8-2013

Ecology and Physiology of Aerobic Aromatic Catabolism in Roseobacters

Christopher Adam Gulvik
University of Tennessee, cgulvik@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Christopher Adam Gulvik entitled "Ecology and Physiology of Aerobic Aromatic Catabolism in Roseobacters." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Microbiology.

Alison Buchan, Major Professor

We have read this dissertation and recommend its acceptance:

Steven W. Wilhelm, Erik R. Zinser, John Sanseverino, Jennifer M. DeBruyn

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

ECOLOGY AND PHYSIOLOGY OF AEROBIC AROMATIC CATABOLISM IN ROSEOBACTERS

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee

Christopher Adam Gulvik
August 2013

DEDICATION

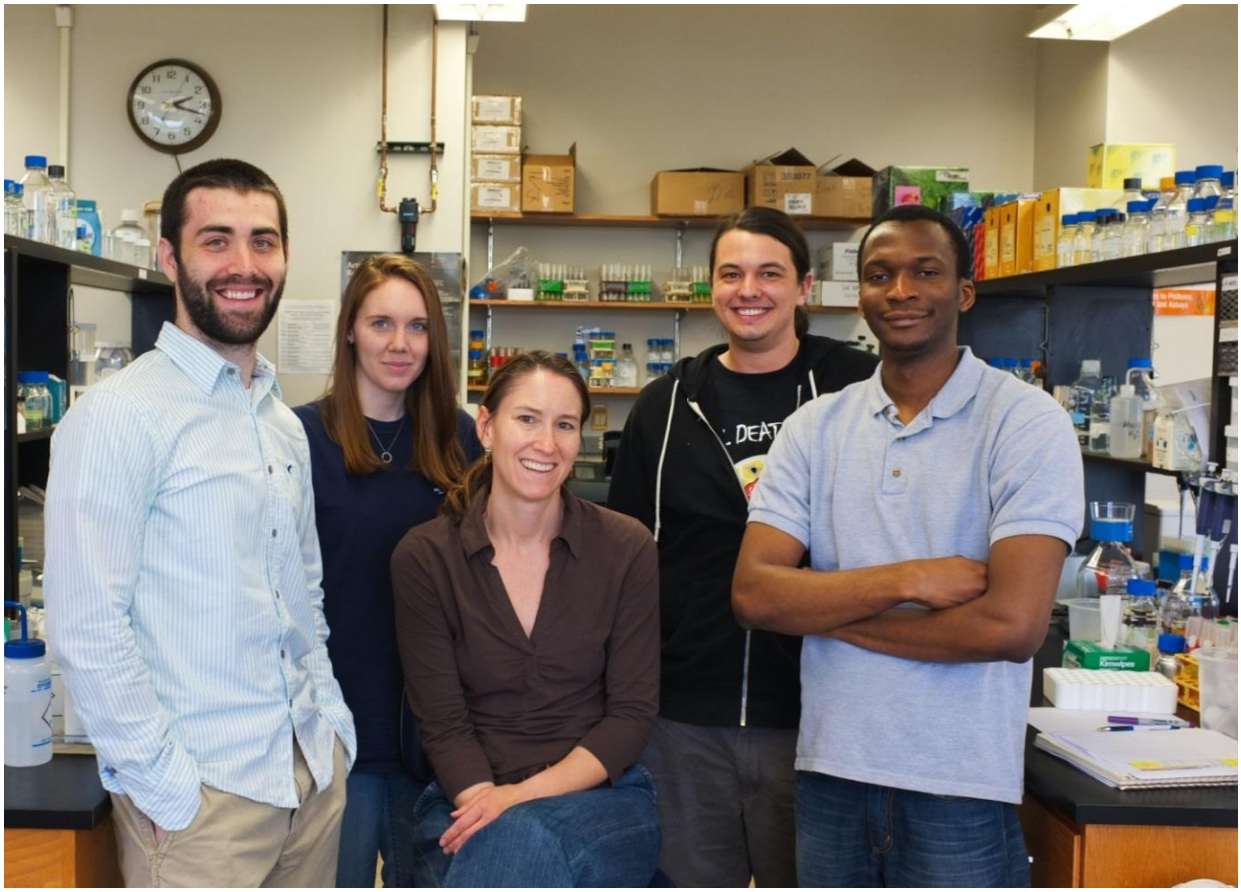
To my maternal grandfather.

ACKNOWLEDGEMENTS

I would have never had the confidence to pursue education in microbiology without so many encouraging years of interacting with others who have prepared me to accomplish this. I will not elaborate on all aspects of influential teachers I have had, but I feel so deeply indebted to the many spectacular science teachers that I must minimally list their names in gratitude for their hard work in training me. Beginning with several *highly* enthusiastic teachers in high school, I would like to thank Ralph Zebell for his hilarious stunts in chemistry classes, Tim Sharpe for teaching me environmental science, John Willauer for physics discussions that really stretched my mind, and Jana Gearhart for an introduction to DNA sequencing and other biotechnology applications. As an undergraduate, Professor Toivo Kallas guided me through bacterial genetics and biotechnology courses meanwhile showing me what I aspire to become—a personable and knowledgeable microbiology professor. Dr. Bob Stelzer showed me how relevant biogeochemistry is, and Dr. Todd Sandrin gave me my first taste in bacterial physiology. Professor Greg Kleinheinz showed me how much impact applied microbiology can have on human public health, and Professor Greg Adler provided me with a strong appreciation for evolutionary, community, and population ecology.

Early on as a graduate student, Dr. Erik Zinser gave me a deep appreciation on how early microbiology experiments were approached and performed that still gives me an "aw" feeling in reflecting on his "Foundations in Microbiology" course, which was both the most intense and influential class I have ever taken. Professor Steve Wilhelm has also showed me additional aspects of what it takes to become a successful professor and has encouraged me to evaluate my own professional skills and goals on a regular basis. As my primary doctoral advisor, Dr. Alison Buchan has given me so much support and one-on-one interactions that I have struggled the most

to state all of her influences on me. Her upbeat and encouraging attitude has inspired me in all aspects of life—both academic and personal. I am truly grateful for all the effort she has given to training me, not just as a microbial ecologist but also as an adult in this world. My lab mates have made time spent in the laboratory especially enjoyable. I am grateful for our lab manager, Mary Hadden, who has always eagerly helped me. Ashley Frank and the other two Buchan Boys (Nathan Cude and Nana Ankrah) helped me through so many difficult and exciting events.



Lastly, I would like to thank my family for their continued support and especially my spouse, Amanda. Without the daily support Amanda provided to me and her continual financial and career sacrifices made for me to pursue this education, absolutely none of this work would have been possible.

ABSTRACT

Roseobacters are an abundant and trophically versatile lineage of marine bacteria that are especially dominant in coastal salt marshes. Central to understanding of how members of the *Roseobacter* clade contribute to biogeochemical cycling in the world's oceans is how these bacteria physiologically respond to mixtures of usable growth substrates present in their environmental niches. A prior study provided evidence that bacterioplankton transcripts most closely related to the Roseobacter *Sagitulla stellata* E-37 are among the most abundant in coastal systems for biogeochemically significant processes of N, P, and S cycling. Thus, this strain was used throughout this dissertation as an environmentally relevant model. Most roseobacter isolates contain multiple aerobic ring-cleaving pathways for the degradation of aromatic compounds, yet it was unknown whether cross-regulation occurred between different parallel pathways. In *S. stellata* E-37, benzoate is catabolized through the aerobic benzoyl-CoA oxidation (*box*) pathway, and *p*-hydroxybenzoate proceeds through the protocatechuate (*pca*) branch of the beta-ketoadipate pathway. Temporal analysis of *S. stellata* E-37 was performed in a mixed-substrate environment containing both benzoate and *p*-hydroxybenzoate and showed that both substrates were simultaneously catabolized at the same rate. Computational and further experimental studies suggest this phenotype appears unique to roseobacters and is anticipated to confer an ecological growth advantage. Additional studies focusing on the environmental relevance of the *box* pathway revealed it is an abundant yet taxonomically constrained pathway to *Betaproteobacteria* and *Roseobacter*. Finally, a study was performed to examine the response of *S. stellata* E-37 to a more chemically complex environment. In coastal salt marshes where roseobacters are dominant, *Spartina alterniflora* (cordgrass) is the most abundant primary producer and *Phaeosphaeria spartinicola* is an abundant and active lignocellulose-degrading

fungus. Cordgrass degradation products (mediated by *P. spartinicola*), representing a salt marsh-like dissolved organic carbon (DOC) mixture, were provided to *S. stellata* E-37 to link metabolic processes to mineralization of cordgrass. Multiple ring-cleaving and carbohydrate metabolism genes were induced during growth on the DOC mixture. Furthermore, the overexpression of central metabolism pathways such as the tricarboxylic acid (TCA) cycle, ATPase, and NADH dehydrogenase suggest a hyperactive trophic strategy for this metabolically versatile roseobacter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1-

INTRODUCTION

I. CARBON CYCLING OF AROMATIC COMPOUNDS

The element carbon is central to life and is present in all four major biomolecule classes: amino acids, carbohydrates, lipids, and nucleic acids. Synthesis of these biomolecules requires a variety of different carbonaceous precursors, and cells use many different enzymes to create these essential compounds for life using available carbon molecules. In the environment, microscopic organisms rarely (if ever) are exposed to all essential compounds, so cells must exploit other enzymes to transform compounds not directly usable for anabolic processes into usable precursors. In recognition that the abundance of certain carbonaceous compounds influences the specific functional processes that occur, much research has focused on chemical composition and metabolic pathways used by microorganisms in natural systems.

Sources of Aromatics

Aliphatic compounds can form into aromatic molecules through abiotic reactions (Wang et al., 2011), though the conditions required for this process are scarce on Earth. An increasingly important source of aromatics is synthetic chemical processing by humans. Microorganisms have been identified that can synthesize and secrete aromatic compounds (Hazelwood et al., 2006; Seyedsayamdost et al., 2011), but by far the largest natural source of aromatic compounds is vascular (lignocellulosic) plants such as trees, shrubs, and grasses. When lignin is degraded, a variety of small molecular weight aromatics are made available as a carbon and energy source for microorganisms. Plants have also been shown to exude aromatics during carbon fixation (Turner and Rice, 1975; Inderjit et al., 2002). Because microorganisms are ubiquitous and plants are the major source of aromatics on Earth, there is strong precedence in studying aromatic degradation by microbes in environments influenced by lignocellulosic plants.

Aromatic Degradation

One class of carbonaceous compounds that environmental microbiologists have studied the degradation of for over a century is aromatic compounds (*e.g.*, Fowler et al., 1910; Happold and Key, 1932; Evans, 1947). Aromatic degradation is of interest biochemically due to the strong π bonds that impart especially high resonance energy and are therefore particularly difficult to break apart (Gilbert et al., 2004). Abiotic reactions can break down aromatic compounds (Huber et al., 2007), but biologically mediated ring fission is the principle way they are degraded in the environment. Enzymes that catalyze ring cleavage are also of interest, in part, due to the relatively few that have been described compared to the large variety of aromatics present in the environment (Fuchs et al., 2011). Large aromatic structures, such as lignin, are unable to be imported across cell membranes and are generally degraded by extracellular enzymes that fungi secrete (*e.g.*, Glenn et al., 1983; Tien and Kirk, 1983), although extracellular ring fission has been observed in some bacteria (Crawford et al., 1983; Ramachandra et al., 1988).

Aromatic compound degradation by microorganisms has been primarily studied in bacteria derived from terrestrial environments. During the initial degradation of lignin from vascular plants (usually mediated by extracellular fungal enzymes), a large variety of smaller molecular weight aromatic compounds are released from heterogeneous lignin structures (Bergbauer and Newell, 1992). With so many aromatic compounds being usable as growth substrates, a surprisingly low number or ring-cleaving routes are normally present in microorganisms. To degrade a large number of aromatic compounds with relatively few ring-cleaving pathways, many different aromatic compounds are often first converted into a relatively limited number of central intermediate aromatic compounds. Then, the energy-intensive steps of mediating ring

fission are accomplished by a more limited number of pathways. This concept of catabolic funneling of aromatic compounds into a central ring-cleaving pathway is best described for the protocatechuate branch of the β-ketoadipate pathway (Figure 1.1). It is not uncommon for bacteria to contain multiple ring-cleaving pathways for the degradation of aromatic compounds, but often only a limited number of ring-cleaving routes are necessary for an organism to degrade a large number of aromatic compounds. In several studies of bacteria derived from soils provided a mixture of usable aromatic compounds (known to proceed through different ring-cleaving pathways), hierarchical substrate preferences have been identified to be the norm (Brzostowicz et al., 2003; Donoso et al., 2011). For example, benzoate is preferentially consumed over *p*-hydroxybenzoate whereby a metabolite or protein in the catechol branch cross-represses the activity of the protocatechuate branch in the β-ketoadipate pathway (Figure 1.2).

II. SIGNIFICANCE OF SALT MARSHES

Beginning in the 1950s, research by Professor Eugene P. Odum of the University of Georgia began studies on salt marsh systems off the coast of the Southeastern United States that are dominated by the primary producer *Spartina alterniflora* (cordgrass) (Figure 1.3). Motivation to study these salt marshes was, in part, due to their especially high productivity that rivals tropical rain forests in terms of the organic carbon made by cordgrass and subsequently degraded and recycled by other biota (Long and Mason, 1983). Early work indicated the importance of macrofauna such as grasshoppers and snails in recycling the abundant cordgrass material (Odum and Smalley, 1959). Additional research considering biotic interactions in these salt marsh systems off the coast of Georgia enabled Odum, often considered the "Father of Ecology", to write the first book in Ecology (Odum, 1953), which outlined seminal ecological concepts.

Further research into the recycling of the cordgrass material in the productive Georgian salt marshes in the 1980s and 1990s was launched by Professor Robert E. Hodson of the University of Georgia. During this time, Professor Hodson published over two dozen studies of cordgrass-derived lignin degradation in coastal salt marshes (*e.g.*, Maccubbin and Hodson, 1980 ; Benner et al., 1985). Seminal work by his students Ronald Benner and Mary Ann Moran revealed aromatics derived from lignocellulosic plants (1) greatly contribute to the dissolved organic carbon (DOC) pool in these Georgia salt marsh systems (Moran and Hodson, 1990, 1994) and (2) are readily degraded by active microbial decomposer communities (Benner et al., 1984; Benner et al., 1986). It was not until the mid-1990s when Moran established her own laboratory to study roseobacter physiologies that group members were specifically linked to aromatic carbon cycling in these systems.

III. SIGNIFICANCE OF ROSEOBACTERS

With the rich history in ecology of salt marshes off the coast of Georgia and recent evidence that microorganisms existed in these systems that actively were degrading the cordgrass material, Professor Mary Ann Moran had a student go back to the salt marshes to isolate lignin-degrading bacteria (González et al., 1996). Using traditional microbiological enrichment approaches with high molecular weight lignin preparations, bacteria degrading lignin from the salt marshes were brought into the laboratory for culturing. Among the most active and dominant bacteria isolated was the obligate aerobe *Sagittula stellata* E-37 which taxonomically belongs to the *Roseobacter* clade. Further studies of this isolate confirmed it is capable of degrading and partially mineralizing synthetic lignin (González et al., 1997) and attaching to lignocellulose particles

(González et al., 1996). During this same time in the mid-1990s, Moran's laboratory also provided evidence roseobacters are some of the most abundant bacteria present in these salt marsh systems based on 16S rRNA gene phylogeny quantitation (González and Moran, 1997).

The term *Roseobacter* can be used to describe the genus *Roseobacter*. In the literature, however, it is most commonly used in reference to the larger *Roseobacter* clade of bacteria that shares 16S rRNA gene sequence identity and clusters members within the α-3 subgroup (within Family *Rhodobacteraceae*) of the class *Alphaproteobacteria* (Buchan et al., 2005). Members of the *Roseobacter* clade contain several genera, and the clade itself was named after the first two cultured isolates, *Roseobacter denitrificans* and *Roseobacter litoralis* that have a pink/rose-colored pigmentation resulting from the production of bacteriochlorophyll *a* (Shiba, 1991). Roseobacters are abundant members of bacterioplankton populations in the world's oceans but are most dominant in coastal marine systems in the oxic, photic layer where they comprise upwards of 30% of the bacterial population (González and Moran, 1997; Suzuki et al., 2001). The *Roseobacter* clade is considered a lineage which doubled their genome size 255 ± 7 million years ago (Moran, 2012). Genetic promiscuity and horizontal gene transfer in roseobacters is apparent (Tang et al., 2010; Luo et al., 2012) and provides one explanation of the high metabolic diversity encoded in their relatively large genomes compared to other abundant marine bacterial clades (*e.g.*, *Prochlorococcus*, SAR116, SAR11).

More recent evidence suggests *Roseobacter* is not only an abundant (Suzuki et al., 2001) but highly active bacterial group in coastal marine environments (Campbell and Kirchman, 2013). Metatranscriptome work by Moran's group also suggests *S. stellata* in particular is among the

most active bacteria in a Marsh Landing, Georgia for biogeochemically relevant cycling of nitrogen, phosphorous, and sulfur (Gifford et al., 2011). Thus, *S. stellata* E-37 is an especially attractive model organism for studying nutrient cycling in coastal salt marsh environments.

IV. STATUS OF ROSEOBACTER PHYSIOLOGY UNDERSTANDING

Microbial degradation of lignocellulosic plants was studied in salt marshes decades before roseobacters were identified as abundant and active taxa in these environments (*e.g.*, Gosselink and Kirby, 1974). Perhaps the most critical step in linking roseobacters to aromatic degradation physiologies was isolation efforts of these bacteria from the environment by Moran's laboratory. For example, a classical enrichment approach was fruitful in isolating individual roseobacters from the salt marshes capable of degrading lignin (González et al., 1996; González et al., 1997). Parallel research in Moran's laboratory by Alison Buchan used newly isolated roseobacters to identify a specific aromatic degradation pathway, the intradiol-cleaving protocatechuate (*pca*) pathway, present and active in many roseobacter isolates capable of degrading aromatic compounds (Buchan et al., 2000; Buchan et al., 2001, 2004). As additional roseobacters with interesting phenotypes were isolated and cultured in the laboratory, genome sequencing of over 50 roseobacters was accomplished. The availability of genomes of the *Roseobacter* lineage has made possible detailed analyses on the metabolic diversity of group members and the development of new hypotheses (Moran et al., 2007; Newton et al., 2010; Tang et al., 2010).

Aside from aromatic catabolism studies of roseobacters initiated by Buchan and colleagues, other major areas of roseobacter physiology have been studied for several years such as sulfur metabolism, physiology in biofilms, and quorum sensing. Moran has continued a strong direction

in understanding how roseobacters contribute to biogeochemical cycling, in particular degradation of sulfur-containing compounds such as dimethylsulphoniopropionate (DMSP) (*e.g.*, González et al., 2003; Moran et al., 2003; Bürgmann et al., 2007; Reisch et al., 2011; Rinta-Kanto et al., 2011), an abundant osmolyte released by phytoplankton (Simo et al., 2002). Intracellular communication among roseobacters and interactions with other marine bacteria has also been studied by several research groups. Using different model roseobacters, tremendous diversity is apparent in terms of the ability of roseobacters to colonize other taxa and diverse types of particulate matter (*e.g.*, Alavi et al., 2001; Gram et al., 2002; Porsby et al., 2008). Biofilm studies of roseobacters indicate they are prolific surface colonizers (Dang and Lovell, 2000, 2002; Dang et al., 2008), and they outcompete other marine bacteria for surfaces (Cude et al., 2012). Complex quorum sensing systems exist in different roseobacters comprised of one or multiple sets of quorum sensing genes (Wagner-Döbler et al., 2005; Cicirelli et al., 2008; Berger et al., 2011). These research areas of roseobacter physiology have all been enhanced by genome sequences for molecular probing (*e.g.*, targeted primer design for PCR) and for genetic manipulation (*e.g.*, Tn-mutagenesis). Development of bioinformatics tools has more recently been able to link specific biogeochemical cycling activities of carbon (Poretsky et al., 2010), nitrogen (Gifford et al., 2011; Mou et al., 2011), and sulfur (Vila-Costa et al., 2010) to roseobacters in natural systems, but laboratory culture-based experiments remain essential to identify specific physiological mechanisms not revealed in culture-independent approaches.

V. OBJECTIVES

Aromatic compounds are prevalent throughout coastal marine systems, where roseobacters exist as a dominant and active portion of the bacterioplankton community. Other than the *pca* route of

the β-ketoadipate pathway, it is unclear which other ring-cleaving pathways roseobacters use. With the availability of many roseobacter genomes, bioinformatics investigations suggest roseobacters contain several other aerobic ring-cleaving pathways. The objectives of this dissertation are to describe physiological characteristics of roseobacters during growth in mixed aromatic substrate environments and to explain ecological implications for such physiologies and aromatic degradation pathways.

VI. REFERENCES

Alavi, M., Miller, T., Erlandson, K., Schneider, R., and Belas, R. (2001) Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. *Environmental Microbiology* **3**: 380-396.

Benner, R., Maccubbin, A.E., and Hodson, R.E. (1984) Preparation, characterization, and microbial degradation of specifically radiolabeled [^{14}C]lignocelluloses from marine and freshwater macrophytes. *Applied and Environmental Microbiology* **47**: 381-389.

Benner, R., Moran, M.A., and Hodson, R.E. (1985) Effects of pH and plant source on lignocellulose biodegradation rates in two wetland ecosystems, the Okefenokee Swamp and a Georgia salt marsh. *Limnology and Oceanography* **30**: 489-499.

Benner, R., Moran, M.A., and Hodson, R.E. (1986) Biogeochemical cycling of lignocellulosic carbon in marine and freshwater ecosystems: Relative contributions of prokaryotes and eukaryotes. *Limnology and Oceanography* **31**: 89-100.

Bergbauer, M., and Newell, S.Y. (1992) Contribution to lignocellulose degradation and DOC formation from a salt marsh macrophyte by the aseomycete *Phaeosphaeria spartinicola*. *FEMS Microbiology Ecology* **86**: 341-347.

Berger, M., Neumann, A., Schulz, S., Simon, M., and Brinkhoff, T. (2011) Tropodithietic acid production in *Phaeobacter gallaeciensis* is regulated by N-acyl homoserine lactone-mediated quorum sensing. *Journal of Bacteriology* **193**: 6576-6585.

Brzostowicz, P.C., Reams, A.B., Clark, T.J., and Neidle, E.L. (2003) Transcriptional cross-regulation of the catechol and protocatechuate branches of the beta-ketoadipate pathway contributes to carbon source-dependent expression of the *Acinetobacter* sp strain ADP1 *pobA* gene. *Applied and Environmental Microbiology* **69**: 1598-1606.

Buchan, A., Neidle, E.L., and Moran, M.A. (2001) Diversity of the ring-cleaving dioxygenase gene *pcaH* in a salt marsh bacterial community. *Applied and Environmental Microbiology* **67**: 5801-5809.

Buchan, A., Neidle, E.L., and Moran, M.A. (2004) Diverse organization of genes of the beta-ketoadipate pathway in members of the marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **70**: 1658-1668.

Buchan, A., González, J.M., and Moran, M.A. (2005) Overview of the marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **71**: 5665-5677.

Buchan, A., Collier, L.S., Neidle, E.L., and Moran, M.A. (2000) Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **66**: 4662-4672.

Bürgmann, H., Howard, E.C., Ye, W.Y., Sun, F., Sun, S.L., Napierala, S., and Moran, M.A. (2007) Transcriptional response of *Silicibacter pomeroyi* DSS-3 to dimethylsulfoniopropionate (DMSP). *Environmental Microbiology* **9**: 2742-2755.

Campbell, B.J., and Kirchman, D.L. (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME Journal* **7**: 210-220.

Cicirelli, E.M., Williamson, H., Tait, K.M., and Fuqua, C. (2008) Acylated homoserine lactone signaling in marine bacterial systems. In *Chemical Communication Among Bacteria*. Washington, D.C.: ASM Press, pp. 251-272.

Crawford, D.L., Pometto, A.L., and Crawford, R.L. (1983) Lignin degradation by *Streptomyces viridosporus*: Isolation and characterization of a new polymeric lignin degradation intermediate. *Applied and Environmental Microbiology* **45**: 898-904.

Cude, W.N., Mooney, J., Tavanaei, A.A., Hadden, M.K., Frank, A.M., Gulvik, C.A. et al. (2012) Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization in the marine roseobacter *Phaeobacter* sp. strain Y4I. *Applied and Environmental Microbiology* **78**: 4771–4780.

Dang, H.Y., and Lovell, C.R. (2000) Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis of 16S rRNA genes. *Applied and Environmental Microbiology* **66**: 467-475.

Dang, H.Y., and Lovell, C.R. (2002) Seasonal dynamics of particle-associated and free-living marine *Proteobacteria* in a salt marsh tidal creek as determined using fluorescence in situ hybridization. *Environmental Microbiology* **4**: 287-295.

Dang, H.Y., Li, T.G., Chen, M.N., and Huang, G.Q. (2008) Cross-Ocean distribution of *Rhodobacterales* bacteria as primary surface colonizers in temperate coastal marine waters. *Applied and Environmental Microbiology* **74**: 52-60.

Donoso, R.A., Pérez-Pantoja, D., and González, B. (2011) Strict and direct transcriptional repression of the *pobA* gene by benzoate avoids 4-hydroxybenzoate degradation in the pollutant degrader bacterium *Cupriavidus necator* JMP134. *Environmental Microbiology* **13**: 1590-1600.

Evans, W.C. (1947) Oxidation of phenol and benzoic acid by some soil bacteria. *Biochemical Journal* **41**: 373-382.

Fowler, G.J., Ardern, E., and Lockett, W.T. (1910) The oxidation of phenol by certain bacteria in pure culture. *Proceedings of the Royal Society of London Series B-Containing Papers of a Biological Character* **83**: 149-156.

Fuchs, G., Boll, M., and Heider, J. (2011) Microbial degradation of aromatic compounds - from one strategy to four. *Nature Reviews Microbiology* **9**: 803-816.

Gifford, S.M., Sharma, S., Rinta-Kanto, J.M., and Moran, M.A. (2011) Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME Journal* **5**: 461-472.

Gilbert, T.R., Kirss, R.V., and Davies, G. (2004) Chemical Bonding and Atmospheric Molecules. In *Chemistry: The Science in Context*. Drake-Johnson, V. (ed). New York, NY: W W Norton & Co Inc, p. 947.

Glenn, J.K., Morgan, M.A., Mayfield, M.B., Kuwahara, M., and Gold, M.H. (1983) An extracellular H_2O_2-requiring enzyme preparation involved in lignin biodegradation by the white rot basidiomycete *Phanerochaete chrysosporium*. *Biochemical and Biophysical Research Communications* **114**: 1077-1083.

González, J.M., and Moran, M.A. (1997) Numerical dominance of a group of marine bacteria in the alpha-subclass of the class *Proteobacteria* in coastal seawater. *Applied and Environmental Microbiology* **63**: 4237-4242.

González, J.M., Whitman, W.B., Hodson, R.E., and Moran, M.A. (1996) Identifying numerically abundant culturable bacteria from complex communities: An example from a lignin enrichment culture. *Applied and Environmental Microbiology* **62**: 4433-4440.

González, J.M., Mayer, F., Moran, M.A., Hodson, R.E., and Whitman, W.B. (1997) *Sagittula stellata* gen. nov, sp. nov, a lignin-transforming bacterium from a coastal environment. *International Journal of Systematic Bacteriology* **47**: 773-780.

González, J.M., Covert, J.S., Whitman, W.B., Henriksen, J.R., Mayer, F., Scharf, B. et al. (2003) *Silicibacter pomeroyi* sp nov and *Roseovarius nubinhibens* sp nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *International Journal of Systematic and Evolutionary Microbiology* **53**: 1261-1269.

Gosselink, J.G., and Kirby, C.J. (1974) Decomposition of salt marsh grass, *Spartina alterniflora* Loisel. *Limnology and Oceanography* **19**: 825-832.

Gram, L., Grossart, H.P., Schlingloff, A., and Kiorboe, T. (2002) Possible quorum sensing in marine snow bacteria: Production of acylated homoserine lactones by Roseobacter strains isolated from marine snow. *Applied and Environmental Microbiology* **68**: 4111-4116.

Happold, F.C., and Key, A. (1932) The bacterial purification of gasworks' liquors. The action of the liquors on the bacterial flora of sewage. *Journal of Hygiene* **32**: 573-580.

Hazelwood, L.A., Tai, S.L., Boer, V.M., de Winde, J.H., Pronk, J.T., and Daran, J.M. (2006) A new physiological role for Pdr12p in *Saccharomyces cerevisiae*: export of aromatic and

branched-chain organic acids produced in amino acid catabolism. *FEMS Yeast Research* **6**: 937-945.

Huber, S.G., Kilian, G., and Schöler, H.F. (2007) Carbon suboxide, a highly reactive intermediate from the abiotic degradation of aromatic compounds in soil. *Environmental Science & Technology* **41**: 7802-7806.

Inderjit, Streibig, J.C., and Olofsdotter, M. (2002) Joint action of phenolic acid mixtures and its significance in allelopathy research. *Physiologia Plantarum* **114**: 422-428.

Long, S.P., and Mason, C.F. (1983) *Saltmarsh Ecology (Tertiary Level Biology)*: Chapman and Hall.

Luo, H.W., Löytynoja, A., and Moran, M.A. (2012) Genome content of uncultivated marine Roseobacters in the surface ocean. *Environmental Microbiology* **14**: 41-51.

Maccubbin, A.E., and Hodson, R.E. (1980) Microbial degradation of detrital lignocelluloses by salt marsh sediment microflora. *Applied and Environmental Microbiology* **40**: 735-740.

Moran, M.A. (2012) Cloud services: Bacteria and sulfur cycling in the surface ocean. In *American Society for Microbiology 112th General Meeting*. San Francisco, California, USA.

Moran, M.A., and Hodson, R.E. (1990) Contributions of degrading *Spartina alterniflora* lignocellulose to the dissolved organic carbon pool of a salt marsh. *Marine Ecology Progress Series* **62**: 161-168.

Moran, M.A., and Hodson, R.E. (1994) Dissolved humic substance of vascular plant origin in a coastal marine environment. *Limnology and Oceanography* **39**: 762-771.

Moran, M.A., González, J.M., and Kiene, R.P. (2003) Linking a bacterial taxon to sulfur cycling in the sea: Studies of the marine *Roseobacter* group. *Geomicrobiology Journal* **20**: 375-388.

Moran, M.A., Belas, R., Schell, M.A., González, J.M., Sun, F., Sun, S. et al. (2007) Ecological genomics of marine roseobacters. *Applied and Environmental Microbiology* **73**: 4559-4569.

Mou, X.Z., Vila-Costa, M., Sun, S.L., Zhao, W.D., Sharma, S., and Moran, M.A. (2011) Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. *Environmental Microbiology Reports* **3**: 798-806.

Newton, R.J., Griffin, L.E., Bowles, K.M., Meile, C., Gifford, S., Givens, C.E. et al. (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME Journal* **4**: 784-798.

Odum, E.P. (1953) *Fundamentals of Ecology*. Philadelphia, PA: W. B. Saunders Company.

Odum, E.P., and Smalley, A.E. (1959) Comparison of population energy flow of a herbivorous and a deposit-feeding invertebrate in a salt marsh ecosystem. *Proc Natl Acad Sci* **45**: 617-622.

Poretsky, R.S., Sun, S.L., Mou, X.Z., and Moran, M.A. (2010) Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environmental Microbiology* **12**: 616-627.

Porsby, C.H., Nielsen, K.F., and Gram, L. (2008) *Phaeobacter* and *Ruegeria* species of the *Roseobacter* clade colonize separate niches in a Danish turbot (*Scophthalmus maximus*)-rearing farm and antagonize *Vibrio anguillarum* under different growth conditions. *Applied and Environmental Microbiology* **74**: 7356-7364.

Ramachandra, M., Crawford, D.L., and Hertel, G. (1988) Characterization of an extracellular lignin peroxidase of the lignocellulolytic actinomycete *Streptomyces viridosporus*. *Applied and Environmental Microbiology* **54**: 3057-3063.

Reisch, C.R., Stoudemayer, M.J., Varaljay, V.A., Amster, I.J., Moran, M.A., and Whitman, W.B. (2011) Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. *Nature* **473**: 208-+.

Rinta-Kanto, J.M., Bürgmann, H., Gifford, S.M., Sun, S.L., Sharma, S., del Valle, D.A. et al. (2011) Analysis of sulfur-related transcription by Roseobacter communities using a taxon-specific functional gene microarray. *Environmental Microbiology* **13**: 453-467.

Seyedsayamdost, M.R., Case, R.J., Kolter, R., and Clardy, J. (2011) The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nature Chemistry* **3**: 331-335.

Shiba, T. (1991) *Roseobacter litoralis* gen. nov., sp. nov., and *Roseobacter denitrificans* sp. nov., Aerobic Pink-Pigmented Bacteria which Contain Bacteriochlorophyll *a*. *Systematic and Applied Microbiology* **14**: 140-145.

Simo, R., Archer, S.D., Pedros-Alio, C., Gilpin, L., and Stelfox-Widdicombe, C.E. (2002) Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. *Limnology and Oceanography* **47**: 53-61.

Suzuki, M.T., Preston, C.M., Chavez, F.P., and DeLong, E.F. (2001) Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay. *Aquatic Microbial Ecology* **24**: 117-127.

Tang, K., Huang, H.Z., Jiao, N.Z., and Wu, C.H. (2010) Phylogenomic analysis of marine roseobacters. *PLoS One* **5**: e11604.

Tien, M., and Kirk, T.K. (1983) Lignin-degrading enzyme from the Hymenomycete *Phanerochaete chrysosporium* Burds. *Science* **221**: 661-662.

Turner, J.A., and Rice, E.L. (1975) Microbial decomposition of ferulic acid in soil. *Journal of Chemical Ecology* **1**: 41-58.

Vila-Costa, M., Rinta-Kanto, J.M., Sun, S.L., Sharma, S., Poretsky, R., and Moran, M.A. (2010) Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. *ISME Journal* **4**: 1410-1420.

Wagner-Döbler, I., Thiel, V., Eberl, L., Allgaier, M., Bodor, A., Meyer, S. et al. (2005) Discovery of complex mixtures of novel long-chain quorum sensing signals in free-living and host-associated marine alphaproteobacteria. *Chembiochem* **6**: 2195-2206.

Wang, Y., Yang, J., Lee, O.O., Dash, S., Lau, S.C.K., Al-Suwailem, A. et al. (2011) Hydrothermally generated aromatic compounds are consumed by bacteria colonizing in Atlantis II Deep of the Red Sea. *ISME Journal* **5**: 1652-1659.

APPENDIX

VII. FIGURES

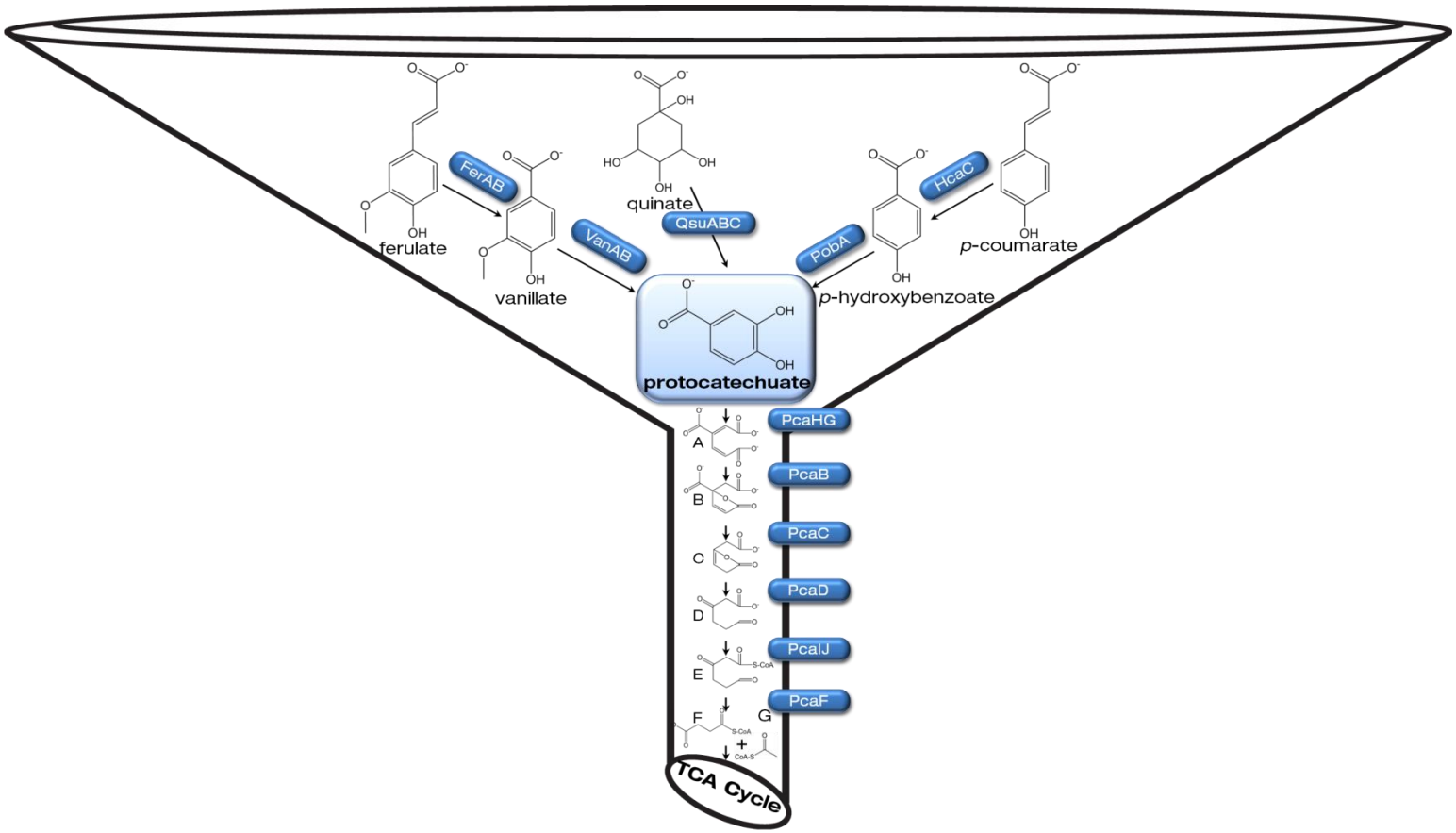


Figure 1.1. Catabolic funneling of aromatic compounds proceeding through the *pca* pathway in *Sagittula stellata* E-37. Ovals represent functional enzymes. A, β-carboxy-cis-cis-muconate; B, γ-carboxymuconolactone; C, β-ketoadipate enol-lactone; D, β-ketoadipate, E, β-ketoadipyl-CoA; F, succinyl-CoA; G, acetyl-CoA. TCA, tricarboxylic acid.

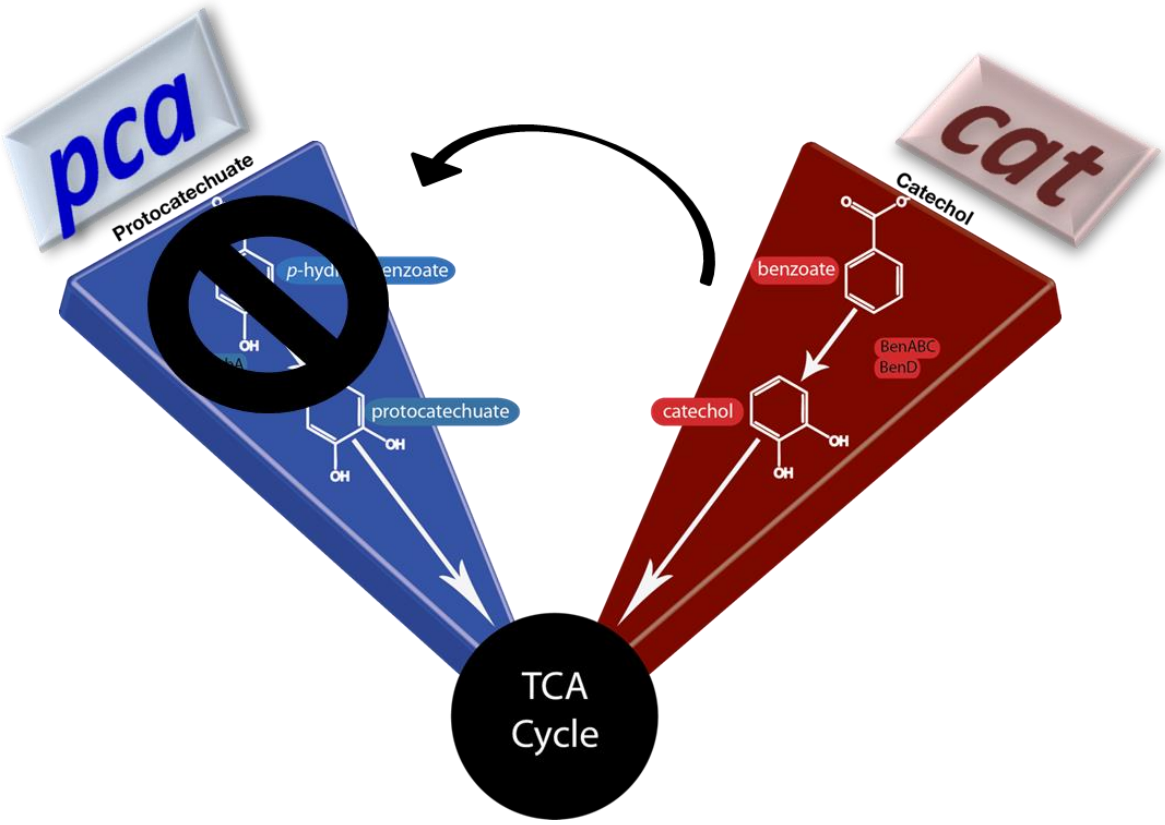


Figure 1.2. Cross-regulation between ring-cleaving pathways. The catechol branch (*cat*) represses the activity of the protocatechuate branch (*pca*). Diagnostic enzymes in the pathways area shown in ovals. TCA, tricarboxylic acid.

Figure 1.3. *Spartina alterniflora* (cordgrass) dominating the coastal landscape of a salt marsh along Charleston, South Carolina. Standing cordgrass thriving in the background and a patch in the foreground are shown as well as an area containing fallen dead cordgrass, part of which is wet (likely harboring active microbial decomposers). This photo was taken in October 12, 2012.

CHAPTER 2 -

THE AEROBIC BENZOYL-COA (*BOX*) PATHWAY IS ABUNDANT AND APPEARS TAXONOMICALLY CONSTRAINED IN BIOGEOGRAPHICALLY DISPERSED ENVIRONMENTAL SAMPLES

I. ABSTRACT

Aromatic compounds are abundant in terrestrial and marine systems. As such, microbial remineralization of these compounds is an important component of the global carbon cycle. The benzoyl-CoA oxidation (*box*) pathway is a recently described aerobic route for benzoate catabolism that has been experimentally identified in the soil betaproteobacterium *Azoarcus evansii* KB740 and the coastal marine alphaproteobacterium *Sagitulla stellata* E-37, a member of the abundant *Roseobacter* lineage. However, the prevalence and genetic diversity of *box* in the environment are unknown. To address this gap in our knowledge, we have developed a degenerate qPCR primer set targeting the diagnostic epoxidase-encoding *boxB* gene and applied it to natural samples collected from a Norwegian fjord. Using this gene assay, abundances of *boxB* were found to range from 0.33-7.2% (expressed as bacterial 16S rRNA gene copies per ng DNA) in the fjord. Sequence analysis of clone library representatives suggests BoxB is strongly conserved at the amino acid level (74% identity). The majority of the clones (20/29) grouped within sequences derived from *Alphaproteobacteria* and 31% grouped within those from *Betaproteobacteria*. Because no other clones grouped outside of these two classes, the *box* pathway may be taxonomically constrained. In total, 69% of the clones grouped strongly within sequences derived from *Roseobacter* clade members. These data suggest roseobacters are the most abundant taxon in the fjord possessing this pathway. We also probed publicly available metagenomes datasets representing geographically distinct environments with the degenerate primer set using De-MetaST-BLAST. Based on the aquatic and terrestrial metagenomes studied, the *box* pathway appears taxonomically constrained and more abundant than other aerobic benzoate-degrading pathways, suggesting it is an environmentally relevant pathway in natural systems.

II. INTRODUCTION

Benzoate is an aromatic compound present in both aquatic and terrestrial systems that originates from abiotic, anthropogenic, or biotic processes (Wakeham et al., 1980; Higuchi, 1990; Wang et al., 2011). There are several different strategies microorganisms use to degrade benzoate. The anaerobic benzoyl-CoA pathway involves thioesterification which delocalizes electrons from the aromatic ring, and ring cleavage is mediated by several reduction steps (Harwood and Gibson, 1997; Heider and Fuchs, 1997; Breese et al., 1998; Ismail and Gescher, 2012). Some aerobic routes for benzoate catabolism require a ring hydroxylating dioxygenase system that generates catechol, a dihydroxylated intermediate, susceptible to either *ortho*- or *meta*-ring cleavage. *Ortho*-cleavage is mediated by enzymes of the catechol (*cat*) branch of the β-ketoadipate pathway (Ornston, 1966). This pathway is well studied and has a broad taxonomic distribution, extending from several bacterial phyla to some fungi (Harwood and Parales, 1996). The *meta*-cleaving catechol route (*xyl*) can proceed through either dehydrogenative or hydrolytic branches and is present in two bacterial phyla (*Actinobacteria* and *Proteobacteria*) (Harayama et al., 1987; Harayama and Rekik, 1990). Additionally, abundance measurements of *cat* (Junca and Pieper, 2004) and *xyl* (Brennerova et al., 2009) indicate correlation with benzoate presence. More recently, the aerobic benzoyl-CoA (*box*) pathway has been identified. The *box* pathway invokes a novel epoxidation mechanism of aerobic ring cleavage that acts on a CoA-thiolated benzoate (generated earlier in the *box* pathway) (Rather et al., 2010). Elucidation of the biochemistry of the *box* pathway challenged the existing paradigm of aerobic aromatic compound catabolism in which preparation for ring cleavage was considered hydroxylation dependent (Rather et al., 2010; Fuchs et al., 2011).

While *box* homologs have been identified in the genomes of several cultivated actinobacteria and proteobacteria (Rather et al., 2010), empirical demonstration of benzoate catabolism via this pathway is limited. To date, *box* activity has only been experimentally documented in one marine alphaproteobacterium (the marine roseobacter isolate *Sagittula stellata* E-37 (Chapter 4) and two terrestrial betaproteobacteria (the terrestrial *Azoarcus evansii* KB740 (Mohamed et al., 2001; Gescher et al., 2002; Ismail, 2008; Rather et al., 2010) and *Azoarcus* sp. CIB (Durante-Rodríguez et al., 2010; Andres Valderrama et al., 2012)). The taxonomic distribution, genetic diversity, and abundance of the *box* pathway in the environment were mysteries we aimed to address in this study. We sought to develop and apply a *boxB* degenerate primer set to environmental samples and to analyze existing metagenomes for *box* pathway homologs.

III. MATERIALS AND METHODS

boxB *Primer Set Development and Validation*

The BoxB sequence (ZP_01745220.1) of the roseobacter isolate *Sagittula stellata* E-37, which uses the aerobic benzoyl-CoA (*box*) pathway to catabolize benzoate, was used to identify other BoxB sequences. In two separate BLASTp queries using default algorithm parameters in v2.2.25+, hits in both the non-redundant (nr) protein sequences and metagenomic (env_nr) proteins databases in NCBI were retrieved. Sequences with e-values ≤2e-85 and <1e-40 to ZP_01745217.1 were considered putative BoxB sequences from the nr and env_nr databases, respectively. To develop conservative sets of degenerate primers, further trimming, removal of distantly related BoxB to ZP_01745217.1, and creation of new phylogenetic trees were performed. Protein sequences were aligned with Clustal Omega 1.0.3 (Sievers et al., 2011). Conserved amino acid residues were manually identified in Clustal X 2.1 (Larkin et al., 2007),

and all ≥18 bp sequences with <1,025 nucleotide degeneracies were considered as possible primers.

Using the De-MetaST-BLAST software package (Chapter 3), metagenome databases available from CAMERA (http://camera.calit2.net/) were queried with developed *boxB* primer sets in Table 2.1 to determine if degenerate primer set(s) designed retrieve sequences that are not *boxB.* The nucleotide hits retrieved with the *boxB* primer sets using De-MetaST had each function predicted by performing in-house BLASTx searches. Each nucleotide sequence was translated in all three frames in both the sense and antisense strand, so each of the six peptides was queried against NCBI's nr protein database. If at least one of the six peptides had one of its top three hits match to the predicted BoxB function, the sequence was considered *boxB.*

Creation of qPCR Standards

To determine the abundance of *boxB* using qPCR, a clone was generated from *boxB* in *Sagitulla stellata* E-37 which was previously documented to use the *box* pathway (Chapter 4). E-37 was grown to late exponential growth phase in 10 ml of YTSS (containing 2.5% yeast extract, 4% tryptone, and 1.5% sea salts [Sigma]), and genomic DNA was extracted using the silica column-based DNeasy® Blood & Tissue Kit [Qiagen]. In a 25 µl PCR, 0.5 µl of each 100 µM primer (*boxB*171F and *boxB*265R) was used to amplify a 300 bp region of *boxB* (hereafter referred to as "*boxB*300") with 1.25 µl containing 20 ng of E-37 genomic DNA μl^{-1} (~8×10^{10} genomes) using 0.5 µl of 10 µM dNTPs, 5 µl of 5X GoTaq Buffer, and 0.125 µl of 5 U μl^{-1} GoTaq® polymerase [Promega]. The PCR involved an initial 95°C denaturation for 15 min, followed by 25 cycles of 95°C denaturation for 45 s, 56.5°C annealing for 45 s, and 72°C elongation for 30 s. The

QIAquick® PCR Cleanup Kit [Qiagen] was performed immediately following PCR amplification to remove bulky genomic DNA, unused primers and dNTPs, and the GoTaq polymerase. To maintain the $^{3'}$A-tails on the *boxB*300 amplicons, ligation immediately proceeded the PCR cleanup using the pGEM®-T Vector System I Kit [Promega]. For the 10 µl ligation reaction, the following reagents were used: 3 µl of purified *boxB*300 amplicons, 1 µl of 50 ng µl^{-1} pGEM®-T Vector, 5 µl of 2X Rapid Ligation Buffer, and 1 µl of 3 U µl^{-1} T4 ligase. The ligation reaction was first incubated at room temperature for 1 h followed by incubation at 4°C for 16 h. A volume of 5 µl of ligated plasmids was gently dispensed into 100 µl of chemically competent *Escherichia coli* DH5α cells thawed on ice. After incubating on ice for 30 min, the cells were heat-shocked at 42°C for 1.5 min, and the clones were placed on ice for 2 min. The clones were given 750 µl of S.O.C. medium and incubated horizontally at 37°C with 250 rpm shaking for 1 h. Aliquots of cells were spread onto 37°C pre-warmed LB/Agar (1.5% w/v)/Amp (100 µg ml^{-1})/IPTG (500 µM)/X-gal (80 µg ml^{-1}) plates and grown at 37°C for 16 h. A white-colored putative clone was picked, transferred into 10 ml LB/Amp (100 µg ml^{-1}) test tubes, and propagated at 37°C for 16 h. Plasmids were extracted using the QIAprep® Miniprep Kit [Qiagen]. A clone containing *boxB*300 was verified by sequencing using the M13F primer ($^{5'}$GAGCGGATAACAATTTCACACAGG$^{3'}$) upstream of the insert site using an ABI PRISM 3730 Sanger sequencer. Isolated pGEM®-T/*boxB*300 plasmids were quantified with a ND-1000 spectrophotometer [NanoDrop Technologies], and serial dilutions ranging from 10^1 to 10^9 copies µl^{-1} were performed using 10 mM Tris-Cl, pH 8 for qPCR standards.

The plasmid standards (pMH10) used for 16S rRNA gene (rDNA) quantitation were generated in a previous report (Buchan et al., 2009) which contains the 27F

($^{5'}$AGAGTTTGATCMTGGCTCAG$^{3'}$) and 1522R ($^{5'}$AAGGAGGTGATCCANCCRCA$^{3'}$) region of the 16S rDNA from the bacterium *Phaeobacter* sp. Y3F.

DNA Extraction of Filtered Cells from a Coastal Norwegian Fjord

During June of 2008, water samples were collected at the Espeland Marine Biological Field Station in the Raunefjorden (60°22'N, 5°28'E), located off the coast of Norway. Mesocosm (11 m^3) treatments of P replete (N:P of 15:1) and P deplete (N:P of 75:1) were previously described by others to invoke algal blooms, which resulted in varied microbial community dynamics (Pagarete et al., 2009; Sorensen et al., 2009; Joassin et al., 2011; Pagarete et al., 2011). Open (unmanipulated) fjord water was also harvested throughout the mesocosm as a control. Deep 16S rDNA amplicon sequencing with 454 FLX was previously performed on these same collected samples, which have been taxonomically classified down to the genus level (Budinoff, 2011). During field sampling, two liters of water were filtered to concentrate cells onto 0.2 µm filters and flash frozen in $N_{2\ (l)}$. Cells were lysed with bead beating and DNA was extracted using the DNeasy® Blood & Tissue Kit [Qiagen]. The identical -70°C stored aliquots that had been used for pyrosequencing by Budinoff (2011) were also used for qPCR abundance measurements in this study.

qPCR of Bacterial 16S rDNA in a Coastal Norwegian Fjord

To relate concentration of *boxB* to bacterial abundance, the total copies of bacterial-specific 16S rDNA were determined. A previous report described optimized final concentrations for the Bact1369-for ($^{5'}$CGGTGAATACGTTCYCGG$^{3'}$) and Prok1492-rev ($^{5'}$GGWTACCTTGTTACGACTT$^{3'}$) qPCR primer set being 1.5 mM and 1.0 mM, respectively

(Buchan et al., 2009). In 25 µl reactions, 10 µl of DNA sample (or standard), 12.5 µl of 2X SYBR® *Premix Ex Taq*™ (Perfect Real Time) [Takara], and 1.25 µl of each primer were pooled. Thermocycling conditions for 16S rDNA qPCR were performed on an Opticon 2 [Bio-Rad] and involved an initial denaturation for 15 min at 95°C followed by 35 cycles of 95°C for 45 s, 56°C for 45 s, 72°C for 15 s, and 83°C for 5 s. Fluorescence for quantification was detected after the 83°C heating to circumvent quantitation of shorter amplicons (e.g., primer dimers) other than 16S rDNA. After amplification, a final 5 min extension at 72°C was performed. The ~140 bp qPCR amplicons were melted in 1°C increments from 50°C to 100°C with fluorescence read after a 1 s hold at each temperature. Using the Beer-Lambert-Bouguer law, the fjord samples were fit into a 5-point standard curve ranging from 10^4 to 10^8 copies of 16S rDNA μl^{-1}.

qPCR of boxB *in a Coastal Norwegian Fjord*

The pGEM®-T vector containing *boxB*300 from E-37 was used at 2.5×10^5 copies $reaction^{-1}$ to determine the optimal concentration of both the forward and reverse primer. A matrix of varied concentrations for each primer (0.5-10 µM) was used in 25 µl reactions containing 12.5 µl of 2X SYBR® *Premix Ex Taq*™ (Perfect Real Time) [Takara]. The samples were thermocycled in an Opticon 2 [Bio-Rad] with an initial 2 min at 95°C, followed by 40 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 56.5°C, and elongation for 30 s at 72°C. To ensure only a single product was amplified in each reaction, a melting curve of the ~300 bp amplicons was performed by measuring fluorescence after a 1 s hold for every 1°C increase from 50°C to 100°C.

A 6-point dilution series of pGEM®-T/*boxB*300 (from 10^1 to 10^6 copies μl^{-1}) was used following the Beer-Lambert-Bouguer law for quantifying *boxB* in the samples collected from the

Norwegian fjord. In each 25 µl qPCR, the following reagents were added: 10 µl of DNA sample (or standard), 12.5 µl of 2X SYBR® *Premix Ex Taq*™ (Perfect Real Time) [Takara], and 1.25 µl of both forward (*boxB*171F [$^{5'}$CARGGNGAYACNGARCCN$^{3'}$]) and reverse (*boxB*265R [$^{5'}$ACNGAYMGNGAYGGNAAR$^{3'}$]) degenerate primers at 20 µM. The same thermocycling conditions were performed as described in the primer optimization of *boxB*300, and again melting curves were used to evaluate primer specificity. The qPCR amplicons were then chilled to 3°C, and cloning of the amplicons was immediately performed.

Creation of boxB *Clone Libraries*

The qPCR products of *boxB*300 were cleaned up with the PCR Cleanup Kit (described previously) immediately following the completion of the melting curves. Cloning was again performed as described earlier using the T4 ligase, pGEM®-T, and *E. coli* DH5α. All clones were plated, propagated, and preserved. Randomly selected clones (n=29) from the Fjord Day 2 sample were chosen for plasmid purification using the QIAprep® Miniprep Kit (previously described) and sequencing. With short (~300 bp inserts), the Sanger sequencing reads extend into part of the vector, which served as an internal sequencing control for each insert because the vector sequence is known.

boxB *Phylogenetic Analyses*

In order to determine the taxonomic identify of *boxB* sequences retrieved from the environment, phylogenetic construction was performed with *boxB* sequences with known taxonomy. The full-length nucleotide sequences from *boxB* genes retrieved from NCBI's genomes using De-MetaST were combined with Norwegian fjord *boxB*300 clone sequences as well as *boxB*300 *in silico*

amplicons retrieved from all metagenomes available from CAMERA (Sun et al., 2011). The sequences were aligned with Clustal W2.1 (Larkin et al., 2007), and the best substitution model was calculated by jModelTest v2.1 where 203 substitution schemes were considered (1624 total models) (Posada, 2008). Using the University of Tennessee's Newton high-performance computing cluster running RAxML-MPI v7.2.6 (Stamatakis, 2006), 1000 maximum-likelihood phylogenetic trees were constructed, and 50000 bootstrap iterations were performed to provide a confidence to the best ML tree topology. After bipartitions were performed, Dendroscope v3.2.1 (Huson et al., 2007) was used to create a graphical tree image, which was then imported into ARB v5.3 (Ludwig et al., 2004) to visualize the relationship of each environmental sequence to *boxB* sequences of cultured strains with known taxonomy. Sequence identities were manually tagged according to either taxonomic group (*e.g.*, *Betaproteobacteria*, *Roseobacter*) or metagenome sampling site (*e.g.*, Antarctica, Farm Soil, HOT) and loaded into UNIFRAC along with the ML bootstrapped tree. A PCoA was performed with 1,000 replicates in UNIFRAC to visualize group relatedness (Lozupone and Knight, 2005).

IV. RESULTS

boxB Primer Set as a Molecular Probing Tool

There were several conserved sites in the BoxB amino acid alignment with varying degrees of degeneracy for each candidate primer sequence (Table 2.1). To determine if any primer set would retrieve sequences other than *boxB*, each set was queried against a soil metagenome (Waseca Farm Soil Site) and three marine metagenomes (Botany Bay, GOS, and HOT) where benzoate is expected to be present as a lignocellulosic plant degradation product (Klap et al., 1998). Many primer pairs were considered out of the seven conserved regions identified (Table

2.1), but the primer pair *boxB*171F and *boxB*265R was favored because it yielded a relatively large amplicon to investigate sequence diversity while also not excluding certain *box*-containing taxa. Additionally, De-MetaST-BLAST queries of the four CAMERA metagenomes with the *boxB*171F and *boxB*265R primer pair revealed all *in silico* amplicons were predicted to be *boxB* homologs (data not shown).

boxB is *Abundant in the Fjord*

Since there are no quantitative estimates on the abundance of *box* pathway in natural environments, we employed the qPCR primer set to a collection of samples from a coastal marine environment (Raunefjorden, Norway). Relative to the total number of bacterial 16S rDNA copies, a range of 0.33 to 7.2% *boxB*300 was observed in the unmanipulated Fjord samples (Figure 2.1). This wide range observed for *boxB* indicates members capable of degrading benzoate through the *box* pathway can fluctuate by nearly 22 fold. Regardless of the treatment type (being P deplete, P replete, or unmanipulated), the maximum frequency of *boxB* per bacterial 16S rDNA occured around 10% at day 8. In all three treatments, the same temporal abundance pattern emerged—an increase from day 5 to peaking at day 8, then a successive decline in days 11 and 14 (Figure 2.1).

box *is Taxonomically Constrained in Natural Bacterial Populations Throughout the World*

Initial phylogenetic tree topologies of *boxB* from cultured taxa alone indicated groupings according to phylogeny (data not shown). Once *boxB*300 sequences from the fjord and other probed metagenomes were combined with the *boxB* from cultured taxa and aligned, the GTR+GAMMA+I substitution model was identified as the best substitution model using

jModelTest (Posada, 2008). The substitution model was then used in generating a maximum likelihood phylogenetic tree to classify individual *boxB*300 sequences into seven taxonomic categories that emerged: *Actinobacteria*, *Alphaproteobacteria*, *Roseobacter*, *Betaproteobacteria*, *Deltaproteobacteria*, *Gammaproteobacteria*, and unclassified (Figure 2.2). Figure 2.3 indicates clones from the Norwegian fjord primarily (69%, 20/29) group within *Roseobacter* and only 31% (9/29) grouped within *Betaproteobacteria*. Nonetheless, 32% of sequences from all available CAMERA metagenomes clustered within the *Roseobacter* clade and 53% grouped with *Betaproteobacteria*; these values change to 37% and 50% (respectively) when sequences from the Norwegian fjord are also considered (Figure 2.2).

V. DISCUSSION

Only three organisms have been shown to use the *box* pathway for catabolism of benzoate, one of which is an alphaproteobacterium *Sagittula stellata* E-37 (Chapter 4) and the other two are betaproteobacteria: *Azoarcus evansii* KB740 and *Azoarcus* sp. CIB (Rather et al., 2010; Valderrama et al., 2012). Based on homology searches of the NCBI Genomes database, it was predicted some Gram-positive actinobacteria contain *box* (Rather et al., 2010). To date, no organism has been shown to contain more than one copy of *boxB* in the NCBI Genomes database, and BoxB shares very low homology to other known proteins (Rather et al., 2011; Ismail and Gescher, 2012), which make *boxB* an attractive genetic target. During the development of a *boxB* primer set, deltaproteobacteria and gammaproteobacteria members also were found to contain a *boxB* homolog. Further support for these taxa containing the full *box* pathway was obtained by observing other *box* genes adjacently located to the putative *boxB*. In this study, we sought to estimate the taxonomic breadth of *box* via phylogenetics.

Using the topology of a maximum-likelihood tree created by RAxML, most (20/29) clones from the fjord group with other *boxB* from roseobacters. This suggests the majority (69%) of bacteria in the fjord containing the *box* pathway are roseobacters. Because most *boxB*-containing bacteria group according to their phylogeny, it seems horizontal gene transfer of *boxB* is not pervasive despite five *boxB*300 identified in a phage metagenome (Figure 2.2). Abundances of *Roseobacter* and *Betaproteobacteria* were relatively high in the fjord samples (Budinoff 2011). Other lineages such as *Flavobacteria* were also abundant, yet we have not been able to identify a flavobacterium containing the *box* pathway. Collectively, 87% of the *boxB*300 sequences retrieved from various metagenomes are either *Roseobacter*-related or *Betaproteobacteria*-related, which suggests the *box* pathway is rather taxonomically constrained in these environments despite neither of these two lineages being the most abundant in the communities (Figure 2.2).

Using a qPCR-based approach, *boxB*300 amplicons were used as a proxy for the abundance of the entire *box* pathway. Abundance measurements of the unmanipulated fjord indicated a large range of 0.33 to 7.2% *boxB*300 per bacterial 16S rDNA (Figure 2.1). The *boxB*300 degenerate primer set was based off *boxB*-containing taxa (totaling 67 organisms) from different phyla, but nonetheless it is possible the primer set did not capture all *boxB* for qPCR measurements. With each bacterium minimally containing one copy of 16S rDNA and others containing more than one copy, these abundance values are likely represent an underestimation.

The bacterioplankton communities throughout the time series in the P deplete, P replete, and fjord control mesocosms changed in composition over the 14 d sampled (Budinoff, 2011). Days 5, 8 and 11 had approximately the same *box* frequency values irrespective of the treatment influenced, suggesting taxa which remained unchanged in these days but varied at the other days (2 and 14) are perhaps those which harbor the *box* pathway in the fjord. No clear trend of a taxon correlates with the *boxB*300 frequencies measured, so it is likely the compositional changes of multiple phylotypes containing *box* explains the temporal *box* frequency fluctuations observed (Figure 2.1).

Overall, the abundance of *boxB* (serving as a proxy for the *box* pathway) is relatively high compared to the previously estimated 2-5% range (Rather et al., 2010; Ismail and Gescher, 2012). The *meta*-cleaving benzoate degradation pathway that proceeds through a catechol pathway (*xyl*) exists in <3% of the bacterial population in an uncontaminated soil system (Junca and Pieper, 2004), which is on the lower side of what we have observed for the *box* abundance in the fjord. A different aerobic ring-cleaving pathway (*pca*), known to be active in the degradation of several aromatics other than benzoate, occurs in 0.2-10.9% of the bacterial population in soil systems (El Azhari et al., 2008), which is a similar abundance range we observed in the fjord for *box*. In both cases, the *xyl* and *pca* pathways are involved in the degradation of many different aromatic compounds (*e.g*., (Stanier et al., 1966; Haigler et al., 1992; Grund and Kutzner, 1998; Wojcieszyńska et al., 2011)), however benzoate is the only known compound to proceed through *box*.

It is well established the anaerobic benzoyl-CoA pathway serves as a major pathway to degrade a variety of aromatic compounds into the central aromatic intermediate benzoyl-CoA (Heider and Fuchs, 1997; Carmona et al., 2009). This concept of many peripheral conversion steps of aromatics formed into a central aromatic compound, termed "catabolic funneling," might also be occurring in aerobic systems with taxa identified in Figure 2.2. Many of the oxygen insensitive peripheral pathways used to convert aromatics into benzoyl-CoA by anaerobes might also be exploited by aerobes to funnel them into the epoxide-cleaving *box* pathway in oxygenated environments. It remains to be determined whether any other aromatic compounds proceed through the *box* pathway in a similar fashion as the anaerobic benzoyl-CoA peripheral routes. The high abundance measurements of *box* correspond to similar frequencies of the well-documented *pca* catabolic funnel for aerobic aromatic catabolism (El Azhari et al., 2008), which stimulates the question of whether *box* might also serve as a central catabolic for processing additional aromatic compounds in addition to benzoate.

VI. ACKNOWLEDGEMENTS

Research with Reantha Pillay was made possible by The University of Tennessee Pre-Collegiate Research Scholars Program. C.A.G. and A.B. were supported as part of the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Award Number DE-SC0000997. C.A.G also acknowledges the 2013 American Society for Microbiology (ASM) Scientific Writing and Publishing Institute (SWPI), especially Professor Steven R. Blanke of the University of Illinois and Dr. Beronda L. Montgomery of Michigan State University, for constructive assistance in writing this chapter.

VII. REFERENCES

Andres Valderrama, J., Durante-Rodríguez, G., Blázquez, B., Luis García, J., Carmona, M., and Díaz, E. (2012) Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. *Journal of Biological Chemistry* 287: 10494-10508.

Breese, K., Boll, M., Alt-Morbe, J., Schagger, H., and Fuchs, G. (1998) Genes coding for the benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium *Thauera aromatica*. *European Journal of Biochemistry* 256: 148-154.

Brennerova, M.V., Josefiova, J., Brenner, V., Pieper, D.H., and Junca, H. (2009) Metagenomics reveals diversity and abundance of *meta*-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environmental Microbiology* 11: 2216-2227.

Buchan, A., Hadden, M., and Suzuki, M.T. (2009) Development and application of quantitative-PCR tools for subgroups of the *Roseobacter* clade. *Applied and Environmental Microbiology* 75: 7542-7547.

Budinoff, C.R. (2011) Diversity and activity of roseobacters and roseophage. In *Department of Microbiology*. Knoxville, TN: The University of Tennessee, p. 187.

Carmona, M., Zamarro, M.T., Blázquez, B., Durante-Rodríguez, G., Juárez, J.F., Valderrama, J.A. et al. (2009) Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *Microbiology and Molecular Biology Reviews* 73: 71-113.

Durante-Rodríguez, G., Valderrama, J.A., Mancheno, J.M., Rivas, G., Alfonso, C., Arias-Palomo, E. et al. (2010) Biochemical characterization of the transcriptional regulator BzdR from *Azoarcus* sp. CIB. *Journal of Biological Chemistry* 285: 35694-35705.

El Azhari, N., Bru, D., Sarr, A., and Martin-Laurent, F. (2008) Estimation of the density of the protocatechuate-degrading bacterial community in soil by real-time PCR. *European Journal of Soil Science* 59: 665-673.

Fuchs, G., Boll, M., and Heider, J. (2011) Microbial degradation of aromatic compounds - from one strategy to four. *Nature Reviews Microbiology* 9: 803-816.

Gescher, J., Zaar, A., Mohamed, M., Schagger, H., and Fuchs, G. (2002) Genes coding for a new pathway of aerobic benzoate metabolism in *Azoarcus evansii*. *Journal of Bacteriology* 184: 6301-6315.

Grund, E., and Kutzner, H.J. (1998) Utilization of quinate and *p*-hydroxybenzoate by actinomycetes: key enzymes and taxonomic relevance. *Journal of Basic Microbiology* 38: 241-255.

Haigler, B.E., Pettigrew, C.A., and Spain, J.C. (1992) Biodegradation of mixtures of substituted benzenes by *Pseudomonas* sp. strain JS150. *Applied and Environmental Microbiology* 58: 2237-2244.

Harayama, S., and Rekik, M. (1990) The *meta* cleavage operon of TOL degradative plasmid pWW0 comprises 13 genes. *Molecular & General Genetics* 221: 113-120.

Harayama, S., Mermod, N., Rekik, M., Lehrbach, P.R., and Timmis, K.N. (1987) Roles of the divergent branches of the *meta*-cleavage pathway in the degradation of benzoate and substituted benzoates. *Journal of Bacteriology* 169: 558-564.

Harwood, C.S., and Parales, R.E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annual Review of Microbiology* 50: 553-590.

Harwood, C.S., and Gibson, J. (1997) Shedding light on anaerobic benzene ring degradation: A process unique to prokaryotes? *Journal of Bacteriology* 179: 301-309.

Heider, J., and Fuchs, G. (1997) Anaerobic metabolism of aromatic compounds. *European Journal of Biochemistry* 243: 577-596.

Higuchi, T. (1990) Lignin biochemistry: Biosynthesis and biodegradation. *Wood Science and Technology* 24: 23-63.

Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8.

Ismail, W. (2008) Benzoyl-coenzyme A thioesterase of *Azoarcus evansii*: properties and function. *Archives of Microbiology* 190: 451-460.

Ismail, W., and Gescher, J. (2012) The epoxy coenzyme-A thioesters pathways for degradation of aromatic compounds. *Applied and Environmental Microbiology* 78: 5043-5051.

Joassin, P., Delille, B., Soetaert, K., Harlay, J., Borges, A.V., Chou, L. et al. (2011) Carbon and nitrogen flows during a bloom of the coccolithophore *Emiliania huxleyi*: Modelling a mesocosm experiment. *Journal of Marine Systems* 85: 71-85.

Junca, H., and Pieper, D.H. (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. *Environmental Microbiology* 6: 95-110.

Klap, V.A., Boon, J.J., Hemminga, M.A., and Van Soelen, J. (1998) Chemical characterization of lignin preparations of fresh and decomposing *Spartina anglica* by pyrolysis mass spectrometry. *Organic Geochemistry* 28: 707-727.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.

Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* 71: 8228-8235.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* 32: 1363-1371.

Mohamed, M.E., Zaar, A., Ebenau-Jehle, C., and Fuchs, G. (2001) Reinvestigation of a new type of aerobic benzoate metabolism in the proteobacterium *Azoarcus evansii*. *Journal of Bacteriology* 183: 1899-1908.

Ornston, L.N. (1966) Conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida* .III. Enzymes of the catechol pathway. *Journal of Biological Chemistry* 241: 3795-3799.

Pagarete, A., Allen, M.J., Wilson, W.H., Kimmance, S.A., and de Vargas, C. (2009) Host-virus shift of the sphingolipid pathway along an *Emiliania huxleyi* bloom: survival of the fattest. *Environmental Microbiology* 11: 2840-2848.

Pagarete, A., Le Corguille, G., Tiwari, B., Ogata, H., de Vargas, C., Wilson, W.H., and Allen, M.J. (2011) Unveiling the transcriptional features associated with coccolithovirus infection of natural *Emiliania huxleyi* blooms. *FEMS Microbiology Ecology* 78: 555-564.

Posada, D. (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253-1256.

Rather, L.J., Knapp, B., Haehnel, W., and Fuchs, G. (2010) Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. *Journal of Biological Chemistry* 285: 20615-20624.

Rather, L.J., Weinert, T., Demmer, U., Bill, E., Ismail, W., Fuchs, G., and Ermler, U. (2011) Structure and mechanism of the diiron benzoyl-coenzyme A epoxidase BoxB. *Journal of Biological Chemistry* 286: 29241-29248.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7.

Sorensen, G., Baker, A.C., Hall, M.J., Munn, C.B., and Schroeder, D.C. (2009) Novel virus dynamics in an *Emiliania huxleyi* bloom. *Journal of Plankton Research* 31: 787-791.

Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.

Stanier, R.Y., Palleron, N.J., and Doudorof, M. (1966) The aerobic pseudomonads: a taxonomic study. *Journal of General Microbiology* 43: 159-271.

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S. et al. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Research* 39: D546-D551.

Valderrama, J.A., Durante-Rodríguez, G., Blázquez, B., García, J.L., Carmona, M., and Díaz, E. (2012) Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. *Journal of Biological Chemistry* 287: 10494-10508.

Wakeham, S.G., Schaffner, C., and Giger, W. (1980) Polycyclic aromatic hydrocarbons in Recent lake sediments-I. Compounds having anthropogenic origins. *Geochimica Et Cosmochimica Acta* 44: 403-413.

Wang, Y., Yang, J., Lee, O.O., Dash, S., Lau, S.C.K., Al-Suwailem, A. et al. (2011) Hydrothermally generated aromatic compounds are consumed by bacteria colonizing in Atlantis II Deep of the Red Sea. *ISME Journal* 5: 1652-1659.

Wojcieszyńska, D., Guzik, U., Greń, I., Perkosz, M., and Hupert-Kocurek, K. (2011) Induction of aromatic ring: cleavage dioxygenases in *Stenotrophomonas maltophilia* strain KB2 in cometabolic systems. *World Journal of Microbiology & Biotechnology* 27: 805-811.

APPENDIX

VIII. TABLES

Table 2.1. Potential degenerate primer sequences to target *boxB*.

Beginning Position in Amino Acid Alignment	Conserved Residues	Forward Nucleotide Sequence ($^{5'}$ to $^{3'}$)	Reverse Complement Sequence ($^{5'}$ to $^{3'}$)	Degeneracy
171	QGDTEP[a]	car ggn gay acn gar ccn	ngg ytc ngt rtc ncc ytg	512
200	QVNVEE[b]	car gtn aay gtn gar gar	ytc ytc nac rtt nac ytg	256
208	HLWAMVY[c]	cay ytn tgg gcn atg gtn tay	rta nac cat ngc cca nar rtg	512
210	WAMVYL[c]	tgg gcn atg gtn tay ytn	nar rta nac cat ngc cca	256
265	TDRDGK[d]	acn gay mgn gay ggn aar	ytt ncc rtc nck rtc ngt	1024
290	FMLTEEA[e]	tty atg ytn acn gar gar gcn	ngc ytc ytc nkt nar cat raa	1024
294	EEAHHM[f]	gar gar gcn cay cay atg	cat rtg rtg ngc ytc ytc	128

[a] The G residue is not completely conserved; the exceptions (ZP_01912152, ZP_01906776) contains an A.

[b] The V residue is replaced by a I in several taxa (EEFV86179, EGP47819, YP_001632180, YP_003981052, YP_548463, YP_004152443, YP_002942023, YP_995533).

[c] The V residue is replaced with an A in ZP_01912152.

[d] All are 100% conserved (identical).

[e] The T is replaced with a K residue in 5 taxa (ZP_08119004, YP_004331581, ZP_07283004, YP_003112768, YP_003770707).

[f] The first H residue is replaced with a F in ZP_01912152.

IX. FIGURES

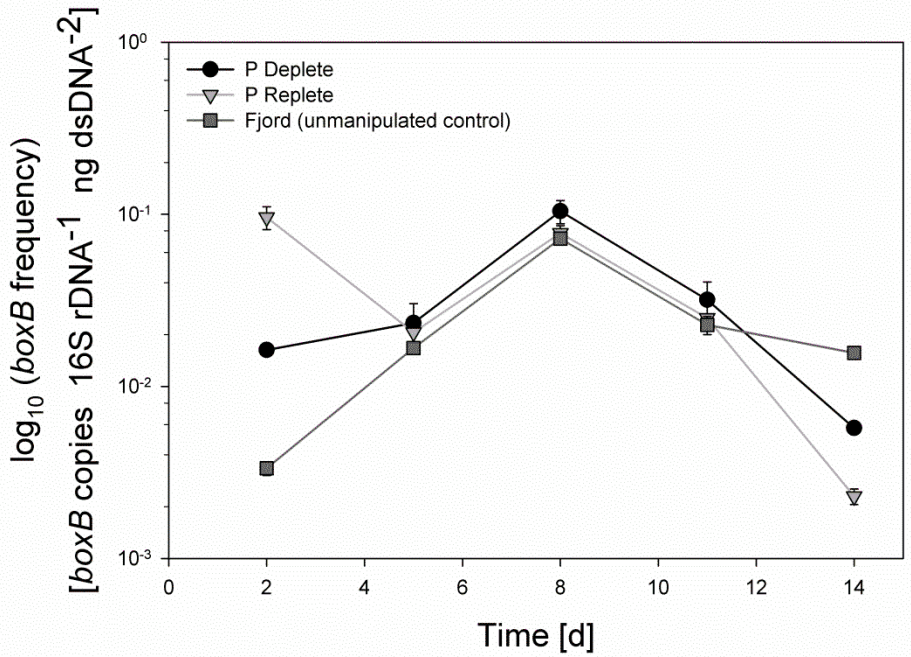


Figure 2.1. Abundance of *boxB* for three different mesocosm types and expressed as *boxB*300 copies per bacterial 16S rDNA per ng dsDNA. Error bars represent standard deviation of technical qPCR triplicates. P, Phosphorous.

Figure 2.2. Maximum-likelihood phylogenetic tree of *boxB*300 determined from 1,000 trees with 50,000 bootstrap iterations. Only cultured members shown, but the tree was constructed with all CAMERA metagenome-originating *boxB*300 sequences as well. Bootstrap values at nodes <70% are not shown, and the outgroup is an uncultured archaeon sequence (not shown). Upper rows for each taxonomic clade represent the number of *boxB*300 reads retrieved from each CAMERA metagenome. Lower rows correspond to the percentage of total *boxB*300 identified within each metagenome. α, *Alphaproteobacteria*; β, *Betaproteobacteria*; δ, *Deltaproteobacteria*; γ, *Gammaproteobacteria*; *Actino*, *Actinobacteria*; *Roseo*; *Roseobacter*; Antarctica, Antarctic lakes metagenome; Broad, Broad phage metagenome; Botany Bay, Botany Bay metagenome; EBPR, Enhanced biological phosphorus removal sludge metagenome; Farm Soil, Waseca County farm soil metagenome; Fjord, Norwegian Raunefjorden clone library; GOS, Global Ocean Survey metagenome; HOT, Hawaii Ocean Time-series metagenome; PBSM, Pacific Beach sand metagenome; Uncl., unclassified; W. Channel, Western Channel Observatory metagenome.

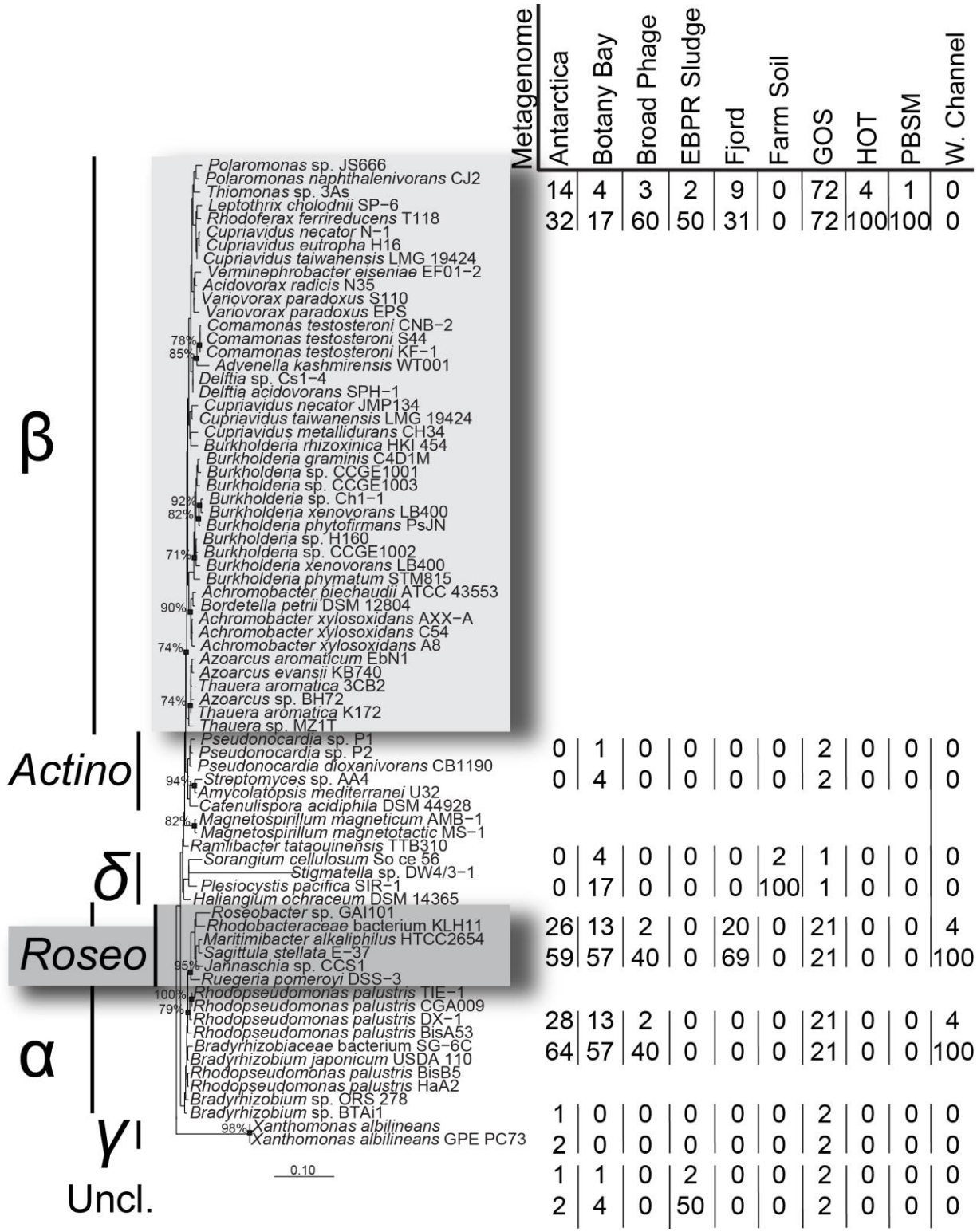

Metagenome
Antarctica
Botany Bay
Broad Phage
EBPR Sludge
Fjord
Farm Soil
GOS
HOT
PBSM
W. Channel
Polaromonas sp. JS666
Polaromonas naphthalenivorans CJ2
Thiomonas sp. 3As
Leptothrix cholodnii SP−6
Rhodoferax ferrireducens T118
Cupriavidus necator N−1
Cupriavidus eutropha H16
Cupriavidus taiwanensis LMG 19424
Verminephrobacter eiseniae EF01−2
Acidovorax radicis N35
Variovorax paradoxus S110
Variovorax paradoxus EPS
Comamonas testosteroni CNB−2
78%
Comamonas testosteroni S44
85%
Comamonas testosteroni KF−1
Advenella kashmirensis WT001
Delftia sp. Cs1−4
Delftia acidovorans SPH−1
Cupriavidus necator JMP134
Cupriavidus taiwanensis LMG 19424
Cupriavidus metallidurans CH34
Burkholderia rhizoxinica HKI 454
Burkholderia graminis C4D1M
Burkholderia sp. CCGE1001
Burkholderia sp. CCGE1003
92%
Burkholderia sp. Ch1−1
82%
Burkholderia xenovorans LB400
Burkholderia phytofirmans PsJN
Burkholderia sp. H160
71%
Burkholderia sp. CCGE1002
Burkholderia xenovorans LB400
Burkholderia phymatum STM815
Achromobacter piechaudii ATCC 43553
90%
Bordetella petrii DSM 12804
Achromobacter xylosoxidans AXX−A
Achromobacter xylosoxidans C54
74%
Achromobacter xylosoxidans A8
Azoarcus aromaticum EbN1
Azoarcus evansii KB740
Thauera aromatica 3CB2
74%
Azoarcus sp. BH72
Thauera aromatica K172
Thauera sp. MZ1T
β
Pseudonocardia sp. P1
Pseudonocardia sp. P2
Pseudonocardia dioxanivorans CB1190
94%
Streptomyces sp. AA4
Amycolatopsis mediterranei U32
Catenulispora acidiphila DSM 44928
Actino
82%
Magnetospirillum magneticum AMB−1
Magnetospirillum magnetotactic MS−1
Ramlibacter tataouinensis TTB310
Sorangium cellulosum So ce 56
Stigmatella sp. DW4/3−1
δ
Plesiocystis pacifica SIR−1
Haliangium ochraceum DSM 14365
Roseobacter sp. GAI101
Rhodobacteraceae bacterium KLH11
Maritimibacter alkaliphilus HTCC2654
Roseo
Sagittula stellata E−37
95%
Jannaschia sp. CCS1
Ruegeria pomeroyi DSS−3
100%
Rhodopseudomonas palustris TIE−1
79%
Rhodopseudomonas palustris CGA009
Rhodopseudomonas palustris DX−1
Rhodopseudomonas palustris BisA53
Bradyrhizobiaceae bacterium SG−6C
α
Bradyrhizobium japonicum USDA 110
Rhodopseudomonas palustris BisB5
Rhodopseudomonas palustris HaA2
Bradyrhizobium sp. ORS 278
Bradyrhizobium sp. BTAi1
98%
Xanthomonas albilineans
Xanthomonas albilineans GPE PC73
γ
0.10
Uncl.
14 4 3 2 9 0 72 4 1 0
32 17 60 50 31 0 72 100 100 0
0 1 0 0 0 0 2 0 0 0
0 4 0 0 0 0 2 0 0 0
0 4 0 0 0 2 1 0 0 0
0 17 0 0 0 100 1 0 0 0
26 13 2 0 20 0 21 0 0 4
59 57 40 0 69 0 21 0 0 100
28 13 2 0 0 0 21 0 0 4
64 57 40 0 0 0 21 0 0 100
1 0 0 0 0 0 2 0 0 0
2 0 0 0 0 0 2 0 0 0
1 1 0 2 0 0 2 0 0 0
2 4 0 50 0 0 2 0 0 0

Figure 2.2

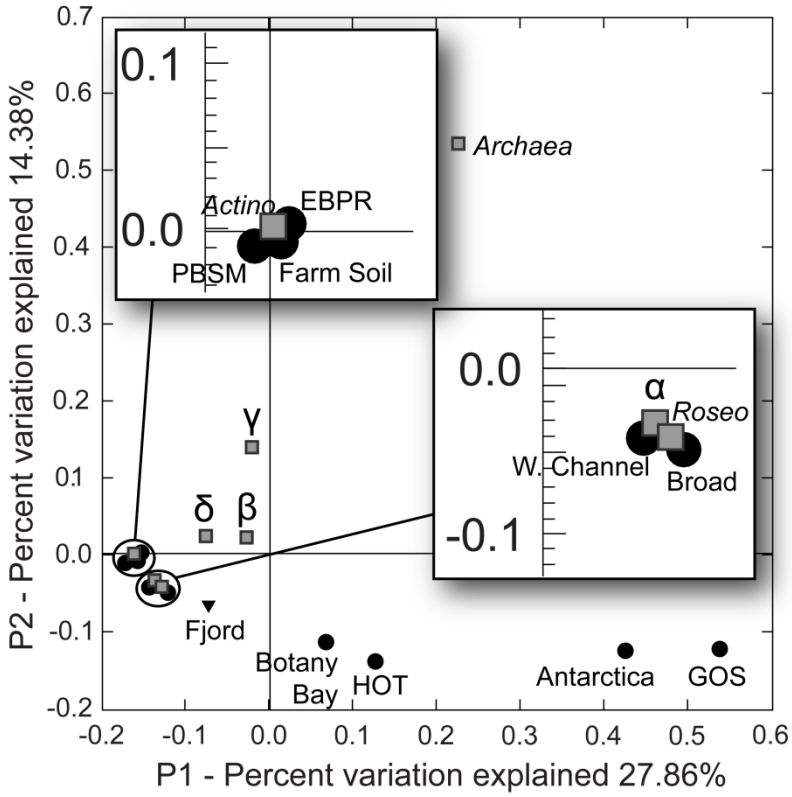


Figure 2.3. PCoA of *boxB*300 amplicon relatedness among taxonomic groups (squares), the fjord clones (triangle), and other environmental locations (circles). Both square insets reflect the organization of closely related groups. Abbreviations are identical to those used in Figure 2.2.

CHAPTER 3 -

DE-METAST-BLAST: A TOOL FOR THE VALIDATION OF DEGENERATE PRIMER SETS AND DATA MINING OF PUBLICLY AVAILABLE METAGENOMES

A VERSION OF THIS CHAPTER HAS BEEN PUBLISHED:

Gulvik CA, Effler TC, Wilhelm SW, Buchan A (2012) De-MetaST-BLAST: A tool for the validation of degenerate primer sets and data mining of publicly available metagenomes. PLOS ONE 7: e50362. DOI: 10.1371/journal.pone.0050362

Four individuals significantly contributed to this chapter: Christopher A. Gulvik, T. Chad Effler, Steven W. Wilhelm, and Alison Buchan. Conceived the program: CAG. Wrote the scripts: TCE. Analyzed the data: CAG AB. Wrote the paper: CAG SWW AB.

I. ABSTRACT

Development and use of primer sets to amplify nucleic acid sequences of interest is fundamental to studies spanning many life science disciplines. As such, the validation of primer sets is essential. Several computer programs have been created to aid in the initial selection of primer sequences that may or may not require multiple nucleotide combinations (*i.e.*, degeneracies). Conversely, validation of primer specificity has remained largely unchanged for several decades, and there are currently few available programs that allows for an evaluation of primers containing degenerate nucleotide bases. To alleviate this gap, we developed the program De-MetaST that performs an *in silico* amplification using user defined nucleotide sequence dataset(s) and primer sequences that may contain degenerate bases. The program returns an output file that contains the *in silico* amplicons. When De-MetaST is paired with NCBI's BLAST (De-MetaST-BLAST), the program also returns the top 10 nr NCBI database hits for each recovered *in silico* amplicon. While the original motivation for development of this search tool was degenerate primer validation using the wealth of nucleotide sequences available in environmental metagenome and metatranscriptome databases, this search tool has potential utility in many data mining applications.

II. INTRODUCTION

PCR is one of the most fundamental and powerful molecular biology tools available. PCR primer sets that contain degenerate bases allow for the amplification of homologous sequences and have been used in various applications, including genetic diversity analyses (*e.g.*, Luton et al., 2002; Mehta et al., 2003; Bürgmann et al., 2004; Jarman et al., 2004; Brown et al., 2005; Chadhain et al., 2006; Schmalenberger and Kertesz, 2007; El Azhari et al., 2008; Brennerova et al., 2009; Wang et al., 2010a; Wang et al., 2010b; Matteson et al., 2011). Several software packages that use a nucleotide or amino acid alignment of the genetic target are available to aid in the initial development of degenerate primer sets (*e.g.*, Amplicon (Jarman, 2004), CODEHOP (Rose et al., 1998; Rose et al., 2003; Staheli et al., 2011), DEFOG (Fuchs et al., 2002), DePiCt (Wei et al., 2003), HYDEN (Linhart and Shamir, 2007), MAD-DPD (Najafabadi et al., 2008), PhiSiGns (Dwivedi et al., 2012), and Primaclade (Gadberry et al., 2005)). In addition, manual identification of conserved regions from aligned sequences generated using software such as ARB (Ludwig et al., 2004), ClustalX (Larkin et al., 2007), and MEGA (Tamura et al., 2011) is also common practice (*e.g.*,Kwok et al., 1994; Lyons et al., 2003; Junca and Pieper, 2004; Allen et al., 2005; Hendrickx et al., 2006; Malmstrom et al., 2010). Once candidate primers are developed, thermodynamic properties and self-complementarity tests can be obtained using online tools (*e.g.*, OligoCalc; Kibbe, 2007).

Despite the utility and common use of degenerate primers, there are no software programs specifically designed to facilitate validation of their specificity. The most common practice for initial validation of degenerate primers is by direct sequence analysis of PCR amplicons (*e.g.*, Rotthauwe et al., 1997; Kirchman et al., 2003; Ochsenreiter et al., 2003; Stach et al., 2003;

López-López et al., 2005). This can be both laborious and costly, and does not take advantage of the ever-increasing publicly available nucleotide data, including that derived from natural samples. In fact, environmental metagenomes and metatranscriptomes are especially attractive reference databases (*e.g.,* CAMERA (Sun et al., 2011) [http://camera.calit2.net/] and MG-RAST [http://metagenomics.anl.gov/]) to perform *in silico* tests *en masse* to identify sequences a degenerate primer set might amplify.

To address this gap in available bioinformatic tools, we have developed a program termed De-MetaST. This program accepts primers that are degenerate using a meta-genome and –transcriptome search tool to retrieve *in silico* PCR amplicons. When paired with BLAST (Altschul et al., 1990), the output provides the most homologous sequences in GenBank for each recovered *in silico* amplicon. In this report, we provide an overview of the program and outline its utility as a tool to validate the specificity of degenerate primer sets. This program is designed to be user-friendly for non-bioinformatics specialists and is publicly available; as are screencast video tutorials demonstrating installation and implementation.

III. DESIGN AND PROGRAM OVERVIEW

De-MetaST is written in C++ and is provided as an executable wrapper to include BLAST (De-MetaST-BLAST) as well as an independent executable (De-MetaST). The function of De-MetaST is to implement a search routine based on bitwise comparisons. Initial steps translate the degenerate nucleotide sequences of each primer, as well as their reverse complement sequences, into unique and specific binary representations. This approach facilitates rapid searches of large

databases that are also transformed into binary representations. The specific computational steps of De-MetaST are outlined in Figure 3.1.

How De-MetaST Works

The De-MetaST program initially converts the inputted primer sequences into 4-digit binary code, where the 16 possible combinations of nucleotides include: A, T, C, G, B, D, H, K, M, N (or X), R, S, V, W, and Y (Figure 3.2). Then, each sequence read within a user defined, FASTA formatted database is converted to 4-digit binary codes and scanned using a bitwise searching operation for the presence of both primer sequences in the appropriate orientation. Limited memory is necessary for this action because each sequence read is individually transformed to binary and immediately scanned for the presence of the primer sequences. The program searches using both the original user inputted primers as well as the reverse and complement of those sequences. This latter search is done to insure identification of target sequences regardless of whether the sense or antisense strand is represented by the database sequence read scanned. The search feature also allows a single primer to serve as both the forward and reverse primer. When primers identify their respective target(s) within a sequence read, the nucleotide sequence delimited by the two primers, termed the *in silico* amplicon, is retrieved. The primer(s) yielding each amplicon are reported in the output. De-MetaST is written to parse *in silico* amplicons >5000 bp into a separate FASTA formatted file that is not subject to BLASTx; users can modify this length restriction by editing the code. All *in silico* amplicons provided in the output represent the sense strand in a 5' to 3' orientation. Thus, when positive hits are made to reads representing antisense strands, the complement and reverse of those reads are generated. Any identifying features (*e.g.*, unique read number) as well as the file name for each predicted hit is recovered.

Although developed to accept degenerate primers, non-degenerate primers can also be input into De-MetaST.. Furthermore, the nucleotide query database(s) themselves may contain sequence reads with degenerate or ambiguous nucleotides (*e.g.,* N). Finally, De-MetaST accepts multiple primer sets as input; the *in silico* amplicons from each set are output into separate FASTA files. As De-MetaST accepts degeneracies in the input primer sequences, it requires absolute conservation in the target sequences; it does not allow for any mismatches between the primer sequence and target. In this way, the user controls the level of primer specificity.

De-MetaST paired with BLAST

Once the database sequence files have been queried for predicted PCR amplicons, each *in silico* amplicon is subject to a BLASTx analysis, which translates the nucleotide sequence in all six frames and performs queries for each translation against the non-redundant (nr) NCBI protein database. The top 10 BLASTx hits for each amplicon are formatted as an XML file. The final step of De-MetaST-BLAST compiles all of the meta-information of the BLASTx results for each amplicon retrieved (*e.g.,* hit accession number, E-value, predicted function, nucleotide sequence, database file name, the primer combination that retrieved the amplicon, unique read number) into a single, tab-delimited TXT file. The BLASTx results file can also be exported as an XLS file format for direct use in Microsoft Excel or other suitable program. A graphical overview of the De-MetaST-BLAST workflow is shown in Figure 3.3.

IV. RESULTS AND DISCUSSION

We have developed a computational method to generate *in silico* amplifications from degenerate primer sets searched against user defined nucleotide databases. To illustrate the utility of De-

MetaST-BLAST, we demonstrate its performance using a novel degenerate primer set designed for use on environmental samples. This primer set targets the bacterial *boxB* gene, which encodes the oxygenase component of a multi-enzyme epoxidase (EC 1.14.13) that is specific to a benzoate catabolic pathway (Rather et al., 2010). Three metagenome libraries representing different environments, library size and DNA sequencing methods were searched and found to contain putative *boxB* amplicons of the appropriate size (300 bp) (Table 3.1). Figure 3.4 shows the typical output of De-MetaST-BLAST for one of those database searches, which includes for each *in silico* amplicon the top 10 BLASTx hits with their corresponding E-value and GenBank accession number.

In order to retrieve an *in silico* amplicon, the program requires both primers to match their respective targets in a single sequence read or sequence assembly (contig). Thus, an important consideration in terms of selection of appropriate searchable databases is the average length of the sequence read or assembly contained within it, as well as the desired amplicon size. This concern may be alleviated as longer read sequencing technologies are developed and/or as sequence coverage and assembly algorithms improve. Interestingly, our analysis demonstrates that *in silico* amplicons of ~300 bp and ~190 bp, representing *boxB* and 16S rRNA gene amplicons, respectively, can be readily recovered from databases dominated by short read length sequences (*e.g.* AntarcticaAquatic; Table 3.1). In fact, the 44 *boxB* amplicons derived from the AntarcticaAquatic dataset were found in reads that ranged from 348-541 bp in length. This result suggests that sequence coverage, or depth, is also a contributing factor to *in silico* amplicon recovery. Incidentally, all of the *in silico* amplicons recovered in this demonstration run were found to be homologous to the desired target (E-value $\leq$ 1e-4).

In terms of data mining, De-MetaST can provide complementary sequence data for gene diversity studies. For example, *in silico* amplicons retrieved from existing sequence datasets can augment experimentally derived clone library sequences. As the De-MetaST output provides the sequence from the same genetic positions as that derived from a companion clone library, downstream analysis, such as sequence alignment and subsequent phylogenetic analysis, is streamlined. Furthermore, as the nucleotide sequences targeted by the primers are returned in the De-MetaST output, users can draw on that information to further refine their primers according to a desired level of functional and/or phylogenetic specificity. The program also has utility beyond searches of environmental sequence databases. It can be used to query any nucleotide dataset, including those derived from single organisms. Thus, it has use in assessing the specificity of primers targeting multi-copy or homologous genes within a single organism or group of organisms.

Benchmarks and System Requirements

De-MetaST-BLAST has been developed for the long-term support (LTS) Ubuntu operating systems 10.04 LTS and 12.04 LTS. While De-MetaST does not make use of multi-core processors, BLAST maintains that capability. Benchmarks were performed on an Intel i7-2600 processor (3.4GHz quad-core, 8-thread) desktop using the developed degenerate *boxB* primer set against the Waseca Farm Soil metagenome (AAFX01000000) (Tringe et al., 2005). This search took approximately 11.7 s (Table 3.2). When the database size was artificially and incrementally increased up to five-fold (772 Mb) by replication of the original dataset, the processing time remained <1 min. Furthermore, to determine the effect of increased numbers of positive hits on

run time, the libraries were seeded with additional sequences containing the target. Doubling of targets within the databases had no effect on run time (Table 3.2). In contrast to the relatively rapid processing speed of De-MetaST, implementation of the BLAST function can add significant processing time to the process, particularly if a local custom database is used. As an example, for the initial benchmark search against the locally installed Farm Soil metagenome that recovered two hits, the BLASTx function added 39.3 s using two threads. Thus, computational requirements and processing speed are primarily dictated by BLAST. When BLAST is performed remotely—the default setting (see below)—the return time is dependent upon availability and processing speeds of the NCBI servers.

Both De-MetaST and De-MetaST-BLAST can be run on any operating system with a C++ compiler (*e.g.*, standard Windows and Mac OS). However, users would need to ensure the BLAST installation is compatible with their processor.

Availability of De-MetaST-BLAST

The De-MetaST package and the De-MetaST-BLAST wrapper are made freely available at http://sourceforge.net/p/de-metast-blast/ and http://code.google.com/p/de-metast-blast/. Along with the program, screencast tutorial videos describe how to install the necessary programs as well as implement the software package with the example dataset provided. The De-MetaST package is self-contained and has no external dependencies, except a C++ compiler, such as g++. De-MetaST-BLAST requires a local BLAST+ suite installation that supports direct query of the NCBI nr protein database using NCBI servers via the –remote option. However, the program can also be configured to query a custom local database. Both approaches are described in tutorial

videos provided. Installation of the De-MetaST program is estimated at 5 min, whereas installation of the BLAST+ suite is estimated to take 3 min, excluding download and extraction times, which are dependent on the user's internet speed and processing power.

V. CONCLUSIONS

It was recently predicted that the increasing amounts of metagenome sequences will likely serve as a valuable resource in evaluation of the coverage and specificity of previously developed primer sets (Iwai et al., 2011). De-MetaST-BLAST will provide users with a useful tool in such evaluations. De-MetaST is designed to provide *in silico* amplicons generated by user defined degenerate primers found within a user defined nucleotide database. When paired with BLAST, the program returns the most homologous GenBank hits, which are useful in assessing the specificity of degenerate primers. However, the program does not evaluate PCR kinetics and efficiencies with degenerate primers. Thus, users are encouraged to consult appropriate references on the use and design of degenerate primers (*e.g.*, Kwok et al., 1994; Preston, 1997), including those that discuss the merits of utilizing base analogs (*e.g.*, inosine; Liu and Nichols, 1994) that can reduce the overall degeneracy of primers.

VI. ACKNOWLEDGMENTS

We thank Dr. Charles R. Budinoff for insightful discussions and Ashley M. Frank, P. Jackson Gainer, and W. Nathan Cude for providing valuable feedback on program installation and use. TCE was supported by an REU award from the NSF (MCB1112001 to SWW). SWW and AB acknowledge NSF award OCE1061352 for support.

VII. REFERENCES

Allen, A.E., Ward, B.B., and Song, B.K. (2005) Characterization of diatom (Bacillariophyceae) nitrate reductase genes and their detection in marine phytoplankton communities. *Journal of Phycology* **41**: 95-104.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Brennerova, M.V., Josefiova, J., Brenner, V., Pieper, D.H., and Junca, H. (2009) Metagenomics reveals diversity and abundance of *meta*-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environmental Microbiology* **11**: 2216-2227.

Brown, M.V., Schwalbach, M.S., Hewson, I., and Fuhrman, J.A. (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environmental Microbiology* **7**: 1466-1479.

Bürgmann, H., Widmer, F., Von Sigler, W., and Zeyer, J. (2004) New molecular screening tools for analysis of free-living diazotrophs in soil. *Applied and Environmental Microbiology* **70**: 240-247.

Chadhain, S.M.N., Schaefer, J.K., Crane, S., Zylstra, G.J., and Barkay, T. (2006) Analysis of mercuric reductase (merA) gene diversity in an anaerobic mercury-contaminated sediment enrichment. *Environmental Microbiology* **8**: 1746-1752.

Dwivedi, B., Schmieder, R., Goldsmith, D.B., Edwards, R.A., and Breitbart, M. (2012) PhiSiGns: an online tool to identify signature genes in phages and design PCR primers for examining phage diversity. *BMC Bioinformatics* **13**.

El Azhari, N., Bru, D., Sarr, A., and Martin-Laurent, F. (2008) Estimation of the density of the protocatechuate-degrading bacterial community in soil by real-time PCR. *European Journal of Soil Science* **59**: 665-673.

Fuchs, T., Malecova, B., Linhart, C., Sharan, R., Khen, M., Herwig, R. et al. (2002) DEFOG: A practical scheme for deciphering families of genes. *Genomics* **80**: 295-302.

Gadberry, M.D., Malcomber, S.T., Doust, A.N., and Kellogg, E.A. (2005) Primaclade - a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* **21**: 1263-1264.

Hendrickx, B., Junca, H., Vosahlova, J., Lindner, A., Ruegg, I., Bucheli-Witschel, M. et al. (2006) Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: Distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. *Journal of Microbiological Methods* **64**: 250-265.

Iwai, S., Johnson, T.A., Chai, B., Hashsham, S.A., and Tiedje, J.M. (2011) Comparison of the specificities and efficacies of primers for aromatic dioxygenase gene analysis of environmental samples. *Applied and Environmental Microbiology* **77**: 3551-3557.

Jarman, S.N. (2004) Amplicon: software for designing PCR primers on aligned DNA sequences. *Bioinformatics* **20**: 1644-1645.

Jarman, S.N., Deagle, B.E., and Gales, N.J. (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology* **13**: 1313-1322.

Junca, H., and Pieper, D.H. (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. *Environmental Microbiology* **6**: 95-110.

Kibbe, W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* **35**: W43-W46.

Kirchman, D.L., Yu, L.Y., and Cottrell, M.T. (2003) Diversity and abundance of uncultured Cytophaga-like bacteria in the Delaware Estuary. *Applied and Environmental Microbiology* **69**: 6587-6596.

Kwok, S., Chang, S.Y., Sninsky, J.J., and Wang, A. (1994) A guide to the design and use of mismatched and degenerate primers. *Genome Research* **3**: S39-S47.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Linhart, C., and Shamir, R. (2007) Degenerate primer design: theoretical analysis and the HYDEN program. *Methods in molecular biology (Clifton, NJ)* **402**: 221-244.

Liu, H., and Nichols, R. (1994) PCR amplification using deoxyinosine to replace entire codon and at ambiguous positions. *Biotechniques* **16**: 24-26.

López-López, A., Bartual, S.G., Stal, L., Onyshchenko, O., and Rodríguez-Valera, F. (2005) Genetic analysis of housekeeping genes reveals a deep-sea ecotype of Alteromonas macleodii in the Mediterranean Sea. *Environmental Microbiology* **7**: 649-659.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* **32**: 1363-1371.

Luton, P.E., Wayne, J.M., Sharp, R.J., and Riley, P.W. (2002) The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology* **148**: 3521-3530.

Lyons, J.I., Newell, S.Y., Buchan, A., and Moran, M.A. (2003) Diversity of ascomycete laccase gene sequences in a southeastern US salt marsh. *Microbial Ecology* **45**: 270-281.

Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., and Chisholm, S.W. (2010) Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans. *ISME Journal* **4**: 1252-1264.

Matteson, A.R., Loar, S.N., Bourbonniere, R.A., and Wilhelm, S.W. (2011) Molecular enumeration of an ecologically important cyanophage in a laurentian great lake. *Applied and Environmental Microbiology* **77**: 6772-6779.

Mehta, M.P., Butterfield, D.A., and Baross, J.A. (2003) Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca ridge. *Applied and Environmental Microbiology* **69**: 960-970.

Muyzer, G., Dewaal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology* **59**: 695-700.

Najafabadi, H.S., Saberj, A., Torabi, N., and Chamankhah, M. (2008) MAD-DPD: designing highly degenerate primers with maximum amplification specificity. *Biotechniques* **44**: 519-526.

Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L., and Schleper, C. (2003) Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environmental Microbiology* **5**: 787-797.

Preston, G.M. (1997) Cloning gene family members using the polymerase chain reaction with degenerate oligonucleotide primers. *Methods in molecular biology (Clifton, NJ)* **69**: 97-113.

Rather, L.J., Knapp, B., Haehnel, W., and Fuchs, G. (2010) Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. *Journal of Biological Chemistry* **285**: 20615-20624.

Rose, T.M., Henikoff, J.G., and Henikoff, S. (2003) CODEHOP (COnsensus-DEgenerate hybrid oligonucleotide primer) PCR primer design. *Nucleic Acids Research* **31**: 3763-3766.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Research* **26**: 1628-1635.

Rotthauwe, J.H., Witzel, K.P., and Liesack, W. (1997) The ammonia monooxygenase structural gene amoA as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations. *Applied and Environmental Microbiology* **63**: 4704-4712.

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. *Applied and Environmental Microbiology* **75**: 7537-7541.

Schmalenberger, A., and Kertesz, M.A. (2007) Desulfurization of aromatic sulfonates by rhizosphere bacteria: high diversity of the *asfA* gene. *Environmental Microbiology* **9**: 535-545.

Stach, J.E.M., Maldonado, L.A., Ward, A.C., Goodfellow, M., and Bull, A.T. (2003) New primers for the class *Actinobacteria*: application to marine and terrestrial environments. *Environmental Microbiology* **5**: 828-841.

Staheli, J.P., Boyce, R., Kovarik, D., and Rose, T.M. (2011) CODEHOP PCR and CODEHOP PCR primer design. *Methods in molecular biology (Clifton, NJ)* **687**: 57-73.

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S. et al. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Research* **39**: D546-D551.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731-2739.

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.

Wang, G.Z., Wang, Y.R., Yang, P.L., Luo, H.Y., Huang, H.Q., Shi, P.J. et al. (2010a) Molecular detection and diversity of xylanase genes in alpine tundra soil. *Applied Microbiology and Biotechnology* **87**: 1383-1393.

Wang, L.P., Wang, W.P., Lai, Q.L., and Shao, Z.Z. (2010b) Gene diversity of CYP153A and AlkB alkane hydroxylases in oil-degrading bacteria isolated from the Atlantic Ocean. *Environmental Microbiology* **12**: 1230-1242.

Wei, X., Kuhn, D., and Narasimhan, G. (2003) Degenerate primer design via clustering. In *Proceedings of the IEEE Computer Society Bioinformatics Conference*. Stanford, CA, pp. pp 75-83.

APPENDIX

VII. TABLES

Table 3.1. *boxB* and 16S rRNA gene *in silico* amplicons identified in representative metagenomes using De-MetaST-BLAST.

CAMERA Metagenome Database Queried	*boxB in silico* amplicons[a]	Unique *boxB* reads[b]	16S rRNA gene *in silico* amplicons[c]	Unique 16S rDNA reads[b]	Database Size [Mbp]	Number of Reads	Average Read Length [bp][d]	Sequencing Method(s)
CAM_PROJ_FarmSoil.read.fa	2	2	6	6	155	1.38E+05	1117	dideoxysequencing (Sanger)
CAM_PROJ_GOS.read.fa	100	86	3710	965	11598	1.36E+07	915	dideoxysequencing (Sanger) and pyrosequencing (454)
CAM_PROJ_AntarcticaAquatic.read.fa	44	43	4758	1665	23819	6.46E+07	369	dideoxysequencing (Sanger) and pyrosequencing (454)

[a] The primers *boxB*171F ($^{5'}$ CARGGNGAYACNGARCC $^{3'}$) and *boxB*265R ($^{5'}$ YTTNCCRTCNCKRTCNGT $^{3'}$) were used to target an approximately 300 bp region of *boxB.*

[b] Unique reads were identified using MOTHUR (v.1.27.0) (Schloss et al., 2009).

[c] The primers 358f ($^{5'}$ CCTACGGGAGGCAGCAG $^{3'}$) and 517r ($^{5'}$ ATTACCGCGGCTGCTGG $^{3'}$) (Muyzer et al., 1993) were used to target an approximately 190 bp amplicon in the 16S rRNA gene.

[d] Average read length was estimated by dividing the database size by number of reads. The AntarcticaAquatic database is dominated by pyrosequencing derived reads (98% of all reads), while the GOS dataset is dominated by Sanger derived reads; the exact distribution for GOS reads is not available.

Table 3.2. Runtime duration of De-MetaST.

Files Input	Database size [Mbytes][a]	Sequences in database [*10^5]	Nucleotides in Database [Mbp]	Hits	Real Time [s]	User Time [s]	System Time [s]
1	206.1	1.4	154	2	11.7	11.7	0.02
2	412.2	2.8	309	4	23.5	23.4	0.05
3	618.3	4.2	463	6	35.2	35.1	0.07
4	824.4	5.5	618	8	47.6	47.5	0.10
5	1030.5	7.0	772	10	58.6	58.5	0.12
1	206.1	1.4	154	4	11.9	11.9	0.02
2	412.2	2.8	309	8	23.3	23.3	0.05
3	618.3	4.2	463	12	35.6	35.5	0.08
4	824.4	5.5	618	16	47.3	47.1	0.10
5	1030.5	7.0	772	20	58.2	58.0	0.12

[a] The datasets used for benchmarking were manipulations of the Waseca Farm Soil metagenome (AAFX01000000); the average sequence read length in these datasets is 1117 bp.

VIII. FIGURES

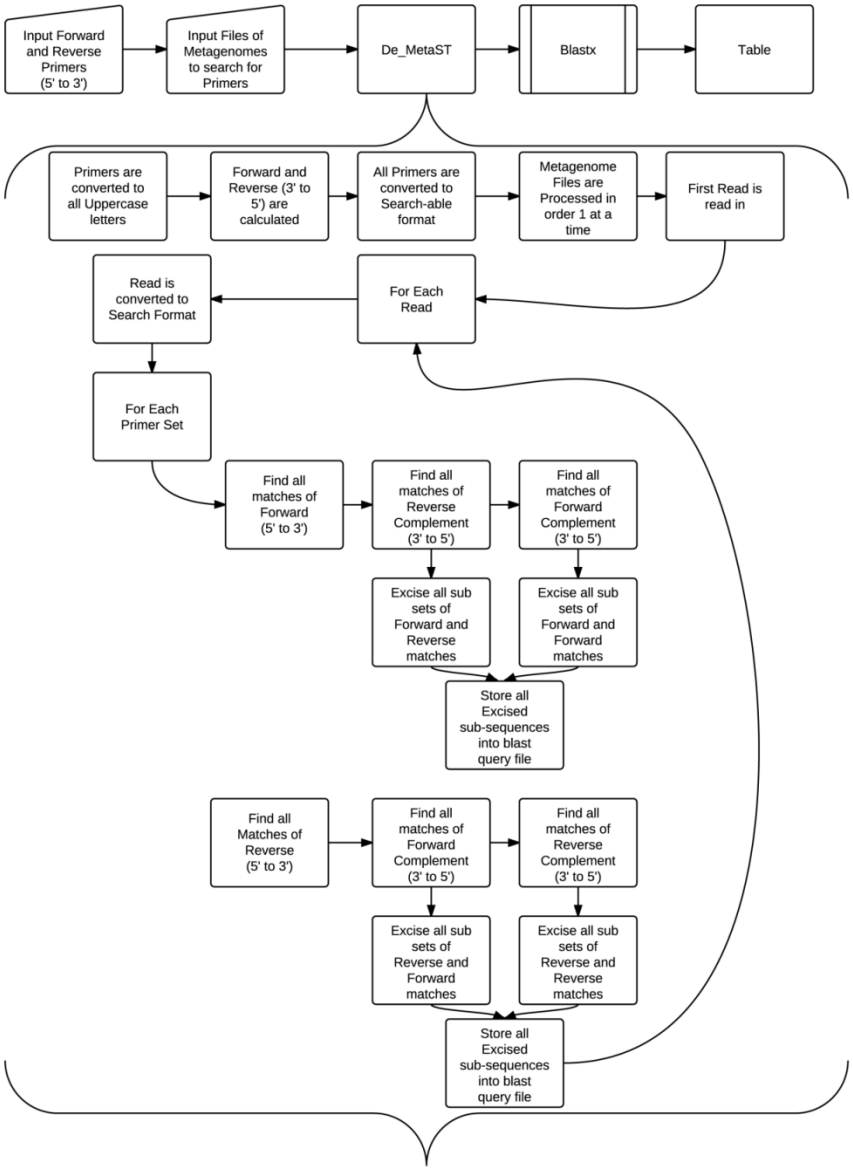


Figure 3.1. Computational procedures of De-MetaST are illustrated within the De-MetaST-BLAST wrapper.

(A)

Nucleotide	Binary Representation
A	0001
C	0010
G	0100
T	1000
B	1110
D	1101
H	1011
K	1100
M	0011
R	0101
S	0110
V	0111
W	1001
Y	1010
N or X	1111

(B)

```
Primer Sequence Input: CAG TCD SWW ABN RYV ACC
   Binary Position #1: 000 101 011 011 010 000
   Binary Position #2: 001 001 100 011 101 000
   Binary Position #3: 100 010 100 011 011 011
   Binary Position #4: 010 001 011 101 101 100

Primer Sequence Input:  C    A    G    T    C    D    S    W    W    A    B    N    R    Y    V    A    C    C
Binary Representation:  0010 0001 0100 1000 0010 1101 0110 1001 1001 0001 1110 1111 0101 1010 0111 0001 0010 0010
```

Figure 3.2. De-MetaST transformation of nucleotide sequences into a binary representation. The binary representation for each of the 16 possible nucleotide character inputs is shown in the upper box. The lower box provides an example of the transformation using a mock primer sequence. Spaced gaps are shown for instructional purposes and do not occur in the De-MetaST search routine.

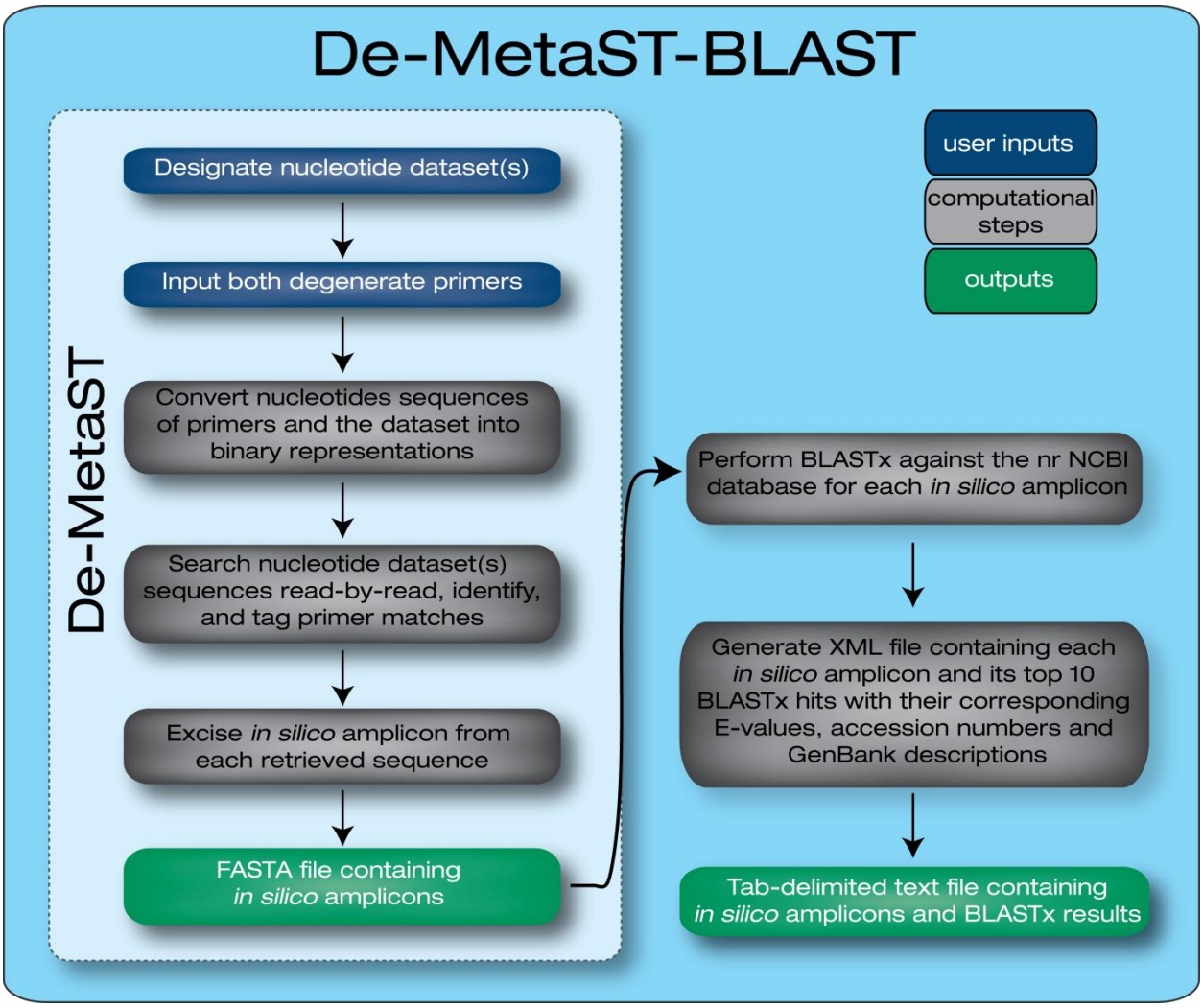
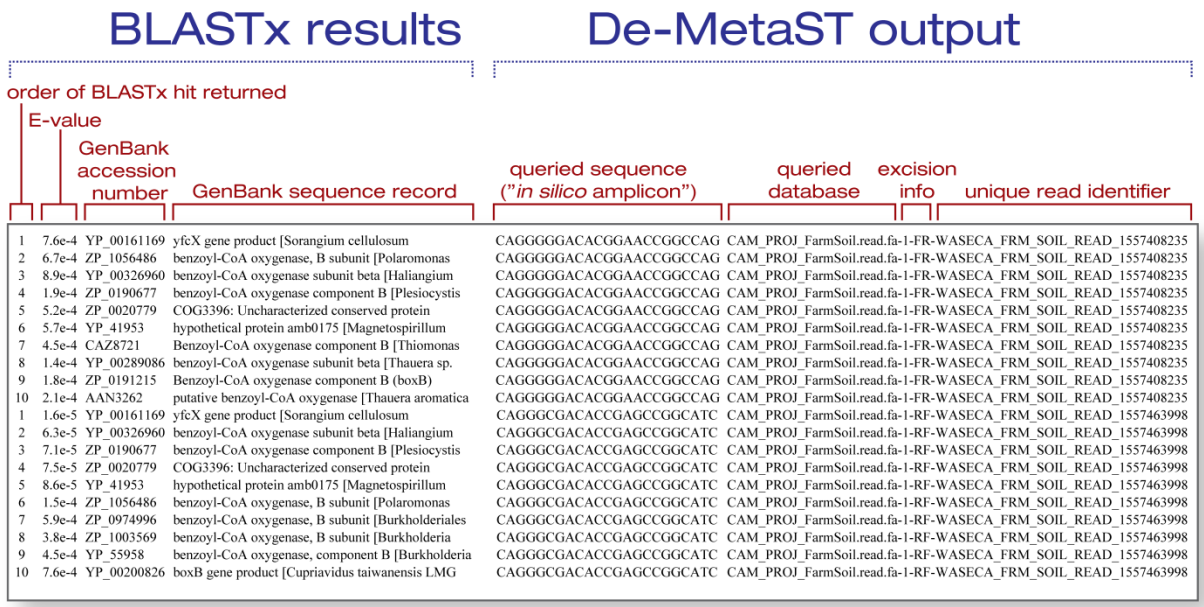


Figure 3.3. Flowchart outlining De-MetaST-BLAST user actions and corresponding computational processes. Fwd, Forward; Rev, Reverse; NCBI, National Center for Biotechnology Information.



1	7.6e-4	YP_00161169	yfcX gene product [Sorangium cellulosum	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
2	6.7e-4	ZP_1056486	benzoyl-CoA oxygenase, B subunit [Polaromonas	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
3	8.9e-4	YP_00326960	benzoyl-CoA oxygenase subunit beta [Haliangium	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
4	1.9e-4	ZP_0190677	benzoyl-CoA oxygenase component B [Plesiocystis	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
5	5.2e-4	ZP_0020779	COG3396: Uncharacterized conserved protein	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
6	5.7e-4	YP_41953	hypothetical protein amb0175 [Magnetospirillum	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
7	4.5e-4	CAZ8721	Benzoyl-CoA oxygenase component B [Thiomonas	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
8	1.4e-4	YP_00289086	benzoyl-CoA oxygenase subunit beta [Thauera sp.	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
9	1.8e-4	ZP_0191215	Benzoyl-CoA oxygenase component B (boxB)	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
10	2.1e-4	AAN3262	putative benzoyl-CoA oxygenase [Thauera aromatica	CAGGGGGACACGGAACCGGCCAG	CAM_PROJ_FarmSoil.read.fa-1-FR-WASECA_FRM_SOIL_READ_1557408235	
1	1.6e-5	YP_00161169	yfcX gene product [Sorangium cellulosum	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
2	6.3e-5	YP_00326960	benzoyl-CoA oxygenase subunit beta [Haliangium	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
3	7.1e-5	ZP_0190677	benzoyl-CoA oxygenase component B [Plesiocystis	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
4	7.5e-5	ZP_0020779	COG3396: Uncharacterized conserved protein	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
5	8.6e-5	YP_41953	hypothetical protein amb0175 [Magnetospirillum	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
6	1.5e-4	ZP_1056486	benzoyl-CoA oxygenase, B subunit [Polaromonas	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
7	5.9e-4	ZP_0974996	benzoyl-CoA oxygenase, B subunit [Burkholderiales	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
8	3.8e-4	ZP_1003569	benzoyl-CoA oxygenase, B subunit [Burkholderia	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
9	4.5e-4	YP_55958	benzoyl-CoA oxygenase, component B [Burkholderia	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	
10	7.6e-4	YP_00200826	boxB gene product [Cupriavidus taiwanensis LMG	CAGGGCGACACCGAGCCGGCATC	CAM_PROJ_FarmSoil.read.fa-1-RF-WASECA_FRM_SOIL_READ_1557463998	

Figure 3.4. Example of De-MetaST-BLAST output. Text within the box denotes the spreadsheet output for a *boxB* primer set search against the WASECA Farm Soil Metagenome (AAFX01000000) [41] that recovers two *in silico* amplicons. Column descriptors are shown in color; select columns have been truncated due to space constraints. For the "excision info" column, the first alphanumeric character reports the "hit" number within a read (i.e. "1" indicates it is the first *in silico* amplicon found within a single read). The subsequent alphanumeric characters denote the primer orientation yielding the amplicon (F = forward, R = reverse). Whether a unique read identifier is returned is contingent upon the database itself.

CHAPTER 4 -

SIMULTANEOUS CATABOLISM OF PLANT-DERIVED AROMATIC COMPOUNDS RESULTS IN ENHANCED GROWTH FOR MEMBERS OF THE *ROSEOBACTER* LINEAGE

A VERSION OF THIS CHAPTER HAS BEEN ACCEPTED FOR PUBLICATION:

Gulvik CA and A Buchan. (2013) Simultaneous catabolism of plant-derived aromatic compounds results in enhanced growth for members of the *Roseobacter* lineage. Appl Environ Microbiol. DOI: 10.1128/AEM.00405-13

I. ABSTRACT

Plant-derived aromatic compounds are important components of the dissolved organic carbon pool in coastal salt marshes, and their mineralization by resident bacteria contributes to carbon cycling in these systems. Members of the roseobacter lineage of marine bacteria are abundant in coastal salt marshes and several characterized strains, including *Sagittula stellata* E-37, utilize aromatic compounds as primary growth substrates. The genome sequence of *S. stellata* contains multiple, potentially competing, aerobic ring-cleaving pathways. Preferential hierarchies in substrate utilization and complex transcriptional regulation have been demonstrated to be the norm in many soil bacteria that also contain multiple ring-cleaving pathways. The purpose of this study was to ascertain whether substrate preference exists in *S. stellata* when provided a mixture of aromatic compounds that proceed through different ring-cleaving pathways. We focused on the protocatechuate (*pca*) and the aerobic benzoyl-CoA (*box*) pathways and the substrates known to proceed through them, *p*-hydroxybenzoate (POB) and benzoate, respectively. When these two substrates were provided at non-carbon limiting concentrations, temporal patterns of cell density, gene transcript abundance, enzyme activity, and substrate concentrations indicated *S. stellata* simultaneously catabolized both substrates. Furthermore, enhanced growth rates were observed when *S. stellata* was provided both compounds simultaneously compared to cells grown singly with an equimolar concentration of either substrate alone. This simultaneous catabolism phenotype was also demonstrated in another lineage member, *Ruegeria pomeroyi* DSS-3. These findings contradict the paradigm of sequential aromatic catabolism reported for soil bacteria and contribute to the growing body of physiological evidence demonstrating the metabolic versatility of roseobacters.

II. INTRODUCTION

The structural diversity of aromatic compounds in the environment is influenced by the various mechanisms that produce them, including naturally occurring abiotic (Wang et al., 2011) and biotic processes (Higuchi, 1990; Boerjan et al., 2003) as well as those of anthropogenic origin (Wakeham et al., 1980). Regardless of the source, the dissolved organic carbon pool containing aromatic compounds in nature is typically structurally heterogeneous (Crawford et al., 1983; Christian and Anderson, 2002). This is an important consideration for microbial degradation and has received significant attention in studies examining the catabolism of aromatic compound mixtures classified as environmental pollutants (Alvarez and Vogel, 1991). Considerably less attention has focused on microbial physiology of mixtures of naturally occurring aromatic compounds, such as those derived from lignin, the structural component of vascular plants (Kirk and Farrell, 1987).

Microbial mineralization of aromatic compounds plays an important role in global carbon cycling and bioremediation. Bacterial aromatic catabolism is described as "catabolic funneling" where upper (also called peripheral) pathways transform a diverse suite of aromatic compounds into one of a limited number of intermediates that are then subject to ring cleavage (Fuchs et al., 2011). The β-ketoadipate pathway is one such pathway and is a paradigm for aerobic catabolism of plant-derived aromatics (Stanier and Ornston, 1973; Harwood and Parales, 1996). In organisms possessing one or both its branches, peripheral pathways generate dihydroxylated intermediates, either catechol or protocatechuate. Alternative aerobic ring-cleaving mechanisms are also present in bacteria, including epoxidation of CoA-thioesterified aromatics, as occurs in the benzoyl-CoA (*box*) pathway for benzoate degradation (Rather et al., 2010). These catabolic

pathways are typically under tight transcriptional regulation (Wheelis and Ornston, 1972; Romero-Steiner et al., 1994) and subject to catabolite repression (Cánovas and Stanier, 1967; Dal et al., 2002; Mazzoli et al., 2007).

Previous studies reveal substrate preferences are the norm when bacterial strains are presented with a mixture of aromatic compounds (Fuchs et al., 2011). This phenomenon has been best characterized for select soil-derived bacteria provided mixtures of benzoate and *p*-hydroxybenzoate (Gaines et al., 1996; Brzostowicz et al., 2003; Donoso et al., 2011), with each ultimately processed through parallel ring-cleavage branches of the β-ketoadipate pathway, referred to as the catechol (*cat*) and protocatechuate (*pca*) branches (Ornston and Stanier, 1966), respectively. The hierarchical nature of substrate utilization profiles has been mechanistically explained as cross-regulation and typically involves transcriptional regulation by pathway metabolites or regulatory proteins (*e.g.*, Siehler et al., 2007; Bleichrodt et al., 2010). However, the extent to which similar hierarchies exist in environmentally relevant microbes is not yet clear. Furthermore, as environmental bacteria are dependent upon growth substrate pools that are highly heterogeneous in composition (McCarthy et al., 1996; del Giorgio and J., 2003), mixed substrate studies may provide the foundation for a better understanding of bacterial catabolism in nature.

The *Roseobacter* lineage of marine bacteria is numerically abundant and active in the world's oceans (Suzuki et al., 2001; Selje et al., 2004). Group members are most dominant in coastal environments, including salt marshes heavily influenced by lignocellulosic vascular plant material (Moran et al., 2007). Roseobacter ubiquity and success in the oceans has been

attributed, in part, to their ability to use a large repertoire of growth substrates, including aromatic compounds (Mou et al., 2008; Newton et al., 2010). Genome analyses have identified several ring-cleaving pathways in roseobacters, including the *box* and *pca* pathways (Newton et al., 2010). Yet, it is unknown whether roseobacters show evidence of substrate preference when presented with mixtures of aromatic compounds representative of compounds derived from vascular plants abundant in coastal marine habitats.

III. MATERIALS AND METHODS

Media and growth conditions. A marine basal medium (MBM) containing 1.5% (w/v) sea salts [Sigma-Aldrich, St. Louis, MO] with 225 nM K_2HPO_4, 13.35 µM NH_4Cl, 71.4285 mM Tris-Cl (pH 7.5), 68 µM Fe-EDTA, trace metals, and vitamins was used to culture *Ruegeria pomeroyi* DSS-3 and *Sagittula stellata* E-37 at 30°C (González et al., 1997). *Variovorax paradoxus* EPS was also cultured at 30°C in M9 basal media (Jamieson et al., 2009). Benzoic acid and *p*-hydroxybenzoic acid were obtained from Sigma-Aldrich, and sodium acetate was obtained from Thermo Fisher Scientific [Waltham, MA]. All glassware was combusted minimally for 6 h at 450°C to remove trace carbon, and negative (non-carbon amended) controls of cells in basal media were also performed. All cell density measurements were manually performed using a GENESYS 20 Spectrophotometer [Thermo Fisher Scientific] at 540 nm.

Nucleic acid isolation. For RNA preservation and isolation, approximately 10^8 cells were pelleted at 5000 x *g* for 5 min, resuspended in 1 ml of RNA*later* [Ambion, Austin, TX], and preserved for 1 h at room temperature. RNA*later* was removed by aspiration following 6000 x *g* centrifugation for 5 min, and the cells were flash frozen in liquid N_2 and stored at -70°C until

processed. Cells were lysed by agitation in the presence of low-binding 200 µm zirconium beads [OPS Diagnostics, L.L.C., Lebanon, NJ], and RNA was extracted using the RNeasy Mini Kit [Qiagen, Valencia, CA]. Genomic DNA was removed using the vigorous Turbo DNase (4 U) [Ambion] method as described in the product manual. Nucleic acids were quantified and purity was assessed with an ND-1000 spectrophotometer [NanoDrop Technologies Inc., Wilmington, DE]. Reverse transcription was carried out in 60 µl volumes containing 180 ng RNA, 600 U M-MLV reverse transcriptase [Invitrogen, Carlsbad, CA], 500 µg ml^{-1} random hexamers [Promega, Madison, WI], 120 U RNaseOUT [Invitrogen], 500 µM dNTPs [Promega], and 10 mM DTT [Invitrogen]. Initially the random hexamers in the presence of dNTPs were annealed to the RNA for 5 min at 65°C, followed by an immediate transfer to ice to maintain the binding interaction. Next, to protect the mRNA and increase full-length cDNA yields, DTT and RNaseOUT were added and incubated for 2 min at 37°C. Finally cDNAs were generated from the RNA by M-MLV reverse transcriptase by first activating for 10 min at 25°C followed by 50 min of synthesis at 37°C. The enzymes were denatured and inactivated for 15 min with 70°C, and the remaining cDNA was stored at -20°C.

Gene transcription assays. RT-qPCR was used to assess relative gene expression. Transcripts diagnostic of the benzoyl-CoA pathway (*boxA*) and the protocatechuate branch of the β-ketoadipate pathway (*pcaH*) were measured and normalized to the expression of three reference genes (*alaS*, *map*, and *rpoC*). Primers were designed for each of these 5 genes and are shown in Table 4.1. All primer sets were optimized for quantitative (qPCR) using the following method. In 25 µl qPCR volumes, a matrix of forward and reverse primer concentrations (ranging from 100-1500 nM final concentrations) was used along with a fixed concentration of E-37 genomic DNA

(2.5×10^5 genomes reaction^{-1}) and 1X SYBR *Premix Ex Taq* (Perfect Real Time) [Takara Bio Inc., Otsu, Japan]. The qPCR amplification included an initial 95°C denaturation for 15 min, followed by 40 cycles of amplification (95°C denaturation for 45 s, 58°C annealing for 45 s, and 72°C elongation and fluorescence detection for 15 s), and a final 72°C extension for 5 min. A melting curve from 50°C to 100°C read every 1°C was performed after each reaction to ensure only a single product melted around 90°C. Within each primer set matrix, the combination of forward and reverse primer concentrations that yielded the lowest C_q was used in the subsequent RT-qPCR.

The 25 µl qPCR volumes consisted of $15\%_{(aq)}$ cDNA template, 577 µM oligonucleotide primers, and SYBR *Premix Ex Taq* (Perfect Real Time) [Takara]. Each cDNA template sample was amplified with primer sets to quantitate *alaS*, *boxA*, *map*, *pcaH*, and *rpoC*. Technical qPCR triplicates were performed for each of the five primer sets used to amplify the three biological triplicates at five time points. The cycling conditions were also the same as described for the qPCR optimization. Non reverse-transcribed aliquots were also performed as negative controls to ensure the RT-qPCR measurements represented cDNA concentration and other nucleic acids' C_q were negligible (>5 difference). Reference genes for normalization remained unchanged during the sampling time points and have been successfully applied as reference genes to another roseobacter in our laboratory (Cude et al., 2012). All normalized *boxA* and *pcaH* RT-qPCR data were relativized to their basal expression of E-37 cells grown on 7 mM acetate according to calculations described by Hellemans *et al*. (Hellemans et al., 2007). Technical replicate errors were propagated with a truncated first-order Taylor series expansion.

Protocatechuate 3,4 dioxygenase (PcaHG) enzyme assays. Approximately 10^{10} cells were washed in a 4°C solution containing 1.5% sea salts [Sigma-Aldrich] and 50 mM Tris-acetate (pH 7.5). Rinsed and pelleted cells were suspended in 396 µl of Bugbuster Protein Extraction Reagent [Novagen, Inc., Madison, WI]. Lysozyme (4 µl) was added (to a final concentration of 0.1 ng ml^{-1}) and the cells were incubated at 30°C for 15 min. After the cell debris was removed via centrifugation (21000 x *g* for 30 min at 4°C), crude cell lysates were assayed for protocatechuate 3,4-dioxygenase (PcaHG, EC 1.13.11.3) activity by measuring the kinetic loss of protocatechuate at 290 nm with a DU 800 UV/Vis spectrophotometer [Beckman Coulter, Inc., Brea, CA] (Stanier and Ingraham, 1954). Briefly, the 1 ml assays included 400 µM protocatechuate and 50 mM Tris-acetate (pH 7.5) along with ≥3 different volumes of lysate tested below the saturation point. Protein concentrations were determined with the Coomassie Plus Protein Assay Reagent Kit [Thermo Scientific Pierce, Rockford, IL] (Bradford, 1976). The specific activity for each sample was calculated using the $\Delta\varepsilon$ of $\varepsilon_{protocatechuate}$ and $\varepsilon_{\beta\text{-}carboxymuconolactone}$ (2280 cm^{-1} M^{-1}). E-37 cells grown solely on 7 mM acetate and on 2 mM protocatechuate served as negative and positive controls.

HPLC-PDA analysis of substrate concentrations. Aromatic substrates in the spent media were separated with a Waters 2695 high-performance liquid chromatography (HPLC) instrument containing a reverse-phase 3.9 x 150 mm Novak-Pak C_{18} column [Waters Corp., Milford, MA] coupled to a Waters 2996 photodiode array (PDA) detector. For the spent MBM, an isocratic elution of 0.8 ml min^{-1} at 25°C with the mobile phase containing 30% $MeCN_{(aq)}$ and 0.07% $phosphoric\ acid_{(aq)}$ produced distinct peaks for benzoate (3.62 min) and POB (2.12 min). The same separation conditions were used for M9 spent media with the exception of increased (2.5%)

phosphoric acid$_{(aq)}$. A ten-point serial dilution curve of authentic standards was used to determine the concentration of each compound at their λ_{max} in the MBM solution that were 230.3 and 256.2 nm for benzoate and POB. The peak area of each eluate was calculated with ApexTrack's integration tool using the Empower 2 Pro software package [Waters] for each of the technical (HPLC-PDA machine) triplicates performed on each sample. Linear regression of the temporally paired substrate concentrations was used to assess the statistical correlation between the catabolic/disappearance rate of each substrate using SigmaPlot 11.0 [Systat Software, Inc., Chicago, IL].

Ash-free dry mass (AFDM) measurements. E-37 cells grown at 30°C with 200 rpm agitation in 250 ml baffled flasks containing 100 ml of 14 mM carbon in MBM were used to estimate the total carbon biomass yields on different growth substrates. Cells grown to early stationary phase on 2 mM benzoate, on 2 mM POB, and on 1 mM benzoate + 1 mM POB were harvested by centrifugation at 6000 x *g* for 10 min. To capture any cells that did not pellet, supernatants were passed through pre-ashed and -weighed glass fiber (GF/F) filters with a nominal 0.7 µm pore size [Whatman Ltd., Maidstone, ME]. Cell pellets were resuspended in 1 ml of spent media and filtered onto their respective GF/F filters. All biomass was dried on the filters for 24 h at 60°C, after which their masses were measured. The dried organic material was then combusted for 4 h at 450°C. The total carbon biomass for each population was determined taking into account the combusted filter masses. Three biological replicates were performed for each growth substrate, and a one-way ANOVA was performed with Tukey's post-hoc tests in SigmaPlot 11.0. One-tailed Student's t-tests were performed with MS Excel 2010 [Microsoft Corp., Redmond, WA].

Genome analyses. Using the NCBI Genomes database (http://blast.ncbi.nlm.nih.gov/), co-occurrences of *box* and *pca* pathways were identified in cultured taxa. Protein sequences from organisms with experimental data were used as tBLASTn queries: BenA (benzoate 1,2-dioxygenase alpha subunit; YP_046122.1) and BenB (benzoate 1,2-dioxygenase beta subunit; YP_046123.1), CatA (catechol 1,2-dioxygenase; YP_046127.1) and CatB (muconate cycloisomerase I; YP_046131.1), BoxB (benzoyl-CoA 2,3-epoxidase; Q9AIX7.1) and BoxC (2,3-epoxybenzoyl-CoA dihydrolase; Q84HH6.1), PcaG (protocatechuate 3,4-dioxygenase alpha subunit; YP_046376.1) and PcaH (protocatechuate 3,4-dioxygenase beta subunit; YP_046375.1), and AraC-type PobR (transcriptional regulator; YP_299213.1) and IclR-type PobR (transcriptional regulator; YP_046382.1).

IV. RESULTS

Sagittula stellata E-37, the representative Roseobacter selected for these studies, is a coastal seawater strain that can catabolize a variety of plant-derived sugars as well as aromatic compounds representative of lignin breakdown products (González et al., 1997; Buchan et al., 2000; Buchan et al., 2004). Furthermore, it has been shown to selectively attach to lignocellulose particles as well as transform and partially mineralize synthetic lignin (González et al., 1997). Its genome contains six ring-cleaving pathways (Newton et al., 2010) (Figure 4.1), making it an especially attractive model system to examine mixed substrate growth during aromatic compound catabolism. In order to facilitate comparisons with previous studies conducted in soil bacteria, we focused our efforts here on the structurally similar growth substrates benzoate and *p*-hydroxybenzoate.

Catabolic funneling in *Sagitulla stellata* E-37. Before addressing the primary question of mixed substrate growth, it was first useful to have a broader understanding of aromatic

compound catabolism in E-37. Experimental studies in roseobacters are largely restricted to the protocatechuate (*pca*) branch of the β-ketoadipate pathway. Activity of the ring-cleaving enzyme, protocatechuate 3,4-dioxygenase (PcaHG) has been previously shown to be inducible by growth on POB in E-37 and other roseobacters (Buchan et al., 2000). Furthermore, *pobA*, the gene encoding a hydroxylase that mediates the conversion of POB to protocatechuate, appears to be coordinately expressed with the *pca* genes in E-37 and other roseobacters (Buchan et al., 2004). To address which additional aromatic substrates capable of supporting growth are processed via the *pca* pathway, PcaHG enzyme assays were performed on extracts of E-37 grown on different substrates, only one of which (*p*-coumarate) is predicted to generate POB as a catabolic intermediate. Basal level activity (<0.04 μmol pca min^{-1} μg $protein^{-1}$) was observed for cells grown on acetate, benzoate, and phenylacetate suggesting that the *pca* pathway is not induced when provided these growth substrates. Conversely, specific activity of PcaHG was detected in cells grown on ferulate, *p*-coumarate, *p*-hydroxybenzoate, protocatechuate, quinate, and vanillate (>0.50 μmol pca min^{-1} μg $protein^{-1}$), suggesting that all of these compounds are catabolized via the *pca* pathway (Figure 4.2). In some cases, these data are supported by the presence of gene homologs to known catabolic enzymes from soil bacteria (Table 4.2 and Figure 4.2). It is not yet clear what, if any, compounds other than benzoate are degraded through the *box* pathway. Transcriptional assays of *pcaH* were also performed on E-37 grown solely on acetate, benzoate, or POB and corroborate PcaHG enzyme assay data (Figure 4.3).

Growth on benzoate induces *boxA* expression in *Sagitulla stellata* E-37. Like many alphaproteobacteria, roseobacters lack the catechol (*cat*) branch of the β-ketoadipate pathway that is a well-described route for benzoate degradation in other taxa. Instead E-37 contains the

complete complement of *box* genes encoding for enzymes that convert benzoate to the TCA cycle intermediates acetyl-CoA and succinyl-CoA (Rather et al., 2010). Growth on benzoate as a sole carbon source has been demonstrated previously in E-37 (González et al., 1997), but experimental observations linking the *box* pathway to this physiology have not yet been performed. Here, we observed increased abundance of *boxA* transcripts when the strain was grown on benzoate relative to cells grown solely on acetate or POB, indicating that E-37 uses the benzoyl-CoA pathway for benzoate catabolism (Figure 4.3). These results also suggest the *box* pathway is strongly inducible and therefore likely subject to transcriptional regulation. The approximate 5-fold difference in transcript abundance between *boxA* and *pcaH* when grown on benzoate and POB, respectively, may be the result of a combination of factors, including differences in promoter regulation and strength, as well as transcript stability. Given these differences, transcript abundance for the mixed substrate experiments described below is expressed relative to those obtained for cells grown on each substrate alone.

Simultaneous catabolism of benzoate and *p*-hydroxybenzoate. In order to assess whether there was a preferential use of either benzoate or POB in non-limiting carbon conditions, E-37 was provided a mixture of both compounds at equal concentrations (1 mM each). Cell density, transcript abundance, PcaHG enzyme activity, and substrate concentration were monitored throughout the growth cycle for each of three biological replicates (Figure 4.4). Collectively, these data provided no evidence for substrate preference. Replicate growth curves were monophasic, and while both *boxA* and *pcaH* transcript abundance are on average 58-95% of that found in cultures grown on either substrate alone, the normalized relative quantities (NRQ) of both transcript abundances were significantly higher than basal levels and remained high even as

substrate concentrations fell. Furthermore, PcaHG specific activities remained unchanged over time and were equivalent to that found in POB-grown cells. Both benzoate and POB were removed from the growth medium at approximately the same rate (Figure 4.6). However, substrate concentrations in early logarithmic phase cultures indicate that POB is removed from the medium at a slightly faster rate than benzoate during the initial growth phase.

Simultaneous aromatic catabolism confers enhanced growth rates. Growth kinetics were monitored to compare the physiology of cells simultaneously catabolizing benzoate and POB to cells grown on each substrate individually at the same carbon concentrations. The growth rate for E-37 grown on both compounds simultaneously ($\mu = 0.129$) was significantly higher ($p < 0.001$) than cells grown on benzoate ($\mu = 0.048$) and POB ($\mu = 0.048$) alone (Figure 4.5). Interestingly, there was a difference in cell yield when this strain was grown on equimolar concentrations of substrates that differ by a single hydroxyl group. Early stationary phase biomass yields of 100 ml cultures of E-37 grown on POB (69.13 ± 2.28 mg C) were significantly lower ($p < 0.015$) than benzoate-grown cells (79.57 ± 4.90 mg C), while the mixed substrate grown cells yielded an intermediary value (74.27 ± 4.89 mg C). One explanation for the observed differences in biomass of cultures grown on equimolar carbon concentrations is there is a difference in the ATP yield for the substrates. In fact, theoretical energy calculations suggest more ATP is made per molecule of benzoate proceeding through *box* than ATP per molecule of POB proceeding through *pob-pca* (Table 4.3).

Genome analysis and growth assays with other strains. To ascertain whether a diagnostic genetic signature might be evident in bacteria demonstrating the simultaneous catabolism

phenotype, a bioinformatics analysis of bacterial genomes containing the *box* and *pca* pathways was undertaken. Searches indicate that all putative *box* operons contain a gene, typically designated *boxR*, with high homology to a XRE-type transcriptional regulator, suggesting functional similarity among the bacteria harboring this pathway (data not shown). Conversely, gene synteny and genetic regulators for POB catabolism are much less conserved. A focused analysis of bacterial genomes that contain both the *box* and *pca* pathways (as of Jan 2013) revealed significant variation in the organization of genes for POB catabolism (*pobA* and the *pca* genes) and their transcriptional regulators (Table 4.4, Figure 4.7). To date, co-localization of *pobA* with the *pca* genes is unique to roseobacters, and *pobA* transcription appears to be coordinate with the *pca* genes and under control of a LysR-type regulator, denoted PcaQ (Buchan et al., 2004). Outside of the *Roseobacter* clade, gene organization indicates *pobA* transcription is mediated by the activity of an adjacently located regulatory protein-encoding *pobR*, belonging to either the AraC or IclR family (DiMarco et al., 1993; DiMarco and Ornston, 1994; Quinn et al., 2001). Of the strains analyzed, simultaneous catabolism studies have only been performed on *Cupriavidus necator* JMP 134, which possesses an AraC-type PobR protein and exhibits a substrate preference for benzoate over POB (Donoso et al., 2011).

Given the unique *pobA* gene organization identified in roseobacters, we hypothesized that the absence of a PobR homolog facilitates the growth phenotypes described here. As a first step in exploring this hypothesis, we performed additional mixed substrate growth experiments with the bacterium *Variovorax paradoxus* EPS that contains both the *box* and *pca* pathways and has an IclR-type *pobR*, as well as another roseobacter, *Ruegeria pomeroyi* DSS-3, that lacks a *pobR* homolog (Figure 4.7). It was first confirmed that both organisms could grow on the two

substrates individually (data not shown). In mixed substrate experiments, simultaneous catabolism of benzoate and POB was observed in DSS-3, however, preferential consumption of POB was observed in EPS (Figure 4.8).

V. DISCUSSION

When a bacterium is provided a mixture of growth supporting compounds, the utilization profile is generally substrate concentration dependent. Under non-limiting (replete) concentrations, sequential utilization of substrates by bacteria is typically observed, with the substrate supporting the highest growth rate receiving preference (Harder and Dijkhuizen, 1982). This response has been well studied for bacteria provided a mixture of two sugars (Harder and Dijkhuizen, 1982), but has also been demonstrated with substrate mixtures from other compound classes, including organic acids and aromatic compounds (*e.g.*, Zylstra et al., 1989; Hugouvieux-Cotte-Pattat et al., 1990). Control mechanisms that are responsible for substrate utilization hierarchies include substrate transport into the cell (Kamogawa and Kurahash.K, 1967; Nichols and Harwood, 1995), transcriptional regulation (Ampé and Lindley, 1995; Fujihara et al., 2006), or post-translational modification of enzyme(s) (Zwaig and Lin, 1966). However, these regulatory controls are often relieved in substrate limiting concentrations, when mixed substrate use is essential for growth (*e.g.*, van der Kooij et al., 1982; Heinaru et al., 2001). In fact, as oligotrophic conditions dominate the microbial landscape (Poindexter, 1981), it is possible that substrate preferences are the exception rather than the rule in most natural environments (Egli, 2010). However, we demonstrate here simultaneous catabolism of two aromatic compounds processed through separate ring-cleaving pathways in roseobacter representatives occurs under carbon replete conditions (C:N:P ratio of >30,000:59:1). Furthermore, we show that this metabolic

versatility leads to an increased growth rate, which may contribute to the competitiveness of these organisms in natural systems, particularly in coastal salt marshes, in which aromatic compounds are a significant component of the dissolved organic carbon pool (Moran and Hodson, 1990). Consistent with this hypothesis is the observation that roseobacters are one of the most abundant groups of bacteria associated with decaying *Spartina alterniflora* (Buchan et al., 2003), a primary source of lignin-derived aromatic compounds to these coastal systems (Moran and Hodson, 1994).

The cross-regulatory mechanisms resulting in substrate preference when strains are provided mixtures of POB and benzoate are complex and difficult to predict from genome sequences alone. Variations in the regulatory proteins and gene organization of each pathway can result in differences in substrate preferences (Tables 4.4 and 4.5). For example, PobR proteins belong to one of two families (AraC and IclR) and activate *pobA* transcription in response to the inducer POB (DiMarco et al., 1993; Donoso et al., 2011). Yet the regulatory mechanisms that dictate repression of *pobA* transcription in the presence of benzoate vary among the organisms possessing different PobR representatives. For example, in the soil bacterium *Cupriavidus necator* JMP134 benzoate serves as structurally similar anti-inducer to prevent AraC-type PobR-mediated expression of *pobA* (Donoso et al., 2011). Conversely, in the soil bacterium *Acinetobacter baylyi* ADP1 benzoate does not directly modulate the activity of the IclR-type PobR protein found in this strain (DiMarco et al., 1993). Instead, repression of the *pca* catabolic genes, whose products are necessary for complete degradation of POB, appears to be primarily mediated by the activities of regulatory proteins of benzoate and catechol catabolic gene loci (BenM and CatM) that upon binding an intermediate of catechol catabolism (*cis, cis* muconate)

repress transcription of the *pca* operon, which includes a gene encoding for a POB permease (Gerischer et al., 1998; Brzostowicz et al., 2003). Our observation that *V. paradoxus* EPS preferentially prefers POB over benzoate in mixed substrate experiments is consistent with the notion that the activities of IclR-type PobR proteins are not directly and negatively influenced by benzoate. However, additional studies are needed to confirm this model and to better understand the underlying regulatory mechanisms that lead to the novel POB > benzoate phenotype evidenced in this strain. The unique *pobA-pca* gene organization found in roseobacters is suggestive of a more simplified regulatory scheme for *pobA*, which may be manifested in the simultaneous catabolism phenotype described here and may indicate that POB serves as an important substrate in the environmental niches that roseobacters occupy. Furthermore, most roseobacter genomes contain multiple, potentially competing ring-cleaving pathways for the degradation of aromatic compounds (Buchan and González, 2010) and raising the possibility that the simultaneous catabolism phenotype demonstrated here is representative of group members' utilization of a broader class of aromatic compounds.

While the specific mechanism(s) that facilitate enhanced growth on mixed substrates are not clear, the answer may lie at the cell membrane. Prior studies have suggested distinct and dedicated transport mechanisms contribute to enhanced growth during mixed-substrate use of glucose and sucrose (Wood and Kelly, 1977, 1982a, b). In these previous studies, catabolic enzyme activities were unchanged in mixed substrate relative to single-substrate experiments. The temporal PcaHG activities reported here are in accordance with this conclusion (Figure 4.4), although the specific transporters of POB and benzoate in E-37 are not yet known. In other taxa, PcaK has been shown to transport POB (Harwood et al., 1994), and BenK or BenP transport

benzoate (Thayer and Wheelis, 1976; Collier et al., 1997). However, homologs to these genes are absent in the E-37 genome (GenBank Accession, AAYA00000000), as well as those of other roseobacters containing the *box* and *pca* pathways. This suggests an alternative transport system for these substrates exists in lineage members. No putative transport genes are found in the local vicinity (10 kb) of the *pobA-pca* operon in E-37. However, a candidate system for benzoate transport is found directly adjacent to the *box* operon (SSE37_24379 and SSE37_24389), which putatively encodes a dicarboxylate TRAP transporter. A thorough investigation of the POB and benzoate transport systems is needed to confirm that enhanced growth by strains simultaneously utilizing these compounds is due to the presence and activities of separate, non-competing transport systems for these substrates.

In the coastal salt marshes in which roseobacters are abundant, the dissolved organic carbon pool is highly aromatic and heterogeneous in structure and in distribution (Moran and Hodson, 1990). A successful ecological strategy for bacteria in such environments is metabolic versatility, a strategy exemplified by cultivated roseobacter representatives (*e.g.* Moran et al., 2007). The simultaneous catabolism of growth substrates demonstrated here further illustrates the flexible metabolic characteristic of this abundant group of marine bacteria.

VI. ACKNOWLEDGMENTS

We thank Dr. Paul M. Orwin of California State University, San Bernardino for providing us with the *Variovorax paradoxus* EPS strain. We are also grateful to James J. Daleiden and Professor Joseph J. Bozell for technical assistance and use of their HPLC-PDA.

C.G. and A.B. were supported as part of the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Award Number DE-SC0000997.

VII. REFERENCES

Alvarez, P.J.J., and Vogel, T.M. (1991) Substrate interactions of benzene, toluene, and *para*-xylene during microbial degradation by pure cultures and mixed culture aquifer slurries. *Applied and Environmental Microbiology* **57**: 2981-2985.

Ampé, F., and Lindley, N.D. (1995) Acetate utilization is inhibited by benzoate in *Alcaligenes eutrophus*: evidence for transcriptional control of the expression of acoE coding for acetyl coenzyme A synthetase. *Journal of Bacteriology* **177**: 5826-5833.

Bleichrodt, F.S., Fischer, R., and Gerischer, U.C. (2010) The beta-ketoadipate pathway of *Acinetobacter baylyi* undergoes carbon catabolite repression, cross-regulation and vertical regulation, and is affected by Crc. *Microbiology* **156**: 1313-1322.

Boerjan, W., Ralph, J., and Baucher, M. (2003) Lignin biosynthesis. *Annual Review of Plant Biology* **54**: 519-546.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254.

Brzostowicz, P.C., Reams, A.B., Clark, T.J., and Neidle, E.L. (2003) Transcriptional cross-regulation of the catechol and protocatechuate branches of the beta-ketoadipate pathway contributes to carbon source-dependent expression of the *Acinetobacter* sp strain ADP1 *pobA* gene. *Applied and Environmental Microbiology* **69**: 1598-1606.

Buchan, A., and González, J.M. (2010) Roseobacter. In *Handbook of hydrocarbon and lipid microbiology*. Timmis, K.N. (ed). Berlin Heidelberg: Springer-Verlag, pp. 1336-1343.

Buchan, A., Neidle, E.L., and Moran, M.A. (2004) Diverse organization of genes of the beta-ketoadipate pathway in members of the marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **70**: 1658-1668.

Buchan, A., Collier, L.S., Neidle, E.L., and Moran, M.A. (2000) Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **66**: 4662-4672.

Buchan, A., Newell, S.Y., Butler, M., Biers, E.J., Hollibaugh, J.T., and Moran, M.A. (2003) Dynamics of bacterial and fungal communities on decaying salt marsh grass. *Applied and Environmental Microbiology* **69**: 6676-6687.

Cánovas, J.L., and Stanier, R.Y. (1967) Regulation of the enzymes of the beta-ketoadipate pathway in *Moraxella calcoacetica*. I. General aspects. *European Journal of Biochemistry* **1**: 289-300.

Christian, J.R., and Anderson, T.R. (2002) Modeling DOM Biogeochemistry. In *Biogeochemistry of Marine Dissolved Organic Matter*. Hansell, D.A., and Carlson, C.A. (eds). San Diego, California: Academic Press, pp. 717-755.

Collier, L.S., Nichols, N.N., and Neidle, E.L. (1997) *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. *Journal of Bacteriology* **179**: 5943-5946.

Crawford, D.L., Pometto, A.L., and Crawford, R.L. (1983) *Streptomyces viridosporus*: isolation and characterization of a new polymeric lignin degradation intermediate. *Applied and Environmental Microbiology* **45**: 898-904.

Cude, W.N., Mooney, J., Tavanaei, A.A., Hadden, M.K., Frank, A.M., Gulvik, C.A. et al. (2012) Production of the antimicrobial secondary metabolite indigoidine contributes to

competitive surface colonization in the marine roseobacter *Phaeobacter* sp. strain Y4I. *Applied and Environmental Microbiology* **78**: 4771–4780.

Dal, S., Steiner, I., and Gerischer, U. (2002) Multiple operons connected with catabolism of aromatic compounds in *Acinetobacter* sp strain ADP1 are under carbon catabolite repression. *Journal of Molecular Microbiology and Biotechnology* **4**: 389-404.

del Giorgio, P.A., and J., D. (2003) Patterns in dissolved organic matter lability and consumption across aquatic ecosystems. In *Aquatic ecosystems: Interactivity of dissolved organic matter*. Findlay, S.E.G., and Sinsabaugh, R.L. (eds). San Diego, CA: Academic Press, pp. 399-424.

DiMarco, A.A., and Ornston, L.N. (1994) Regulation of *p*-hydroxybenzoate hydroxylase synthesis by PobR bound to an operator in *Acinetobacter calcoaceticus*. *Journal of Bacteriology* **176**: 4277-4284.

DiMarco, A.A., Averhoff, B., and Ornston, L.N. (1993) Identification of the transcriptional activator *pobR* and characterization of its role in the expression of *pobA*, the structural gene for *p*-hydroxybenzoate hydroxylase in *Acinetobacter calcoaceticus*. *Journal of Bacteriology* **175**: 4499-4506.

Donoso, R.A., Pérez-Pantoja, D., and González, B. (2011) Strict and direct transcriptional repression of the *pobA* gene by benzoate avoids 4-hydroxybenzoate degradation in the pollutant degrader bacterium *Cupriavidus necator* JMP134. *Environmental Microbiology* **13**: 1590-1600.

Egli, T. (2010) How to live at very low substrate concentration. *Water Research* **44**: 4826-4837.

Fuchs, G., Boll, M., and Heider, J. (2011) Microbial degradation of aromatic compounds - from one strategy to four. *Nature Reviews Microbiology* **9**: 803-816.

Fujihara, H., Yoshida, H., Matsunaga, T., Goto, M., and Furukawa, K. (2006) Cross-regulation of biphenyl- and salicylate-catabolic genes by two regulatory systems in *Pseudomonas pseudoalcaligenes* KF707. *Journal of Bacteriology* **188**: 4690-4697.

Gaines, G.L., Smith, L., and Neidle, E.L. (1996) Novel nuclear magnetic resonance spectroscopy methods demonstrate preferential carbon source utilization by *Acinetobacter calcoaceticus*. *Journal of Bacteriology* **178**: 6833-6841.

Gerischer, U., Segura, A., and Ornston, L.N. (1998) PcaU, a transcriptional activator of genes for protocatechuate utilization in *Acinetobacter*. *Journal of Bacteriology* **180**: 1512-1524.

González, J.M., Mayer, F., Moran, M.A., Hodson, R.E., and Whitman, W.B. (1997) *Sagittula stellata* gen. nov, sp. nov, a lignin-transforming bacterium from a coastal environment. *International Journal of Systematic Bacteriology* **47**: 773-780.

Harder, W., and Dijkhuizen, L. (1982) Strategies of mixed substrate utilization in microorganisms. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **297**: 459-480.

Harwood, C.S., and Parales, R.E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annual Review of Microbiology* **50**: 553-590.

Harwood, C.S., Nichols, N.N., Kim, M.K., Ditty, J.L., and Parales, R.E. (1994) Identification of the *pcaRKF* gene cluster from *Pseudomonas putida* - Involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. *Journal of Bacteriology* **176**: 6479-6488.

Heinaru, E., Viggor, S., Vedler, E., Truu, J., Merimaa, M., and Heinaru, A. (2001) Reversible accumulation of *p*-hydroxybenzoate and catechol determines the sequential

decomposition of phenolic compounds in mixed substrate cultivations in pseudomonads. *FEMS Microbiology Ecology* **37**: 79-89.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**: R19.

Higuchi, T. (1990) Lignin biochemistry: Biosynthesis and biodegradation. *Wood Science and Technology* **24**: 23-63.

Hugouvieux-Cotte-Pattat, N., Köhler, T., Rekik, M., and Harayama, S. (1990) Growth-phase-dependent expression of the *Pseudomonas putida* TOL plasmid pWW0 catabolic genes. *Journal of Bacteriology* **172**: 6651-6660.

Jamieson, W.D., Pehl, M.J., Gregory, G.A., and Orwin, P.M. (2009) Coordinated surface activities in *Variovorax paradoxus* EPS. *Bmc Microbiology* **9**.

Kamogawa, A., and Kurahash.K (1967) Inhibitory effect of glucose on the growth of a mutant strain of *Escherichia coli* defective in glucose transport system. *Journal of Biochemistry* **61**: 220-230.

Kirk, T.K., and Farrell, R.L. (1987) Enzymatic "combustion": the microbial degradation of lignin. *Annual Review of Microbiology* **41**: 465-505.

Mazzoli, R., Pessione, E., Giuffrida, M.G., Fattori, P., Barello, C., Giunta, C., and Lindley, N.D. (2007) Degradation of aromatic compounds by *Acinetobacter radioresistens* S13: growth characteristics on single substrates and mixtures. *Archives of Microbiology* **188**: 55-68.

McCarthy, M., Hedges, J., and Benner, R. (1996) Major biochemical composition of dissolved high molecular weight organic matter in seawater. *Marine Chemistry* **55**: 281-297.

Moran, M.A., and Hodson, R.E. (1990) Contributions of degrading *Spartina alterniflora* lignocellulose to the dissolved organic carbon pool of a salt marsh. *Marine Ecology Progress Series* **62**: 161-168.

Moran, M.A., and Hodson, R.E. (1994) Dissolved humic substance of vascular plant origin in a coastal marine environment. *Limnology and Oceanography* **39**: 762-771.

Moran, M.A., Belas, R., Schell, M.A., González, J.M., Sun, F., Sun, S. et al. (2007) Ecological genomics of marine roseobacters. *Applied and Environmental Microbiology* **73**: 4559-4569.

Mou, X., Sun, S., Edwards, R.A., Hodson, R.E., and Moran, M.A. (2008) Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**: 708-711.

Newton, R.J., Griffin, L.E., Bowles, K.M., Meile, C., Gifford, S., Givens, C.E. et al. (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME Journal* **4**: 784-798.

Nichols, N.N., and Harwood, C.S. (1995) Repression of 4-hydroxybenzoate transport and degradation by benzoate: a new layer of regulatory control in the *Pseudomonas putida* beta-ketoadipate pathway. *Journal of Bacteriology* **177**: 7033-7040.

Ornston, L.N., and Stanier, R.Y. (1966) Conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida*. I. Biochemistry. *Journal of Biological Chemistry* **241**: 3776-3786.

Poindexter, J.S. (1981) Oligotrophy. Fast and famine existence. *Advances in Microbial Ecology* **5**: 63-89.

Quinn, J.A., McKay, D.B., and Entsch, B. (2001) Analysis of the *pobA* and *pobR* genes controlling expression of *p*-hydroxybenzoate hydroxylase in *Azotobacter chroococcum*. *Gene* **264**: 77-85.

Rather, L.J., Knapp, B., Haehnel, W., and Fuchs, G. (2010) Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. *Journal of Biological Chemistry* **285**: 20615-20624.

Romero-Steiner, S., Parales, R.E., Harwood, C.S., and Houghton, J.E. (1994) Characterization of the *pcaR* regulatory gene from *Pseudomonas putida*, which is required for the complete degradation of *p*-hydroxybenzoate. *Journal of Bacteriology* **176**: 5771-5779.

Selje, N., Simon, M., and Brinkhoff, T. (2004) A newly discovered *Roseobacter* cluster in temperate and polar oceans. *Nature* **427**: 445-448.

Siehler, S.Y., Dal, S., Fischer, R., Patz, P., and Gerischer, U. (2007) Multiple-level regulation of genes for protocatechuate degradation in *Acinetobacter baylyi* includes cross-regulation. *Applied and Environmental Microbiology* **73**: 232-242.

Stanier, R.Y., and Ingraham, J.L. (1954) Protocatechuic acid oxidase. *Journal of Biological Chemistry* **210**: 799-808.

Stanier, R.Y., and Ornston, L.N. (1973) The beta-ketoadipate pathway. *Advances in Microbial Physiology* **9**: 89-151.

Suzuki, M.T., Preston, C.M., Chavez, F.P., and DeLong, E.F. (2001) Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay. *Aquatic Microbial Ecology* **24**: 117-127.

Thayer, J.R., and Wheelis, M.L. (1976) Characterization of a benzoate permease mutant of *Pseudomonas putida*. *Archives of Microbiology* **110**: 37-42.

van der Kooij, D., Oranje, J.P., and Hijnen, W.A.M. (1982) Growth of *Pseudomonas aeruginosa* in tap water in relation to utilization of substrates at concentrations of a few micrograms per liter. *Applied and Environmental Microbiology* **44**: 1086-1095.

Wakeham, S.G., Schaffner, C., and Giger, W. (1980) Polycyclic aromatic hydrocarbons in Recent lake sediments-I. Compounds having anthropogenic origins. *Geochimica Et Cosmochimica Acta* **44**: 403-413.

Wang, Y., Yang, J., Lee, O.O., Dash, S., Lau, S.C.K., Al-Suwailem, A. et al. (2011) Hydrothermally generated aromatic compounds are consumed by bacteria colonizing in Atlantis II Deep of the Red Sea. *ISME Journal* **5**: 1652-1659.

Wheelis, M.L., and Ornston, L.N. (1972) Genetic control of enzyme induction in the β-ketoadipate pathway of *Pseudomonas putida*: deletion mapping of *cat* mutations. *Journal of Bacteriology* **109**: 790-795.

Wood, A.P., and Kelly, D.P. (1977) Heterotrophic growth of *Thiobacillus* A2 on sugars and organic acids. *Archives of Microbiology* **113**: 257-264.

Wood, A.P., and Kelly, D.P. (1982a) Kinetics of sugar transport by *Thiobacillus* A2. *Archives of Microbiology* **131**: 156-159.

Wood, A.P., and Kelly, D.P. (1982b) Mechanisms of sugar transport by *Thiobacillus* A2. *Archives of Microbiology* **131**: 160-164.

Zwaig, N., and Lin, E.C.C. (1966) Feedback inhibition of glycerol kinase, a catabolic enzyme in *Escherichia coli*. *Science* **153**: 755-757.

Zylstra, G.J., Olsen, R.H., and Ballou, D.P. (1989) Cloning, expression, and regulation of the *Pseudomonas cepacia* protocatechuate 3,4-dioxygenase genes. *Journal of Bacteriology* **171**: 5907-5914.

APPENDIX

VIII. TABLES

Table 4.1. Oligonucleotide primers used in RT-qPCR.

Name [a b]	**Sequence (5' to 3')**
*alaS*97F	GATCCGACGCTTATGTTCGT
*alaS*247R	TGTAACCGACGTTGTCCAGA
*boxA*508F	GTGCTCGAAGGTCAGTCCAT
*boxA*694R	CCTTCAGGTCGCACAGGTAG
*map*620F	GCATGTTCTTCACCATCGAG
*map*785R	GCGGGAGAGAGGGTAAAGAT
*pcaH*167F	AGGACAACGACCTGATCACC
*pcaH*333R	TTCCTTCTTGTGGCGGTAAC
*rpoC*340F	GCCCATATCTGGTTCCTCAA
*rpoC*498R	TTCCTCTTCGGTCAGCATCT

[a]The first three or four italicized letters represent the gene targeted.

[b]Suffixes F and R denote the forward and reverse primers, respectively.

Table 4.2. Phenotypic and genetic evidence for aromatic compound catabolism in *S. stellata* E-37.

Carbon Substrate	Concentration [mM]	Growth[b]	PcaHG Specific Activity[c]	Catabolic Gene(s) in E-37	Locus Tag(s)[f]
Acetate	3.5, 7, and 25	+	0.017±0.008	tricarboxylic acid cycle (*gltA*, *acnA*, *icd*, *sucABCD*, *shdABCD*, *fumC*, & *mdh*)	SSE37_06879, SSE37_25138, SSE37_06274, SSE37_11324, SSE37_11329, SSE37_11294, SSE37_11309, SSE37_11244, SSE37_11229, SSE37_11254, SSE37_11249, SSE37_05510, SSE37_11289
Benzoate	0.5, 1, and 2	+	0.030±0.016	*bcl-boxDCBAE*	SSE37_24404, SSE37_24409, SSE37_24419, SSE37_24424, SSE37_24439, SSE37_24444
Caffeate	1,2, and 3	-	n.d.	*hcaABC*	SSE37_01000, SSE37_01050, SSE37_12324
Catechol	1,2, and 3	-	n.d.	*xylE*	SSE37_12149
Chlorogenate	1,2, and 3	-	n.d.	*hcaG*, *hcaABC*, & *pobA*	SSE37_18822, SSE37_01000, SSE37_01050, SSE37_12324, SSE37_18837
Ferulate	2	+	(+)[d]	unk[d] & *vanAB*	SSE37_02815, SSE37_00265
Gentisate	1,2, and 3	-	n.d.	*nagILK*	SSE37_02745[g], SSE37_02750, SSE37_02755
Homogentisate	n.d.[a]	n.d.	n.d.	*hmgAB* & *hmgC*	SSE37_24544, SSE37_24549, SSE37_14734
Homoprotocatechuate	n.d.	n.d.	n.d.	*hpaBC* & *hpaDEF*	SSE37_00435, SSE37_00425, SSE37_23379, SSE37_23374, SSE37_23369
p-coumarate	2	+	1.205±0.702	*hcaC* & *pobA*	SSE37_12324, SSE37_18837
p-hydroxybenzoate	0.5, 1, and 2	+	0.477±0.051	*pobA*	SSE37_18837
Phenylacetate	2	+	0.005±0.017	*paaABCDEZ* & *paaIJK*	SSE37_00390, SSE37_00385, SSE37_00380, SSE37_00375, SSE37_00370, SSE37_00360, SSE37_15758, SSE37_15753, SSE37_15748
Protocatechuate	2	+	0.632±0.141	*pcaCHG*, *pcaDB*, *pcaIJ*, & *pcaF*	SSE37_18842, SSE37_18847, SSE37_18852, SSE37_21470, SSE37_21460, SSE37_23014, SSE37_23019, SSE37_16533
Quinate	2	+	0.694±0.140	*qsuBCD*	SSE37_03985, SSE37_06077, SSE37_03980
Salicylate	1,2, and 3	-	n.d.	*salA*	SSE37_01015
Vanillate	2	+	(+)	*vanAB*	SSE37_02815, SSE37_00265

a. n.d. = not determined.

b. Growth was assessed by monitoring the O.D. at 540 nm of cultures in MBM.

c. PcaHG activity was scored as "-" or "+" if the specific activity was <0.04 or >0.50 µmol pca min^{-1} µg $protein^{-1}$, respectively.

d. (+) indicates specific activity was not quantified but levels were >0.50 µmol pca min^{-1} µg $protein^{-1}$

e. unk = Gene(s) involved in the pathway are unknown.

f. Locus tags unique to the E-37 genome are listed in respective order of the catabolic genes.

g. A disparately located *nagI* duplicate exists (SSE37_02425).

Table 4.3. Stoichiometric reaction summaries with theoretical net ATP yields for different aerobic routes to degrade benzoate and *p*-hydroxybenzoate.[e] [f]

Substrate	Pathway	Expected Taxonomic Range[a]	Reactants	Key Intermediate(s)	Products	ATP Yield[d]
Benzoate	**Benzoyl-CoA (*box*)**	***Actinobacteria, Proteobacteria***	**1 Benzoate + 2 CoA + 1 ATP + 1 O_2 + 3 H_2O**	**1 Acetyl-CoA + 1 Formate[b] + 1 Succinyl-CoA**	**1 AMP + 1 H^+ + 1 PPi + 1 NADH**	**23**
Benzoate	β-Ketoadipate, Catechol (*cat*) Branch (*ortho*-cleavage)	Fungi, *Proteobacteria*	1 Benzoate + 1 CoA + 2 O_2 + 1 H_2O	1 Acetyl-CoA + 1 Succinate	1 CO_2 + 1 H^+	17
Benzoate	β-Ketoadipate, Protocatechuate (*pca*) Branch (*ortho*-cleavage)	*Bacteria*	1 Benzoate + 1 CoA + 3 O_2 + 2 NADPH	1 Acetyl-CoA + 1 Succinate	2 $NADP^+$ + 1 H_2O	11
Benzoate	Catechol I (*meta*-cleavage)	*Actinobacteria*, *Proteobacteria*	1 Benzoate + 1 CoA + 2 O_2 + 2 H_2O + 1 NAD^+	1 Acetyl-CoA + 1 Formate + 1 Pyruvate[c]	1 CO_2 + 2 H^+ + 1 NADH	33
Benzoate	Catechol II (*meta*-cleavage)	*Actinobacteria*, *Proteobacteria*	1 Benzoate + 1 CoA + 2 O_2 + 2 H_2O + 2 NAD^+	1 Acetyl-CoA + 1 Pyruvate	2 CO_2 + 2 H^+ + 2 NADH	33
Benzoate	Protocatechuate (*meta*-cleavage)	*Proteobacteria*	1 Benzoate + 3 O_2 + 1 NADPH	2 Pyruvate	1 CO_2 + 1 $NADP^+$	27
***p*-Hydroxybenzoate**	**β-Ketoadipate, Protocatechuate (*pca*) Branch (*ortho*-cleavage)**	***Bacteria***	**1 POB + 1 CoA + 2 O_2 + 1 NADPH**	**1 Acetyl-CoA + 1 Succinate**	**2 CO_2 + 1 H^+ + 1 $NADP^+$ + 1 H_2O + 2 $FADH_2$ + 4 NADH + 1 GTP**	**14**

Substrate	Pathway	Expected Taxonomic Range[a]	Reactants	Key Intermediate(s)	Products	ATP Yield[d]
p-Hydroxybenzoate	Protocatechuate (*meta*-cleavage)	*Proteobacteria*	1 POB + 2 O_2 + 1 H_2O + <u>1 NADPH</u>	<u>2 Acetyl-CoA</u> + <u>2 Pyruvate</u>	7 CO_2 + 3 H^+ + <u>2 $FADH_2$</u> + <u>8 NADH</u> + <u>2 GTP</u>	30
p-Hydroxybenzoate	Protocatechuate (*para*-cleavage)	Bacilli	1 POB + 1 CoA + 2 O_2 + 1 H_2O + <u>1 NADPH</u> + 2 NAD^+	<u>2 Acetyl-CoA</u> + <u>1 Pyruvate</u>	7 CO_2 + 2 H^+ + <u>2 $FADH_2$</u> + <u>9 NADH</u> + <u>2 GTP</u>	33

a. Retrieved from http://www.biocyc.org

b. Formate dehydrogenation yields 1 NADH.

c. Pyruvate dehydrogenation yields 1 Acetyl-CoA and 1 NADH.

d. Net ATP yields per molecule of substrate were calculated from the underlined products going through the TCA cycle using 3 ATP/NAD(P)H, 2 ATP/$FADH_2$, 1 ATP/GTP.

e. *box* and *pca* pathways present in E-37 are bolded.

f. Underlined compounds were used for energy calculations.

Table 4.4. List of bacterial strains predicted to contain both *box* and *pca* pathways.[a]

Taxonomy	Strain	Isolation Source	*box*	*pca*	PobR[c]	*ben*[d]	*cat*
Actinobacteria	*Pseudonocardia dioxanivorans* CB1190	Soil	+	+	n.p.	+	+
Actinobacteria	*Streptomyces* sp. AA4	Soil	+	+	n.p.	+	+
Alphaproteobacteria	Bradyrhizobiaceae bacterium SG-6C	Soil	+	+	n.p.	-	-
Alphaproteobacteria	*Bradyrhizobium japonicum* USDA 110	Soil	+	+	AraC	-	-
Alphaproteobacteria	*Roseobacter* sp. GAI101	Coastal Marine	+	+	AraC	-	-
Alphaproteobacteria	***Jannaschia* sp. CCS1**	**Coastal Marine**	**+**	**+**	**n.p.**	-	-
Alphaproteobacteria	***Ruegeria pomeroyi* DSS-3**	**Coastal Marine**	**+**	**+**	**n.p.**	-	-
Alphaproteobacteria	***Sagittula stellata* E-37**[d]	**Coastal Marine**	**+**	**+**	**n.p.**	-	-
Alphaproteobacteria	**Rhodobacterales bacterium KLH11**	**Coastal Marine**	**+**	**+**	**n.p.**	-	-
Betaproteobacteria	*Achromobacter piechaudii* ATCC 43553	Host, Human airways	+	+	n.p.	+	+
Betaproteobacteria	*Bordetella petrii* DSM 12804	Fresh water	+	+	n.p.	+	+
Betaproteobacteria	*Burkholderia phymatum* STM815	Soil	+	+	AraC	+	+
Betaproteobacteria	*Burkholderia phytofirmans* PsJN	Soil	+	+	n.p.	+	+
Betaproteobacteria	*Burkholderia* sp. CCGE1001	Soil	+	+	n.p.	+	+
Betaproteobacteria	*Burkholderia* sp. CCGE1002	Soil	+	+	AraC	+	+
Betaproteobacteria	*Burkholderia* sp. CCGE1003	Soil	+	+	n.p.	-	+
Betaproteobacteria	*Burkholderia xenovorans* LB400	Soil	+	+	n.p.	+	+
Betaproteobacteria	*Cupriavidus metallidurans* CH34	Fresh water, Soil	+	+	AraC	+	+
Betaproteobacteria	*Cupriavidus necator* JMP134[d]	Fresh water, Soil	+	+	AraC	+	+
Betaproteobacteria	*Cupriavidus necator* N-1	Fresh water, Soil	+	+	AraC	+	+
Betaproteobacteria	*Ralstonia eutropha* H16	Fresh water, Soil	+	+	AraC	+	+
Betaproteobacteria	*Polaromonas* sp. JS666	Fresh water	+	+	IclR	-	+
Betaproteobacteria	*Variovorax paradoxus* EPS	Soil	+	+	IclR	-	-
Betaproteobacteria	*Variovorax paradoxus* S110	Soil	+	+	IclR	-	-

[a]Predicted or [b]experimentally shown taxa that simultaneously catabolize POB and benozate are bolded; non-bolded are predicted to have a substrate preference phenotype.

[c]Predicated PobR family; n.p. = not present to indicate a *pobR* homolog was not found in the immediate vicinity of *pobA*.

[d]Evidence for presence of alternate route for benzoate degradation via catechol.

Table 4.5. Different genotypes proposed to have different phenotypes for aerobic growth in a benzoate + POB mixture.

Phenotype	benzoate > POB	benzoate > POB	benzoate > POB	benzoate > POB	benzoate > POB	benzoate < POB	benzoate = POB
Genotype	*ben*/*cat*, *box*, *pca*	*ben*/*cat*, *pca*	*ben*/*cat*, *pca*	*ben*/*cat*, *box*, *pca*	*box*, *pca*	*box*, *pca*	*box*, *pca*
PobR Family	AraC	AraC	IclR (or absent)	IclR (or absent)	AraC	IclR	(absent)
Regulation Mechanism(s)	PobR binds benzoate and represses *pobA*, consequently no *pca* induction	PobR binds benzoate and represses *pobA*, consequently no *pca* induction -AND- BenM & CatM bind to promoter-operator region of *pcaU* to repress *pca* transcription; PCA accumulation consequently represses *pob* activity	BenM & CatM bind to promoter-operator region of *pcaU* to repress *pca* transcription; PCA accumulation consequently represses *pob* activity	BenM & CatM bind to promoter-operator region of *pcaU* to repress *pca* transcription; PCA accumulation consequently represses *pob* activity	PobR binds benzoate and represses *pobA*, consequently no *pca* induction	unknown	no "cross-talk" strong enough to repress or induce other pathway
Source(s)	(Donoso et al., 2011)	based off (Brzostowicz et al., 2003), (Donoso et al., 2011), and (Gerischer et al., 1998)	(Brzostowicz et al., 2003) and (Gerischer et al., 1998)	based off (Brzostowicz et al., 2003) and (Gerischer et al., 1998)	based off (Donoso et al., 2011)	this study	this study
Model Strain	*Cupriavidus necator* JMP134[a b]	unknown	*Acinetobacter baylyi* ADP1[a b]	*Burkholderia xenovorans* LB400[b]	*Bradyrhizobium japonicum* USDA 110[b]	*Variovorax paradoxus* EPS [a b]	*Sagittula stellata* E-37[a b]

[a]experimentally verified phenotype.

[b]genome is publicly available.

IX. FIGURES

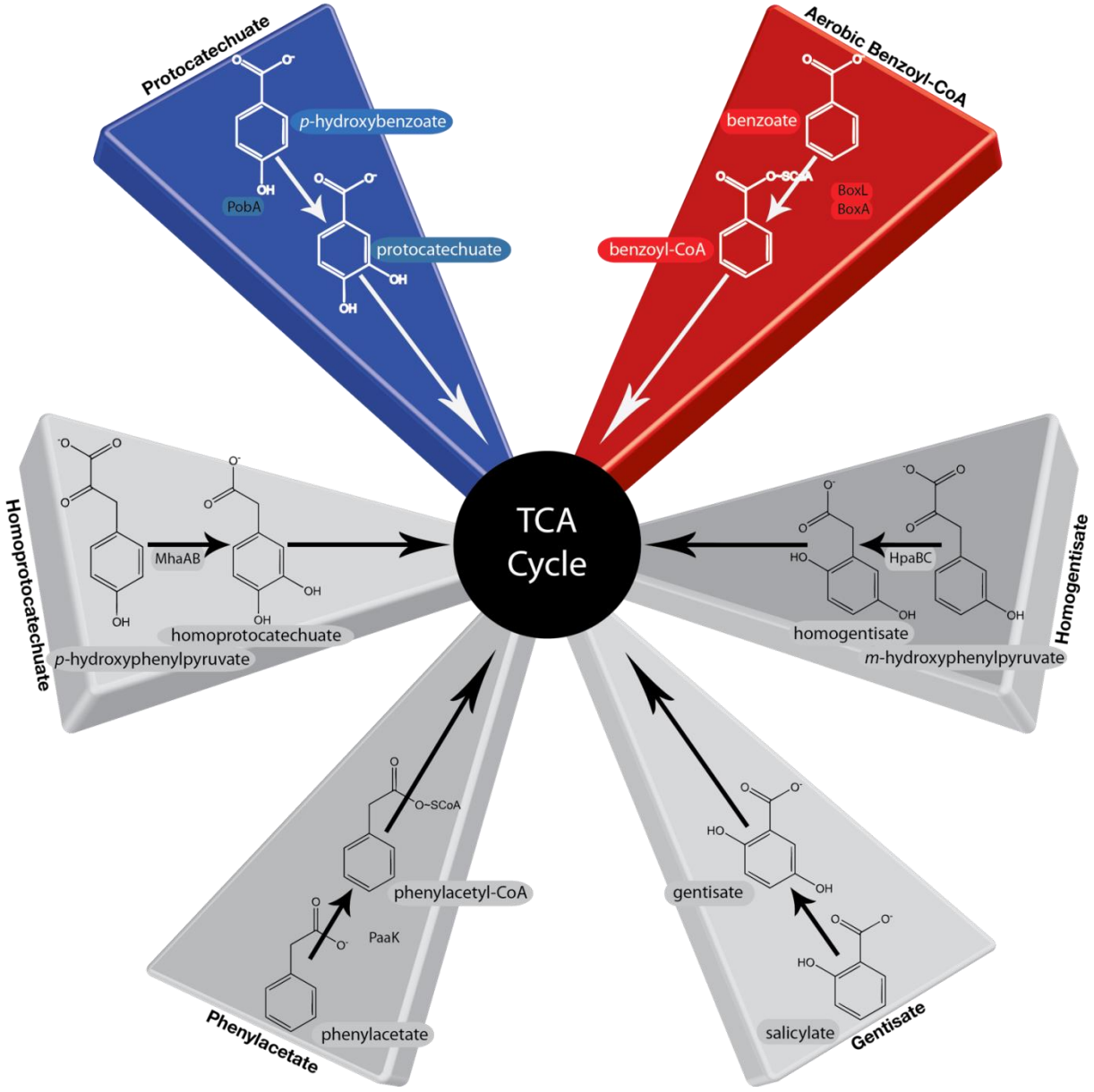


Figure 4.1. Six ring-cleaving pathways present in *Sagittula stellata* E-37. TCA, tricarboxylic acid.

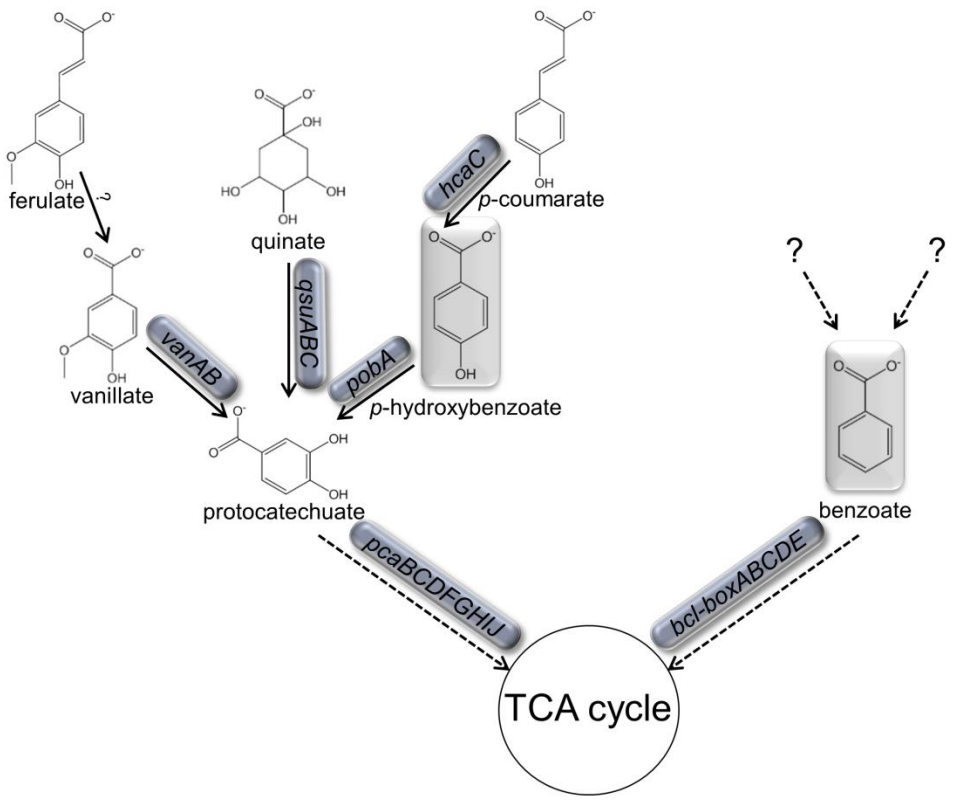


Figure 4.2. Substrates proceeding through the protocatechuate branch of the β-ketoadipate and the aerobic benzoyl CoA pathway in *Sagittula stellata* E-37. Gene designations are based on sequence homology to characterized proteins. Question marks denote uncharacterized genes or substrates. TCA, tricarboxylic acid.

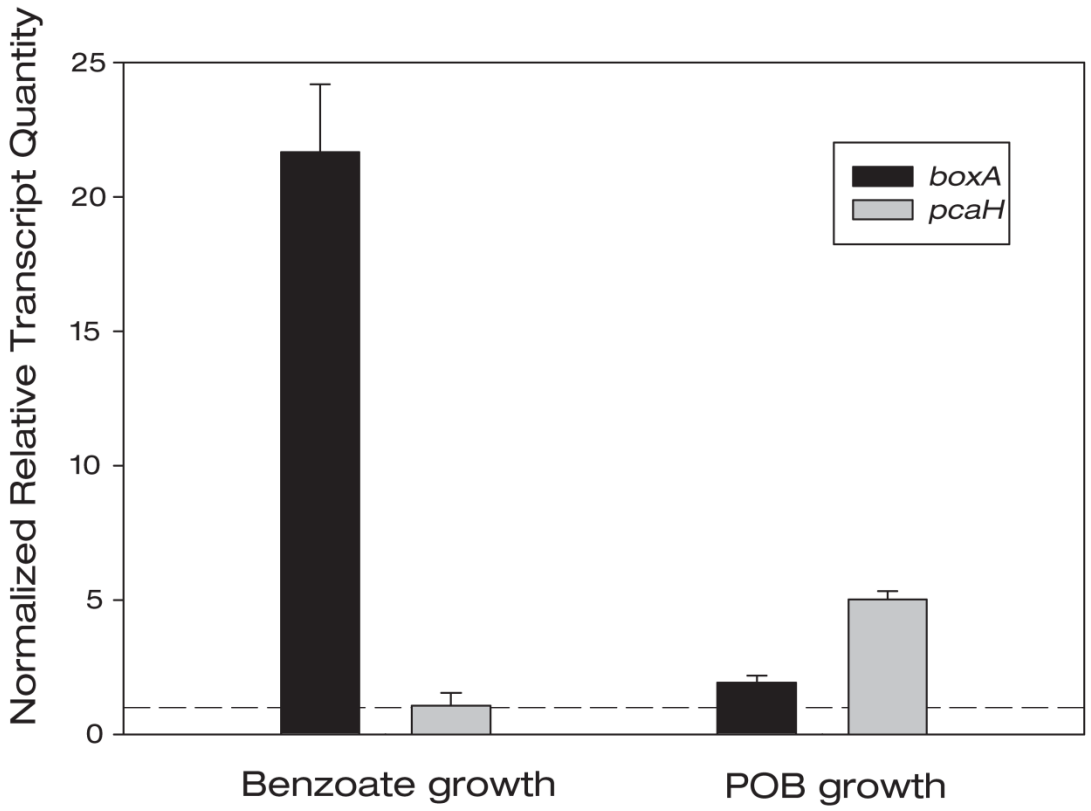


Figure 4.3. *boxA* and *pcaH* transcript abundance for E-37 grown solely on POB (2 mM) or benzoate (2 mM). cDNA copy numbers were relativized to those from cells grown on an equimolar C concentration of acetate (7 mM). The horizontal dashed line represents unchanged expression. Error bars represent the standard error of the mean of biological triplicates.

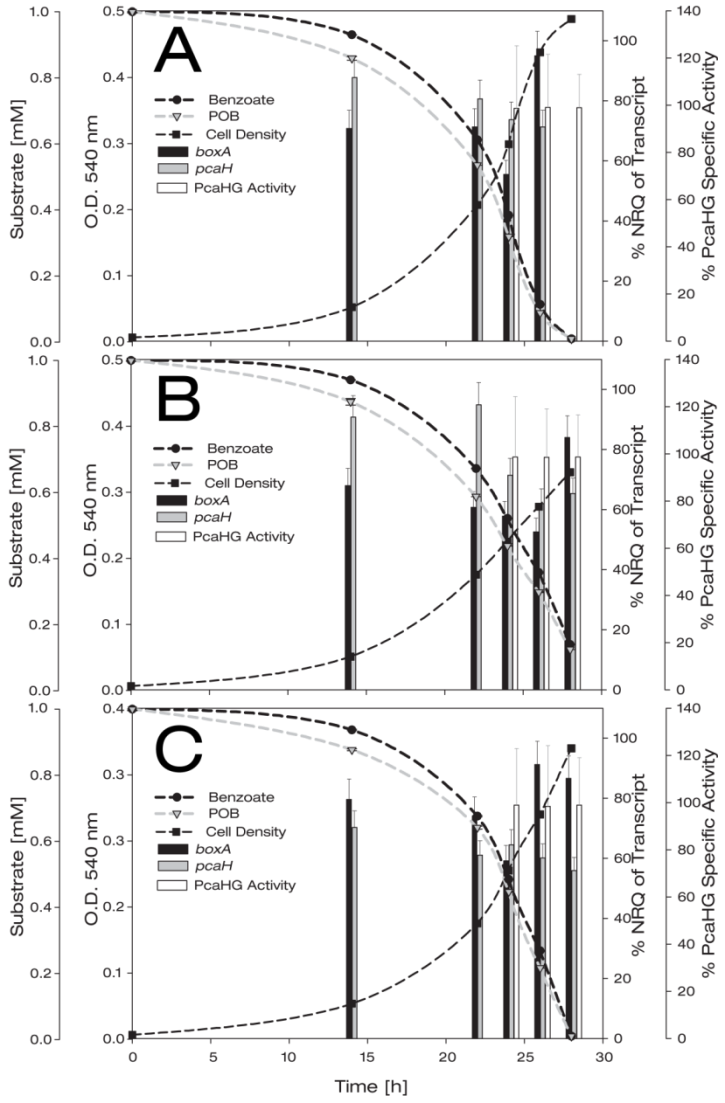


Figure 4.4. Simultaneous catabolism of benzoate (1 mM) and POB (1 mM) by E-37. Panels A-C represent three biological replicates. Data for all three biological replicates are shown on separate panels (A-C) due to slight variation in growth phase and sample time points among parallel cultures. The *boxA* and *pcaH* transcript abundances are expressed as a percentage of those when E-37 was grown solely on benzoate or POB, respectively. Error bars represent the standard error of the mean (n = 3 per gene). PcaHG specific activity is also expressed as the percentage of activity obtained from POB-grown cells; error bars represent standard deviation of at least 3 replicates. Standard deviation of the technical variation (n = 3) for benzoate and POB concentrations are smaller than the symbol. NRQ, Normalized Relative Quantity. O.D., Optical Density.

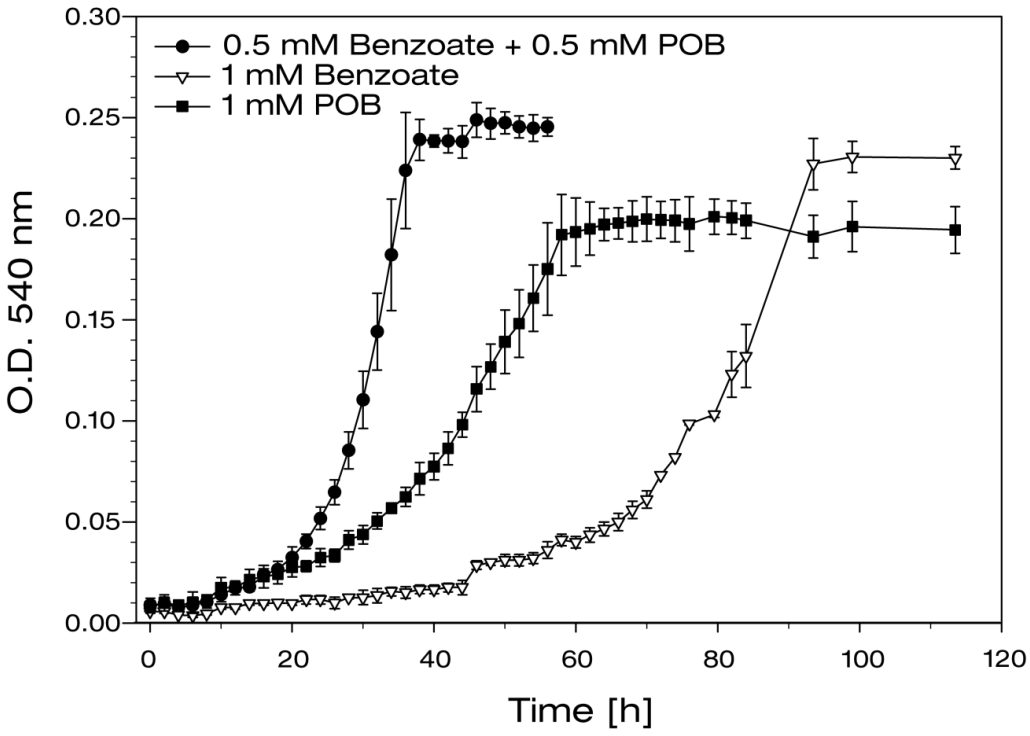


Figure 4.5. Growth responses to benzoate (1 mM), POB (1 mM), and a mixture of benzoate (0.5 mM) and POB (0.5 mM) by *S. stellata* E-37. Error bars for O.D. and substrate concentrations represent the standard deviation of biological replicates (n = 3) for each. O.D., Optical Density.

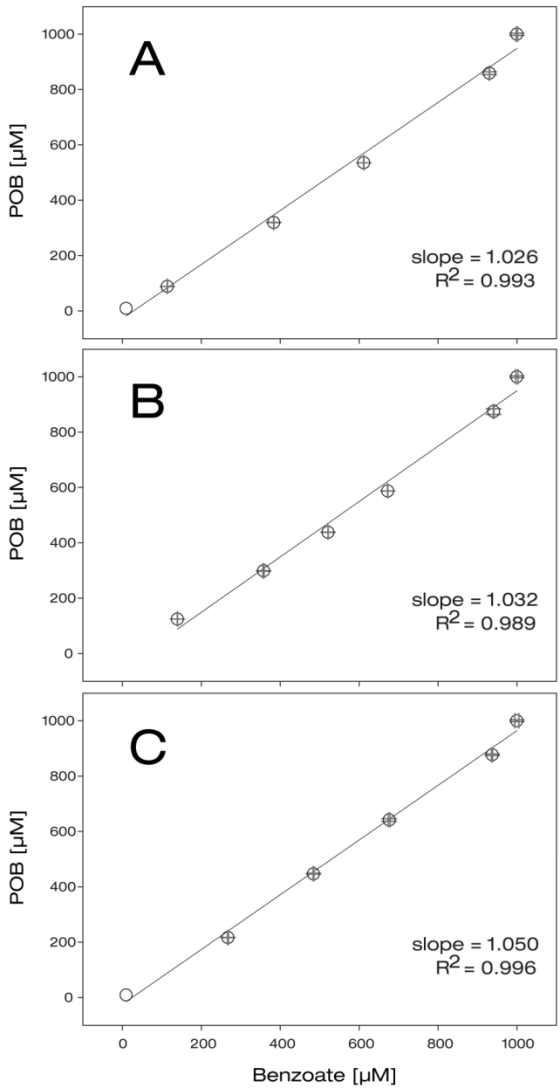


Figure 4.6. Linear regression of temporally paired POB and benzoate concentrations derived from Figure 4.4 of *S. stellata* E-37 during simultaneous catabolism. Error bars represent the standard deviation of triplicate technical replicates.

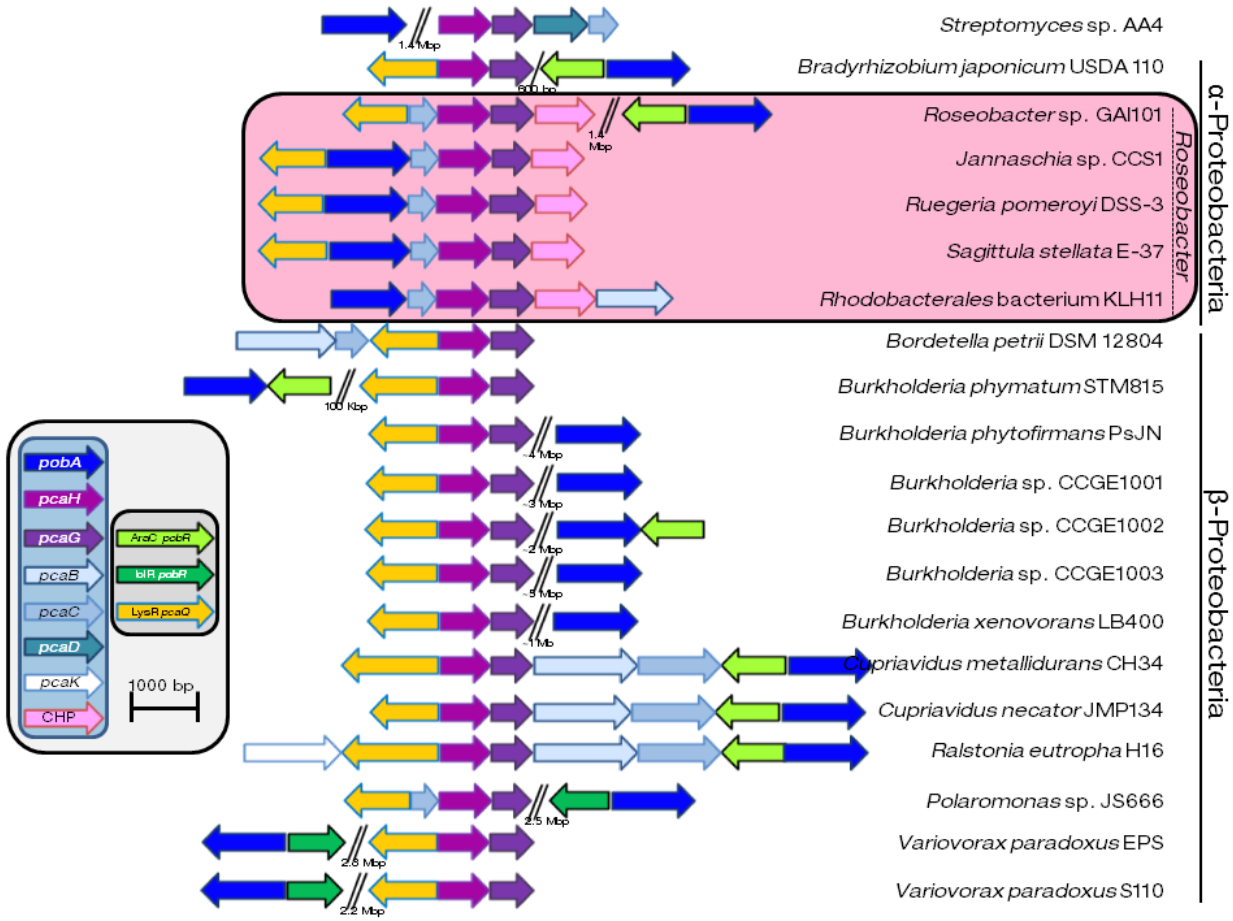


Figure 4.7. Gene neighborhoods of *pob* and *pca* operons from representative taxa. All neighborhoods are centered around *pcaHG* and drawn to scale. Each color represents a unique COG number (with the exception of PcaG and PcaH): PobA (COG0654), PcaH (COG3485), PcaG (COG3485), PcaB (COG0015), PcaC (COG0599), PcaD (COG0596), PcaK (COG2271), AraC-type PobR (COG2207), IclR-type PobR (COG1414), LysR-type PcaQ (COG0583) and conserved hypothetical protein [CHP] in *Roseobacter* (COG3246).

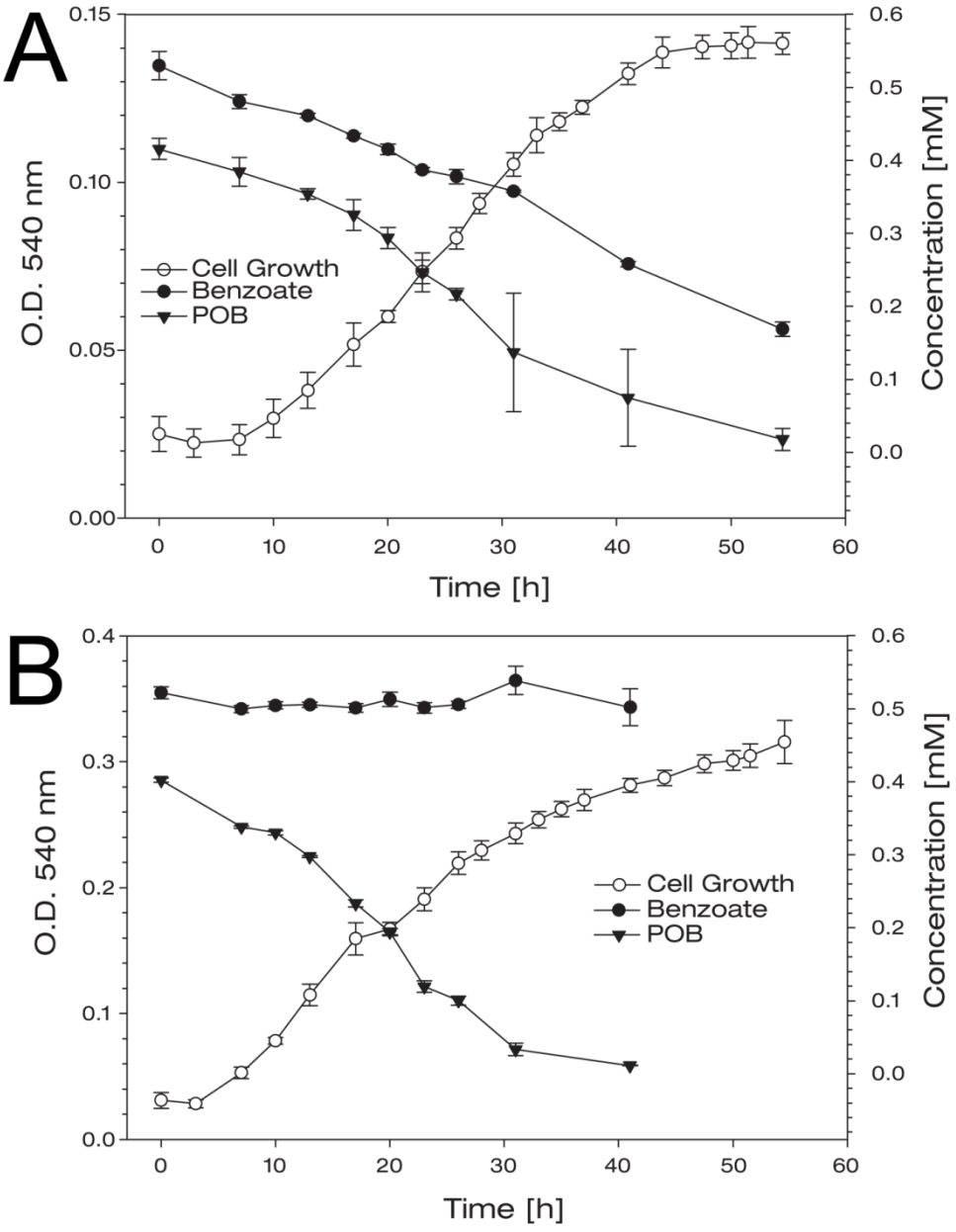


Figure 4.8. Simultaneous catabolism of benzoate and POB by (A) *R. pomeroyi* DSS-3 and preferential POB consumption by (B) *V. paradoxus* EPS. Error bars for O.D. and substrate concentrations represent the standard deviation of biological replicates (n = 3) for each. O.D., Optical Density.

CHAPTER 5 -

HYPERINDUCTION OF MULTIPLE CATABOLIC AND CENTRAL METABOLISM PATHWAYS OF A ROSEOBACTER IN A MODELED SALT MARSH ENVIRONMENT CONTAINING CORDGRASS-DERIVED DOC

I. ABSTRACT

Members of the *Roseobacter* lineage are abundant throughout the world's oceans but are most prevalent in coastal salt marshes where they can comprise upwards of 30% of the bacterioplankton population. Salt marshes in the southeastern United States are dominated by the lignocellulose-containing *Spartina alterniflora* (cordgrass), and a variety of microorganisms are responsible for remineralization of *Spartina* detritus. Fauna (*e.g.*, crustacea and insects) shred cordgrass material into smaller pieces, while fungi are involved in enzymatically attacking the structural components into smaller molecular weight compounds that some bacteria then completely degrade. An abundant and active microbial lignocellulose decomposer of cordgrass is *Phaeosphaeria spartinicola*. In this study, this fungus was grown in a cordgrass-containing medium to degrade and release a mixture of compounds representative of a salt marsh system. We fed this material to the roseobacter *Sagittula stellata* E-37 to determine which pathways are involved in the degradation of compounds likely originating from cordgrass. Transcriptomics profiling of *S. stellata* during growth on the cordgrass-degraded DOC mixture demonstrated that many genes for degradation pathways were over-represented, relative to acetate-grown cultures, in cDNA libraries. Numerous ring-cleaving genes are predicted to be present in roseobacter genomes and we now show evidence that several genes were expressed (*hcaB*, 24-fold; *hpaB*, 41-fold; *nagI*, 2-fold) from distinct pathways (*e.g.*, gentisate [*nag*], homoprotocatechuate [*hpa*], and hydroxycinnamate [*hca*]) during growth on cordgrass material. Carbohydrate metabolism also appeared highly active, as suggested by a 320-fold overabundance of alpha-amylase (*amyA*) gene transcripts in cordgrass-DOC fed cultures. The mixed-substrate utilizing phenotype appears to have a consequence of overexpressed central metabolism genes as well, such as the tricarboxylic acid cycle (4-fold for *icd*), ATPase (8-fold for *atpA*), and NADH dehydrogenase (3-

fold for *nuoA*). The hyperactive phenotype suggested by this work provides a physiological understanding of mixed-substrate use and highlights a trophic strategy for a generalist marine bacterium.

II. INTRODUCTION

The enormous diversity of carbonaceous compounds on Earth is influenced by the various mechanisms that produce them and require the activity of numerous enzymes for biogeochemical cycling of carbon. Regardless of the source, the dissolved organic carbon (DOC) pool containing aromatic compounds in nature is typically structurally heterogeneous (Crawford et al., 1983; Christian and Anderson, 2002). This is an important consideration for microbial degradation and has received significant attention in studies examining the catabolism of mixtures of aromatic compounds classified as environmental pollutants (Grbić-Galić and Vogel, 1987; Arvin et al., 1989; Alvarez and Vogel, 1991; Reardon et al., 2000). However, considerably less attention has been paid to the microbial catabolism of mixtures of naturally occurring aromatic compounds, such as those derived from lignin, the structural component of vascular plant tissue (Kirk and Farrell, 1987). Lignin is chemically complex, and during its biologically mediated degradation a variety of compounds are produced that are available as primary growth substrates for microorganisms capable of their catabolism.

In the southeastern United States coastal salt marshes, the primary producer is *Spartina alterniflora*, a lignocellulose-containing cordgrass (Pomeroy et al., 1981; Currin et al., 1995). Over the past few decades this cordgrass has become prevalent in estuaries along the western United States (Callaway and Josselyn, 1992) as well as the gulf coasts of Northern America (Blum et al., 2007). In these systems, decomposition of lignocellulose is temporally partitioned, with fungi serving as the initial decomposers followed by bacteria (Schubauer and Hopkinson, 1984). The dominant fungus found on standing dead *S. alterniflora* leaves is *Phaeosphaeria spartinicola* (Newell et al., 1989; Leuchtmann and Newell, 1991). This ascomyceteous fungus

can degrade the polysaccharide (*i.e*., cellulose and hemicellulose) and lignin components of plant material, yielding oxidation products that serve as primary growth substrates for bacteria (*e.g*., *p*-coumarate, ferulate, *p*-hydroxybenzoate, vanillate, and vanillin) (Bergbauer and Newell, 1992).

Bacterial catabolism of aromatic compounds is often initialized by upper, or peripheral, pathways that transform a diverse suite of aromatic compounds into one of a limited number of intermediates that are then subject to ring cleavage (Fuchs et al., 2011). Bacteria capable of utilizing aromatic compounds as primary growth substrates frequently possess multiple ring-cleaving pathways that are typically under strong transcriptional regulation. Contrary to the established paradigm of substrate preference, we recently demonstrated that the salt marsh bacterium *Sagitulla stellata* E-37 is able to simultaneously catabolize two aromatic compounds present in nature as a result of lignin degradation (Chapter 4). In fact, *S. stellata* displays enhanced growth on a mixture of these two chemically similar compounds over each provided individually at equivalent concentrations. The purpose of this study was to determine which pathways this roseobacter uses during growth in natural, chemically complex mixtures of DOC representative of the salt marsh system it inhabits.

III. MATERIALS AND METHODS

Generation of Cordgrass-Derived DOC

The ascomycetous fungus *Phaeosphaeria spartinicola* 9005 (Leuchtmann and Newell, 1991) was always cultured at 25°C in the dark. This strain has a lignin-degrading (laccase, EC 1.10.3.2) phenotype as confirmed using a colorimetric assay based on transformation of syringaldazine [Sigma-Aldrich Corp., St. Louis, MO] as previously described (Lyons et al., 2003). The fungus

was first grown in a complex and rich yeast extract (0.2% w/v), malt extract (2% w/v), and sea salts (1.5% w/v) (YMSS) medium for 14 d, then transferred into a glucose sea salts (G-SS) minimal medium containing 0.3% (w/v) D-glucose [Thermo Fisher Scientific, Hampton, NH], 2% (w/v) sea salts [Sigma], 23.5 mM $NaNO_3$, 10.3 mM KH_2PO_4, and 17.2 mM K_2HPO_4 (Bergbauer and Newell, 1992; Lyons et al., 2003).

Spartina alterniflora (cordgrass) was harvested off the coast of Charleston, South Carolina and dried in the sun for 1 mo. Dried material was ground with a Model 4 Wiley Mill [Thomas Scientific, Swedesboro, NJ] and collected through a U.S. Standard No. 20 (850 µm cutoff) sieve [Fisher]. The cordgrass was sterilized with ethylene oxide and added to autoclaved G-SS. Once the fungi acclimated to the G-SS minimal medium (30 d), 4-5 ml (approximately 60 mg of dry fungal biomass) was used to inoculate 300 ml volumes of 1% (w/v) cordgrass-amended G-SS (CG-SS) media in 1 liter baffled flasks. The fungi were allowed to degrade the cordgrass for 35 d, after which they were removed via 0.22 µm Express PLUS filtration [EMD Millipore Corp., Billerica, MA]. For controls, a fungal inoculated G-SS medium (no cordgrass) and a sterile (fungus-free) CG-SS medium were incubated in parallel for 35 d. All glassware for fungal and bacterial growth in this study was minimally combusted for 6 h at 450°C to remove organic carbon prior to media preparation.

Sagitulla stellata E-37 Growth in Cordgrass-Derived DOC.

S. stellata was grown in each of the fungal-degraded CG-SS biological triplicates. A 1 liter volume of marine basal medium (MBM) (Chapter 4) supplemented with 10% (v/v) fungal-degraded CG-SS (hereafter referred to as cordgrass-derived DOC) was provided to E-37 (n = 3)

in 2 liter flasks. During growth at 30°C with 200 rpm agitation for 28 h, O.D. at 540 nm was used to monitor cell density and temporal subsamples were harvested for analysis: preserved cells for gene expression, washed cells for enzyme assays, and spent media for chemical characterization.

Approximately $1x10^8$ cells were harvested at each time point for transcriptional studies. Multiple technical replicates were collected at each time point for each biological replicate and were preserved with RNA*later* [Invitrogen, Carlsbad, CA]. E-37 cells were resuspended in 1 ml volumes of RNA*later*, incubated at room temperature for 1 h, then pelleted at 13100 *x g* for 15 min at 4°C. The solution was removed and cells were flash-frozen in liquid nitrogen. Preserved cells were maintained at -70°C until nucleic acid extraction. For baseline expression values, *S. stellata* cells were also grown to corresponding growth phases and cell densities in 7 mM acetate-MBM (Chapter 4), and preserved in the same way.

RNA-seq Transcriptomics.

RNA was isolated from late exponential (14 h growth, $O.D._{540\ nm} = 0.350$) *S. stellata* cells as previously described (Chapter 4). An additional column-based cleanup was performed to increase RNA extract quality with Zymo-Spin IC columns [Zymo Research Corp., Irvine, CA] and Qiagen's EB buffer as eluate. RNA quality was assessed with a ND-1000 NanoDrop, HS RNA Qubit assay with a Qubit 2.0 fluorometer [Invitrogen], and a RNA 600 chip [Agilent Technologies, Santa Clara, CA] with a 2100 Bioanalyzer [Agilent]. Depletion of rRNA was accomplished with the Ribo-Zero Gold magnetic kit [Epicentre Biotechnologies, Madison, WI]. Library preparation involved first and second strand synthesis with 500 ng RNA, and a unique

heptamer barcoding sequence was used for each sample. Prior to loading the barcoded cDNAs for sequencing, each library was quantified with qPCR using the KAPA Library Quantification Kit [Kapa Biosystems, Woburn, MA]. Single-end sequencing was performed by Hudson-Alpha [Huntsville, AL] using a HiSeq 2500 [Illumina, Inc., San Diego, CA] and yielded approximately 25 M reads for each technical sequencing replicate (n = 2 or 3), equating to ~50-75 M reads per biological replicate. For comparison, *S. stellata* has roughly a 5 Mbp genome with ~5500 genes.

RNA-seq Analysis.

For RNA-seq data, two independent approaches were performed. CLC Genomics Workbench v6.0.1 [CLC Bio, Århus, Denmark] and a combination of open-source packages (hereafter referred to as "OSA" for open source analysis) which use a proprietary alignment algorithm and the Burrows-Wheeler alignment algorithm (Li and Durbin, 2009, 2010) (respectively). Data normalization was also different in these two approaches—RPKM (reads per kilobase per million mapped reads [Mortazavi et al., 2008]) for CLC and DESeq (Anders and Huber, 2010) for OSA. To determine the KEGG orthology (KO) numbers, the E-37 GenBank file was retrieved from Roseobase (http://www.roseobase.org/) with GBrowse, and it was converted with a custom perl script into a multi-protein FastA file containing each protein's corresponding locus tag. KEGG Automatic Annotation Server v2.0 (Moriya et al., 2007) used 10 *Rhodobacteraceae* family genomes (sil [*Ruegeria pomeroyi* DSS-3], sit [*Ruegeria* sp. TM1040], jan [*Jannaschia* sp. CCS1], pga [*Phaeobacter gallaeciensis* DSM 17395], pgl [*Phaeobacter gallaeciensis* 2.10], rsq [*Rhodobacter sphaeroides* ATCC 17025], rsk [*Rhodobacter sphaeroides* KD131], rcp [*Rhodobacter capsulatus*], rde [*Roseobacter denitrificans*], dsh [*Dinoroseobacter shibae*]) in bi-directional best hit BLASTp queries to link *S. stellata* locus tags to KO numbers. The KO

numbers along with gene expression data were used in iPath2.0 for visualization (Yamada et al., 2011).

PcaHG Enzyme Activity Assays.

Approximately 10^{10} *S. stellata* cells were harvested at 8, 10, 14, 18, and 28 h of CG-SS growth to determine the activity of protocatechuate 3,4-dioxygenase (PcaHG, EC 1.13.11.3) as previously described (Chapter 4). Cells were harvested and rinsed in a 4°C cold 50 mM Tris-acetate with 1.5% (w/v) sea salts [Sigma-Aldrich] at pH 7.5 to remove aromatic compounds, and a lysozyme (0.1 ng ml^{-1}) with Bugbuster solution [Novagen, Inc., Madison, WI] was used to lyse the cell pellets. Crude lysate activities of removing protocatechuate (A290 nm) were determined with a DU 800 UV/Vis spectrophotometer [Beckman Coulter, Inc., Brea, CA]. A Bradford assay (Coomassie Plus Protein Assay Reagent Kit [Thermo Scientific Pierce, Rockford, IL]) was used to determine protein concentrations for specific activity calculations. *S. stellata* cells grown to late exponential phase in 14 mM acetate-MBM and in 3 mM protocatechuate-MBM were used as negative and positive controls, respectively.

Cordgrass-DOC Characterization.

All spent media samples were harvested, passed through a 0.22 µm filter, and stored at 4°C. Total DOC was quantified with a GC-MS.

IV. RESULTS AND DISCUSSION

S. stellata E-37 can grow on a variety of cordgrass-derived compounds (González et al., 1997; Buchan et al., 2000; Buchan et al., 2004), and N, P, and S cycle transcripts best recruited to the

S. stellata E-37 genome suggesting closely related organisms are among the most active microorganisms in a salt marsh system for these biogeochemical transformation activities (Gifford et al., 2011). We therefore selected *S. stellata* as a model roseobacter isolate based on its catabolic capacity (Newton et al., 2010) and its relevance to salt marsh systems. To determine the activity of a roseobacter isolate, *S. stellata*, in transforming compounds derived from the abundant cordgrass *S. alterniflora*, we developed a model salt marsh system in the laboratory. With this system we hoped to model a multi-phylum interaction and gain insight into a particular environmentally relevant function—roseobacter mineralization of cordgrass-derived DOC (Figure 5.1).

Prior to individual gene expression analysis of *S. stellata* grown in the cordgrass-derived DOC mixture, several quality assessments were performed using acetate grown cells for basal expression. In order to assess the quality of the normalization procedures, the RPKM normalized expression of all ~5500 genes were compared among technical and biological replicates, represented as a boxplot (Figure 5.2). The gene expression medians across technical and biological replicates are similar after RPKM normalization within each technical sequencing replicate. Overall, a wider range of expression values were observed across CS-SS grown cells compared to growth in acetate, which is likely due to the relatively simple catabolic functions performed during acetate metabolism having only the tricarboxylic acid (TCA) catabolic route invoked. Validation of gene expression data via RT-qPCR with primers developed (Table 5.1) will strengthen this and other conclusions drawn, however two fundamentally different transcriptomics analysis revealed the same trends described throughout this study. Variation among the CG-SS biological replicates was much more pronounced than in acetate growth

(Figure 5.3). Due to the anticipated chemical complexity in terms of the variety of carbonaceous compounds released from *P. spartinicola* during degradation of cordgrass, the CG-SS media themselves were slightly different (Table 5.2) and thus transcriptional responses of *S. stellata* were observed to be more variable than growth in acetate. Even with such biological variation captured during growth in the cordgrass-derived DOC, 3237 genes had significantly changed ($p<0.05$) RPKM expression values compared to acetate-grown cells shown in Figure 5.4. With such physiological variation observed, a focused analysis of metabolic pathways was performed to determine which pathways are used during growth of *S. stellata* in cordgrass-derived DOC.

Multiple ring-cleaving pathways are invoked.

The *P. spartinicola* fungi were initially confirmed to have a lignin-degrading phenotype with a laccase assay (Figure 5.5), thus smaller molecular weight aromatics were expected to be produced in the cordgrass-degraded DOC mixtures fed to *S. stellata*. Lignin degradation by *P. spartinicola* yields a variety of aromatic compounds (Bergbauer and Newell, 1992), and multiple ring-cleaving pathways are predicted in *S. stellata* E-37 (Newton et al., 2010). Therefore, we studied the transcriptomes for a variety of aromatic degradation genes to determine if any were invoked. The gentisate pathway requires the activity of three protein-encoding genes (*nagIJK*) (Zhou et al., 2001), however *S. stellata* contains a second *nagI* (SSE37_02425) which is separate and distantly located from the other three genes that occur adjacent to each other. The *nagIJK* cluster were all significantly overexpressed ($p<0.0005$) in CG-SS-grown cells with approximately 2-fold overexpression (Table 5.3). Previous studies have found the gentisate pathway to be an important ring-fission route for the degradation of environmental contaminants (*e.g.*, Grund et al., 1992; Rong et al., 2009), however *nagIJK* also appear to be used in the

degradation of plant material. Culturing *S. stellata* on gentisate in MBM has never been fruitful in our laboratory. While *S. stellata* might not be able to use gentisate as a sole carbon source, a mixed substrate utilizing phenotype affords a cell's ability to invoke additional pathways to make use of other growth substrates that are not growth sustaining by themselves (Ihssen and Egli, 2005). NagI acts by itself as a dimer to cleave gentisate (Lack, 1959), and substituted gentisates are also catabolized by this pathway, so it is also possible that *S. stellata* responded to these substrates. The duplicated *nagI* SSE37_02425 gene is highly divergent (36% amino acid identity and 51% similarity) from the other, which suggests like other taxa with more than one *nagI* (Jeon et al., 2006; Lee et al., 2011) it may have additional roles in assuaging gentisate toxicity. Salicylate monooxygenase (SalA) converts salicylate into gentisate, and *salA* was significantly ($p<0.005$) induced in *S. stellata* by 29-fold (Table 5.3). A similar (20-fold) expression value was observed in an *Acinetobacter* strain during salicylate catabolism (Jung et al., 2011), suggesting salicylate is metabolized by *S. stellata*. Whether gentisate or salicylate are directly taken up by *S. stellata* in a salt marsh environment is unknown. It is also possible these are only generated intracellularly during catabolism of other substituted gentisates and salicylates.

Several other genes involved in aromatic catabolism were overexpressed during *S. stellata* growth in the cordgrass-derived DOC. The protocatechuate (*pca*) pathway is a well-documented route for the aerobic degradation of many lignin-derived aromatics (Harwood and Parales, 1996). The *pca* genes essential for protocatechuate degradation in *S. stellata* are split into four separate regions across the chromosome, and the ring-fission encoding *pcaHG* genes were only modestly overexpressed (Table 5.3). *S. stellata* invokes only a 5-fold overexpression when grown on a

pcaHG-inducing substrate (Chapter 4), so we considered an enzyme assay to further estimate how active this pathway was during growth in cordgrass-derived DOC. Activity assays indicate significant PcaHG activity throughout the time series compared to activity measured during acetate growth (Table 5.4). Lower activities than protocatechuate-only grown cells are at least partially due to additional pathways invoked, which likely resulted in more non-PcaHG proteins per cell and therefore lower specific activities.

Central metabolic pathways are overexpressed.

Multiple catabolic pathways were invoked in *S. stellata* E-37, so we considered whether the sum of this mixed-substrate utilizing phenotype activity also increased central metabolism activity. Alternatively, *S. stellata* E-37 could maintain a similar level of central metabolism, so we targeted the TCA cycle and oxidative phosphorylation genes. Fourteen genes involved in the TCA cycle were targeted for expression analysis (*gltA, acnA, icd, sucABCCD, sdhABCD, fum,* and *mdh*). *S. stellata* contains a duplication of *sucC,* and the gene (SSE37_22250) distantly located from the other *suc* genes was underexpressed when grown in the cordgrass-derived DOC. The 13 other genes were overexpressed, and all but two had significance values of $p<0.05$ (Table 5.3). To further investigate increased induction of central metabolic pathways, we considered the genes involved in oxidative phosphorylation. All of the 48 genes we considered as part of the five complexes used in oxidative phosphorylation were overexpressed, and all but two genes had $p<0.05$ significance values. In particular, ATPase genes (*atpABCDEFGHI*) averaged a 7-fold (± 2.3) overinduction. Similarly, the NADH dehydrogenase operon (*nuoABCDEFGHIJKLMN*) had a 4-fold (± 0.9) overexpression with the highest p-value being 0.002. Collectively these data suggest *S. stellata* cells were hyperactive in central metabolic pathway induction, especially

those involved in ATP and NADH biosynthesis during growth in cordgrass-derived DOC (Figure 5.6, Table 5.3). Intracellular pools of ATP and NADH were not quantified due to complications with flux analysis, because additional energy equivalents would be required to fuel the mixed-substrate utilizing phenotype observed in the cordgrass-degraded DOC.

Carbohydrate metabolism.

The marine fungus *P. spartinicola* is known to act on cellulose and hemicellulose components of cordgrass, so an abundance of sugars was expected to have been released during the 35 day pre-exposure period. The fungi also use these for growth substrates themselves (Bergbauer and Newell, 1992), thus it is unknown in what abundance these compounds would be present in the mixture. The glycolysis pathway contains several reversible steps that can complicate interpretation of transcriptional and translational activities. For example, *pgi* (encoding for the bidirectional glucose-6-phosphate isomerase) was dramatically underexpressed in *S. stellata* grown on the cordgrass DOC relative to acetate-grown cultures (Table 5.3). Irreversible reactions encoded by genes (*pdhABC*) were underrepresented in the transcript pool indicating pyruvate fermentation was not significantly invoked. Energetically committed (due to ATP hydrolysis), the three irreversible reactions encoded by genes (*glk*, *pfk*, and *pyk*) in glycolysis were all significantly ($p<0.005$) overexpressed, suggesting glycolysis was indeed used during growth in the cordgrass-derived DOC. Further evidence that carbohydrate metabolism occurred in tandem with aromatic degradation was other sugar degradation genes invoked. For example, the α-amylase encoding gene (*amyA*, SSE37_00800) was among the top 50 most expressed genes observed (Table 5.5). Seven other genes that surround *amyA* were also highly induced, three of

which are putative transporter encoding genes and one is a hypothetical protein of unknown function.

Unknown function genes induced and repressed.

There were many genes we were unable to determine putative functions for with BLAST queries, but it appears some hypothetical protein encoding genes (*e.g.*, SSE37_00815) might be assigned a putative phenotype given the proximity to other annotated genes. While the exact function of SSE37_00815 remains unknown, it seems likely it is involved in metabolism of large α-linked carbohydrates. There is another group of genes (SSE37_21570, SSE37_21575, and SSE37_21580) adjacent to each other that were highly expressed (17 ± 0.3), but the role of these could not be estimated. Similarly two other groups of genes all encoding hypothetical proteins were observed being significantly ($p<2.8\times10^{-12}$) underrepresented in the transcript pool (Table 5.6). These strong expression changes suggest such genes do serve metabolic functions, although it is also possible they happen to be under the transcriptional regulation whereby other function genes are transcribed. Experimental work such as targeted mutagenesis is required to determine their functions.

Conclusion.

Here we show a representative roseobacter isolate grown in cordgrass-derived DOC invokes multiple catabolic pathways and leads to overactive central metabolic pathways such as the TCA cycle and oxidative phosphorylation. Prior to this study, we have only observed the activity of two ring-cleaving pathways (*box* and *pca*) in the roseobacter *S. stellata* E-37 (Chapter 4). The transcriptomics data in this study indicate several additional aromatic degradation genes are

induced during growth in the cordgrass-derived DOC pool, which may provide leads for identifying roles of specific functional pathways that relate to degradation of aromatic and carbohydrate components of plant material. Growth activity of roseobacters in salt marsh systems is much higher compared to other abundant marine bacterial groups such as SAR11 (Campbell and Kirchman, 2013), and this hyperactive physiology of a roseobacter member is perhaps a fundamental physiological reasoning for such findings. Evidence is emerging that indicate even functional genes can be differentiated to roseobacters apart from other taxa (Chapter 2), and computational tools (*e.g.*, Gifford et al., 2011; Luo et al., 2012) are being developed which will be useful in future research addressing the extent to which this increased activity in roseobacters occurs in the environment.

V. ACKNOWLEDGEMENTS

C.A.G. and A.B. were supported in part from the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Award Number DE-SC0000997 and from a Microbiology across Campuses Educational and Research Venture (M-CERV) grant.

VI. REFERENCES

Alvarez, P.J.J., and Vogel, T.M. (1991) Substrate interactions of benzene, toluene, and *para*-xylene during microbial degradation by pure cultures and mixed culture aquifer slurries. *Applied and Environmental Microbiology* **57**: 2981-2985.

Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biology* **11**.

Arvin, E., Jensen, B.K., and Gundersen, A.T. (1989) Substrate interactions during aerobic biodegradation of benzene. *Applied and Environmental Microbiology* **55**: 3221-3225.

Bergbauer, M., and Newell, S.Y. (1992) Contribution to lignocellulose degradation and DOC formation from a salt marsh macrophyte by the aseomycete *Phaeosphaeria spartinicola*. *FEMS Microbiology Ecology* **86**: 341-347.

Blum, M.J., Bando, K.J., Katz, M., and Strong, D.R. (2007) Geographic structure, genetic diversity and source tracking of *Spartina alterniflora*. *Journal of Biogeography* **34**: 2055-2069.

Buchan, A., Neidle, E.L., and Moran, M.A. (2004) Diverse organization of genes of the beta-ketoadipate pathway in members of the marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **70**: 1658-1668.

Buchan, A., Collier, L.S., Neidle, E.L., and Moran, M.A. (2000) Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **66**: 4662-4672.

Callaway, J.C., and Josselyn, M.N. (1992) The introduction and spread of smooth cordgrass (*Spartina alterniflora*) in South San Francisco Bay. *Estuaries* **15**: 218-226.

Campbell, B.J., and Kirchman, D.L. (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME Journal* **7**: 210-220.

Christian, J.R., and Anderson, T.R. (2002) Modeling DOM Biogeochemistry. In *Biogeochemistry of Marine Dissolved Organic Matter*. Hansell, D.A., and Carlson, C.A. (eds). San Diego, California: Academic Press, pp. 717-755.

Crawford, D.L., Pometto, A.L., and Crawford, R.L. (1983) Lignin degradation by *Streptomyces viridosporus*: Isolation and characterization of a new polymeric lignin degradation intermediate. *Applied and Environmental Microbiology* **45**: 898-904.

Currin, C.A., Newell, S.Y., and Paerl, H.W. (1995) The role of standing dead *Spartina alterniflora* and benthic microalgae in salt marsh food webs: considerations based on multiple stable isotope analysis. *Marine Ecology Progress Series* **121**: 99-116.

Fuchs, G., Boll, M., and Heider, J. (2011) Microbial degradation of aromatic compounds - from one strategy to four. *Nature Reviews Microbiology* **9**: 803-816.

Gifford, S.M., Sharma, S., Rinta-Kanto, J.M., and Moran, M.A. (2011) Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME Journal* **5**: 461-472.

González, J.M., Mayer, F., Moran, M.A., Hodson, R.E., and Whitman, W.B. (1997) *Sagittula stellata* gen. nov, sp. nov, a lignin-transforming bacterium from a coastal environment. *International Journal of Systematic Bacteriology* **47**: 773-780.

Grbić-Galić, D., and Vogel, T.M. (1987) Transformation of toluene and benzene by mixed methanogenic cultures. *Applied and Environmental Microbiology* **53**: 254-260.

Grund, E., Denecke, B., and Eichenlaub, R. (1992) Naphthalene degradation via salicylate and gentisate by *Rhodococcus* sp. strain B4. *Applied and Environmental Microbiology* **58**: 1874-1877.

Harwood, C.S., and Parales, R.E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annual Review of Microbiology* **50**: 553-590.

Ihssen, J., and Egli, T. (2005) Global physiological analysis of carbon- and energy-limited growing *Escherichia coli* confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. *Environmental Microbiology* **7**: 1568-1581.

Jeon, C.O., Park, M., Ro, H.S., Park, W., and Madsen, E.L. (2006) The naphthalene catabolic (*nag*) genes of *Polaromonas naphthalenivorans* CJ2: Evolutionary implications for two gene clusters and novel regulatory control. *Applied and Environmental Microbiology* **72**: 1086-1095.

Jung, J., Madsen, E.L., Jeon, C.O., and Park, W. (2011) Comparative genomic analysis of *Acinetobacter oleivorans* DR1 to determine strain-specific genomic regions and gentisate biodegradation. *Applied and Environmental Microbiology* **77**: 7418-7424.

Kirk, T.K., and Farrell, R.L. (1987) Enzymatic "combustion": the microbial degradation of lignin. *Annual Review of Microbiology* **41**: 465-505.

Lack, L. (1959) The enzymatic oxidation of gentisic acid. *Biochimica Et Biophysica Acta* **34**: 117-123.

Lee, H.J., Kim, J.M., Lee, S.H., Park, M., Lee, K., Madsen, E.L., and Jeon, C.O. (2011) Gentisate 1,2-dioxygenase, in the third naphthalene catabolic gene cluster of *Polaromonas naphthalenivorans* CJ2, has a role in naphthalene degradation. *Microbiology-SGM* **157**: 2891-2903.

Leuchtmann, A., and Newell, S.Y. (1991) *Phaeosphaeria spartinicola*, a new species on *Spartina*. *Mycotaxon* **41**: 1-7.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li, H., and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589-595.

Luo, H.W., Löytynoja, A., and Moran, M.A. (2012) Genome content of uncultivated marine Roseobacters in the surface ocean. *Environmental Microbiology* **14**: 41-51.

Lyons, J.I., Newell, S.Y., Buchan, A., and Moran, M.A. (2003) Diversity of ascomycete laccase gene sequences in a southeastern US salt marsh. *Microbial Ecology* **45**: 270-281.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**: W182-W185.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621-628.

Newell, S.Y., Fallon, R.D., and Miller, J.D. (1989) Decomposition and microbial dynamics for standing, naturally positioned leaves of the saltmarsh grass *Spartina alterniflora*. *Marine Biology* **101**: 471-481.

Newton, R.J., Griffin, L.E., Bowles, K.M., Meile, C., Gifford, S., Givens, C.E. et al. (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME Journal* **4**: 784-798.

Pomeroy, L.R., Darley, W.M., Dunn, E.L., Gallagher, J.L., Haines, E.B., and Whitney , D.M. (1981) Primary production. In *The ecology of a salt marsh*. Pomeroy, L.R., and Wigert, R.G. (eds): Springer, pp. 39-67.

Reardon, K.F., Mosteller, D.C., and Rogers, J.D.B. (2000) Biodegradation kinetics of benzene, toluene, and phenol as single and mixed substrates for *Pseudomonas putida* F1. *Biotechnology and Bioengineering* **69**: 385-400.

Rong, L., Guo, X.Q., Chen, K., Zhu, J.C., Li, S.P., and Jiang, J.D. (2009) Isolation of an isocarbophos-degrading strain of *Arthrobacter* sp. scl-2 and identification of the degradation pathway. *Journal of Microbiology and Biotechnology* **19**: 1439-1446.

Schubauer, J.P., and Hopkinson, C.S. (1984) Above- and belowground emergent macrophyte production and turnover in a coastal marsh ecosystem, Georgia. *Limnology and Oceanography* **29**: 1052-1065.

Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., and Bork, P. (2011) iPath2.0: interactive pathway explorer. *Nucleic Acids Research* **39**: W412-W415.

Zhou, N.Y., Fuenmayor, S.L., and Williams, P.A. (2001) *nag* genes of *Ralstonia* (formerly *Pseudomonas*) sp strain U2 encoding enzymes for gentisate catabolism. *Journal of Bacteriology* **183**: 700-708.

APPENDIX

VII. TABLES

Table 5.1. Oligonucleotide primers used for RT-qPCR expression validation. Locus tags and gene names of the target are listed next to each primer pair.

E-37 Locus Target	Gene	Forward Primer (5' to 3)	Reverse Primer (5' to 3')
SSE37_05000	*alaS*	GATCCGACGCTTATGTTCGT	TGTAACCGACGTTGTCCAGA
SSE37_13553	*map*	GCATGTTCTTCACCATCGAG	GCGGGAGAGAGGGTAAAGAT
SSE37_15096	*rpoC*	GCCCATATCTGGTTCCTCAA	TTCCTCTTCGGTCAGCATCT
SSE37_00800	*amyA*	GACGAAAACGGCTACGACAT	CTCGTTTTTCACCCACTGGT
SSE37_19837	*atpA*	AAGGTGTCGGCCTATGAATG	AAGGTGTCGGCCTATGAATG
SSE37_01050	*hcaB*	TCATCGAGATGTGCTGGAAG	TCATCGAGATGTGCTGGAAG
SSE37_00435	*hpaB*	ACTACTGGAACCACGCCATC	CCGTAGTGGGCAATGAAGTT
SSE37_02745	*nagI*	CGACACGATCCTCGAGTACA	GAGAGGTTCTCGTGGGAATG
SSE37_03720	*nuoA*	AAGGTGTCGGCCTATGAATG	TCATCGAGATGTGCTGGAAG
SSE37_18847	*pcaH*	AGGACAACGACCTGATCACC	TTCCTTCTTGTGGCGGTAAC
SSE37_03985	*qsuB*	CCTGGTCCTGATGTGTTCCT	GTCGAGGATCAGACCGATGT
SSE37_12149	*xylE*	GGATCAACCGTATCGACCAC	CCTTTTTCAGGATGCTCTCG

Table 5.2. Total organic carbon (TOC) concentrations determined for media after 35 d of *Phaeosphaeria spartinicola* 9005 growth.

Medium	Biological Replicate #	TOC (ppm)
CG-SS	1	1522.71
	2	1564.71
	3	1532.16
G-SS	1	1149.54
	2	1121.61

Table 5.3. Differential expression of select metabolic pathways in *S. stellata* E-37.

Pathway	Gene[a]	Locus Tag[b]	Expression (RPKM)	p-value
aerobic benzoyl-CoA	*bcl*	SSE37_24404	1.85	1.17E-07
	boxD	SSE37_24409	1.53	2.71E-04
	boxC	SSE37_24419	1.38	1.88E-05
	boxB	SSE37_24424	1.47	2.17E-05
	boxA	SSE37_24439	1.04	0.49
	boxE	SSE37_24444	1.03	0.87
	boxR	SSE37_24449	-1.03	0.85
catechol (*meta*-cleaving)	*xylE*	SSE37_12149	36.42	0.01
gentisate	*nagI*	SSE37_02425	-1.28	0.10
	nagI	SSE37_02745	1.95	1.36E-05
	nagL	SSE37_02750	1.64	4.70E-03
	nagK	SSE37_02755	2.12	1.99E-03
homogentisate	*hmgA*	SSE37_24544	1.10	0.35
	hmgB	SSE37_24549	1.15	7.55E-04
	hmgC	SSE37_14734	-1.31	0.02
homoprotocatechuate	*hpaB*	SSE37_00435	41.12	0.01
	hpaC	SSE37_00425	3.01	0.05
	hpaD	SSE37_23379	1.07	0.07
	hpaE	SSE37_23374	-1.05	0.27
	hpaF	SSE37_23369	-1.27	0.12
hydroxycinnamate	*hcaA*	SSE37_01000	4.37	0.01
	hcaB	SSE37_01050	24.43	0.03
	hcaC	SSE37_12324	-1.10	0.23
	hcaG	SSE37_18822	3.49	1.45E-03
p-hydroxybenzoate	*pobA*	SSE37_18837	1.64	0.26
phenylacetate	*paaA*	SSE37_00390	-1.67	7.29E-03
	paaB	SSE37_00385	-1.44	0.06
	paaC	SSE37_00380	1.03	0.92
	paaD	SSE37_00375	-1.02	0.95
	paaE	SSE37_00370	1.04	0.88
	paaZ	SSE37_00360	1.24	0.22
	paaI	SSE37_15758	1.04	0.44
	paaJ	SSE37_15753	1.11	0.30
	paaK	SSE37_15748	1.58	1.30E-07
protocatechuate (*ortho*-cleaving)	*pcaQ*	SSE37_18832	3.59	0.14
	pcaC	SSE37_18842	-1.08	0.44
	pcaH	SSE37_18847	1.18	0.03
	pcaG	SSE37_18852	1.26	0.06
	pcaD	SSE37_21470	2.64	1.41E-03

Pathway	Gene[a]	Locus Tag[b]	Expression (RPKM)	p-value
	pcaI	SSE37_23014	-1.06	0.70
	pcaJ	SSE37_23019	1.41	0.09
	pcaF	SSE37_16533	2.43	1.17E-06
quinate	*qsuB*	SSE37_03985	2.20	7.54E-05
	qsuC	SSE37_06077	2.72	1.03E-04
	qsuD	SSE37_03980	1.99	2.13E-07
salicylate	*salA*	SSE37_01015	29.41	9.43E-03
vanillate	*vanA*	SSE37_02815	1.03	0.79
	vanB	SSE37_00265	-1.70	0.23
TCA	*gltA*	SSE37_06879	8.18	2.10E-06
	acnA	SSE37_25138	5.43	4.81E-05
	icd	SSE37_06274	3.52	2.86E-07
	sucA	SSE37_11324	1.10	0.59
	sucB	SSE37_11329	1.22	0.33
	sucC	SSE37_11294	1.46	0.08
	sucC	SSE37_22250	-2.18	0.09
	sucD	SSE37_11309	1.53	0.06
	sdhA	SSE37_11244	2.37	3.66E-03
	sdhB	SSE37_11229	1.56	5.85E-03
	sdhC	SSE37_11254	1.59	0.05
	sdhD	SSE37_11249	1.48	0.07
	fumC	SSE37_05510	6.60	9.79E-05
	mdh	SSE37_11289	6.33	2.09E-03
NADH dehydrogenase (Complex I, oxidative phosphorylation)	*nuoA*	SSE37_03720	3.46	1.23E-03
	nuoB	SSE37_03725	3.33	1.90E-04
	nuoC	SSE37_03735	3.49	1.94E-05
	nuoD	SSE37_03740	3.54	8.60E-06
	nuoE	SSE37_03745	3.98	9.37E-07
	nuoF	SSE37_03770	2.30	3.04E-04
	nuoG	SSE37_03795	4.24	2.39E-03
	nuoH	SSE37_03805	2.66	1.69E-03
	nuoI	SSE37_03815	4.26	8.97E-05
	nuoJ	SSE37_03825	3.33	1.62E-03
	nuoK	SSE37_03830	3.04	1.88E-03
	nuoL	SSE37_03840	5.34	7.89E-04
	nuoM	SSE37_03845	5.83	5.76E-04
	nuoN	SSE37_03850	3.87	5.75E-04
	nuoE-2	SSE37_03750	4.27	1.34E-04
	nuoK-2	SSE37_03835	5.39	4.68E-04

Pathway	Gene[a]	Locus Tag[b]	Expression (RPKM)	p-value
	COG3761	SSE37_10652	3.16	4.93E-04
	COG0702	SSE37_18567	2.14	1.33E-03
Succinate dehydrogenase (Complex II, oxidative phosphorylation)	*sdhA*	SSE37_11244	2.37	3.66E-03
	sdhB	SSE37_11229	1.56	5.85E-03
	sdhC	SSE37_11254	1.59	0.05
	sdhD	SSE37_11249	1.48	0.07
Cyt reductase (Complex III, oxidative phosphorylation)	*petA*	SSE37_19202	6.56	1.44E-03
	petB	SSE37_19197	8.94	8.11E-04
	petC	SSE37_19192	9.67	8.27E-04
	cydA	SSE37_19237	2.17	7.61E-05
	cydB	SSE37_19232	2.09	0.03
Cyt oxidase (Complex IV, oxidative phosphorylation)	*ccoN*	SSE37_10979	6.74	1.15E-04
	ccoO	SSE37_10984	4.73	2.46E-05
	ccoP	SSE37_10994	5.97	2.77E-05
	ccoQ	SSE37_10989	5.98	1.75E-05
	cyoE	SSE37_22969	4.86	5.96E-04
	ctaG	SSE37_22959	4.92	2.13E-05
	COG1612	SSE37_05145	1.32	0.02
	ctaD	SSE37_08623	4.40	4.54E-04
	ctaC	SSE37_22974	3.31	1.22E-04
	ctaE	SSE37_22954	3.69	3.26E-04
ATPase (Complex V, oxidative phosphorylation)	*atpA*	SSE37_19837	7.89	2.04E-03
	atpB	SSE37_15541	4.76	5.85E-04
	atpC	SSE37_19852	8.75	2.34E-03
	atpD	SSE37_19847	8.77	1.66E-03
	atpE	SSE37_15536	4.68	7.61E-03
	atpF	SSE37_15526	5.57	7.86E-03
	atpG	SSE37_19842	10.60	1.66E-03
	atpH	SSE37_19832	8.53	1.36E-03
	atpI	SSE37_15546	4.06	5.74E-03
	ppk	SSE37_08808	1.10	0.03
	ppaC	SSE37_13608	4.15	3.30E-03
Glycolysis	*galM*	SSE37_03995	1.16	0.38
	glk	SSE37_07473	3.58	6.08E-06
	pfk	SSE37_14559	2.05	7.07E-05
	pgi	SSE37_06914	-100.08	1.22E-24
	fda	SSE37_07908	2.89	5.78E-03
	gap-1	SSE37_25228	5.38	1.26E-03
	gap-2	SSE37_06549	2.20	0.06
	gap-3	SSE37_24534	1.51	0.09

Pathway	Gene[a]	Locus Tag[b]	Expression (RPKM)	p-value
	pgk	SSE37_07968	2.17	6.28E-03
	gpmI	SSE37_15888	4.02	2.12E-03
	eno	SSE37_07483	6.15	2.08E-03
	pyk	SSE37_20142	3.47	1.04E-03
	pdhA	SSE37_07988	-1.06	0.36
	pdhB	SSE37_07998	-1.25	2.51E-06
	pdhC	SSE37_08008	-1.15	2.22E-03

a . Unknown gene names are reported as clustered orthologous group (COG) numbers.

b. Ascending order of locus tags within each pathway is the order on the *S. stellata* E-37 chromosome.

Table 5.4. Protocatechuate 3,4-dioxygenase (PcaHG) activities of crude cell lysates.

Sample	Specific Activity (nmol pca/min/µg protein)
Acetate-MBM	49 ± 5.0
CG-SS (8 h)	294 ± 152
CG-SS (10 h)	249 ± 76.0
CG-SS (14 h)	335 ± 103
CG-SS (18 h)	410 ± 254
CG-SS (28 h)	353 ± 87.0
Protocatechuate-MBM	2333 ± 501

Table 5.5. Top 50 over-expressed genes during growth in CG-SS with predicted annotations sorted according to gene order of the *S. stellata* E-37 chromosome.[a]

Locus_Tag	KEGG KO#	NCBI GI#	Annotated Product Name	Fold Change (RPKM)	p-value
SSE37_00430		126731664	riboflavin biosynthesis protein, truncated	16.57	0.01
SSE37_00435		126731665	putative phenol hydroxylase large subunit	41.12	0.01
SSE37_00800		126731738	alpha-glucosides-binding periplasmic abc	439.37	7.81E-06
SSE37_00805	K10233	126731739	binding-protein-dependent transport systems	385.53	5.32E-06
SSE37_00810	K10234	126731740	binding-protein-dependent transport systems	266.02	1.22E-06
SSE37_00815		126731741	hypothetical protein	421.45	4.12E-07
SSE37_00820	K10235	126731742	ABC alpha-glucoside transporter, ATPase subunit	282.07	1.00E-06
SSE37_00825		126731743	Alpha amylase	320.26	5.04E-07
SSE37_00830		126731744	oligosaccharide alpha-1,6-glucosidase protein	248.67	5.39E-07
SSE37_00835		126731745	putative two-component response regulator	30.11	3.02E-05
SSE37_00950		126730555	TRAP transporter solute receptor, TAXI family	24.21	0.02
SSE37_01010		126730567	hypothetical protein	17.18	6.21E-03
SSE37_01015		126730568	putative monooxygenase	29.41	9.43E-03
SSE37_01020		126730569	TRAP-T family transporter, DctP (periplasmic	25.67	4.91E-03
SSE37_01025		126730570	TRAP-T family transporter, DctM (12 TMs)	20.1	8.79E-04
SSE37_01030		126730571	TRAP-T family transporter, DctQ (4 TMs) subunit	23.81	6.02E-06
SSE37_01035		126730572	TRAP-T family transporter, DctP (periplasmic	29.17	0.01
SSE37_01040		126730573	Glucose-methanol-choline oxidoreductase	21.65	2.88E-03
SSE37_01045	K00020	126730574	probable 6-phosphogluconate dehydrogenase	22.4	7.75E-03
SSE37_01050	K00128	126730575	Aldehyde dehydrogenase (NAD+)	24.43	0.03
SSE37_01060		126730577	putative carboxymethylenebutenolidase	21.22	0.08
SSE37_01070		126730579	enoyl-CoA hydratase	15.31	1.09E-03
SSE37_01105		126730586	hypothetical protein	73.58	1.07E-03
SSE37_01815	K07080	126730728	immunogenic protein	47.58	1.35E-05
SSE37_04150		126729634	hypothetical protein	18.26	9.48E-04

Locus_Tag	KEGG KO#	NCBI GI#	Annotated Product Name	Fold Change (RPKM)	p-value
SSE37_04155	K00390	126729635	phosphoadenosine phosphosulfate reductase	21.27	4.10E-03
SSE37_04160	K00381	126729636	sulfite reductase, putative	22.41	6.88E-03
SSE37_04165		126729637	hypothetical protein	15.54	0.01
SSE37_04800		126729764	hypothetical protein	∞	0.30
SSE37_05762	K10439	126732431	putative sugar uptake ABC transporter	44.43	2.23E-09
SSE37_12139		126730506	putative hydrolase	16.55	0.05
SSE37_12144		126730507	hypothetical protein	29.18	0.03
SSE37_12149	K00446	126730508	putative catechol 2,3-dioxygenase	36.42	0.01
SSE37_13763		126730884	zinc/manganese/iron ABC transporter, permease	∞	2.71E-03
SSE37_15081		126732282	hypothetical protein	19.15	1.42E-03
SSE37_16598	K09125	126730016	hypothetical protein	15.19	1.17E-03
SSE37_19297	K02965	126728168	ribosomal protein S19	17.84	2.34E-03
SSE37_21570		126731071	hypothetical protein	17.33	7.05E-11
SSE37_21575		126731072	hypothetical protein	17.42	2.03E-12
SSE37_21580		126731073	hypothetical protein	17.87	1.65E-07
SSE37_21600		126731077	ABC transporter ATP-binding protein	15.38	2.75E-05
SSE37_21615		126731080	alcohol dehydrogenase, iron-containing family	15.64	4.90E-13
SSE37_21620		126731081	hypothetical protein	63.9	1.18E-09
SSE37_24649	K10227	126729448	ABC sorbitol/mannitol transporter, periplasmic	41.71	0.01
SSE37_24654	K10228	126729449	binding-protein-dependent transport systems	52.03	0.04
SSE37_24659	K10229	126729450	binding-protein-dependent transport systems	56.7	0.04
SSE37_24664	K10230	126729451	ABC transporter related protein	40.38	0.06
SSE37_24669	K08261	126729452	sorbitol dehydrogenase	77.21	0.06
SSE37_24674	K00045	126729453	Mannitol dehydrogenase-like protein	40.47	0.06
SSE37_25193		126732612	YeeE/YedE family protein	20.47	0.01

a. Like-colored genes appear as a putative operon or pathway.

Table 5.6. Top 50 under-expressed genes during growth in CG-SS with predicted annotations sorted according to gene order of the *S. stellata* E-37 chromosome.[a]

Locus_Tag	KEGG KO#	NCBI GI#	Annotated Product Name	Fold Change (RPKM)	p-value
SSE37_03775		126729559	hypothetical protein	-52.75	3.28E-10
SSE37_03875	K04063	126729579	osmotically inducible protein OsmC	-53.29	6.23E-17
SSE37_04305		126729665	hypothetical protein	-109.84	3.31E-22
SSE37_04435		126729691	hypothetical protein	-30.7	2.62E-13
SSE37_04675		126729739	hypothetical protein	-153.55	2.96E-25
SSE37_05025		126729809	outer membrane protein, putative	-55.48	1.50E-19
SSE37_05180		126729840	penicillin binding protein B	-32.69	4.94E-12
SSE37_06294		126731254	hypothetical protein	-38.66	8.01E-13
SSE37_06754	K03781	126731346	Catalase	-68.26	2.83E-13
SSE37_06914		126731378	glucose-6-phosphate isomerase	-100.08	1.22E-24
SSE37_07398		126728633	hypothetical protein	-83.4	1.52E-09
SSE37_07618		126728677	Outer membrane protein, OmpA/MotB family	-65.43	9.69E-24
SSE37_08118		126728777	hypothetical protein	-65.88	7.11E-13
SSE37_08738		126728901	hypothetical protein	-44.06	1.94E-15
SSE37_09088		126728971	hypothetical protein	-35.63	3.85E-15
SSE37_09883		126731961	hypothetical protein	-33.14	2.43E-16
SSE37_10447		126732511	hypothetical protein	-23.27	2.67E-14
SSE37_10794		126730237	hypothetical protein	-48.88	6.59E-11
SSE37_11574		126730393	polynucleotide phosphorylase/polyadenylase	-23.52	1.19E-18
SSE37_11789		126730436	hypothetical protein	-28.16	5.08E-13
SSE37_13111		126731542	hypothetical protein	-47.45	5.42E-18
SSE37_13513		126730834	putative fatty acid oxidation complex alpha	-27.71	4.75E-13
SSE37_13518		126730835	acetyl-CoA acetyltransferase	-48.72	6.71E-12
SSE37_13523		126730836	acyl-CoA dehydrogenase	-65.46	1.07E-14
SSE37_13643	K05520	126730860	Peptidase C56, PfpI	-50.9	7.80E-18

Locus_Tag	KEGG KO#	NCBI GI#	Annotated Product Name	Fold Change (RPKM)	p-value
SSE37_13928		126730917	putative oxidoreductase	-52.33	1.34E-20
SSE37_14128		126730957	hypothetical protein	-34.09	2.05E-08
SSE37_15091		126732284	Putative glucose dehydrogenase B	-59.98	6.03E-14
SSE37_15146		126732295	hypothetical protein	-26.4	1.13E-12
SSE37_15151		126732296	hypothetical protein	-69.26	2.15E-15
SSE37_15988		126733008	hypothetical protein	-67.35	5.44E-21
SSE37_16423		126729981	hypothetical protein	-22.62	7.46E-14
SSE37_16428		126729982	hypothetical protein	-26.13	2.84E-12
SSE37_16433		126729983	hypothetical protein	-38.75	7.71E-23
SSE37_16438		126729984	hypothetical protein	-42.57	5.78E-15
SSE37_16478		126729992	hypothetical protein	-79.53	1.66E-21
SSE37_16483		126729993	CsbD-like protein	-143.04	8.57E-17
SSE37_16793		126730055	hypothetical protein	-184.34	3.46E-16
SSE37_17685	K07029	126732954	hypothetical protein	-23.24	2.42E-08
SSE37_18432		126732883	DedA family protein	-29.77	3.22E-15
SSE37_18592		126728027	Beta-Ig-H3/Fasciclin	-23.16	2.44E-15
SSE37_19432		126728195	hypothetical protein	-67.11	8.78E-19
SSE37_19437		126728196	hypothetical protein	-70.63	5.38E-18
SSE37_19442		126728197	hypothetical protein	-51.36	3.86E-20
SSE37_20432		126728395	probable exopolysaccharide synthesis protein	-23.42	2.42E-13
SSE37_21790	K07054	126731115	hypothetical protein	-22.91	6.93E-11
SSE37_23309		126729180	hypothetical protein	-23.5	9.37E-17
SSE37_24714		126729461	hypothetical protein	-27.16	1.88E-10
SSE37_25023		126732578	hypothetical protein	-41.21	4.75E-15
SSE37_25028		126732579	hypothetical protein	-41.83	1.63E-09

a. Like-colored genes appear as a putative operon or pathway.

VIII. FIGURES

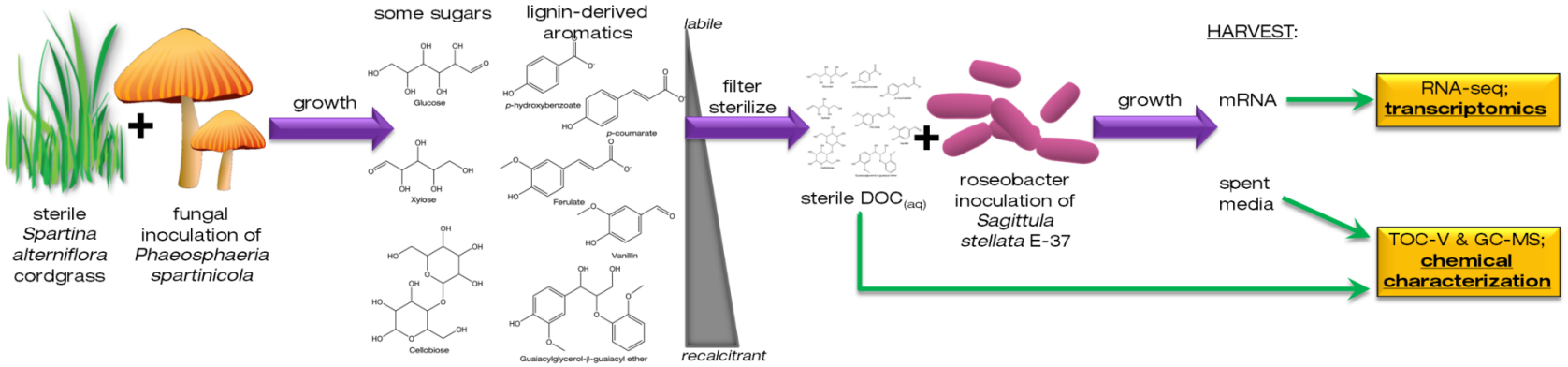


Figure 5.1. Flow chart of experimental approach to determine *S. stellata* E-37 activity in cordgrass-derived DOC.

Basidiomycetous fungi are shown for illustrative purposes; *P. spartinicola* 9005 is a microscopic ascomycete.

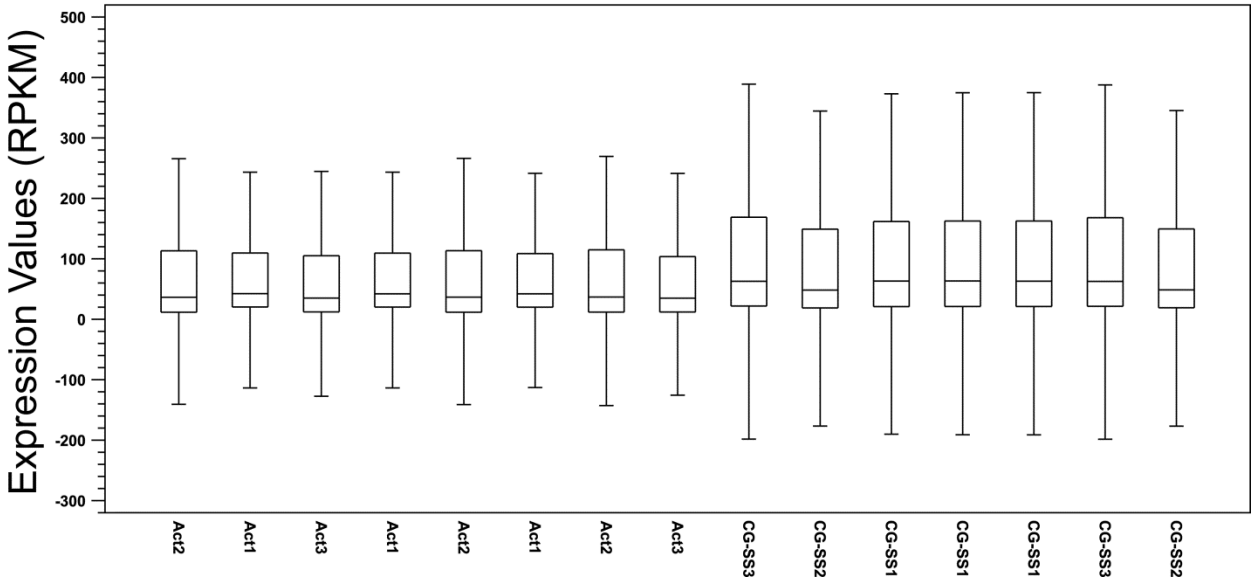


Figure 5.2. **Boxplot of RPKM normalized expression values for each technical sequencing replicate.** Biological replicates (n = 3) of the technical sequencing replicates are numbered 1-3. Medians are shown as the horizontal lines centered within each box. Act, 7 mM acetate grown; CG-SS, 10% Cordgrass-Glucose-Sea Salts grown.

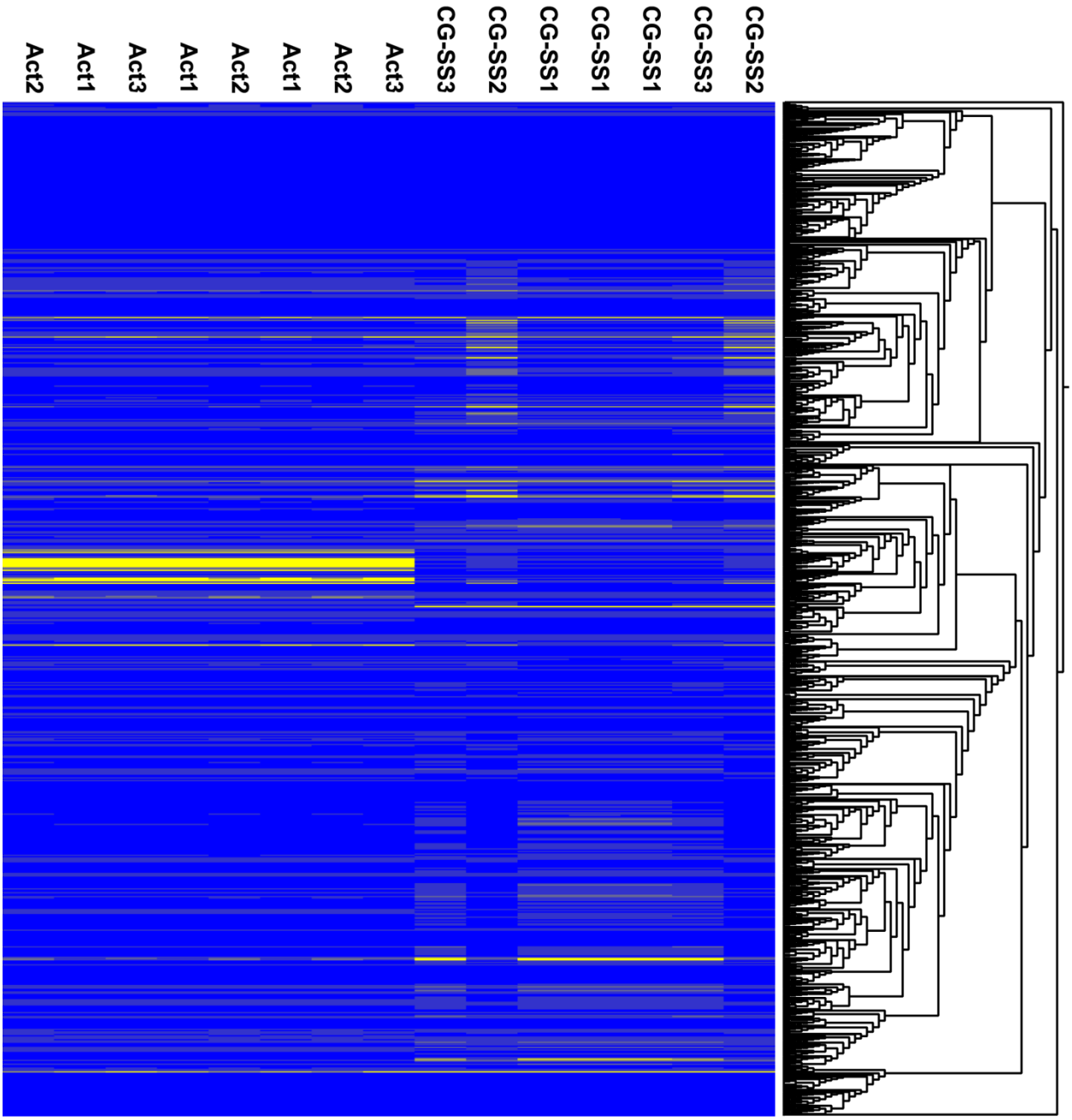


Figure 5.3. **Heat map of each RPKM normalized sequencing technical replicate.** Numbered 1-3 are biological replicates. Average linkage of clusters was used to construct transcript relatedness, and branch lengths reflect Pearson correlations. Act, 7 mM acetate grown; CG-SS, 10% Cordgrass-Glucose-Sea Salts grown.

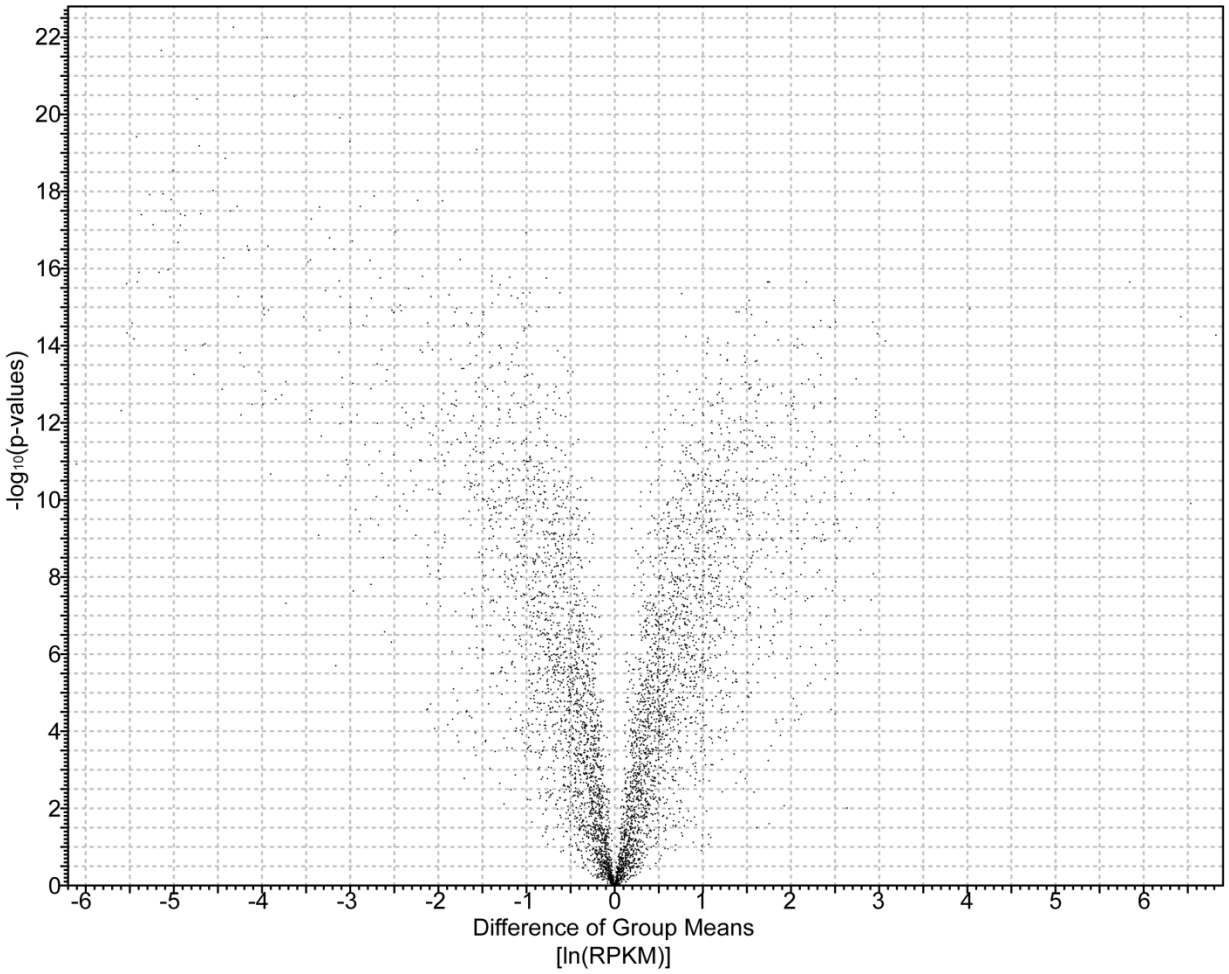


Figure 5.4. Volcano plot of significantly different genes between acetate and CG-SS growth.

P-values were calculated from unpaired Student's T-tests on RPKM averages of CG-SS grown versus Acetate grown replicates.

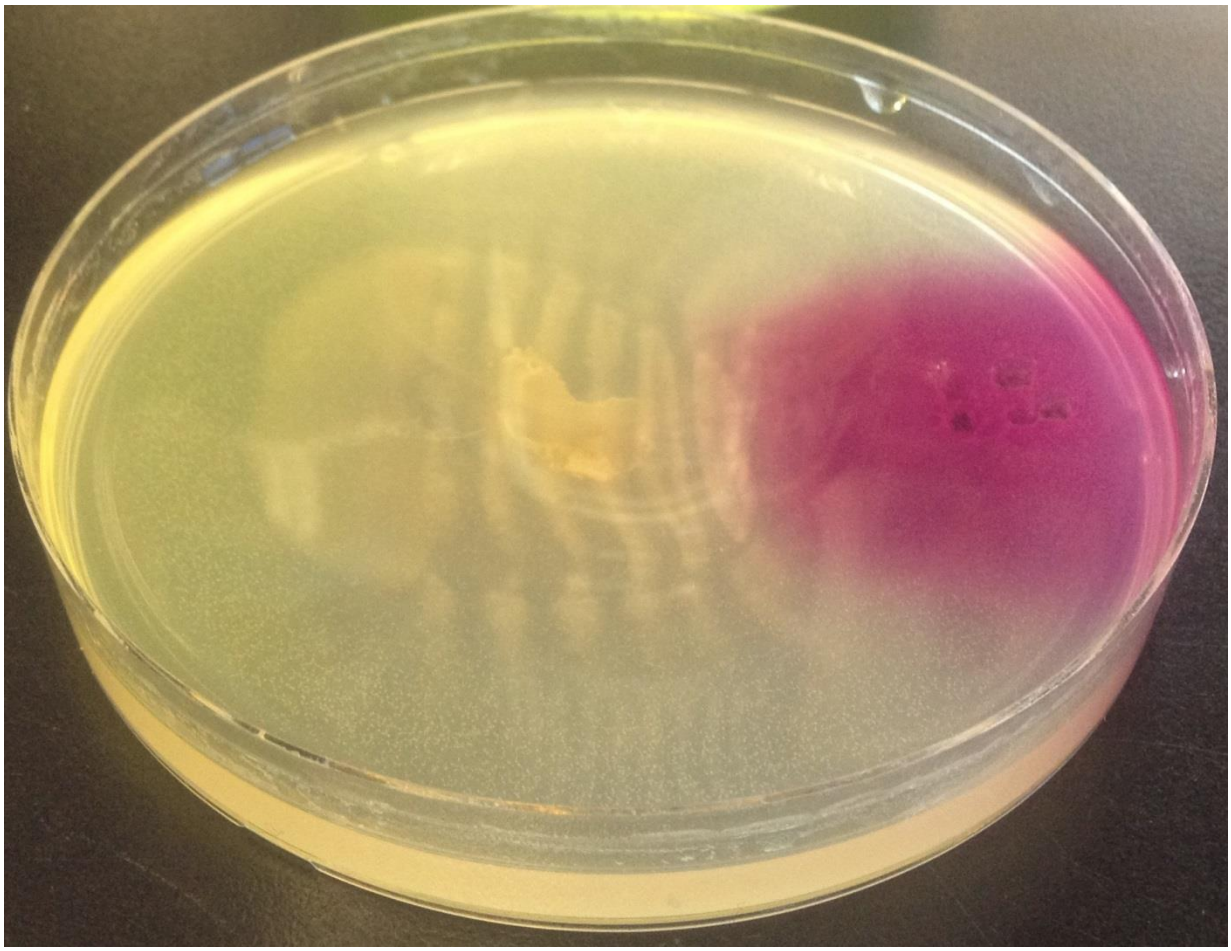


Figure 5.5. Laccase (lignin-degrading) assay of *Phaeosphaeria spartinicola* 9005.

Syringaldazine poured at the edge of the plate turned color immediately. Strong pink/purple color indicated a positive result for laccase.

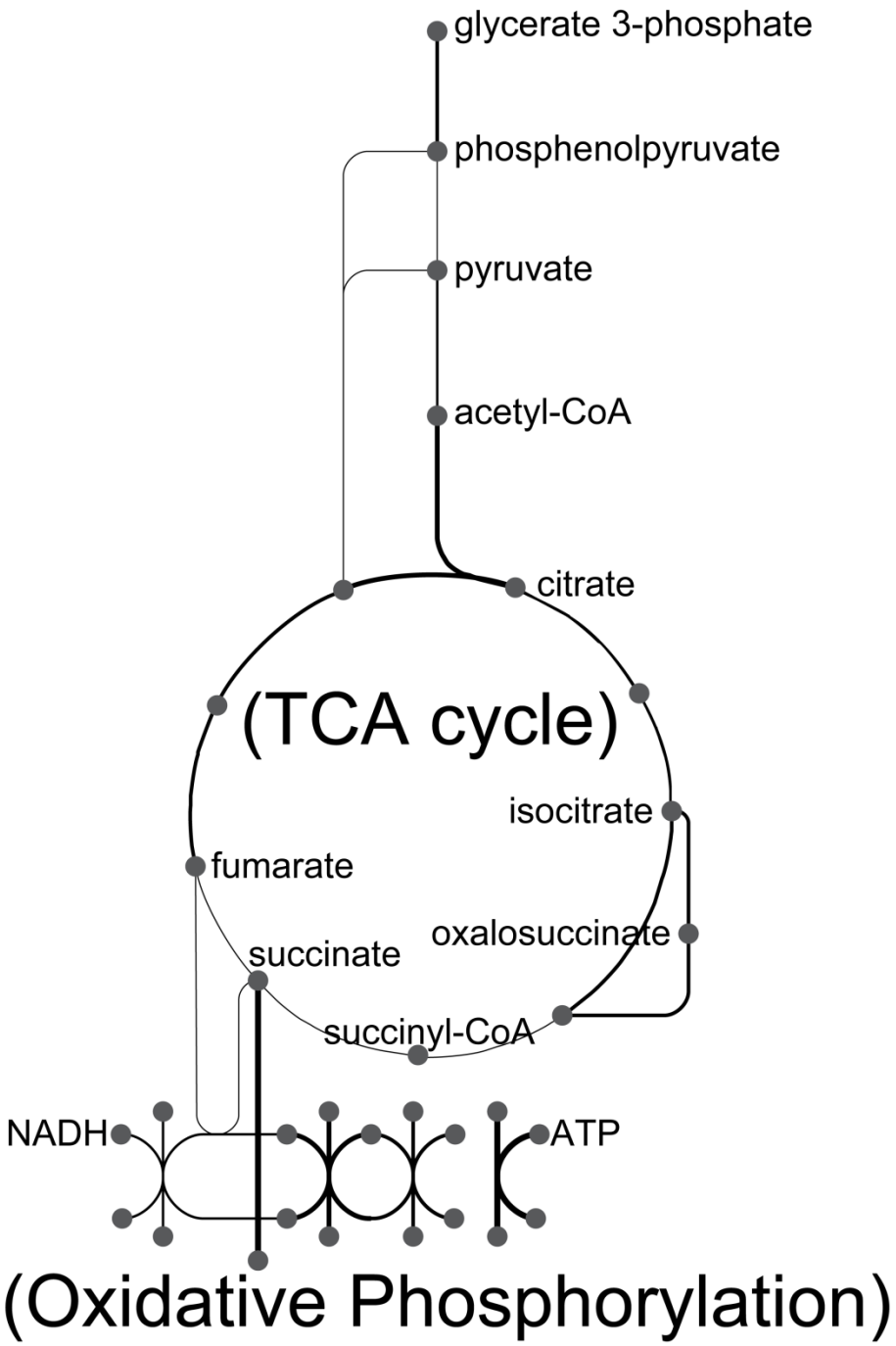


Figure 5.6. Overexpression of central metabolism genes in *S. stellata* E-37. All reactions are shown as lines where the thickness is proportional to overexpression values; none were underrepresented. Select compounds (grey dots) are labeled. ATP, adenosine-5'-triphosphate; CoA, coenzyme A; NADH, nicotinamide adenine dinucleotide; TCA, tricarboxylic acid.

CHAPTER 6 -

CONCLUSIONS

The primary intent of this dissertation was to document and communicate how roseobacters, particularly *Sagittula stellata* E-37, physiologically respond to aromatic compounds and the ecological implications of such activities.

A major route used by roseobacters for aerobic aromatic catabolism is the protocatechuate (*pca*) pathway (Buchan et al., 2000), but there are many other ring-cleaving pathways predicted in roseobacter isolate genomes based on homology searches (Newton et al., 2010). To determine how abundant another aromatic degradation pathway (the *box* pathway) is in natural roseobacter populations required two separate, but related, projects. Molecular-based probing techniques continue to evolve, and reliance on oligonucleotide primer-utilizing approaches is relatively common for studying environmental samples. With a theme in microbial ecology currently being diversity and continual unexpected variations on previous paradigms, I thought it would be imprudent *not* to make use of the growing number of publicly available metagenomes for primer development. Developing the De-MetaST-BLAST program (Figure 3.3) was my approach to do this in creating a diagnostic primer set for the *box* pathway (Table 2.1). Probing metagenome datasets from samples collected throughout the world resulted in increased confidence that the primer set was only retrieving *box* genes, especially because several coastal marine metagenomes are available, where I used the *boxB* primer set. Application of the primer set revealed *box* is surprisingly in high frequency (Figure 2.1), and it seems likely the pathway is being used for the breakdown of additional aromatic compounds other than benzoate. Pursuing this further might be best suited with a strain from *Azoarcus* because members contain both an aerobic (*box*) (Valderrama et al., 2012) and an anaerobic (*bzd*) pathway (Barragán et al., 2004) to

degrade benzoate as well as a variety of peripheral pathways known to create benzoyl-CoA as an intermediate (Heider and Fuchs, 1997).

Overall, I have focused most of my efforts on using model systems in the laboratory rather than direct analysis of natural systems. Upon entering the University of Tennessee as a graduate student, I had never seen the ocean. When I visited Charleston, South Carolina, I was completely shocked at how abundant and beautiful the cordgrass *Spartina alterniflora* really is along the coast (Figure 1.3). So, it is clear after seeing the huge salt marshes dominated by this grass that microorganisms that degrade it are of great ecological importance. Experimental work in the laboratory has been useful in identifying physiological characteristics (*e.g.*, simultaneous aromatic catabolism) of roseobacter isolates degrading plant-related material, which would have likely been overlooked in complex environmental experiments. Identification of this phenotype existing in cultured roseobacters enables future research to test how frequent simultaneous aromatic catabolism occurs in roseobacters inhabiting salt marsh systems.

Trophic specialization and genome minimization are popular topics of study in marine microbial ecology today (*e.g.*, Dupont et al., 2012; Grzymski and Dussaq, 2012; Kwan et al., 2012), yet from an ecological standpoint roseobacters are very different from many other abundant marine taxa in this regard. In many ways, their generalist physiology is a contrastive characteristic that makes roseobacters especially interesting to study. The simultaneous aromatic catabolism observed in *S. stellata* (Figure 4.4) further distinguishes the activities of these organisms from other abundant taxa in the oceans. Significantly enhanced growth rates were observed in *S. stellata* during simultaneous aromatic catabolism (Figure 4.5), and the result of multiple ring-

cleaving pathways being invoked is overexpression of central metabolic pathways (Figure 5.6). This further suggests multiple transporters are used to import the variety of substrates simultaneously consumed, but identification of these transporters are especially challenging for aromatic compounds that exist as both acids (which freely diffuse across cell membranes) and carboxylates (which require transport). There were many genes induced in *S. stellata* during growth in cordgrass-derived DOC that have unknown functions (*e.g.*, in Table 5.5). In some cases operons and transcriptional regulators can be predicted from the RNA-seq data, however biochemical characterization of these to assign functions will require much stronger biochemistry knowledge than I currently have. With clever experimental approaches and generation of *S. stellata* mutants, identification of the many transporters present in *S. stellata* should be possible.

REFERENCES

Barragán, M.J.L., Carmona, M., Zamarro, M.T., Thiele, B., Boll, M., Fuchs, G. et al. (2004) The *bzd* gene cluster, coding for anaerobic benzoate catabolism, in *Azoarcus* sp strain CIB. *Journal of Bacteriology* **186**: 5762-5774.

Buchan, A., Collier, L.S., Neidle, E.L., and Moran, M.A. (2000) Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. *Applied and Environmental Microbiology* **66**: 4662-4672.

Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.J., Richter, R.A., Valas, R. et al. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME Journal* **6**: 1186-1199.

Grzymski, J.J., and Dussaq, A.M. (2012) The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME Journal* **6**: 71-80.

Heider, J., and Fuchs, G. (1997) Anaerobic metabolism of aromatic compounds. *European Journal of Biochemistry* **243**: 577-596.

Kwan, J.C., Donia, M.S., Han, A.W., Hirose, E., Haygood, M.G., and Schmidt, E.W. (2012) Genome streamlining and chemical defense in a coral reef symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 20655-20660.

Newton, R.J., Griffin, L.E., Bowles, K.M., Meile, C., Gifford, S., Givens, C.E. et al. (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME Journal* **4**: 784-798.

Valderrama, J.A., Durante-Rodríguez, G., Blázquez, B., García, J.L., Carmona, M., and Díaz, E. (2012) Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. *Journal of Biological Chemistry* **287**: 10494-10508.

VITA

Christopher Adam Gulvik received his high school diploma in 2004 from Monona Grove High School in Madison, Wisconsin. In 2008, he earned a Bachelor of Science in Microbiology and a minor in Chemistry with *cum laude* honors from the University of Wisconsin – Oshkosh in Oshkosh, Wisconsin. Beginning in the fall of 2008, Chris started working in the Department of Microbiology under the guidance of his advisor Dr. Alison Buchan at the University of Tennessee – Knoxville (UTK) in Knoxville, Tennessee. In addition to research, Chris also taught microbiology laboratory courses eight of his ten semesters at UTK with advisement from Elizabeth McPherson. On Wednesday, April 24th, 2013, Chris had an oral defense for this dissertation.