pathway providing additional evidence implicating HGT as an important contributor to chemical diversification [6,28].

It has been noted that the *pfaA* homolog in *Elusimicrobium minutum* Pei191 (Type R) has an aberrant GC content (46%) compared to the rest of the genome (39%), prompting the suggestion that it entered the genome through horizontal gene transfer [44]. To further investigate the potential novelty of these gene clusters within their host genomes, DNA compositional analyses were performed (%GC analysis and AlienHunter [45]). No conclusive evidence based on DNA compositional metrics was revealed, indicating these gene clusters are not aberrant within their respective genomes. Our analyses of all secondary lipid synthase gene clusters indicate that the GC content of the *pfaA* homolog is consistently higher than the genome average (data not shown), most likely a reflection of the amino acid composition of these proteins. While these results do not rule out possible ancient acquisition of these pathways, comparative genomic analysis, as presented above, provides better resolution in support of horizontal acquisition of these gene clusters. Improved genomic sampling within individual phylogenetic groups will enhance our ability to resolve the evolutionary history and mobility of secondary lipid production potential.

Conclusions

Secondary lipid synthase gene clusters have been detected in 10 microbial phyla, representing 86 species across two domains of life. The genetic potential to produce long-chain fatty acids via a FAS/PKS mechanism appears to be scattered throughout the bacterial domain and has been co-opted by some Eukarya. Results presented here demonstrate that these biosynthetic pathways are not relegated solely to a narrow group of marine bacteria, as previously believed. Instead, this third mechanism of bacterial fatty acid synthesis may be involved in the production of specialized lipid products across the bacterial tree. As additional genomes are sequenced and their physiologies explored, additional bacterial lineages harboring novel secondary lipid pathways and producing novel fatty acyl products will be revealed. The linking of these pathways to chemical products and determining the physiological role of these biosynthetic processes is a crucial next step.

Methods

Identification of FAS/PKS candidate sequences in sequenced microbial genomes

Genomes were downloaded from NCBI and the Joint Genome Institute's Integrated Microbial Genomes (IMG) database (http://img.jgi.doe.gov). A local TBLASTN of *pfa* genes known to produce EPA and DHA was performed against the downloaded genomes and results were curated by e-value ($\leq$1e-30) to identify possible candidate sequences. To minimize false positive hits (e.g. polyketide synthase or nonribosomal peptide synthase genes), a refined set of query sequences was designed to exploit the conserved domain architecture of secondary lipid synthase gene products (e.g. PfaA–E). A subsequent TBLASTN of individual domains (KS, MAT, ACP, KR, DH, ER) was performed. These analyses provided a broad list of candidate genes. The most

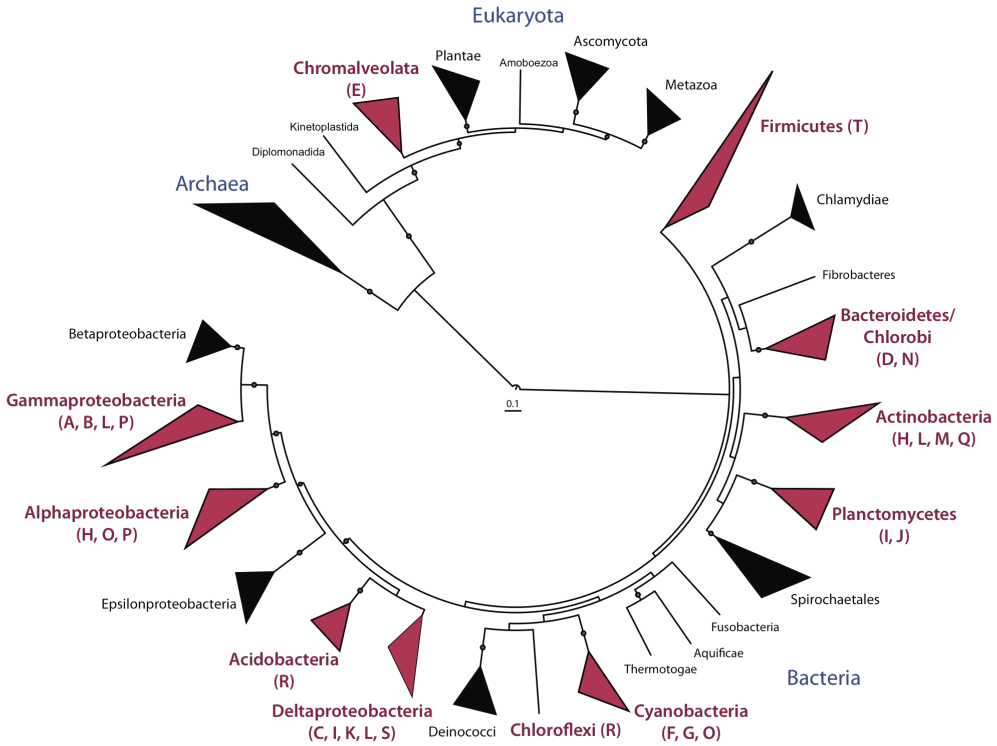

Figure 3. Phylogenetic distribution of *pfa* gene clusters and homologues. Clades in red contain putative secondary lipid gene clusters. Tree modified from the Interactive Tree of Life.
doi:10.1371/journal.pone.0020146.g003

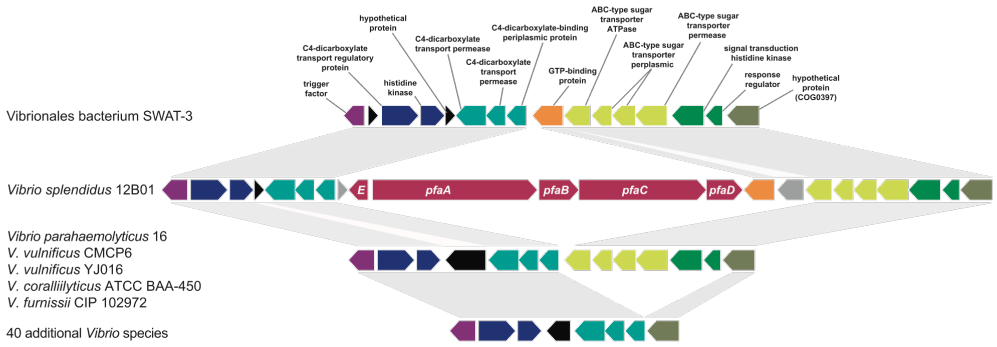

Figure 4. Plot of *Vibrio splendidus* 12B01 *pfa* region versus corresponding genomic regions of multiple *Vibrio* species. Gray tracts indicate regions of homology between genomes. The *pfa* genes are labeled in the *Vibrio splendidus* 12B01 genome and are absent in the other *Vibrio* genomes.
doi:10.1371/journal.pone.0020146.g004

informative query sequences were tandem ACP sequences, a hallmark characteristic of secondary lipid products. All candidate genomes and specific gene neighborhoods were inspect for adjacent genes using the sequence visualization tool Artemis [46] and via IMG's Gene Ortholog Browser tool. Protein sequences were annotated using InterProScan [47] and manually curated. **File S1** is presented as an Excel worksheet and provides a complete list of annotations and GenBank accession numbers for each protein sequence investigated in this study.

Phylogenetic analyses

Proximal KS amino acid sequences were aligned using MAFFT [48] and viewed and edited in Jalview [49]. Poorly aligned and/or divergent positions were excluded using Gblocks [50] with a minimum block of five and allowed gap positions equal to half. The best model for the resulting 364 amino acid alignment was evaluated using ProtTest [51], a program for selecting a model of protein evolution that uses PHYML [52] and PAL [53]. ProtTest chose LG+G+F based on Akaike criterion. A maximum likelihood phylogenetic tree was constructed using RAxML rapid bootstrapping [54] through the CIPRES portal [55] cluster at the San Diego Supercomputer Center. The phylogenetic tree graphic was produced using the Interactive Tree of Life [56]. The DH tree was constructed in an identical manner from the full alignment however using the LG+G model.

Multidimensional scaling analysis

Component domains from each gene cluster were letter coded (e.g. KS-MAT-ACP$_2$-KR-ER-PPTase becomes ABCCDEF). Coded gene clusters were aligned using MAFFT and viewed and edited in Jalview. The alignment was used to generate a pairwise distance matrix in MOTHUR [57] and subsequently converted to a similarity matrix. This similarity matrix was imported into Primer 6 [58] and used to create a multidimensional scaling (MDS) plot, using default parameters (25 random starts, Krustal fit scheme of 1, and a minimum stress value of 0.01). Subsequently, a cluster analysis of the data was performed, using average group linkage, and overlaid in order to define "Types" based on domain content and organization.

Permutational analysis of variance and canonical analysis of principal coordinates

A set of 33 life history traits were allocated to each bacterial species to create a binary matrix. A permutational analysis of variance (PERMANOVA) [59,60] was used to test the association between gene cluster type and bacterial life history traits. The PERMANOVA was based on unrestricted permutations of the raw data and a partial sums of squares. To visualize and test which life history traits were associated with which gene clusters we used a constrained canonical analysis of principal coordinates (CAP) [61,62]. CAP analyses were based on 10,000 random permutations of the raw data and a Bray-Curtis similarity matrix. Individual life history traits that might be responsible for any differences in multivariate space were investigated by calculating Spearman Rank correlations of canonical ordination axes with the original genera variables. Traits with strong correlations (defined as $\geq$0.4 in this study) were then overlaid as a bi-plot. We used the leave-one-out procedure in the CAP analysis to calculate allocation success for each *a priori* defined gene cluster. This essentially gave us a measure of distinctness for the life history traits associated with each gene cluster. Allocation success was considered indicative of a more distinct set of life history traits than expected by chance alone when values exceeded 14.3%. This threshold came from the possibility of each individual observation having a 14.3% chance of being placed into one of the 7 *a priori* defined groups (cluster Types). As replication within each *a priori* defined gene cluster group varied (ranging from n = 1 to n = 30), formal tests were only run on those gene clusters for which n$\geq$5 (A, B, F, I, L, M, R).

Supporting Information

Figure S1 Multidimensional scaling (MDS) plot of similarity matrix of domain order and count in *pfa* gene clusters and homologues. See File S1 for numerical key describing species abbreviations.
(EPS)

Figure S2 Maximum likelihood phylogenetic tree of dehydratase/isomerase protein domains (158 amino acid alignment). Gene cluster Types are colored and given a letter label. Asterisks represent Types first clustered and described in this study. Bootstrap values $\geq$50% are indicated by dots. The *E. coli* DH10B FabA protein was used as the outgroup.
(EPS)

Figure S3 Multiple sequence alignment showing sections of PPTase domains from representative organisms. Conserved domains are labeled at the top of the alignment. Motifs P2 and P3 do not show a clear pattern of variation among different secondary lipid types.
(EPS)

Figure S4 CAP ordination of the similarity among seven gene clusters (A, B, F, I, L, M, R) based on 33 life history traits. Group centroids are displayed for each gene cluster to ease interpretation. Vector lines in the bi-plot represent Spearman Rank correlations, with the direction indicating the relationship of each trait to the gene clusters in multivariate space. The length of each vector line is proportional to the strength of the correlation, with the blue circle showing the threshold for a correlation of one.
(EPS)

Figure S5 CAP ordination of the similarity among all gene clusters (A–T) based on 33 life history traits. Group centroids are displayed for each gene cluster to ease interpretation. Vector lines in the bi-plot represent Spearman Rank correlations, with the direction indicating the relationship of each trait to the gene clusters in multivariate space. The length of each vector line is proportional to the strength of the correlation, with the blue circle showing the threshold for a correlation of one.
(EPS)

File S1 Excel spreadsheet providing GenBank accession numbers and domain annotations for putative secondary lipid synthase genes. Numbers in column B provide a key for numerical abbreviations used in Figure S1.
(XLS)

Acknowledgments

The authors would like to sincerely thank Gareth Williams for statistical assistance with the PERMANOVA and CAP analyses. Additionally, we thank Sheila Podell for insightful discussion and providing comments on the manuscript.

Author Contributions

Conceived and designed the experiments: CNS EEA. Performed the experiments: CNS EEA. Analyzed the data: CNS EEA. Wrote the paper: CNS EEA.

References

1. Magnuson K, Jackowski S, Rock CO, Cronan JE (1993) Regulation of fatty acid biosynthesis in *Escherichia coli*. Microbiological Reviews 57: 522–542.
2. Schweizer E, Hofmann J (2004) Microbial type I fatty acid synthases (FAS): Major players in a network of cellular FAS systems. Microbiol Mol Biol Rev 68: 501–517.
3. Kolattukudy PE, Fernandes ND, Azad AK, Fitzmaurice AM, Sirakova TD (1997) Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. Mol Microbiol 24: 263–270.
4. Nichols DS (2003) Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. FEMS Microbiol Lett 219: 1–7.
5. Nichols DS, McMeekin TA (2002) Biomarker techniques to screen for bacteria that produce polyunsaturated fatty acids. J Microbiol Methods 48: 161–170.
6. Metz JG, Roessler P, Facciotti D, Levering C, Dittrich F, et al. (2001) Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. Science 293: 290–293.
7. Yazawa K (1996) Production of eicosapentaenoic acid from marine bacteria. Lipids 31: S297–S300.
8. Tanaka M, Ueno A, Kawasaki K, Yumoto I, Ohgiya S, et al. (1999) Isolation of clustered genes that are notably homologous to the eicosapentaenoic acid biosynthesis gene cluster from the docosahexaenoic acid-producing bacterium *Vibrio marinus* strain MP-1. Biotechnol Lett 21: 939–945.
9. Allen EE, Bartlett DH (2002) Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9. Microbiology 148: 1903–1913.
10. Campbell EL, Cohen MF, Meeks JC (1997) A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. Arch Microbiol 167: 251–258.
11. Bauersachs T, Compaore J, Hopmans EC, Stal LJ, Schouten S, et al. (2009) Distribution of heterocyst glycolipids in cyanobacteria. Phytochemistry 70: 2034–2039.
12. Miyanaga A, Funa N, Awakawa T, Horinouchi S (2008) Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. Proc Natl Acad Sci USA 105: 871–876.
13. Tang Y, Tsai SC, Khosla C (2003) Polyketide chain length control by chain length factor. J Am Chem Soc 125: 12708–12709.
14. Shulse CN, Allen EE (2011) Diversity and distribution of microbial long-chain fatty acid biosynthetic genes in the marine environment. Environ Microbiol 13: 684–695.
15. Jiang H, Zirkle R, Metz JG, Braun L, Richter L, et al. (2008) The role of tandem acyl carrier protein domains in polyunsaturated fatty acid biosynthesis. J Am Chem Soc 130: 6336–6337.
16. Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. Chemical Reviews 106: 3468–3496.
17. Okuyama H, Orikasa Y, Nishida T, Watanabe K, Morita N (2007) Bacterial genes responsible for the biosynthesis of eicosapentaenoic and docosahexaenoic acids and their heterologous expression. Appl Environ Microbiol 73: 665–670.
18. Wang F, Xiao X, Ou HY, Gai YB, Wang FP (2009) Role and regulation of fatty acid biosynthesis in the response of *Shewanella piezotolerans* WP3 to different temperatures and pressures. J Bacteriol 191: 2574–2584.
19. Venkateswaran K, Moser DP, Dollhopf ME, Lies DP, Saffarini DA, et al. (1999) Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. Int J Syst Bacteriol 49: 705–724.
20. Moser DP, Nealson KH (1996) Growth of the facultative anaerobe *Shewanella putrefaciens* by elemental sulfur reduction. Appl Environ Microbiol 62: 2100–2105.
21. Perry KA, Kostka JE, Luther GW, Nealson KH (1993) Mediation of sulfur speciation by a Black Sea facultative anaerobe. Science 259: 801–803.
22. Orikasa Y, Tanaka M, Sugihara S, Hori R, Nishida T, et al. (2009) *pfaB* products determine the molecular species produced in bacterial polyunsaturated fatty acid biosynthesis. FEMS Microbiol Lett 295: 170–176.
23. Methe BA, Nelson KE, Deming JW, Momen B, Melamud E, et al. (2005) The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. Proc Natl Acad Sci USA 102: 10913–10918.
24. Riley M, Staley JT, Danchin A, Wang TZ, Brettin TS, et al. (2008) Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. BMC Genomics 9: 210.
25. Delong EF, Yayanos AA (1986) Biochemical function and ecological significance of novel bacterial lipids in deep-sea procaryotes. Appl Environ Microbiol 51: 730–737.
26. Auman AJ, Breezee JL, Gosink JJ, Kampfer P, Staley JT (2006) *Psychromonas ingrahamii* sp nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. Int J Syst Evol Microbiol 56: 1001–1007.
27. Bowman JP, McCammon SA, Lewis T, Skerratt JH, Brown JL, et al. (1998) *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson et al. 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. Microbiology 144: 1601–1609.
28. Lippmeier JC, Crawford KS, Owen CB, Rivas AA, Metz JG, et al. (2009) Characterization of both polyunsaturated fatty acid biosynthetic pathways in *Schizochytrium* sp. Lipids 44: 621–630.
29. Singh A, Ward OP (1996) Production of high yields of docosahexaenoic acid by *Thraustochytrium roseum* ATCC 28210. J Ind Microbiol 16: 370–373.
30. Yokochi T, Honda D, Higashihara T, Nakahara T (1998) Optimization of docosahexaenoic acid production by *Schizochytrium limacinum* SR21. Appl Microbiol Biotechnol 49: 72–76.
31. Raghukumar S (2002) Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). Eur J Protistol 38: 127–145.
32. Starkey RL, De PK (1939) A new species of *Azotobacter*. Soil Science 47: 329–343.
33. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417: 141–147.
34. Normand P, Lapierre P, Tisa LS, Gogarten JP, Alloisio N, et al. (2007) Genome characteristics of facultatively symbiotic *Frankia* sp strains reflect host range and host plant biogeography. Genome Res 17: 7–15.
35. Sangkhobol V, Skerman VBD (1981) *Chitinophaga*, a new genus of chitinolytic myxobacteria. Int J Syst Bacteriol 31: 285–293.
36. Funa N, Ozawa H, Hirata A, Horinouchi S (2006) Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in *Azotobacter vinelandii*. Proc Natl Acad Sci USA 103: 6356–6361.
37. Sugihara S, Orikasa Y, Okuyama H (2008) An EntD-like phosphopantetheinyl transferase gene from *Photobacterium profundum* SS9 complements *pfa* genes of *Moritella marina* strain MP-1 involved in biosynthesis of docosahexaenoic acid. Biotechnol Lett 30: 411–414.
38. Hantash FM, Ammerlaan M, Earhart CF (1997) Enterobactin synthase polypeptides of *Escherichia coli* are present in an osmotic-shock-sensitive cytoplasmic locality. Microbiology-Uk 143: 147–156.
39. White SW, Zheng J, Zhang YM, Rock CO (2005) The structural biology of type II fatty acid biosynthesis. Annu Rev Biochem 74: 791–831.
40. Wiens GD, Rockey DD, Wu ZN, Chang J, Levy R, et al. (2008) Genome sequence of the fish pathogen *Renibacterium salmoninarum* suggests reductive evolution away from an environmental *Arthrobacter* ancestor. J Bacteriol 190: 6970–6982.
41. Sukovich DJ, Seffernick JL, Richman JE, Hunt KA, Gralnick JA, et al. (2010) Structure, function, and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1. Appl Environ Microbiol 76: 3842–3849.
42. Sukovich DJ, Seffernick JL, Richman JE, Gralnick JA, Wackett LP (2010) Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. Appl Environ Microbiol 76: 3850–3862.
43. Nishida T, Orikasa Y, Ito Y, Yu R, Yamada A, et al. (2006) *Escherichia coli* engineered to produce eicosapentaenoic acid becomes resistant against oxidative damages. FEBS Lett 580: 2731–2735.
44. Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, et al. (2009) Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly Termite Group 1). Appl Environ Microbiol 75: 2841–2849.
45. Vernikos GS, Parkhill J (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. Bioinformatics 22: 2196–2203.
46. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al. (2000) Artemis: sequence visualization and annotation. Bioinformatics 16: 944–945.
47. Zdobnov EM, Apweiler R (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847–848.
48. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics 9: 286–298.
49. Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. Bioinformatics 20: 426–427.
50. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552.
51. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105.
52. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
53. Drummond A, Strimmer K (2001) PAL: an object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics 17: 662–663.
54. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57: 758–771.
55. Miller MA, Holder MT, Vos R, Midford PE, Liebowitz T, et al. (2009) The CIPRES Portals. CIPRES.
56. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127–128.
57. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75: 7537–7541.
58. Clarke K, Warwick R (2001) Change in marine communities: an approach to statistical analysis and interpretation. Plymouth: PRIMER-E.
59. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecol 26: 32–46.
60. McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology 82: 290–297.
61. Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. Ecology 84: 511–525.
62. Anderson MJ, Robinson J (2003) Generalized discriminant analysis based on distances. Australian & New Zealand Journal of Statistics 45: 301–318.

Supplementary Material

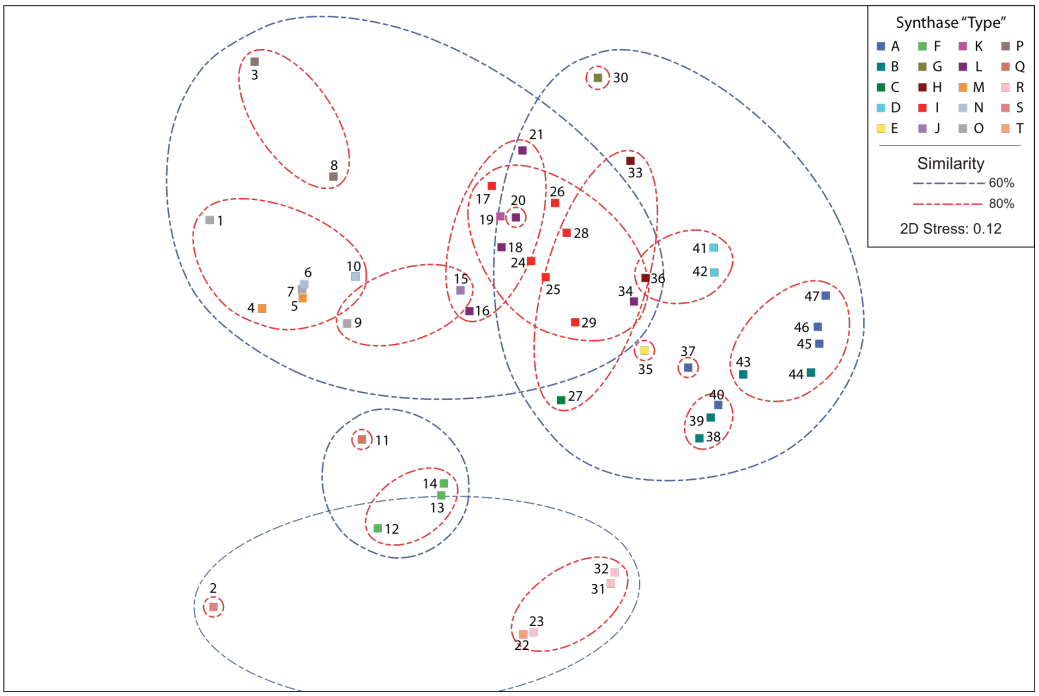

Figure S1. Multidimensional scaling (MDS) plot of similarity matrix of domain order and count in *pfa* gene clusters and homologues. See Table S1 for numerical key describing species abbreviations.

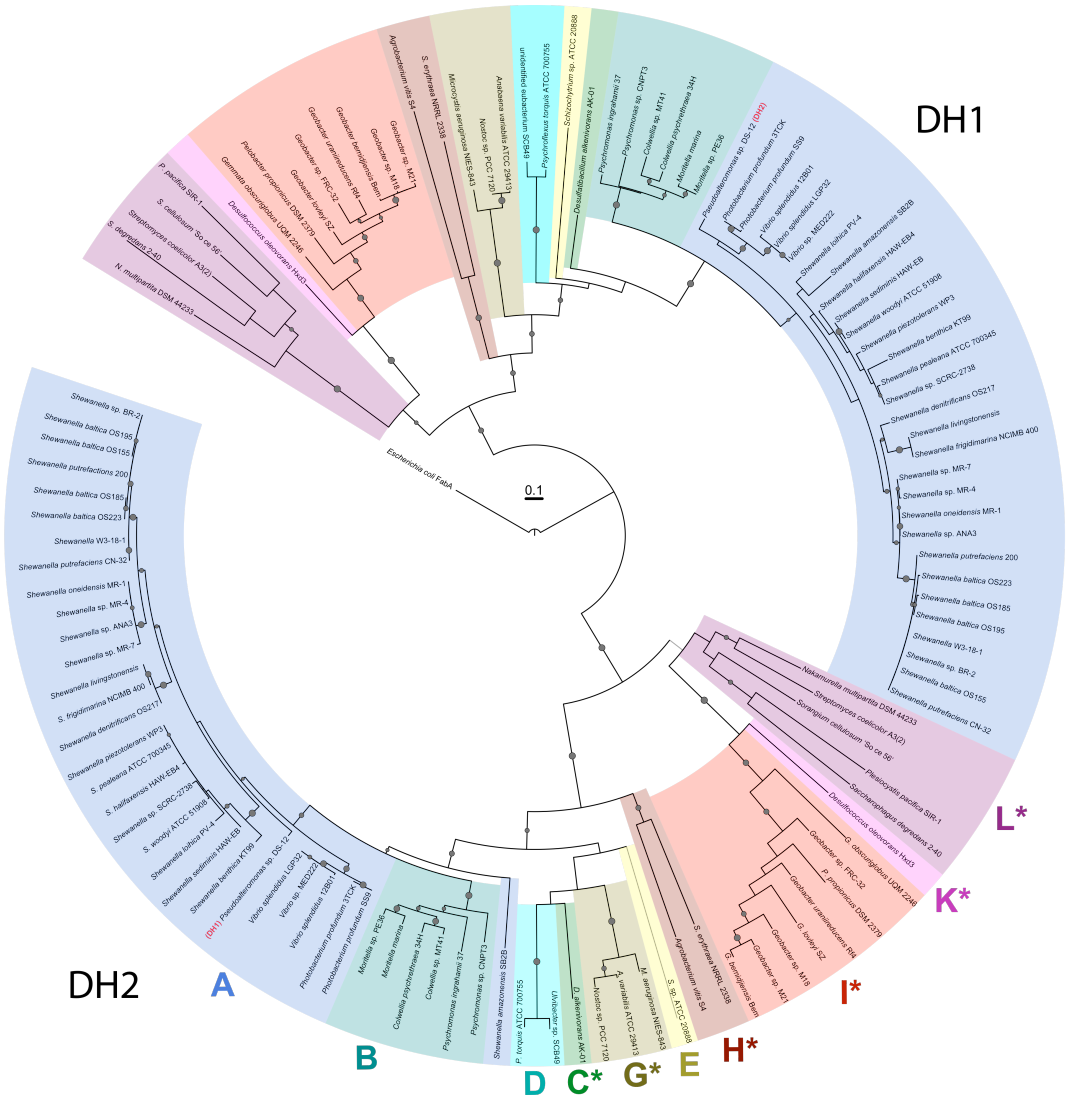

Figure S2. Maximum likelihood phylogenetic tree of dehydratase/isomerase protein domains (158 amino acid alignment). Gene cluster Types are colored and given a letter label. Asterisks represent Types first clustered and described in this study. Bootstrap values ≥50% are indicated by dots. The *E. coli* DH10B FabA protein was used as the outgroup.

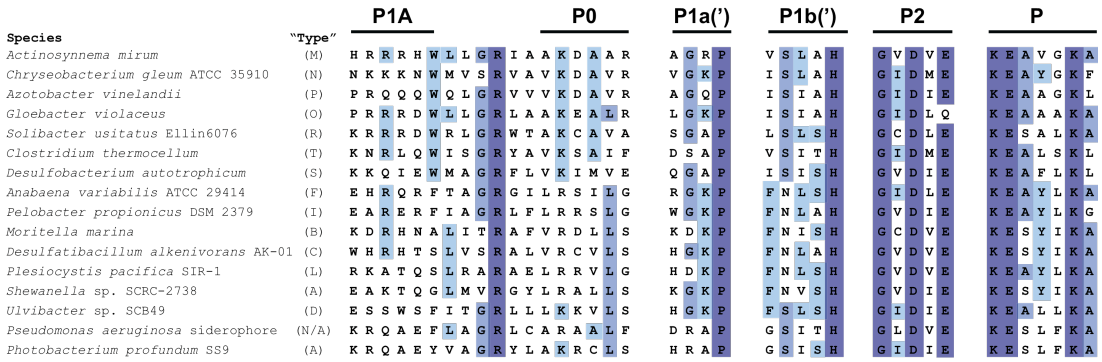

Figure S3. Multiple sequence alignment showing sections of PPTase domains from representative organisms. Conserved domains are labeled at the top of the alignment. Motifs P2 and P3 do not show a clear pattern of variation among different secondary lipid types.

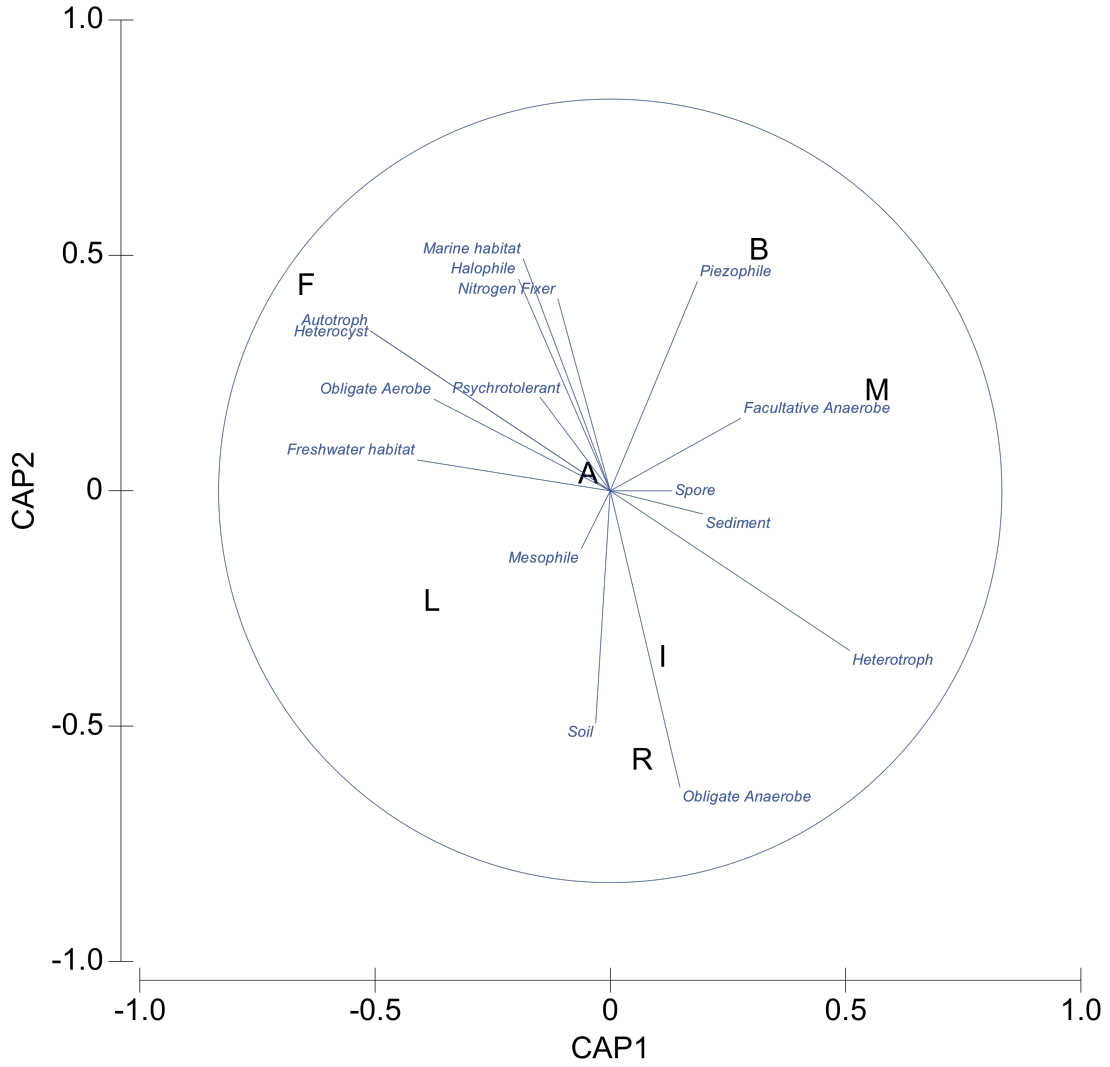

Figure S4. CAP ordination of the similarity among seven gene clusters (A, B, F, I, L, M, R) based on 33 life history traits. Group centeroids are displayed for each gene cluster to ease interpretation. Vector lines in the bi-plot represent Spearman Rank correlations, with the direction indicating the relationship of each trait to the gene clusters in multivariate space. The length of each vector line is proportional to the strength of the correlation, with the blue circle showing the threshold for a correlation of one.

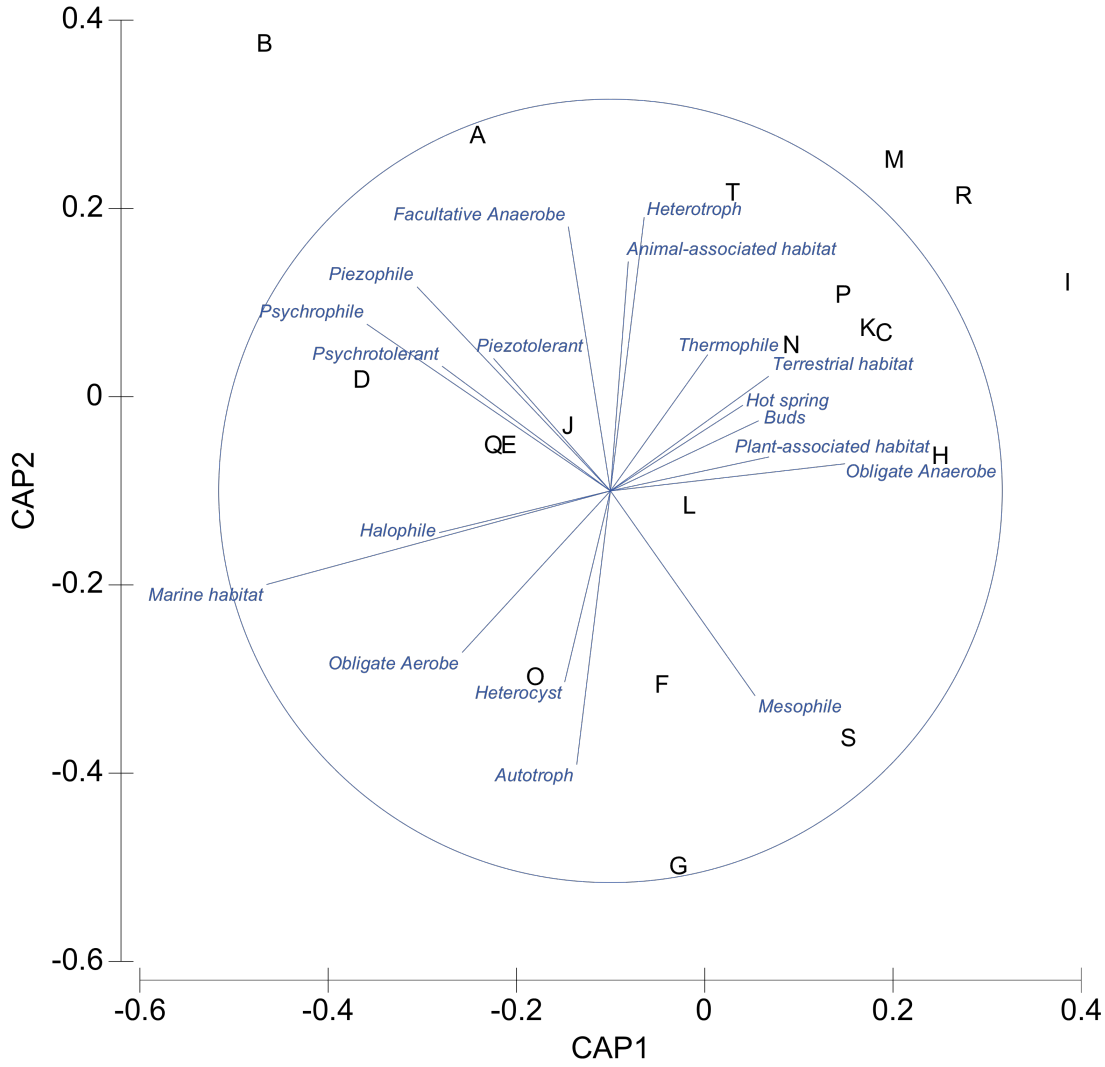

Figure S5. CAP ordination of the similarity among all gene clusters (A-T) based on 33 life history traits. Group centroids are displayed for each gene cluster to ease interpretation. Vector lines in the bi-plot represent Spearman Rank correlations, with the direction indicating the relationship of each trait to the gene clusters in multivariate space. The length of each vector line is proportional to the strength of the correlation, with the blue circle showing the threshold for a correlation of one.

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases.

Species Name	Figure S1 Key	GenBank Accession No.	Domain Order
Parvularcula bermudensis HTCC2503	**1**	CP002156 REGION: 1464266..1472761	ER-KS-MAT-ACP-KR
Parvularcula bermudensis HTCC2503	**1**	CP002156 REGION: 1472770..1476528	KS-AT-PPTase
Desulfobacterium autotrophicum HRM2	**2**	YP_002604438	ER-KS-CLF-AT-KS-ACP_3-KR
Desulfobacterium autotrophicum HRM2	**2**	YP_002604439	PPTase
Azotobacter vinelandii DJ	**3**	YP_002800097	ER-KS-MAT-ACP-KR
Azotobacter vinelandii DJ	**3**	YP_002800094	PPTase
Actinosynnema mirum DSM 43827	**4**	YP_003100276	ACS-ACP
Actinosynnema mirum DSM 43827	**4**	YP_003100277	ER-KS-MAT-ACP_2-KR
Actinosynnema mirum DSM 43827	**4**	YP_003100278	KS-AT-PPTase
Actinosynnema mirum DSM 43827	**4**	YP_003100279	ACP
Frankia alni ACN14a	**5**	YP_711795	ER-KS-MAT-ACP_2-KR
Frankia alni ACN14a	**5**	YP_711796	KS-AT-PPTase
Frankia alni ACN14a	**5**	YP_711797	ACP
Frankia sp. CcI3	**5**	YP_480039	ER-KS-MAT-ACP_2-KR
Frankia sp. CcI3	**5**	YP_480040	KS-AT-PPTase
Frankia sp. CcI3	**5**	YP_480041	ACP
Frankia sp. EAN1pec	**5**	YP_001509872	ER-KS-MAT-ACP_2-KR
Frankia sp. EAN1pec	**5**	YP_001509871	KS-AT-PPTase
Frankia sp. EAN1pec	**5**	YP_001509870	ACP
Renibacterium salmoninarum ATCC 33209	**5**	YP_001625679	ER-KS-MAT-ACP_2-KR
Renibacterium salmoninarum ATCC 33209	**5**	YP_001625680	KS-AT-PPTase
Renibacterium salmoninarum ATCC 33209	**5**	YP_001625681	ACP
Chryseobacterium gleum ATCC 35910	**6**	EFK33535	ER-KS-MAT-ACP_2-KR
Chryseobacterium gleum ATCC 35910	**6**	EFK33534	KS-AT-PPTase
Chryseobacterium gleum ATCC 35910	**6**	EFK33533	ACP
Dinoroseobacter shibae DFL 12	**7**	YP_001533304	ER-KS-MAT-ACP_2-KR
Dinoroseobacter shibae DFL 12	**7**	YP_001533303	KS-AT-PPTase
Dinoroseobacter shibae DFL 14	**7**	YP_001533302	ACP

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Streptomyces avermitilis MA-4680	**7**	NP_828538	ER-KS-MAT-ACP_2-KR
Streptomyces avermitilis MA-4680	**7**	NP_828537	KS-AT-PPTase
Streptomyces avermitilis MA-4680	**7**	NP_828536	ACP
Beijerinckia indica subsp. Indica ATCC 9039	**8**	YP_001830916	ER-KS-MAT-ACP_2-KR
Beijerinckia indica subsp. Indica ATCC 9039	**8**	YP_001830913	PPTase
Gloeobacter violaceus PCC 7421	**9**	NP_927172	ER-KS-MAT-ACP_3-KR
Gloeobacter violaceus PCC 7421	**9**	NP_927171	KS-AT-PPTase
Gloeobacter violaceus PCC 7421	**9**	NP_927170	ACP
Chitinophaga pinensis DSM 2588	**10**	YP_003121148	ER-KS-MAT-ACP_2-KR
Chitinophaga pinensis DSM 2588	**10**	YP_003121147	KS-PPTase
Chitinophaga pinensis DSM 2588	**10**	YP_003121146	ACP
Rhodococcus erythropolis PR4	**11**	YP_002763824	KS-MAT-ACP_2-KR
Rhodococcus erythropolis PR4	**11**	YP_002763825	TE
Anabaena variabilis ATCC 29413 (F)	**12**	YP_323100	KS-MAT-ACP_2-KR
Anabaena variabilis ATCC 29413 (F)	**12**	YP_323101	KR
Anabaena variabilis ATCC 29413 (F)	**12**	YP_323102	KS
Anabaena variabilis ATCC 29413 (F)	**12**	YP_323103	CLF-AT
Anabaena variabilis ATCC 29413 (F)	**12**	YP_323104	ER
Anabaena variabilis ATCC 29413 (F)	**12**	YP_323105	ACP-TE
Anabaena variabilis ATCC 29413 (F)	**12**	YP_323107	PPTase
Nodularia spumigena CCY 9414	**12**	ZP_01630197	KS-MAT-ACP_2-KR
Nodularia spumigena CCY 9414	**12**	ZP_01630198	KR
Nodularia spumigena CCY 9414	**12**	ZP_01630200	KS-CLF-AT
Nodularia spumigena CCY 9414	**12**	ZP_01630201	ER
Nodularia spumigena CCY 9414	**12**	ZP_01630202	ACP-TE
Nodularia spumigena CCY 9414	**12**	ZP_01630204	PPTase

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Nostoc punctiforme PCC 73102	**12**	YP_001863788	KS-MAT-ACP_2-KR
Nostoc punctiforme PCC 73102	**12**	YP_001863787	KR
Nostoc punctiforme PCC 73102	**12**	YP_001863786	KS-CLF-AT
Nostoc punctiforme PCC 73102	**12**	YP_001863785	ER
Nostoc punctiforme PCC 73102	**12**	YP_001863784	ACP-TE
Nostoc punctiforme PCC 73102	**12**	YP_001863782	PPTase
Nostoc (Anabaena) azollae' 0708	**13**	ADI65420	KS-MAT-ACP_2
Nostoc (Anabaena) azollae' 0708	**13**	ADI65419	KR
Nostoc (Anabaena) azollae' 0708	**13**	ADI65418	KS
Nostoc (Anabaena) azollae' 0708	**13**	ADI65417	CLF-AT
Nostoc (Anabaena) azollae' 0708	**13**	ADI65416	ER
Nostoc (Anabaena) azollae' 0708	**13**	ADI65415	ACP-TE
Nostoc sp. PCC 7120 (F)	**14**	NP_489391	KS-MAT-ACP_2
Nostoc sp. PCC 7120 (F)	**14**	NP_489393	KR
Nostoc sp. PCC 7120 (F)	**14**	NP_489394	KS
Nostoc sp. PCC 7120 (F)	**14**	NP_489395	CLF-AT
Nostoc sp. PCC 7120 (F)	**14**	NP_489396	ER
Nostoc sp. PCC 7120 (F)	**14**	NP_489397	ACP-TE
Nostoc sp. PCC 7120 (F)	**14**	NP_489399	PPTase
Planctomyces limnophilus DSM 3776	**15**	YP_003628512	ER
Planctomyces limnophilus DSM 3776	**15**	YP_003628511	KS-MAT-ACP_3-KR
Plesiocystis pacifica SIR-1	**16**	ZP_01907911	ER
Plesiocystis pacifica SIR-1	**16**	ZP_01907910	KS-MAT
Plesiocystis pacifica SIR-1	**16**	ZP_01907909	ACP_2-KR
Plesiocystis pacifica SIR-1	**16**	ZP_01907908	KS-CLF-AT-DH_2
Plesiocystis pacifica SIR-1	**16**	ZP_01911082	PPTase
Gemmata obscuriglobus UQM 2246	**17**	ZP_02733785	ER
Gemmata obscuriglobus UQM 2246	**17**	ZP_02733786	KS-MAT-ACP_2-KR
Gemmata obscuriglobus UQM 2246	**17**	ZP_02733787	KS-CLF-AT-DH_2
Saccharophagus degradans 2-40	**18**	YP_526161	ER
Saccharophagus degradans 2-40	**18**	YP_526162	KS-MAT-ACP_3-KR
Saccharophagus degradans 2-40	**18**	YP_526163	KS-CLF-AT-DH_2

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Desulfococcus oleovorans Hxd3	**19**	YP_001529556	ER
Desulfococcus oleovorans Hxd3	**19**	YP_001529557	KS-MAT-ACP_3-KR
Desulfococcus oleovorans Hxd3	**19**	YP_001529558	KS-CLF-AT-DH_2
Nakamurella multipartita DSM 44233	**20**	YP_003203517	ER
Nakamurella multipartita DSM 44233	**20**	YP_003203516	KS-MAT-ACP_2-KR
Nakamurella multipartita DSM 44233	**20**	YP_003203515	KS-CLF-AT-DH_2-AT
Streptomyces coelicolor A3(2)	**21**	NP_624464	ER
Streptomyces coelicolor A3(2)	**21**	NP_624465	KS-MAT-ACP-KR
Streptomyces coelicolor A3(2)	**21**	NP_624466	KS-CLF-AT-DH_2
Streptomyces coelicolor A3(2)	**21**	ZP_04690736	ACP-KR
Clostridium thermocellum ATCC 27405	**22**	YP_001036569	KS-MAT-CLF-ACP_3-KR
Clostridium thermocellum ATCC 27405	**22**	YP_001036570	PPTase
Elusimicrobium minutum Pei191	**23**	YP_001875990	KS-MAT-CLF-ACP_4-KR
Elusimicrobium minutum Pei191	**23**	YP_001875989	PPTase
Geobacter sp. FRC-32	**24**	YP_002538590	ER
Geobacter sp. FRC-32	**24**	YP_002538591	KS-MAT-ACP_4-KR
Geobacter sp. FRC-32	**24**	YP_002538592	KS-CLF-AT-DH_2
Geobacter sp. FRC-32	**24**	YP_002538597	PPTase
Geobacter uraniireducens Rf4	**24**	YP_001230776	ER
Geobacter uraniireducens Rf4	**24**	YP_001230775	KS-MAT-ACP_4-KR
Geobacter uraniireducens Rf4	**24**	YP_001230774	KS-CLF-AT-DH_2
Geobacter uraniireducens Rf4	**24**	YP_001230769	PPTase
Pelobacter propionicus DSM 2379	**24**	YP_902751	ER
Pelobacter propionicus DSM 2379	**24**	YP_902752	KS-MAT-ACP_4-KR
Pelobacter propionicus DSM 2379	**24**	YP_902753	KS-CLF-AT-DH_2
Pelobacter propionicus DSM 2379	**24**	YP_902770	PPTase
Geobacter bemidjiensis Bem	**25**	YP_002138915	ER
Geobacter bemidjiensis Bem	**25**	YP_002138914	KS-MAT-ACP_5-KR
Geobacter bemidjiensis Bem	**25**	YP_002138913	KS-CLF-AT-DH_2
Geobacter bemidjiensis Bem	**25**	YP_002138907	PPTase

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Geobacter sp. M21	**25**	YP_003021922	ER
Geobacter sp. M21	**25**	YP_003021923	KS-MAT-ACP_5-KR
Geobacter sp. M21	**25**	YP_003021924	KS-CLF-AT-DH_2
Geobacter sp. M21	**25**	YP_003021930	PPTase
Geobacter lovleyi SZ	**26**	YP_001952502	ER
Geobacter lovleyi SZ	**26**	YP_001952503	KS-MAT-ACP_4-KR
Geobacter lovleyi SZ	**26**	YP_001952504	KS-CLF-AT-DH_2
Desulfatibacillum alkenivorans AK-01	**27**	YP_002430227	KS-MAT-ACP_3-KR
Desulfatibacillum alkenivorans AK-01	**27**	YP_002430226	KS-CLF-AT-DH_2
Desulfatibacillum alkenivorans AK-01	**27**	YP_002430225	ER
Desulfatibacillum alkenivorans AK-01	**27**	YP_002430224	PPTase
Geobacter sp. M18	**28**	ZP_05311357	ER
Geobacter sp. M18	**28**	ZP_05311356	KS-MAT-ACP_5-KR
Geobacter sp. M18	**28**	ZP_05311355	KS-CLF-AT-DH_2
Desulfuromonas acetoxidans DSM 684	**29**	ZP_01312928	ER
Desulfuromonas acetoxidans DSM 684	**29**	ZP_01312929	KS-MAT-ACP_5-KR
Anabaena variabilis ATCC 29413 (G)	**30**	YP_325239	KS-MAT-ACP
Anabaena variabilis ATCC 29413 (G)	**30**	YP_325241	KR
Anabaena variabilis ATCC 29413 (G)	**30**	YP_325242	KS-AT-DH_2
Anabaena variabilis ATCC 29413 (G)	**30**	YP_325243	ER
Microcystis aeruginosa NIES-843	**30**	YP_001657798	KS-ACP
Microcystis aeruginosa NIES-843	**30**	YP_001657800	KR
Microcystis aeruginosa NIES-843	**30**	YP_001657801	KS-AT-DH_2
Microcystis aeruginosa NIES-843	**30**	YP_001657802	ER
Nostoc sp. PCC 7120 (G)	**30**	NP_485686	KS-MAT-ACP
Nostoc sp. PCC 7120 (G)	**30**	NP_485684	KR
Nostoc sp. PCC 7120 (G)	**30**	NP_485683	KS-AT-DH_2
Nostoc sp. PCC 7120 (G)	**30**	NP_485682	ER
Roseiflexus sp. RS-1	**31**	YP_001277223	KS-MAT-CLF-ACP_6-KR
Roseiflexus sp. RS-1	**31**	YP_001277224	PPTase
Roseiflexus castenholzii DSM 13941	**32**	YP_001432959	KS-MAT-CLF-ACP_5-KR
Roseiflexus castenholzii DSM 13941	**32**	YP_001432960	PPTase
Candidatus *(Acidobacteria) Korebacter versatilis* Ellin345	**33**	YP_589544	KS-MAT-CLF-ACP_3-KR
Candidatus *(Acidobacteria) Korebacter versatilis* Ellin345	**33**	YP_589546	PPTase

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Candidatus *Solibacter usitatus* Ellin6076	**33**	YP_824341	KS-MAT-CLF-ACP_3-KR
Candidatus *Solibacter usitatus* Ellin6076	**33**	YP_824340	PPTase
Saccharopolyspora erythraea NRRL 2338	**33**	ZP_06563680	KS-MAT-ACP
Saccharopolyspora erythraea NRRL 2338	**33**	ZP_06563679	KR
Saccharopolyspora erythraea NRRL 2338	**33**	ZP_06563678	KS-CLF-AT-DH_2
Saccharopolyspora erythraea NRRL 2338	**33**	ZP_06563677	ER
Sorangium cellulosum 'So ce 56'	**34**	YP_001611455	ER
Sorangium cellulosum 'So ce 56'	**34**	YP_001611456	KS-MAT-ACP_5
Sorangium cellulosum 'So ce 56'	**34**	YP_001611457	KS-CLF-DH_2-AT
Schizochytrium sp. ATCC 20888	**35**	AF378327	KS-MAT-ACP_9-KR
Schizochytrium sp. ATCC 20888	**35**	AF378328	KS-CLF-AT-ER
Schizochytrium sp. ATCC 20888	**35**	AF378329	DH_2-ER
Agrobacterium vitis S4	**36**	YP_002547429	ACP
Agrobacterium vitis S4	**36**	YP_002547431	KS-MAT-ACP_4-KR
Agrobacterium vitis S4	**36**	YP_002547432	KS-CLF-AT-DH_2
Agrobacterium vitis S4	**36**	YP_002547433	ER
Pseudoalteromonas sp. DS-12	**37**	DQ469875	KS-MAT-ACP_5-KR-KS-CLF-PPTase-DH_2
Moritella sp. PE36	**38**	ZP_01898228	KS-MAT-ACP_7-KR
Moritella sp. PE36	**38**	ZP_01898227	KS-AT
Moritella sp. PE36	**38**	ZP_01900673	KS-CLF-DH_2
Moritella sp. PE36	**38**	ZP_01900672	ER
Moritella sp. PE36	**38**	ZP_01899831	PPTase
Moritella marina	**39**	BAF02836	PPTase
Moritella marina	**39**	BAA89382	KS-MAT-ACP_5-KR
Moritella marina	**39**	BAA89383	KS-AT
Moritella marina	**39**	BAA89384	KS-CLF-DH_2
Photobacterium profundum 3TCK	**40**	ZP_01221855	KS-MAT-ACP_5-KR
Photobacterium profundum 3TCK	**40**	ZP_01221854	AT
Photobacterium profundum 3TCK	**40**	ZP_01221853	KS-CLF-DH_2
Photobacterium profundum 3TCK	**40**	ZP_01221852	ER
Photobacterium profundum 3TCK	**40**	ZP_01219528	PPTase

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Photobacterium profundum SS9	**40**	YP_129673	KS-MAT-ACP_5-KR
Photobacterium profundum SS9	**40**	YP_129672	AT
Photobacterium profundum SS9	**40**	YP_129671	KS-CLF-AT-DH_2
Photobacterium profundum SS9	**40**	YP_129670	ER
Photobacterium profundum SS9	**40**	YP_133485	PPTase
Ulvibacter sp. SCB49	**41**	ZP_01889252	PPTase
Ulvibacter sp. SCB49	**41**	ZP_01889250	KS-MAT-ACP_4
Ulvibacter sp. SCB49	**41**	ZP_01889249	KR
Ulvibacter sp. SCB49	**41**	ZP_01889248	KS-CLF-AT-DH_2
Ulvibacter sp. SCB49	**41**	ZP_01889247	ER
Psychroflexus torquis ATCC 700755	**42**	ZP_01255376	PPTase
Psychroflexus torquis ATCC 700755	**42**	ZP_01255378	KS-MAT-ACP_6
Psychroflexus torquis ATCC 700755	**42**	ZP_01255379	KR
Psychroflexus torquis ATCC 700755	**42**	ZP_01255380	KS-CLF-AT-DH_2
Psychroflexus torquis ATCC 700755	**42**	ZP_01255381	ER
Psychromonas ingrahamii 37	**43**	YP_943075	KS-MAT-ACP_6-KR
Psychromonas ingrahamii 37	**43**	YP_943076	KS-AT
Psychromonas ingrahamii 37	**43**	YP_943077	KS-CLF-DH_2
Psychromonas ingrahamii 37	**43**	YP_943078	ER
Psychromonas ingrahamii 37	**43**	YP_944031	PPTase
Psychromonas sp. CNPT3	**43**	ZP_01216282	KS-MAT-ACP_6-KR
Psychromonas sp. CNPT3	**43**	ZP_01216281	KS-AT
Psychromonas sp. CNPT3	**43**	ZP_01216280	KS-CLF-DH_2
Psychromonas sp. CNPT3	**43**	ZP_01216279	ER
Colwellia psychrerythraea 34H	**44**	YP_269804	PPTase
Colwellia psychrerythraea 34H	**44**	YP_269802	KS-MAT-ACP_6-KR
Colwellia psychrerythraea 34H	**44**	YP_269801	KS-AT
Colwellia psychrerythraea 34H	**44**	YP_269800	KS-CLF-DH_2
Colwellia psychrerythraea 34H	**44**	YP_269797	ER
Colwellia sp. MT41	**44**	D. Bartlett-unpublished	PPTase
Colwellia sp. MT41	**44**	D. Bartlett-unpublished	KS-MAT-ACP_6-KR
Colwellia sp. MT41	**44**	D. Bartlett-unpublished	KS-AT
Colwellia sp. MT41	**44**	D. Bartlett-unpublished	KS-CLF-AT-DH_2
Colwellia sp. MT41	**44**	D. Bartlett-unpublished	ER

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Shewanella halifaxensis HAW-EB4	**45**	YP_001675213	PPTase
Shewanella halifaxensis HAW-EB4	**45**	YP_001675216	KS-MAT-ACP_6-KR
Shewanella halifaxensis HAW-EB4	**45**	YP_001675217	AT
Shewanella halifaxensis HAW-EB4	**45**	YP_001675218	KS-CLF-DH_2
Shewanella halifaxensis HAW-EB4	**45**	YP_001675219	ER
Shewanella piezotolerans WP3	**45**	YP_002312825	PPTase
Shewanella piezotolerans WP3	**45**	YP_002312829	KS-MAT-ACP_6-KR
Shewanella piezotolerans WP3	**45**	YP_002312830	AT
Shewanella piezotolerans WP3	**45**	ACJ30245	KS-CLF-DH_2
Shewanella piezotolerans WP3	**45**	ACJ30246	ER
Shewanella sediminis HAW-EB3	**45**	YP_001474918	PPTase
Shewanella sediminis HAW-EB3	**45**	YP_001474934	KS-MAT-ACP_6-KR
Shewanella sediminis HAW-EB3	**45**	YP_001474935	AT
Shewanella sediminis HAW-EB3	**45**	YP_001474936	KS-CLF-DH_2
Shewanella sediminis HAW-EB3	**45**	YP_001474937	ER
Shewanella sp. SCRC-2738	**45**	U73935	PPTase-KS-MAT-ACP_6-KR-AT-KS-CLF-DH_2-ER
Shewanella woodyi ATCC 51908	**45**	YP_001761668	PPTase
Shewanella woodyi ATCC 51908	**45**	YP_001759810	KS-MAT-ACP_6-KR
Shewanella woodyi ATCC 51908	**45**	YP_001759809	AT
Shewanella woodyi ATCC 51908	**45**	YP_001759808	KS-CLF-DH_2
Shewanella woodyi ATCC 51908	**45**	YP_001759807	ER
Shewanella amazonensis SB2B	**46**	YP_926996	PPTase
Shewanella amazonensis SB2B	**46**	YP_926992	KS-MAT-ACP_5-KR
Shewanella amazonensis SB2B	**46**	YP_926991	AT
Shewanella amazonensis SB2B	**46**	YP_926990	KS-CLF-DH_2
Shewanella amazonensis SB2B	**46**	YP_926989	ER

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Shewanella baltica OS155	**46**	YP_001049814	PPTase
Shewanella baltica OS155	**46**	YP_001049812	KS-MAT-ACP_5-KR
Shewanella baltica OS155	**46**	YP_001049811	AT
Shewanella baltica OS155	**46**	YP_001049810	KS-CLF-DH_2
Shewanella baltica OS155	**46**	YP_001049809	ER
Shewanella baltica OS185	**46**	YP_001365634	PPTase
Shewanella baltica OS185	**46**	YP_001365632	KS-MAT-ACP_5-KR
Shewanella baltica OS185	**46**	YP_001365631	AT
Shewanella baltica OS185	**46**	YP_001365630	KS-CLF-DH_2
Shewanella baltica OS185	**46**	YP_001365629	ER
Shewanella baltica OS195	**46**	YP_001553889	PPTase
Shewanella baltica OS195	**46**	YP_001553887	KS-MAT-ACP_5-KR
Shewanella baltica OS195	**46**	YP_001553886	AT
Shewanella baltica OS195	**46**	YP_001553885	KS-CLF-DH_2
Shewanella baltica OS195	**46**	YP_001553884	ER
Shewanella baltica OS223	**46**	YP_002358836	PPTase
Shewanella baltica OS223	**46**	YP_002358838	KS-MAT-ACP_5-KR
Shewanella baltica OS223	**46**	YP_002358839	AT
Shewanella baltica OS223	**46**	YP_002358840	KS-CLF-DH_2
Shewanella baltica OS223	**46**	YP_002358841	ER
Shewanella denitrificans OS217	**46**	YP_563627	PPTase
Shewanella denitrificans OS217	**46**	YP_563629	KS-MAT-ACP_5-KR
Shewanella denitrificans OS217	**46**	YP_563630	AT
Shewanella denitrificans OS217	**46**	YP_563631	KS-CLF-DH_2
Shewanella denitrificans OS217	**46**	YP_563632	ER
Shewanella frigidimarina NCIMB	**46**	YP_749936	PPTase
Shewanella frigidimarina NCIMB	**46**	YP_749912	KS-MAT-ACP_5-KR
Shewanella frigidimarina NCIMB	**46**	YP_749911	AT
Shewanella frigidimarina NCIMB	**46**	YP_749910	KS-CLF-DH_2
Shewanella frigidimarina NCIMB	**46**	YP_749909	ER
Shewanella livingstonensis	**46**	AB284098	KS-MAT-ACP_5-KR
Shewanella livingstonensis	**46**	AB373983	AT
Shewanella livingstonensis	**46**	AB373984	KS-CLF-DH_2
Shewanella livingstonensis	**46**	AB373985	ER
Shewanella livingstonensis	**46**	AB284096	PPTase
Shewanella loihica PV-4	**46**	YP_001094781	PPTase
Shewanella loihica PV-4	**46**	YP_001094790	KS-MAT-ACP_5-KR
Shewanella loihica PV-4	**46**	YP_001094791	AT
Shewanella loihica PV-4	**46**	YP_001094792	KS-CLF-DH_2
Shewanella loihica PV-4	**46**	YP_001094793	ER

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Shewanella pealeana ATCC 700345	**46**	YP_001502758	PPTase
Shewanella pealeana ATCC 700345	**46**	YP_001502760	KS-MAT-ACP_5-KR
Shewanella pealeana ATCC 700345	**46**	YP_001502761	AT
Shewanella pealeana ATCC 700345	**46**	YP_001502762	KS-CLF-DH_2
Shewanella pealeana ATCC 700345	**46**	YP_001502763	ER
Shewanella putrefaciens 200	**46**	ZP_01704386	PPTase
Shewanella putrefaciens 200	**46**	ZP_01704383	KS-MAT-ACP
Shewanella putrefaciens 200	**46**	ZP_01704382	ACP_4-KR
Shewanella putrefaciens 200	**46**	ZP_01704381	AT
Shewanella putrefaciens 200	**46**	ZP_01704380	KS-CLF-DH_2
Shewanella putrefaciens 200	**46**	ZP_01704379	ER
Shewanella sp. ANA-3	**46**	YP_870471	PPTase
Shewanella sp. ANA-3	**46**	YP_870473	KS-MAT-ACP_5-KR
Shewanella sp. ANA-3	**46**	YP_870474	AT
Shewanella sp. ANA-3	**46**	YP_870475	KS-CLF-DH_2
Shewanella sp. ANA-3	**46**	YP_870476	ER
Shewanella sp. BR-2	**46**	EU719604	PPTase-KS-MAT-ACP_5-KR-AT-KS-CLF-DH_2-ER
Shewanella sp. MR-4	**46**	YP_734792	PPTase
Shewanella sp. MR-4	**46**	YP_734794	KS-MAT-ACP_5-KR
Shewanella sp. MR-4	**46**	YP_734795	AT
Shewanella sp. MR-4	**46**	YP_734796	KS-CLF-DH_2
Shewanella sp. MR-7	**46**	YP_738773	PPTase
Shewanella sp. MR-7	**46**	YP_738775	KS-MAT-ACP_5-KR
Shewanella sp. MR-7	**46**	YP_738776	AT
Shewanella sp. MR-7	**46**	YP_738777	KS-CLF-DH_2
Shewanella sp. MR-7	**46**	YP_738778	ER
Vibrio sp. MED222	**46**	ZP_01065943	PPTase
Vibrio sp. MED222	**46**	ZP_01065944	KS-MAT-ACP_5-KR
Vibrio sp. MED222	**46**	ZP_01065945	AT
Vibrio sp. MED222	**46**	ZP_01065946	KS-CLF-DH_2
Vibrio sp. MED222	**46**	ZP_01065947	ER
Vibrio splendidus 12B01	**46**	ZP_00990742	PPTase
Vibrio splendidus 12B01	**46**	ZP_00990741	KS-MAT-ACP_5-KR
Vibrio splendidus 12B01	**46**	ZP_00990740	AT
Vibrio splendidus 12B01	**46**	ZP_00990739	KS-CLF-DH_2
Vibrio splendidus 12B01	**46**	ZP_00990738	ER
Vibrio splendidus LGP32	**46**	YP_002417803	PPTase
Vibrio splendidus LGP32	**46**	YP_002417804	KS-MAT-ACP_5-KR
Vibrio splendidus LGP32	**46**	YP_002417805	AT
Vibrio splendidus LGP32	**46**	YP_002417806	KS-CLF-DH_2
Vibrio splendidus LGP32	**46**	YP_002417807	ER

Table S1. GenBank Accession numbers and domain annotations of organisms possessing putative secondary lipid synthases, continued.

Shewanella benthica KT99	**47**	EDQ00220	PPTase
Shewanella benthica KT99	**47**	EDQ00213	KS-MAT-ACP_4-KR
Shewanella benthica KT99	**47**	EDQ00212	AT
Shewanella benthica KT99	**47**	EDQ00211	KS-CLF-DH_2
Shewanella benthica KT99	**47**	EDQ00210	ER
Shewanella oneidensis MR-1	**47**	NP_717216	PPTase
Shewanella oneidensis MR-1	**47**	NP_717214	KS-MAT-ACP_4-KR
Shewanella oneidensis MR-1	**47**	ACCESSION # UNAVAILABLE	AT
Shewanella oneidensis MR-1	**47**	NP_717212	KS-CLF-DH_2
Shewanella oneidensis MR-1	**47**	NP_717210	ER
Shewanella putrefaciens CN32	**47**	YP_001182855	PPTase
Shewanella putrefaciens CN32	**47**	YP_001182853	KS-MAT-ACP_4-KR
Shewanella putrefaciens CN32	**47**	YP_001182852	AT
Shewanella putrefaciens CN32	**47**	YP_001182851	KS-CLF-DH_2
Shewanella putrefaciens CN32	**47**	YP_001182850	ER
Shewanella sp. W3-18-1	**47**	YP_964146	PPTase
Shewanella sp. W3-18-1	**47**	YP_964148	KS-MAT-ACP_4-KR
Shewanella sp. W3-18-1	**47**	YP_964149	AT
Shewanella sp. W3-18-1	**47**	YP_964150	KS-CLF-DH_2
Shewanella sp. W3-18-1	**47**	YP_964151	ER
Streptomyces ghanaensis ATCC 14672	**N/A***	ZP_04690735	KS-CLF-AT

**Streptomyces ghanaensis* ATCC 14672 is a draft genome and the secondary lipid synthase gene cluster may be truncated as it occurs at the end of a scaffold; thus it was excluded from the NMDS analysis.

Chapter 1, in full, is a reprint of the material as it appears in PLoS ONE, 2011, Shulse, Christine; Allen, Eric. The dissertation author was the primary investigator and author of this paper.

Chapter 2:

Diversity and distribution of microbial long-chain fatty acid biosynthetic genes in the marine environment

environmental microbiology

Environmental Microbiology (2010) doi:10.1111/j.1462-2920.2010.02373.x

Diversity and distribution of microbial long-chain fatty acid biosynthetic genes in the marine environment

Christine N. Shulse[1] and Eric E. Allen[1,2]*

[1]*Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA.*

[2]*Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0202, USA.*

Summary

Bacterial production of long-chain fatty acids via a polyketide synthase-related mechanism has thus far only been investigated in isolate-based studies. Here, the genetic capacity for production of long-chain fatty acids was investigated using a culture-independent approach. PCR primers targeting the keto-acyl synthase (KS) domain of the *pfaA* gene involved in omega-3 polyunsaturated fatty acid (PUFA) biosynthesis were used to construct clone libraries to investigate KS sequence diversity in disparate marine habitats. Of the 446 sequences recovered, 123 (27.6%) clustered with KS sequences involved in the synthesis of eicosapentaenoic acid (EPA, C20:5*n*-3), docosahexaenoic acid (DHA, C22:6*n*-3) and arachidonic acid (AA, C20:4*n*-6). The remaining 72.4% of clones formed environmental-only groups or grouped with the KS domains of *pfaA* homologues from organisms producing unidentified products. In total, 17 groups were recovered – four known and 13 newly identified. A query of metagenomic data sets revealed sequences related to EPA KS domains, as well as sequences related to four environmental-only groups discovered in the clone libraries. The phylogenetic affiliation and end product of these environmental-only KS clusters is unknown. These findings reveal a widespread capacity for long-chain fatty acid production in marine microorganisms, including biosynthetic pathways not yet characterized.

Introduction

Core fatty acid biosynthesis is essential for the formation of biological membranes in all bacterial and eukaryotic cells. The majority of *Bacteria* produce fatty acids via the type II fatty acid synthesis system (FAS) in which discrete enzymes encoded by separate genes catalyse specific steps of the biosynthetic pathway (White *et al*., 2005). The dominant cellular fatty acids produced via the type II FAS typically contain between 14 and 18 carbons. An exception to the type II FAS paradigm is found in certain coryneform bacteria of the order *Actinomycetales* that possess iterative type I FAS systems, multifunctional enzyme complexes harbouring catalytic activities as discrete functional domains (Schweizer and Hofmann, 2004). A third bacterial fatty acid synthase system is recognized in a narrow group of predominately marine γ-*Proteobacteria* that includes species of the *Shewanella*, *Photobacterium*, *Moritella, Colwellia* and *Vibrio* genera (Nichols and McMeekin, 2002; Nichols, 2003). This system is responsible for the specific *de novo* synthesis of the long-chain omega-3 polyunsaturated fatty acids (PUFAs) eicosapentaenoic acid (EPA, C20:5*n*-3) and docosahexaenoic acid (DHA, C22:5*n*-3), and the omega-6 PUFA arachidonic acid (AA, C20:4*n*-6) (Russell and Nichols, 1999). Colloquially termed the 'Pfa Synthase', this mechanism of PUFA synthesis proceeds via a novel type I iterative fatty acid synthase/polyketide synthase (FAS/PKS) enzyme complex (Metz *et al*., 2001). The gene cluster dedicated to PUFA production consists of five genes, designated *pfaE-ABCD* (Allen and Bartlett, 2002). The genetic structure of the PUFA biosynthetic genes consists of a four-gene operon that includes *pfaABCD*, while *pfaE*, which encodes the phosphopantetheinyl transferase required for catalytic activation of acyl carrier protein domains, may be unlinked and found elsewhere in the genome (Fig. 1A). Domain organization within *pfa* gene products is also highly conserved. Specifically, *pfaA* [KS-MAT-$ACP_{(2\text{-}9)}$-KR], *pfaD* [ER] and *pfaE* [PPTase] show strict domain conservation whereas *pfaB* and *pfaC* homologues possess slight domain rearrangements or insertions depending on pathway : product and/or species (Fig. 1B).

In addition to polyenoic fatty acyl products, related FAS/PKS-like gene clusters have been shown to synthesize other long-chain fatty acid products. The C_{26}–C_{32} alkyl chains containing hydroxyl and ketone functional groups that comprise the aglycone moiety of cyanobacterial heterocyst glycolipids and the C_{22}–C_{26} fatty acyl chains of phenolic lipids comprising the dormant cysts of the Gram-negative bacterium *Azotobacter vinelandii* (order

Received 12 August, 2010; accepted 7 October, 2010. *For correspondence. E-mail eallen@ucsd.edu; Tel. (+1) 858 534 2570; Fax (+1) 858 534 7313.

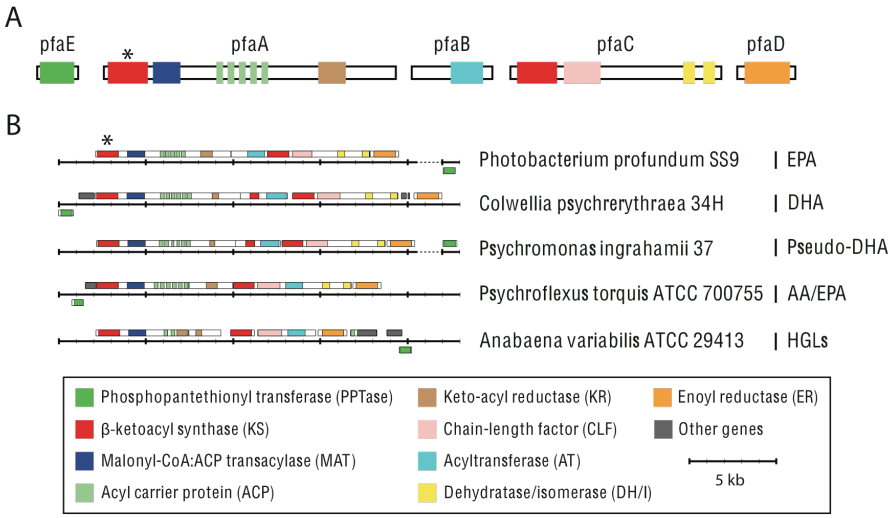

Fig. 1. A. Organization of the five gene cluster, *pfaEABCD*, involved in polyunsaturated fatty acid synthesis.
B. Enzymatic domain architecture within PfaEABCD from representative organisms that produce eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), arachidonic acid (AA) and the aglycone moiety of heterocyst glycolipids (HGLs). Asterisk denotes *pfaA* KS domain targeted in this study.

Pseudomonadales) are both synthesized via an analogous iterative type I FAS/PKS mechanism (Campbell *et al.*, 1997; Miyanaga *et al.*, 2008). Despite sharing a highly similar biosynthetic reaction mechanism, the broad chemical diversity of lipid products observed across the few disparate microbial groups analysed thus far serves to underscore the potential novelty of this biosynthetic pathway.

Although the structural role of heterocyst glycolipids and the phenolic lipid components of resting stage cysts has been known for decades (Reusch and Sadoff, 1983; Murry and Wolk, 1989), the function of PUFAs in the microbial membrane remains enigmatic. It has been suggested that they are a homeoviscous adaptive response, as the majority of isolated PUFA producing bacteria are characterized as psychrophilic and/or piezophilic/piezotolerant (Kato and Nogi, 2001). This theory is supported by the few available studies that find *Shewanella* mutants deficient in EPA production are cold and/or pressure sensitive (Sato *et al.*, 2008; Wang *et al.*, 2009). However in contrast to this result, EPA defective mutants in *Photobacterium profundum* SS9 show no appreciable growth deficit at elevated pressure or reduced temperature (Allen *et al.*, 1999). It has additionally been suggested that bacterial EPA may play an antioxidative role (Nishida *et al.*, 2006). The few culture-based studies investigating the genetics and physiology of bacterial PUFA production have utilized strains recovered using isolation strategies on rich media (Bowman *et al.*, 1997), often from the organs of metazoans (DeLong, 1986; Leonardo *et al.*, 1999) and thus not likely representative of free-living bacteria in the open ocean. Because our current knowledge of bacterial PUFA synthesis is based solely on a few cultivated representatives, our understanding of the phylogenetic breadth and ecological distribution of this biosynthetic process is severely limited. It is not known if bacterial PUFA synthesis is a common adaptation to deep sea and permanently cold environments in the 'uncultured majority' or if it is also prevalent in less extreme habitats. The ability of bacteria to produce unique long-chain fatty acid products, including PUFAs, via the Pfa synthase or related FAS/PKS mechanism may thus be greatly underestimated.

To circumvent biases associated with the narrow phylogenetic affiliation of cultured PUFA producing species, we investigated the genetic potential for long-chain fatty acid production via the *pfa* gene FAS/PKS mechanism in six diverse marine habitats using a culture-independent, sequence-based approach. Specifically, the diversity and distribution of *pfaA* keto-acyl synthase (KS) domain sequences was examined via PCR-based analyses and the querying of metagenomic data sets. The KS domain performs a generic condensing function responsible for extending the growing acyl-ACP chain. It has been shown

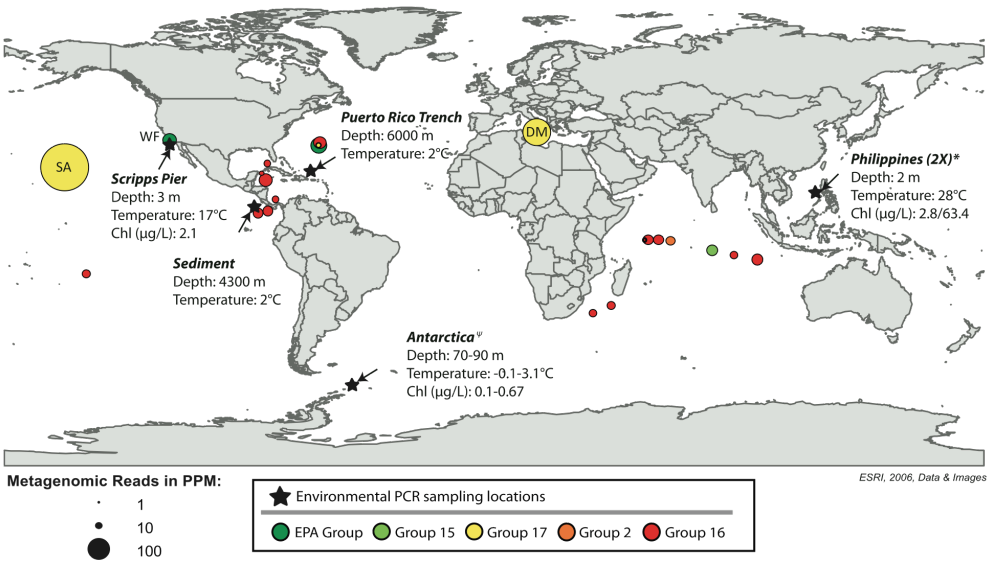

Fig. 2. Map showing location of metagenomic hits to environmental and reference KS reads as well as metadata for sites analysed via PCR. Colours correspond to functional groups shown in Fig. 4. Circle size is proportional to the number of hits retrieved from a given location in parts per million. Abbreviations: SA, Station Aloha; WF, Whale Fall; DM, Deep Mediterranean; no label, Global Ocean Sampling expedition. *Metadata from Garren and colleagues (2009); ψMetadata from Manganelli and colleagues (2009).

that the KS domains of a given PKS generally cluster as a monophyletic group (Ginolhac *et al.*, 2005), thus the KS domain was chosen to provide insight into long-chain fatty acid product diversity in the various environments sampled. Complementing these directed analyses, all publically available metagenomic data sets were queried for the presence of fatty acid producing iterative type I FAS/PKS genetic signatures. In addition to verified fatty acid products from known phylogenetic groups, we report the identification of widespread functional classes likely representing novel products from unidentified phylogenetic groups. The present study expands the catalogue of candidate biosynthetic signatures for cellular hydrocarbon (lipid) production with implications in marine microbial ecophysiology as well as applications in the nascent field of biofuel production from microbial sources.

Results and discussion

Recovery of 'pfa-like' genes from physically and chemically diverse environments

In order to test the hypothesis that PUFA-producing microorganisms are not relegated to high-pressure, low temperature habitats, we interrogated seawater and sediment samples from diverse geographic regions for the presence of '*pfa*-like' genes. Samples were collected from the Puerto Rico Trench, deep-ocean sediments of the Costa Rica Margin, Antarctic waters, two sites in the Philippines, one in close proximity to an aquaculture facility (high chlorophyll a site) and one offshore (low chlorophyll a site), and waters off the Scripps Pier in La Jolla, CA, USA (Fig. 2). These environments cover a range of temperatures and pressures, the two main factors often implicated in PUFA production. Antarctic waters were slightly below 0°C (Manganelli *et al.*, 2009) while waters off the Scripps Pier were 17°C at the time of sampling. Puerto Rico Trench waters were sampled at 6000 m while Philippines waters were sampled at approximately 2 m depth (Garren *et al.*, 2009). Significantly, the water temperature at the Philippines site was 28°C, over 10°C warmer than the other coastal site in La Jolla, CA, USA. DNA was extracted from these samples and used in combination with degenerate PCR primers designed to target the *pfaA* KS domain to investigate *pfa* gene diversity in each of these habitats.

Phylogenetic analysis of environmental samples

446 KS sequences were retrieved from the six sites investigated. Sequences retrieved clustered into 17 functional

groups at 20% genetic distance, representing our empirically derived sequence identity cut-off between different fatty acid products. In our analyses, this was the identity threshold at which known EPA-producing KS domains from species of the orders *Alteromonadales* and *Vibrionales* grouped together. At 20% genetic distance the rarefaction curves for most of these environments approach asymptotic saturation, indicating that nearly all *pfaA* KS product diversity in these environments has likely been sampled (Fig. S1). An exception is the high chlorophyll site in the Philippines, which consists of 42 clones and does not approach asymptotic saturation at this depth of library sequencing. Thirteen of the 17 clades reported, representing 72.4% of all sequences retrieved, did not associate with KS sequences having verified products. Consequently, the majority of sequences recovered are associated with biosynthetic clusters with unknown reaction products. Of the sequences that could be assigned to homologous groups with known phylogenetic and product affinity, 13.6% grouped with known EPA-producing KS domains, 11.6% grouped with DHA-producing KS domains, and 2.2% grouped with KS domains from members of the order *Flavobacteriales* that produce both AA and EPA. The remaining sequences grouped with KS domains found in the family *Psychromondaceae*, which, despite harbouring a coherent *pfa* gene cluster, either does not produce PUFAs, or produces DHA only in trace amounts (Kawasaki *et al*., 2002; Auman *et al*., 2006).

Analysis of KS domains retrieved

No single functional group was shared across all six marine habitats (Fig. 3). Analysis of the co-occurrence of functional groups and the compositional relatedness of the sites sampled suggest sporadic distribution patterns. KS sequences clustering with known EPA producers were retrieved from the Scripps Pier, Antarctic waters, and deep ocean floor sediments. Bacteria producing DHA have been isolated from deep sea sediments (Delong and Yayanos, 1986) and it has been suggested that this bacterial community could provide an important source of PUFAs for benthic animals (Nichols, 2003). Perhaps surprisingly, none of the 89 clones retrieved from the Puerto Rico Trench, a cold and high-pressure environment, was related to EPA-producing KS domains. The percentage of EPA or DHA in the bacterial membrane has been shown to increase as a function of increasing pressure in some deep-sea isolates (Delong and Yayanos, 1986; Allen

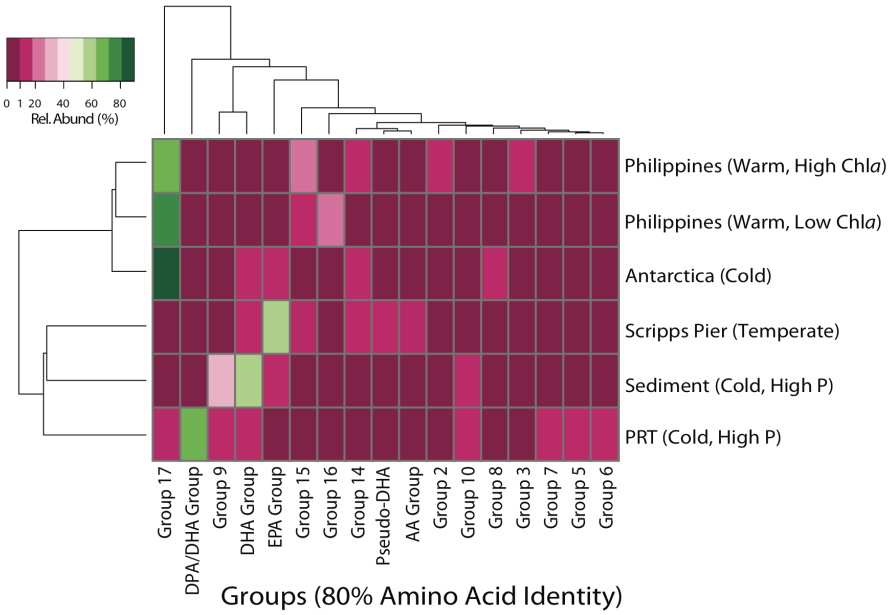

Fig. 3. Heatmap depicting the relative abundance of sequences grouped at 20% phylogenetic distance and normalized to total clones per library. Hierarchical clustering shows relationship among the various environments based on group abundance (*y*-axis) as well as the relationship among the groups based on where they are recovered (*x*-axis). The phylogenetic relationship among groups is shown in Fig. 4.

et al., 1999). The absence of EPA-producing KS domains in a permanently cold and high-pressure environment demonstrates that EPA-producing bacteria may be restricted to specific niches within these environments, such as the organic rich guts of metazoans from which they are routinely cultured. Given that the production of reactive oxygen species (ROS) is among the first host defences against bacterial challenge (Ha *et al*., 2005), the production of EPA, with its demonstrated anti-oxidative effect (Nishida *et al*., 2006), may be a specialized adaptation to living within a metazoan host.

Isolate-based studies have suggested that the genes for EPA synthesis are constitutively transcribed, even under conditions in which no EPA is produced (Allen and Bartlett, 2002; our data not shown). The proteins encoded by these genes are large, multi-domain enzymes and the production of PUFAs (or related long-chain products) is a likely energetically expensive process. Thus it is suggestive that EPA-producing KS regions were only recovered from nutrient-rich sites. The incorporation of polyunsaturated fatty acids into the microbial membrane to retain fluidity at low temperature and high pressure may be an adaptation that only bacteria in the most nutrient-rich environments can afford energetically. However, in contrast to omega-3 PUFA regulation, the production of other long-chain fatty acids shows tight regulation, despite the biosynthetic mechanisms sharing structural and mechanistic homology. The genes encoding the multidomain FAS/PKS product responsible for the aglycone alkyl chains of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133 are only transcribed following a step-down in nitrogen level in its environment (Campbell *et al*., 1997), a condition that precipitates the induction of heterocyst formation. It remains to be seen if the widespread environmental clades show similar environmental control over production in a free-living state.

Seven clones retrieved from waters off the Scripps Pier formed a clade with the KS domains of PfaA from *Psychromonas ingrahamii* and *Psychromonas* sp. CNPT3, which, although they contain *pfaEABCD* (Riley *et al*., 2008), do not produce PUFAs (Delong and Yayanos, 1986; Auman *et al*., 2006).

The DHA clade was the most widely distributed group among the clone libraries, with sequences originating from the Scripps Pier, the Puerto Rico Trench, Antarctic waters and deep ocean sediment. Fifty-eight per cent of the sediment library was composed of clones related to DHA-producing KS domains. The discovery of potential PUFA-producing bacteria in sediment from the Costa Rica Margin implies that PUFA production by autochthonous sediment bacteria may be a widespread phenomenon. PUFA-producing bacteria have been isolated from other deep ocean sites as well as anoxic intertidal flats (Freese *et al*., 2009), and PUFAs have been chemically detected in a variety of sediment samples (Bowman *et al*., 2003; Freese *et al*., 2008). It is unknown whether chemically detected PUFAs are derived from eukaryotes or prokaryotes, although the depths at which they have been found in cores suggest a prokaryotic origin (Freese *et al*., 2008).

Clones related to a KS domain from the *Flavobacteriaceae* species *Psychroflexus torquis*, which produces both AA and EPA, originally isolated from sea ice (Bowman *et al*., 1998), were only recovered from off the Scripps Pier. *Flavobacteria* are some of the most abundant and diverse bacteria in coastal waters (Cottrell and Kirchman, 2000; Alonso *et al*., 2007). Together with γ-*Proteobacteria*, they appear to make up the primary bacterial input of PUFAs into temperate coastal waters. This input may have important implications for the trophic transfer of essential fatty acids in ocean ecosystems. Studies in ponds and lakes have suggested that EPA plays an important role in zooplankton growth and fecundity (Muller-Navarra, 1995; Muller-Navarra *et al*., 2000), noting that an increased EPA/carbon ratio leads to an increase in weight and clutch size in the model organism *Daphnia magna*. The source of this EPA is often assumed to be eukaryotic phytoplankton. The discovery of bacterial EPA-producing KS sequences in temperate coastal waters lends credence to the often-overlooked idea of prokaryotes as sources of essential nutrients in the marine environment (Nichols, 2003).

This study retrieved an unexpected number of sequences that clade with no known long-chain fatty acid producing KS. Phylogenetic identification of the novel environmental clades based on BLAST analysis of the corresponding mate-pair read from metagenomic data sets was mostly ambiguous due to the limited resolution provided by short sequences, although some tentative assignments are discussed below. Group 18 was the most abundant group in the Puerto Rico Trench, representing 63% of the sequences recovered from this environment. No reference sequences associated with Group 18 are available. However, the KS from the PfaA homologue of a marine fungoid protist, *Schizochytrium* sp. ATCC 20888, which produces DHA and the omega-6 product docosapentaenoic acid (DPA) via the same bacterial FAS/PKS mechanism (Metz *et al*., 2001), branched just outside Group 18, and the monophyleny of these two clades had moderate but significant boostrap support (60%) (Fig. 4). Although *Schizochytrium*, of the *Thraustochytriidae* family, also contain components of the canonical eukaryotic elongation-desaturation pathway to produce PUFAs, the PKS pathway is essential for DHA/DPA production in at least one strain (Lippmeier *et al*., 2009). Sequencing has revealed the presence of *pfa* genes in additional strains of *Schizochytrium* (Huang *et al*., 2008; our data not shown), implying that eukaryotic

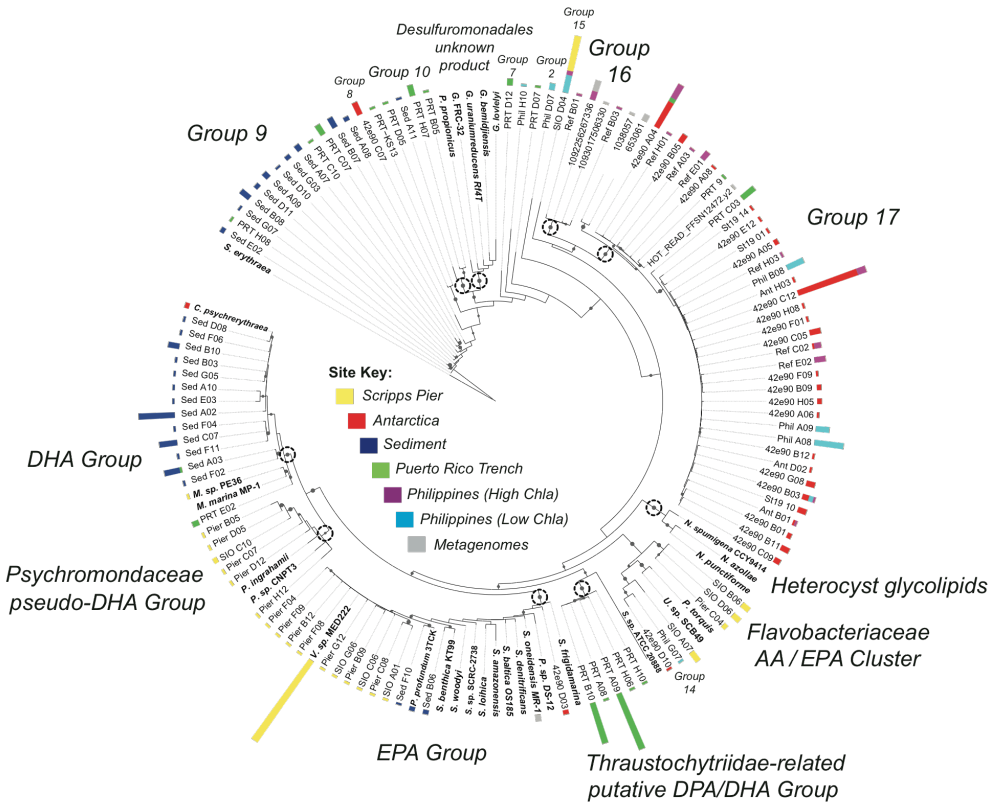

Fig. 4. Maximum likelihood phylogenetic tree of keto-acyl synthase protein domains (161 amino acid alignment, primer regions removed). Source of sequences analysed are shown in the site key. Bar graph heights along the tree's periphery are proportional to the number of sequences recovered in that 99% identity cluster. Terminal labels denote a single representative sequence for that cluster. See Table S3 for a list of clones in each cluster and Table 1 for clone naming conventions. Circled nodes correspond to origination points for labelled groups (based on 80% amino acid identity clustering). All groups consisting of three or more clones are labelled. Reference strain sequences are shown in bold. Bootstrap values 50% are indicated by dots. The *Saccharopolyspora erythraea* EryAI KS domain was used as an outgroup.

production of PUFAs via a PKS mechanism may be more widespread than previously believed. Thraustochytrid species can be easily cultured from deep-sea water (Raghukumar and Raghukumar, 1999) and their 18S sequences are often seen in molecular surveys of deep ocean environments (Lopez-Garcia *et al.*, 2001). Thraustochytrids are prolific PUFA producers with PUFAs comprising as much as 80% of total fatty acids (Burja *et al.*, 2006). Results presented here indicate that the eukaryotic FAS/PKS mechanism's contribution to PUFA input to the pelagic marine food web in deep ocean habitats may be far more significant than bacterial input.

The second most abundant clade in sediments after DHA producers was Group 9, which falls out at the base of the tree (Fig. 4). Additionally, the novel groups 3, 6, 7, 8 and 10 fall near the base of the tree and contain no cultured, verified reference species. However, although it contains no known KSs, Group 10 branches with bootstrap support (79%) with a 'PfaA-like' KS found in the *Deltaproteobacteria* whose product has not been verified. Recent studies have indicated that in *Shewanella*, the *pfa* gene cluster is involved both in the production of polyunsaturated fatty acids (EPA) as well the production of long-chain olefinic hydrocarbons (Sugihara *et al.*, 2010; Sukovich *et al.*, 2010a). It has also been shown that a strain of the deltaproteobacterium *Geobacter* has the ability to produce these long-chain olefinic hydrocarbons (Sukovich *et al.*, 2010b). Therefore, it seems likely that the

'Pfa-like' KS in the *Deltaproteobacteria* is involved in olefinic hydrocarbon production. The novel PfaA KS-like groups identified at the base of the phylogenetic tree in the current study (Fig. 4) may also produce a fatty acid product that is further modified to produce an olefinic hydrocarbon.

Sequence conservation of recovered KS domains

A Cys-His-His active site is required for carbon–carbon bond creation by ketoacyl synthase domains. All but one putative environmental KS protein retrieved in this study contains the catalytic cysteine residue (Fig. S2) which, in the type II FAS homologue, binds the primer substrate so the Claisen condensation reaction can occur, eventually yielding an extended carbon chain (Olsen *et al.*, 2001). The reverse primer designed for this study is anchored in the HGTGT active site, of which the histidine residue is also necessary for KS action (Davies *et al.*, 2000). Furthermore, BLASTX of the recovered putative PfaA KS domain proteins against the NCBI non-redundant database gave top hits exclusively to known PfaA KSs and PfaA-like KSs. Therefore, analysis of the conserved active sites as well as BLAST analysis indicates that these are likely active PfaA-related KSs producing long-chain fatty acids in the marine environment rather than remnants of defunct gene clusters.

Geographic distribution of metagenomic reads

To complement the targeted analysis of *pfaA* KS diversity, we queried all available metagenomic sequence data for reads possessing *pfaA* KS homologues. All *pfaA* KS sequences from reference genomes present in GenBank plus all recovered sequences from the PCR survey were used as our query data set. A total of 67 metagenomic reads were recovered from four data sets representing 23 samples: the Global Ocean Sampling (GOS) Expedition (Rusch *et al.*, 2007), a whale fall metagenome (Tringe *et al.*, 2005), the Hawaii Ocean Time Series (HOT) Station ALOHA (DeLong *et al.*, 2006), and the Mediterranean Bathypelagic Habitat (DEEPMED; Martin-Cuadrado *et al.*, 2007). The GOS hits derived from the North Atlantic, Caribbean, Pacific and Indian Oceans, demonstrating a widespread capacity for bacterial long-chain fatty acid production. Of these metagenomic reads 14 sequences covered the full KS fragment amplified in this study and were included in our phylogenetic analyses (Fig. 4). Because each of the metagenomic reads derived from forward/reserve mate-paired libraries, we interrogated the corresponding mate-pair read for all metagenomic hits in attempts to classify possible taxonomic affinity for each metagenomic sequence as described below.

Long-chain fatty acid biosynthetic genes in the ocean

Metagenomic hits to EPA-producing KS domains were recovered from two sites, 24 from Sargasso Sea Station 11 and one from the rib bone of a grey whale carcass from 1674 meters depth off the California coast (Fig. 2). A near complete genome closely related to *Shewanella* MR-1 was recovered from the Sargasso Sea sample (Venter *et al.*, 2004), although this may represent allochthonous contamination rather than the autochthonous microbial community at this site (DeLong, 2005). *Shewanella* MR-1 is known to contain the EPA-producing gene cluster *pfaE-ABCD* (Wang *et al.*, 2009), and the three Sargasso Sea metagenomic sequences covering the full KS fragment amplified in this study shared 100% amino acid identity with the fragment from the sequenced genome of *Shewanella* MR-1. The 'whalefall' is a nutrient-rich site with relatively low microbial species diversity (Tringe *et al.*, 2005). Given that EPA-producing bacteria have previously been cultured from nutrient-rich metazoan guts, these bacteria may be particularly adapted to exploit cold, nutrient-rich niches in the open ocean.

The remaining 42 metagenomic reads were identified as hits to four environmental only clades, Group 2, Group 15, Group 16 and Group 17 (Fig. 4). Group 2 consists of three clones retrieved from fish pens in the Philippines and the sole metagenomic hit came from a GOS clone recovered from the Indian Ocean, 500 miles west of the Seychelles. The corresponding mate-pair of this read was searched against the NCBI non-redundant database using BLASTX and the top five hits were retrieved. These hits were to four disparate bacterial phyla (*Proteobacteria*, *Chloroflexi*, *Planctomycetes* and *Verrucomicrobia*) and thus no conclusive phylogenetic information could be assigned to Group 2 (Table S1).

Group 15 consists of 16 Scripps Pier clones and 10 clones from the Philippines (both sites), all of which shared 99% amino acid identity, and the sole metagenomic hit to clones from Group 15 came from the Indian Ocean. The discovery of Group 15 in both metagenomic data sets and wide-ranging environments such as the Philippines and the Scripps Pier points to a widely dispersed and highly conserved long-chain fatty acid FAS/PKS product produced by unidentified mesophilic marine bacteria. Mate-pair analysis was inconclusive due to the extremely limited sample size (Table S1).

Group 16 represented the most abundant and widely dispersed clade in the metagenomic data, with 25 hits to the GOS metagenomic data set at sites throughout the Indian, Atlantic and Pacific Oceans. From the clone libraries, Group 16 was only identified at the 'pristine' site off the coast of the Philippines. The mate-pairs to these 25 hits were analysed for phylogenetic information. Thirteen of these 25 mate-pair reads hit within a *pfa*-like gene cluster, including PfaA or PfaC (Fig. 1), which is to be expected as the GOS clones have 2–6 kilobase pair (kbp)

inserts and the *pfa* gene cluster typically spans up to 20 kbp. Six mate-pair reads were missing and an additional one sequence did not have any hits in the database. Two of the remaining five mate-pairs had top scoring BLAST matches to *Deltaproteobacteria* (Table S1). Most known bacterial PUFA biosynthetic gene clusters are found in marine *Gammaproteobacteria*; these results indicate the presence of a *pfa*-like gene cluster in marine bacteria within multiple proteobacterial classes suggesting possible significant phylogenetic extent of this process.

Group 17, with 163 clones recovered from four sites (Antarctic waters, both Philippines sites, and the Puerto Rico Trench), was the most abundant and widespread group in the clone libraries. The far-ranging prevalence of this clade was upheld in the metagenomic searches. The fifteen metagenomic reads identified derived from a fosmid from 3000 m depth in the Mediterranean (Martin-Cuadrado *et al*., 2007), fosmids from all sites sampled below the mesopelagic at station ALOHA (500, 770 and 4000 m) (DeLong *et al*., 2006), and clones from GOS Sargasso stations 3, 11 and 13. In total, Group 17 was found in three major ocean basins and two seas. In mate-pair analysis, all but one of the Sargasso clones had best hits to the *Proteobacteria*, while the three of the four clones from deeper sites had best matches to *Chlorobia* (Table S1). The *Chlorobia* are commonly known as the green sulfur bacteria and are anoxygenic phototrophs. Thus their presence in the deep sea seems incongruous, although not perhaps unprecedented (Beatty *et al*., 2005), and they may not be metabolically active at depth. If confirmed, this would be the first report of *pfa*-like genes in the *Chlorobia*. The mate-pair from the fourth clone had only one hit to a predicted transcriptional regulator in the *Firmicutes* with an *E*-value of 3.18E-06 and thus may not be a good indicator of phylogeny.

Results presented here demonstrate that culture-free studies can provide enormous insight into *in situ* metabolic potential that culture dependent studies may under or overestimate. The genetic potential for PUFA production observed here is unexpected, with the greatest relative abundance of EPA-producing KSs off the Scripps Pier rather than in the cold, high-pressure environment of the Puerto Rico Trench as might be expected from prior culture-based studies. Furthermore, we have revealed a widespread and diverse capacity for long-chain fatty acid production in the marine environment. This diversity likely encompasses novel lipid functional classes as well as novel phylogenetic groups producing unique fatty acyl products via a FAS/PKS mechanism. Lipids are energy dense molecules. From a biotechnology standpoint, this diversity represents a wealth of microbial hydrocarbon synthesis mechanisms that may be exploited for the production of specialized products for biofuel, petrochemical, nutraceutical or biotechnological applications. Furthermore, these results indicate that the bacterial membrane, a system extensively studied in model organisms such as *E. coli*, still contains much to be revealed in uncultured organisms. We have barely scratched the surface of how bacteria are able to interact with their environment via membrane-based adaptations. These unique lipid products synthesized by novel biosynthetic pathways may have important implications in ecologically relevant domains including intra- and interspecies communication, niche colonization, nutrient acquisition and the trophic dynamics of marine ecosystems.

Experimental procedures

Sample collection, DNA extraction and storage

The samples used in this study are shown in Table 1 and the associated metadata is found in Fig. 2. Samples originated from the Southern Ocean (70–90 m depth) (Manganelli *et al*., 2009), the Pacific Ocean off the Scripps Pier (3 m depth) in La Jolla, CA, the Puerto Rico Trench (6000 m depth – Antarctic Bottom Water) (Eloe *et al*., 2010), the Pacific Ocean at the Costa Rica convergent margin (4000 m depth), and coastal waters off the Philippines (Garren *et al*., 2009). Puerto Rico Trench water samples were collected in October 2008. Water (210 l) was collected using standard 12 l Niskin bottles, filtered through a 3 μm membrane filter (Millipore, Billerica, MA, USA) and collected on a 0.2 μm membrane filter, which was stored frozen until DNA extraction. Genomic DNA extraction was performed as previously described (Rusch *et al*., 2007) excluding the CTAB purification step. DNA was purified using a PureLink Genomic DNA Mini Kit column. Costa Rica convergent margin sediment samples were collected in February 2009 and stored frozen until DNA extraction. DNA from 0.25 g of sediment from 5 cm depth within the core was extracted using a MoBio Power Soil DNA extraction kit (Carlsbad, CA, USA) following the manufacturer's protocol. Fifty litres of Scripps Pier water was collected in May 2008, passed through a 20 μm nylon pre-filter, a 3 μm membrane filter (Millipore) and collected on a 0.2 μm polyether sulfone membrane filter, which was stored frozen until DNA was extracted as described previously (Rusch *et al*., 2007), with the omission of the CTAB purification step. Metadata pertaining to the Scripps Pier was downloaded from the Southern California Coastal Ocean Observing System (http://sccoos.org). Additional DNA from the Antarctic and the

Table 1. Sampling locations and clone names used in this study.

Sample	Number of clones	Prefix of clone name
Scripps Pier, La Jolla, CA	93	SIO or Pier
Costa Rica Margin Sediment	77	Sed
Puerto Rico Trench	89	PRT
Antarctic	106	42e90, St19 or Ant
Philippines (high Chl*a*)	42	Phil
Philippines (low Chl*a*)	39	Ref

Philippines was kindly provided by another laboratory and was previously extracted as described (Manganelli *et al*., 2009 and (Garren *et al*., 2009 respectively).

pfaA *KS amplification and amplicon library construction*

pfaA KS domain fragments (Fig. 1) were amplified from DNA extracts obtained from environmental samples using degenerate *pfaA* KS forward and reverse primers (5′-TGGGAAGA RAAYTCWTTCCC-3′) and (5′-GTRCCNGTRCCRTGNG CTTC-3′). These primers correspond to base pair positions 589–608 and 1090–1109, respectively, of the *pfaA* gene from *Photobacterium profundum* SS9 (GenBank Accession No. AF409100, Region: 7454..15175). Degenerate primers were designed based on known *pfaA* KS sequences in NCBI's GenBank and the Joint Genome Institute's (JGI) Integrated Microbial Genomes (IMG) database (http://img.jgi.doe.gov) (Table S2).

Polymerase chain reactions were performed using Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA). PCR amplification was carried out in a total volume of 50 μl containing 85 ng of template DNA, 1.5 mM $MgCl_2$, 100 μM dNTPs, 0.2 μM dNTPs, 0.2 μM primers and 2 U Invitrogen Taq DNA polymerase. The amplification conditions were 94°C for 3 min, 30 cycles at 94°C for 45 s, 50.3°C for 30 s, and 72°C for 1 min 30 s. A second PCR amplification was carried out under the same conditions using 2 μl of the first reaction as template, if necessary. Amplicon products of the expected size (520 bp) were gel purified using the QIAquick gel extraction kit according to the manufacturer's instructions with the modification that DNA was eluted in 30 μl of Buffer EB. PCR products were cloned using the TOPO-TA cloning kit for sequencing (Invitrogen).

Sequencing and sequence analyses

Plasmids were sequenced using the plasmid-targeted primer Universal T7 or Universal T3. Bellerophon (Huber *et al*., 2004) was used to identify potentially chimeric sequences. Of the 470 nearly full-length sequences recovered, 24 putative chimeras were identified and removed from further analysis. Rarefaction analysis, cluster analysis and library comparisons were performed using MOTHUR (Schloss *et al*., 2009).

PfaA KSs were clustered at 99% sequence identity and representative sequences were chosen using MOTHUR. The clones in each cluster and the name of the representative clone are provided in Table S3. KSs were aligned using MUSCLE (Edgar, 2004) and viewed and edited in Jalview (Clamp *et al*., 2004). The best model for the 161 amino acid alignment (excluding primer regions) was evaluated using ProtTest (Abascal *et al*., 2005), a program for selecting a model of protein evolution that uses PHYML (Guindon and Gascuel, 2003) and PAL (Drummond and Strimmer, 2001). ProtTest chose WAG+I based on Akaike criterion. A maximum likelihood phylogenetic tree was constructed using RAxML rapid bootstrapping (Stamatakis *et al*., 2008) through the CIPRES portal (Miller *et al*., 2009) cluster at the San Diego Supercomputer Center. The phylogenetic tree graphic was produced using the Interactive Tree of Life (Letunic and Bork, 2007). Cluster analysis was performed and a heatmap was created using the function heatmap.2 in the R package gplots (R Development Core Team, 2009; Warnes, 2009). Colours came from the R package RColorBrewer (Neuwirth, 2007).

Identification of pfaA *KS homologues in metagenomic data sets and taxonomic analysis*

A BLASTN search of all metagenomic reads available on the CAMERA website (http://camera.calit2.net/) was performed using all recovered environmental *pfaA*-like KSs as well as all available known PUFA-producing KSs as queries with a strict *E*-value cut-off of 10^{-30}. This strict *E*-value was chosen to greatly reduce the possibility of false positive hits. Hits were normalized to the number of reads per site and mapped using the GIS software ArcGIS Desktop. Extracted *pfaA*-like KS sequences and mate-pairs were then used as queries in a BLASTX search against NCBI's non-redundant database, and hits outside the *pfa* gene cluster were used to assist in phylogenetic identification (Table S1).

Nucleotide sequence accession numbers

These sequence data have been submitted to the GenBank database under accession numbers HM637293–HM637738.

Acknowledgements

The authors gratefully acknowledge E. Eloe, D. Fadrosh, M. Garren, T. Hammer, F. Malfatti and G. Subbarao for providing DNA samples. We thank T. Hughes, coordinator of the Geographic Information Systems lab at UC San Diego, for assistance in the production of Fig. 4. We would also like to thank all members of the Allen Lab for their support, comments and critiques. This work was supported in part by a generous gift from the Legler Benbough Foundation to E.E.A.

References

Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21:** 2104–2105.

Allen, E.E., and Bartlett, D.H. (2002) Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9. *Microbiology* **148:** 1903–1913.

Allen, E.E., Facciotti, D., and Bartlett, D.H. (1999) Monounsaturated but not polyunsaturated fatty acids are required for growth of the deep-sea bacterium *Photobacterium profundum* SS9 at high pressure and low temperature. *Appl Environ Microbiol* **65:** 1710–1720.

Alonso, C., Warnecke, F., Amann, R., and Pernthaler, J. (2007) High local and global diversity of *Flavobacteria* in marine plankton. *Environ Microbiol* **9:** 1253–1266.

Auman, A.J., Breezee, J.L., Gosink, J.J., Kampfer, P., and Staley, J.T. (2006) *Psychromonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. *Int J Syst Evol Microbiol* **56:** 1001–1007.

Beatty, J.T., Overmann, J., Lince, M.T., Manske, A.K., Lang, A.S., Blankenship, R.E., *et al.* (2005) An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. *Proc Natl Acad Sci USA* **102:** 9306–9310.

Bowman, J.P., McCammon, S.A., Nichols, D.S., Skerratt, J.H., Rea, S.M., Nichols, P.D., and McMeekin, T.A. (1997) *Shewanella gelidimarina* sp. nov. and *Shewanella frigidimarina* sp. nov., novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5ω3) and grow anaerobically by dissimilatory Fe(III) reduction. *Int J Syst Bacteriol* **47:** 1040–1047.

Bowman, J.P., McCammon, S.A., Lewis, T., Skerratt, J.H., Brown, J.L., Nichols, D.S., and McMeekin, T.A. (1998) *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson *et al.* 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiology* **144:** 1601–1609.

Bowman, J.P., McCammon, S.A., Gibson, J.A.E., Robertson, L., and Nichols, P.D. (2003) Prokaryotic metabolic activity and community structure in Antarctic continental shelf sediments. *Appl Environ Microbiol* **69:** 2448–2462.

Burja, A.M., Radianingtyas, H., Windust, A., and Barrow, C.J. (2006) Isolation and characterization of polyunsaturated fatty acid producing *Thraustochytrium* species: screening of strains and optimization of omega-3 production. *Appl Microbiol Biotechnol* **72:** 1161–1169.

Campbell, E.L., Cohen, M.F., and Meeks, J.C. (1997) A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. *Arch Microbiol* **167:** 251–258.

Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. (2004) The Jalview Java alignment editor. *Bioinformatics* **20:** 426–427.

Cottrell, M.T., and Kirchman, D.L. (2000) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Appl Environ Microbiol* **66:** 5116–5122.

Davies, C., Heath, R.J., White, S.W., and Rock, C.O. (2000) The 1.8 angstrom crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. *Structure* **8:** 185–195.

DeLong, E.F. (1986) Adaptations of deep-sea bacteria to the abyssal environment. PhD Thesis. San Diego, CA, USA: University of Calfornia.

DeLong, E.E. (2005) Microbial community genomics in the ocean. *Nat Rev Microbiol* **3:** 459–469.

Delong, E.F., and Yayanos, A.A. (1986) Biochemical function and ecological significance of novel bacterial lipids in deep-sea procaryotes. *Appl Environ Microbiol* **51:** 730–737.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496–503.

Drummond, A., and Strimmer, K. (2001) PAL: an object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics* **17:** 662–663.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Eloe, E.A., Shulse, C.N., Fadrosh, D.W., Williamson, S.J., Allen, E.E., and Bartlett, D.H. (2010) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep ocean environment. *Environ Microbiol Rep* (in press).

Freese, E., Koster, J., and Rullkotter, J. (2008) Origin and composition of organic matter in tidal flat sediments from the German Wadden Sea. *Org Geochem* **39:** 820–829.

Freese, E., Rutters, H., Koster, J., Rullkotter, J., and Sass, H. (2009) *Gammaproteobacteria* as a possible source of eicosapentaenoic acid in anoxic intertidal sediments. *Microb Ecol* **57:** 444–454.

Garren, M., Raymundo, L., Guest, J., Harvell, C.D., and Azam, F. (2009) Resilience of coral-associated bacterial communities exposed to fish farm effluent. *PLoS ONE* **4:** e7319.

Ginolhac, A., Jarrin, C., Robe, P., Perriere, G., Vogel, T., Simonet, P., and Nalin, R. (2005) Type I polyketide synthases may have evolved through horizontal gene transfer. *J Mol Evol* **60:** 716–725.

Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

Ha, E.M., Oh, C.T., Ryu, J.H., Bae, Y.S., Kang, S.W., Jang, I.H., *et al.* (2005) An antioxidant system required for host protection against gut infection in *Drosophila*. *Dev Cell* **8:** 125–132.

Huang, J.Z., Jiang, X.Z., Zhang, X.W., Chen, W.H., Tian, B.Y., Shu, Z.Y., and Hu, S.N. (2008) Expressed sequence tag analysis of marine fungus *Schizochytrium* producing docosahexaenoic acid. *J Biotechnol* **138:** 9–16.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20:** 2317–2319.

Kato, C., and Nogi, Y. (2001) Correlation between phylogenetic structure and function: examples from deep-sea *Shewanella*. *FEMS Microbiol Ecol* **35:** 223–230.

Kawasaki, K., Nogi, Y., Hishinuma, M., Nodasaka, Y., Matsuyama, H., and Yumoto, I. (2002) *Psychromonas marina* sp. nov., a novel halophilic, facultatively psychrophilic bacterium isolated from the coast of the Okhotsk Sea. *Int J Syst Evol Microbiol* **52:** 1455–1459.

Leonardo, M.R., Moser, D.P., Barbieri, E., Brantner, C.A., MacGregor, B.J., Paster, B.J., *et al.* (1999) *Shewanella pealeana* sp. nov., a member of the microbial community associated with the accessory nidamental gland of the squid *Loligo pealei*. *Int J Syst Bacteriol* **49:** 1341–1351.

Letunic, I., and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:** 127–128.

Lippmeier, J.C., Crawford, K.S., Owen, C.B., Rivas, A.A., Metz, J.G., and Apt, K.E. (2009) Characterization of both polyunsaturated fatty acid biosynthetic pathways in *Schizochytrium* sp. *Lipids* **44:** 621–630.

Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C., and Moreira, D. (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409:** 603–607.

Manganelli, M., Malfatti, F., Samo, T.J., Mitchell, B.G., Wang, H., and Azam, F. (2009) Major role of microbes in carbon fluxes during austral winter in the Southern Drake Passage. *PLoS ONE* **4:** e6941.

Martin-Cuadrado, A.B., Lopez-Garcia, P., Alba, J.C., Moreira, D., Monticelli, L., Strittmatter, A., *et al.* (2007) Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS ONE* **2:** e914.

Metz, J.G., Roessler, P., Facciotti, D., Levering, C., Dittrich, F., Lassner, M., *et al.* (2001) Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science* **293:** 290–293.

Miller, M.A., Holder, M.T., Vos, R., Midford, P.E., Liebowitz, T., Chan, L., *et al.* (2009) *The CIPRES Portals*. CIPRES [WWW document]. URL http://www.phylo.org/sub_sections/portal.

Miyanaga, A., Funa, N., Awakawa, T., and Horinouchi, S. (2008) Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc Natl Acad Sci USA* **105:** 871–876.

Muller-Navarra, D. (1995) Evidence that a highly unsaturated fatty acid limits *Dapnhia* growth in nature. *Arch Hydrobiol* **132:** 297–307.

Muller-Navarra, D.C., Brett, M.T., Liston, A.M., and Goldman, C.R. (2000) A highly unsaturated fatty acid predicts carbon transfer between primary producers and consumers. *Nature* **403:** 74–77.

Murry, M.A., and Wolk, C.P. (1989) Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two layers of the cell envelope. *Arch Microbiol* **151:** 469–474.

Neuwirth, E. (2007) RColorBrewer: ColorBrewer palettes. R package version 1.0-2.

Nichols, D.S. (2003) Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. *FEMS Microbiol Lett* **219:** 1–7.

Nichols, D.S., and McMeekin, T.A. (2002) Biomarker techniques to screen for bacteria that produce polyunsaturated fatty acids. *J Microbiol Methods* **48:** 161–170.

Nishida, T., Orikasa, Y., Ito, Y., Yu, R., Yamada, A., Watanabe, K., and Okuyama, H. (2006) *Escherichia coli* engineered to produce eicosapentaenoic acid becomes resistant against oxidative damages. *FEBS Lett* **580:** 2731–2735.

Olsen, J.G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., and Larsen, S. (2001) Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **9:** 233–243.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0 [WWW document]. URL http://www.R-project.org.

Raghukumar, S., and Raghukumar, C. (1999) Thraustochytrid fungoid protists in faecal pellets of the tunicate *Pegea confoederata*, their tolerance to deep-sea conditions and implication in degradation processes. *Mar Ecol Prog Ser* **190:** 133–140.

Reusch, R.N., and Sadoff, H.L. (1983) Novel lipid components of the *Azotobacter vinelandii* cyst membrane. *Nature* **302:** 268–270.

Riley, M., Staley, J.T., Danchin, A., Wang, T.Z., Brettin, T.S., Hauser, L.J., *et al.* (2008) Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. *BMC Genomics* **9:** 210.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007) The *Sorcerer II* Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5:** 398–431.

Russell, N., and Nichols, D. (1999) Polyunsaturated fatty acids in marine bacteria – a dogma rewritten. *Microbiology* **145:** 767–779.

Sato, S., Kurihara, T., Kawamoto, J., Hosokawa, M., Sato, S.B., and Esaki, N. (2008) Cold adaptation of eicosapentaenoic acid-less mutant of *Shewanella livingstonensis* Ac10 involving uptake and remodeling of synthetic phospholipids containing various polyunsaturated fatty acids. *Extremophiles* **12:** 753–761.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Schweizer, E., and Hofmann, J. (2004) Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol Mol Biol Rev* **68:** 501–517.

Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* **57:** 758–771.

Sugihara, S., Hori, R., Nakanowatari, H., Takada, Y., Yumoto, I., Morita, N., *et al.* (2010) Possible biosynthetic pathways for all *cis*-3,6,9,12,15,19,22,25,28-hentriacontanonaene in bacteria. *Lipids* **45:** 167–177.

Sukovich, D.J., Seffernick, J.L., Richman, J.E., Hunt, K.A., Gralnick, J.A., and Wackett, L.P. (2010a) Structure, function, and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1. *Appl Environ Microbiol* **76:** 3842–3849.

Sukovich, D.J., Seffernick, J.L., Richman, J.E., Gralnick, J.A., and Wackett, L.P. (2010b) Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Appl Environ Microbiol* **76:** 3850–3862.

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., *et al.* (2005) Comparative metagenomics of microbial communities. *Science* **308:** 554–557.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Wang, F., Xiao, X., Ou, H.Y., Gai, Y.B., and Wang, F.P. (2009) Role and regulation of fatty acid biosynthesis in the response of *Shewanella piezotolerans* WP3 to different temperatures and pressures. *J Bacteriol* **191:** 2574–2584.

Warnes, G.R. Includes R source code and/or documentation contributed by (in alphabetical order): Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz and Bill Venables. (2009) *gplots: Various R programming tools for plotting data*. R package version 2.7.4 [WWW document]. URL http://CRAN.R-project.org/package=gplots.

White, S.W., Zheng, J., Zhang, Y.M., and Rock, C.O. (2005) The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* **74:** 791–831.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Rarefaction curves for environmental libraries at an 80% identity bin.
Fig. S2. Multiple alignment of a portion of the PfaA KS amino acid alignment from representative clones and isolates. The conserved catalytic cysteine is marked with a red star.
Table S1. Phylogenetic analysis of clones containing '*pfa*-like' genes retrieved from metagenomic data sets by BLASTX of mate pair sequence against NCBI's nonredundant database. Mate pairs in italics hit within the *pfa* gene cluster or found no matches and were not used for taxonomic purposes. Mate pair read ID denotes GenBank accession number or, if unavailable, read identifier present within the CAMERA database.
Table S2. Bacterial *pfaA* genes used to design degenerate forward and reverse primers.
Table S3. Table showing membership of clusters at 99% identity, and the representative clone displayed on the phylogenetic tree. Reference sequences and metagenomic read are identified by GenBank accession number or, if unavailable, read identifier present within the CAMERA database.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplementary Materials

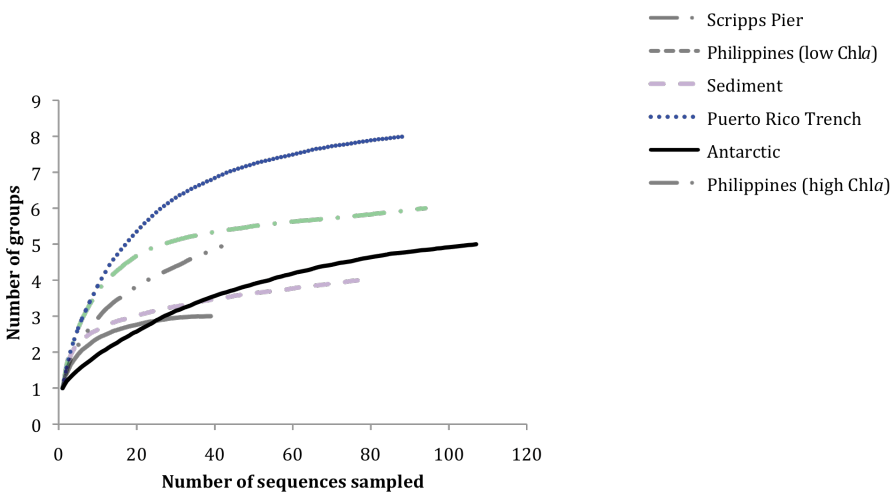

Figure S1. Rarefaction curves for environmental libraries at an 80% identity bin.

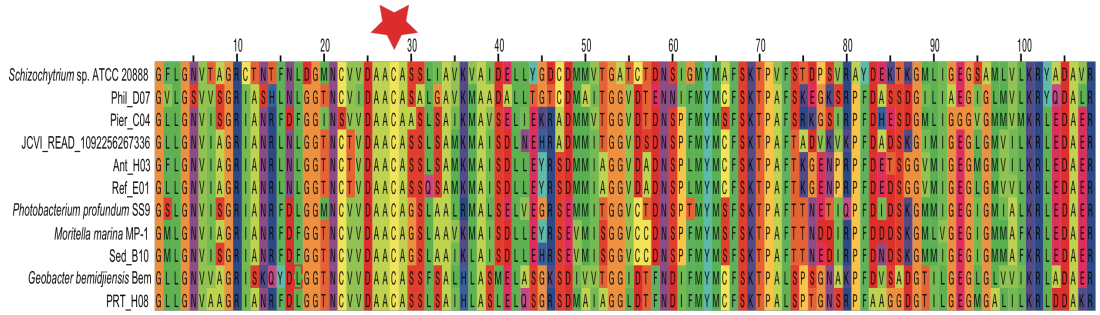

Figure S2. Multiple alignment of a portion of the PfaA KS amino acid alignment from representative clones and isolates. The conserved catalytic cysteine is marked with a red star.

Table S1. Phylogenetic analysis of clones containing "*pfa*-like" genes retrieved from metagenomic datasets by BLASTX of mate pair sequence against NCBI's nonredundant database. Mate pairs in italics hit within the *pfa* gene cluster or found no matches and were not used for taxonomic purposes. Mate pair read ID denotes GenBank Accession number or, if unavailable, read identifier present within the CAMERA database.

Group ID	***Sample (Mate Pair Read ID)***	***Top BLAST Annotation***	***Taxonomy***	***e-value***	***Consensus Taxonomy***
Group 2	**1 Match with Mate Pair Sequence**				
	GS114 – Indian Ocean (JCVI_READ_1108829989016)	haloalkane dehalogenase	Deltaproteobacteria	2.16E-42	Bacteria (5/5)
Group 15	**1 Match with Mate Pair Sequence**				
	GS112b – Indian Ocean (JCVI_READ_1105333670678)	two-component response regulator	Deltaproteobacteria	3.61E-14	Bacteria (5/5)
Group 16	**19 Matches with Mate Pair Sequence**				
	GS048a - Polynesia Archipelagos (JCVI_READ_1099650466684)	*heterocyst glycolipid synthase*	*Cyanobacteria*	*3.98E-97*	*Cyanobacteria (5/5)*
	GS029 – Galapagos Islands (JCVI_READ_1092399702601)	*beta-ketoacyl synthase*	*Cyanobacteria*	*1.00E-65*	*Cyanobacteria (5/5)*
	GS108a – Indian Ocean (JCVI_READ_1103180593508)	*beta-ketoacyl synthase*	*Cyanobacteria*	*2.43E-40*	*Cyanobacteria (4/5)*
	GS110a – Indian Ocean (JCVI_READ_1103242333476)	outer membrane autotransporter barrel "domain," putative	Deltaproteobacteria	4.41E-43	Proteobacteria (2/3)

Table S1. Phylogenetic analysis of clones containing "*pfa*-like" genes retrieved from metagenomic datasets by BLASTX of mate pair sequence against NCBI's nonredundant database. Mate pairs in italics hit within the *pfa* gene cluster or found no matches and were not used for taxonomic purposes. Mate pair read ID denotes GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

	GS114 – Indian Ocean (JCVI_READ_1103242950473)	hypothetical protein Cflav_PD2800	Verrucomicrobia	1.66E-25	Bacteria (5/5)
	GS114 – Indian Ocean (JCVI_READ_1108839692212)	hypothetical protein Anae109_2925	Deltaproteobacteria	1.24E-36	Bacteria (4/4)
	GS114 – Indian Ocean (JCVI_READ_1108839757303)	hypothetical protein Cflav_PD2800	Verrucomicrobia	2.21E-36	Bacteria (4/4)
	GS115 – Indian Ocean (JCVI_READ_1103120881160)	*erythronolide synthase*	*Gammaproteobacteria*	*3.68E-60*	*Gammaproteobacteria (5/5)*
	GS116 – Indian Ocean (JCVI_READ_1103769098047)	*Beta-ketoacyl synthase*	*Cyanobacteria*	*3.06E-52*	*Cyanobacteria (4/5)*
	GS117a – Indian Ocean (JCVI_READ_1108800265420)	*Beta-ketoacyl synthase*	*Cyanobacteria*	*3.39E-39*	*Cyanobacteria (4/5)*
	GS123 – Indian Ocean (JCVI_READ_1103359315152)	*Beta-ketoacyl synthase*	*Cyanobacteria*	*4.30E-24*	*Cyanobacteria (5/5)*
	GS000c – Sargasso Sea (JCVI_READ_1151992)	*polyunsaturated fatty acid synthase PfaA*	*Gammaproteobacteria*	*5.65E-48*	*Gammaproteobacteria (3/5)*
	GS000c – Sargasso Sea (JCVI_READ_1180069)	bifunctional "BirA," biotin operon repressor/biotin--acetyl-CoA-carboxylase ligase	Firmicutes	8.25E-37	Firmicutes (5/5)
	GS017 – Yucatan Channel (JCVI_READ_1092256493236)	*no matches found*			

Table S1. Phylogenetic analysis of clones containing "*pfa*-like" genes retrieved from metagenomic datasets by BLASTX of mate pair sequence against NCBI's nonredundant database. Mate pairs in italics hit within the *pfa* gene cluster or found no matches and were not used for taxonomic purposes. Mate pair read ID denotes GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

	GS121 – Indian Ocean (JCVI_READ_11042301 30002)	*omega-3 polyunsaturated fatty acid synthase PfaA*	*Gammaproteobacteria*	*1.43E-59*	*Gammaproteobacteria (5/5)*
	GS015 – Caribbean Sea (JCVI_READ_10930174 63022)	*heterocyst glycolipid synthase*	*Cyanobacteria*	*1.39E-70*	*Cyanobacteria (5/5)*
	GS018 – Caribbean Sea (JCVI_READ_10930184 07574)	*heterocyst glycolipid synthase*	*Cyanobacteria*	*3.20E-06*	*Cyanobacteria (1/1)*
	GS018 – Caribbean Sea (JCVI_READ_10930186 07686)	*heterocyst glycolipid synthase*	*Cyanobacteria*	*3.08E-102*	*Cyanobacteria (5/5)*
	GS000b – Sargasso Sea (JCVI_READ_866221)	*polyunsaturated fatty acid synthase PfaC*	*Gammaproteobacteria*	*1.09E-58*	*Gammaproteobacteria (5/5)*
Group 17	**15 Matches**				
	HF4000_12-21-03 – North Pacific Subtropical Gyre (DU771976)	adenylylsulfate reductase	Chlorobi	4.37E-104	Chlorobi (5/5)
	HF770_12-21-03 – North Pacific Subtropical Gyre (DU786493)	COG1396: Predicted transcriptional regulators	Firmicutes	3.18E-06	Firmicutes (1/1)
	HF770_12-21-03 – North Pacific Subtropical Gyre (DU794959)	sulfur relay protein TusD/DsrE	Chlorobi	9.29E-39	Chlorobi (5/5)
	DEEPMED – Mediterranean Sea (EI946242)	biotin/lipoate A/B protein ligase	Chlorobi	1.22E-48	Bacteria (5/5)

Table S1. Phylogenetic analysis of clones containing "*pfa*-like" genes retrieved from metagenomic datasets by BLASTX of mate pair sequence against NCBI's nonredundant database. Mate pairs in italics hit within the *pfa* gene cluster or found no matches and were not used for taxonomic purposes. Mate pair read ID denotes GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

	HF_SMPL_HOT179_500M_SG – North Pacific Subtropical Gyre (NCBI_HOT_READ_FFSN12472.x2)	*omega-3 polyunsaturated fatty acid synthase PfaA*	*Gammaproteobacteria*	*3.55E-93*	*Cyanobacteria (4/5)*
	HF_SMPL_HOT179_500M_SG – North Pacific Subtropical Gyre (NCBI_HOT_READ_FFSN36479.g1)	*polyketide synthase phosphopantetheine-binding HglE*	*Cyanobacteria*	*3.31E-67*	*Cyanobacteria (3/5)*
	HF_SMPL_HOT179_500M_SG – North Pacific Subtropical Gyre (NCBI_HOT_READ_FFSN72515.b1)	*beta-ketoacyl synthase*	*Cyanobacteria*	*8.90E-69*	*Cyanobacteria (4/5)*
	GS000b – Sargasso Sea (JCVI_READ_653062)	acriflavin resistance protein	Bacteroidetes	2.27E-43	Bacteroidetes (3/5)
	GS000c – Sargasso Sea (JCVI_READ_1038058)	hypothetical protein sce2292	Deltaproteobacteria	1.94E-33	Deltaproteobacteria (4/5)
	GS000c – Sargasso Sea (JCVI_READ_1160944)	*Beta-ketoacyl synthase*	*Cyanobacteria*	*7.20E-47*	*Cyanobacteria (5/5)*
	GS000c – Sargasso Sea (JCVI_READ_1265546)	RND family efflux transporter MFP subunit	Gammaproteobacteria	8.08E-31	Gammaproteobacteria (4/5)
	GS000c – Sargasso Sea (JCVI_READ_1337222)	protein of unknown function DUF395 YeeE/YedE	Alphaproteobacteria	4.35E-12	Alphaproteobacteria (3/5)
	GS000c – Sargasso Sea (JCVI_READ_1341453)	*polyketide-type polyunsaturated fatty acid synthase PfaA*	*Gammaproteobacteria*	*1.54E-51*	*Gammaproteobacteria (4/4)*
	GS000c – Sargasso Sea (JCVI_READ_1363117)	efflux "transporter," RND "family," MFP subunit	Betaproteobacteria	9.61E-20	Proteobacteria (3/5)

Table S1. Phylogenetic analysis of clones containing "*pfa*-like" genes retrieved from metagenomic datasets by BLASTX of mate pair sequence against NCBI's nonredundant database. Mate pairs in italics hit within the *pfa* gene cluster or found no matches and were not used for taxonomic purposes. Mate pair read ID denotes GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

	GS000d – Sargasso Sea (JCVI_READ_1656027)	*polyunsaturated fatty acid synthase PfaA*	*Gammaproteobacteria*	*1.7 2E-59*	*Gammaproteobacteria (3/5)*

Table S2. Bacterial *pfaA* genes used to design degenerate forward and reverse primers.

Organism	***pfaA* Gene Accession Number**
Photobacterium profundum SS9	AAL01060
Photobacterium profundum 3TCK	ZP_01221855
Vibrio splendidus 12B01	ZP_00990741
Shewanella pealeana ATCC 700345	YP_001502760
Shewanella sp. MR-4	YP_734794
Shewanella sp. MR7	YP_738775
Shewanella sp. ANA-3	YP_870473.1
Shewanella baltica OS195	YP_001553887.1
Shewanella baltica OS185	YP_001365632
Shewanella baltica OS155	YP_001049812.1
Shewanella sp. W3-18-1	YP_964148.1
Shewanella putrefaciens CN-32	YP_001182853.1
Shewanella denitrificans OS217	YP_563629.1
Colwellia psychrerythraea 34H	YP_269802
Moritella sp. PE36	ZP_01898228.1
Psychromonas ingrahamii 37	YP_943075.1

Table S3. Table showing membership of clusters at 99% identity, and the representative clone displayed on the phylogenetic tree. Reference sequences and metagenomic read are identified by GenBank Accession number or, if unavailable, read identifier present within the CAMERA database.

Clones and/or Reference Sequences in Cluster	**Representative on Tree**
Phil_F10, Phil_F11, Phil_D07	Phil_D07
Phil_H10	Phil_H10
PRT_D07	PRT_D07
PRT_D05	PRT_D05
PRT-KS13	PRT-KS13
PRT_H05, PRT_E08, PRT_D12	PRT_D12
42e90_G03, 42e90_C07, 42e90_D04, 42e90_D07, St19_03, Ant_B03	42e90_C07
Sed_E05, Sed_A08	Sed_A08
PRT_E09, PRT_F10, PRT_C12, PRT_F11, PRT_C07	PRT_C07
PRT_C10	PRT_C10
Sed_C11, Sed_C06, Sed_C09, Sed_C08, Sed_B07	Sed_B07
Sed_E01, Sed_A07, Sed_F03	Sed_A07
Sed_D10	Sed_D10
Sed_G12, Sed_D03, Sed_A09	Sed_A09
Sed_G04, Sed_G03	Sed_G03
Sed_F07, Sed_C04, Sed_C10, Sed_B08, Sed_C05	Sed_B08
PRT_H08	PRT_H08
Sed_H11, Sed_E02	Sed_E02
Sed_E09, Sed_D11	Sed_D11
Sed_A11	Sed_A11
PRT_B05	PRT_B05
PRT_G01, PRT_H09, PRT_H07, PRT_D03, PRT_A10	PRT_H07
Geobacter lovleyi SZ (YP_001952503.1)	*Geobacter lovleyi* SZ (YP_001952503.1)
Pelobacter propionicus DSM 2379 (YP_902752.1)	*Pelobacter propionicus* DSM 2379 (YP_902752.1)
Geobacter bemidjiensis Bem (YP_002138914.1)	*Geobacter bemidjiensis* Bem (YP_002138914.1)
Geobacter sp. FRC-32 (YP_002538591.1)	*Geobacter* sp. FRC-32 (YP_002538591.1)
Geobacter uraniireducens Rf4 (YP_001230775.1)	*Geobacter uraniireducens* Rf4 (YP_001230775.1)
Nostoc punctiforme PCC 73102 (YP_001863788.1)	*Nostoc punctiforme* PCC 73102 (YP_001863788.1)
Nodularia spumigena CCY9414 (ZP_01630197.1)	*Nodularia spumigena* CCY9414 (ZP_01630197.1)
SIO_B08, SIO_G10, SIO_B06, SIO_E04	SIO_B06
Pier_C04, Pier_C12	Pier_C04
Pier_B06, Pier_H11, SIO_D06, Pier_C06	SIO_D06
Psychroflexus torquis ATCC 700755 (ZP_01255378)	*Psychroflexus torquis* ATCC 700755 (ZP_01255378)
Unidentified eubacterium SCB49 (ZP_01889250)	Unidentified eubacterium SCB49 (ZP_01889250)
SIO_H11, SIO_A09, SIO_A07, Pier_G05, Pier_F07	SIO_A07
Phil_G07	Phil_G07
42e90_D10	42e90_D10

Table S3. Table showing membership of clusters at 99% identity, and the representative clone displayed on the phylogenetic tree. Reference sequences and metagenomic read are identified by GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

Pier_E12, Pier_F06, SIO_E10, SIO_D04, Pier_G11, Pier_D08, Pier_D10, Pier_C05, Pier_F10, Pier_E11, Pier_A08, Pier_A07, Pier_A12, SIO_D09, SIO_D05, Pier_H09, Phil_D10, Phil_F12, Ref_H05, Phil_E11, Phil_D12, Phil_D08, Phil_C08, Phil_F07, Ref_G02, Phil_C07	SIO_D04
Ref_B03	Ref_B03
Ref_B01	Ref_B01
JCVI_READ_1092256267336, JCVI_READ_1093018364746, Ref_D03, Ref_D01, JCVI_READ_1092256565508, JCVI_READ_1093017714646, JCVI_READ_1092257235826, Ref_D05, Ref_B02, Ref_A02, Ref_A04	JCVI_READ_1092256267336
JCVI_READ_1341454, JCVI_READ_1337223, JCVI_READ_653061	JCVI_READ_653061
JCVI_READ_1038057	JCVI_READ_1038057
Ref_A03	Ref_A03
PRT_9	PRT_9
NCBI_HOT_READ_FFSN12472.y2	NCBI_HOT_READ_FFSN12472.y2
42e90_A08	42e90_A08
Ref_H01	Ref_H01
Ref_C05, Ref_G06, Ref_E01, Ref_E03	Ref_E01
42e90_C02, 42e90_E04, 42e90_H11, 42e90_B05	42e90_B05
Ant_F01, Ref_F04, Ant_C02, Ant_A02, 42e90_G07, 42e90_A04, 42e90_G06, Ref_G04, 42e90_C03, Ref_F06, Ref_F01, Ref_C03, 42e90_H09, Ref_A05, 42e90_E06, Ref_H06, 42e90_E10, Ant_F02, Ant_G03, Ant_E03, PRT_C09, PRT_A07	42e90_A04
PRT_G02, PRT_F12, PRT_F03, PRT_C11, PRT_C03, PRT_B09, PRT_F09	PRT_C03
St19_14	St19_14
Ref_E06, Ref_F03, Ref_G05, Ref_E02, Ref_D02	Ref_E02
42e90_E12	42e90_E12
42e90_B01, St19_13	42e90_B01
42e90_E09, 42e90_B11, 42e90_F07	42e90_B11
42e90_E02, 42e90_C09, Ant_D01	42e90_C09
Phil_F09, Phil_G12, Phil_E10, Phil_C09, Phil_E08, Phil_G09, Phil_H09, Phil_B11, Phil_G11, Phil_D09, Phil_A12, Phil_C11, Phil_A08	Phil_A08
Phil_B10, Phil_H07, Phil_G10, Phil_C10, Phil_H11, Phil_E12, Phil_B08, Phil_E07	Phil_B08
42e90_H08	42e90_H08
42e90_H05, St19_04	42e90_H05
Ant_H03	Ant_H03
Ref_C06, Ant_B01	Ant_B01

Table S3. Table showing membership of clusters at 99% identity, and the representative clone displayed on the phylogenetic tree. Reference sequences and metagenomic read are identified by GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

42e90_F09	42e90_F09
42e90_C04, 42e90_B09	42e90_B09
42e90_F01	42e90_F01
St19_01	St19_01
Ref_H03	Ref_H03
42e90_H03, 42e90_A05	42e90_A05
Ant_D02	Ant_D02
42e90_B03, Phil_C12, Phil_A07, Ref_B05, 42e90_A10, 42e90_F02, 42e90_D08	42e90_B03
St19_11, St19_10, 42e90_C08, 42e90_B08	St19_10
42e90_F11, 42e90_G04, 42e90_B04, 42e90_G10, St19_02, 42e90_H02, 42e90_G08, Ant_G01, 42e90_H06, 42e90_B10	42e90_G08
St19_09, 42e90_C05, 42e90_D09, St19_07, 42e90_C06	42e90_C05
42e90_G12, 42e90_B12	42e90_B12
Phil_B09, Phil_B07, Phil_A10, Phil_A09, Phil_H12, Phil_A11	Phil_A09
Ref_F05, Ref_F02, Ref_C02, 42e90_E05	Ref_C02
42e90_G09, Ant_H02, 42e90_F10, Ref_B04, Ref_A06, St19_08, 42e90_D05, St19_05, Ref_D04, St19_12, 42e90_D11, 42e90_E11, Ant_E02, 42e90_E08, 42e90_E07, Ref_E05, Ant_H01, 42e90_H10, 42e90_B06, 42e90_E03, Ant_C01, 42e90_D01, Ant_E01, 42e90_G11, St19_15, 42e90_A09, 42e90_A03, 42e90_F05, 42e90_F06, 42e90_D02, 42e90_C12, Ref_C01	42e90_C12
PRT_H10	PRT_H10
PRT_H06	PRT_H06
PRT_H02, PRT_G03, PRT_G04, PRT_G07, PRT_G10, PRT_F02, PRT_D11, PRT_D04, PRT_G05, PRT_D01, PRT_G06, PRT_G11, PRT_D10, PRT_E07, PRT_B11, PRT_G08, PRT_A11, PRT_B06, PRT_B12, PRT_C05, PRT_B04, PRT_E10, PRT_D09, PRT_B01, PRT_D08, PRT_A09, PRT_F05, PRT_H03, PRT_G09, PRT_E04, PRT_C02, PRT_B08, PRT_A06	PRT_A09
PRT_A08	PRT_A08
PRT_C01, PRT_A02, PRT_F08, PRT_E05, PRT_F06, PRT_E12, PRT_D02, PRT_C08, PRT_B07, PRT_B03, PRT_C06, PRT_H04, PRT_F07, PRT_E11, PRT_G12, PRT_C04, PRT_A04, PRT_B02, PRT_B10, PRT_F04	PRT_B10
Shewanella amazonensis SB2B (YP_926992.1)	*Shewanella amazonensis* SB2B (YP_926992.1)
Pseudoalteromonas sp. DS-12 (ABF00127.1)	*Pseudoalteromonas* sp. DS-12 (ABF00127.1)
Shewanella frigidimarina NCIMB 400 (YP_749912.1)	*Shewanella frigidimarina* NCIMB 400 (YP_749912.1)
42e90_E01, 42e90_D03	42e90_D03

Table S3. Table showing membership of clusters at 99% identity, and the representative clone displayed on the phylogenetic tree. Reference sequences and metagenomic read are identified by GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

Shewanella loihica PV-4 (YP_001094790)	*Shewanella loihica* PV-4 (YP_001094790)
Sed_H08, Sed_B06	Sed_B06
Shewanella denitrificans OS217 (YP_563629.1)	*Shewanella denitrificans* OS217 (YP_563629.1)
Shewanella benthica KT99 (ZP_02158284.1)	*Shewanella benthica* KT99 (ZP_02158284.1)
Shewanella sp. ANA-3 (YP_870473.1), *S.* sp. MR7 (YP_738775), *S.* sp. MR-4 (YP_734794), JCVI_READ_25744, JCVI_READ_475388, *S. oneidensis* MR-1 (AAN54658.1), JCVI_READ_420490	*Shewanella oneidensis* MR-1 (AAN54658.1)
Shewanella baltica OS185 (YP_001365632), *S. putrefaciens* 200 (ZP_01704383), *S. baltica* OS155 (YP_001049812.1), *S.* sp. W3-18-1 (YP_964148.1), *S. baltica* OS195 (YP_001553887.1), *S. putrefaciens* CN-32 (YP_001182853.1)	*Shewanella baltica* OS185 (YP_001365632)
Shewanella woodyi ATCC 51908 (YP_001759810.1)	*Shewanella woodyi* ATCC 51908 (YP_001759810.1)
Shewanella pealeana ATCC 700345 (YP_001502760), *S.* sp. SCRC-2738 (AAB81123.1)	*Shewanella* sp. SCRC-2738 (AAB81123.1)
SIO_A01	SIO_A01
Pier_C08	Pier_C08
Sed_F10	Sed_F10
Photobacterium profundum 3TCK (ZP_01221855), *P. profundum* SS9 (AAL01060), Sed_C03, Sed_H10	*Photobacterium profundum* 3TCK (ZP_01221855)
Pier_B09	Pier_B09
Pier_F04	Pier_F04
Pier_G12	Pier_G12
Pier_F08	Pier_F08
SIO_C06	SIO_C06
Pier_B12	Pier_B12
Pier_F09	Pier_F09
SIO_B07, Pier_B07, Pier_H08, Pier_D11, Pier_E10, Pier_C09, Pier_H05, Pier_A10, Pier_E09, Pier_G04, Pier_E04, Pier_C11, SIO_A05, SIO_H08, Pier_D04, SIO_D12, Pier_H10, SIO_H12, Pier_G06, *Vibrio splendidus* 12B01 (ZP_00990741), Pier_C10, SIO_F03, SIO_A08, SIO_E08, Pier_B11, Pier_G08, Pier_F05, Pier_B10, Pier_B08, SIO_F08, SIO_F06, SIO_G08, *V.* sp. MED222 (ZP_01065944), SIO_E03, Pier_H06, SIO_F02, SIO_G07, Pier_F12, Pier_E08, Pier__A11, Pier_B04, Pier_A04, Pier_E07, Pier_G09, Pier_A09, Pier_A06	*Vibrio* sp. MED222 (ZP_01065944)
Sed_A10	Sed_A10
Sed_E03	Sed_E03
Sed_F11	Sed_F11
Sed_F04	Sed_F04

Table S3. Table showing membership of clusters at 99% identity, and the representative clone displayed on the phylogenetic tree. Reference sequences and metagenomic read are identified by GenBank Accession number or, if unavailable, read identifier present within the CAMERA database, continued.

Sed_B02, Sed_F12, Sed_F09, Sed_C12, Sed_H03, Sed_C07, Sed_G08, Sed_H07	Sed_C07
Sed_D12, Sed_D02, Sed_B12, Sed_B04, Sed_B11, Sed_B01, Sed_E12, Sed_D07, Sed_G06, Sed_F05, Sed_E06, Sed_H06, Sed_H09, Sed_B05, Sed_A02, Sed_A04	Sed_A02
Sed_D08	Sed_D08
42e90_C10, *Colwellia psychrerythraea* 34H (YP_269802), 42e90_C01	*Colwellia psychrerythraea* 34H (YP_269802)
Sed_F06	Sed_F06
Sed_G05	Sed_G05
Sed_B03	Sed_B03
Sed_E08, Sed_G11, Sed_G02, Sed_B10, Sed_D04	Sed_B10
Sed_F02	Sed_F02
Sed_D09, Sed_C01, Sed_E07, Sed_A05, Sed_B09, Sed_C02, Sed_A03, PRT_A12	Sed_A03
PRT_E03, PRT_E02, PRT_D06	PRT_E02
Moritella sp. PE36 (ZP_01898228.1), SIO_C12	*Moritella* sp. PE36 (ZP_01898228.1)
Moritella marina (BAA89382)	*Moritella marina* (BAA89382)
Pier_H12	Pier_H12
Psychromonas sp. CNPT3 (ZP_01216282.1)	*Psychromonas* sp. CNPT3 (ZP_01216282.1)
Psychromonas ingrahamii 37 (YP_943075.1)	*Psychromonas ingrahamii* 37 (YP_943075.1)
Pier_B05	Pier_B05
SIO_E12, SIO_C10	SIO_C10
Pier_D05	Pier_D05
Pier_C07	Pier_C07
Pier_D12	Pier_D12
Schizochytrium sp. ATCC_20888 (AAK72879.2)	*Schizochytrium* sp. ATCC_20888 (AAK72879.2)
Sed_G07	Sed_G07
Nostoc azollae 0708 (YP_003722543.1)	*Nostoc azollae* 0708 (YP_003722543.1)
SIO_G06	SIO_G06
JCVI_READ_1093017506330	JCVI_READ_1093017506330
42e90_A06	42e90_A06

Chapter 2, in full, is a reprint of the material as it appears in Environmental Microbiology, 2011, Shulse, Christine; Allen, Eric. The dissertation author was the primary investigator and author of this paper.

Chapter 3:

Comparative analysis of polyunsaturated fatty acid and hydrocarbon production in the genus *Shewanella*

Abstract

Bacterial production of long-chain omega-3 polyunsaturated fatty acids (PUFAs), such as eicosapentaenoic acid (EPA, 20:5*n*-3) and docosahexaenoic acid (DHA, 22:6*n*-3), is a distinct biosynthetic activity constrained to a narrow subset of predominately marine γ-proteobacteria. The genes responsible for *de novo* PUFA biosynthesis, designated *pfaEABCD*, encode large, multi-domain protein complexes akin to type I iterative fatty acid and polyketide synthases (herein referred to as "Pfa synthases"). Recently, the *pfa* gene cluster in *Shewanella* species has been implicated in the production of a very long chain polyunsaturated hydrocarbon compound, hentriacontanonaene (31:9), via the synthesis of a precursor fatty acyl substrate for *ole* (olefin) gene product-mediated hydrocarbon biosynthesis. Interestingly, despite universal conservation of the *pfa* gene cluster in all sequenced *Shewanella* strains, not all species synthesize PUFA (EPA) products. To further investigate the relationship between PUFA and hydrocarbon biosynthesis via the Pfa synthase mechanism, we have analyzed the phospholipid (fatty acid) and neutral lipid (hydrocarbon) fractions of multiple *Shewanella* strains possessing varying PUFA production potential. Results indicate that in contrast to PUFA production, 31:9 is produced in all strains analyzed. Chemical profiling and growth analysis of *Shewanella* wild-type and mutant (Δ*pfaA* and Δ*oleB*) strains indicate that loss of EPA and loss of hydrocarbon produce differential growth defects at low temperatures.

Introduction

A narrow subset of γ-proteobacteria is known to produce the long-chain omega-3 polyunsaturated fatty acid (PUFA) eicosapentaenoic acid (EPA, 20:5*n*-3), including species of the *Shewanella*, *Photobacterium*, *Moritella, Colwellia* and *Vibrio* genera [1]. The genes responsible for *de novo* PUFA biosynthesis, designated *pfaEABCD* [2], encode large, multi-domain protein complexes akin to type I iterative fatty acid and polyketide synthases [3], known as the Pfa synthase. Bacterial EPA has been hypothesized to provide various advantages to the cell, including resistance to oxidative stress [4] and resistance to membrane stressors such as high pressure and low temperature [5]. Recently, the *pfa* gene cluster in *Shewanella* has also been implicated in the production of a very long chain polyunsaturated hydrocarbon compound, 3,6,9,12,15,19,22,25,28-hentriacontanonaene (31:9) [6,7]. The current model for 31:9 biosynthesis suggests two molecules of an intermediate in EPA biosynthesis, putatively hexadeca-4,7,10,13-tetraenoic acid (16:4*n*-3), serve as substrates for the *ole* gene products, OleABCD, to produce the polyolefinic 31:9 hydrocarbon molecule [7]. The role of OleA in catalyzing the head-to-head condensation of two fatty acyl substrates has been experimentally verified [8,9], while the functions of OleB (a/b hydrolase), OleC (acyl-CoA synthetase-like), and OleD (reductase/dehydrogenase) are predicted from their homology to characterized proteins [8].

It has long been known that the production of PUFAs in marine bacteria increases with decreasing temperature and, in piezophilic species, as a function of increasing hydrostatic pressure [5,10]. Given the multifunctional role of the Pfa

synthase in PUFA and hydrocarbon biosynthesis, previous studies present conflicting results with regard to hydrocarbon production as a function of growth temperature. One study suggests that levels of EPA increase and hydrocarbon decreases with decreased temperature [6]. Conversely, another study shows the amount of hydrocarbon produced increases with decreasing temperature [7]. The regulatory and ecophysiological factors that determine the flux of fatty acyl intermediates through the Pfa synthase towards PUFA or hydrocarbon production are unknown but temperature seems to play an as yet undetermined role.

Various facets of EPA production, including function [11,12,13] and dynamics [14], have been investigated in members of the genus *Shewanella*. One factor that has been correlated with variable levels of EPA production in *Shewanella* species is phylogeny [15]. Kato and Nogi recognized two sub-genus clades of *Shewanella* based on 16S rRNA gene sequence analysis: *Shewanella* group 1 members are cold adapted and produce high levels of EPA (>10% of total fatty acid (TFA)); whereas *Shewanella* group 2 members are mesophilic and produce no or low levels of EPA (≤5% TFA). Despite the universal conservation for EPA biosynthetic potential (*pfaA-E*) in all sequenced group 1 and group 2 species, variable PUFA synthesis suggests differential regulatory phenomena coordinating flux through the Pfa synthase complex. How this throughput correlates with 31:9 hydrocarbon biosynthesis has not been investigated yet it is possible that hydrocarbon production could fluctuate based on phylogeny as well.

The current study investigates multiple phylogenetically diverse strains of Shewanella using chemical profiling as a function of growth temperature to show that

hydrocarbon and EPA content increase with decreasing temperature in both groups of *Shewanella*. Furthermore, analysis of wild-type versus mutant (Δ*pfaA* and Δ*oleB*) strains reveals that hydrocarbon and EPA play temporally distinct roles in adaptation to low temperature.

Materials and Methods

Bacterial Strains and Culture Conditions

All bacterial strains used in this study are listed in Table 1. The following wild-type strains were used: *Shewanella amazonensis* SB2B [16], *Shewanella baltica* OS155 [17], *Shewanella frigidimarina* NCIMB 400 [18], *Shewanella oneidensis* MR-1 [19], *Shewanella pealeana* ATCC 700345 [20], and *Shewanella woodyi* ATCC 51908 [21]. All strains were routinely cultured aerobically in 250-mL culture flasks in Conjugation medium (5 g/L yeast extract and 18 g/L marine broth 2216 (Difco)) with shaking at 225 rpm to late log phase (OD_{600} = 0.7-0.9). To investigate the effects of temperature on the production of C31:9 and EPA, all strains were cultured at 4, 15, 20, and 25°C (n=3). The following mesophilic strains were additionally cultured at higher temperatures: *Shewanella amazonensis* SB2B (30, 37, and 40°C), *Shewanella baltica* OS155 (30°C), and *Shewanella oneidensis* MR-1 (30 and 37°C).

Temperature-dependent growth studies

Each strain was grown to stationary phase in 2216 marine medium at 25°C. Stationary-phase cultures were diluted 1/200 into 5 mL of Conjugation medium in 18

X 150 mm glass culture tubes. Tubes were incubated at 4, 15, or 25°C with shaking at 225 rpm.

Extraction and analysis of lipids

Cells grown at various temperatures were harvested in late exponential phase via centrifugation at 5,000 X g, washed in 10 ml of 50% artificial seawater (16 g/L Sigma sea salts, Sigma-Aldrich), frozen at -80°C, and lyophilized prior to lipid extraction. Lipids were first extracted from a lyophilized cell sample with a mixture of dichloromethane/methanol (2:1, by vol). This crude extract was applied to a 0.4-g celite column, eluted with 10 ml of the dichloromethane/methanol mixture, evaporated completely under a gentle stream of N_2, and extracted with 6 ml of hexane/methanol (1:1, by vol). The hexane/methanol extraction was repeated three times. Both the combined hexane fractions and the methanol fractions were evaporated completely under N_2 and the hexane fraction residues dissolved in 1 ml (final volume) of hexane and stored at -20°C until analysis. Fatty acid methyl esters (FAMEs) were prepared from the methanol fraction as previously described [10]. Briefly, the methanol fraction residue was reacted with 5% H_2SO_4 in anhydrous methanol at 90°C for 90 min, FAMEs were extracted with hexane, and nonesterified fatty acids were saponified with 10% NaCl. The hexane layer was removed and evaporated completely under N_2, redissolved in 1 ml of hexane, and stored at -20°C until analysis.

Gas chromatography-mass spectrometry (GC-MS) analysis

GC-MS analyses were performed with an Agilent Technologies model 7890A GC equipped with a 30 m X 0.32 mm internal diameter (5%-phenyl)-methylpolysiloxane fused silica capillary column and flame ionization detector, connected to an Agilent Technologies model 5975C MS. FAME samples were injected at 140ºC in the splitless mode with a venting time of 0.5 minutes. After five minutes the oven was temperature-programmed from 140ºC to 280ºC at 4ºC/minute. The final temperature of 280ºC was maintained for 5 minutes. Hydrogen was used as the carrier gas, and the injector and detector were both maintained at 260ºC. Peak areas were quantified and mass spectra were acquired and processed using MSD ChemStation software (Agilent Technologies, Santa Clara, California). MS operating conditions were as follows: electron multiplier, 1718 V; transfer line, 250ºC; scan threshold, 150; 2.12 scans/second; mass range, 50-750 atomic mass units; solvent delay, 4 min. Neutral lipid fractions were injected at 50ºC. After 0.5 minutes the oven was temperature programmed from 50ºC to 330ºC at 10ºC/minute. The final temperature of 330ºC was maintained for 10 minutes. The injector and detector were maintained at 250ºC and 340ºC respectively. MS operating conditions were as follows: 3.21 scans/second; mass range, 50-500 atomic mass units; solvent delay, 6 minutes. Tetratriacontane (C34:0) was used as an internal standard. FAMEs were identified by comparing their retention times to authentic FAME standards as well as by electron impact ionization (EI)-GC/MS analysis. The hydrocarbon hentriacontanonaene (31:9) was identified from its previously determined fragmentation pattern [22], as an authentic standard is not available for this compound.

Mutagenesis

Deletion of *oleB* from *S. oneidensis* MR-1 and *S. pealeana* ATCC 700345, and deletion of *pfaA* from *S. pealeana* ATCC 700345 was achieved by the Saffarini Laboratory at the University of Wisconsin – Milwaukee, as previously described [23].

Results

The long-chain hydrocarbon hentriacontanonaene is produced in all analyzed *Shewanella*

The hydrocarbon and EPA content of six strains of *Shewanella* from two phylogenetically distinct groups [15] was analyzed by GC-MS after growth at 4, 15, 20, and 25°C (as well as 30, 37, and 40°C in mesophilic strains, depending on maximum growth temperature).

In contrast to PUFA production, which is completely absent in *S. amazonensis* SB2B at all temperatures and absent in several additional strains at high temperatures despite the ubiquitous presence of the *pfa* genes, 31:9 is produced in all strains analyzed, at almost all temperatures investigated. The exception is the psychrotolerant bacterium *Shewanella woodyi* ATCC 51908, which produces the hydrocarbon at 4, 15, and 20°C but not 25°C.

Figure 1 shows the hydrocarbon content and the amount of EPA expressed as a percentage of total fatty acids from six *Shewanella* strains grown at various temperatures. Both the amount of EPA as a percentage of total fatty acids, as well as the hydrocarbon content, generally decreased with increasing temperature. One exception occurs in *S. baltica* OS155, which shows the typical trend of decreasing

EPA and hydrocarbon with increasing temperature, until 30°C, the strain's optimal growth temperature. At 30°C EPA production falls to zero as expected but hydrocarbon content increases to 1.57 ± 0.57 μg/mg dry cell weight (Figure 1D), the highest content in any strain analyzed.

The maximum production of EPA as a percentage of total fatty acids occurred in all strains at 4°C, except in *S. pealeana* ATCC 700345, where it occurred at 15°C (Figure 1A). In *S. pealeana* EPA represented 17.1% of total fatty acids at 15°C, which was the highest production percentage of any strain analyzed. No other strain produced EPA greater than 6% of total fatty acids at any temperature analyzed.

The absolute amount of hydrocarbon produced at an intermediate temperature of 20ºC was compared across strains (Figure 2). At this temperature there were no clear trends in hydrocarbon based on phylogeny.

Growth defects in hydrocarbon(-) and hydrocarbon(-)/EPA(-) mutants at low temperature

In order to differentiate between the roles of EPA and the hydrocarbon at low temperature, the prolific EPA-producer *Shewanella pealeana* ATCC 700345, a hydrocarbon-deficient mutant (Δ*oleB*), and an EPA- /hydrocarbon-deficient mutant (Δ*pfaA*) were grown at 25ºC and then inoculated into medium at 4, 15, and 25ºC. There was a slight growth defect in both the Δ*oleB* and Δ*pfaA* mutants at 25ºC (Figure 3A). At 15ºC, both the Δ*oleB* mutant and the Δ*pfaA* mutant had an equally prolonged lag phase prior to exponential growth when compared with the wild type strain (Figure 3B), a phenotype that had been previously reported for a hydrocarbon-less strain of the

trace EPA producer *S. oneidensis* MR-1 [7]. At 4ºC the differential effects of EPA and the hydrocarbon became apparent. The hydrocarbon-deficient Δ*oleB* mutant showed the same prolonged lag phase as was observed at 15ºC, while EPA-/hydrocarbon-deficient Δ*pfaA* mutant had both an increased lag phase and a decreased exponential growth rate (Figure 3C). At 4ºC the doubling times based on optical density measurements were 9 h, 12 h, and 18 h for the parental, hydrocarbon-deficient, and hydrocarbon-/EPA-deficient strains respectively. Lastly, the Δ*pfaA* mutant had a modest reduction in growth yield at 4ºC but not at 15ºC or 25ºC (data not shown).

Loss of hydrocarbon has differential effects on EPA production

OleB has been identified as a member of the alpha/beta-hydrolase superfamily [8,24], and is a likely candidate to catalyze the first step diverting the predicted 16:4*n*-3 intermediate from EPA to hydrocarbon production. Therefore, deletions of *oleB* were generated in both the trace EPA producer *S. oneidensis* MR-1 and the prolific EPA producer *S. pealeana* ATCC 700345 to ablate hydrocarbon production and evaluate the effects on EPA production in these strains. The hydrocarbon elutes at 23.7 min in wild-type *S. oneidensis* MR-1 and *S. pealeana* ATCC 700345. The gas chromatograph of the *oleB* mutants showed no peak in this region, indicating that hydrocarbon production had been knocked out in these strains. The hydrocarbon-deficient mutants *S. oneidensis* Δ*oleB* and *S. pealeana* Δ*oleB* were grown at 15ºC and 4ºC and their EPA content was analyzed (Table 2). The *S. oneidensis* Δ*oleB* mutant had elevated EPA content at both 15ºC and 4ºC when compared with the wild type. However, the *S. pealeana* Δ*oleB* mutant only exhibited increased EPA content

compared to wild-type at 4°C with a significant overall decrease in EPA percent composition at 15ºC.

Discussion

The involvement of the Pfa synthase in the production of the long chain hydrocarbon product hentriacontanonaene in addition to its canonical role in PUFA biosynthesis raises several questions about the role and physiological tradeoffs in the production of these secondary lipid molecules.

Previous studies on the role of EPA in adaptation to low temperature and protection against oxidative stress in EPA-deficient (*pfa*) mutant strains must be reevaluated, keeping in mind that disruption of *pfa* gene functions disrupts both EPA and hydrocarbon biosynthesis. Sukovich *et. al*. showed a role for the hydrocarbon in adaptation to low temperature in the trace-EPA producer *S. oneidensis* MR-1 [7]. Because *S. oneidensis* MR-1 produces trace EPA (Table 2), the question remained whether EPA could compensate for the loss of the hydrocarbon in more prolific EPA-producing *Shewanella* strains.

In this study, we show that the 31:9 hydrocarbon and EPA both play important yet distinct roles in adaptation to low temperature in the prolific EPA producer *S. pealeana* ATCC 700345. High levels of EPA (19.5 ± 0.7 of TFA, Table 2) cannot compensate for the loss of the hydrocarbon in the growth of *S. pealeana* Δ*oleB* at 4ºC (Figure 3). This result is consistent with the hydrocarbon product having a unique role in the growth of *Shewanella* at low temperatures, specifically during lag phase. Likewise, the loss of EPA and hydrocarbon (Δ*pfaA*) has detrimental effects on growth

beyond those seen for the loss of the hydrocarbon alone in *S. pealeana*. This growth defect becomes apparent during exponential growth, as the doubling time of the mutant is twice that of the wild-type strain. Previous biophysical studies have suggested that the Pfa synthase product (presumed to be EPA) provides membrane stability, although the mutant used in these studies would also be hydrocarbon-deficient [25]. Further experiments are required to untangle the differential roles of EPA and hydrocarbon during growth.

The abolition of 31:9 hydrocarbon synthesis by *oleB* deletion increases the amount of EPA produced at 4ºC in both *S. oneidensis* MR-1 and *S. pealeana* ATCC 700345 and at 15ºC in *S. oneidensis* (Table 2). Interestingly, the loss of the ability to produce hydrocarbons in *S. pealeana* at 15ºC decreases rather than increases the amount of EPA produced as a percentage of total fatty acids. The mechanism behind this result is unknown and merits further investigation, as very little is known about the regulation of EPA production. The purpose of generating the *oleB* deletion in *S. pealeana* ATCC 700345 and *S. oneidensis* MR-1 was to create strains with no hydrocarbon production and observe the effects of this on EPA production, rather than to demonstrate a specific function for the *oleB* gene product. It is incidental that the data presented here are consistent with (but do not prove) a role for OleB in catalyzing the first step diverting a 16:4 intermediate from PUFA to hydrocarbon production (model shown in Figure 4). Further studies, including complementation of the *oleB* deletion, are needed to conclusively demonstrate the role of OleB (as well as OleC and OleD) in hydrocarbon biosynthesis.

The Pfa synthase seems to produce both less EPA and less 31:9 with increasing temperature across both groups of *Shewanella* (Figure 3). These results indicate that temperature is not a factor determining the final end product of the Pfa synthase in the strains examined here, as was suggested for *Shewanella* sp. strain osh8 [6]. Rather, our results are consistent with a model in which the Pfa synthase itself is less active at higher temperatures, leading to a decrease in both end products, instead of a constitutively active Pfa synthase that shunts intermediates to one pathway or another depending on temperature.

In conclusion, our results demonstrate consistent dynamics of EPA and hydrocarbon production across six strains of *Shewanella* at varying temperatures. Further, we uncover temporally distinct roles for EPA and the hydrocarbon in growth at low temperatures. The hydrocarbon apparently aids cell growth during lag phase, perhaps by stabilizing membrane proteins necessary to cell division during this phase, while EPA provides a growth advantage during exponential phase.

Acknowledgements

We thank K Nealson and A Cheung for providing a selection of *Shewanella* strains used in this study. The GC-MS analyses were carried out in Lihini Aluwihare's laboratory at Scripps Institution of Oceanography, UC San Diego, with the assistance of N Shaul. The *oleB* and *pfaA* deletion mutants were generated by Daad Saffarini's laboratory at the University of Wisconsin, Milwaukee.

References

1. Nichols DS (2003) Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. FEMS Microbiol Lett 219: 1-7.

2. Allen EE, Bartlett DH (2002) Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9. Microbiology 148: 1903-1913.

3. Metz JG, Roessler P, Facciotti D, Levering C, Dittrich F, et al. (2001) Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. Science 293: 290-293.

4. Nishida T, Orikasa Y, Ito Y, Yu R, Yamada A, et al. (2006) *Escherichia coli* engineered to produce eicosapentaenoic acid becomes resistant against oxidative damages. FEBS Lett 580: 2731-2735.

5. Delong EF, Yayanos AA (1986) Biochemical function and ecological significance of novel bacterial lipids in deep-sea procaryotes. Appl Environ Microbiol 51: 730-737.

6. Sugihara S, Hori R, Nakanowatari H, Takada Y, Yumoto I, et al. (2010) Possible biosynthetic pathways for all *cis*-3,6,9,12,15,19,22,25,28-hentriacontanonaene in bacteria. Lipids 45: 167-177.

7. Sukovich DJ, Seffernick JL, Richman JE, Hunt KA, Gralnick JA, et al. (2010) Structure, function, and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1. Appl Environ Microbiol 76: 3842-3849.

8. Sukovich DJ, Seffernick JL, Richman JE, Gralnick JA, Wackett LP (2010) Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. Appl Environ Microbiol 76: 3850-3862.

9. Frias JA, Richman JE, Erickson JS, Wackett LP (2011) Purification and characterization of OleA from *Xanthomonas campestris* and demonstration of a non-decarboxylative Claisen condensation reaction. J Biol Chem 286: 10930-10938.

10. Allen EE, Facciotti D, Bartlett DH (1999) Monounsaturated but not polyunsaturated fatty acids are required for growth of the deep-sea bacterium *Photobacterium profundum* SS9 at high pressure and low temperature. Appl Environ Microbiol 65: 1710-1720.

11. Sato S, Kurihara T, Kawamoto J, Hosokawa M, Sato SB, et al. (2008) Cold adaptation of eicosapentaenoic acid-less mutant of *Shewanella livingstonensis* Ac10 involving uptake and remodeling of synthetic phospholipids containing various polyunsaturated fatty acids. Extremophiles 12: 753-761.

12. Kawamoto J, Kurihara T, Yamamoto K, Nagayasu M, Tani Y, et al. (2009) Eicosapentaenoic acid plays a beneficial role in membrane organization and cell division of a cold-adapted bacterium, *Shewanella livingstonensis* Ac10. J Bacteriol 191: 632-640.

13. Nishida T, Hori R, Morita N, Okuyama H (2010) Membrane eicosapentaenoic acid is involved in the hydrophobicity of bacterial cells and affects the entry of hydrophilic and hydrophobic compounds. FEMS Microbiol Lett 306: 91-96.

14. Wang F, Xiao X, Ou HY, Gai YB, Wang FP (2009) Role and regulation of fatty acid biosynthesis in the response of *Shewanella piezotolerans* WP3 to different temperatures and pressures. J Bacteriol 191: 2574-2584.

15. Kato C, Nogi Y (2001) Correlation between phylogenetic structure and function: examples from deep-sea *Shewanella*. FEMS Microbiol Ecol 35: 223-230.

16. Venkateswaran K, Dollhopf ME, Aller R, Stackebrandt E, Nealson KH (1998) *Shewanella amazonensis* sp. nov., a novel metal-reducing facultative anaerobe from Amazonian shelf muds. Int J Syst Bacteriol 48: 965-972.

17. Ziemke F, Hofle MG, Lalucat J, Rossello-Mora R (1998) Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. Int J Syst Bacteriol 48: 179-186.

18. Bowman JP, McCammon SA, Nichols DS, Skerratt JH, Rea SM, et al. (1997) *Shewanella gelidimarina* sp. nov. and *Shewanella frigidimarina* sp. nov., novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5ω3) and grow anaerobically by dissimilatory Fe(III) reduction. Int J Syst Bacteriol 47: 1040-1047.

19. Venkateswaran K, Moser DP, Dollhopf ME, Lies DP, Saffarini DA, et al. (1999) Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. Int J Syst Bacteriol 49: 705-724.

20. Leonardo MR, Moser DP, Barbieri E, Brantner CA, MacGregor BJ, et al. (1999) *Shewanella pealeana* sp nov., a member of the microbial community associated with the accessory nidamental gland of the squid *Loligo pealei*. Int J Syst Bacteriol 49: 1341-1351.

21. Makemson JC, Fulayfil NR, Landry W, VanErt LM, Wimpee CF, et al. (1997) *Shewanella woodyi* sp. nov., an exclusively respiratory luminous bacterium isolated from the Alboran Sea. Int J Syst Bacteriol 47: 1034-1039.

22. Nichols DS, Nichols PD, McMeekin TA (1995) A new n-C31:9 polyene hydrocarbon from Antarctic bacteria. FEMS Microbiol Lett 125: 281-285.

23. Shroff NP, Charania MA, Saffarini DA (2010) ArcB1, a homolog of *Escherichia coli* ArcB, regulates dimethyl sulfoxide reduction in *Shewanella oneidensis* MR-1. J Bacteriol 192: 3227-3230.

24. Beller HR, Goh E-B, Keasling JD (2010) Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. Appl Environ Microbiol 76: 1212-1223.

25. Usui K, Hiraki T, Kawamoto J, Kurihara T, Nogi Y, et al. (2012) Eicosapentaenoic acid plays a role in stabilizing dynamic membrane structure in the deep-sea piezophile Shewanella violacea: A study employing high-pressure time-resolved fluorescence anisotropy measurement. Biochimica Et Biophysica Acta-Biomembranes 1818: 574-583.

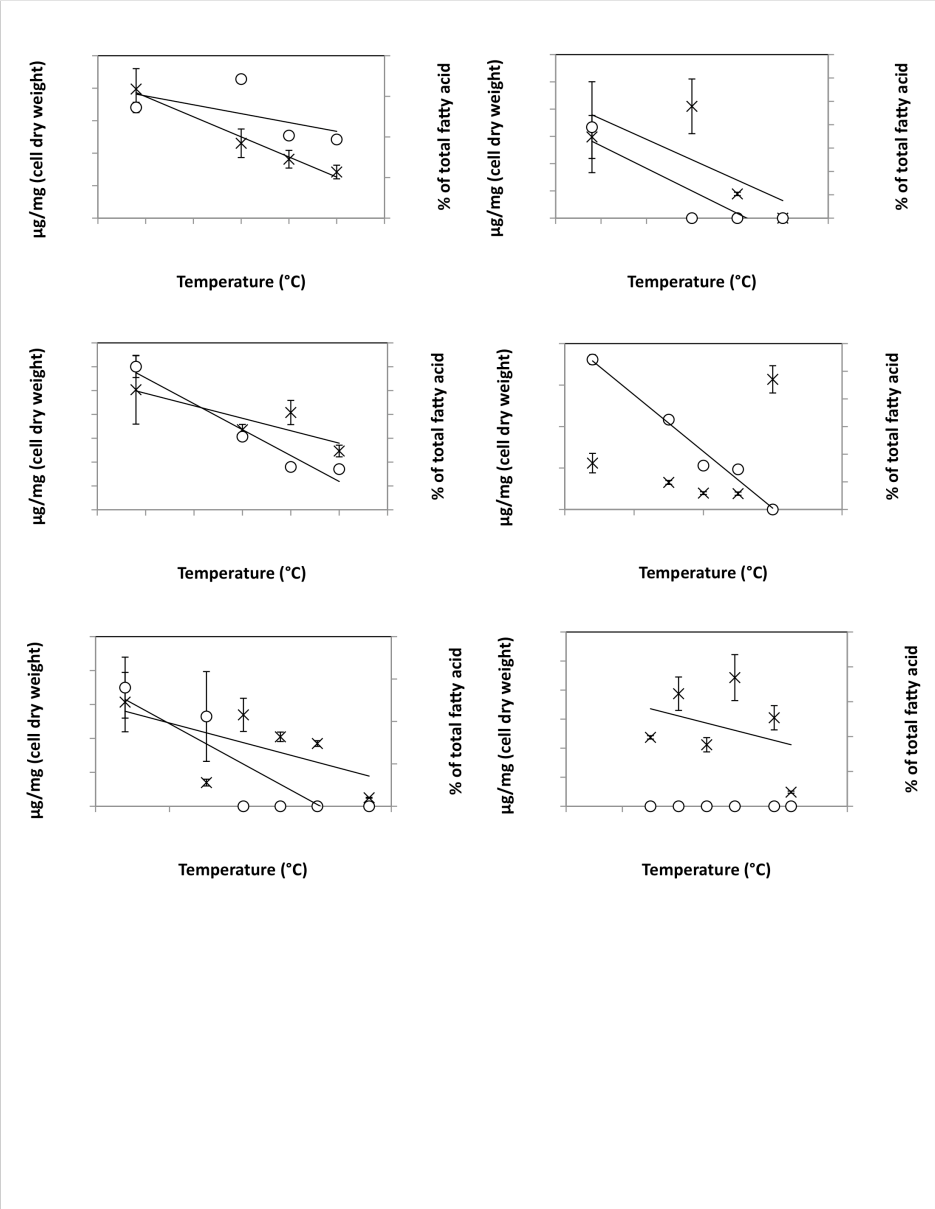

Figure 1. Trends in EPA and hydrocarbon production in phylogenetically diverse strains of *Shewanella* as a function of temperature. Open circles, EPA as a percentage of total fatty acid; X, hydrocarbon in µg/mg (cell dry weight). Error bars indicate standard error of three biological replicates. **(A)** *S. pealeana* ATCC 700345 (Group 1) **(B)** *S. woodyi* ATCC 51908 (Group 1) **(C)** *S. frigidimarina* NCIMB 400 (Group 2) **(D)** *S. baltica* OS155 (Group 2) **(E)** *S. oneidensis* MR-1 (Group 2) **(F)** *S. amazonensis* SB2B (Group 2)

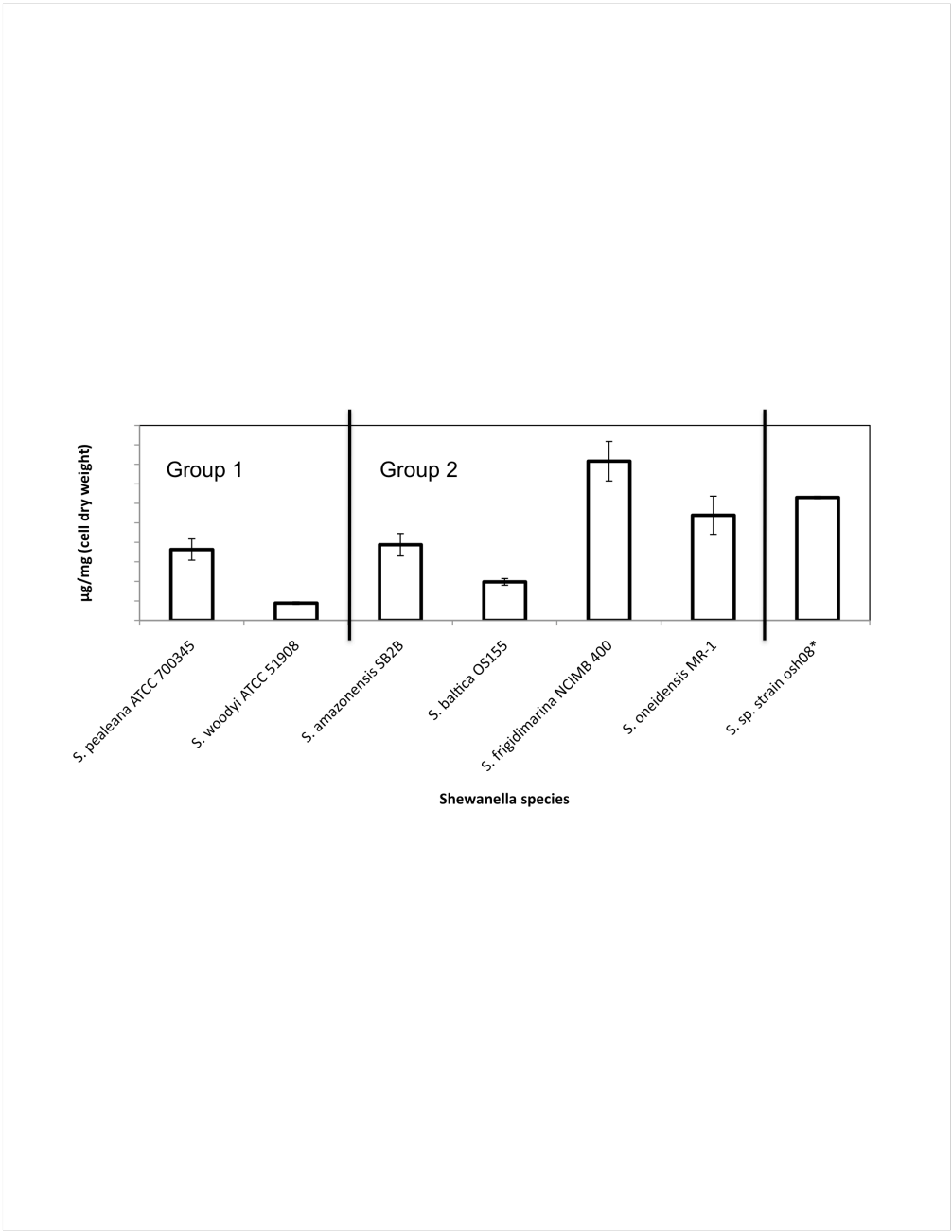

Figure 2. Hydrocarbon production across strains at 20ºC. Error bars indicated standard error of three biological replicates. *Data from Sugihara *et. al*.

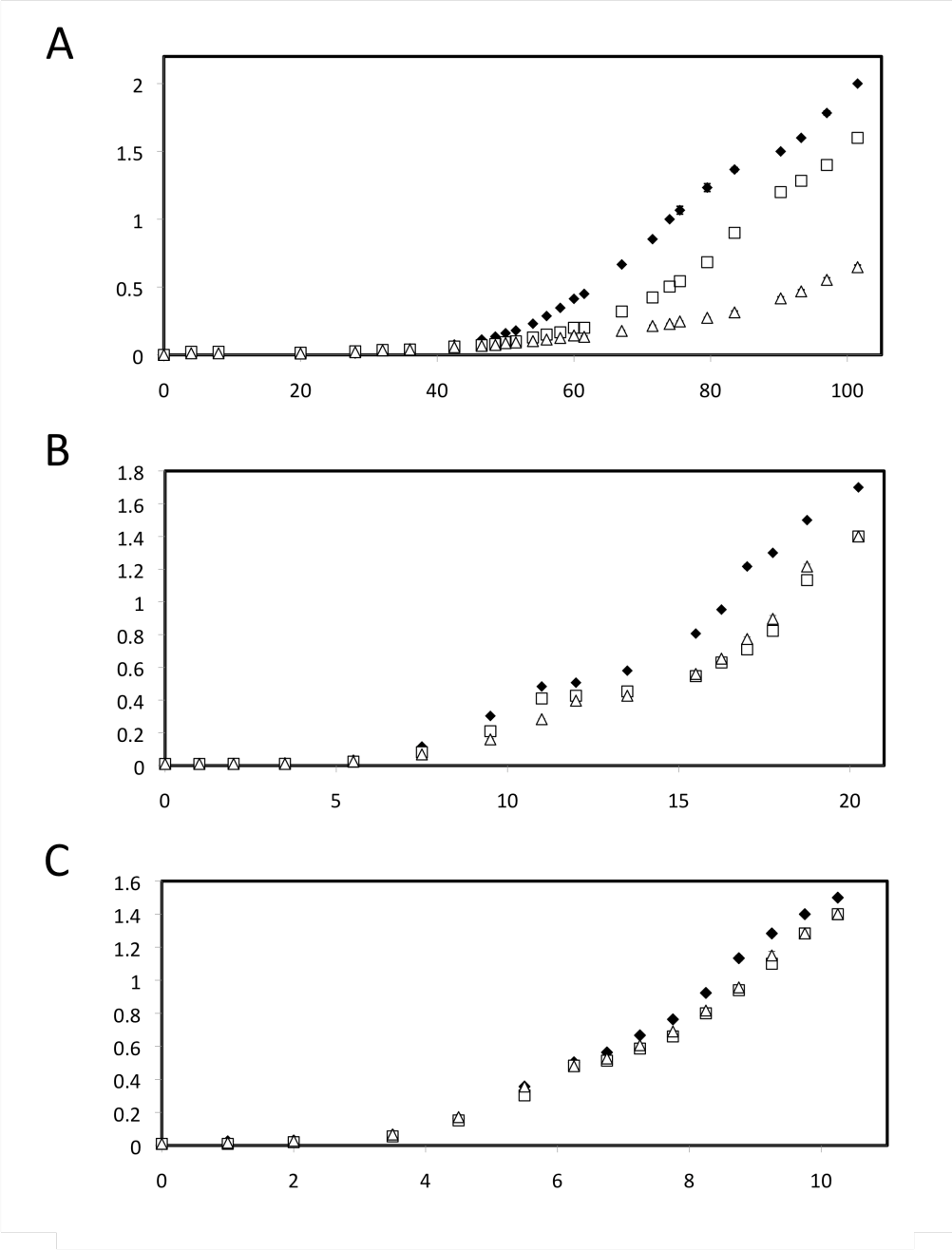

Figure 3. Growth of *S. pealeana* ATCC 700345 and mutants at (A) 4, (B) 15, and (C) 25°C. Closed diamonds, wild-type *S. pealeana* ATCC 700345; open squares, hydrocarbon-less mutant *S. pealeana* Δ*oleB*; open triangles, hydrocarbon-less/EPA-less mutant *S. pealeana* Δ*pfaA*. Error bars indicate standard deviation of three biological replicates.

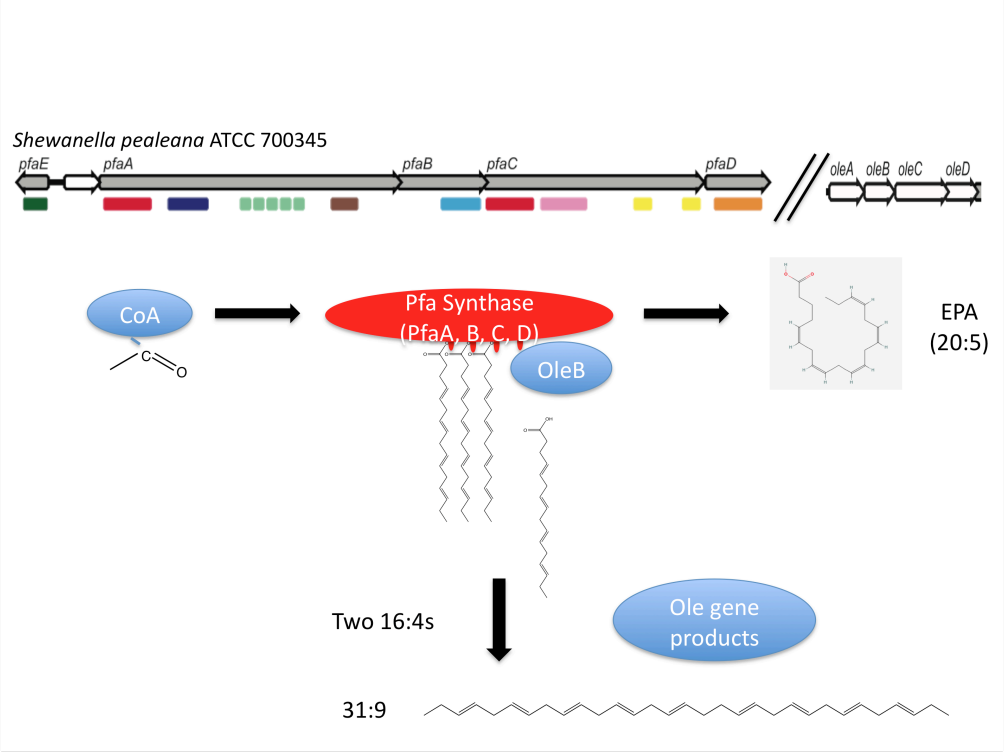

Figure 4. Model for the interactions of the *ole* gene products and the Pfa synthase.

Table 1. Organisms used in this study.

Organism	*Genotype or relevant characteristic(s)*	*Reference*
Wild-type organisms		
S. amazonensis SB2B	Kato and Nogi Group 2	[1]
S. baltica OS155	Kato and Nogi Group 2	[2]
S. frigidimarina NCIMB 400	Kato and Nogi Group 2	[3]
S. oneidensis MR-1	Kato and Nogi Group 2	[4]
S. pealeana ATCC 700345	Kato and Nogi Group 1	[5]
S. woodyi ATCC 51908	Kato and Nogi Group 1	[6]
Genetically modified organisms		
S. oneidensis Δ*oleB*	*S. oneidensis* MR-1, Δ*oleB*; hydrocarbon minus	This study
S. pealeana Δ*oleB*	*S. pealeana* ATCC 700345, Δ*oleB*; hydrocarbon minus	This study
S. pealeana Δ*pfaA*	*S. pealeana* ATCC 700345, Δ*pfaA*; hydrocarbon and EPA minus	This study

1. Venkateswaran K, Dollhopf ME, Aller R, Stackebrandt E, Nealson KH (1998) *Shewanella amazonensis* sp. nov., a novel metal-reducing facultative anaerobe from Amazonian shelf muds. Int J Syst Bacteriol 48: 965-972.
2. Ziemke F, Hofle MG, Lalucat J, Rossello-Mora R (1998) Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. Int J Syst Bacteriol 48: 179-186.
3. Bowman JP, McCammon SA, Nichols DS, Skerratt JH, Rea SM, et al. (1997) *Shewanella gelidimarina* sp. nov. and *Shewanella frigidimarina* sp. nov., novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5ω3) and grow anaerobically by dissimilatory Fe(III) reduction. Int J Syst Bacteriol 47: 1040-1047.
4. Venkateswaran K, Moser DP, Dollhopf ME, Lies DP, Saffarini DA, et al. (1999) Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. Int J Syst Bacteriol 49: 705-724.
5. Leonardo MR, Moser DP, Barbieri E, Brantner CA, MacGregor BJ, et al. (1999) *Shewanella pealeana* sp nov., a member of the microbial community associated with the accessory nidamental gland of the squid *Loligo pealei*. Int J Syst Bacteriol 49: 1341-1351.
6. Makemson JC, Fulayfil NR, Landry W, VanErt LM, Wimpee CF, et al. (1997) *Shewanella woodyi* sp. nov., an exclusively respiratory luminous bacterium isolated from the Alboran Sea. Int J Syst Bacteriol 47: 1034-1039.

Table 2. EPA levels (% of TFA) in *oleB* deletion mutants versus wild type strains as a function of temperature.

	S. pealeana ATCC 700345	*S. pealeana* Δ*oleB*	*S. oneidensis* MR-1	*S. oneidensis* Δ*oleB*
4°C	13.6 ± 2.2	19.5 ± 0.7	0.3 ± 0.2	0.7 ± 0.0
15°C	17.1 ± 0.1	13.4 ± 0.1	0.2 ± 0.4	0.3 ± 0.0

Conclusions and Suggestions for Future Research

Conclusions

Structurally divergent fatty acids were first noted in bacteria some 35 years ago [1]. Only in the past 15 years has the unique mechanism that gives rise to these diverse fatty acids been elucidated [2,3,4]. Although presently methods for identification of microbial fatty acids are quite advanced, all work on the genetics and mechanisms of fatty acid synthesis has been conducted on a modest few microbial isolates due to technical barriers. However, greater than 99% of all microbes remain uncultured, and it is these microbes whose interactions, metabolisms, and physiologies affect life on planet Earth. This dissertation expands our understanding of microbial secondary lipid synthesis to systematically include all sequenced isolates as well as uncultured environmental microbes, and further reveals the trade-offs associated with secondary lipid production in the genus *Shewanella*.

Suggestions for future research

The phylogenetically and geographically widespread presence of putative secondary lipid biosynthetic gene clusters identified in this dissertation was surprising and leads to additional questions. Studies linking the pathways discovered herein with their chemical products are of utmost importance, for both ecological and biotechnological reasons. We already know that the omega-3 polyunsaturated fatty acids structurally identical to those produced via the Pfa synthase in bacteria are important for metazoan health and development; the unique fatty acids produced by additional secondary lipid synthases could likewise have as yet unknown therapeutic or biotechnological benefits. Both the dissertation author and others [5] have had

success driving production of secondary lipids off cosmids and fosmids in *E. coli*. This approach provides a known background against which to look for novel gas chromatograph (GC) peaks. Once noted using GC, novel products can be purified using high performance liquid chromatography (HPLC) and subject to further characterization. Additionally, the genes responsible can be manipulated using genetic tools available in *E. coli* to further evaluate the biosynthetic mechanism. Successful heterologous expression of secondary lipids could also provide such advantages as high levels of production not possible in native strains, or expression of valuable natural products in autotrophic microorganisms.

Thirteen novel groups of putative secondary lipid synthases were discovered in environmental samples using targeted PCR queries, and their environmental relevance was subsequently reaffirmed when four of these groups were found in metagenomic surveys. The phylogenetic affiliations of the microbes harboring these unique synthase genes remain unknown. Future sequencing of metagenomic fosmid clones containing the identified gene clusters could give greater genomic context, allowing us to tentatively ascribe the genes to a specific microbial lineage. Others have had success in linking genes to phylogeny using this approach, perhaps most famously linking proteorhodopsin genes to a wide variety of bacterial lineages from the Global Ocean Survey [6].

One of the 13 novel clades recovered from the environment was found to form a greater monophyletic clade with the KS domain known to be involved in the production of DPA and DHA from the marine fungoid protist, *Schizochytrium* sp. ATCC 20888, of the *Thraustochytriidae* family. Thraustochytrids are prolific

producers of hydrolytic enzymes and important degraders of particulate organic material in marine and estuarine environments [7]. The putative "*Thraustochytriidae*-related" environmental clade was not only exclusive to the Puerto Rico Trench; it was also the most abundant clade in that environment. This result is intriguing, because it is basically unknown how widespread the Pfa synthase is in marine protists. It has previously been identified in only the *Schizochytrium*, where it is presumably the result of a horizontal gene transfer event from bacteria [2]. To further explore the dynamics of the Pfa synthase in marine protists, 18 strains belonging to the *Thraustochytriidae* family were isolated from Southern California coastal waters. In addition to *Schizochytrium* sensu stricto, PCR- based analyses revealed evidence for *PFA* genes in *Thraustochytrium* and unclassified Thraustochytriidae strains. Transcriptome sequencing of one of these isolates, *Thraustochytrium* sp. LLF1b (ATCC PRA-367; NCBI Taxonomy ID 1112570) is currently underway. These results will help us understand the relative contribution of the Pfa synthase and the canonical eukaryotic desaturase-elongase pathway to overall long-chain (lc) polyunsaturated fatty acid production in a model organism harboring both systems.

The physiological significance of the various secondary lipid molecules is an expanding area of contemporary interest. This dissertation provides further support for the role of both a long-chain (lc) polyunsaturated hydrocarbon and a long-chain polyunsaturated fatty acid in cold-adaptation in the genus *Shewanella*. Furthermore, these products play a role at different phases of growth. Others have previously hypothesized EPA specifically interacts with proteins necessary for growth at low temperatures, based on the observation that the levels of four outer-membrane proteins

decrease in a *Shewanella livingstonensis* Δ*pfaA* mutant [8]. However, this mutant presumably lacks both EPA and the lc-hydrocarbon, thus either product could be responsible for this effect. Whole proteome analyses of the *Shewanella pealeana* Δ*pfaA* and *S. pealeana* Δ*oleB* mutants as well as the parental strain *S. pealeana* ATCC 700345 could begin to differentiate between the roles of EPA and lc-hydrocarbon in cellular protein levels. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) is routinely used for proteomics of complex samples and would be the tool of choice for the experiments proposed here. Additionally, a new high pH reverse phase separation approach that could be taken in these experiments has been shown to increase peptide identification substantially [9]. High-resolution proteomic analysis of these mutant strains may identify other potentially important proteins affected by hydrocarbon and EPA synthesis thus significantly expanding the physiological role of these unique secondary metabolites.

Looking forward, the products of the various secondary lipid synthases could prove to have important ecological roles. For instance, some bacteria are known to use fatty acid derivatives as signaling molecules [10]. It is not unreasonable to hypothesize that the products of some of these secondary lipid synthases identified in this dissertation could be further modified to serve as signaling molecules in these understudied phyla.

References

1. Johns RB, Perry GJ (1977) Lipids of the marine bacterium *Flexibacter polymorphus*. Arch Microbiol 114: 267-271.

2. Metz JG, Roessler P, Facciotti D, Levering C, Dittrich F, et al. (2001) Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. Science 293: 290-293.

3. Campbell EL, Cohen MF, Meeks JC (1997) A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. Arch Microbiol 167: 251-258.

4. Miyanaga A, Funa N, Awakawa T, Horinouchi S (2008) Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. Proc Natl Acad Sci USA 105: 871-876.

5. Yazawa K (1996) Production of eicosapentaenoic acid from marine bacteria. Lipids 31: S297-S300.

6. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66-74.

7. Raghukumar S (2002) Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). Eur J Protistol 38: 127-145.

8. Kawamoto J, Kurihara T, Yamamoto K, Nagayasu M, Tani Y, et al. (2009) Eicosapentaenoic acid plays a beneficial role in membrane organization and cell division of a cold-adapted bacterium, *Shewanella livingstonensis* Ac10. J Bacteriol 191: 632-640.

9. Gilar M, Olivova P, Daly AE, Gebler JC (2005) Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. Journal of Separation Science 28: 1694-1703.

10. Ryan RP, Dow JM (2008) Diffusible signals and interspecies communication in bacteria. Microbiology-Sgm 154: 1845-1858.