ALISON BUCHAN
Ecology and Genetics of Aromatic Compound Degradation in the Ecologically Important Roseobacter Lineage of Marine Bacteria
(Under the direction of MARY ANN MORAN)

The degradation of lignin-related aromatic compounds is an important ecological process in the highly productive salt marshes of the southeastern U. S., yet little is known of the mediating organisms or their catabolic pathways. An initial study of six isolates suggested that members of the roseobacter lineage, a dominant marine clade within the α-Proteobacteria, could degrade aromatic compounds via the β-ketoadipate pathway, a catabolic route that has been well characterized in soil microbes. Four of the roseobacter group isolates had inducible protocatechuate 3,4-dioxygenase activity, a key enzyme in the pathway, in cell-free extracts when grown on *p*-hydroxybenzoate. The *pcaHG* genes encoding this ring-cleavage enzyme were cloned and sequenced from two isolates and in both cases the genes could be expressed in *E. coli* to yield dioxygenase activity. Evidence of genes encoding for protocatechuate 3,4-dioxygenase was found in all six roseobacter isolates by detection of *pcaH* by Southern hybridization or PCR amplification. These results suggested this ecologically important marine lineage compose a significant fraction of the aromatic compound degrading community in coastal systems. To test this hypothesis we investigated the diversity of *pcaH* amplified from bacterial communities associated with decaying *Spartina alterniflora*, the salt marsh grass dominating these coastal systems, as well as from enrichment cultures with aromatic substrates. Sequence analysis of 149 *pcaH* clones revealed 85 unique sequences. Fifty-eight percent of the clones matched sequences amplified from a collection of 36 bacterial isolates obtained from seawater or from senescent *Spartina*.

Fifty-two percent of the *pcaH* clones could be assigned to the roseobacter group. Another 6% matched genes retrieved from non-roseobacter isolates cultured from decaying *Spartina* and 42% could not be assigned to a cultured bacterium based on sequence identity. These findings revealed a high diversity of genes encoding a single step in aromatic compound degradation in this coastal marsh and that many of the genes were indeed harbored by members of the roseobacter lineage. In a final study, we explored the genetic diversity of the β-ketoadipate pathway among eight members of the roseobacter lineage (pairwise sequence identities of the 16S rDNA gene ranged from 92 to 99%). Genomic fragments containing gene clusters of this pathway were isolated and characterized by targeting *pcaH*. Sequence analysis revealed five unique gene arrangements. Identical gene clusters were found mostly between isolates demonstrating species-level identity (i.e. >99% similarity of 16S rDNA). In one isolate, six functionally related genes were identified: *pcaQ*, *pobA*, *pcaD*, -*C*, -*H*, and -*G*. The remaining seven isolates lacked at least one of these six genes within their respective gene clusters, however, gene order was consistent with this isolate s orientation. These results portray the dynamic nature of a set of genes that potentially play a central catabolic role in coastal marine environments.

INDEX WORDS: aromatic compound degradation, marine bacteria, protocatechuate dioxygenase, microbial ecology

ECOLOGY AND GENETICS OF AROMATIC COMPOUND DEGRADATION IN THE ROSEOBACTER LINEAGE OF MARINE BACTERIA

by

ALISON BUCHAN

B.S., James Madison University, 1994

M.S., The University of Georgia, 1997

A Dissertation Submitted to the Graduate Faculty of

The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2001

ECOLOGY AND GENETICS OF AROMATIC COMPOUND DEGRADATION IN THE ROSEOBACTER LINEAGE OF MARINE BACTERIA

by

ALISON BUCHAN

Approved:

Major Professor: Mary Ann Moran

Committee: Ellen L. Neidle
Brian Binder
James T. Hollibaugh
William Whitman

Electronic Version Approved:

Gordan L. Patel
Dean of the Graduate School
The University of Georgia
December 2001

ACKNOWLEDGMENTS

I would like to extend my deepest gratitude to my advisor Mary Ann Moran and to my co-advisor Ellen Neidle. They both provided a tremendous amount of support, guidance and encouragement throughout my studies. I also appreciate the assitance and patience of my committee members past and present: Brian Binder, James. T. Hollibaugh, Mark Schell, William Whitman and Feng Chen. I owe tremendous gratitude to Michael Krieger, Jennifer Brofft, the members of the Moran and Neidle labs, as well as the countless individuals in both the Departments of Marine Sciences and Microbiology. Finally, I wish to thank my family for always being there.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In the highly productive salt marshes of the southeastern U. S., lignin-related aromatic compounds comprise almost a third of the dissolved organic carbon pool (18). These compounds arise primarily from the decay of lignin, a complex aromatic polymer that is a major structural component of vascular plant tissue (12). In these coastal environments, aromatic compounds principally originate from *Spartina alterniflora* (18), a grass responsible for more than 80% of the total primary production in these systems (21). Substantial contributions also come from terrestrial environments and are delivered by riverine transport (18). While it is well established that bacteria contribute significantly to the transformation of this plant-derived material (16-18), the organisms responsible for catabolism of the chemically stable aromatic structures and the enzymatic pathways involved have yet to be properly characterized.

I. The β-ketoadipate pathway: a catabolic route for naturally-occurring aromatic compounds.

Despite the presence of a vast array of aromatic compounds in the environment, only a limited number of catabolic pathways for degrading these compounds have been identified. The central metabolic obstacle to degradation is the chemical stability of the aromatic ring, and only a few known enzymes can cleave the aromatic ring aerobically by

taking advantage of the oxidative potential of molecular oxygen (4). Most aromatic compounds, therefore, are first converted into a di- or trihydroxylated substrate of one of these ring-cleaving enzymes. The most prevalent ring-cleaving substrates are catechol and protocatechuate (10). Cleavage of the aromatic ring is catalyzed by dioxygenases and is termed *ortho*-cleavage (intradiol) when it occurs between the two hydroxyl groups and *meta*-cleavage (extradiol) when it occurs adjacent to one of the hydroxyl groups (10). Following ring cleavage, the products are typically converted into tricarboxylic acid cycle intermediates. *meta*-cleavage pathways are often plasmid-borne and are usually employed in the catabolism of environmental pollutants (e.g., toluene and xylene) (1). Modified versions of the *ortho*-cleavage pathway have evolved to degrade chlorinated substrates and are also typically localized on catabolic plasmids (22).

The β-ketoadipate is a primarily chromosomally-encoded convergent pathway that plays an integral role in the catabolism of naturally-occurring aromatic compounds derived from lignin and other plant components, as well as some environmental pollutants. The two branches of the pathway convert numerous aromatic monomers to β-ketoadipate via *ortho*-cleavage of either protocatechuate or catechol. Two additional steps complete the conversion of β-ketoadipate to succinyl- and acetyl-CoA. This conversion has been illustrated in a variety of diverse bacterial genera, including *Acinetobacter*, *Alicaligenes*, *Azotobacter*, *Agrobacterium*, *Burkholderia*, *Bacillus*, *Pseudomonas*, *Rhizobium* and *Rhodococcus* (10).

II. The roseobacter group: an ecologically important marine lineage.

In 1997, a group of marine bacteria was identified in the estuaries of Georgia that is both numerically important and readily cultivated in the laboratory (6). This lineage, designated the "roseobacter group" because of its phylogenetic affiliation with members

of the well-characterized genus *Roseobacter*, forms a monophyletic clade within the α-3 subclass of the Proteobacteria. In addition to contributing up to 30% of the bacterioplankton 16S ribosomal RNA (rRNA) genes in coastal Georgia seawater (6), members of the lineage are well represented in bacterial 16S rRNA clone libraries from several oceanic and coastal regions (7, 19, 24). Cultured representatives of the group have been isolated from diverse coastal and open ocean sites, including hydrothermal vents and sea ice (2, 3, 9, 15, 23, 25), as well as saline lakes (13, 20). Group members demonstrate a variety of metabolic capabilities, including phototrophy, aerobic sulfite oxidation, organic sulfur compound degradation, and lignin degradation (2, 11, 14, 15, 23).

The group was first discovered when strains from the clade were repeatedly isolated from marine enrichment cultures with lignin derivatives and marine humic substances (8). Detailed studies of one such isolate, *Sagittula stellata*, revealed that it was able to mineralize and solubilize polymeric synthetic lignin and use several lignin-related aromatic monomers (e.g. benzoate, *p*-hydroxybenzoate, vanillate, coumarate, cinnamate and ferulate) as a sole carbon and energy source (5). Subsequent screens of several other roseobacter isolates revealed growth on several of those same lignin-related aromatic compounds, and suggested this group may be ecologically important in the degradation of naturally-occurring aromatic compounds found in marshes and estuaries of the southeastern U.S.

III. Objective

Despite the abundance of aromatic compounds in coastal marshes and estuaries, investigations of the aerobic degradation of these compounds by marine bacteria have

been rare. The primary objective of this dissertation is to identify and characterize the catabolic pathways utilized by indigenous marine bacteria in the degradation of aromatic compounds.

References

1. Assinder, S. J. and P. A. Williams. 1990. The TOL plasmids: determinants of the catabolism of toluene and xylenes. Adv. Microbiol. Physiol. 31:1-69.

2. Bowman, J. P., S. A. McCammon, M. V. Brown, D. S. Nichols, and T. A. McMeekin. 1997. Diversity and association of psychrophilic bacteria in Antarctic sea ice. Appl. Environ. Microbiol. 63:3068-3078.

3. Burnett, W. J. and J. D. McKenzie. 1997. Subcuticular bacteria from the brittle star *Ophiactis balli* (Echinodermata: Ophiuroidea) represent a new lineage of extracellular marine symbionts in the α-subdivision of the class Proteobacteria. Appl. Environ. Microbiol. 63:1721-1724.

4. Dagley, S. 1986. Biochemistry of aromatic hydrocarbon degradation in Pseudomonads. P 527-555. In J.R. Sokatch (Ed.), The Bacteria, Vol 10. Academic Press Inc., New York.

5. González, J. M., F. Mayer, M. A. Moran, R. E. Hodson, and W. B. Whitman. 1997. *Sagittula stellata* gen. nov., sp. nov., a lignin transforming bacterium from a coastal environment. Int. J. Syst. Bacteriol. 47:773-780.

6. González, J. M., and M. A. Moran. 1997. Numerical dominance of a group of marine bacteria in the α-subclass of the class Proteobacteria in coastal seawater. Appl. Environ. Microbiol. 63:4237-4242.

7. González, J. M., R. Simó, R. Massana, J. S. Covert, E. O. Casamayor, C. Pedrós-Alió, and M. A. Moran. 2000. Bacterial community structure associated with a

dimethylsulfoniopropionate-producing north Atlantic algal bloom. Appl. Environ. Microbiol. 66:4237-4246.

8. González, J. M., W. B. Whitman, R. E. Hodson, and M. A. Moran. 1996. Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. Appl. Environ. Microbiol. 62:4433-4440.

9. Gosink, J. J. and J. T. Staley. 1995. Biodiversity of gas vacuolate bacteria from Antarctic sea ice and water. Appl. Environ. Microbiol. 61:3486-3489.

10. Harwood, C. S., and R. E. Parales. 1996. The β-ketoadipate pathway and the biology of self-identity. Ann. Rev. Microbiol. 50:553-590.

11. Holmes, A. J., D. P. Kelly, S. C. Baker, A. S. Thompson, P. DeMarco, E. M. Kenna, and J. C. Murrell. 1997. *Methylosulfonomonas methylovora* gen. nov., sp. nov., and *Marinosulfonomonas methylotropha* gen. nov., sp. nov: novel methyltrophs able to grow on methanesulfonic acid. Arch. Microbiol. 167:46-53.

12. Kirk, T. K., T. Higuchi, and H.-M. Chang (eds). 1980. Lignin biodegradation: microbiology, chemistry and potential applications, CRC Press, Boca Raton, FL.

13. Labrenz, M., B. J. Tindall, P. A. Lawson, M. D. Collins, P. Schumann, and P. Hirsch. 2000. *Staleya guttiformis* gen. nov., sp. nov. and *Sulfitobacter brevis* sp. nov., α-3-Proteobacteria from the hypersaline, heliothermal and meromictic antarctic Ekho Lake. Int. J. Syst. Evol. Microbiol. 50:303-313.

14. Lafay, B., R. Ruimy, C. Rausch de Traubenberg, V. Breittmayer, M. J. Gauthier, and R. Christen. 1995. *Roseobacter algicola* sp. nov., a new marine bacterium isolated from

the phycosphere of the toxin-producing dinoflagellate *Prorocentrum lima*. Int. J. Syst. Bacteriol. 45:290-296.

15. Ledyard, K. M., E. F. DeLong, and J. W. H. Dacey. 1993. Characterization of a DMSP-degrading bacterial isolate from the Sargasso Sea. Arch. Microbiol. 160:312-318.

16. Moran, M. A. and R. E. Hodson. 1989. Formation and bacterial utilization of dissolved organic carbon derived from detrital lignocellulose. Limnol. Oceanogr. 34:1034-1047.

17. Moran, M. A., and R. E. Hodson. 1990. Bacterial production on humic and nonhumic components of dissolved organic carbon. Limnol. Oceanogr. 35:1744-1756.

18. Moran, M. A., and R. E. Hodson. 1994. Dissolved humic substances of vascular plant origin in a coastal marine environment. Limnol. Oceanogr. 39:762-771.

19. Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni. 1995. Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnol. Oceanogr. 40:148-158.

20. Petursdottir, S. K. and J. K. Kristjansson. 1997. *Silicibacter lacuscaerulensis* gen nov., sp. nov., a mesophilic moderately halophilic bacterium characteristic of the Blue Lagoon geothermal lake in Iceland. Extremophiles. 1:94-99.

21. Pomeroy, L. R., W. M. Darley, E. L. Dunn, J. L. Gallagher, E. B. Haines, and D. M. Whitney. 1981. Primary production, p. 39-68. *In* L. R. Pomeroy and R. G. Wiegert (ed.), The ecology of a salt marsh. Springer-Verlag, New York, N.Y.

22. Schlömann, M. 1994. Evolution of chlorocatechol catabolic pathways. Biodegradation. 5:301-321.

23. Sorokin, D. Y. 1995. *Sulfitobacter pontiacus* gen. nov., sp. nov.: a new heterotrophic bacterium from the Black Sea, specialized on sulfite oxidation. Microbiology 64:354-365.

24. Suzuki, M. T., M. S. Rappé, Z. W. Haimberger, H. Winfield, N. Adair, J. Ströbel, and S. J. Giovannoni. 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl. Environ. Microbiol. 63:983-989.

25. Teske, A., T. Brinkhoff, G. Muyzer, D. P. Moser, J. Rethmeier, and H. W. Jannasch. 2000. Diversity of thiosulfate-oxidizing bacteria from marine sediments and hydrothermal vents. Appl. Envir. Microbiol. 66: 3125-3133.

CHAPTER 2

KEY AROMATIC RING-CLEAVING ENZYME, PROTOCATECHUATE 3,4-DIOXYGENASE, IN THE ECOLOGICALLY IMPORTANT MARINE ROSEOBACTER LINEAGE[1]

[1]Buchan, A., L. S. Collier, E. L. Neidle, and M. A. Moran. 2000. *Applied and Environmental Microbiology*. 66:4662-4672.

Reprinted here with permission of publisher.

Abstract

Aromatic compound degradation was investigated in six bacteria representing an ecologically important marine taxon of the α-proteobacteria. Initial screens suggested that isolates in the roseobacter lineage can degrade aromatic compounds via the β-ketoadipate pathway, a catabolic route that has been well characterized in soil microbes. Six roseobacter isolates were screened for the presence of protocatechuate 3,4-dioxygenase, a key enzyme in the β-ketoadipate pathway. All six isolates were capable of growth on at least three of the eight aromatic monomers presented (anthranilate, benzoate, *p*-hydroxybenzoate, salicyclate, vanillate, ferulate, protocatechuate and coumarate). Four of the roseobacter group isolates had inducible protocatechuate 3,4-dioxygenase activity in cell-free extracts when grown on *p*-hydroxybenzoate. The *pcaGH* genes encoding this ring-cleavage enzyme were cloned and sequenced from two isolates, *Sagittula stellata* E-37 and isolate Y3F, and in both cases the genes could be expressed in *Escherichia coli* to yield dioxygenase activity. Additional genes involved in the protocatechuate branch of the β-ketoadipate pathway (*pcaC*, *pcaQ*, *pobA*) were found to cluster with *pcaGH* in these two isolates. Pairwise sequence analysis of the *pca* genes revealed greater similarity between the two roseobacter group isolates than when these genes from either roseobacter strain were compared to genes from soil bacteria. A degenerate PCR primer set targeting a conserved region within PcaH successfully amplified a fragment of *pcaH* from two additional roseobacter group isolates, and Southern hybridization indicated the presence of *pcaH* in the remaining isolates from which DNA sequences were not determined. Evidence of protocatechuate 3,4-dioxygenase and the β-ketoadipate pathway was found in

all six roseobacter isolates and suggests widespread abilities to degrade aromatic compounds in this ecologically important marine lineage.

Introduction

The roseobacter lineage in the α-proteobacteria is abundant in southeastern U.S. estuaries (4, 16) and other coastal environments (28, 47). In the expansive salt marshes of the southeastern U.S., where many roseobacter strains have been isolated, organic matter is strongly influenced by naturally occurring aromatic compounds in the form of lignin and humic substances (26, 27). Studies of roseobacter group isolate *Sagittula stellata* E-37 (17) and preliminary screens of other cultured roseobacter isolates revealed capabilities for the transformation of synthetic lignin and degradation of lignin-related aromatic monomers. Because the roseobacter lineage is one of the few dominant marine clades that is amenable to culturing (14), the group presents a unique opportunity to investigate the catabolism of aromatic compounds by a cluster of bacteria that is ecologically important and exclusively marine.

Despite the vast array of aromatic compounds in aquatic and terrestrial environments, the degradation of different compounds usually proceeds through a limited number of metabolic pathways. Most aromatic compounds are first converted to one of several di- or tri-hydroxylated substrates, such as catechol or protocatechuate (Fig. 2.1), whose aromatic ring can be enzymatically cleaved (3). In the β-ketoadipate pathway, a primarily chromosomally-encoded catabolic route that is widely distributed in soil bacteria and fungi, catechol and protocatechuate are cleaved between their two hydroxyl groups by catechol 1,2 dioxygenase (1,2-CTD) or protocatechuate 3,4 dioxygenase (3,4-PCD), respectively (18). This branched pathway converges when the ring-cleavage products of either catechol or protocatechuate are converted to β-ketoadipate, the

metabolite for which the pathway was named. Two additional steps complete the conversion of β-ketoadipate to tricarboxylic cycle intermediates.

Studies of soil bacteria have revealed that the catechol and protocatechuate branches of the β-ketoadipate pathway contain analogous enzymes that are similar in sequence (21). Both 3,4-PCD and 1,2-CTD belong to a large class of nonheme iron-containing dioxygenases (18). 3,4-PCD is comprised of equimolar amounts of two nonidentical α- and β-subunits that are encoded by the usually cotranscribed *pcaG* and *pcaH* genes. The PcaG and PcaH proteins are similar to each other at both the structural level and that of amino acid sequence, with approximately 30% of their aligned residues being identical (21, 30, 49). Significant sequence similarity is also observed between these proteins and the subunits of 1,2-CTD which is usually comprised of homodimers encoded by the *catA* gene. The similar sequences of these proteins from soil bacteria such as *Pseudomonas putida*, *Agrobacterium tumefaciens,* and *Acinetobacter* species suggest that the PcaG, PcaH, and CatA proteins all arose from a common ancestor that diverged fairly recently (40).

Although aromatic compounds are abundant in coastal marshes and estuaries, investigations of the aerobic degradation of these compounds by marine bacteria have been rare. A spate of recent articles indicates an increasing interest in this topic (11, 12, 23, 35), although most current knowledge is nevertheless based on studies of soil bacteria. Here we describe investigations of the protocatechuate branch of the β-ketoadipate pathway, a route that in coastal marine environments may be used to degrade aromatic monomers that arise during the decay of lignin and other vascular plant components (e.g.,

vanillate, coumarate, cinnamate, ferulate, benzoate, and *p*-hydroxybenzoate) (5, 22, 36) (Fig. 2.1). Some environmental aromatic pollutants may be degraded by this pathway as well (3). In this report, evidence is provided that a key enzyme of this pathway, 3,4-PCD, is present in six marine bacteria affiliated with the roseobacter group. We suggest that aromatic ring cleavage may be characteristic of this ecologically important lineage.

Material and Methods

Bacterial isolation and phylogenetic analysis. Isolates used in this study were cultured from seawater or sediments collected in the estuaries or coastal waters of the southeastern U.S., either isolated from lignin or aromatic monomer enrichments (*Sagittula stellata* E-37, *Sulfitobacter* sp. EE-36, isolate Y3F, isolate IC4, isolate S25com04) or cultured directly from coastal seawater using non-selective, low nutrient seawater plates (isolate GAI-16) (16, 17). All isolates were shown to be members of the roseobacter group by sequencing of rRNA regions as previously described (González and Moran 1997). Full 16S rDNA sequences were previously reported for *S. stellata* E-37 (U58356), GAI-37 (AF007260), and *Sulfitobacter* sp. EE-36 (AF007254). The complete 16S rDNA sequence for Y3F and partial sequences for isolates IC4 and S25com04 were determined using universal primers as outlined in González et al. (15) and submitted to GenBank (AF253467, AF254098, AF254099).

Screening of roseobacter group isolates for growth on aromatic compounds. Isolates were initially grown on Marine Basal Medium (MBM: 100 mM NaCl, 25 mM $MgSO_4$, 5 mM KCl, 5 mM $CaCl_2$, 25 mM FeEDTA; pH 7.5) agar plates amended with

0.4% yeast extract. Fresh cultures were transferred to MBM plates supplemented with one of a suite of aromatic monomers as the sole carbon source (anthranilate, benzoate, *p*-hydroxybenzoate [POB], salicyclate, vanillate, ferulate, protocatechuate or coumerate) at 4 mM concentration and 0.1% vitamins (17). Plates were incubated at 28-30°C in the dark for up to 14 d.

Preparation of cell-free extracts and enzyme assays. Isolates were grown in 100 ml MBM with 4 mM of either pob or sodium acetate as the sole carbon source at 28-30°C for 48-96 hr, shaking and in the dark. Cells were harvested by centrifugation, washed once with sterile, deionized water, and stored at -20°C. Cell pellets were suspended in 100-200 µl breaking buffer (50 mM Tris-HCl, 10% glycerol, 5mM $(NH_4)_2SO_4$, 2.5 mM EDTA, and 1 mM dithiothreitol, pH 7.5). Cell-free extracts were prepared as previously described (44) and 3,4-PCD activity was determined spectrophotometrically (45). Activity of the dioxygenase was determined by measuring the decrease in absorbance at 290 nm over a 120 sec period at 12 sec intervals on a Beckman DU 640 spectrophotometer (Beckman Instruments Inc., Fullerton, CA). Protein concentrations were determined by the method of Bradford (1) or Lowry (25), using the Protein Assay Kit II or the DC Protein Assay Kit, respectively (Bio-Rad Laboratories, Hercules, CA). One unit of activity (U) is defined as the amount of enzyme that transforms 1 µmol of substrate per min per mg protein.

Detection and isolation of catabolic genes from *S. stellata* E-37 and isolate Y3F. Three degenerate oligonucleotides, pcaHF2 ($^{5'}$CCRMYGCYGARCGY-RCRTYGT$^{3'}$), pcaHF1 ($^{5'}$GARRTRTGGCARGCSAAYGC$^{3'}$), and pcaHR1 ($^{5'}$CCVGGRTTYATS

GTYKGGAA[3]) were designed based on conserved regions found within the amino acid alignments of PcaH from *Acinetobacter* sp. ADP1 (M33798), *Pseudomonas putida* (ATCC 23975) (L26294), and *Burkholderia cepacia* DB01 (M30791), where R = A + G, M = A + C, S = G + C, V = G + A + C, Y = C + T, and K = G + T. The degenerate oligonucleotides correspond to residue positions 74-79, 99-104, and 140-145, respectively, of PcaH from *Acinetobacter* sp. ADP1. Chromosomal digestions of *S. stellata* E-37 were used in Southern hybridization analysis with the degenerate oligonucleotides. Nonradioative probes made by 3 tailing of the degenerate oligonucleotides with digoxigenin (DIG) were used in hybridizations (Genius System; Roche Molecular Biochemicals, Indianapolis, IN). Adjacent fragments housing other proximal, putative catabolic genes were identified in *S. stellata* E-37 by Southern hybridization analysis using DIG labeled fragments generated from distal ends of the primary fragment (see Fig. 2.2A).

The catabolic gene cluster was initially identified in isolate Y3F by PCR amplification of a portion of *pcaG* using *S. stellata* E-37 specific primers P34O1321R ([5] GGATGTCGAAGCGGT[3]), designed from an apparently conserved region spanning residues 183-187 in PcaG (*S. stellata* E-37 numbering), and pcaHF1ND ([5] GAGGTCTGGCAGGCCAAT[3']), a non-degenerate version of pcaHF1. The PCR was carried out in a reaction mixture containing: 1X buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), 2 mM dNTP, 30 ng template DNA, 36 pmole P34O1321R primer, 50 pmole pcaHF1ND primer, and 1U Taq DNA polymerase. PCR was performed in a DNA Thermal Cycler 480 (Perkin-Elmer Corp., Foster City, CA) with an initial 10 min

at 95°C followed by 30 cycles of: 1 min at 95°C, 1 min at 47°C, and 1 min at 72°C. The PCR product from isolate Y3F was then labeled with DIG by a random priming reaction and used in Southern hybridizations with isolate Y3F chromosomal digestions.

Hybridizations with DIG-labeled probes produced by the random priming method were always carried out at 42°C. Hybridization temperatures for the DIG-labeled oligonucleotide probes were determined empirically for each probe and ranged between 55-65°C. Lower stringency hybridization conditions consisted of decreasing the hybridization temperature 5-10°C below the empirically determined optimal temperature and carrying out all post-hybridization washes at that temperature.

DNA preparation and plasmid construction. To isolate genomic fragments housing putative catabolic genes, recombinant plasmid libraries were constructed. DNA fragments of the sizes corresponding to positive hybridization signals in Southern blot analysis were gel excised and ligated into either the pZERO (Invitrogen Corp., Carlsbad, CA) or pT7Blue-2 (Novagen Inc., Madison, WI) vector. Colony hybridization of genomic libraries with DIG-labeled DNA probes identified positive clones.

Expression of roseobacter group isolate DNA in *E. coli*. To express the *pcaHG* genes from *S. stellata* E-37 and isolate Y3F under control of the *lac* promoter in *E. coli*, pABX07 and pABX26 were constructed as follows using PCR primers that would introduce restriction sites for optimal positioning in the expression vector pCYB1 (New England Biolabs). A *Nde*I cleavage site was introduced just before the ATG start codon of the *pcaH* gene and a *Hind*III cleavage site was introduced downstream of the *pcaG* stop codon by PCR amplification using the high fidelity *pwo* DNA Polymerase (Roche

Molecular Biochemicals). The 1.3 kbp *Nde*I-*Hind*III *pcaHG* DNA fragments were then ligated into the corresponding sites on the pCYB1 vector. The resultant recombinant plasmids were sequenced on one strand to verify that no mutations had been introduced during PCR. L Broth cultures (100 ml) of plasmid carrying *E. coli* Top10F' (Invitrogen) cells were grown at 30°C for 13 h. IPTG was added at the time of inoculation. Isolate Y3F required higher a higher concentration of IPTG (800 µM) relative to *S. stellata* E-37 (100 µM).

Detection of putative *pcaH* in roseobacter isolates. A degenerate PCR primer pair was designed based on conserved regions within deduced PcaH sequence alignments of all genes in the database: P34OIDf ($^{5'}$YTIGTIGARRTITGGCARGCIAA YGC$^{3'}$) and P34OIDr ($^{5'}$ICYIAIRTGIAYRTGIGCIGGICKCCA$^{3'}$). For PCR analysis, genomic DNA was prepared from each isolate by scraping fresh colonies (ca. 50 mg) from an agar plate. Colonies were washed twice with dH_2O and then boiled in 500 µl dH_2O for 7-10 min. Cellular debris was collected by centrifugation at 15,000 x *g* for 2 min. The supernatant fluid was drawn off and used directly in PCR reactions using the MasterAmp PCR Optimization Kit (Epicentre Technologies, Madison, WI) to establish 12 separate reactions buffers for each isolate. To each 50 µl reaction, 50 pmole of each primer, 1U Taq Polymerase (Roche Molecular Biochemicals), and 3 µl of cell lysate were added. PCR was performed with an initial 4 min at 98°C, followed by 30 cycles of: 95°C for 1 min, 49°C for 45 sec, and 72°C for 45 sec. Amplification products of appropriate size were ligated into the pCR 2.1 vector (Invitrogen) and sequenced.

Sequence determination and analysis. DNA sequences were determined with double-stranded templates and sequencing primers that recognized the cloning vector. When necessary, adjacent portions were sequenced by designing new oligonucleotide primers from the sequenced portion. Nucleotide sequencing was performed on either an ABI373A (at the University of Georgia Molecular Genetics Instrumentation Facility) or an ABI310 automated DNA sequencer (PE Applied Biosystems). In all cases, both strands of reported genes were sequenced. The DNA and deduced amino acid sequences were analyzed using the Genetics Computer Group (GCG) program package 8.0 (6). Homology searches (BLAST) were carried out at the network server of the National Center for Biotechnology Information.

Phylogenetic trees were constructed for PcaGH sequences with the PHYLIP package (9) by using evolutionary distances (Kimura distances) and the neighbor-joining method.

Sequences were deposited into the Genbank database (AF253465, AF253466, AF253538, AF253539).

Results

Choice of marine strains. Roseobacter isolates were chosen for this study because they are abundant in the vascular-plant dominated coastal marshes of the southeastern U.S. (15, 16). Six strains were selected from a collection of approximately 30 obtained from Georgia's diverse salt marshes. Four of the six isolates (*S. stellata* E-37, *Sulfitobacter* sp. EE-36, isolate Y3F, and isolate IC4) were obtained by enrichment with lignin-rich

pulp mill effluent (Indulin) as the sole carbon and energy source. Three of these Indulin-grown isolates (*S. stellata* E-37, *Sulfitobacter* sp. EE-36, and isolate Y3F) came from sea water of two different Georgia intertidal marshes (Skidaway River, GA and the Dublin River, GA) that have high concentrations of natural plant lignins. The fourth of these Indulin-grown strains, isolate IC4, was from marine sediment collected in the St. Mary's Estuary, GA, near the outfall of a paper mill plant in a region with high concentrations of industrially-modified lignin wastes. Isolate S25com04, from the Satilla River, was selected for its ability to grow on a single aromatic-ring containing compound (pob). Finally, isolate GAI-16 was obtained directly from coastal seawater near Skidaway, GA by non-selective means (16). These isolates thus represent four different coastal regions of the southeastern U.S. and three different culture approaches.

Growth on aromatic compounds. The ability of the marine bacteria to grow on aromatic substrates was tested on solid media containing as the sole carbon source one of eight single-ring containing compounds known to be degraded by soil microbes via the β-ketoadipate pathway. These compounds, arising in large part in nature from the decay of vascular plant material, may be converted either to catechol or protocatechuate. Compounds such as protocatechuate, POB, vanillate, ferulate, and coumerate are typically degraded through the protocatechuate branch of the β -ketoadipate pathway, whereas benzoate is converted to catechol by some soil microbes and protocatechuate by others. All six *Roseobacter* group isolates were able to grow well on at least two of the individual aromatic monomers provided as a sole carbon source, and one strain, *Sulfitobacter* sp. EE-36, was able to grow at the expense of all compounds tested (Table

2.1). All isolates grew well on POB, a characteristic typical of organisms containing the β-ketoadipate pathway. Most also grew on protocatechuate and coumerate. Of the two compounds typically converted to catechol by soil microbes, anthranilate was a good growth substrate for most of the roseobacter group isolates, and salicylate was a good growth substrate for half of these marine strains (Table 2.1).

3,4-PCD enzyme assays. The ability to grow on POB was universal for the six roseobacter isolates, and since this compound might be expected to generate protocatechuate during catabolism, 3,4-PCD enzyme activity was assayed. When grown with 4 mM pob, 3,4-PCD activity was detected in cell-free extracts of four of the isolates (*S. stellata* E-37, GAI-16, S25com04, and IC4) (Table 2.2). This activity appeared to be inducible since it was not detected when any of these isolates were grown in the presence of acetate, a non-aromatic growth substrate. No 3,4-PCD activity was detected when Y3F or *Sulfitobacter* sp. EE-36 were grown with POB or when Y3F was grown with vanillate.

Identification of catabolic genes in *S. stellata* E-37 and isolate Y3F. To identify the *pcaGH* genes from the roseobacter group isolates we relied on the DNA sequence of soil bacteria available in the GenBank database. Efforts focused on two isolates that grew well on pob and protocatechuate, one having detectable 3,4-PCD enzyme activity in cell-free extracts (*S. stellata* E-37), and the other not (isolate Y3F). Three degenerate DNA probes for 3,4-PCD were designed based on sequence alignments of the genes encoding the α and β subunits of 3,4-PCD from *Acinetobacter* sp. ADP1, *P. putida*(ATCC 23975), and *Burkholderia cepacia* DB01. In dot blot hybridizations, one of the three probes,

pcaHF1, successfully hybridized with DNA from *Acinetobacter* sp. ADP1, *Pseudomonas putida* F1, and *S. stellata* E-37, but not with DNA from Y3F or the other *Roseobacter* strains (data summarized in Table 2.3). DNA from *E. coli* , which does not contain the 3,4-PCD enzyme, did not hybridize with this probe. Southern blot analysis revealed that a 1.7 kbp *Bam*HI fragment from *S. stellata* E-37 hybridized with the pcaHF1 probe, as did the expected 2.4 kbp *Hin*dIII fragment from *Acinetobacter* sp. ADP1 (19). Under low stringency conditions, a 1.5 kbp *Bam*HI fragment from isolate GAI-16 gave a positive hybridization signal with the pcaHF1 probe.

Isolation, cloning, and sequencing of the fragment from *S. stellata* E-37 genomic DNA showed it contained an open reading frame (ORF) of 723 bp immediately followed by an incomplete ORF of 429 bp. Homology searches with sequences from the database revealed these two ORFs had highest similarity to the two subunits of 3,4-PCD (Table 2.4), and we designated these ORFs *pcaH* and *pcaG*. Southern hybridizations were carried out to identify and isolate the adjacent DNA fragment containing the terminal portion of the putative *pcaG*. ORFs showing significant similarity to other catabolic genes, γ-carboxymuconolactone decarboyxlase (*pcaC*) and *p*-hydroxybenzoate hydroxylase (*pobA*), and a LysR trancriptional regulator (*pcaQ*) were also isolated and identified on adjacent DNA fragments from *S. stellata* E-37 (Fig. 2.2A and Table 2.4A).

Although Southern hybridization of genomic DNA from isolate Y3F did not yield a detectable signal using the pcaHF1 probe, the possibility remained that this isolate could encode 3,4-PCD. To test this possibility, a new oligonucleotide probe (pcaHF1ND) was designed to anneal to a position approximately 350 nucleotides into

pcaH, based on the sequence information obtained from *S. stellata* E-37. This probe was found to hybridize with DNA from isolate Y3F. A 2.0 kbp *Bam*HI fragment from isolate S25com04 also gave a positive signal with the pcaHF1ND probe (summarized in Table 2.3). This oligonucleotide was then used in a PCR amplification reaction as the forward primer, with a reverse primer based on the sequence of a terminal region in the *S. stellata* E-37 *pcaG* gene. Using either *S. stellata* E-37 or Y3F template DNA and the same amplification conditions, *S. stellata* E-37 yielded the expected product of approximately 950 bp while isolate Y3F yielded a significantly smaller product of 390 bp. Sequence determination of this smaller fragment and comparisons with known *pcaGH* sequences indicated that the pcaHF1ND primer had not annealed in Y3F to *pcaH* as predicted but rather to the homologous portion of *pcaG*. A DIG-labeled probe made from the PCR product was used in Southern blot analysis and detected a genomic *Bam*HI fragment of approximately 3.0 kbp size. This fragment was isolated on plasmid pABX20 after appropriately-sized fragments were ligated into a standard cloning vecor. A series of subclones facilitated the sequencing of the fragment (Fig. 2.2B). Sequence determination of these subclones revealed ORFs with homology to *pcaG* and *pcaH*, as well as two additional ORFs showing homology to *pcaC* and *pcaQ* (Table 2.4B).

Sequence analysis of *pcaG* and*H* from *S. stellata* E-37 and isolate Y3F. The deduced sequences of PcaG and PcaH, the putative α and β subunits of 3,4-PCD, indicated molecular masses of 21.9 and 26.7 kDa in *S. stellata* E-37, and 22.9 and 26.5 kDa in isolate Y3F. These values correspond well with those reported for 3,4-PCDs in *Acinetobacter* sp. ADP1, *P. putida* (ATCC 23975)*, P. marginata* (ATCC 10248)*, B.*

cepacia DB01*, Azotobacter vinelandii*, *Brevibacterium fuscum*, and a *Moraxella* sp. (i.e. 23.2 ± 1.1 for PcaG and 28.8 ± 5.1 kDa for PcaH) (41).

Pairwise comparisons of the amino acid sequences deduced from *pcaG* and *pcaH* from the two roseobacter isolates were carried out with the corresponding sequences from the soil microbes *Acinetobacter* sp. ADP1, *B. cepacia* DB01, *P. putida* (ATCC 23975), and *R. opacus* 1CP. The percentages of identical residues for PcaG were 39-59% with isolate Y3F and 42-52% with *S. stellata* E-37. Comparisions of PcaH sequences revealed 47-62% identity for isolate Y3F and 47-61% identity for *S. stellata* E-37. In these alignments, residues known to be important for catalytic function (32) were well conserved. The close arrangement of the *pcaG* and *pcaH* genes, which are separated by 1 bp in isolate Y3F and overlap by 1 bp in *S. stellata* E-37, are consistent with the likelihood that these genes are cotranscribed and could be translationally coupled to form non-identical subunits of the same enzyme (29). Pairwise comparisons with all PcaGH sequences in the database revealed the deduced amino acid similarity was highest between *S. stellata* E-37 and isolate Y3F (i.e. 67% similarity; 60% identity). High similarity was also evident at the nucleotide level, with *pcaGH* from the two roseobacter group isolates showing 66% identity.

Expression of the *pcaGH* genes in *E. coli*. The *pcaGH* genes from the two roseobacter group isolates were expressed in *E. coli*, a bacterium that does not encode 3,4-PCD. These genes from *S. stellata* E-37 and isolate Y3F were placed under the control of an IPTG-inducible promoter on the expression vector pCYB1. The presence of *pcaGH* from either marine bacterium resulted in IPTG-inducible 3,4-PCD activity in cell-free

extracts of the plasmid-bearing *E. coli* strains. The Y3F-encoded activity (19±8 U)was lower than that of *S. stellata* E-37 (113±10 U), and required different IPTG induction conditions. Nevertheless, in each case this activity was inducible and linear with respect to the amount of cell extract added in the assay, indicating that the *pcaGH* genes of both marine isolates encode 3,4-PCD. No activity was detected with the pCYB1 vector alone or by Top10F' (pABX07) and Top10F'(pABX26) in the absence of IPTG.

Detection of *pcaH* in other members of the roseobacter group. Southern hybridization analysis using two different *pcaH* probes failed to yield a positive hybridization signal with DNA from two of the six roseobacter isolates, IC4 and *Sulfitobacter* sp. EE-36. To investigate further the possible presence of *pcaGH* in these strains, a PCR based approach was used. A degenerate primer pair was designed from alignments of the deduced amino acid sequences of PcaH (Fig. 2.3). This primer pair sucessfully amplifed a 159 bp product from both *Sulfitobacter* sp. EE-36 and isolate IC4. In addition, a PCR fragment of the expected size was amplified from DNA of *S. stellata* E-37. The sequence of these PCR fragments indicated homology to *pcaH*. Comparisons of the deduced amino acid sequence over this 53 residue stretch of PcaH showed significant conservation within this region among the roseobacter group isolates, as well as among other bacteria in the database (Table 2.5). The entire *pcaH* regions of the marine bacteria were only studied from isolate Y3F and *S. stellata* E-37, but the full sequences of *pcaH* and *pcaG* from these strains as well as analysis of the adjacent genetic regions support the presence of the β-ketoadipate pathway in bacteria of the roseobacter group.

Genes adjacent to *pcaGH* that may encode proteins of the β-ketoadipate pathway. The two 390 bp ORFs immediately upstream of *pcaH* in both *S. stellata* E-37 and isolate Y3F were homologous to a portion of *pcaL* from *Streptomyces coelicolor* A3(2) (Table 2.4). In the Gram positive bacteria *S. coelicolor*A3(2) and *Rhodococcus opacu* 1CP, 3-oxoadipate enol-lactone hydrolase and 4-carboxymuconolactone decarboxylase, termed PcaC and PcaD in Gram negative soil microorganisms (Fig. 2.1), appear to have fused into one protein demonstrating both activities, designated PcaL (8, 24). The ORFs of the two *Roseobacter* group were homologous to the decarboxylase (PcaC-like) segment of PcaL, which comprises the C-terminal third of the protein, and no similarity was found to the hydrolase portion of the protein. In light of the size and sequence of the ORFs, we designated the roseobacter genes *pcaC*. Their deduced amino acid sequences were 35-54% identical to PcaC sequences from soil bacteria *Acinetobacter* sp. ADP1 (P20370), *Bradyrhizobium japonicum* USDA110 (Y10223), and *P. putida* (ATCC 23975) (P0081). The *pcaC* genes from the two roseobacter group isolates showed high similarity to each other at the nucleic acid (77% identity) and amino acid (80% similarity; 74% identity) levels. The stop codon of *pcaC* in *S. stellata* E-37 is located 14 bp upstream of the initation codon for *pcaH*, while 1 bp separates *pcaC* and *pcaH* in isolate Y3F.

In *S. stellata* E-37, an 1179-bp ORF, designated *pobA*, was 53 bp upstream of and transcribed in the same direction as *pcaC*. It may encode a hydroxylase for the conversion of POB to protocatechuate (Fig. 2.1). Its deduced sequence was 56-61% identical to PobA sequences from soil bacteria *Azotobacter chroococcum* (AF019891), *Rhizobium leguminosarum* biovar viciae (L23969), *P. fluorescens* (X68438), *P. putida*

PAO1C (P20586), and *Acinetobacter* sp. ADP1 (Q03298). The regions associated with flavin adenine dinucleotide (FAD) and substrate binding were highly conserved among all the sequences. The 21 amino acids involved with FAD binding in *P. putida* PAO1C PobA showed 100% identity to the deduced protein from *S. stellata* E-37, while seven of the eight amino acids associated with substrate binding were also identical (48). As with PobA from *Acinetobacter* sp. ADP1, the single discrepancy is at residue position 124, where the alanine residue that typically occupies this position is replaced by valine in *S. stellata* E-37 and by serine in *Acinetobacter* sp. ADP1. The putative protein in *S. stellata* E-37 appears to contain an additional 3 residues between the completely conserved amino acids at position 127 (Val) and position 137 (Pro) of the corresponding proteins from soil bacteria.

Upstream of and divergently transcribed from the *pca* catabolic genes in both *S. stellata* E-37 and isolate Y3F were ORFs that we designated *pcaQ* on the basis of the similarity of their deduced amino acid sequences to the LysR-type transcriptional activator protein PcaQ from *A. tumefaciens* A348. These putative roseobacter regulatory proteins showed 39-43% identity to PcaQ from *A. tumefaciens* A348 and 18-24% identity to several LysR-type regulators of the catechol branch of the β-ketoadipate pathway, including CatR from *P. putida* (P20667), and BenM and CatM from *Acinetobacter* sp. ADP1 (AF009224). The region of highest similarity among the PcaQ proteins was in the N-terminus, an area presumed to comprise a helix-turn-helix motif for DNA binding (50). In *S. stellata* E-37, *pcaQ* was 78 bp upstream of *pobA*, while in isolate Y3F, it was 86 bp upstream of *pcaH*. By analogy to PcaQ of *A. tumefaciens* A348

and other LysR type regulators, the roseobacter PcaQ proteins might be expected to regulate their own synthesis and also control the expression of those genes downstream of and divergently transcribed from *pcaQ* (37).

Discussion

Despite the ecological importance of the β-ketodipate pathway for the degradation of aromatic compounds in a variety of environments, studies of this pathway have focused primarily on soil microorganisms, with an emphasis on bacterial groups associated with plants (e.g. *Pseudomonas*, *Agrobacterium*, *Rhizobium*). Here we report the identification of a key enzyme of this pathway, 3,4-PCD, in members of a marine lineage in the α-proteobacteria. Although no single method was successful with all isolates, multiple approaches using DNA probes and PCR primer sets indicated that all six roseobacter group members encode 3,4-PCD (Table 2.3). The identification of a cluster of ORFs in *S. stellata* E-37 and isolate Y3F with homology to genes encoding proteins of the β-ketodipate pathway in addition to 3,4-PCD strengthens the argument for the presence of the pathway. These initial characterizations of the marine bacteria provide a framework for comparisions of catabolism with their well-studied soil counterparts. Moreover, the presence of the β-ketoadipate pathway in marine roseobacter isolates, including an isolate cultured using non-selective methods, suggests that these bacteria are contributing to the degradation of aromatic components of vascular plant material in coastal rivers and marshes.

3,4-PCD in roseobacter group isolates. The level of 3,4-PCD activity in cell-free extracts from four of the roseobacter isolates (Table 2.2) was comparable to that observed in similar studies of POB-grown soil microbes (33, 52). It was surprising, therefore, that this activity was not detectable in POB- or vanillate-induced cell-free extracts of the Y3F isolate, from which the entire *pcaGH* genes were isolated, nor in POB-induced extracts of *Sulfitobacter* sp. EE-36, from which a portion of the *pcaH* gene was characterized. It may be that some of these environmental isolates are less sensitive to our method of cell lysis and/or that the functional enzymes were damaged by heat during sonication. Since the *pcaGH* genes of Y3F when expressed in *E. coli* resulted in a 6-fold lower 3,4-PCD specific activity than comparable experiments with *S. stellata* E-37, the Y3F enzyme could be more heat labile. It seemed unlikely that protocatechuate 4,5-dioxygenase rather than 3,4-PCD cleaved protocatechuate during growth with POB since no activity for this enzyme was detected in cell-free extracts from any of the six isolates (A. Buchan, unpublished data). The activity of another enzyme known to cleave protocatechuate, protocatechuate 2,3-dioxygenase, was not determined, although this enzyme has only been identified in one isolate (51). The possibility that 3,4-PCD failed to be induced by the presence of POB (and vanillate in the case of Y3F) because metabolism proceeds via a ring-cleavage substrate other than protocatechuate for these particular compounds in Y3F or *Sulfitobacter* sp. EE-36 was not investigated.

Phylogenetic analysis of the α and β subunits of 3,4-PCD obtained from *S. stellata* E-37 and isolate Y3F indicated that the genes from the roseobacter isolates cluster with low bootstrap values with the genes from *Acinetobacter* sp. ADP1 (Fig. 2.4). The six

residues shown to be involved in ligand binding and the four residues demonstrated to be involved in the active site of *P. putida* PRS2000 (30) are conserved in both roseobacter isolates. The only exception is in *S. stellata* E-37, for which the ligand binding residue in PcaH located at position 148 (*S. stellata* E-37 numbering) is replaced by proline, where typically either an arginine or glycine is found (Fig. 2.3).

The α and β subunits of *S. stellata* E-37 show a high degree of similarity to one another (45% similarity, 37% identity) as do the α and β subunits of isolate Y3F (36% similarity, 42% identity). This indicates a common ancestry of the two subunits, as previously suggested for the α and β subunits of other 3,4-PCD enzymes (30). Among soil bacteria, and now marine roseobacter isolates, the amino acids in the β subunit (PcaH) are more conserved than those in the α subunit (PcaG), a fact that has been attributed to the presence of the four Fe ligand residues within the β subunit (32, 49).

Genetic Organization. The *pcaGH* genes in *S.stellata* E-37 and isolate Y3F were linked to other genes likely to be involved in protocatechuate degradation in a manner similar to that of previously characterized bacteria (Fig. 2.5). Genes encoding enzymes of the β-ketoadiapte pathway are generally clustered on the chromosome in supraoperonic units, arrangements of linked genes and operons having related physiological functions (8, 21, 31, 33). In *Acinetobacter* sp. ADP1, the operonic units needed for POB, protocatechuate, shikimate and quinate degradation are located within a 18 kbp region of the chromosome (13), although in other organisms these operonic units can be separated by more than 10 kbp, as is the case for *P. putida* and *R. opacus* 1CP (21). The *pcaG* and *H* genes are adjacent to each other in all characterized bacteria, most likely because they encode a two-

subunit enzyme and hence would be expected to evolve as a unit (21). The genetic arrangement in *S. stellata* E-37 differs from that of isolate Y3F in having *pobA* between *pcaQ* and *pcaC*.

At the level of nucleic and deduced amino acids, the catabolic genes identified from *S. stellata* E-37 and isolate Y3F are more similar to each other than to analogous genes from other organisms. Whether these genes were obtained before the roseobacter isolates and other α-proteobacteria diverged, or whether they were obtained by horizontal gene transfer at a later time is unclear. The discontinuities in evolutionary distances of *pca* genes relative to presumably stable markers such as 16S rRNA gene sequences (Fig. 2.4) make these questions difficult to address, but clearly portray the dynamic nature of this group of catabolic genes.

Regulation. This dynamic nature is also reflected in a variety of different classes of transcriptional regulators that have been found to modulate gene expression in the β-ketoadipate pathway of *Acinetobacter* sp. ADP1, *P. putida* PRS2000, *R. opacus* 1CP and *A. tumefaciens* A348 (7, 13, 20, 34, 40, 42)(Fig. 2.5). Investigations of the LysR-type regulator PcaQ in *A. tumefaciens* A348 (37) as well as studies demonstrating the widespread distribution of PcaQ homologs among members of the Rhizobiaceae family, prompted Parke (39) to suggest that the absence of PcaQ may be the exception rather than the rule for the protocatechuate branch of the pathway within the α-proteobacteria. This suggestion is further supported by the identification of PcaQ homologs in our two *Roseobacter* group isolates (Table 2.4). Nevertheless, a putative PcaQ was recently

reported in a *Pseudomonas* strain (a member of the α-proteobacteria) directly upstream of *pcaHG* (33), perhaps indicating a less recent dispersal of this regulator.

In *S. stellata* E-3, *pobA*, encoding a putative enzyme for the conversion of POB to protocatechuate, is located immediately upstream of the *pca* catabolic genes (Fig 2.5). Thus far the *pcaQ-pobA* orientation in *S. stellata* E-37 appears to be unique and raises the possibility that in this organism, *pobA* is under the regulatory control of *pcaQ* and regulated in conjunction with the other *pca* genes identified. In *Acinetobacter* sp. ADP1 and *A. tumefaciens* A348, *pobA* is activated by a transcriptional regulator, PobR, that responds to *p*-hydroxybenzoate and is not a member of the LysR-type transcriptional regulatory family (7, 40). A PobR homolog has also been found directly upstream of *pobA* in *Pseudomonas* sp. HR199 (33) (Fig. 2.5).

Ecological Significance. Marine bacteria affiliated with the roseobacter clade have recently been found to be abundant in the estuaries of the southeastern U.S., where they can contribute up to 30% of the bacterioplankton 16S rRNA genes (16). To date, the roseobacter clade is one of only a few dominant lineages of marine bacteria known to be readily amenable to culturing (14), making laboratory studies of their physiology and genetics particularly important.

Five of the six roseobacter isolates examined in this study were cultured following enrichment or selection for growth on lignin. Although not common in many marine ecosystems, lignin can be an important source of organic matter in vascular plant-dominated coastal marshes, including the *Spartina alterniflora* marshes from which these isolates were obtained (27). As vascular plant material decays, the lignin is converted to

smaller aromatic compounds such as benzoate, catechol, POB, cinnamate, *p*-coumarate, vanillate, ferulate, quinate, and shikimate (46). Many soil bacteria convert these lignin-related monomers to protocatechuate and then degrade them via the β-ketoadipate pathway (5, 22, 36). Marine bacteria in the roseobacter clade apparently degrade lignin-related compounds in the same manner. The presence of ring-cleavage dioxygenases and associated genes of the β-ketoadipate pathway in marine roseobacter isolates provides a valuable resource for comparative studies on the regulation and ecology of aromatic compound degradation in diverse environments.

Acknowledgements

José González kindly provided some of the roseobacter isolates and advice on culturing methods. Nathaniel Cosper helped with attempted efforts to complement *Acinetobacter* mutants, and he provided helpful discussions concerning the 3,4-PCD enzyme. This work was supported by the NSF through the Biological Oceanography Program OCE-9730745 (to M.A.M.), MCB-9808784 (to E.L.N), and a traineeship (to A.B.) provided through a research training grant in prokaryotic diversity BIR-9413235.

References

1. Bradford, M.M. 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248-254

2. Buchan, A., E. L. Neidle, and M. A. Moran. Unpublished data.

3. Dagley, S. 1986. Biochemistry of aromatic hydrocarbon degradation in Pseudomonads. P 527-555. In J.R. Sokatch (Ed.), The Bacteria, Vol 10. Academic Press Inc., New York.

4. Dang, H. and C. R. Lovell. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis 16S rRNA genes. Appl. Envir. Microbiol. 66: 467-475.

5. Delneri, D., G. Degrassi, R. Rizzo, C. V. Bruschi. 1995. Degradation of trans-ferulic acid and *p*-courmaric acid by *Acinetobacter calcoaceticus* DSM 586. Biochim. Biophys. Acta. 1244:363-367.

6. Devereaux, J., P. Haebverli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

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

8. Eulberg, D., S. Lakner, L. A. Golovleva, and M. Schlömann. 1998. Characterization of a protocatechuate catabolic gene cluster from *Rhodococcus opacus* 1CP: evidence for a merged enzyme with 4-carboxymuconolactone-decarboxylating and 3-oxoadipate enol-lactone-hydrolyzing activity. J. Bacteriol. 180:1072-1081.

9. Felsenstein, J. 1989. PHYLIP-Phylogeny Inference Package (version 3.2). Cladistics 5:164-166.

10. Frazee, K. W., D. M. Livingston, D. C. LaPorte, and J. D. Lipscomb. 1993. Cloning, sequencing, and expression of *Pseudomonas putida* protocatechuate 3,4-dioxygenase genes. J. Bacteriol. 175:6194-6202.

11. Garcia, E. N., I. G. Siegert, and P. Suarez. 1998. Toxicity assays and naphthalene utilization by natural bacteria selected in marine environments. Bull. Environ. Contam. Toxocol. 61:370-377.

12. Geiselbrecht, A. D., B. P. Hedlund, M. A. Tichi, and J. T. Staley. 1998. Isolation of marine polycyclic aromatic hydrocarbon (PAH)-degrading *Cycloclasticus* strains from the Gulf of Mexico and comparision of their PAH degradation ability with that of Puget Sound *Cycloclasticus* strains. Appl. Environ. Microbiol. 64:4703-4710.

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

14. Giovannoni, S., and M. Rappé. 2000. Evolution, diversity, and molecular ecology of marine prokaryotes, p47-84. *In* D. Kirchman (ed.), Microbial ecology of the oceans. Wiley Interscience, New York, N. Y.

15. González, J. M., W. B. Whitman, R. E. Hodson, and M. A. Moran. 1996. Identifying numerically abundant culturable bacteria from complex communities: An Example from a lignin enrichment culture. Appl. Environ. Microbiol. 62:4433-4440.

16. González, J. M. and M. A. Moran. 1997. Numerical dominance of a group of marine bacteria in the α-subclass of the class proteobacteria in coastal seawater. Appl. Environ. Microbiol. 63:4237-4242.

17. González, J. M., F. Mayer, M. A. Moran, R. E. Hodson, and W. B. Whitman. 1997. *Sagittula stellata* gen. nov., sp. nov., a lignin-transforming bacterium from a coastal environment. Int. J. Syst. Bacteriol. 47(3):773-780.

18. Harayama, S., M. Kok, and E. L. Neidle. 1992. Functional and evolutionary relationships among diverse oxygenases. Ann. Rev. Microbiol. 46:565-601.

19. Hartnett, C., E. L. Neidle, K. L. Ngai, and L. N. Ornston. 1990. DNA sequences of genes encoding *Acinetobacter calcoaceticus* protocatechuate 3,4-dioxygenase: evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence. J.Bacteriol. 172:956-966.

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

21. Harwood, C. S. and R. E. Parales. 1996. The β-ketoadipate Pathway and the Biology of Self-Identity. Ann. Rev. Microbiol. 50:553-590.

22. Hawkins, A. R., H. K. Lamb, J. D. Moore, I .G. Charles and C. F. Roberts. 1993. The pre-chorimsate (shikimate) and quinate pathways in filamentous fungi: theoretical and practical aspects. J. Gen. Microbiol. 139:2891-2899.

23. Hedlund, B. P., A. D. Geiselbrecht, T. J. Blair, and J. T. Staley. 1999. Polycyclic aromatic hydrocarbon degradation by a new marine bacterium, *Neptunomonas naphtovorans* gen. nov., sp. nov. Apple. Environ. Microbiol. 65:251-259.

24. Iwagami, S. G., Y, Keqian, and J. Davies. 2000. Characterization of the protocatechuic acid catabolic gene cluster from *Stretomyces* sp. strain 2065. Appl Environ Microb. 66:1499-1508.

25. Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall. 1951. Protein Measurement with the Folin Phenol Reagent. J. Biol. Chem. 193:265-275.

26. Moran, M. A. and R. E. Hodson. 1989. Formation and bacterial utilization of dissolved organic carbon derived from detrital lignocellulose. Limnol. Oceanogr. 35:1744-1756.

27. Moran, M. A. and R. E. Hodson. 1994. Dissolved humic substances of vascular plant origin in a coastal marine environment. Limnol. Oceanogr. 39:762-771.

28. Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni. 1995. Genetic comparisions reveal the same unknown bacerial lineages in Atlantic and Pacific bacterioplankton communities. Limnol. Oceaogr. 40:148-158.

29. Normark, S., S. Bergstrom, T. Edlund, T. Grundstrom, B. Jaurin, F. P. Lindberg, and O. Olsson. 1983. Overlapping genes. Annu. Rev. Genet. 17: 499-525.

30. Ohlendorf, D.H., A. M. Orville, and J. D. Lipscomb. 1994. Structure of protocatechuate 3,4-dioxygenase from *Pseudomonas aeruginosa* at 2.15Å resolution. J. Mol. Biol. 244:586-608.

31. Ornston, L. N., J. Houghton, E. L. Neidle, and L. A. Gregg. 1990. Subtle selection and novel mutation during evolutionary divergence of the β-ketoadiapte. In: *Pseudomonas*: Biotransformations, Pathogenesis and Evolving Biotechnology. p 207-225.

32. Orville, A. M., J. D. Lipscomb, and D. H. Ohlendorf. 1997. Crystal structures of substrate and substrate analog complexes of protocatechuate 3,4-dioxygenase: Endogenous Fe super (3+) ligand displacement in response to substrate binding. Biochem. 36:10052-10066.

33. Overhage, J. A., U. Kresse, H. Priefert, H. Sommer, G. Krammer, J. Rabenhorst, and A. Steinbüchel. 1999. Molecular characterization of the genes *pcaG* and *pcaH*, encoding protocatechate 3,4-dioxygenase, which are essential for vanillin catabolism in *Pseudomonas* sp. strain HR 199. Appl. Environ. Microbiol. 65:951-960.

34. Parales, R. E., and C. S. Harwood. 1993. Regulation of the *pcaIJ* genes for aromatic acid degradation in *Pseudomonas putida*. J. Bacteriol. 175:5829-5838.

35. Pardos, J., E. Pelletier, R. Siron, and D. Delille. 1999. Fate of a new silicone-based oil treating agent and its effect on marine microbial communities. Environ. Tox. Chem. 18:819-827.

36. Parke, D., F. Rynne, and A. Glenn. 1991. Regulation of phenolic catabolism in *Rhizobium leguminosarum* biovar trifolii. J. Bacteriol. 173:5546-5550.

37. Parke, D. 1993. Positive regulation of phenolic catabolism in *Agrobacterium tumefaciens* by the *pcaQ* gene in response to β-carboxy-cis-cis muconate. J. Bacteriol. 175:3529-3535.

38. Parke D., 1996. Characterization of PcaQ, a LysR-type transcriptional activator required for catabolism of phenolic compounds, from *Agrobacterium tumefaciens*. J. Bacteriol. 178:266-272.

39. Parke D., 1996. Conservation of PcaQ, a transcriptional activator of *pca* genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. J. Bacteriol. 178:3671-3675.

40. Parke D., 1997. Acquistion, reorganization, and merge of genes: Novel management of the β-ketoadipate pathway in *Agrobacterium tumefaciens*. FEMS Microbiol. Letters. 146:3-12.

41. Petersen, E. I., J. Zuegg, D. W. Ribbons, H. Schwab. 1996. Molecular cloning and homology modeling of protocatechuate 3,4-dioxygenase from *Pseudomonas marginata*. Microbiol. Res. 151:359-370.

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

43. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

44. Shanley, M. S., E. L. Neidle, R. E. Parales, and L. N. Ornston. 1986. Cloning and expression of *Acinetobacter calcoaceticus catBCDE* genes in *Pseudomonas putida* and *Escherichia coli*. J. Bacteriol. 165.557-563.

45. Stanier, R. Y. and J. I. Ingraham. 1954. Protocatechuic acid oxidase. J. Biol. Chem. 210:799-808

46. Subba Rao, P. V., A. M. D. Nambudiri and J. V. Ghat. 1971. Microbial degradation of phenylpropaoid compounds. J. Sci. Indust. Res. 30:663-679.

47. Suzuki, M. T., M. S. Rappé, Z. W. Haimberger, H. Winfield, N. Adair, J. Ströbel, and S. J. Giovannoni. 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl. Environ. Microbiol. 63:983-989.

48. Van Berkel, W., A. Westphal, K. Eschrich, M. Eppink, and A. Dekok. 1992. Substitution of Arg214 at the substrate-binding site of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. Euro. J. Biochim. 210:411-419.

49. Vetting. M. W., D. A. D'Argenio, L. N. Ornston, and D. H. Ohlendorf. 2000. Structure of *Acinetobacter* sp. ADP1 protocatechuate 3,4-dioxygenase at 2.2 Å resolution: implications for the mechanisms of an intradiol dioxygenase. Biochem. 39:7943-7955.

50. Viale, A. M., H. Kobayashi, T. Akazawa, and S. Henikoff. 1991. *rcbR*, a gene coding for a member of the LysR family of transcriptional regulators, is located upsteram of the expressed set of ribulose 1,5-bisphosphate carboxylase oxygenase genes in the photosynthetic bacterium *Chromatium vinosum*. J. Bacteriol. 173(16):5224-5229.

51. Wolgel, S. A., J. E. Dege, P. E. Perkinsolson, C. H. Juarezgarcia, R. L. Crawford, E. Munck, and J. D. Lipscomb. 1993. Purification and characterization of protocatechuate 2,3-dioxygenase from *Bacillus macerans*-a new extradiol catecholic dioxygenase. J. Bacteriol. 175:4414-4426.

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

Table 2.1. Growth on aromatic compounds by roseobacter group isolates

Isolate	Growth[a] of an isolate on:							
	anthranilate	benzoate	POB	salicylate	vanillate	ferulate	protocatechuate	coumarate
S. stellata E-37	+/—	—	+	—	—	+	+	+
Y3F	+	—	+	—	—	—	+	—
GAI-16	+	+/—	+	—	—	—	—	+/—
Sulfitobacter sp. EE-36	+	+	+	+	+	+	+	+
S25com04	+/—	+	+	+	+	+	+	+
IC4	+/—	+	+	+	+	+	+	+

[a]Growth was determined on MBM plates containing 4 mM growth substrate and 0.1% vitamins. +, growth within 4 days; +/—, growth within 14 days; —, no growth within 14 days.

Table 2.2. Specific activity of protocatechuate 3,4 dioxygenase (3,4-PCD) in cell-free extracts of roseobacter group isolates[a]

Isolate	Specific Activity (± standard deviation) (nmol/min/mg protein)	
	POB	acetate
S. stellata E-37	460 (± 110)	<5
GAI-16	310 (± 60)	<5
S25com04	730 (± 90)	<5
IC4	880 (± 120)	<5

[a]All isolates were grown in Marine Basal Medium with 4 mM substrate.

Table 2.3. Indicators of 3,4-PCD in the six roseobacter isolates

Isolate	Isolation substrate	Growth on POB	3,4-PCD activity[a]	Hybridization with *pcaH* probe	*pcaGH* retrieved[e]	*pcaGH* expression in *E. coli*[e]	*pcaH* fragment amplified[b]
S. stellata E-37	Indulin	+	+	+[c]	+	+	+
Y3F	Indulin	+		+[d]	+	+	
GAI-16	yeast extract	+	+	+[c]	na	na	
Sulfitobacter sp. EE-36	Indulin	+			na	na	+
S25com04	POB	+	+	+[d]	na	na	
IC4	Indulin	+	+		na	na	+

[a] Enzyme activity was detected spectrophotometrically in cell-free extracts of POB-grown cultures.

[b] A 159 bp PCR fragment was amplified using the P34OID primer set.

[c] Positive Southern hybridization signal with pcaHF1 probe was obtained.

[d] Positive Southern hybridization signal with pcaHF1ND probe was obtained.

[e] na = not attempted.

Table 2.4. β–Ketoadipate pathway genes identified in (A) *S. stellata* E-37 and (B) Y3F

Gene designation	Putative function of gene product	Size of gene (bp)	% G+C	Size of deduced gene product		Most similar gene (% amino acid similarity/identity) (GenBank/SwissProt accession no.)
				Residues	**kDa**	
A. *S. stellata* E-37:						
pobA	*p*-hydroxybenzoate hydroxylase	1179	63.6	393	44.1	*pobA Pseudomonas fluorescens* (70/61) (640342)
pcaC	γ-carboxymuconolactone decarboyxlase	393	64.9	131	14.4	*pcaL Steptomyces coelicolor* (55/44) (AL079355)
pcaG	α subunit 3,4-protocatechuate dioxygenase	603	65.1	201	21.9	*pcaG Acinetobacter* sp. ADP1 (61/52) (M33798)
pcaH	β subunit 3,4-protocatechuate dioxygenase	723	66.7	241	26.7	*pcaH Pseudomonas* sp. HR199 (70/61) (Y18527)
pcaQ	LysR-type transcriptional regulator	957	65.7	319	34.4	*pcaQ Agrobacterium tumefaciens*A348 (49/39) (P526681)
B. Y3F:						
pcaC	γ-carboxymuconolactone decarboyxlase	396	67.6	132	14.3	*pcaL Steptomyces coelicolor* (50/43) (AL079355)
pcaG	α subunit 3,4 protocatechuate dioxygenase	621	64.2	207	22.9	*pcaG Acinetobacter* sp. ADP1 (67/59) (M33798)
pcaH	β subunit 3,4-protocatechuate dioxygenase	729	63.9	243	26.5	*pcaH Pseudomonas putida* NCIMB 9869 (69/63) (U96339)
pcaQ	LysR-type transcriptional regulator	927	69.1	309	32.9	*pcaQ Agrobacterium tumefaciens* A348 (50/43) (P526681)

Table 2.5. Similarity among corresponding 53 amino acid regions of PcaH*

Organism	% Similarity (identity) with:									
	IC4[a]	Y3F[b]	EE36[c]	HR199[d]	PP[e]	E-37[f]	ADP1[g]	BC[h]	RO[i]	2065[j]
IC4	100(100)									
Y3F	100(98)	100(100)								
EE36	96(94)	96(94)	100(100)							
HR199	96(94)	96(94)	96(96)	100(100)						
PP	94(93)	94(93)	93(91)	93(91)	100(100)					
E-37	81(76)	81(76)	79(74)	81(76)	85(77)	100(100)				
ADP1	79(70)	79(72)	79(70)	77(68)	79(72)	79(62)	100(100)			
BC	70(66)	70(66)	72(68)	72(68)	72(66)	68(60)	93(91)	100(100)		
RO	70(66)	70(66)	70(66)	72(70)	72(68)	72(60)	76(72)	76(68)	100(100)	
2065	68(62)	68(62)	68(62)	66(60)	68(60)	64(57)	66(57)	72(66)	66(59)	100(100)

*Similarity and identities are shown for residues encoding the 159 bp located between the two P34OID primers.

[a]isolate IC4 (input accession number), [b]isolate Y3F (input accession number). [c]*Sulfitobacter* sp. EE36 (input accession number), [d]*Pseudomonas* sp. HR199 (418527), [e]*Pseudomonas putida* (L14836), [f]*Sagittula stellata* E-37 (input accession number), [g]*Acinetobacter* sp. ADP1 (M33798), [h]*Burkholderia cepacia* (M30791), [i]*Rhodococcus opacus* 1CP (AF003947), [j]*Streptomyces* sp. 2065. (AF019386).

Figure 2.1. Schematic of the protocatechuate branch of the β-ketoadipate pathway in some prokaryotes (21). Gene designations are shown in italics.

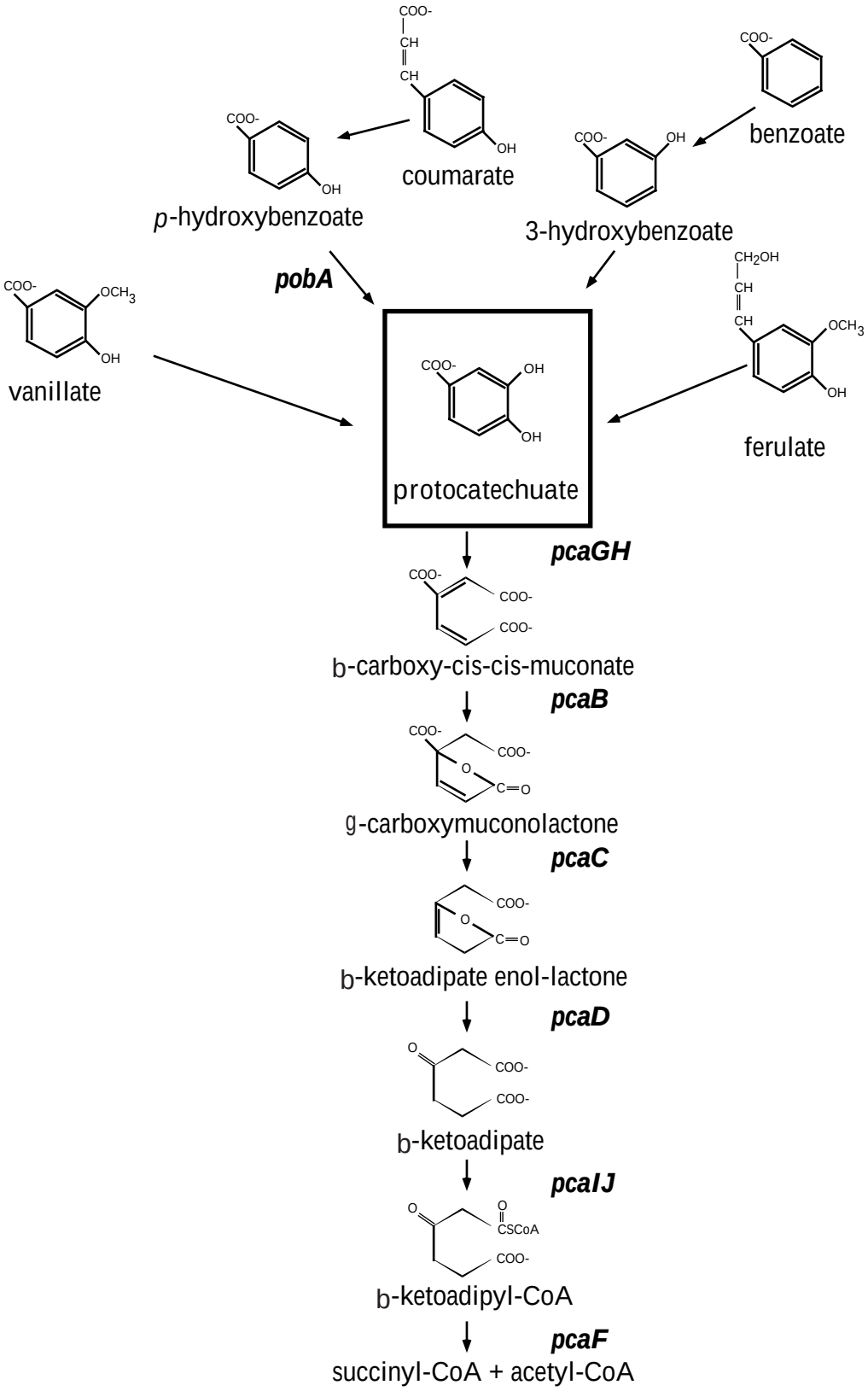
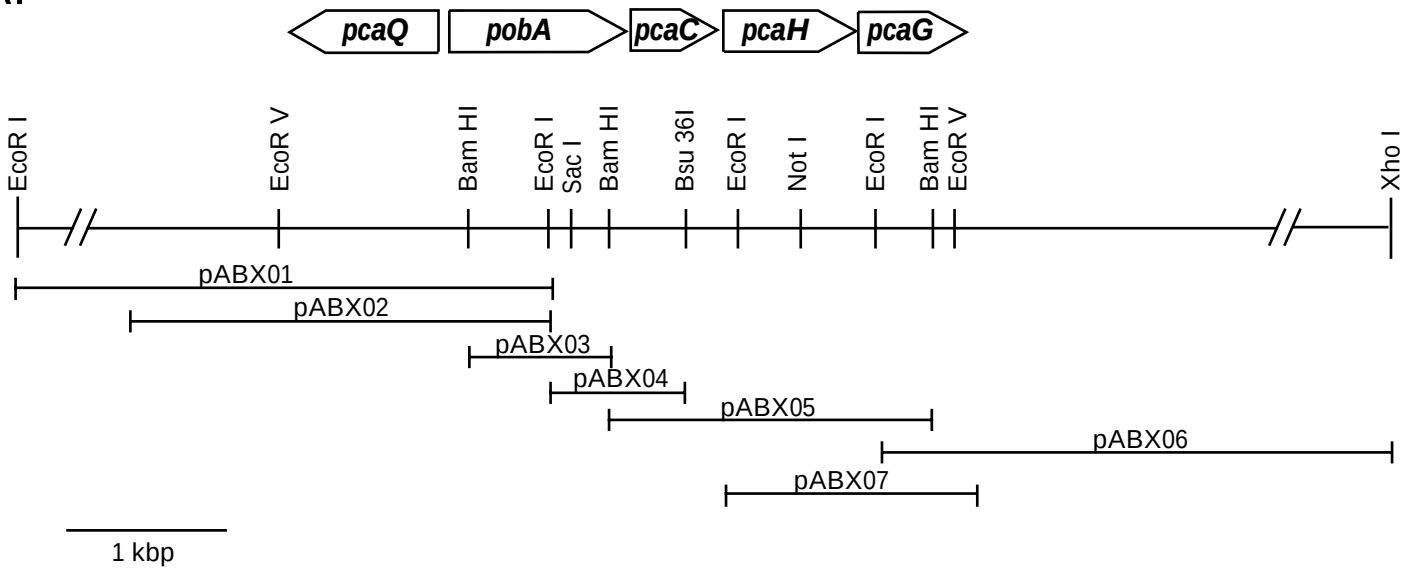

coumarate
p-hydroxybenzoate
benzoate
3-hydroxybenzoate
pobA
vanillate
protocatechuate
ferulate
pcaGH
b-carboxy-cis-cis-muconate
pcaB
g-carboxymuconolactone
pcaC
b-ketoadipate enol-lactone
pcaD
b-ketoadipate
pcaIJ
b-ketoadipyl-CoA
pcaF
succinyl-CoA + acetyl-CoA

Figure 2.2. Restriction map of the chromosomal *pca* regions from (A) *S. stellata* E-37 and (B) Y3F. The location of genes and their transcriptional directions are shown relative to selected restriction endonuclease recognition sites. Horizontal lines indicate the DNA regions contained on recombinant plasmids, whose designations are shown above the corresponding line.

A.


pcaQ
pobA
pcaC
pcaH
pcaG
EcoR I
EcoR V
Bam HI
EcoR I
Sac I
Bam HI
Bsu 36I
EcoR I
Not I
EcoR I
Bam HI
EcoR V
Xho I
pABX01
pABX02
pABX03
pABX04
pABX05
pABX06
pABX07
1 kbp

B.

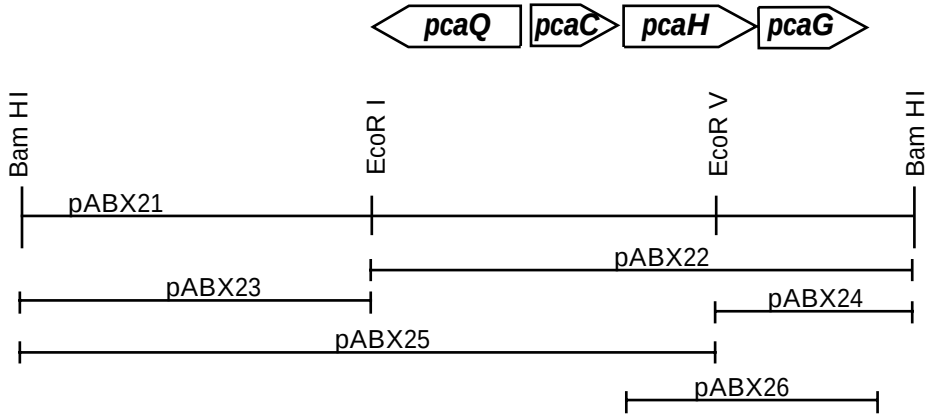

pcaQ
pcaC
pcaH
pcaG
Bam HI
EcoR I
EcoR V
Bam HI
pABX21
pABX22
pABX23
pABX24
pABX25
pABX26

Figure 2.3. Protein sequence alignments of PcaH. Aligned residues that are identical or similar are shown with black backgrounds or boxed, respectively. Residues presumed to be involved in substrate specificity are indicated by a dot, and those demonstrated to be involved in catalysis and Fe^{2+} binding in *P. putida* are indicated by an asterisk (32). Gray regions underscored with arrows indicate residues used for design of the P34OID degenerate PCR primers, with identical and similiar residues shaded gray or boxed, respectively. Alignments are shown for *Pseudomonas putida* (ATCC 23975) (L14836), *Acinetobacter* sp. ADP1(M33798), *Burkholderia cepacia* DB01(M30799), *Rhodococcus opacus* 1CP (AF003947), and *Streptomyces* sp. strain 2056 (AF019386). Alignments were conducted using the PILEUP program of the Genetics Computer Group package.

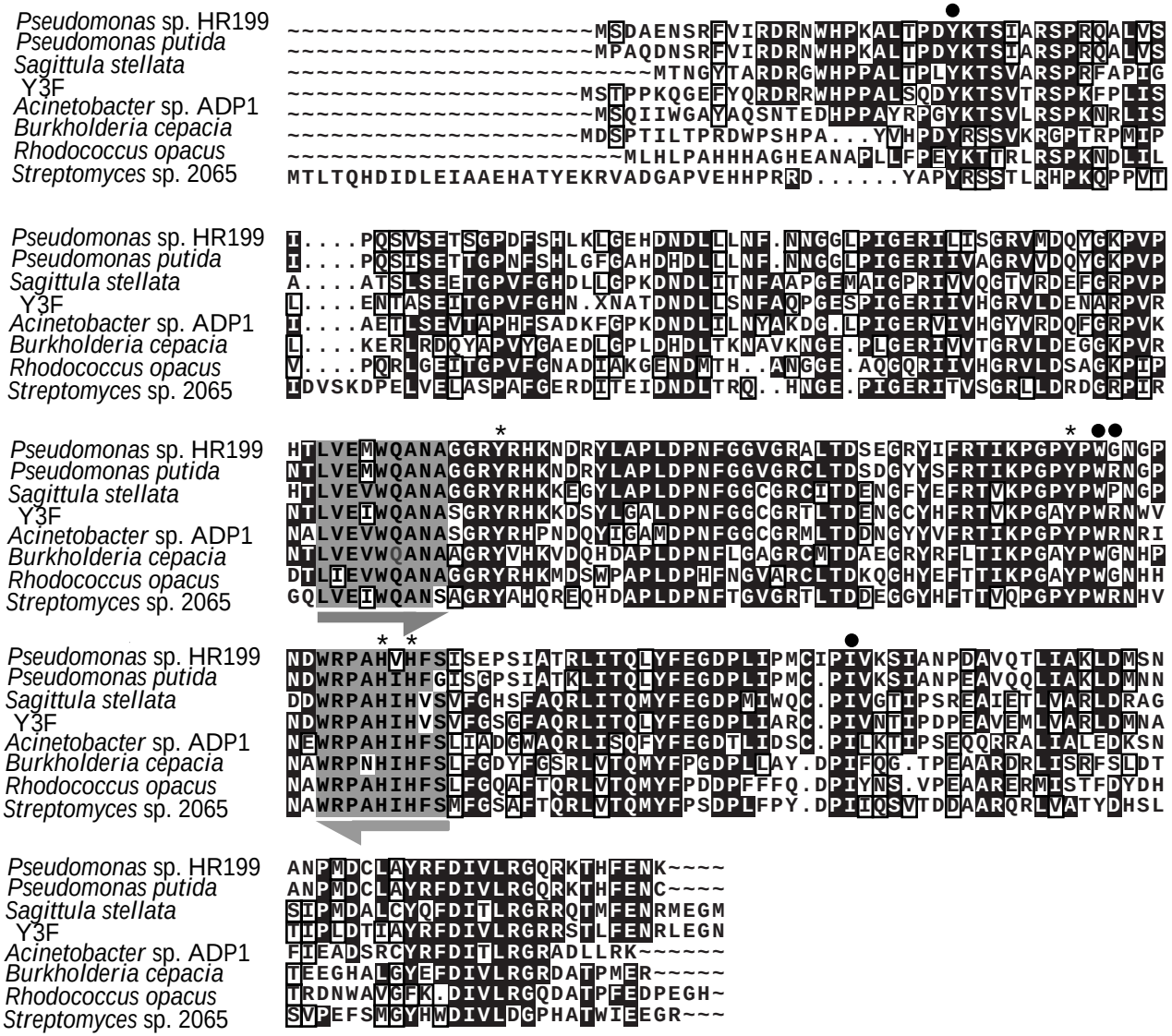

Pseudomonas sp. HR199
Pseudomonas putida
Sagittula stellata
Y3F
Acinetobacter sp. ADP1
Burkholderia cepacia
Rhodococcus opacus
Streptomyces sp. 2065

Figure 2.4. Phylogenetic tree of PcaGH protein sequences. The tree is based on the deduced amino acids of the *pcaGH* genes and is unrooted, with CatA from *Acinetobacter* sp. ADP1 (Z36909) as the outgroup. Bootstrap values greater than 50% are indicated at branch nodes. The scale bar indicates amount of genetic change measured as number of amino acid substitutions per site.

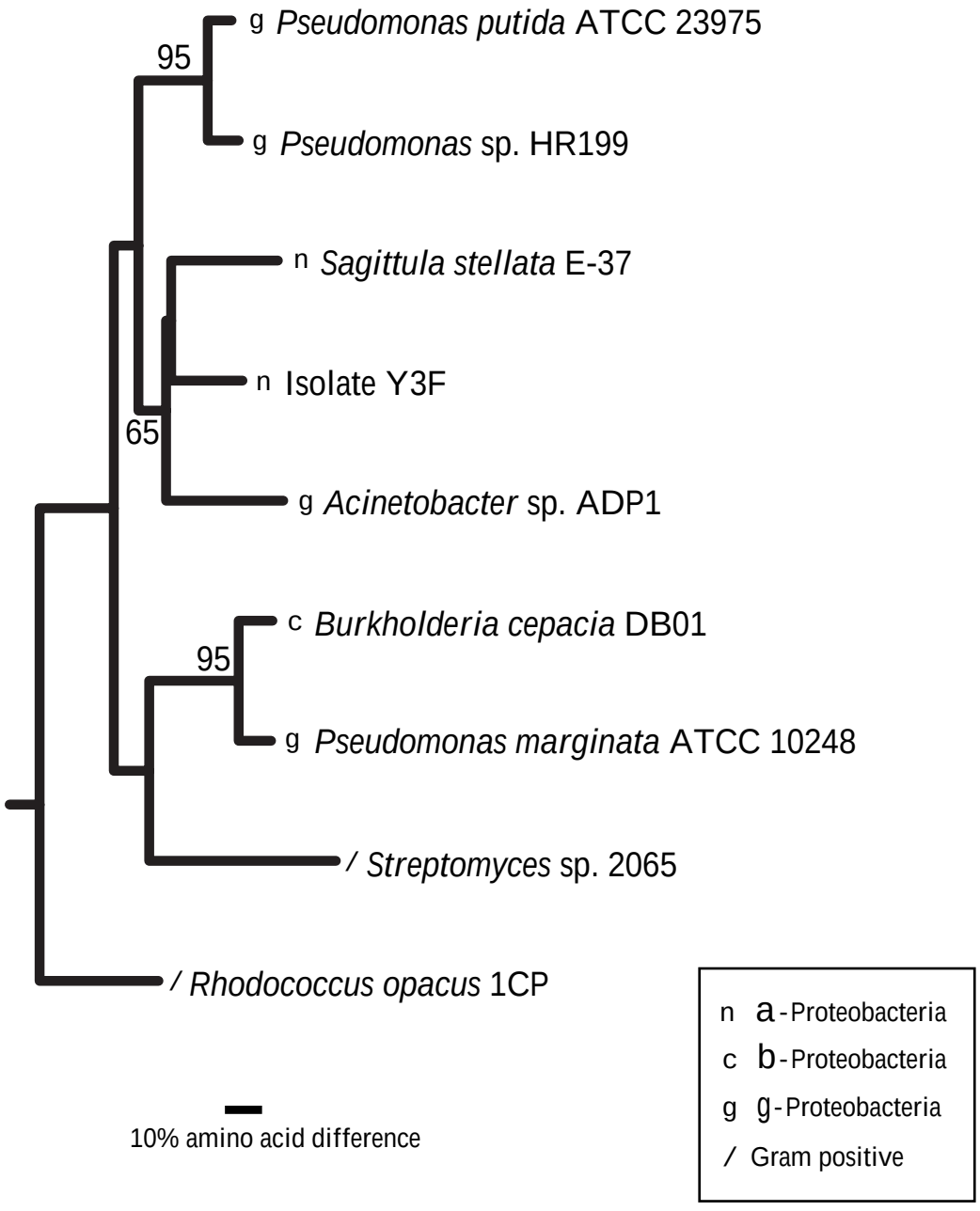
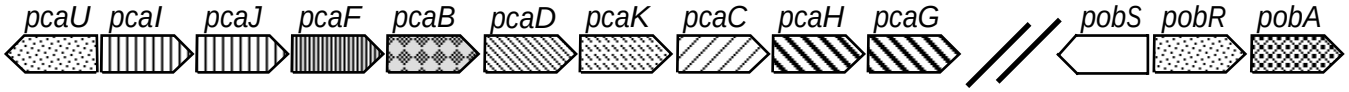
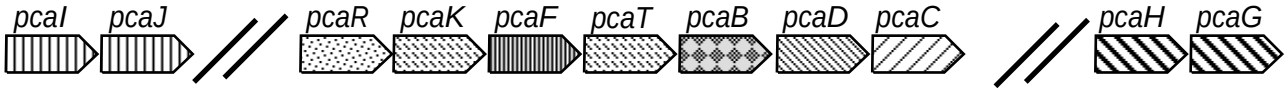

g Pseudomonas putida ATCC 23975
95
g Pseudomonas sp. HR199
n Sagittula stellata E-37
n Isolate Y3F
65
g Acinetobacter sp. ADP1
c Burkholderia cepacia DB01
95
g Pseudomonas marginata ATCC 10248
/ Streptomyces sp. 2065
/ Rhodococcus opacus 1CP
10% amino acid difference
n a-Proteobacteria
c b-Proteobacteria
g g-Proteobacteria
/ Gram positive

Figure 2.5. Organization of gene clusters for protocatechuate metabolism in selected bacteria. Arrows indicate direction of transcription. Bold double lines indicate genes separated by <10 kbp. The information was compiled from the following sources: *Acinetobacter* ADP1 (13), *P. putida* (8, 21, 40), *Pseudomonas* sp. HR199 (32), *R. opacus* 1CP (8), *A. tumefaciens* A348 (39), *S. stellata* E-37 and isolate Y3F (this study).

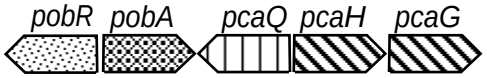
Acinetobacter sp. ADP1 (g-Proteobacteria)



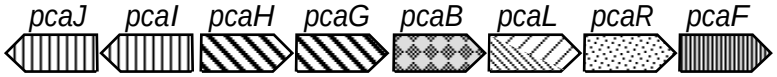
Pseudomonas putida (g-Proteobacteria)



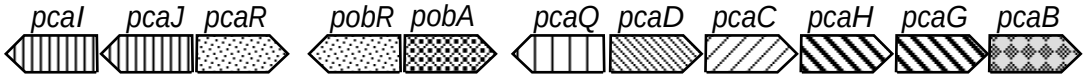
Pseudomonas sp. HR199 (g-Proteobacteria)



Rhodococcus opacus 1CP (Gram Positive)

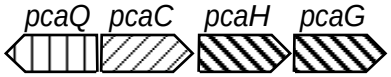


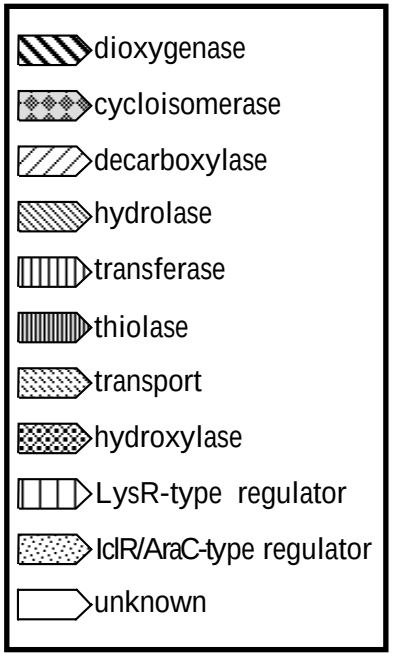
Agrobacterium tumefaciens A348 (a-Proteobacteria)



Sagittula stellata E-37 (a-Proteobacteria)

Y3F (a-Proteobacteria)





CHAPTER 3

DIVERSITY OF A RING-CLEAVING DIOXYGENASE GENE *pcaH* IN A SALT MARSH BACTERIAL COMMUNITY[2]

[2]Buchan, A., E. L. Neidle, and M. A. Moran. 2001. *Applied and Environmental Microbiology*. 67:5801-5809.

Reprinted by permission of publisher.

Abstract

The degradation of lignin-related aromatic compounds is an important ecological process in the highly productive salt marshes of the southeastern U. S., yet little is known of the mediating organisms or their catabolic pathways. Here we report the diversity of a gene encoding a key ring-cleaving enzyme of the β-ketoadipate pathway, *pcaH*, amplified from bacterial communities associated with decaying *Spartina alterniflora*, the salt marsh grass dominating these coastal systems, as well as from enrichment cultures with aromatic substrates (*p*-hydroxybenzoate, anthranilate, vanillate and dehydroabietate). Sequence analysis of 149 *pcaH* clones revealed 85 unique sequences. Thirteen amino acid residues out of 53 compared were invariant in the PcaH proteins, suggesting a required catalytic or structural function for these residues. Fifty-eight percent of the clones matched sequences amplified from a collection of 36 bacterial isolates obtained from seawater, marine sediments, or senescent *Spartina*. Fifty-two percent of the *pcaH* clones could be assigned to the roseobacter group, a marine lineage of α-subclass of the Proteobacteria abundant in coastal ecosystems. Another 6% matched genes retrieved from isolates in the genera *Acinetobacter*, *Bacillus,* and *Stappia*, and 42% could not be assigned to a cultured bacterium based on sequence identity. These results suggest a high diversity of genes encoding a single step in aromatic compound degradation in this coastal marsh.

Introduction

In southeastern U.S. saltmarshes, lignin-related aromatic compounds comprise a significant fraction of the total organic carbon pool. These compounds arise primarily from *Spartina alterniflora*, a grass responsible for >80% of the total primary production (34), and from other vascular plants that decompose within the marsh sediments. While it is widely recognized that bacteria play a major role in the transformation of vascular plant material (25-27), the bacteria responsible and the enzymatic pathways involved have yet to be properly characterized.

In terrestrial soils a major catabolic route for lignin-related aromatic compounds is the β-ketoadipate pathway (32). This primarily chromosomally encoded convergent pathway plays an integral role in the catabolism of a vast array of phenolic compounds and is widespread in phylogenetically diverse soil bacteria and fungi (18). In this pathway, polycyclic and homocyclic aromatic compounds are transformed into one of two dihydroxylated intermediates, catechol or protocatechuate. These phenolic compounds are then cleaved between their two hydroxyl groups (*ortho*-cleavage) by either catechol 1,2 dioxygenase (1,2-CTD) or protocatechuate 3,4 dioxygenase (3,4-PCD). Following ring cleavage the products are converted to β-ketoadipate, the intermediate for which the pathway is named. Two additional steps complete the conversion of β-ketoadipate to tricarboxylic acid cycle intermediates (Fig. 3.1). While this pathway has been identified in a number of bacterial genera, including *Acinetobacter*, *Alicaligenes*, *Azotobacter*, *Bacillus*, *Pseudomonas*, *Rhodococcus* and *Streptomyces* (7, 18), it is not known whether or not it is prevalent in marine communities.

The β-ketoadipate pathway is biochemically conserved and the structural genes encoding enzymes in this pathway are similar among the phylogenetically diverse organisms that possess it (18). Both 3,4-PCD and 1,2-CTD belong to a large class of non-heme iron-containing dioxygenases. 3,4-PCD is composed of equimolar concentrations of two non-identical subunits, termed α and β, encoded by the usually cotranscribed *pcaG* and *pcaH* genes, respectively. The β-subunit contains all of the ligands required for formation of the catalytic site, which may explain the greater similarity of PcaH sequences relative to PcaG sequences among various organisms (30). This conservation within PcaH facilitates the use of molecular tools to detect the corresponding gene in isolates and environmental samples.

Although the β-ketoadipate pathway is an important catabolic pathway in soil bacteria, alternative routes of aromatic compound degradation, including *meta*- and *para*-cleavage pathways, have been identified (18). However, since studies of these pathways have also focused primarily on soil organisms, their relevance in marine systems remains relatively unexplored. In this report, we investigate the potential ecological role of the β-ketoadipate pathway in coastal marine environments by assessing the presence and diversity of *pcaH* gene pools in natural bacterial communities associated with decaying *Spartina*. We also identify *pcaH* gene fragments in marine isolates cultured from seawater, marine sediments and decomposing *Spartina* for comparative studies with genes from uncultivated organisms. Our results suggest that the β-ketoadipate pathway is widespread in southeastern U.S. coastal bacteria and that members of the roseobacter

lineage, an ecologically important marine clade, may be the dominant aromatic compound-degrading bacteria in these systems.

Materials and Methods

Natural communty DNA. *Spartina* detritus was collected from a marsh at the Skidaway Institute of Oceanography (Savannah, GA) in April 2000. *Spartina* leaves were vigorously agitated in 0.2 µm filter-sterilized seawater to dislodge bacteria. The rinse water was passed through a series of Nitrex filters (140 µm, 70 µm and finally 30 µm) to remove larger plant pieces and sediment. The bacterial community was captured by passing 100 ml of the screened rinse water through a 0.2 µm filter and DNA was extracted from the filter using the Soil DNA Extraction Kit Mega Size (MoBio, Solana Beach, CA). The remaining rinse water was used as the inoculum for the enrichments as outlined below.

Amplification of *pcaH* from the natural community. A degenerate PCR primer set based on conserved regions in PcaH (P340IDf [5'YTI GTI GAR RTI TGG CAR CGI AAY GC 3'] and P340IDr [5' ICY IAI RTG IAY RTG IGC IGG ICK CCA 3']; Y = C or T, R = A or G, K = T or G) was used to amplify a 212 bp fragment of *pcaH* (3). The PCR mixture contained 1X buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCL [pH 8.3]), 2 mM deoxynucleotide triphosphates, 1.0 µM of each primer, 50 ng of DNA, and 1 U of *Taq* polymerase. PCR was performed in a DNA Engine (MJ Research, Incline Village, NV) with an initial cycle of 3 min at 95°C followed by 30 cycles of 45 sec at 95°C, 45 sec at 60°C, and 45 sec at 72°C. Products of the appropriate size were

recovered from the gel using a QiaSpin Gel Extraction Kit (Qiagen, Valencia, CA) and PCR products were cloned using a TA cloning kit (Invitrogen Corp., Carlsbad, CA).

Enrichment design. Enrichment cultures consisting of 10 L of filter-sterilized seawater (salinity, 27 practical salinity units [psu]) amended with a single substrate were established in 20 L polycarbonate carboys. The natural community described above was used as the inoculum for the enrichments at a 1:40 dilution. Substrates (acetate, *p*-hydroxybenzoate, anthranilate, vanillate or dehydroabietate) were added on Day 0 (final concentration 10 µM final) and again on Days 2, 5, 8, and 11. A treatment that received no substrate was also included. The enrichments were run in duplicate and carried out at room temp in the dark, with manual shaking of the carboys every other day. On Day 14, bacterial cells were collected on 293-mm-diameter, 0.2 µm pore-size polycarbonate filters. The filters were cut in half; one half was processed immediately and the other half was stored at -70°C. DNA was extracted from the filter halves using the Soil DNA Extraction Kit Mega Size (MoBio). Bacterial abundance in the enrichments was determined by Acridine Orange direct counts (19) on Day 0 and Day 14. Dissolved organic carbon concentration was measured on Day 0 using a TOC-5000 (Shimadzu Corp., Norcross, GA).

Amplification of *pcaH* from enrichment communities. *pcaH* clone libraries were established for the enrichment communities following the same protocol as for the natural community. Clone sequences were named based on the substrate used in the enrichment (Table 3.1) and a sequential numerical designation.

A non-degenerate version of the P340 primer set based on the *pcaH* sequence previously obtained from isolate Y3F (3) was also designed. The primers Y3Ffor ($^{5'}$ CTG GTG GAG ATC TGGCAG GCC AAT GC $^{3'}$) and Y3Frev ($^{5'}$ CGA AAC GTG GATATG CGC GGG CCG CCA $^{3'}$) were used to amplify a product from one replicate of the *p*-hydroxybenzoate enrichments and the natural community. The PCR and cloning procedures were as outlined above. Clones obtained using this PCR primer set are indicated by the letter "Y" preceding the clone number designation.

T-RFLP analysis. 16S rRNA genes were amplified from enrichment DNA using the general bacterial primers 8F and 1522R (12). The 8F primer was fluorescently-labeled on the 5' end with either FAM or TET. PCR was carried out with Ready-To-Go PCR Beads (Amersham Pharmacia, Piscataway, NJ) with 0.2 µM of each primer and 50 ng of DNA. An initial 3 min at 95°C was followed by 25 cycles of 1 min at 95°C, 1 min at 60°C, and 1.5 min at 72°C. Products of the correct size (ca. 1500 bp) were recovered from a 1.0% agarose gel with the QiaSpin Gel Extraction Kit (Qiagen), followed by an additional purification step using the PCR Purification Kit (MoBio). Restriction digests were carried out in a 10 µl total volume containing 100 ng purified PCR product and 10 U of either *Cfo*I or *Rsa*I (Roche, Indianapolis, IN). Digests were carried out at 37°C for 3 h, after which samples were precipitated in ethanol and suspended in 12 µl of deionized formamide with 1 µl of fluorescently-labeled DNA fragment length standard Genescan-2500 (TAMRA; Applied Biosystems). The terminal restriction fragments lengths were determined on an ABI PRISM 310 (Applied Biosystems) in GeneScan mode. Typically, DNA extracted from replicate enrichments was analyzed simultaneously by using the

FAM label for one replicate and the TET label for the other and coinjecting the samples. Similarilites among the enrichment assemblage T-RFLP profiles were determined by cluster analysis using the KyPlot v 2.0 (http://ftp.vector.co.jp/pack/Win95/business /calc/graph).

Bacterial isolation and 16S rDNA analysis. Most isolates examined in this study were cultured from seawater, sediments, or decaying salt marsh grass collected in the estuaries and coastal waters of the southeastern U. S. Several of the strains were previously described, having been isolated from lignin or aromatic monomer enrichment cultures (isolates Y3F, Y4I, IC4, *Sagittula stellata* E-37, and *Sulfitobacter* sp. strain EE-36) (3, 17). Some strains were cultured directly from coastal seawater using nonselective, low-nutrient seawater plates (all isolates with GAI prefix) (15, 17). Some were derived from a marine dimethylsulfoniopropionate (DMSP) enrichment (isolate DSS-3) (16). Additional strains were isolated for this study from *Spartina* detritus collected at the Skidaway Institute of Oceanography during October 1999 (all isolates with an SE prefix). SE Isolates (176 in total) were obtained by grinding *Spartina* leaves in a blender with filter-sterilized seawater and spreading the liquid onto low-nutrient seawater plates containing, per liter, 10 mg of peptone (Difco Laboratories, Detriot, Mich.), 5 mg of yeast extract (Difco) and 1.5% purified agar (Difco) in filter-sterilized diluted Sargasso Sea water that had been aged for more than one year in the dark (final salinity, 24 psu) (15). Finally, two isolates that were not obtained from the southeastern U.S. coast were examined: *Sulfitobacter pontiacus* ChLG 10, which was cultured from the Black Sea (41), and strain ISM, which was cultured from the Caribbean Sea (10).

16S rRNA gene sequences were reported previously (accession numbers follow parenthetically) for the following isolates: DSS-3 (AF098491), EE-36 (AF007254), GAI-05 (AF007256), GAI-37 (AF007260), GAI-111 (AF098494), IC4 (AF254098), ISM (AF098495), Y3F (AF253467). If not already available, 16S rRNA gene sequences for the isolates were obtained by PCR amplification using the general bacterial primers 27F and 1522R (12). Genomic DNA was prepared from each isolate by a colony boil method as previously described (3). Approximately 500 bp of the PCR product were directly sequenced using the 27F primer on an ABI PRISM 310 (Applied Biosystems) following purification with an Ultra Clean PCR Clean-up kit (MoBio). Sequences were analyzed using the Genetics Computer Group program package 10.0 (Wisconsin Package Version, Madison, WI.).

pcaH genes were amplified from isolates using the degenerate primer set as outlined above, except that 3 µl of cell lysate was used in the PCR. Both strands of *pcaH* gene fragments were sequenced.

Sequence and phylogenetic analysis. Sequence analysis was performed on an ABI PRISM 310 genetic analyzer using the BigDye terminator cycle-sequencing kit (Applied Biosystems). DNA sequences were determined with M13 primers that recognized the cloning vectors. Phylogenetic trees were constructed with the PHYLIP package by using evolutionary distances (Jukes-Cantor) and the neighbor-joining method. Sequences reported here were submitted to the GenBank database under the following accession numbers: AF388307, AF388308, AY038900 to AY038926, and AY040248 to AY040273.

Results

***pcaH* diversity in the salt marsh community.** A *pcaH* clone library was established for the natural bacterial community associated with decaying *Spartina* (hereafter referred to as the salt marsh community) by amplification of DNA with the degenerate primer set. Twenty-one clones were sequenced, yielding 14 unique sequences (Fig. 3.2). Homology searches with sequences from GenBank confirmed that deduced amino acid sequences from the PCR products of the clones had highest similarity to the approximately 240 residue PcaH from various bacterial genera. The deduced amino acid similarities ranged between 82 and 100% and identities between 73 and 100%. Furthermore, two residues demonstrated to be involved in Fe^{2+} binding, Tyr408 and Tyr447, and a residue involved in substrate specificity, Trp449, (30, 44) were conserved in all sequences.

Aromatic substrate enrichments. Enrichments were established to monitor the response of the bacterial community from decaying *Spartina* to specific aromatic substrates representative of compounds associated with vascular plant decay. Anthranilate, *p*-hydroxybenzoate, and vanillate are aromatic monomers that have been demonstrated to be degraded through the β-ketoadipate pathway in soil microorganisms (18). Studies of soil microbes indicate *p*-hydroxybenzoate and vanillate are converted to protocatechuate, whereas anthranilate is typically converted to catechol prior to intradiol ring cleavage (18). Dehydroabietate is a plant diterpenoid, commonly associated with pulp and paper mill effluent, for which an extradiol cleavage pathway has only recently been elucidated (23). Enrichments with acetate, a non-aromatic compound, and no carbon controls were established for comparison with the aromatic compound enrichments.

The natural dissolved organic carbon concentration in the filter-sterilized coastal seawater was 365 µM C and the four 10 µM substrate additions during the course of the enrichments increased this by <11%. Direct counts conducted on Day 0 and Day 14 showed an increase in bacterial cell numbers in all enrichments, with an average increase of 2.8-fold during the 2 week enrichment period for the enrichments with substrate additions and 1.4-fold for the no-carbon controls (Table 3.1). The greater increase in cell numbers in the presence of added substrates suggests that bacteria capable of metabolizing those compounds became established in the enrichment treatments.

16S rDNA T-RFLP analysis of enrichment communities. The enriched bacterial communities were characterized using the 16S rDNA T-RFLP procedure (21). Independent PCR amplification and GeneScan analysis of each sample on at least two occasions confirmed the reproducibility of T-RFLP profiles. Replicate enrichments with the same substrate typically developed very similar bacterial communities. A cluster analysis using the relative peak area of each of the major peaks in the T-RFLP chromatograms digested with *Cfo*I (31 peaks) and *Rsa*I (32 peaks) confirmed that replicates supplemented with the same aromatic compound had the greatest similarity in amplifable 16S rRNA genes (Fig. 3.3). The vanillate and *p*-hydroxybenzoate enrichment culture communities formed a subgroup in this analysis, perhaps due to the structural similarity of these two compounds. The two treatments that were not supplemented with an aromatic compound (acetate and no-carbon addition) also formed a distinct cluster.

***pcaH* in enrichment communities.** To characterize the ring cleavage genes harbored by the enriched bacterial communities, *pcaH* clone libraries were established for 11 of the 12 enrichments by amplification with the degenerate primer set. The remaining sample (no carbon replicate NocB) did not yield a PCR product when amplified with the degenerate primer set (although it did produce a product when amplified with 16S rDNA primers). Repeated attempts to obtain a PCR product from this replicate (including carrying out a second DNA extraction on the unused filter half) were not successful, and thus this sample was not characterized further. From all other samples, a total of 120 *pcaH* clones were sequenced with at least 10 clones sequenced from each library (Table 3.1).

Seventy-six unique sequences were identified, five of which matched sequences retrieved from the natural community (Fig. 3.2). The *pcaH* sequences did not exhibit segregation according to enrichment substrate. For example, the 20 *pcaH* sequences retrieved from the replicate vanillate enrichments were distributed throughout the *pcaH* tree and 19 clustered with sequences from other enrichment types. Similarly, 14 of the 20 sequences obtained from the anthranilate enrichments clustered with those from other enrichments (Fig. 3.2).

Phylogeny of *Spartina*-associated isolates. For comparative purposes, a collection of culturable marine bacteria harboring the *pcaH* gene was assembled. An initial screen of the 176 SE isolates obtained from decaying *Spartina* was carried out with the degenerate PCR primers targeting *pcaH*. For all 28 isolates (16%) that gave a PCR product of the correct size, phylogenetic analysis of 16S rRNA gene sequences was carried out. Twenty-three of these isolates showed close phylogenetic affinities to α-Proteobacteria

previously isolated from marine environments. The majority of these (18) fell into the *Rhizobium-Agrobacterium* group, showing ≥ 96.7% similarity to a symbiont isolated from the Eastern oyster, *Crassostrea virginica* (CV 902-700) (1). The closest described relative of these CV 902-700-like isolates is *Stappia stellulata* (originally described as an *Agrobacterium*), an organism isolated from marine sediments and seawater (38). The remaining α-Proteobacterial isolates were affiliated with the rhodobacter or roseobacter groups. Two of the isolates showed affiliations with γ-Proteobacteria and three were closely related to *Bacillus* spp. (Table 3.2).

Many *pcaH*-containing SE isolates were related to organisms from which 3,4-PCD activity has been previously reported. These include the roseobacter group (3), *Agrobacterium* species (5, 31, 33), *Acinetobacter* species (11, 32, 43), and *Bacillus* species (24). 3,4-PCD activity has not been reported in the rhodobacter or Halomonadaceae groups (in which isolates SE37 and SE96 cluster), although both of these group contain members capable of metabolizing aromatic compounds (9, 36).

***pcaH* in marine isolates.** In addition to the 28 SE isolates described above, strains previously isolated from seawater or sediments and belonging to the roseobacter clade were also screened for *pcaH*. Nine roseobacter strains yielded a PCR product of the correct size using the *pcaH* primers (ISM, Y4I, DSS-3, GAI-05, GAI-21, GAI-109, GAI-111, GAI-37, and *Sulfitobacter pontiacus*). We had previously identified this gene in four additional roseobacter isolates (*Sagitulla stellata* E-37, EE-36, Y3F and IC4) (3), and include these organisms in all analyses presented here. The PCR products from all SE isolates and roseobacter isolates were sequenced, except for the SE isolates showing very

high similarity by 16S rRNA gene analysis to the *C. virginica* symbiont CV902-700. Due to the strain level identity of these 19 isolates, only 8 were selected for *pcaH* sequence analysis. In all, 26 *pcaH* sequences from marine isolates were obtained.

Similarity in *pcaH* sequences was typically found for closely related isolates. *S. pontiacus*, EE-36, GAI-37, and GAI-21 form a cluster within the roseobacter lineage based on 16S rRNA gene analysis, and *pcaH* genes from these isolates also cluster with a high bootstrap value (Fig. 3.2), sharing ≥ 81.8% sequence similarity at the nucleotide level. Isolates Y3F and Y4I have a 16S rRNA gene sequence similarity of 100% and *pcaH* sequence similarity of 97.5%. Isolate SE197 and *Acinetobacter calcoaceticus* share 99.7% 16S rRNA gene sequence similarity, and their *pcaH* sequences form a distinct cluster that is supported by a high bootstrap value (Fig. 3.2). All of the *C. virginica* CV902-700-like isolates had *pcaH* sequences that were ≥ 97.5% similar and deduced amino acid sequence that were ≥ 94.3% identical. Finally, two pairs of isolates, GAI-109 and GAI-111 and SE45 and SE95, had identical sequences for both the 16S rRNA gene and *pcaH*.

The 16S rRNA gene and *pcaH* phylogenies were not always congruent, however. Comparisons of isolate SE37 to the CV902-700-like strains show only ca. 84% sequence similarity based on 16S rDNA analysis, but as little as one base pair difference when comparing *pcaH* sequences. Furthermore, *pcaH* gene sequences available for two agrobacterial strains related to the CV902-700-like isolates do not appear to cluster with the *pcaH* genes from these isolates in our analysis (Fig. 3.2). Finally, the obvious lack of similarity among *pcaH* sequences retrieved from the *Bacillus* isolates SE98, SE105, and SE165 suggests this gene may be highly divergent within these organisms, although no

other *Bacillus pcaH* genes are available for comparison (i.e. this is the first report of a *pcaH* sequence for a *Bacillus* isolate).

Comparisons of PcaH sequences from all isolates examined in this study and those previously deposited in GenBank show sequence similarity of ≥ 52.2% at the nucleotide level and ≥ 52.8% and ≥ 43.4% similarity and identity at the deduced amino acid level, respectively. Furthermore, the conservation of 13 residues in all clone and isolate sequences suggests a required catalytic or structural function.

Comparison of *pcaH* genes from clones and isolates. Of the 21 *pcaH* clones obtained from the salt marsh community using the degenerate primer set, 10 (44%) were considered matches (i.e. ≤ 1 bp difference) with genes from roseobacter group isolates. One additional *pcaH* clone, SMC5, had a nucleotide similarity of > 98% to roseobacter group isolate *S. pontiacus*. Finally, clones SMC1 and SMC7 shared > 94% sequence similarity with *Bacillus* isolate SE165, bringing the total number of clones that clustered with *pcaH* sequences from isolates to 13 (56%).

Of the 120 clones obtained from the enrichments, 67 (54%) were considered matches (i.e. ≤ 1 bp difference) with genes from roseobacter group isolates. A number of the remaining clones differed from isolate *pcaH* sequences at more than one position, but nonetheless showed notable sequence similarity with isolates (Fig. 3.2). Three clones (VanlA12, AcetB8, and VanlA15) grouped with the *S. stellata* E-37 *pcaH* sequence and shared ≥ 96.9% sequence similarity. Four clones clustered with *S. pontiacus* and demonstrated a within-group nucleotide sequence similarity of ≥ 96.9%, bringing the total number of clones that grouped with roseobacter isolates to 74 (60%). Two other clones

had sequences which were identical to those of two *Bacillus* isolates, SE165 and SE98, five clones had identical sequences to the *Acinetobacter* isolate SE197, and three clones had sequences identical to CV 902-700-like isolates and the rhodobacter isolate SE37. Clone DhaA9 was 93.1% identical at nucleotide level and 98.1% identical at the amino acid level to the *pcaH* fragment of γ-Proteobacteria isolate SE96.

Only a few *pcaH* clone sequences grouped with sequences from isolates not identified in this study. NocA2 and DhaB19 shared > 85% sequence identity and clustered with the *Pseudomonas putida* and *P. aeruginosa pcaH* sequences in GenBank. AcetA3 was 84% identical to the *pcaH* sequence from the β-Proteobacterium *Burkholderia cepacia*.

Clustering of clone *pcaH* sequences from the same enrichment type was generally not found. One of the few exceptions was that clones resembling *Acinetobacter* sp. were recovered only the from the acetate- and anthranilate-amended enrichments. In addition, 7 clones from the dehydroabietate and *p*-hydroxybenzoate enrichments formed a cluster with *pcaH* genes from isolates with varying phylogenetic affinities (β- and γ-Proteobacteria and *Streptomyces* sp.). Finally, 7 of the 10 clones obtained from the NocA library showed sequence similarity to *pcaH* from *S. stellata* E-37. This relatively low clonal diversity may suggest the *pcaH*-containing community in this treatment was composed of only a few organisms. Indeed, low abundance of *pcaH* genes in the absence of aromatic substrates may also explain our inability to obtain a PCR product from the second no-substrate replicate (NocB).

Non-degenerate *pcaH* primers. Due to the phylogenetic differences among the organisms from which *pcaH* had previously been sequenced, the design of universal *pcaH* primers required a high degree of DNA sequence degeneracy. In an attempt to investigate potential bias of the degenerate primers, a nondegenerate primer set was designed based on the *pcaH* sequence from roseobacter isolate Y3F, an isolate for which no similar sequences were found among the 141 *pcaH* clones obtained with the degenerate primer set. This second primer set was used to amplify *pcaH* gene fragments from both the salt marsh community and one replicate of the *p*-hydroxybenzoate enrichments (PhbA). Four representatives of the cloned PCR products were sequenced from each sample. One of the clones analyzed, SMCY6, had a *pcaH* sequence identical to that for isolate Y3F. In addition, SMCY1 was 96.2% similar at the nucleotide level and identical at the deduced amino acid level to the *pcaH* fragment of SE62, another roseobacter group isolate. Finally, clone PobY3 shared 87.4% nucleotide similarity and 98.1% amino acid identity to roseobacter group isolates SE45 and SE95. The remaining five clones had no identifiable sequence similarity with either an isolate or clone.

In total, 86 (58%) of the 149 *pcaH* clones obtained from the salt marsh community and enrichments were considered a match (≤1 bp mismatch) to the gene sequence from an isolate examined in this study, with 78 (52%) matching one of five roseobacter group isolates. Sixty-three (42%) of the cloned *pcaH* sequences did not closely match that of an isolated bacterium. In almost all cases, the branch topology of trees based on nucleotide sequences were maintained when deduced amino acids were analyzed (data not shown).

Discussion

The ecological significance of the β-ketoadipate pathway for the degradation of naturally occurring aromatic compounds has been inferred from studies of a select group of soil microorganisms. While these studies have been instrumental in characterizing structural and sequence similarities, as well as the regulation and function of the pathway in certain bacteria, they have yet to establish this catabolic route as widely distributed in many natural systems. Through the development of a degenerate primer set targeting all known *pcaH* sequences, we can now begin to investigate the importance and diversity of this key aromatic ring cleavage gene in a variety of natural bacterial communities.

***pcaH* diversity in salt marsh and enrichment communities.** The *pcaH* gene diversity in the bacterial communities associated with decaying *Spartina* was high. Of the 21 clones derived using the degenerate primer set, 14 unique sequences were identified. Enrichment cultures were established to assess the diversity of *pcaH* genes harbored in marine bacterial assemblages by varying the amount and type of aromatic substrates available, and T-RFLP analysis of 16S rRNA genes indicates that distinct bacterial communities indeed developed in each treatment (Fig. 3.3). Analysis of these communities revealed an additional 76 gene sequences out of 120 partial *pcaH* sequences. Five of these matched sequences found in direct amplifications from the salt marsh community, but 71 were novel sequences. The Y3F-specific primers yielded even more novel *pcaH* sequences from the *Spartina*-associated bacterial community; 8 new sequences were obtained from 8 clones, only one of which was identical to the sequence from the isolate for which this primer set was specifically designed. The possibility that

some fraction of the *pcaH* diversity found in this study results from chimeric artifacts generated during PCR amplification does not change the overall conclusion that significant diversity exists for this key gene in aromatic compound degradation. Further, the small size of the amplified product reduces the likelihood of heteroduplex formation (45).

High levels of functional gene diversity in environmental samples are not unprecedented, and have been noted previously for genes involved in denitrification (2, 39), bisulfite reduction (6), and nitrogen fixation (22, 29). However, it is not typical that functional genes retrieved directly from environmental samples have such high sequence similarity to those from cultured bacteria. For example, Scala and Kerkof (39) identified 37 unique *nosZ* genes from marine sediments, none of which resembled those of cultivated organisms. Similarly, Lovell et al. (22) report 43 unique *nifH* sequences out of 59 clones analyzed, none of which matched those of known nitrogen fixers. Yet for *pcaH* genes retrieved from decaying *Spartina*, 58% of the clones matched (i.e. $\leq$ 1 bp difference) those sequences found in a companion collection of marine isolates. Nearly half of the 25 genes amplified from the salt marsh community (44%) could be matched to one of five roseobacter isolates cultured from decaying *Spartina* detritus or seawater. Similarly, over half the 124 clones from the enrichment communities (54%) could be matched to a roseobacter isolate (Table 3.3). A more conservative definition requiring no mismatches between sequences still results in 32% and 28% of the salt marsh and enrichment communities *pcaH* sequences, respectively, matching those of cultured roseobacter species. It is unlikely that this predominance of roseobacter-like *pcaH* sequences is due to a particular bias in the degenerate primers, since the primers were designed to target

PcaH in 14 phylogenetically diverse organisms representing several bacterial lineages (e.g. α-, β-, γ-Proteobacteria, gram positive). Moreover, the *pcaH* gene was previously found to be quite common among culturable members of the roseobacter clade (3).

Members of the roseobacter clade are abundant in many coastal and open-ocean environments (16, 28, 42) and have been found to contribute up to 30% of the bacterioplankton 16S rRNA genes in southeastern U. S. coastal systems (15). Unlike other dominant marine bacterial lineages that have no close relatives in culture, roseobacter group members are readily isolated from coastal and open-ocean systems (10, 13, 15, 20). Roseobacter group members have also been shown to be primary colonists on surfaces in coastal salt marshes (8). Both surface colonization and plant degradation typically involve the production of exopolysaccharide, holdfast structures or fibrils which can assist in cellular attachment (35). The largest number of *pcaH* clones clustered with the gene from roseobacter group isolate *S. stellata* E-37, a bacterium able to attach selectively to the surface of lignocellulose particles and mineralize cellulose and synthetic lignin (14).

The aromatic compounds used in the enrichment experiments represent fused-ring and hydroxy, methyl and amino-substituted structures. The presence of roseobacter–like *pcaH* genes in all enrichments with aromatic substrates suggests these bacteria are capable of converting a variety of ring structures and therefore may contain multiple sets of catabolic genes. Anthranilate, *p*-hydroxybenzoate, and vanillate have each been shown to require a unique set of upper pathway genes for conversion into a dihydroxylated intermediate such as protocatechuate or catechol (4, 18, 37, 40).

While T-RFLP profiles of 16S rDNA amplicon pools from enrichment community DNA indicated that distinctive bacterial communities developed in response to each enrichment substrate (Fig. 3.3), there was surprisingly little evidence that *pcaH* gene sequences likewise segregated according to enrichment type (Fig. 3.2). This absence of *pcaH* clustering by enrichment type suggests that the marine bacteria responsible for aromatic ring cleavage are nutritional generalists, able to funnel a variety of different aromatic structures through the protocatechuate branch of the β-ketoadipate pathway. Alternative explanations for the lack of apparent *pcaH* clustering by enrichment substrate are that the diversity of *pcaH* clones is high relative to the sample size of the clone libraries (i.e. clustering may have been evident if more clones were sequenced per enrichment treatment), or that the distinct T-RFLP patterns in enrichment treatments reflect compositional differences of that component of the bacterial community that is not involved in aromatic ring cleavage.

Ecological significance. Despite its significance in the processing and degradation of aromatic compounds in a variety of systems, the ecological role of the β-ketoadipate pathway has not yet been demonstrated outside of soil ecosystems. Here we report the importance and diversity of a gene encoding a key enzyme of the pathway, 3,4-PCD, in both natural and enriched bacterial communities from a southeastern salt marsh. If we presume that successful amplification of a portion of *pcaH* is indicative of the presence of a functional 3,4-PCD enzyme (i.e. *pcaH* and *pcaG*), these results suggest the involvement of taxonomically diverse marine bacteria, some of which have yet to be identified, in the processing of aromatic compounds via a mechanism well described in soil bacteria. In the

environment from which these genes were amplified, lignin and lignin degradation products are the most likely sources of naturally occurring aromatic substrates.

The degenerate primer set used here was based on previously retrieved *pcaH* genes, and therefore may target only a subset of the ring cleavage dioxygenases present in this system. Furthermore, at least 6 different ring cleavage dioxygenases in addition to 3,4-PCD have been identified in soil bacteria, and these or other novel dioxygenases may also be present in coastal marine marshes. Nonetheless, the *pcaH* gene is present in the decomposer community of this coastal marsh and a minimum of 85 different versions are present, as indicated by sequence differences in the 159 bp fragment amplified from *pcaH*.

The radiation of similar *pcaH* sequences raises questions about phenotypic microheterogeneity in the salt marsh bacterial community, and potentially has interesting implications for population dynamics and ecological function. The sequence microheterogeneity observed in this *pcaH* fragment may reflect genetic divergence within this phylogenetically broad clade that has little ecological significance. Alternatively, it may serve as the basis for slight differences in enzyme activity or stability under varying environmental conditions. Because members of the roseobacter lineage are amenable to culturing, it will be possible to pursue laboratory-based physiological and genetic studies on aromatic compound degradation by members of this bacterial clade possessing distinctive *pcaGH* sequences. Access to the physiology of ecologically relevant bacteria via culturing is uncommon in microbial ecology and may lead to unique insights into the role of functional gene microheterogeneity in natural environments.

Acknowledgments

We are grateful to Dr. José González for advice on experimental design. This work was supported by the NSF through grants from the Microbial Observatory Program (MCB-0084164 to M.A.M) and the Molecular and Cellular Biosciences Program (MCB-9808784 to E.L.N.) and a traineeship (to A.B.) provided through a research training grant in prokaryotic diversity (BIR-9413235).

References

1. Boettcher, K. J., B. J. Barber, and J. T. Singer. 2000. Additional evidence that juvenile oyster disease is caused by a member of the *Roseobacter* group and colonization of nonaffected animals by *Stappia stellulata*-like strains. Appl. Environ. Microbiol. 66:3924-3930.

2. Braker, G., J. Zhou, L. Wu, A. H. Devol, and J. M. Tiedje. 2000. Nitrite reductase genes (*nirK* and *nirS*) as functional markers to investigate diversity of denitrifying bacteria in Pacific northwest marine sediment communities. Appl. Environ. Microbiol. 66:2096-2104.

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

4. Bundy, B. M., A. L. Campbell, and E. L. Neidle. 1998. Similarities between the *antABC*-encoded anthranilate dioxygenase and the *benABC*-encoded benzoate dioxygenase of *Acinetobacter* sp. strain ADP1. J. Bacteriol. 180:4466-4474.

5. Contzen, M., and A. Stolz. 2000. Characterization of the genes for two protocatechuate 3,4-dioxygenases from the 4-sulfocatechol-degrading bacterium *Agrobacterium radiobacter* strain S2. J. Bacteriol. 182:6123-6129.

6. Cottrell, M., and S. C. Cary. 1999. Diversity of dissimilatory bisulfite reductase genes of bacteria associated with the deep-sea hydrothermal vent polychaete annelid *Alvinella pompejana*. Appl. Environ. Microbiol. 65:1127-1132.

7. Dagley, S. 1986. Biochemistry of aromatic hydrocarbon degradation in Pseudomonads. P 527-555. In J.R. Sokatch (Ed.), The Bacteria, Vol 10. Academic Press Inc., New York

8. Dang, H., and C. R. Lovell. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis. Appl. Environ. Microbiol. 66:467-475.

9. Esham, E. C., W. Y. Ye, and M. A. Moran. 2000. Identification and characterization of humic substances-degrading bacterial isolates from an estuarine environment. FEMS Microbiol. Ecol. 34:103-111.

10. Fuhrman, J. A., S. H. Lee, Y. Masuchi, A. A. Davis, and R. M. Wilcox. 1994. Characterization of marine prokaryotic communities via DNA and RNA. Microb. Ecol. 28:133-145.

11. Gaines, G. L., L. Smith, and E. L. Neidle. 1996. Novel nuclear magnetic resonance spectroscopy methods demonstrate preferential carbon source utilization by *Acinetobacter calcoaceticus.* J. Bacteriol. 178:6833-6841.

12. Giovannoni, S., and M. Rappé. 2000. Evolution, diversity, and molecular ecology of marine prokaryotes, p47-84. *In* D. Kirchman (ed.), Microbial ecology of the oceans. Wiley Interscience, New York, N. Y.

13. González, J. M., R. P. Kiene, and M. A. Moran. 1999. Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α-subclass of the class proteobacteria. Appl. Environ. Microbiol. 65:3810-3819.

14. González, J. M., F. Mayer, M. A. Moran, R. E. Hodson, and W. B. Whitman. 1997. *Sagittula stellata* gen. nov., sp. nov., a lignin transforming bacterium from a coastal environment. Int. J. Syst. Bacteriol. 47:773-780.

15. González, J. M., and M. A. Moran. 1997. Numerical dominance of a group of marine bacteria in the α-subclass of the class Proteobacteria in coastal seawater. Appl. Environ. Microbiol. 63:4237-4242.

16. González, J. M., R. Simó, R. Massana, J. S. Covert, E. O. Casamayor, C. Pedrós-Alió, and M. A. Moran. 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing north Atlantic algal bloom. Appl. Environ. Microbiol. 66:4237-4246.

17. González, J. M., W. B. Whitman, R. E. Hodson, and M. A. Moran. 1996. Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. Appl. Environ. Microbiol. 62:4433-4440.

18. Harwood, C. S., and R. E. Parales. 1996. The β-ketoadipate pathway and the biology of self-identity. Ann. Rev. Microbiol. 50:553-590.

19. Hobbie, J. E., R. J. Daley, and S. Jasper. 1977. Use of Nucleopore filters for counting bacteria by fluorescence microscopy. Appl. Environ. Microbiol. 33:1225-1228.

20. Ledyard, K. M., E. F. Delong, and J. W. H. Dacey. 1993. Characterization of a DMSP-degrading bacterial isolate from the Sargasso Sea. Arch. Microbiol. 160:312-318.

21. Liu, W.-T., T. L. Marsh, H. Cheng, and L. J. Forney. 1997. Characterization of microbial diversity by determining terminal fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. 63:4516-4522.

22. Lovell, C. R., Y. M. Piceno, J. M. Quattro, and C. E. Bagwell. 2000. Molecular analysis of diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. Appl. Environ. Microbiol. 66:3814-3822.

23. Martin, V. J. J., and W. W. Mohn. 2000. Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. J. Bacteriol. 182:3784-3793.

24. Mashetty, S. B., S. Manohar, and T. B Karegoudar. 1996. Degradation of 3-hydroxybenzoic acid by a *Bacillus* species. Indian J. Biochem. Bio. 33:145-148.

25. Moran, M. A. and R. E. Hodson. 1989. Formation and bacterial utilization of dissolved organic carbon derived from detrital lignocellulose. Limnol. Oceanogr. 34:1034-1047.

26. Moran, M. A., and R. E. Hodson. 1990. Bacterial production on humic and nonhumic components of dissolved organic carbon. Limnol. Oceanogr. 35:1744-1756.

27. Moran, M. A., and R. E. Hodson. 1994. Dissolved humic substances of vascular plant origin in a coastal marine environment. Limnol. Oceanogr. 39:762-771.

28. Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni. 1995. Genetic comparisions reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. Limnol. Oceanogr. 40:148-158.

29. Ohkuma, M., S. Noda, and T. Kudo. 1999. Phylogenetic diversity of the nitrogen fixation genes in the symbiotic community in the gut of diverse termites. Appl. Environ. Microbiol. 65:4926-4934.

30. Ohlendorf, D. H., A. M. Orville, and J. D. Lipscomb. 1994. Structure of protocatechuate 3,4-dioxygenase from *Pseudomonas aeruginosa* at 2.15 Å Resolution. J. Mol. Biol. 244:586-608.

31. Parke, D. 1997. Acquistion, reorganization, and merger of genes: novel management of the b-ketoadipate pathway in *Agrobacterium tumefaciens*. FEMS Microbiol. Lett. 146:3-12.

32. Parke, D., D. A. D'Argenio, and L. N. Ornston. 2000. Bacteria are not what they eat: That is why they are so diverse. J. Bacteriol. 182:257-263.

33. Parke, D., and L. N. Ornston. 1986. Enzymes of the β-ketoadipate pathway are inducible in *Rhizobium* and *Agrobacterium* spp. and constitutive in *Bradyrhizobium* spp. J. Bacteriol. 165:288-292.

34. Pomeroy, L. R., W. M. Darley, E. L. Dunn, J. L. Gallagher, E. B. Haines, and D. M. Whitney. 1981. Primary production, p. 39-68. *In* L. R. Pomeroy and R. G. Wiegert (ed.), The ecology of a salt marsh. Springer-Verlag, New York, N.Y.

35. Rogers, G. M. and A. A. W. Baecker. 1987. Theories on the degradation in wood associated with glycocalix-producing bacteria. J. Inst. Wood. Sci. 11: 78-84.

36. Rolden, M. D., R. Blasco, F. J. Caballero, and F. Castillo. 1998. Degradation of *p*-nitrophenol by the phototrophic bacterium *Rhodobacter capsulatus*. Arch. Microbiol. 169:36-42.

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

38. Rüger, H.-J., and M. G. Holfe. 1992. Marine star-shaped-aggregate-forming bacteria: *Agrobacterium atlantic* sp. nov.; *Agrobacterium meteor* sp. nov; *Agrobacterium ferruginous* sp. nov., nom. rev.; *Agrobacterium gelatinovorum* sp. nov., nom. rev.; and *Agrobacterium stellulatum* sp. nov., nom. rev. Int. J. Syst. Bacteriol. 42:133-143.

39. Scala, D.J., and L. J. Kerkhof. 1999. Diversity of nitrous oxide reductase (*nos*Z) genes in continental shelf sediments. Appl. Environ. Microbiol. 65:1681-1687.

40. Segura, A., P. V. Bunz, D. A. D'Argenio, and L. N. Ornston. 1999. Genetic analysis of a chromosomal region containing *vanA* and *vanB*, genes required for conversion of either ferulate or vanillate to protocatechuate in *Acinetobacter*. J. Bacteriol. 181:3494-3504.

41. Sorokin, D. Y., and A. M. Lysenko. 1993. Heterotrophic bacteria from the Black Sea oxidizing reduced sulfur compounds to sulfate. Microbiol. *(English Translation Mikrobiologiya)* 62:1018-1031.

42. Suzuki, M. T., M. S. Rappé, Z. W. Haimberger, H. Winfield, N. Adair, J. Ströbel, and S. J. Giovannoni. 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl. Environ. Microbiol. 63:983-989.

43. Vasydevan, N., and A. Mahadevan. 1992. Degradation of nonphenolic β-*ortho*-4-lignin substructure model compounds by *Acinetobacter* sp. Res. Microbiol. 143:333-339.

44. Vetting, M. W., D. A. D'Argenio, L. N. Ornston, and D. H. Ohlendorf. 2000. A structure of *Acinetobacter* sp. ADP1 protocatechuate 3,4-dioxygenase at 2.2Åresolution: implications for the mechanism of an intradiol dioxygenase. Biochem. 39:7943-7955.

45. Wang, G. C.-Y., and Y. Wang. 1996. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology 142:1107-1117.

Table 3.1. Bacterial cell growth during a two week enrichment period with aromatic substrates and *pcaH* clone recovery from the enrichments and original salt marsh community.

Sample designation	**Substrate**	**Increase in no. of cells (fold)**	**No. of *pcaH* clones sequenced**	**% Unique Sequences**
SMC[a]	NA[b]	NA	21	76%
NocA	none	1.6	10	40%[c]
NocB		1.1	NA	
AcetA	acetate	5.1	12	52%
AcetB		3.2	11	
PhbA	*p*-hydroxybenzoate	3.4	15	42%
PhbB		2.7	11	
VanlA	vanillate	2.4	10	60%
VanlB		3.2	10	
AnthA	anthranilate	1.6	10	55%
AnthB		1.7	10	
DhaA	dehydroabietate	2.8	10	57%
DhaB		1.9	11	

[a]SMC, salt marsh community.
[b]NA, not applicable.
[c] Calculated for the NocA sample only.

Table 3.2. Phylogenetic affiliations of marine isolates with amplifiable *pcaH* genes.

Isolate	Major taxa	Group	Closest relative (accession no.)	% 16S rDNA Similarity
Sagittula stellata E-37	α-Proteobacteria	Roseobacter	NA[a]	NA
Sulfitobacter pontiacus	α-Proteobacteria	Roseobacter	NA	NA
DSS-3	α-Proteobacteria	Roseobacter	*Ruegeria* sp. AS-36 (AJ391197)	97
EE-36	α-Proteobacteria	Roseobacter	*Sulfitobacter pontiacus* (Y13155)	99
GAI-05	α-Proteobacteria	Roseobacter	marine isolate JP88.1 (AY007684)	98
GAI-21	α-Proteobacteria	Roseobacter	*Sulfitobacter* sp. GAI-37 (AF007260)	98
GAI-37	α-Proteobacteria	Roseobacter	*Sufitobacter* sp. GAI-21 (AF007257)	98
GAI-111/109	α-Proteobacteria	Roseobacter	*Roseobacter* clone NAC11-6 (AF245634)	94
IC4	α-Proteobacteria	Roseobacter	hydrothermal vent strain TB66 (AF254109)	98
ISM	α-Proteobacteria	Roseobacter	*C. virginica* symbiont CV919-312 (AF114484)	96
Y3F/Y4I	α-Proteobacteria	Roseobacter	marine bacterium PP-154 (AJ296158)	97
SE03[b]	α-Proteobacteria	*Rhizobium*/*Agrobacterium*	*C. virginica* symbiont CV902-700 (AF246615)	97

SE09	α-Proteobacteria	*Rhizobium/Agrobacterium*	*C. virginica symbiont CV902-700 (AF246615)*	98
SE11	α-Proteobacteria	*Rhizobium/Agrobacterium*	*C. virginica symbiont CV902-700 (AF246615)*	97
SE65	α-Proteobacteria	*Rhizobium/Agrobacterium*	*C. virginica symbiont CV902-700 (AF246615)*	99
SE45/SE95	α-Proteobacteria	Roseobacter	*hydrothermal vent strain AG33 (AF254108)*	98
SE62	α-Proteobacteria	Roseobacter	*Isolate GAI-37 (AF007260)*	96
SE37	α-Proteobacteria	*Rhodobacter*	*marine isolate Sippewissett 2-21 (AF055822)*	99
SE197	γ-Proteobacteria	*Moraxellaceae*	*Acinetobacter calcoaceticus (AF159045)*	99
SE96	γ-Proteobacteria	*Halomonadaceae*	*Noctiluca scintillans endocyte (AF262750)*	97
SE98	*Firmicutes*	*Bacillus/Clostridium*	*Bacillus cereus (AF274244)*	99
SE105	*Firmicutes*	*Bacillus/Clostridium*	*Bacillus sp. OS-5 (BSP296095)*	99
SE165	*Firmicutes*	*Bacillus/Clostridium*	*Bacillus subtilis*	91

[a] *NA, not applicable.*

[b] The following isolates had 16S rRNA gene sequences identical to that of SE03: SE22, SE26, SE27, SE32, SE35, SE36, SE39, SE44, SE49, SE55, SE57, SE60, SE83, SE97, SE114.

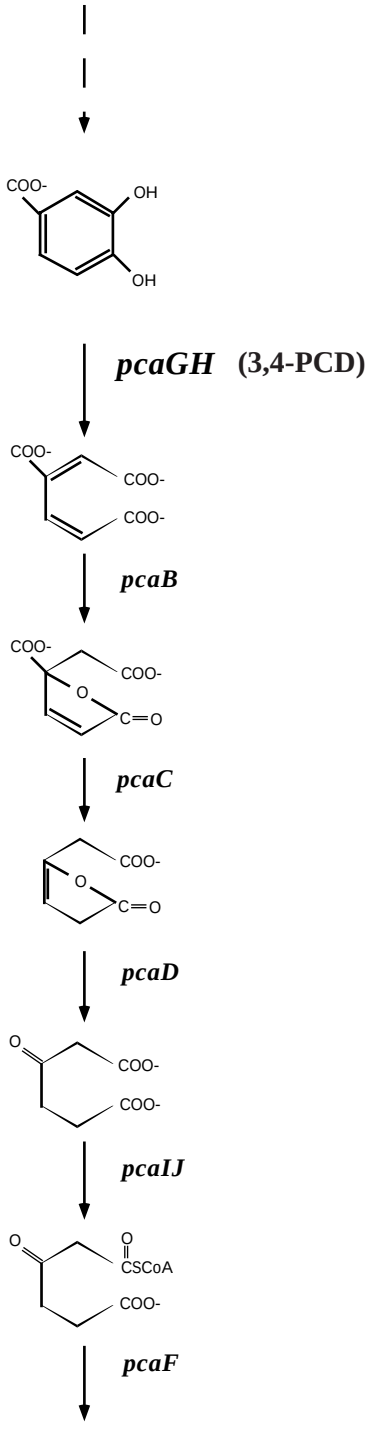
Table 3.3. Phylogenetic affiliation[a] of *pcaH* sequences from cultured and uncultured (salt marsh community clones) members of the bacterial community associated with decaying *Spartina alterniflora* and from enrichments of that community with a variety of aromatic substrates (enrichment clones)[a]

	% of sequences affiliated with:					
	Roseobacter group	*Rhodobacter*	*Acinetobacter*	*Halomonas*	*Bacillus*	Unidentified
Salt marsh community clones (N = 25)	44%	0%	0%	0%	0%	56%
Enrichment clones (N = 124)	54%	2%	4%	0%	2%	38%
SE isolate collection (N = 28)	11%	71%	3.5%	3.5%	11%	NA[b]

[a] Clone affiliations were inferred based on † 1 bp difference to *pcaH* sequences from isolates.
[a]NA, not applicable.

Figure 3.1. Protocatechuate branch of the β-ketoadipate pathway. Gene designations are in italics. CoA, coenzyme A.

polycyclic and homocyclic aromatic compounds



succinyl-CoA + acetyl-CoA

Figure 3.2. Phylogenetic tree of *pcaH* sequences from isolates, the natural salt marsh community, and the enrichment communities. The tree is based on the 159 nucleotides located in between the degenerate primer binding sites and is unrooted; *pcaH* from *Rhodoccocus opacus* 1CP is the outgroup. Major clone groups are indicated, and the numbers in parenthesis are the number of clones. Isolate sequences are color coded with either black type (sequences from roseobacter group isolates) or white type (sequences from members of other phylogenetic groups). Sequences from the salt marsh community and enrichment communities are color coded by treatment and are identified by designations shown in Table 3.1. Boostrap values greater than 50% are indicated at branch nodes.

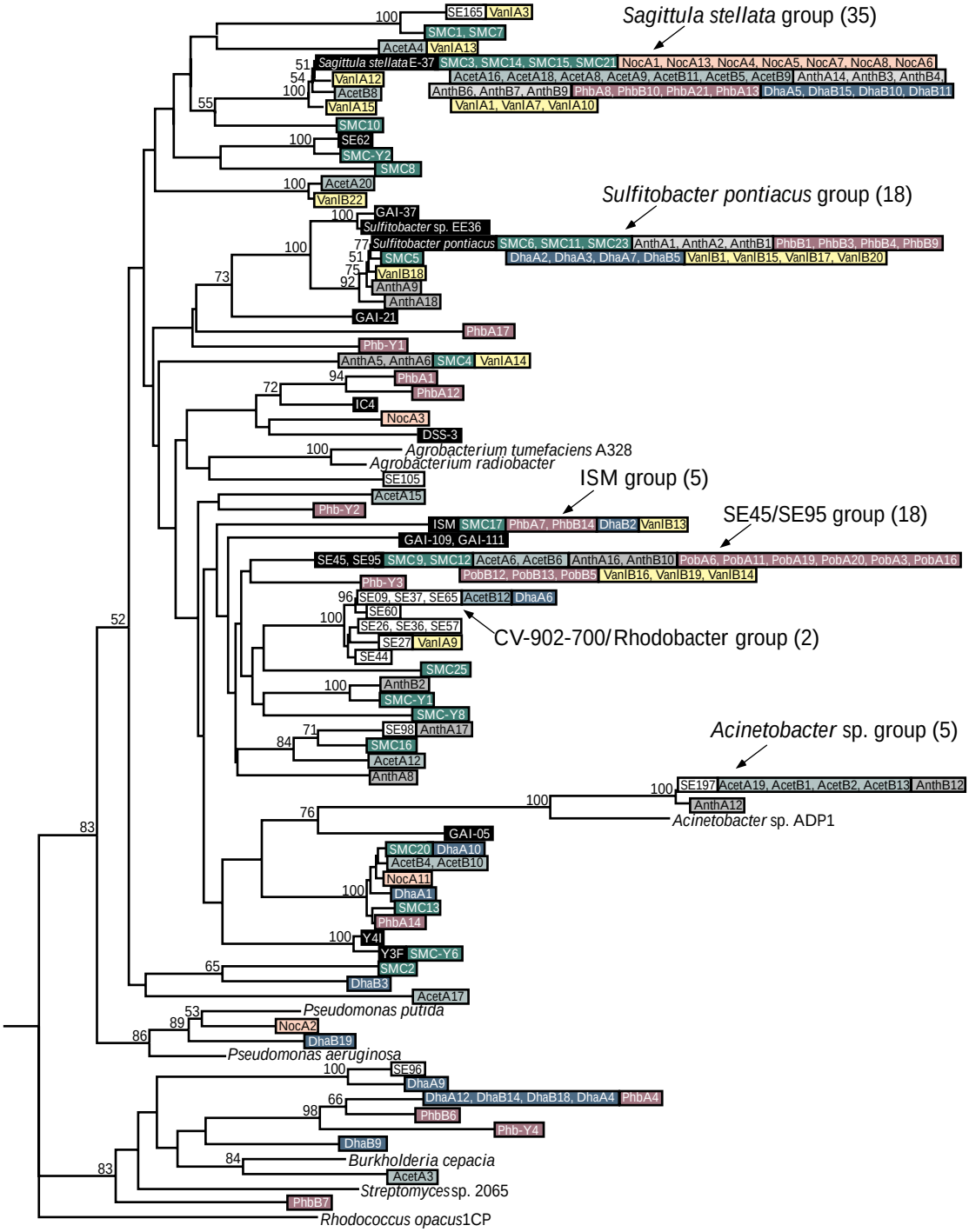
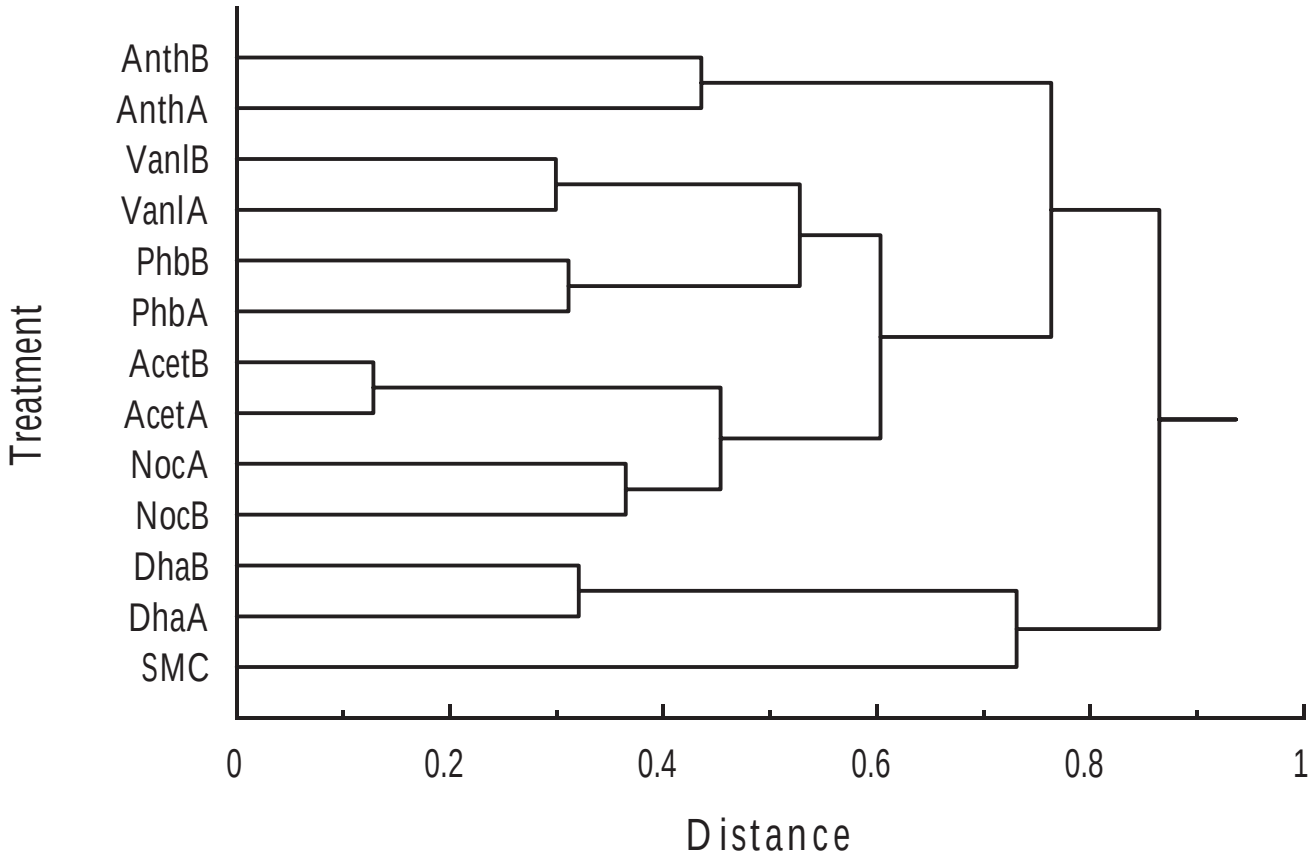

Sagittula stellata group (35)
Sulfitobacter pontiacus group (18)
ISM group (5)
SE45/SE95 group (18)
CV-902-700/ Rhodobacter group (2)
Acinetobacter sp. group (5)
SE165 VanIA3
SMC1, SMC7
AcetA4 VanIA13
Sagittula stellata E-37 SMC3, SMC14, SMC15, SMC21 NocA1, NocA13, NocA4, NocA5, NocA7, NocA8, NocA6
AcetA16, AcetA18, AcetA8, AcetA9, AcetB11, AcetB5, AcetB9 AnthA14, AnthB3, AnthB4,
AnthB6, AnthB7, AnthB9 PhbA8, PhbB10, PhbA21, PhbA13 DhaA5, DhaB15, DhaB10, DhaB11
VanIA1, VanIA7, VanIA10
VanIA12
AcetB8
VanIA15
SMC10
SE62
SMC-Y2
SMC8
AcetA20
VanIB22
GAI-37
Sulfitobacter sp. EE36
Sulfitobacter pontiacus SMC6, SMC11, SMC23 AnthA1, AnthA2, AnthB1 PhbB1, PhbB3, PhbB4, PhbB9
DhaA2, DhaA3, DhaA7, DhaB5 VanIB1, VanIB15, VanIB17, VanIB20
SMC5
VanIB18
AnthA9
AnthA18
GAI-21
PhbA17
Phb-Y1
AnthA5, AnthA6 SMC4 VanIA14
PhbA1
PhbA12
IC4
NocA3
DSS-3
Agrobacterium tumefaciens A328
Agrobacterium radiobacter
SE105
AcetA15
Phb-Y2
ISM SMC17 PhbA7, PhbB14 DhaB2 VanIB13
GAI-109, GAI-111
SE45, SE95 SMC9, SMC12 AcetA6, AcetB6 AnthA16, AnthB10 PobA6, PobA11, PobA19, PobA20, PobA3, PobA16
PobB12, PobB13, PobB5 VanIB16, VanIB19, VanIB14
Phb-Y3
SE09, SE37, SE65 AcetB12 DhaA6
SE60
SE26, SE36, SE57
SE27 VanIA9
SE44
SMC25
AnthB2
SMC-Y1
SMC-Y8
SE98 AnthA17
SMC16
AcetA12
AnthA8
SE197 AcetA19, AcetB1, AcetB2, AcetB13 AnthB12
AnthA12
Acinetobacter sp. ADP1
GAI-05
SMC20 DhaA10
AcetB4, AcetB10
NocA11
DhaA1
SMC13
PhbA14
Y4I
Y3F SMC-Y6
SMC2
DhaB3
AcetA17
Pseudomonas putida
NocA2
DhaB19
Pseudomonas aeruginosa
SE96
DhaA9
DhaA12, DhaB14, DhaB18, DhaA4 PhbA4
PhbB6
Phb-Y4
DhaB9
Burkholderia cepacia
AcetA3
Streptomyces sp. 2065
PhbB7
Rhodococcus opacus 1CP

Figure 3.3. Cluster analysis of 16S rRDNA T-RFLP profiles from the salt marsh and enrichment communities based on the relative areas of the major peaks. A similarity matrix was constucted using Euclidean distances, and clustering was performed by using Ward's method. The enrichment designations are described in Table 3.1.


AnthB
AnthA
VanlB
VanlA
PhbB
PhbA
AcetB
AcetA
NocA
NocB
DhaB
DhaA
SMC
Treatment
0
0.2
0.4
0.6
0.8
1
Distance

CHAPTER 4

THE DYNAMIC NATURE OF GENE ORGANIZATION ASSOCIATED WITH THE β-KETOADIPATE PATHWAY IN MEMBERS OF THE MARINE ROSEOBACTER LINEAGE[3]

[3]Buchan, A., E. L. Neidle, and M. A. Moran. To be submitted to *Molecular Biology and Evolution*.

Abstract

Members of the roseobacter lineage, an ecologically important marine clade within the class α-Proteobacteria, harbor genes encoding enzymes of the protocatechuate branch of the β-ketoadipate pathway, a major catabolic route for lignin-related aromatic compounds. The genes of this chromosomally-encoded pathway clustered in previously studied organisms, although gene order varied among organisms. In this study, we characterized genomic fragments containing gene clusters of the protocatechuate branch of the pathway in eight closely related members of the roseobacter lineage (pairwise 16S rRNA gene sequence identities of 92 to 99%). Sequence analysis revealed five unique gene arrangements. Identical gene organizations were found mostly between isolates demonstrating species-level identity (i.e. >99% similarity of 16S rRNA gene). In one isolate, six functionally related genes were identified: *pcaQ*, *pobA*, *pcaD*, -*C*, -*H*, and -*G*. The remaining seven isolates lacked at least one of these genes, although the relative order of remaining genes was preserved. Three genes (*pcaC, -H, -G*) were found in all isolates. The proximity of the genes to one another, coupled with the presence of a LysR-type regulator (*pcaQ*) in an adjacent upstream position, suggests that these clustered genes are co-regulated. These results portray the dynamic nature of a set of genes that potentially play a central catabolic role in coastal marine environments.

Introduction

The β-ketoadipate pathway provides a model system for studying mechanisms of evolution in an ecologically important catabolic pathway. This widely distributed, chromosomally-encoded convergent pathway plays a central role in the degradation of naturally-occurring aromatic compounds that arise during the decay of lignin and other vascular plant components, as well as environmental pollutants (21). The pathway is biochemically conserved, and the structural genes encoding enzymes of the pathway are similar among phylogenetically diverse organisms (21). Despite this mechanistic conservation, studies of a limited number of soil bacteria demonstrate remarkable diversity of this pathway in terms of gene organization, regulation, inducing metabolites, and transport systems (35). This diversification may be reflective of the distinctive selection pressures faced by the organisms maintaining the pathway and thus may reveal characteristic features that are specific to the bacterial group in which the pathway resides (21, 34)

The degradation of aromatic compounds typically proceeds by conversion of the ring structure to one of several di- or trihydroxylated intermediates in preparation for enzymatic cleavage. The most prevalent intermediates are protocatechuate and catechol. These hydroxylated intermediates then undergo ring fission followed by a series of reactions leading to the generation of substrates that feed into the tricarboxylic acid cycle. In the β-ketoadipate pathway, protocatechuate and catechol are cleaved between their two hydroxyl groups by protocatechuate 3,4-dioxygenase (3,4-PCD) and catechol 1,2-dioxygenase (1,2-CTD), respectively. In five additional steps, the ring cleavage products

are converted to succinyl- and acetyl-CoA (Fig. 4. 1). Both 3,4-PCD and 1,2-CTD belong to a large class of non-heme iron containing dioxygenases. The usually co-transcribed *pcaG* and *pcaH* genes encode the α and β subunits that comprise the heterodimeric 3,4-PCD. 1,2-CTD is usually composed of homodimers encoded by the *catA* gene.

The genes encoding the enzymes, transport, and regulatory proteins of the β-ketoadipate pathway are usually present in chromosomal supraoperonic clusters (34). An remarkable example of such an assemblage is demonstrated by *Acinetobacter* sp. ADP1, where more than 40 genes involved in the catabolism of plant-related compounds, most leading to protocatechuate, form a 30 kbp chromosomal cluster (36). However, the extent of gene clustering varies among the phylogenetically diverse organisms possessing the pathway. For instance, *Pseudomonas putida* demonstrates a contrasting scenario, as the genes required for catabolism of protocatechuate alone are dispersed on the chromosome in three distinct gene clusters (34). Furthermore, with the exception of genes that would be expected to evolve as a single transcriptional unit because they encode subunits of a single enzyme (e.g. *pcaHG*, *pcaIJ*), gene order does not appear to be maintained within the operons (21).

Examination of the sequence and organization of these catabolic genes provides an opportunity to gain insight into the mechanisms of chromosomal rearrangements contributing to the diversification of this pathway in distinct evolutionary lineages. Such studies might also offer further evidence for structural and biochemical interactions between enzymes of the pathway postulated previously (13). Finally, analysis of these

genes could provide a unique perspective on microbial diversity. A substantial degree of sequence diversity was identified within a segment of *pcaH* from natural and enriched bacterial communities, suggesting a significant amount of phenotypic microheterogenity for a single catabolic step of the pathway (3). Whether this sequence diversity extends to other protein-encoding genes of the pathway among co-existing populations would provide evidence relating to the acquisition and evolution of this pathway and how it contributes to microbial diversity within natural communities.

The marine roseobacter lineage provides an ideal system for genetic studies of the β-ketoadipate pathway. Representatives of this lineage are abundant in coastal seawater, aerobic sediments, and decaying plant material from coastal salt marshes (7, 14, 28, 44). Previous studies have demonstrated widespread capabilities for the transformation of lignin-related aromatic monomers, and the protocatechuate branch of the β-ketoadipate pathway has been identified in several isolates (3, 4). Furthermore, more than half of the *pcaH* genes retrieved by PCR amplification from salt marsh bacterial communities (77 of 149), could be traced to members of the roseobacter lineage (3). Thus, this group provides a unique opportunity to investigate the evolution of an important catabolic pathway in an ecologically relevant lineage of marine bacteria.

To gain a better perspective on conservation of sequence and gene arrangement in closely-related organisms, we analyzed protein-encoding genes of the protocatechaute branch of the β-ketoadipate pathway in eight roseobacter isolates. Comparative analyses of chromosomal regions housing *pcaH* suggests that dispersion of the genes of the pathway is characteristic of this lineage, and that diversity in gene organization might

reflect physiological differences that contribute to the overall success of these organisms in their environment.

Materials and Methods

Bacterial isolation and 16S rRNA gene analysis. Most isolates examined in this study were cultured from seawater, sediments, or decaying salt marsh grass collected in the estuaries and coastal waters of the southeastern U. S. Several of the strains were isolated from lignin or aromatic monomer enrichment cultures (isolates Y3F, Y4I, *Sagittula stellata* E-37, and *Sulfitobacter* sp. strain EE-36) (3, 17). SE45 was cultured directly from coastal seawater using nonselective, low-nutrient seawater plates (4). '*Silicibacter pomeroyi*' DSS-3 was derived from a marine dimethylsulfoniopropionate (DMSP) enrichment (16). Finally, two isolates that were not obtained from the southeastern U.S. coast were examined: *Sulfitobacter pontiacus* ChLG 10, which was cultured from the Black Sea (42), and '*Roseovarius nubinhibens*' ISM, which was cultured from the Caribbean Sea (12).

16S rRNA gene sequences were reported previously for all isolates (accession numbers follow parenthetically): '*S. pomeroyi*' DSS-3 (AF098491), *Sulfitobacter* sp. EE-36 (AF007254), '*R. nubinhibens*' ISM (AF098495), Y3F (AF253467), Y4I (AF388307), SE45 (AF388308), *S. stellata* E-37 (U58356), *S. pontiacus* (Y13155).

Detection and isolation of catabolic genes from roseobacter isolates. The catabolic gene clusters were initially identified in isolates by targeting a region of *pcaH*. A degenerate PCR primer pair designed from conserved PcaH regions (3) was used to

amplify a 212 bp product from all isolates. This product was gel purified and labeled with digoxigenin (DIG) by a random priming reaction (Genius system; Roche Molecular Biochemicals, Indianapolis, IN), and used in Southern hybridizations with isolate chromosomal digestions. DNA fragments of the sizes corresponding to bands giving positive hybridization signals in Southern blot analysis were gel purified and ligated into the pZERO vector (Invitrogen Corp., Carlsbad, CA). Colony hybridization of genomic libraries with DIG-labeled DNA probes identified positive clones. In some cases, '*S. pomeroyi*' DSS-3, *Sulfitobacter* sp. EE-36, and SE45, adjacent fragments were identified by Southern hybridization analysis using DIG-labeled probes generated from distal ends of the primary fragment (Fig. 4.2).

Sequence determination of the genomic fragments was facilitated by using the GeneJumper Primer Insertion Kit for Sequencing (Invitrogen Corp.), which randomly inserted a minimal version of the *Mu* transposon containing two primer binding sites and a selectable marker into the target DNA.

Expression of roseobacter DNA in *E. coli*. To express the *pcaHG* genes from isolates SE45 and '*R. nubinhibens*' ISM under control of the *lac* promoter in *Escherichia coli*, p4NK and p6NB were constructed using PCR primers that introduced restriction sites for optimal positioning in the expression vector pCYB1 (New England Biolabs, Beverly, MA). A *Nde*I cleavage site was introduced just before the ATG start codon of the *pcaH* gene, and either a *Kpn*I (SE45) or *Bam*HI ('*R. nubinhibens*' ISM) cleavage site was introduced downstream of the *pcaG* stop codon by PCR amplification using the high-fidelity *pfu* DNA polymerase (Stratagene, La Jolla, CA). The 1.3-kbp fragments were

then ligated into the corresponding sites on the pCYB1 vector. The correct sequences of the resultant recombination plasmids were confirmed. Luria broth cultures (100 ml) of plasmid-carrying *E. coli* Top10F' cells (Invitrogen Corp.) were grown at 30°C for 12 hr. At the time of inoculation, 100 µM isopropyl-γ-D-thiogalacto-pyranoside (IPTG) was added to cultures of Top10F' carrying p4NK and p6NB. The induced cells were harvested by centrifugation, washed once with sterile Tris buffer, and stored at -20°C.

3,4-PCD enzyme assays. Cell pellets were suspended in 200 µl of breaking buffer [50 mM Tris-HCl, 10% glycerol, 5 mM $(NH_4)_2SO_4$, 2.5 mM EDTA, 1 mM dithiothreitol (pH 7.5)]. Cell extracts were prepared as previously described (41), and 3,4-PCD activity was determined spectrophotometrically by measuring the decrease in absorbance at 290 nm (43). Protein concentrations were determined by the method of Bradford (2).

Sequence determination and analysis. DNA sequences were determined with double-stranded templates and primers that recognized the cloning vector or the GeneJumper transposon. When necessary, new oligonucleotide primers were made based on previously sequenced regions. Either an ABI3700 or ABI310 automated DNA sequencer (Applied Biosystems, Foster City, CA) was used. Sequences reported here were submitted to the GenBank database. Homology searches (BLAST) were carried out at the network server of the National Center for Biotechnology Information. Sequences were analyzed using the Genetics Computer Group program package 10.1 (Wisconsin Package Version; Madison, WI.). Phylogenetic trees were constructed for sequences with the PHYLIP package (11) by using evolutionary distances (Jukes Cantor or Kimura) and tree building algorithms (Neighbor or Fitch) or a parsimony method (ProtPars). The

pcaC, -*H*, and -*G* gene sequences were tested for intragenic recombination using the program Partial Likelihoods Assessed through Optimisation (PLATO; http://evolve.zoo.ox.ac.uk/program). The conserved hypothetical proteins were analyzed for the presence of signature sequence motifs using the programs PROSITE (http://ca.expasy.ch/prosite/), PRINTS (http://bioinf.mcc.ac.uk/cgi-bin/dbbrowser/PRINTS), and BLOCKS (www.blocks.fhcrc.org/blocks_search.html). The conserved hypothetical protein was also searched for putative membrane spanning regions using the MEMSAT2 program at http://insulin.brunel.ac.uk.

Results

Choice of roseobacter strains. The isolates examined in this study are part of a larger collection of roseobacter group strains cultured from the vascular-plant dominated coastal marshes of the southeastern U.S. (4, 14, 17). *pcaH* genes in these eight isolates were identical or highly similar to *pcaH* gene fragments retrieved by PCR from salt marsh DNA, suggesting that the isolates are close relatives of the bacteria carrying out aromatic compound degradation in the marsh ecosystem (4). The marine roseobacter lineage forms a monophyletic clade within the α-3 subclass of the Proteobacteria (14); pairwise sequence identities of 16S rRNA genes from the eight roseobacter isolates ranged from 92 to 100%. Two pairs of the isolates (Y3F and Y4I; *Sulfitobacter* sp. EE-36 and *S. pontiacus*) demonstrate species-level similarity. Y3F and Y4I have 100% sequence identity over 1376 bp of the 16S rRNA gene corresponding to positions 48-1484 in

Escherichia coli. The two *Sulfitobacter* isolates have 99.8% identity over 1357 bp corresponding to positions 50-1506 in *E. coli*.

Identification of catabolic genes in roseobacter isolates. Genomic fragments housing genes involved in the catabolism of protocatechuate were identified in *S. pontiacus*, *Sulfitobacter* sp. EE-36, *S. pomeroyi* DSS-3, *R. nubinhibens* ISM, Y4I and SE45 by Southern hybridization analysis targeting *pcaH*. Isolation, cloning, and sequencing of these fragments revealed that in addition to *pcaH* these fragments contained open reading frames (ORFs) with significant sequence similarity to proteins of the protocatechuate branch of the β-ketoadipate pathway. Three of the ORFs, designated *pcaC*, *pcaH*, and *pcaG*, were present in all six roseobacter isolates, and were presumed to encode γ-carboxymuconolactone decarboxylase (PcaC), and protocatechuate 3,4-dioxygenase (PcaHG) (Fig. 4.1). An ORF termed *pcaQ* was also identified upstream of and oriented divergently from the catabolic genes in four of the six isolates examined; however, only a partial sequence was obtained from '*R. nubinhibens*' ISM (Fig. 4.2). This gene putatively encodes a LysR-type transcriptional regulator (PcaQ). No ORFs showing similarity to transcriptional regulators were evident in the immediate vicinity of the *pca* genes in either *Sulfitobacter* strain. An ORF designated *pobA*, based on sequence similarity to *p*-hydroxybenzoate hydroxylase (PobA), was found in two of the isolates ('*R. nubinhibens*' ISM and '*S. pomeroyi*' DSS-3). Finally, an ORF similar to *pcaD* that encodes enol-lactone hydrolase (PcaD) was present in two of the isolates ('*R. nubinhibens*' ISM and SE45). The orientation and proximity of the *pca* genes to one another suggests they are

transcribed as a single unit (Fig. 4.2), as has been demonstrated previously for several soil bacteria (21).

Sequence analysis of the *pcaC*, *-H*, and *-G* catabolic genes from roseobacter isolates. Genes encoding proteins of the protocatchuate branch of the β-ketoadipate pathway were previously identified in two additional roseobacter isolates, *S. stellata* E-37 and isolate Y3F (3), and these genes are included in sequence analyses presented here. The three genes present in all eight isolates encode two distinct enzymes catalyzing reactions in the degradation of protocatechuate (*pcaC*, *pcaH*, and *pcaG*; Fig. 4.1). Among the sequences obtained from roseobacter group isolates, nucleotide sequence similarities were between 69.8 and 97.2% for *pcaC*, 61.6 and 98.9% for *pcaG,* and 63.2 and 97.6% for *pcaH*. Pairwise comparisons of the roseobacter amino acid sequences deduced from *pcaC*, *pcaG* and *pcaH* were carried out with the corresponding sequences from the soil bacteria *Acinetobacter* sp. ADP1, *Pseudomonas aeruginosa*, *P. putida*, *Sinorhizobium meliloti*, *Mesorhizobium loti*, *Caulobacter crescentus*, *Agrobacterium tumefaciens*, and *Rhodococcous opacus*. The *P. aeruginosa*, *S. meliloti*, *M. loti*, and *C. crescentus* gene sequences were obtained from genomic data and their function has not yet been demonstrated experimentally. In these comparisons, the residues were >36.5% similar and >28.2% identical for PcaG and >50.8% similar and >38.5% identical for PcaH. In alignments, residues known to be important for catalytic function (30) were well conserved within all of the roseobacter PcaGH sequences. Sequence analysis indicated a 72 bp insertion in *pcaG* of SE45, which was confirmed by PCR amplification and sequencing of the region from genomic DNA preparations (data not shown). The crystal

structure of PcaHG from *P. putida* has been solved (30), and sequence comparisons with PcaG from SE45 suggests that this insertion is within a loop on the outside face of the enzyme. Pairwise comparisons of roseobacter sequences with PcaC proteins showed >45.5% sequence similarity and >29.4% identity.

The phylogenetic relationship of PcaHG suggests that the roseobacter sequences do not form a cohesive lineage; sequences from *S. stellata* E-37 and the two *Sulfitobacter* strains cluster with α-proteobacterial sequences but not with the other rosebacter isolates (Fig. 4.3). However, the deep branches are difficult to resolve, and thus the phylogenetic relationship of these proteins within the context of other α-proteobacteria is not easily interpreted with the sequences on hand. In contrast, the phylogenetic relationship of PcaC (Fig. 4.4) illustrates a branching order that is consistent with that provided by 16S rRNA gene analysis (Fig. 4.5). Phylogenetic trees constructed for the individual PcaG and PcaH subunits were identical in overall tree topology to one another and to the PcaHG tree (data not shown). Branch order was maintained when PcaC and PcaHG trees were constructed using alternative tree-building algorithms (Fitch) and a parsimony method (ProtPars) (data not shown).

Using the PLATO program of Grassly and Holmes (18) we found no evidence of intragenic recombination within the sets of 16 *pcaC*, *pcaH*, or *pcaG* genes (eight from roseobacter isolates and eight from phylogenetically diverse soil bacteria) analyzed. The relative rates of evolution were assessed using the method outlined in Dykhuizen and Green (9). The similarity in percent divergence of *pcaC*, *-H*, and *-G* among the 16 genes analyzed (32.2%, 34.2%, and 32.7%, respectively) suggested that the genes are evolving

at approximately the same rate. Significant differences were evident for all three genes when compared to the 16S rRNA gene, which had significantly lower sequence divergence as determined by an ANOVA test (9.6%; $p < 0.001$). A regression of pairwise distances of each of the PcaC, PcaH, and PcaG proteins against pairwise distances of the 16S rRNA gene provided slopes that were not significantly different from one another, likewise suggesting similar rates of evolution of the *pca* genes.

Expression of the *pcaHG* genes in *E. coli*. Because cleavage of the aromatic ring is the critical catabolic obstacle in the degradation of protocatechuate, demonstration of activity of PcaHG in roseobacter isolates would provide strong evidence that the pathway is functional in these organisms. *pcaHG* genes of two of the roseobacter isolates, SE45 and *R. nubinhibens* ISM, were transformed into *Escherchia coli*, a bacterium that does not encode 3,4-PCD. The presence of *pcaHG* from isolate SE45 resulted in IPTG-inducible (100 M) 3,4-PCD activity in cell extracts of the plasmid-bearing *E. coli* strains (258 ± 0.053 nmol/min/mg), while activity was not evident from the *pcaHG* genes from '*R. nubinibens*' ISM. No activity was detected in *E. coli* with the cloning vector or with recombinant plasmids in the absence of IPTG. Heterologous expression of *pcaGH* in *E. coli* was demonstrated previously for *S. stellata* E-37 and isolate Y3F (3).

Sequence analysis of additional *pca* related genes. The ORF designated *pobA* was found immediately upstream of and transcribed in the same direction as *pcaC* in *S. stellata* E-37 and '*S. pomeroyi*' DSS-3 and upstream of *pcaD* in '*R. nubinhibens*' ISM. This gene appears to encode a hydroxylase for the conversion of *p*-hydroxybenzoate to protocatechuate. Its deduced amino acid sequence was >64% identical to PobA

sequences from soil bacteria. The regions associated with flavin adenine dinucleotide and substrate binding (45) were highly conserved among all the roseobacter sequences. Pairwise comparisons show the roseobacter sequences have significant nucleotide sequence identity (65.0 - 71.8%) and deduced amino acid similarity (70.3 - 79.5%) and identity (60.5 - 69.5%). The putative protein in *S. stellata* E-37 appears to contain an additional three residues between the completely conserved amino acids at position 127 (Val) and position 137 (Pro) of the corresponding proteins from soil bacteria (3). These additional residues are not present in the deduced PobA sequences from *R. nubinhibens* ISM or *S. pomeroyi* DSS-3.

The ORF designated *pcaD* was immediately upstream of and transcribed in the same direction as *pcaC* in '*R. nubinhibens*' ISM and isolate SE45. Pairwise comparisons show the two roseobacter isolates have 70.6% nucleotide sequence identity and 71.1% deduced amino acid similarity and 66.5% identity. The deduced amino acid sequence also suggests homology to the CatD protein which catalyzes the analogous reaction in the catechol branch of the β-ketoadipate pathway (20). The conserved active site cysteine demonstrated to be critical to hydrolysis in the protein from *Pseudomonas* sp. B13 (37) is present in the sequences of both roseobacter isolates. The amino acids surrounding this residue are also fairly well conserved (GYXXXCXA).

The ORF designated *pcaQ* was upstream and divergently transcribed from the *pca* catabolic genes in all roseobacter isolates except the two *Sulfitobacter* strains. These putative LysR-type roseobacter regulatory proteins show 39 to 76% identity to PcaQ from other α-proteobacteria (*A. tumefaciens*, *S. meliloti*, and *M. loti*) as well as sequence

similarity to several LysR-type regulators of the catechol branch of the β-ketoadipate pathway and of the modified *ortho* chlorocatechol pathway (Fig. 4.6). Pairwise comparisons show the roseobacter isolates have between 48.7 and 98.3% nucleotide sequence identity, and 45.0 and 98.7% deduced amino acid similarity and 38.4 and 98.4% identity. Possible PcaQ binding sites that have the consensus sequence T-N_{11}-A and dyad symmetry were present within the 50 bp immediately upstream of the putative *pcaQ* ATG start codon in all roseobacter isolates. The region of highest similarity among the PcaQ proteins was in the N terminus, an area presumed to comprise a helix-turn-helix motif for DNA binding (40). Most LysR family members show a reduced lysine (Lys) content and an increased arginine (Arg) content compared to other bacterial proteins, possibly related to the DNA-binding function of these transcriptional regulators (47). Relative to the ratios in surrounding proteins, low Lys to Lys-plus-Arg ratios were found in the PcaQ proteins from roseobacter isolates (Table 4.1).

The phylogenetic relationship of PcaQ is shown in the context of other LysR-type transcriptional regulators that modulate expression of genes involved in aromatic compound catabolism (Fig. 4.6). The deduced amino acid sequences of roseobacter PcaQ proteins cluster confidently with other PcaQ designated sequences, although, as seen with the PcaHG phylogeny, the precise ordering of the deep branches is not clear. As might be expected, a 55 residue stretch of amino acids in the central region of the protein proposed to be involved in inducer binding (6) appears more highly conserved among the PcaQ proteins relative to the other LysR-type proteins. However, within the PcaQ proteins the degree of sequence conservation over this region was lower in PcaQ from isolates Y3F

and Y4I relative to these proteins from the other roseobacter isolates. Whether this sequence divergence is significant enough to suggest different inducing metabolites is unknown. By analogy to PcaQ of *A. tumefaciens* and other LysR-type regulators, the roseobacter PcaQ proteins might be expected to regulate their own synthesis and also control the expression of genes downstream of and divergently transcribed from *pcaQ* (33).

The G+C content has been determined to be 65.0 and 66.8 for isolates *S. stellata* E-37 and '*S. pomeroyi*' DSS-3, respectively (15, W. B. Whitman pers. comm). The G+C content of the *pca* genes (Table 4.1) is consistent with that of the genomes of these organisms. The length of the *pobA* and *pca* genes as well as the molecular weights, calculated from the deduced amino acids, (Table 4.1) are consistent with those found for their soil counterparts.

A conserved hypothetical protein present in all roseobacter isolates. A highly conserved ORF was found immediately downstream of and oriented in the same direction as *pcaG* in all roseobacter isolates (Fig. 4.2). The complete gene for this conserved hypothetical protein (CHP) was retrieved from *S. stellata* E-37 and isolate SE45, and is 828 bp and 837 bp in length, respectively. The two genes share 73.9% identity at the nucleotide level. Partial sequences of the CHP were obtained for the remaining six roseobacter isolates, with nearly complete genes obtained from *S. pomeroyi* DSS-3 (688 bp) and *S. pontiacus* (458 bp). Pairwise comparisons of *S. stellata* E-37, SE45, *S. pontiacus*, and *S. pomeroyi* DSS-3 show ≥71.8% nucleotide identity for the CHP**.**

The orientation and proximity of this CHP suggests that it may be co-transcribed with the *pca* genes located immediately upstream. However, the function of the CHP remains unknown. Searches against the database indicated the CHP from the two roseobacters had >47% similarity and >36% identity to a CHP in *Mycobacterium tuberculosis* (G15608856). Exhaustive sequence searches were conducted to identify protein motifs or signature sequences that might reveal catalytic or functional properties of this hypothetical protein (refer to materials and methods section), yet none were found. The absence of transmembrane helices suggests the putative protein is cytoplasmic. In addition to sequence similarity to the putative protein in *M. tuberculosis*, the complete CHP sequences from isolates SE45 and *S. stellata* E-37 demonstrate between 23.6 and 29.3% sequence identity and 34.9 and 41.8% similarity to a CHP in the *M. loti* (AP003015) and *S. meliloti* (AL591790) genomes. ORFs having similarity to this CHP were found in five additional organisms, including the distantly related organisms *Streptomyces coelicolor* and *Archaeoglobus fulgidus*. Comparative analyses of the genes in the proximity of the CHP in these eight organisms indicated the presence of a gene showing homology to 3-hydroxyacyl CoA dehydrogenase, a protein involved in β-oxidation of fatty acids, in all of these organisms. The conservation of the proximity of these two genes among phylogenetically diverse organisms suggests a functional relationship between these genes (31).

Additional ORFs in the vicinity of the *pca* genes. The chromsomal fragments from Y3F and Y4I contained an ORF that showed significant similarity (38.1%) to a protein involved in the degradation of the aromatic herbicide methyl parathion in *Plesiomonas sp.*

DLL-1 (AY029773). Three additional ORFs were sequenced from SE45: a 666 bp ORF had greatest similarity (59.3%) to a putative TetR-like transcriptional regulator from *A. tumefaciens* (G15890456), a partial ORF (287 bp) showed highest similarity to a putative shikimate dehydrogenase from *A. tumefaciens* (G15890454), and a 591 bp ORF had highest similarity to PobR, a regulatory protein that has been demonstrated to modulate the expression of *pobA* in soil bacteria (8, 38). In the *Sulfitobacter* strains, a 1114 bp ORF showed 87% deduced amino acid similarity to a glutathione-dependent alcohol dehydrogenase from *Rhodobacter sphaeroides* (P72324) and a partial ORF showed significant sequence similarity to a putative aldehyde dehydrogenase from *M. loti* (g13473288).

Species-level comparisons. A comparison of sequence similarity in the fragments obtained from the two pairs of isolates demonstrating species level-similarity of the 16S rRNA gene showed Y3F and Y4I share 98.2% sequence identity over 3354 bp. *Sulfitobacter* sp. EE-36 and *S. pontaicus* share 97.6% identity over 3597 bp. The lowest sequence similarity at the nucleotide level between Y3F and Y4I was seen in *pcaC* (97.2%), however, all of these substitutions were synonomous, resulting in 100% identity at the protein level. The highest degree of protein sequence divergence in both sets of comparisons were seen in PcaG. The deduced amino acids sequences were 97.6% identical between Y3F and Y4I and 98.5% identical between *Sulfitobacter* sp. EE-36 and *S. pontaicus*. The remaining catabolic proteins, PcaH and PcaC, showed greater than 99% protein identity.

Discussion

Marine bacteria affiliated with the roseobacter clade are abundant in the highly productive salt marshes of the southeastern U. S., where lignin is a significant source of organic matter (14, 26, 27). Members of this lineage demonstrate capabilities for the degradation of lignin-related aromatic compounds, and catabolism of these substrates appears to occur via the protocatechuate branch of the β-ketadipate pathway. Growth on *p*-hydroxybenzoate, a feature typical of organisms possessing this pathway, is characteristic of all roseobacter group members demonstrated to have *pca* genes (3, 4, unpublished results). Expression of roseobacter *pcaHG* genes in *E. coli,* coupled with identification of additional genes involved in protocatechuate degradation, provide evidence that the pathway is functional in this bacterial clade (3, this study).

The dynamic nature of the organization of genes involved in the catabolism of protocatechuate among closely-related organisms is evident from sequence analysis of the eight roseobacter isolates. Five unique gene arrangements were identified, with identical gene organization in each of the two species pairs. Identical *pca* gene organization was also evident between *S. stellata* E-37 and *S. pomeroyi* DSS-3, which share 93.1% sequence identity of the 16S rRNA gene, 64.1% sequence similarity over a 4.97 kbp chromosomal region containing the *pobA* and *pca* genes.

The absence of a transcriptional regulator in the *Sulfitobacter* isolates is surprising, but not unprecedented, as *Pseudomonas putida* does not appear to have a transcriptional regulator in the immediate vicinity of the *pcaHG* genes (21, Fig. 4.5). These findings suggest that regulation of the metabolic enzymes of the pathway may be unique in the

Sulfitobacter strains relative to the other roseobacter isolates. Similarly, the dispersion of the *pobA* and *pcaD* genes might suggest at least five alternative regulatory schemes for the catabolism of *p*-hydroxybenzoate and protocatechate in the eight roseobacter isolates.

The presence of a CHP gene located immediately downstream of *pcaG* in all eight isolates is intriguing. The proximity and orientation of this putative gene suggests it is co-transcribed with the *pca* genes, and therefore might be expected to play a role in aromatic compound metabolism. Preliminary evidence from analysis of bacterial genome sequences indicates that this CHP might be similar to proteins that play a role in the oxidation of fatty acids. Further genetic and biochemical investigations are necessary to elucidate the function of this CHP.

The importance of transport proteins in defining the biological individuality of organisms harboring the β-ketoadipate pathway is becoming increasingly evident (35, 48). In members of the γ-proteobacteria (i.e. *Acinetobacter* sp. ADP1, *P. putida*, and *P. aeruginosa*), genes for the transport of aromatic compounds are typically located within the same operon as the catabolic genes (21), suggesting an intimate coordination between uptake and catabolism in this lineage. This trait does not appear to be consistent in all organisms, as transport genes are absent in *pcaHG*-containing operons for roseobacter isolates. The absence of adjacent regulatory genes also appears to extend to other members of the α- proteobacteria. For example, no transport genes were found in the *pca* clusters of *A. tumefaciens* (34)*, C. crescentus, S. meliloti,* nor *M. loti*. Studies of cultured (3) and uncultured (4) representatives of the roseobacter group suggest these bacteria are capable of converting a variety of ring structures. Therefore, transport genes may instead

be associated with upper pathway catabolic genes required for conversion of diverse aromatic substrates to protocatechuate.

The presence of both branches of the β-ketoadipate pathway within a single bacterium may be pivotal in dictating the genetic organization of genes in the pathway. Both the catechol and protocatechuate branches of the pathway have been found in soil microorganisms that belong to the γ-proteobacteria, including *Acinetobacter* sp. ADP1, *P. putida*, *P. aeruginosa*, and the Gram positive bacterium *R. opacus*. Alternatively, there is no evidence of the catechol branch of the pathway in the roseobacter isolates (3), nor do they appear to be present in other *pca*-containing α-proteobacteria, including *A. tumefaciens*, *S. meliloti*, *M. loti*, or *C. crescentus,* based on genome sequence analyses. Organisms possessing both branches of the pathway may require additional regulatory mechanisms for dictating the preferential hierarchy in the uptake of substrates when presented with the mixture of aromatic compounds found in natural environments (1, 13, 29). In addition, the presence of both branches of the pathway may lead to a greater selective pressure to maintain the genes of each branch within a limited number of transcriptional units.

The organization of the genes of the β-ketoadipate pathway within operons may facilitate the intimate association of proteins with related functions and contribute to the efficient funneling of metabolites (13). Localization of *pobA* within the *pca* cluster of several of the roseobacter isolates might therefore enhance the enzymatic conversion of *p*-hydroxbenzoate to protocatechuate. In fact, structural studies suggest a physical interaction that facilitates transfer of metabolites between PobA and PcaHG (A. M.

Orville, pers. comm.). This novel gene organization also points to the importance of *p*-hydroxybenzoate as a substrate feeding into the β-ketoadipate pathway in members of the roseobacter lineage.

Since the organization of genes within operons is believed to provide coordinated regulation and production of functionally related genes, one might expect that operon conservation would be a logical and economical strategy. However, this is frequently not the case, implying that destruction of operon structure is largely selectively neutral in long term evolution (22). Conservation of operon organization is often more prevalent among nonessential genes, presumably to facilitate horizontal gene transfer of complete functional units (24). The extent of genetic rearrangements and sequence divergence found within the roseobacter clade suggests an ancient acquisition of the β-ketoadipate pathway and points to its importance among members of this lineage. However, during the course of evolution of this pathway, chromosomal rearrangements appear to have played a significant role in driving organisms to assume a unique genetic identity. This raises questions concerning the broader ecological implications of genetic variation on niche partitioning. However, subtle metabolic variations that may be potentially ecologically relevant, are difficult to address experimentally. Nonetheless, this initial assessment of the genetic diversity of the pathway provides clues as to likely sources of functional diversification.

References

1. Ampe, F., and N. D. Lindley. 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. J. Bacteriol. 177:5826-5833.

2. Bradford, M. M. 1976. A rapid and sensitive method for the quantification of microgram quantitites of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248-254.

3. Buchan, A., L. S. Collier, E. L. Neidle, and M. A. Moran 2000. Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine Roseobacter lineage. Appl. Environ. Microbiol. 66: 4662-4672.

4. Buchan, A., E. L. Neidle, and M. A. Moran. 2001. Diversity of a ring-cleaving dioxygenase gene *pcaH* in a salt marsh bacterial community. Appl. Environ. Microbiol. 67:pages.

5. Coco, W. M., R. K. Rothmel, S. Henikoff, A. M. Chakrabarty. 1993. Nucleotide sequence and initial functional characterization of the *clcR* gene encoding a LysR family activator of the *clcABD* chlorocatechol operon in *Pseudomonas putida*. J. Bacteriol. 175: 417-427.

6. Collier, L. S., G. L. Gaines, and E. L. Neidle. 1998. Regulation of benzoate degradation in *Acinetobacter* sp. strain ADP1 by BenM, a LysR-type transcriptional activator. J. Bacteriol. 180:2493-2501.

7. Dang, H. Y. and C.R. Lovell. 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. Appl. Environ. Microbiol. 66:467-475.

8. DiMarco, A. A., B. Averoff, and L. N. Ornston. 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*. J. Bacteriol. 175:4499-4506.

9. Dykhuizen, D. E. and L. Green. 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. 173:7257-7268.

10. Eulberg, D., S. Lakner, L. A. Golovleva, and M. Schl mann. 1998. Characterization of a protocatechuate catabolic gene cluster from *Rhodococcus opacus* 1CP: evidence for a merged enzyme with 4-carboxymuconolactone-decarboxylating and 3-oxoadipate enol-lactone-hydrolyzing activity. J. Bacteriol. 180:1072-1081.

11. Felsenstein, J. 1989. PHYLIP-Phylogeny Inference Package (version 3.2). Cladistics. 5:164-166.

12. Fuhrman, J. A., S. H. Lee, Y. Masuchi, A. A. Davis, and R. M. Wilcox. 1994. Characterization of marine prokaryotic communities via DNA and RNA. Microb. Ecol. 28:133-145.

13. Gaines, G. L., L. Smith, and E. L. Neidle. 1996. Novel nuclear magnetic resonance spectroscopy methods demonstrate preferential carbon utilization by *Acinetobacter calcoaceticus*. J. Bacteriol. 178:6833-6841.

14. González, J. M., W. B. Whitman, R. E. Hodson, and M. A. Moran. 1996. Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. Appl. Environ. Microbiol. 62:4433-4440.

15. González, J. M. and M. A. Moran. 1997. Numerical dominance of a group of marine bacteria in the α-subclass of the class proteobacteria in coastal seawater. Appl. Environ. Microbiol. 63:4237-4242.

16. González, J. M., F. Mayer, M. A. Moran, R. E. Hodson, and W. B. Whitman. 1997. *Sagittula stellata* gen. nov., sp. nov., a lignin-transforming bacterium from a coastal environment. Int. J. Syst. Bacteriol. 47(3):773-780.

17. González, J. M., R. P. Kiene, and M. A. Moran. 1999. Transformations of sulfur compounds by an abundant lineage of marine bacteria in the α-subclass of the class Proteobacteria. Appl. Environ. Microbiol. 65: 3810-3819.

18. Grassly, N. C. and E. C. Holmes. 1997. A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol. Biol. Evol. 14:239-247.

19. Harayama, S., M. Kok, and E. L. Neidle. 1992. Functional and evolutionary relationships among diverse oxygenases. Ann. Rev. Microbiol. 46:565-601.

20. Hartnett, G. B. and L. N. Ornston. 1994. Acquistion of apparent DNA slippage structures during extensive evolutionary divergence of *pcaD* and *catD* genes encoding identical catalytic activities in *Acinetobacter calcoaceticus*. Gene. 142: 23-29.

21. Harwood, C. S., and R. E. Parales. 1996. The β-ketoadipate pathway and the biology of self-identity. Ann. Rev. Microbiol. 50:553-590.

22. Itoh, T., K. Takemoto, H. Mori and T. Gojobori. 1999. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. Mol. Biol. Evol. 16:332-346.

23. Iwagami, S. G., Y. Keqian, and J. Davies. 2000. Characterization of the protocatechuic acid catabolic gene cluster from *Streptomyces* sp. strain 2065. Appl. Environ. Microbiol. 66:1499-1508.

24. Lawrence, J. G. and J. R. Roth. 1996. Selfish operons: Horizontal transfer may drive the evolution of gene clusters. Genetics. 143:1843-1860

25. Matrubutham, U. and A. R. Harker. 1994. Analysis of duplicated gene sequences associated with *tfdR* and *tfdS* in *Alicaligenes eutrophus* JMP34. J. Bacteriol. 176:2348-2353.

26. Moran, M. A. and R. E. Hodson. 1989. Formation bacterial utilization of dissolved organic carbon derived from detrital lignocellulose. Limnol. Oceanogr. 35:1744-1756.

27. Moran, M. A. and R. E. Hodson. 1994. Dissolved humic substances of vascular plant origin in a coastal marine environment. Limnol. Oceanogr. 39:762-771.

28. Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni. 1995. Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankon communities. Limnol. Oceanogr. 40:148-158.

29. Nichols, N. N., and C. S. Harwood. 1995. Repression of 4-hydroxybenzoate transport and degradation by benzoate: a new layer of regulatory control in the *Pseudomonas putida* β-ketoadipate pathway. J. Bacteriol. 177:7033-7040.

29. Nichols, N. N., and C. S. Harwood. 1995. Repression of 4-hydroxybenzoate transport and degradation by benzoate: a new layer of regulatory control in the *Pseudomonas putida* β-ketoadipate pathway. J. Bacteriol. 177:7033-7040.

30. Orville, A. M., J. D. Lipscomb, and D. H. Ohlendorf. 1997. Crystal structures of substrate and substrate analog complexes of protocatechuate 3,4-dioxygenase: endogenous Fe super (3+) ligand displacement in response to substrate binding. Biochemistry 36:1005-10066.

31. Overbeek, R., M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev. 1999. The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. 96: 2896-2901.

32. Parales, R. E. and C. S. Harwood. 1993. Regulation of the *pcaIJ* genes for aromatic acid degradation in *Pseudomonas putida*. J. Bacteriol. 175:5829-5838.

33. Parke, D. 1996. Characterization of PcaQ, a LysR-type transcriptional activator required for catabolism of phenolic compounds in *Agrobacterium tumefaciens*. J. Bact. 178:266-272.

34. Parke, D. 1997. Acquisition, reorganization, and merger of genes: novel management of the β-ketoadipate pathway in *Agrobacterium tumefaciens*. FEMS Microbial. Lett. 146:3-12.

35. Parke, D., D. A. D'Argenio, and L. N. Ornston. 2000. Bacteia are not what they eat: that is why they are so diverse. J. Bacteriol. 182:257-263.

36. Parke, D, M. A Garcia, and L. N. Ornston. 2001. Cloning and genetic characterization of *dca* genes required for β-oxidation of straight-chain dicarboxylic acids in *Acinetobacter* sp. strain ADP1. Appl. Environ. Microbiol. 67:4817-4827.

37. Pathak, D., G. Ashley, and D. Ollis. 1991. Thioprotease-like active site found in the enzyme dienelactone hydrolase-localization using biochemical, genetic and structural tools. Prot-Struc, Func, Genet. 9:267-279.

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

39. Rothmel, R. K., T. L. Aldrich, J. E. Houghton, W. M. Coco, L. N. Ornston, and A. M. Chakrabarty. 1990. Nucleotide sequencing and characterization of *Pseudomonas putida catR*, a postive regulator of the *catBC* operon is a member of the LysR family. J. Bacteriol. 172:922-931.

40. Schell, M. A. 1993. Molecular biology of the LysR family of transcriptional regulators. Ann. Rev. Microbiol. 47:597-626.

41. Shanley, M. S., E. L. Neidle, R. E. Parales, and L. N. Ornston. 1986. Cloning and expression of *Acinetobacter calcoaceticus catBCDE* genes in *Pseudomonas putida* and *Escherichia coli*. J. Bacteriol. 165:557-563.

42. Sorokin, D. Y., and A. M. Lysenko. 1993. Heterotrophic bacteria from the Black Sea oxidizing reduced sulfur compounds to sulfate. Microbiol. *(English Translation Mikrobiologiya)* 62:1018-1031.

43. Stainer, R. Y. and J. L. Ingraham. 1954. Protocatechuic acid oxidase. J. Biol. Chem. 210:799-808.

44. Suzuki, M. T., M. S. Rapp , Z. W. Haimberger, H. Winfield, N. Adair, J. Str bel, and S. J. Giovannoni. 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Appl. Environ. Microbiol. 63:983-989.

45. Van Berkel, W., A. Westphal, K. Eschrich, M. Eppink, and A. Dekok. 1992. Substitution of Arg214 at the substrate-binding site of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. Eur. J. Biochim. 210:411-419.

46. van der Meer, J. R., A. C. Frijters, J. H. Leveau, R. I. Eggen, A. J. Zehnder, and W. M. de Vos. 1991. Characterization of the *Pseudomonas* sp. strain P51 gene *tcbR*, a LysR-type transcriptional activator of the *tcbCDEF* chlorocatechol oxidative operon, and analysis of the regulatory region. J. Bacteriol. 173:3700-3708.

47. Viale, A. M., H. Kobayashi, T. Akazawa, and S. Henikoff. 1991. *rcbR*, a gene coding for a member of the LysR family of transcriptional regulators, is located upstream of the expressed set of ribulose 1,5-bisphosphate carboxylase/oxygenase genes in the photosynthetic bacterium *Chromatium vinosum*. J. Bact. 173:5224-5229.

48. Young, D. A., D. Parke, D. A. D'Argenio, M. A. Smith and L. N. Ornston. 2001. Evolution of a catabolic pathway. ASM News. 67:362-369.

Table 4.1. Characterization of β-ketoadipate pathway genes identified in roseobacter isolates

Gene		*'Roseovarius nubinhibens'* ISM	SE45	*Sagittula stellata* E-37	*'Silicibacter pomeroyi'* DSS-3	Y3F	Y4I	*Sulfitobacter pontiacus*	*Sulfitobacter* sp. EE36
pcaO	% G+C	62.1	68.5	65.7	66.8	69.1	69.2	-	-
	length (bp)	227[a]	930	960	918	929	932	-	-
	stop codon[b]	-	TGA	TAA	TGA	TAG	TAG	-	-
	K/K+R[c]	-	0.16	0.22	0.17	0.12	0.12	-	-
	MW (kDa)[d]	-	33.7	34.4	33.1	32.9	33.1	-	-
pobA	% G+C	63.2	-	63.7	63.6	-	-	-	-
	length (bp)	1170	-	1181	1170	-	-	-	-
	stop codon	TGA	-	TGA	TAG	-	-	-	-
	K/K+R	0.23	-	0.28	0.19	-	-	-	-
	MW (kDa)	43379	-	44070	43790	-	-	-	-
pcaD	% G+C	71.9	68.7	-	-	-	-	-	-
	length (bp)	789	789	-	-	-	-	-	-
	stop codon	TGA	TGA	-	-	-	-	-	-
	K/K+R	0.05	0.15	-	-	-	-	-	-
	MW (kDa)	27.8	27.6	-	-	-	-	-	-
pcaC	% G+C	66.7	66.7	64.6	64.8	67.3	69.1	62.7	61.9
	length (bp)	411	387	393	378	395	396	381	381
	stop codon[b]	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TGA
	K/K+R[c]	0.13	0.33	0.38	0.27	0.2	0.2	0.46	0.46
	MW (kDa)[e]	15	14.1	14.4	13.8	14.3	14.3	13.7	13.7
pcaH	% G+C	62.7	63.5	66.7	63.0	63.7	64.8	60.1	60.7
	length (bp)	729	726	723	729	738	738	732	732
	stop codon	TGA	TGA	TGA	TGA	TGA	TGA	TGA	TGA
	K/K+R	0.19	0.3	0.2	0.21	0.21	0.23	0.29	0.28
	MW (kDa)	27.3	27.1	26.8	27.3	27.7	27.7	27.1	27
pcaG	% G+C	65.5	66.8	65.1	66.0	64.2	64.4	59.4	58.9
	length (bp)	621	693	602	621	618	621	606	606
	stop codon	TAG	TGA	TGA	TGA	TGA	TGA	TGA	TGA
	K/K+R	0.16	0.36	0.3	0.26	0.23	0.24	0.38	0.39
	MW (kDa)	22.6	25.3	21.9	22.6	22.9	22.8	22.1	22.2

[a]Partial gene sequence as obtained.
[b]Stop codon of deduced gene product.
[c]Lysine to lysine-plus-arginine ratio.
[d]Molecular weight of deduced gene product.

Figure 4.1. Schematic of the protocatechuate branch of the β-ketoadipate pathway. Gene designations are shown in italics. CoA, coenzyme A.

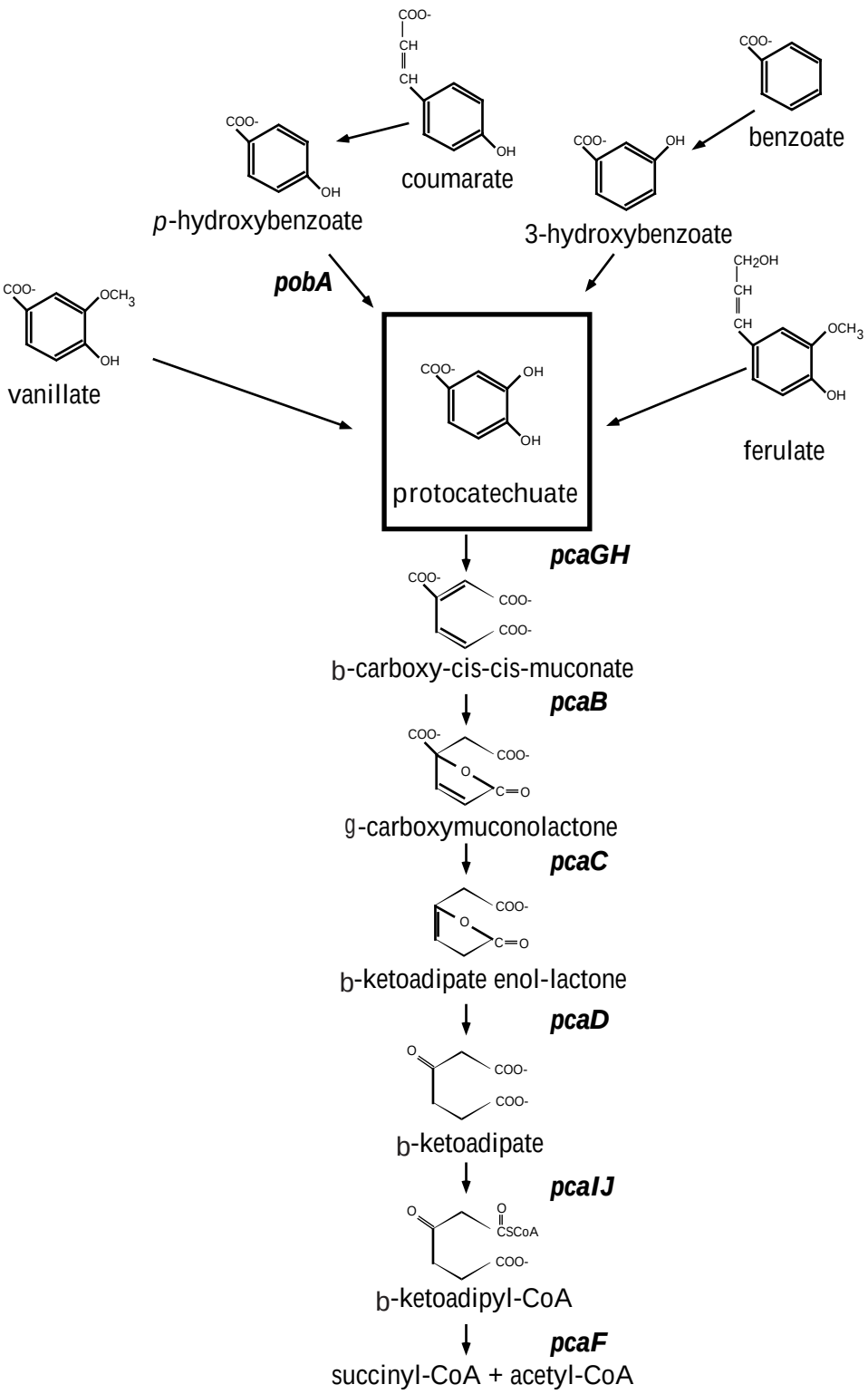

coumarate
benzoate
p-hydroxybenzoate
3-hydroxybenzoate
pobA
vanillate
ferulate
protocatechuate
pcaGH
b-carboxy-cis-cis-muconate
pcaB
g-carboxymuconolactone
pcaC
b-ketoadipate enol-lactone
pcaD
b-ketoadipate
pcaIJ
b-ketoadipyl-CoA
pcaF
succinyl-CoA + acetyl-CoA

Figure 4.2. Restriction map of the chromosomal fragments from isolate SE45 (A), '*R. nubinhibens*' ISM (B), '*S. pomeroyi*' DSS-3 (C), isolate Y4I (D), *Sulfitobacter* sp. EE-36 (E), and *S. pontiacus* (E). The locations of genes and their transcriptional directions are shown relative to selected restriction endonuclease recognition sites; A = *Apa*I, B = *Bam*HI, E = *Eco*RI, N = *Nsi*I, P = *Pst*I, X = *Xho*I. Horizontal lines indicate the DNA regions contained on recombinant plasmids, whose designations are shown above the corresponding line. Gene designations are as follows: SD = shikimate dehydrogenase, TetR = TetR-like transcriptional regulator, MP = methyl parathion degrading protein, ALD = aldehyde dehydrogenase, AD = alcohol dehydrogenase, ORF = open reading frame. *pca* and *pob* gene designations are provided in the text.

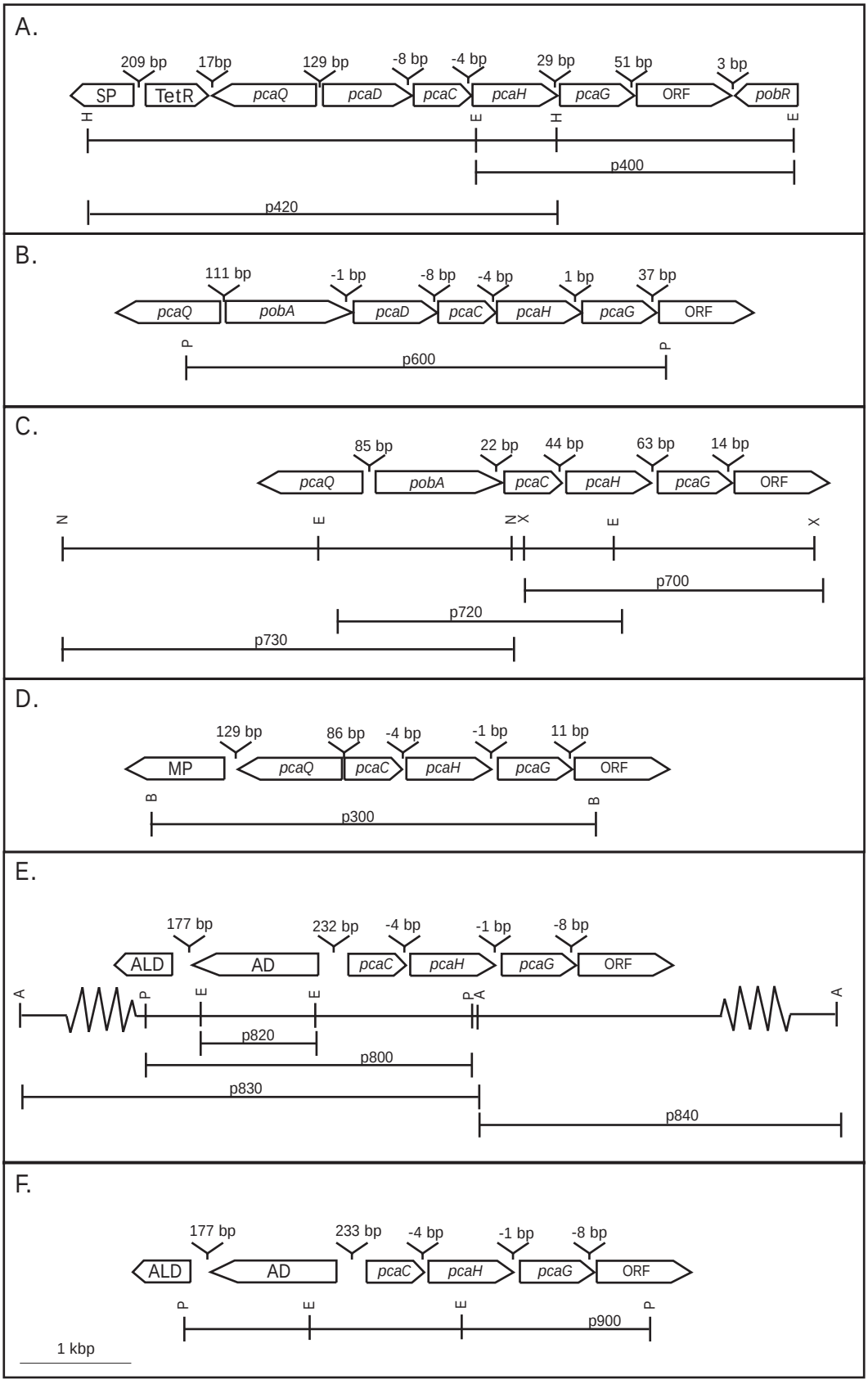

A.
209 bp
17bp
129 bp
-8 bp
-4 bp
29 bp
51 bp
3 bp
SP
TetR
pcaQ
pcaD
pcaC
pcaH
pcaG
ORF
pobR
H
E
H
E
p400
p420
B.
111 bp
-1 bp
-8 bp
-4 bp
1 bp
37 bp
pcaQ
pobA
pcaD
pcaC
pcaH
pcaG
ORF
P
P
p600
C.
85 bp
22 bp
44 bp
63 bp
14 bp
pcaQ
pobA
pcaC
pcaH
pcaG
ORF
N
E
N
X
E
X
p700
p720
p730
D.
129 bp
86 bp
-4 bp
-1 bp
11 bp
MP
pcaQ
pcaC
pcaH
pcaG
ORF
B
B
p300
E.
177 bp
232 bp
-4 bp
-1 bp
-8 bp
ALD
AD
pcaC
pcaH
pcaG
ORF
A
P
E
E
P
A
A
p820
p800
p830
p840
F.
177 bp
233 bp
-4 bp
-1 bp
-8 bp
ALD
AD
pcaC
pcaH
pcaG
ORF
P
E
E
P
p900
1 kbp

Figure 4.3. Phylogenetic tree of PcaHG protein sequences. The tree is based on the deduced amino acid encoded by the *pcaHG* genes and is unrooted, with PcaHG from *Rhodococcus opacus* 1CP (AF003947) as the outgroup. Bootstrap values greater than 50% are indicated at branch nodes. The scale bar indicates Kimura distances. Symbols located at the branch ends indicate phylogenetic affiliations of the bacteria from which protein sequences were obtained: black closed circles = roseobacter group organisms, grey closed circles = other α-Proteobacteria, open circle = γ-Proteobacteria, black square = Gram positive.

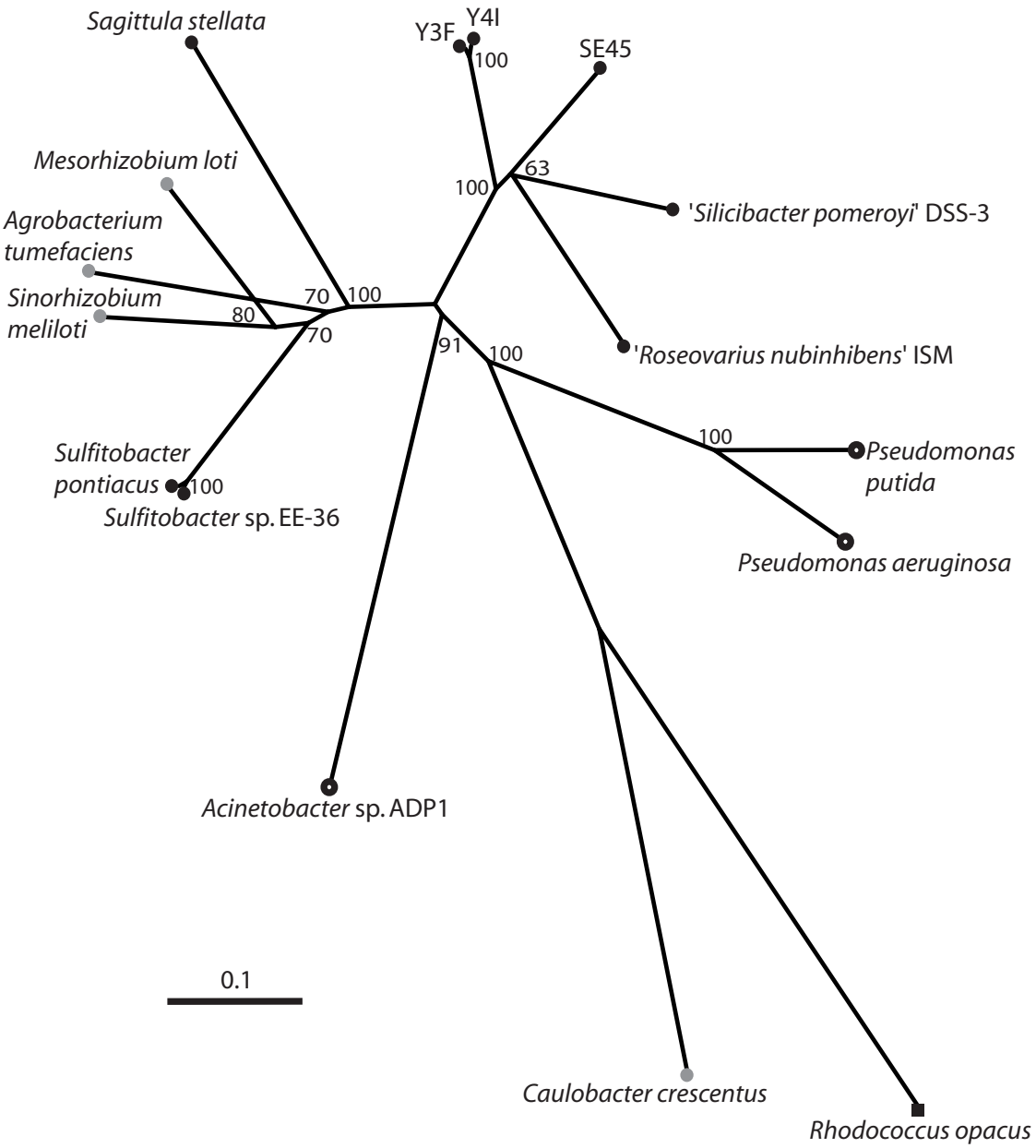

Sagittula stellata
Y3F
Y4I
100
SE45
Mesorhizobium loti
100
63
'Silicibacter pomeroyi' DSS-3
Agrobacterium tumefaciens
Sinorhizobium meliloti
70
100
80
70
91
100
'Roseovarius nubinhibens' ISM
100
Sulfitobacter pontiacus
100
Sulfitobacter sp. EE-36
Pseudomonas putida
Pseudomonas aeruginosa
Acinetobacter sp. ADP1
0.1
Caulobacter crescentus
Rhodococcus opacus

Figure 4.4. Phylogenetic tree of PcaC protein sequences. The tree is based on the deduced amino acid encoded by the *pcaC* genes and is unrooted. In *R. opacus* and *C. crescentus*, *pcaC* and *pcaD* have fused into a single gene (*pcaL*) encoding a protein demonstrating both PcaC and PcaD activities (Eulberg et al., 1999); the PcaC-like segment of PcaL was used in this analysis. The protein sequence from *R. opacus* 1CP (AF003947) was used as the outgroup. Bootstrap values greater than 50% are indicated at branch nodes. The scale bar indicates Kimura distances. Symbols located at the branch ends indicate phylogenetic affiliations of the bacteria from which protein sequences were obtained: black closed circles = roseobacter group organisms, grey closed circles = other α-Proteobacteria, open circle = γ-Proteobacteria, black square = Gram positive.

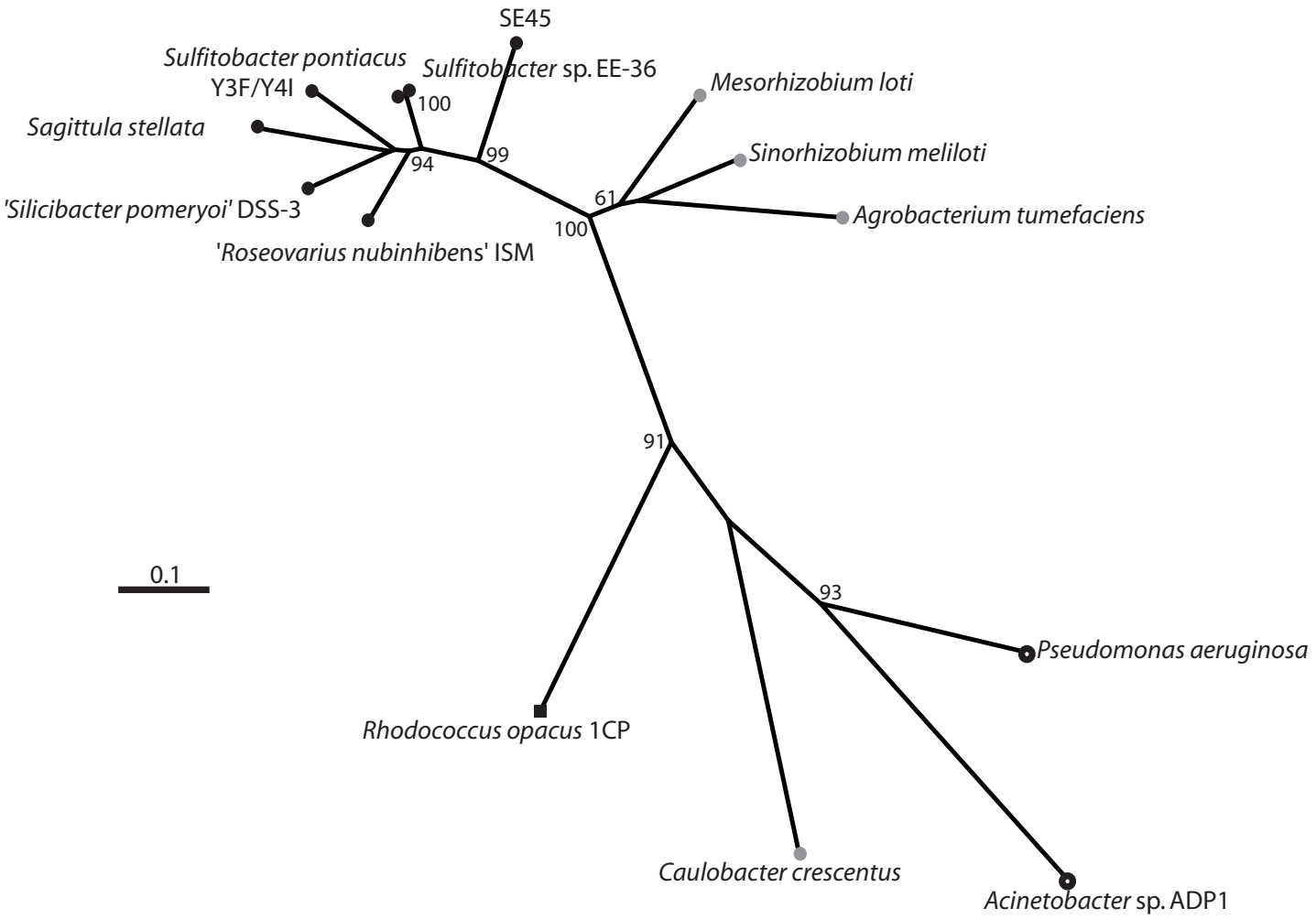

SE45
Sulfitobacter pontiacus Y3F/Y4I
Sulfitobacter sp. EE-36
Mesorhizobium loti
100
Sagittula stellata
99
94
Sinorhizobium meliloti
'Silicibacter pomeryoi' DSS-3
61
Agrobacterium tumefaciens
100
'Roseovarius nubinhibens' ISM
91
0.1
93
Pseudomonas aeruginosa
Rhodococcus opacus 1CP
Caulobacter crescentus
Acinetobacter sp. ADP1

Figure 4.5. Phylogenetic tree of 16S rRNA gene sequences. The tree is based on 1300 nucleotides (positions 111 to 1411; *E. coli* numbering system) with the gene from *Rhodococcus opacus* (AB032565) as the outgroup. Bootstrap values greater than 50% are indicated at branch nodes. The scale bar indicates Jukes-Cantor distances. *pca* gene organization of each isolate is shown; *pcaB* = β-carboxy-*cis*,*cis*-muconate lactonizing enzyme, *pcaC* = γ-carboxy-muconolactone decarboxylase, *pcaD* = β-ketoadipate enol-lactone hydrolase, *pcaHG* = protocatechuate 3,4-dioxygenase, *pcaF* = β-ketoadipyl CoA thiolase, *pcaIJ* = β-ketoadipate succinyl CoA transferase, *pcaK*, *pcaT* = transport proteins, *pcaL* = *pcaDC* gene fusion, *pcaQ* = LysR-type transcriptional regulator, *pobA* = *p*-hydroxybenzoate hydroxylase, *pobR*, *pcaU*, *pcaR* = IclR/AraC-type transcriptional regulator. Heavy double lines indicate >10 kbp separates the transcriptional units.

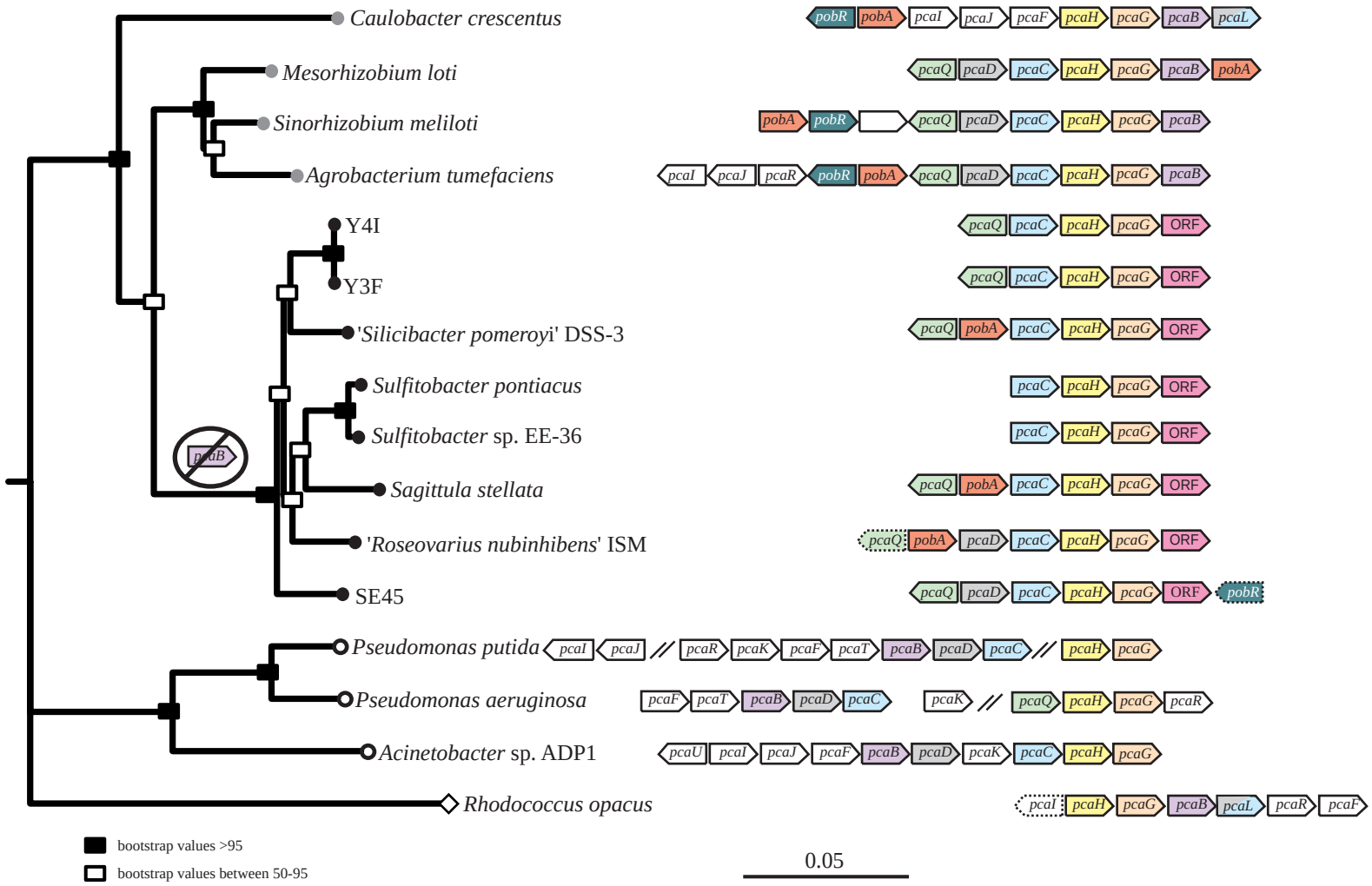
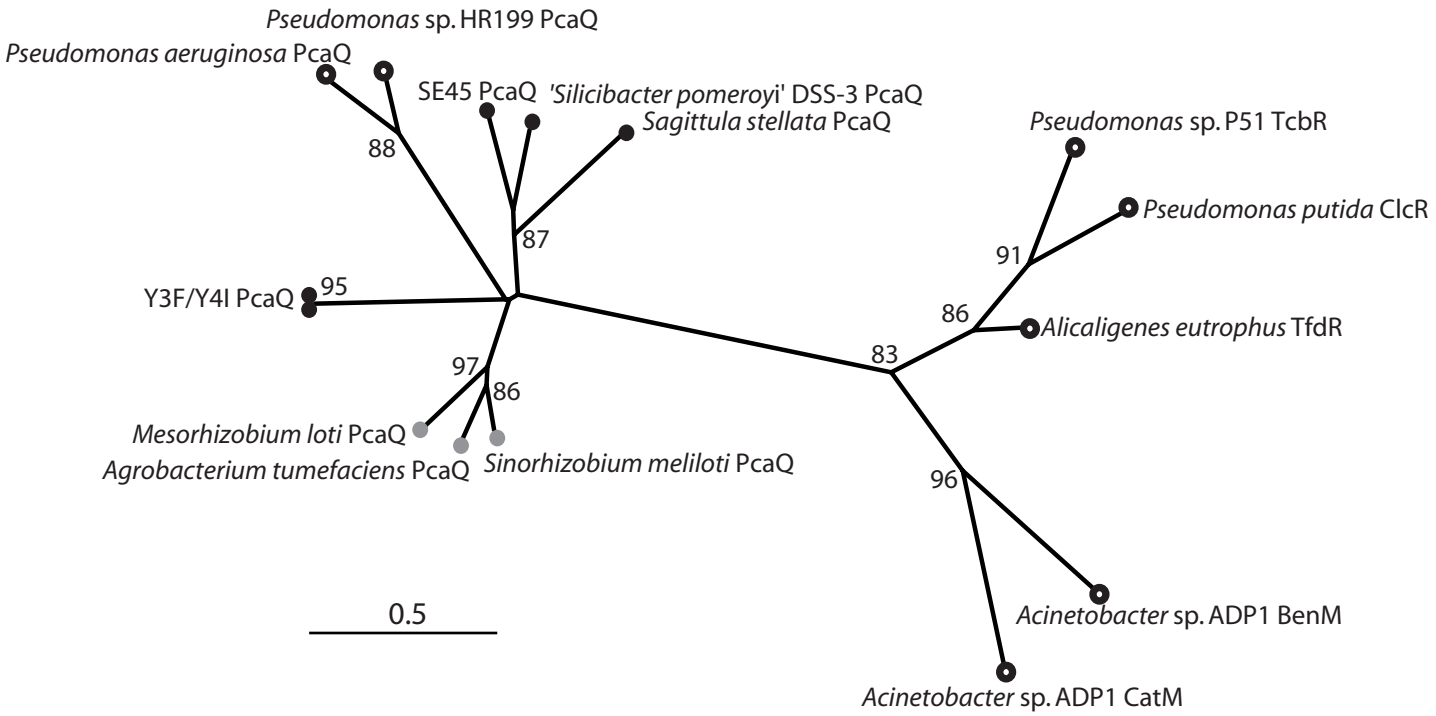

Caulobacter crescentus
Mesorhizobium loti
Sinorhizobium meliloti
Agrobacterium tumefaciens
Y4I
Y3F
'Silicibacter pomeroyi' DSS-3
Sulfitobacter pontiacus
Sulfitobacter sp. EE-36
Sagittula stellata
'Roseovarius nubinhibens' ISM
SE45
Pseudomonas putida
Pseudomonas aeruginosa
Acinetobacter sp. ADP1
Rhodococcus opacus
pcaB
pobR pobA pcaI pcaJ pcaF pcaH pcaG pcaB pcaL
pcaQ pcaD pcaC pcaH pcaG pcaB pobA
pobA pobR pcaQ pcaD pcaC pcaH pcaG pcaB
pcaI pcaJ pcaR pobR pobA pcaQ pcaD pcaC pcaH pcaG pcaB
pcaQ pcaC pcaH pcaG ORF
pcaQ pcaC pcaH pcaG ORF
pcaQ pobA pcaC pcaH pcaG ORF
pcaC pcaH pcaG ORF
pcaC pcaH pcaG ORF
pcaQ pobA pcaC pcaH pcaG ORF
pcaQ pobA pcaD pcaC pcaH pcaG ORF
pcaQ pcaD pcaC pcaH pcaG ORF pobR
pcaI pcaJ pcaR pcaK pcaF pcaT pcaB pcaD pcaC pcaH pcaG
pcaF pcaT pcaB pcaD pcaC pcaK pcaQ pcaH pcaG pcaR
pcaU pcaI pcaJ pcaF pcaB pcaD pcaK pcaC pcaH pcaG
pcaI pcaH pcaG pcaB pcaL pcaR pcaF
bootstrap values >95
bootstrap values between 50-95
0.05

Figure 4.6. Phylogenetic tree of LysR-type protein sequences. The tree is based on deduced amino acids encoded by the *pcaQ*, *tcbR* (van der Meer et al., 1991; P27102), *clcR* (Coco et al., 1993; Q05840), *tfdR* (Matrubutham et al., 1994; P10086), *benM*, and *catM* (Collier et al., 1998; AF009224) genes and is unrooted, with CatR (P206667; Rothmel et al., 1990) as the outgroup. Bootstrap values greater than 50% are indicated at branch nodes. The scale bar indicates Kimura distances. Symbols located at the branch ends indicate phylogenetic affiliations of the bacteria from which protein sequences were obtained: black closed circles = roseobacter group organisms, grey closed circles = other α-Proteobacteria, open circle = γ-Proteobacteria, black square = Gram positive.


Pseudomonas sp. HR199 PcaQ
Pseudomonas aeruginosa PcaQ
SE45 PcaQ
'Silicibacter pomeroyi' DSS-3 PcaQ
Sagittula stellata PcaQ
88
87
Y3F/Y4I PcaQ
95
97
86
Mesorhizobium loti PcaQ
Agrobacterium tumefaciens PcaQ
Sinorhizobium meliloti PcaQ
0.5
Pseudomonas sp. P51 TcbR
Pseudomonas putida ClcR
91
86
Alicaligenes eutrophus TfdR
83
96
Acinetobacter sp. ADP1 BenM
Acinetobacter sp. ADP1 CatM

CHAPTER 5

SUMMARY

The intent of these studies was to identify and characterize the mechanisms by which marine bacteria, and members of the roseobacter lineage in particular, degrade naturally-occurring aromatic compounds. The relevance of these investigations is illustrated by: [1] the abundance of aromatic compounds in coastal systems (5, 6) and [2] the predominance of members of the roseobacter lineage in these environments (1, 3). Since this lineage is one of the few dominant marine clades that is amenable to culturing, investigations of its physiology and genetics are particularly important to our understanding of the ecology of natural bacterial communities (2).

The degradation of aromatic monomers that arise during the decay of lignin and other vascular plant components is an important process in the biogeochemical cycling of organic matter. As this plant material decays, the lignin component is converted to aromatic monomers such as benzoate, catechol, *p*-hydroxybenzoate, cinnamate, *p*-coumarate, vanillate, ferulate, quinate, and shikimate (7). In terrestrial systems, these monomers are typically converted to protocatechuate and degraded via the β-ketoadipate pathway (6). Members of the marine roseobacter lineage appear to degrade lignin-related during growth on *p*-hydroxybenzoate in several of the isolates and that activities in cell extracts were comparable to those found for soil bacteria (Table 2.3). Expression of

roseobacter *pcaHG* genes in *E. coli*, coupled with the identification of additional functionally related genes in adjacent positions (Table 2.2; Fig. 2.5; Fig. 4.5), provides further evidence that this pathway is functional in this ecologically important marine lineage.

An investigation of the presence and diversity of the *pcaH* gene in natural bacterial communities supports the proposal that members of the roseobacter lineage play an important role in the degradation of aromatic compounds in coastal ecosystems. In this study, a collection of cultivated organisms was instrumental in identifying over half (58%) of the *pcaH* genes PCR amplified from natural and enriched communities. The vast majority of these genes could be assigned to a roseobacter group member based on sequence identity (Table 3.3). Identification of roseobacter-like *pcaH* genes in enrichments of the natural community with fused-ring and hydroxy-, methyl-, and amino-substituted structures (Fig. 3.2) complements the growth assays carried out on cultivated organisms (Table 2.1) and indicates these bacteria are capable of metabolizing a variety of ring structures. The number of highly similar *pcaH* sequences obtained in this study also raises questions about the physiological implications of gene microdiversity in natural bacterial assemblages.

The genes involved in the protocatechuate branch of the β-ketoadipate pathway in eight roseobacter isolates demonstrates the dynamic nature of gene organization (Fig. 4.5). The observed genetic arrangements suggest at least five distinct regulatory schemes for the catabolism of *p*-hydroxybenzoate and protocatechuate. The unique localization of the *pobA* gene within the *pca* cluster of several isolates indicates *p*-hydroxybenzoate is an

important substrate feeding into this pathway in roseobacter group members. There is no strong evidence for a recent lateral transfer event for any of the *pobA* or *pca* genes examined (Fig. 4.3, 4.4, 4.6). The extent of genetic rearrangements and sequence divergence points to an ancient acquisition of this pathway in this phylogenetically broad clade and suggests an important role for this pathway in the metabolic repertoire of these organisms. This study also illustrates the significant role of chromosomal rearrangements in the evolution of this pathway and its contributions to microbial diversification. Whether variations in protein sequences or organization of genes of this pathway reflect physiological differences that contribute to the overall success of these organisms in natural populations is difficult to address experimentally. However, a certain degree of specialization at the whole-organism level would be expected in order for individual populations to compete effectively.

References

1. Dang, H. and C. R. Lovell. 2000. Bacterial primary colonization and early succession on surfaces in marine waters as determined by amplified rRNA gene restriction analysis and sequence analysis 16S rRNA genes. Appl. Envir. Microbiol. 66: 467-475.

2. Giovannoni, S., and M. Rappé. 2000. Evolution, diversity, and molecular ecology of marine prokaryotes, p47-84. *In* D. Kirchman (ed.), Microbial ecology of the oceans. Wiley Interscience, New York, N. Y.

3. González, J. M., and M. A. Moran. 1997. Numerical dominance of a group of marine bacteria in the α-subclass of the class Proteobacteria in coastal seawater. Appl. Environ. Microbiol. 63:4237-4242.

4. Harwood, C. S., and R. E. Parales. 1996. The β-ketoadipate pathway and the biology of self-identity. Ann. Rev. Microbiol. 50:553-590.

5. Moran, M. A. and R. E. Hodson. 1989. Formation and bacterial utilization of dissolved organic carbon derived from detrital lignocellulose. Limnol. Oceanogr. 34:1034-1047.

6. Moran, M. A., and R. E. Hodson. 1994. Dissolved humic substances of vascular plant origin in a coastal marine environment. Limnol. Oceanogr. 39:762-771.

7. Subba Rao, P. V., A. M. D. Nambudiri and J. V. Ghat. 1971. Microbial degradation of phenylpropaoid compounds. J. Sci. Indust. Res. 30:663-679.