Genome reduction and evolution in an obligate luminous symbiont.

by

Tory A. Hendry

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2012

Doctoral Committee:

Professor Paul V. Dunlap, Chair
Associate Professor Yin-Long Qiu
Assistant Professor Gregory James Dick
Assistant Professor Patrick D. Schloss

Illustration by Patricia Beals

Copyright Tory A. Hendry
2012

To my grandfather for giving me a love of nature,
my mother for giving me the courage to try,
my husband for giving constant encouragement,
and my son for being the best thing I have done.

ACKNOWLEDGEMENTS

Support and advice from many people have guided me through my graduate career and made this work possible. First and foremost I thank my committee members Paul Dunlap, Greg Dick, Pat Schloss and Yin-Long Qiu for their invaluable assistance and constructive comments. I would especially like to thank my advisor, Paul, for his time, advice, and support. I have been privileged to learn how to be a careful and thoughtful researcher under his guidance, while enjoying the freedom to follow the research directions that inspired me. I am also grateful to lab members past and present, particularly Jennifer Ast, Henryk Urbanczyk, Alison Gould, and Beth Pringle, for their thoughts and assistance. A special thank you goes to Katherine Dougan for all of her meticulous help.

Numerous people went out of their way in helping me obtain samples. I thank Jay Hemdal from the Toledo Zoo and Seth Wolters from the Steinhart Aquarium for allowing me access to their flashlight fish exhibits. I am grateful to Richard Rosenblatt and the Scripps Institution of Oceanography for the loan of flashlight fish specimens. I greatly appreciate the help of Quality Marine Suppliers, Grant Norton from Sustainable Reef Suppliers, Brad Remmer from Sea Dwelling Creatures, and Fish Doctors Aquarium in finding, collecting, and transporting fish. And I would especially like to thank Margo Haygood for the generous gift of DNA.

My graduate career has been greatly enhanced by discussions with, and support from, my amazing cohort and other graduate students in the department. Thanks to the best big sib of all, Wendy Grus, for taking me under her wing. I can not thank Michael Sheehan enough for reading nearly every word I've written and hearing every theory I've come up with. There is also no way to sufficiently thank Susanna Messinger and Rachel Vannette for the gift of sanity, which they gave to me over many cups of coffee. I greatly appreciate the sympathetic ear of Abby Woodroffe, who was there through many rocky starts. And an enormous thank you to my pseudosister, Flynn Boonstra, whose love and support were even more meaningful for having been given long distance.

I have received unlimited help from many people in EEB, especially members of the office staff, Julia Eussen, Jane Sullivan, Cindy Carl, Sonja Botes, and Amber Stadler. I have been funded personally by the departmental Edward's fellowship and this work was funded by departmental Block Grant awards and Rackham graduate student research grants.

TABLE OF CONTENTS

Dedication…….. ii

Acknowledgements…… iii

List of Figures…… vi

List of Tables……. viii

List of Appendices……. ix

Abstract…….. x

Chapter

I. Introduction……. 1

II. The uncultured luminous symbiont of *Anomalops katoptron* represents a new bacterial genus……... 12

III. Genome reduction and host dependence in a luminous symbiont……... 44

IV. The obligate luminous symbionts of flashlight fish are specific to host genera……... 67

V. Extremely low genetic diversity and population dynamics in an obligate luminous symbiont..…….. 84

VI. Genomic divergence between obligate luminous symbiont species…… 117

VII. Conclusion…….. 138

Appendices……. 147

LIST OF FIGURES

Figure 2.1. Maximum likelihood trees from housekeeping genes (16S rRNA gene, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, and *topA*)…… 31

Figure 2.2. Maximum likelihood trees from the *lux* operon……..……..... 32

Figure 2.3. A. The structure of the *lux* operon and flanking regions in the *A. katoptron* symbionts and representative members of *Vibrionaceae*. B. Alignment of the regulatory region upstream of the *lux* operon in the *A. katopron* symbiont and *Vibrio harveyi.*…….. 33

Figure 2.4. Maximum likelihood tree of '*Ca*. Photodesmus katoptron' *luxR* and *Vibrionaceae* homologs……. 34

Fig. 3.1. Graph of genome size versus number of genes……... 57

Fig. 3.2. Chromosome comparisons between '*Ca.* Photodesmus katoptron' and *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116)……... 58

Fig. 3.3. Phylogenetic tree with genome sizes of '*Ca*. Photodesmus katoptron' and relatives…….. 59

Fig. 3.4. Genes in TIGRfam functional categories for '*Ca.* Photodesmus katoptron' and *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116)…….. 60

Fig. 3.5.Amino acid synthesis pathways showing gene loss in '*Ca.* Photodesmus katoptron.'…….. 61

Fig. 3.6. Genomic hierarchical clustering dendogram showing that '*Ca.* Photodesmus katoptron' is more similar in gene content to distantly related obligate symbionts than it is to free-living close relatives……. 62

Fig. 4.1. Maximum likelihood trees from housekeeping genes (16S rRNA gene, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, and *topA*)…… 78

Fig. 4.2. Maximum likelihood trees from *lux* operon (*luxCDABEG*)…… 79

Fig. 4.3. Divergence between strains from different host genera for the symbionts *Aliivibrio fischeri* and '*Ca*. Photodesmus spp.'……. 80

Fig. 4.4. Cladograms comparing anomalopid host phylogeny and anomalopid symbiont phylogeny……... 81

Fig. 5.1. Diversity estimates for the anomalopid symbionts and various facultatively host-associated relatives…… 105

Fig. 5.2. Diversity estimates for anomalopid symbionts, free-living luminous symbionts, and genetically monomorphic obligately host-associated bacteria……. 106

Fig. 5.3. Percentage of sites polymorphic for a given polymorphism index (number polymorphic reads at a site) in '*Ca*. Photodesmus katoptron.'……... 107

Fig. 5.4. Percentage of sites polymorphic for a given polymorphism index (number polymorphic reads at a site) in '*Ca*. Photodesmus blepharus' Ppalp…….. 108

Fig. 5.5. '*Candidatus* Photodesmus katoptron' host distribution, sample sites and genetic divergence……. 109

Fig. 6.1. A. Alignments of syntenic contigs from 'Ca. Photodesmus katoptron' (Pk) and 'Ca. Photodesmus blepharus' Ppalp (Pp) and B. Alignments of chromosome I and chromosome II from *Vibrio cholerae* N16961, *Vibrio parahaemolyticus* RIMD 2210633, and *Vibrio vulnificus* CMCP6……... 133

Fig. 6.2. Genes shared by '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp arranged by percent identity between the anomalopid symbionts compared to percent identity to relatives……... 134

LIST OF TABLES

Table 2.1. GenBank accession numbers and strain designations for all taxa in the housekeeping gene analysis……... 35

Table 2.2. GenBank accession numbers and strain designations for taxa in the *lux* gene analysis…….. 38

Table 3.1. Genome characteristics that vary with lifestyle compared to '*Ca.* Photodesmus katoptron'……. 63

Table 3.2. Contig and assembly information showing size (bp) and read coverage depth of contigs by category…….. 63

Table 5.1. Polymorphism values for anomalopid symbionts. Single nucleotide polymorphisms per kilobase are shown for chromosome I, chromosome II, and the plasmid for each species……. 110

Table 5.2. Polymorphism values for free-living, host-associated *Vibrionaceae* species…… 111

Table 5.3. Rates of nonsynonymous and synonymous substitutions for anomalopid symbionts, facultatively host-associated relatives and *B. aphidicola*… 112

Table 5.4. Genetic divergence between symbiont populations for anomalopid symbionts and free-living luminous symbionts……. 112

Table 5.5. Genes used to investigate divergence between populations of '*Ca.* Photodesmus katoptron.' Numbers of polymorphisms in each population are shown……. 113

Table 6.1. Genome content information for the '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp genomes……... 135

Table 6.2. Comparison of how genes unique to one symbiont species were lost in the other species. Absolute numbers and percentages are shown…….. 135

LIST OF APPENDICES

Appendix 1. List of conserved genes present in the '*Ca*. Photodesmus katoptron' genome……... 147

Appendix 2. List of genes in selected TIGRfam categories found in the '*Ca*. Photodesmus katoptron' genome……... 151

Appendix 3. *Vibrionaceae* strains used in diversity estimates…….. 172

Appendix 4. List of unique genes found in the '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp genomes……... 176

ABSTRACT

Genome reduction and evolution in an obligate luminous symbiont.

by

Tory A. Hendry

Chair: Paul V. Dunlap

The luminous bacteria symbiotic with anomalopid flashlight fish are unusual compared to other luminous symbionts; they cannot be cultured outside the host and are thought to be obligately dependent on their hosts. The aims of this dissertation are to test if anomalopid symbionts are obligately dependent on hosts for growth and to compare the evolutionary history of anomalopid symbionts to symbionts with varying lifestyles. To do this, I present the sequenced genomes of two species of anomalopid symbionts, which I show to have specific interactions with different host species. I find that anomalopid symbionts are obligately dependent on their hosts and that this interaction has had a large affect on their evolution. While anomalopid symbionts are closely related to facultative luminous symbionts, they have multiple characteristics in common with obligate intracellular symbionts. Anomalopid symbionts have reduced genomes due to gene loss.

Gene loss has lead to their inability to synthesize most amino acids and to utilize carbon/energy sources besides glucose. The observation of these metabolic deficiencies supports the obligate dependence of anomalopid symbionts on their hosts. I also find patterns consistent with anomalopid symbionts experiencing high levels of genetic drift. For instance, anomalopid symbionts are evolving at a faster rate than free-living relatives. Additionally, anomalopid symbionts are genetically monomorphic, even across a wide geographic range, and have an excess of rare substitutions. Furthermore, one symbiont species displays a high number of nonsynonymous substitutions. These indications of high genetic drift are consistent with the symbionts undergoing population bottlenecks during transfers between host generations. I also find that the two symbiont species are very similar in gene content but highly divergent at both the nucleotide and amino acid level. This finding is comparable to trends seen in obligate intracellular symbionts. The genomic patterns common to obligate intracellular symbionts are caused by their being intracellular and vertically transmitted, thus it is surprising that the extracellular, environmentally acquired symbionts of flashlight fish would also show so many of these trends. That these commonalities exist between symbionts with different ecologies highlights the importance of host interactions in determining the evolutionary history of symbionts.

CHAPTER I

Introduction

Symbiosis between bacteria and animals is ubiquitous and can have significant evolutionary consequences for both partners. The effect of symbiosis on animals has been well documented; partnerships with bacteria can allow animal lineages to occupy new niches and diversify (Moran, 2006, 2007; Sapp, 2004). The affects of symbiosis on bacteria are less clear. While some bacterial lineages have been greatly impacted by host associations, most remain under-studied. The best examples of bacterial evolution through symbiosis are obligate intracellular bacteria, which typically show extreme evolutionary change compared to free-living relatives. However, many symbiotic bacteria are extracellular and may inhabit multiple niches, including being host-associated. Consequently, bacteria will vary in how dependent on a given host they are and possibly how strongly they are affected by interactions with host. The evolutionary consequences of the full range of symbiotic lifestyles are not known, but elucidating these patterns is important for understanding bacterial evolution and ecology, as well as symbiosis.

This dissertation focuses on understudied symbionts with a potentially unusual ecology. The luminous symbionts of anomalopid flashlight fish (Beryciformes: Anomalopidae) are extracellular and environmentally acquired like other luminous symbionts, but, because they cannot be cultured outside of the host, researchers have hypothesized that they might be obligately dependent on the host (Hastings and Nealson, 1981; Haygood, 1993; Haygood and Distel, 1993; Herring and Morin, 1978). This is not typical for luminous symbionts (Dunlap, 2009). If the ecology of anomalopid symbionts

is different from other luminous symbionts, they may have very different evolutionary patterns. The main goals of this dissertation are to 1) place anomalopid symbionts in a phylogenetic context, 2) determine if anomalopid symbionts are obligately dependent on hosts for growth, and 3) explore the patterns of genome evolution in anomalopid symbionts, when compared to other symbionts with different lifestyles.

I predict that if anomalopid symbionts are obligately dependent on their hosts, they will show more evolutionary trends in common with distantly related obligate symbionts than with closely related facultative symbionts, demonstrating the importance of host interaction on bacteria. I outline my specific predictions below. However, since anomalopid symbionts are phylogenetically independent from other obligate symbionts and differ in aspects of the host interaction, I predict that deviations from common patterns in obligate symbionts will occur. These deviations can be used to infer the affect of different factors on symbiont evolution. For instance, since anomalopid symbionts are environmentally acquired they may not be codiverging with their hosts, whereas intracellular symbionts do codiverge with hosts. If anomalopid symbionts were shown to not be codiverging, this would highlight the importance of an intracellular lifestyle in causing codivergence.

To understand patterns of genome evolution in anomalopid symbionts, this dissertation makes comparisons to two model symbiont systems, bioluminescent symbiosis and the endosymbionts of insects. Here I will provide some general background information on each system and outline the main questions addressed in the dissertation.

Bioluminescent symbiosis

Bioluminescent symbionts are well-studied facultatively symbiotic bacteria. Luminous symbionts form associations with various groups of fish and squid. These interactions have likely evolved independently many times, presumably due to the large benefit conferred to the host by light production Hosts typically use bacterially produced light in a variety of ecological contexts such as obtaining food, avoiding predators, and communicating. For instance, anomalopid flashlight fish are strictly nocturnal and are thought to use light for the purposes of attracting phototrophic plankton, confusing predators, and signaling to mates (Harvey, 1922; Herring and Morin, 1978; Morin et al., 1975). Symbiotic bacteria are housed extracellularly in specialized structures called light organs. These structures are presumed to provide nutrients to the bacteria, though in most cases it is not known which nutrients they are (Haygood, 1993). Anomalopid light organs are located one under each eye and are composed of tubules that are densely packed with bacteria (Bassot, 1966; Kessel, 1977). Pores on the surface of the light organ allow for regular release of bacteria (Haygood et al., 1984). Anomalopids, like other hosts of luminous bacteria, appear to acquire their symbionts from the environment during development (Haygood, 1993).

All known luminous symbionts are members of the marine *Gammaproteobacteria* family *Vibrionaceae*, which contains most luminous bacteria. Members of *Vibrionaceae* are widespread and common in marine environments. They are often associated with host animals and include pathogens, symbionts, commensals, and saprophytes (Dunlap, 2009; Reen et al., 2006; Thompson et al., 2004). Many *Vibrionaceae* genomes have been sequenced and they display genomic signatures of adaptation to diverse and frequently

changing habitats (Lauro et al., 2009; Reen et al., 2006). In keeping with this trend, most luminous symbionts are facultatively symbiotic and inhabit many habitats in addition to being symbiotic (Dunlap, 2009; Dunlap et al., 2007). Possible exceptions to this pattern are the symbionts of anomalopid flashlight fish and deep-sea anglerfish (Ceratioidei) (Haygood and Distel, 1993). These two groups of bacteria cannot be cultured outside the host, suggesting that they are obligate. While some possible examples of bacterial adaptation to the host environment have been found (Wollenberg and Ruby, 2012), it is not known how much symbiotic interactions have impacted evolution of facultative luminous symbionts.

If anomalopid symbionts are not obligately dependent on their hosts, but instead live in a variety of habitats and are simply difficult to culture, I expect their genomes to resemble those of other luminous symbionts. Specifically, they should have large genomes adapted to diverse conditions. For instance, the genomes of facultative symbionts should contain a variety of energy metabolism genes for utilizing diverse carbon/energy sources. I also expect that if they are facultatively symbiotic, anomalopid symbionts should maintain large amounts of genetic diversity in keeping with having relatively large population sizes and adapting to multiple habitats. In this dissertation I will compare the gene content and genetic diversity of multiple anomalopid symbiont genomes to sequenced genomes of facultatively symbiotic luminous bacteria to test these predictions. I will do this within a phylogenetic context after identifying the closest relatives of anomalopid symbionts.

Endosymbionts of insects

One of the best-studied groups of obligate symbionts is the endosymbionts of insects. Many insects harbor obligate intracellular bacteria, which are necessary for host function (Gil et al., 2004; McCutcheon et al., 2009; Moran, 2007; Wernegreen, 2002). Typically, the bacteria provide metabolites that cannot be synthesized by the host. These bacteria have lost the ability to survive outside the host and are vertically transmitted to new host generations. Several evolutionary patterns are found in phylogenetically independent endosymbiont lineages: gene loss and genome reduction, species specificity with hosts, a bias towards AT nucleotide substitutions, and an increased evolutionary rate (Fry and Wernegreen, 2005; Gil et al., 2004; McCutcheon and Moran, 2012; Moran, 1996; Moran et al., 2008; Wernegreen, 2002; Wernegreen and Funk, 2004; Wernegreen and Moran, 1999; Woolfit and Bromham, 2003). These are thought to result from high levels of genetic drift caused by vertical transmission, population bottlenecks between host generations, small population sizes, and possibly relaxed selection and high mutation rates. An additional trend, low intraspecific genetic diversity, has been reported for one genus of endosymbiont and some intracellular pathogens (Abbot and Moran, 2002; Achtman, 2008; Funk et al., 2001).

In addition to finding the genomic patterns described above, comparisons of multiple related species of obligate endosymbionts have uncovered much about the process of genome reduction. Early on in their evolution reduced genomes are highly dynamic, undergoing frequent large-scale gene loss, genomic rearrangements, and high rates of nucleotide substitutions. However, after the loss of many genes and repetitive elements, the rate of change in reduced genomes slows and these genomes become

relatively static. Using multiple species of anomalopid symbionts, I will test for these evolutionary trends common to endosymbionts. These tests will determine if anomalopid symbionts are obligately dependent on their hosts. They will also inform our understanding of how host interactions affect the evolution of symbionts.

The analyses are divided into five chapters described below.

Chapter II

In this chapter I present phylogenetic analyses on anomalopid symbionts. I seek to confirm the placement of anomalopid symbionts in the family *Vibrionaceae* and to identify their closest relatives. To do this I performed multilocus phylogenetic analysis using both conserved housekeeping genes and phylogenetically informative luminescence genes. Placing anomalopid symbionts in a phylogenetic context allowed me to perform phylogenetically informed comparative analyses testing for trends common to obligate symbionts.

Chapter III

Here I present genome comparisons between one species of anomalopid symbiont and free-living relatives to determine if the anomalopid symbiont is obligately dependent on the host for growth. I sequenced and assembled the complete genome of the anomalopid symbiont from high-throughput sequencing data. From this annotated genome, genes were assigned to functional categories using the TIGRfams system. The numbers and types of genes in each functional category were compared to the genomes of

two free-living *Vibrionaceae* species, *Aliivibrio fischeri* and *Vibrio campbellii*. I assessed the anomalopid symbiont for gene loss, specifically the loss of necessary metabolic pathways, as indications of host dependence. I also used the loss of pathways to infer which nutrients the host might be providing to the bacteria.

To infer the ecology of the anomalopid symbiont, I compared the gene content of the anomalopid symbiont to bacteria representing a variety of ecological lifestyles. Free-living bacteria, both symbiotic and not, and obligate symbionts were included in the analysis. Additionally, the comparison species represented both copiotrophs (adapted to high nutrient, highly dynamic conditions) and oligotrophs (adapted to low nutrient, stable conditions). I hypothesized that bacteria with similar lifestyles would have similar gene content and thus the gene content of the anomalopid symbiont could be used to predict its ecology.

Chapter IV

If anomalopid symbionts are obligate they may be more likely than other luminous symbionts to display high levels of host species specificity. To address host specificity, I performed phylogenetic analyses using multilocus sequence data from the symbionts of three host species. These hosts included multiple genera and multiple species within a genus to determine the scale of specificity. I used phylogenetic distance between symbionts as well as divergence of 16S rRNA gene sequences from each symbiont to delineate species boundaries. If there is a high level of specificity I expected that different host species would harbor different species of symbionts.

Chapter V

Host associations may have a large influence on the population dynamics of symbionts and therefore on their evolution. Obligate symbionts are influenced by host associations by being restricted to the host environment and thus highly subject to genetic drift. In addition, some obligate symbiont populations are affected by the dynamics of host populations (Abbot and Moran, 2002; Funk et al., 2001). It is not known how broadly these factors, host restriction and host dynamics, affect bacterial populations. To test for patterns of genetic diversity caused by host restriction and host population dynamics, I compared patterns of genetic diversity in two anomalopid symbiont species to patterns in free-living relatives and unrelated obligate symbionts. I utilized whole genome high-throughput sequencing data to find polymorphisms in anomalopid symbionts from multiple individuals. I determined how wide spread versus rare these substitutions were and the ratio of nonsynonymous to synonymous substitutions. I predicted that if anomalopid symbionts undergo repeated population bottlenecks they should lose diversity and therefore have lower diversity levels than free-living relatives, as well as excess rare and nonsynonymous substitutions. Additionally, I used sequence data from three different populations of one anomalopid symbiont species to investigate genetic divergence across a geographic range.

Chapter VI

Comparisons of multiple strains of obligate symbionts have uncovered patterns in how the genomes of obligate symbionts change over time. To determine if these trends hold true in anomalopid symbionts I compared the genomes of two species of

anomalopid symbionts. I tested for patterns found in other obligate symbionts, including high rates of change soon after becoming obligate and genomic stasis later in evolutionary history. I also used differences in the gene content of the anomalopid symbionts to develop hypothesis about the genetic basis for their different ecologies.

References

Abbot, P., Moran, N.A., 2002. Extremely low levels of genetic polymorphism in endosymbionts (*Buchnera*) of aphids (*Pemphigus*). Mol. Ecol. 11, 2649-2660.

Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. 62, 53-70.

Bassot, J.-M., 1966. On the comparative morphology of some luminous organs. In: Johnson, F.H., Haneda, Y. (Eds.), Bioluminescence in progress. Princeton University, Princeton, NJ.

Dunlap, P.V., 2009. Bioluminescence, microbial. In: Schaechter, M. (Ed.), Encyclopedia of microbiology. Elsevier, Oxford, pp. 45-61.

Dunlap, P.V., Ast, J.C., Kimura, S., Fukui, A., Yoshino, T., Endo, H., 2007. Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbioses. Cladistics 23, 507-532.

Fry, A.J., Wernegreen, J.J., 2005. The roles of positive and negative selection in the molecular evolution of insect endosymbionts. Gene 355, 1-10.

Funk, D.J., Wernegreen, J.J., Moran, N.A., 2001. Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-buchnera association. Genetics 157, 477-489.

Gil, R., Latorre, A., Moya, A., 2004. Bacterial endosymbionts of insects: insights from comparative genomics. Environmental Microbiology 6, 1109-1122.

Harvey, E.N., 1922. The production of light by the fishes *Photoblepharon* and *Anomalops*. Carnegie Institute of Washington, Washington, D.C.

Hastings, J.W., Nealson, K., 1981. The symbiotic luminous bacteria. In: Starr, M.P., Stolp, H., Truper, H.G., Balows, A., Schlegel, H.G. (Eds.), The Prokaryotes: a Handbook on Habitats, Isolation, and Identification of Bacteria. Springer-Verlag, Berlin, pp. 1332-1345.

Haygood, M.G., 1993. Light organ symbioses in fishes. Crit. Rev. Microbiol. 19, 191-216.

Haygood, M.G., Distel, D.L., 1993. Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. Nature 363, 154-156.

Haygood, M.G., Tebo, B.M., Nealson, K.H., 1984. Luminous bacteria of a monocentrid fish (*Monocentris japonicus*) and 2 anomalopid fishes (*Photoblepharon palpebratus* and *Kryptophanaron alfredi*) - populations sizes and growth within the light organs, and rates of release into the seawater. Marine Biology 78, 249-254.

Herring, P.J., Morin, J.G., 1978. Bioluminescence in fishes. In: Herring, P.J. (Ed.), Bioluminescence in Action. Academic Press, London.

Kessel, M., 1977. The ultrastructure of the relationship between the luminous organ of the teleost fish *Photoblepharon palpebratus* and its symbiotic bacteria. Cytobiologie 15, 145-158.

Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., DeMaere, M.Z., Ting, L., Ertan, H., Johnson, J., Ferriera, S., Lapidus, A., Anderson, I., Kyrpides, N., Munk, A.C., Detter, C., Han, C.S., Brown, M.V., Robb, F.T., Kjelleberg, S., Cavicchioli, R., 2009. The genomic basis of trophic strategy in marine bacteria. Proc. Natl. Acad. Sci. U. S. A. 106, 15527-15533.

McCutcheon, J.P., McDonald, B.R., Moran, N.A., 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc. Natl. Acad. Sci. U. S. A. 106, 15394-15399.

McCutcheon, J.P., Moran, N.A., 2012. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10, 13-26.

Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873-2878.

Moran, N.A., 2006. Symbiosis. Curr. Biol. 16, R866-R871.

Moran, N.A., 2007. Symbiosis as an adaptive process and source of phenotypic complexity. Proc. Natl. Acad. Sci. U. S. A. 104, 8627-8633.

Moran, N.A., McCutcheon, J.P., Nakabachi, A., 2008. Genomics and Evolution of Heritable Bacterial Symbionts. Annu. Rev. Genet. 42, 165-190.

Morin, J.G., Harrington, A., Nealson, K., Krieger, N., Baldwin, T.O., Hastings, J.W., 1975. Light for all reasons - versatility in behavioral repertoire of flashlight fish. Science 190, 74-76.

Reen, F.J., Almagro-Moreno, S., Ussery, D., Boyd, E.F., 2006. The genomic code: inferring Vibrionaceae niche specialization. Nature Reviews Microbiology 4, 697-704.

Sapp, J., 2004. The dynamics of symbiosis: an historical overview. Canadian Journal of Botany-Revue Canadienne De Botanique 82, 1046-1056.

Thompson, F.L., Iida, T., Swings, J., 2004. Biodiversity of vibrios. Microbiol. Mol. Biol. Rev. 68, 403-+.

Wernegreen, J.J., 2002. Genome evolution in bacterial endosymbionts of insects. Nature Reviews Genetics 3, 850-861.

Wernegreen, J.J., Funk, D.J., 2004. Mutation exposed: A neutral explanation for extreme base composition of an endosymbiont genome. J. Mol. Evol. 59, 849-858.

Wernegreen, J.J., Moran, N.A., 1999. Evidence for genetic drift in endosymbionts (Buchnera): Analyses of protein-coding genes. Mol. Biol. Evol. 16, 83-97.

Wollenberg, M.S., Ruby, E.G., 2012. Phylogeny and fitness of *Vibrio fischeri* from the light organs of Euprymna scolopes in two Oahu, Hawaii populations. Isme Journal 6, 352-362.

Woolfit, M., Bromham, L., 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20, 1545-1555.

CHAPTER II

The uncultured luminous symbiont of *Anomalops katoptron* represents a new bacterial genus.

Abstract

Flashlight fishes (Beryciformes: Anomalopidae) harbor luminous symbiotic bacteria in subocular light organs and use the bacterial light for predator avoidance, feeding, and communication. Despite many attempts anomalopid symbionts have not been brought into laboratory culture, which has restricted progress in understanding their phylogenetic relationships with other luminous bacteria, identification of the genes of their luminescence system, as well as the nature of their symbiotic interactions with their fish hosts. To begin addressing these issues, we used culture-independent analysis of the bacteria symbiotic with the anomalopid fish, *Anomalops katoptron*, to characterize the phylogeny of the bacteria and to identify the genes of their luminescence system including those involved in the regulation of luminescence. Analysis of the 16S rRNA, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, and *topA* genes resolved the *A. katoptron* symbionts as a clade nested within and deeply divergent from other members of *Vibrionaceae*. The bacterial luminescence (*lux*) genes were identified as a contiguous set (*luxCDABEG*), as found for the *lux* operons of other luminous bacteria. Phylogenetic analysis based on the *lux* genes confirmed the housekeeping gene phylogenetic placement. Furthermore, genes flanking the *lux* operon in the *A. katoptron* symbionts differed from those flanking *lux*

operons of other genera of luminous bacteria. We therefore propose the candidate name '*Candidatus* Photodesmus katoptron' (Greek: photo = light, desmus = servant) for the species of bacteria symbiotic with *A. katoptron*. Results of a preliminary genomic analysis for genes regulating luminescence in other bacteria identified only a *Vibrio harveyi*-type *luxR* gene. These results suggest that expression of the luminescence system might be continuous in '*Ca*. Photodesmus katoptron.'

Introduction

Flashlight fish (Beryciformes: Anomalopidae) are a group of nocturnally active marine fish found in shallow reef and deep-water tropical habitats. Currently there are eight species in genera *Anomalops*, *Kryptophanaron*, *Parmops*, *Phthanophaneron*, *Photoblepharon*, and *Protoblepharon* (Baldwin et al., 1997; Johnson and Rosenblatt, 1988; Johnson et al., 2001; McCosker and Rosenblatt, 1987; Rosenblatt and Johnson, 1991). At night the fish leave cave and crevice shelters to feed on zooplankton, using light from a pair of bean-shaped, subocular light organs for illumination. The light is used also in signaling for communication and to disorient predators (Harvey, 1922; Herring and Morin, 1978; Morin et al., 1975). Light is produced by dense populations of luminous bacteria housed within an array of tubules that form the bulk of the light organ (Bassot, 1966; Kessel, 1977). These tubules converge at pores on the light organ surface and allow release of the bacteria into the surrounding seawater (Haygood et al., 1984).

In contrast to the light organ symbionts of most known bacterially-bioluminescent fish and squids, anomalopid symbionts have not been grown in laboratory culture, despite numerous attempts (Hastings and Nealson, 1981; Haygood, 1993; Herring and Morin,

1978). Microscopic analysis showing the presence of masses of bacterial cells within anomalopid light organ tubules, assays specific for bacterial luciferase, and other studies convincingly demonstrate the bacterial nature of anomalopid light emission (Bassot, 1966; Harvey, 1922; Haygood et al., 1984; Kessel, 1977; Leisman et al., 1980). Furthermore, sequence analysis of the bacterial 16S rRNA gene and of *luxA*, which codes for the alpha subunit of bacterial luciferase, suggested that these symbionts are related to other luminous bacteria in the family *Vibrionaceae* (Haygood, 1990; Haygood and Distel, 1993; Wolfe and Haygood, 1991). However, the inability to culture the bacteria from anomalopid light organs has held back studies of their phylogenetic placement, the gene structure and regulation of their luminescence system, and the nature of their symbiotic interactions with the fish, in comparison with other bioluminescent associations (Dunlap, 2009; Dunlap et al., 2007; Nyholm and McFall-Ngai, 2004; Visick and Ruby, 2006).

The inability to grow anomalopid bacteria in the laboratory has lead to the suggestion that they may be obligately dependent on their hosts and have lost the ability to reproduce outside of the fish light organ (Haygood, 1993; Haygood and Distel, 1993). If so, this would make them unusual as most confirmed obligate symbionts are intracellular. It would also put anomalopid symbionts in striking contrast to other free-living luminous symbionts (Dunlap et al., 2007). Regardless of whether anomalopid symbionts are obligate or not, they represent a significant gap in our knowledge about a biologically, economically and medically important group of bacteria, the family *Vibrionaceae* (Dunlap, 2009; Mead et al., 1999; Reen et al., 2006; Thompson et al., 2004, 2005). As a first step in gaining insight on these issues, we used culture-independent methods to characterize in detail the phylogenetic status of bacteria residing in light

organs of *Anomalops katoptron*. We also sought to define the gene structure and regulation of the luminescence system of these bacteria, as these characteristics can be both phylogenetically and ecologically informative (Ast and Dunlap, 2004, 2005; Ast et al., 2007).

Materials and methods

Sample collection and preparation

Specimens of *A. katoptron* were collected by commercial aquarium suppliers from coastal waters at Christmas Island (one specimen), and near Cebu City, Cebu, the Philippines (five specimens), and were obtained alive. The fish were sacrificed shortly after acquisition, and the light organs, which were strongly luminous at the time of sacrifice, were excised and processed individually. The light organs were rinsed in filter-sterilized 70% artificial seawater containing 25 mM HEPES buffer (pH 7.25), scored multiple times with a sterile scalpel blade, and then gently squeezed to expel bacteria. The bacterial cells were collected and concentrated by centrifugation, and DNA was extracted from the cell pellets using a Promega (Madison, WI) Wizard Genomic DNA extraction kit following the manufacturer's instructions.

Sequencing and alignment of genes

Housekeeping genes used for phylogenetic analysis, the 16S rRNA gene, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, and *topA*, were PCR-amplified from DNA of bacterial cells collected from one light organ of the Christmas Island fish specimen. PCR products were sequenced by staff of the University of Michigan DNA Sequencing Core using PCR

primers and internally designed primers and dye terminator cycle sequencing on a Perkin-Elmer (Waltham, MA) ABI 3730 or 3700 DNA analyzer. Each gene sequence was covered twice, forward and reverse. Analysis of sequencing chromatograms revealed only single peaks for these genes, which is consistent with monoclonality of the bacteria (Ast et al., 2007); the population of bacteria from this single light organ is therefore referred to collectively as Akat2007.1.1. Sequences were aligned using ClustalX (Thompson et al., 1994) and further adjusted manually in MacClade (Maddison and Maddison, 1992). A similar PCR approach, using degenerate primers, was successful in obtaining sequences of the *luxD*, *luxA*, and *luxB* genes; however, the other genes of the *lux* operon, *luxC*, *luxE* and *luxG*, were not amplified by this approach. Instead, the sequences of the *luxC*, *luxE* and *luxG* genes were obtained from preliminary genomic sequence data for bacterial DNA pooled from five light organs of four fish (from Christmas Island and Cebu City), using a Roche 454FLX Genome Sequencer (Chapter III). The initial alignment of the *lux* genes was done with ClustalX (Thompson et al., 1994) and further proofreading was done manually in MacClade (Maddison and Maddison, 1992). The entire *lux* operon and flanking regions was covered 15 to 20 times by reads of 400 bases or greater. Insertion/deletion ambiguities in sequences of the *luxC*, *luxE*, and *luxG* genes (from 454 reads) were called based on the majority of reads or to maintain the reading frame in coding regions. The very low total number of between-specimen polymorphic sites, estimated as six polymorphisms in 16.5 Kb of DNA for the *lux* genes and flanking DNA, permitted DNA from both pooled and individual light organ samples to be analyzed as a single data point. All genes were annotated by BLAST search and alignment with genes from species in *Vibrionaceae*.

Taxon sampling and phylogenetic analysis

Taxa included in the analysis (see Tables 1 and 2 for strains and sequence accession numbers) were chosen to evenly cover the known diversity within *Vibrionaceae* while using only species for which the majority of loci to be analyzed were available. Type strains were used to the extent possible. The loci examined here include functionally conserved housekeeping genes and the bacterial *lux* genes, which have been shown to be phylogenetically informative within *Vibrionaceae* (Ast and Dunlap, 2004; Ast et al., 2009; Kaeding et al., 2007; Thompson et al., 2005). In addition to gene sequences available from GenBank, *topA* sequences of *Aliivibrio sp.* "thorii", and *Aliivibrio sp.* "sifiae" and a*tpA* sequences of *A. sp.* "thorii", *A. sp.* "sifiae", *Photobacterium leiognathi*, *Photobacterium mandapamensis*, *Photobacterium kishitanii*, and *Photobacterium phosphoreum* were obtained in this study by PCR-amplification and sequencing as described above. Genbank accession numbers of all sequences obtained in this study are shown in Tables 1 and 2.

The housekeeping genes were analyzed individually in Modeltest, version 3.07 (Posada and Crandall, 1998) using the Akaike Information Criterion; the same model was found for each gene. A concatenated gene matrix was then analyzed using Maximum Likelihood in Garli (Zwickl, 2006) under the GTR + I + Γ model with all parameter values calculated by Garli. Housekeeping genes were also analyzed individually in Garli and were visually compared to search for topological differences as an indication of horizontal gene transfer. No major topological differences between gene trees were observed. The *lux* operon was analyzed as one locus with non-coding regions removed. The *lux* genes were found to differ from the housekeeping genes in base frequencies as

well as rate of substitution per base in Modeltest, so the *lux* gene matrix was analyzed separately from the housekeeping gene matrix, again using the GTR + I + Γ model with program calculated values. For both matrices, 1000 bootstrap replicates were performed, with each run for 1000 generations. No likelihood improvements were observed after generation 750. The housekeeping gene tree was rooted with *Grimontia*, *Enterovibrio*, and *Salinivibrio* sequences, since other analyses have shown these genera to be basal within the family (Urbanczyk et al., 2008). The *lux* gene tree was not rooted due to the lack of a suitable outgroup. The same matrices and models were used for Bayesian analyses run in MRBAYES v3.1.2 (Huelsenbeck and Ronquist, 2001). For each gene set two runs of four parallel Markov chain Monte Carlo chains were performed. The housekeeping gene matrix was run for 100,000 generations sampling every 100 generations at which point the standard deviation of split frequencies was effectively zero. *Lux* genes were analyzed for 10,000 generations sampling every 10 generations. Each analysis produced a 50% majority rule consensus tree from 1000 samples with a 25% burn-in.

Identification of lux regulatory genes

To determine if the *A. katoptron* symbiont genome contained genes involved in regulating luminescence that have been identified in other bacteria, the 454 sequencing reads described above were assembled into contigs using the program Velvet (Zerbino and Birney, 2008) by staff of the University of Michigan CCDU Bioinformatics Core. The assembly produced 474 contigs, the majority of which had over 50X coverage. This database was then translated into protein and searched with protein sequences using the

tblastn algorithm in BLAST. Hits with a max score above 200 and greater than 70% sequence identity were initially considered to be homologs. These sequence matches from the *A. katoptron* symbiont genome were then searched against the GenBank database using BLAST to confirm that they matched to the gene of interest. The efficacy of this approach for identifying genes was verified by searching for genes previously amplified by PCR. One sequence, *luxR*, was analyzed by alignment and Maximum Likelihood based phylogenetic analysis for specific sequence similarities to related species and this sequence was submitted to Genbank under the accession JN007808.

Results

Phylogenetic analysis based on housekeeping genes

Previously, analysis of 16S rRNA gene sequences indicated that bacteria symbiotic with anomalopids are related to members of *Vibrionaceae* (Haygood and Distel, 1993). To gain deeper insight into the phylogenetic placement of anomalopid bacteria, we sequenced several housekeeping genes, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, and *topA*, and the 16S rRNA gene for the bacteria, Akat2007.1.1, from a single light organ of a specimen of *A. katoptron*. These newly obtained sequences are available in Genbank (accessions listed in Table 2.1). We then carried out a phylogenetic analysis using both Maximum Likelihood (ML) and Bayesian (BA) methods. Both analyses of these housekeeping genes recovered trees of identical topology. The Maximum Likelihood tree with ML bootstrap values and BA posterior probabilities is shown (Fig. 2.1). The tree resolved Akat2007.1.1, together with other anomalopid symbionts, as a clade distinct from known genera within *Vibrionaceae* with high bootstrap and posterior

probability support. The anomalopid symbiont clade was deeply divergent from other members of *Vibrionaceae*. Although only 16S rRNA gene sequences were available for anomalopid symbionts other than Akat2007.1.1 (Haygood and Distel, 1993), the analyses supported the grouping of these taxa. In addition, the anomalopid bacteria were placed as sister to *Vibrio* with 100% bootstrap and posterior probability support. We note parenthetically here, however, that low bootstrap support was found for monophyly of *Vibrio*. Furthermore, branch lengths were much longer for the anomalopid symbiont clade than for other genera (1306 changes along the genus branch compared to 212, 299, and 498 for *Photobacterium*, *Aliivibrio*, and *Vibrio*, respectively), indicating a faster rate of divergence of the anomalopid bacteria. The more rapid divergence was confirmed with a relative rate test (Tajima's Test) (Tajima, 1993) in Mega4 (Tamura et al., 2007). Using *A. fischeri* as an outgroup, Akat2007.1.1 was found to be evolving at a significantly higher rate than *Vibrio harveyi* ($p = 0.000001$); specifically, out of 9179 housekeeping gene sites, Akat2007.1.1 had 963 unique changes compared to 411 for *V. harveyi*. These results indicate that the bacteria symbiotic with anomalopid fish represent a clade within *Vibrionaceae* that is evolutionarily distinct from other genera. We propose the candidate name '*Candidatus* Photodesmus' (Greek: photo = light, desmus = servant) for this genus and the candidate name '*Candidatus* Photodesmus katoptron' for the species of bacteria symbiotic with *A. katoptron*.

Phylogenetic analysis based on luminescence genes

To further test this phylogenetic placement, we also sequenced the luminescence (*lux*) genes and flanking regions of '*Ca*. Photodesmus katoptron' (Genbank accession

numbers in Table 2.2). Phylogenetic analysis based on *lux* gene sequences resolves luminous bacteria effectively and has also provided insight into the frequency of horizontal gene transfer in light-emitting bacteria (Ast and Dunlap, 2004; Ast et al., 2009; Dunlap et al., 2007; Kaeding et al., 2007; Kasai et al., 2007; Urbanczyk et al., 2008). Previously, the *luxA* gene of bacteria symbiotic with *Kryptophanaron alfredi* was sequenced and found to be homologous to *luxA* of *Vibrionaceae* (Haygood, 1990). We found that genes of the '*Ca*. Photodesmus katoptron' *lux* operon are present with the same gene content and order as in other luminous bacteria, *luxCDABEG* (Fig. 2.1). These genes were easily aligned with homologs in other species of luminous bacteria except for 180 bases at the 5' end of *luxC*. Although this area showed no significant similarity to known homologs, a translational start codon is present in the same approximate location as in other species.

Both ML and BA analyses based on *luxCDABEG* gene sequences yielded the same general topology as that based on housekeeping genes, with strong resolution of the anomalopid symbiont clade from *Photobacterium*, *Aliivibrio*, and *Vibrio* (Fig. 2.2). At this time, there are no known luminous representatives of the other *Vibrionaceae* genera, *Salinivibrio*, *Grimontia*, and *Enterovibrio*. The similarity of topologies of housekeeping genes and *lux* genes indicates that '*Ca*. Photodesmus katoptron' apparently did not acquire the *lux* genes by horizontal transfer. Branch lengths for the *lux* gene tree were not distinctly longer, however, for the anomalopid bacteria than for other genera, indicating that the *lux* genes of anomalopid bacteria are not evolving more rapidly than those of other luminous bacteria. These results affirm the deep divergence of the anomalopid

symbiont clade from other members of *Vibrionaceae* identified through analysis of housekeeping genes.

Genes flanking the lux operon

To examine further the relationship between '*Ca*. Photodesmus katoptron', *Vibrio*, and other luminous bacteria, we next compared the regions flanking the *lux* operons of these bacteria. We found that genes that flank the *lux* operon in *V. harveyi* and *V. cholerae* (as demonstrated in Genbank accessions NC_009784, AB119994, and AB120061) do not flank the *lux* operon of '*Ca*. Photodesmus katoptron'. Also, luminescence regulatory genes that flank the *lux* operon in *Aliivibrio*, *luxI* and *luxR* (e.g., (Dunlap, 2009), and genes related to luminescence that flank the *lux* operon of *Photobacterium*, the fluorescent lumazine operon genes *lumP* and *lumQ* and the riboflavin synthesis genes *ribEBHA*, (Ast et al., 2007; Lin et al., 1996; Lin et al., 1995), do not flank the *lux* operon of '*Ca*. Photodesmus katoptron' (Fig. 2.3A). Instead, these regions in '*Ca*. Photodesmus katoptron' contain housekeeping genes identified by sequence similarity as homologs of other genes in *Vibrionaceae*: acetyl-CoA carboxytransferase alpha subunit (*accA*), aspartate transcarbamylase (*pyrIB*), and the iron (III) transporter system genes (*fhuBCD*) (Fig. 2.3A). Additionally one segment of sequence was alignable with 65% similarity to *dsbD*, which codes for a fused thiol:disulfide interchange protein. None of these genes, however, are found flanking the *lux* operons of luminous species for which these regions have been examined (Fig. 2.3A). Because the DNA flanking *lux* operons often is similar within a genus and different between genera (Ast et al., 2007; Dunlap, 2009), these results further support the

classification of the anomalopid symbiont clade as a new *Vibrionaceae* genus, '*Ca*. Photodesmus'.

Regulation of lux operon expression

To determine if the '*Ca*. Photodesmus katoptron' genome contains homologs of genes that regulate luminescence in other luminous bacteria, we next examined a draft sequence of the '*Ca*. Photodesmus katoptron' genome (Chapter III). Homologs of the *A. fischeri luxR*, *luxI*, and *ainS* genes were not found, whereas homologs of *cyaA* and *crp*, which are involved in controlling luminescence in *A. fischeri* and *V. harveyi* (Chatterjee et al., 2002; Ulitzur and Dunlap, 1995) were present. However, a *crp* binding site was not found in the *lux* operon regulatory region. Most interestingly, a homolog of the *V. harveyi* gene for LuxR, an activator of the *lux* operon, was identified in the '*Ca*. Photodesmus katoptron' genome (Genbank accession JN007808). Upstream genes that regulate *luxR* in *V. harveyi*, *luxS*, *luxM*, *cqsA*, *luxO*, *luxU*, *luxP*, *luxQ*, *luxN*, and *cqsS*, however, were not found. The '*Ca*. Photodesmus katoptron' *luxR* gene showed a high degree of sequence similarity to *V. harveyi luxR* homologs from other members of *Vibrio*, specifically 71% sequence similarity to the *V. harveyi* homolog, but not to phylogenetically and functionally distinct *luxR* family genes from other members of *Vibrionaceae* (Fig. 2.4)

The presence of a homolog of the *V. harveyi luxR* gene suggests that luminescence in '*Ca*. Photodesmus katoptron' could be controlled by LuxR. Consistent with this possibility, the *lux* regulatory region upstream of the *lux* operon of '*Ca*. Photodesmus katoptron' was alignable with the same region in *V. harveyi.* Within this

region, the two LuxR binding sites identified for *V. harveyi* (Lee et al., 2008) were present, and the conserved inverted repeats of these sites were very similar to the *V. harveyi* sites (Fig. 2.3B). This similarity and the presence of a *V. harveyi*-like *luxR* gene are consistent with the close evolutionary relationship between '*Ca*. Photodesmus katoptron' and *Vibrio* identified through phylogenetic analysis. However, the apparent lack of other *V. harveyi*-like regulatory genes in the '*Ca*. Photodesmus katoptron' genome suggests that expression of the *lux* operon either is constitutive in '*Ca*. Photodesmus katoptron' or that it is regulated in a manner different from *V. harveyi*. Supporting these possibilities, the LuxR binding site in the *luxR* promoter region of *V. harveyi*, which is involved in autoregulatory control of LuxR synthesis (Chatterjee et al., 1996), was not present in the *luxR* promoter region of '*Ca*. Photodesmus katoptron.'

Discussion

Determining broad evolutionary patterns requires accurately reconstructed phylogenies for trait mapping and phylogenetically independent comparative analyses (Pagel, 1997; Price, 1997). Achieving this goal in bacteria is complicated by an inability to match sequence data to organisms, as most taxa cannot be grown in pure culture. Anomalopid symbionts present an unusual case where large numbers of an uncultured bacterial species are found in the absence of other bacteria. This study presents a detailed assessment of the phylogeny of the uncultured luminous bacterial symbionts from the light organs of the anomalopid fish, *A. katoptron*. The ability to obtain sufficient DNA from these genetically monomorphic bacteria makes it possible to use culture independent means to gain insights into the lifestyle of the bacterium and the nature of

the host-symbiont interaction. Anomalopid symbionts are the first uncultured luminous symbiont to be studied in similar depth to the many free-living, culturable symbionts in the family. Here we have shown that they are significantly divergent from relatives and represent a new genus level clade, and we discuss the implications of our findings within a phylogenetic context.

Earlier studies based on the 16S rRNA and *luxA* genes (Haygood, 1990; Haygood and Distel, 1993) provided evidence that luminous bacteria symbiotic with anomalopids are related to *Vibrionaceae*, an ecologically diverse group of Gram-negative bacteria (Reen et al., 2006; Thompson et al., 2004). To test this placement we analyzed several phylogenetically informative genes, both housekeeping genes and luminescence genes (Ast and Dunlap, 2004; Ast et al., 2009; Thompson et al., 2005). We also established the gene order of the *lux* operon, which has been shown to provide insight into phylogenetic relationships (Ast and Dunlap, 2004). These analyses support the placement of anomalopid symbionts in a new genus and lead us to propose the candidate name '*Ca.* Photodesmus katoptron' for *A. katoptron* symbionts. The low sequence polymorphism found here between samples shows that other specimens of *A. katoptron* are likely to harbor this same species of symbiont, and phylogenetic analysis suggests that those of other anomalopid species are likely to be members of '*Ca.* Photodesmus'. However, this analysis is based only on 16S rRNA gene data for all symbiont sequences except the *A. katoptron* symbiont. Additional sequence data from the symbionts of other anomalopids will be necessary to determine if all anomalopid symbionts should be placed in the same genus and to attempt to draw species level delineations.

The inability to culture anomalopid light organ bacteria on laboratory media (Hastings and Nealson, 1981; Haygood, 1993; Herring and Morin, 1978) suggests that these bacteria have lost the ability to survive and reproduce outside of the fish light organ. Here we confirmed that anomalopid symbionts are evolving at a faster rate than relatives for the housekeeping genes tested. This trend is intriguing because many obligate symbionts have been shown to be evolving much faster than relatives. For example, the bacterial endosymbionts of insects have higher rates of nucleotide substitutions than free-living relatives (Moran, 1996; Spaulding and von Dohlen, 1998; Woolfit and Bromham, 2003). The higher rate of divergence has been shown to result from drift in small, non-recombining bacterial populations that undergo regular bottleneck events as they transfer to new host generations (Balbi and Feil, 2007; Moran, 1996; Wernegreen, 2002; Wernegreen and Moran, 1999; Woolfit and Bromham, 2003). The aphid symbionts Buchnera, for example, undergo sharp decreases in population size during transfer to a new host generations (Mira and Moran, 2002), increasing the strength of genetic drift (Balbi and Feil, 2007; Wernegreen, 2002; Wernegreen and Moran, 1999; Woolfit and Bromham, 2003). Although the mechanism by which anomalopid bacteria transfer between host generations is not yet known, the increased nucleotide substitution rate in these bacteria compared to other members of *Vibrionaceae* is consistent with population bottlenecks at their transfer to members of a new host generation and is therefore suggestive of an obligate relationship.

A second commonality is that, like '*Ca*. Photodesmus katoptron,' obligate symbionts have low levels of genetic polymorphism within a host species. For example, Buchnera sampled from the same aphid host species exhibit very little genetic

polymorphism, even across wide geographic ranges, whereas differences between host species are much greater (Abbot and Moran, 2002; Funk et al., 2001). A similarly low level of variation has been found in the symbionts of deep-sea vestimentiferan tube worms (Di Meo et al., 2000; Harmer et al., 2008) and in intracellular pathogenic bacteria (Achtman, 2008; Achtman et al., 2004; Sreevatsan et al., 1996; Van Ert et al., 2007). This pattern might relate to host demography, with bottlenecks in host populations removing genetic variation from symbiont populations (Funk et al., 2001). Low genetic polymorphism therefore might be indicative of a tight symbiont-host relationship. The levels of polymorphism found in '*Ca*. Photodesmus katoptron' genes analyzed here are lower than those reported for tube worm symbionts (Di Meo et al., 2000; Harmer et al., 2008) and in the same range as values found in Buchnera (Abbot and Moran, 2002; Funk et al., 2001).

Bacterial luminescence genes have been used as systems for the study of bacterial gene evolution and regulation (Ast and Dunlap, 2004; Ast et al., 2007; Kasai et al., 2007; Urbanczyk et al., 2008), specifically quorum sensing (Hasegawa and Hase, 2009; Henke and Bassler, 2004a,b; Jaques and McCarter, 2006; Lee et al., 2008; Lenz et al., 2004; Miller and Bassler, 2001; Miyamoto et al., 2003; Tu et al., 2008; Waters and Bassler, 2005). However, the majority of these studies contain only culturable *Vibrionaceae* species, providing an incomplete view of evolution. Most luminous symbionts are culturable, environmentally acquired, and live in a variety of habitats in addition to being host associated (Ast et al., 2007; Dunlap, 2009; Dunlap et al., 2007; Wada et al., 1999). However, the symbionts of deep sea anglerfish (Ceratioidei) and flashlight fish (Beryciformes: Anomalopidae) have not been cultured despite considerable effort,

leading to speculation that they may be obligately dependent upon their hosts (Haygood, 1993; Haygood and Distel, 1993; Herring and Morin, 1978). The addition of such taxa to comparative analyses within a phylogenetic framework will improve our understanding of the intricacies of evolution in bioluminescent symbiosis in a variety of ways. Here, we investigated the structure and regulation of the '*Ca*. Photodesmus katoptron' *lux* operon in comparison to other members of *Vibrionaceae*.

With respect to the origin and evolution of the bacterial *lux* genes, the results of this study, by including data from an uncultured luminous bacterium, '*Ca*. Photodesmus katoptron,' support the view that the bacterial luminescence system arose once, apparently in the common ancestor of *Vibrionaceae*. In all light-emitting bacteria examined to date, including here the uncultured symbionts of *A. katoptron*, the *lux* operons have very similar gene content and the same gene organization, and the *lux* genes are homologous. On the other hand, the many species of *Aliivibrio*, *Photobacterium*, and *Vibrio* that lack the *lux* genes, and the presence of *lux* genes in certain members of allied bacterial families, *Enterobacteriaceae* and *Alteromonadaceae* indicate that the luminescence system has been lost frequently within *Vibrionaceae* and has been transferred infrequently within and beyond *Vibrionaceae* (Dunlap, 2009; Urbanczyk et al., 2008).

Our results also have implications for the evolution of luminescence gene regulation. The presence of a *V. harveyi*-type *luxR* gene in the '*Ca*. Photodesmus katoptron' genome and *luxR* binding sites in the *lux* operon promoter region (Fig. 2.3B) indicate that LuxR protein is likely to function in '*Ca*. Photodesmus katoptron' to activate *lux* operon expression. This appears to be the first case of a *V. harveyi*-type *luxR*

functioning in luminescence regulation outside of the genus *Vibrio* and suggests that the *V. harveyi*-type *lux* regulatory system evolved in the common ancestor of *Vibrio* and '*Ca*. Photodesmus.' *V. harveyi*-type LuxR is a member of a family of regulatory genes, but within *Vibrio*, *luxR* homologs have been shown to have conserved function regulating a large number of pathways (Fidopiastis et al., 2002; Hasegawa and Hase, 2009; Henke and Bassler, 2004a; Jaques and McCarter, 2006; Jobling and Holmes, 1997; Lee et al., 2008; Lenz et al., 2004; Miyamoto et al., 1998). The results presented here indicate that LuxR may be involved in the regulation of such genes in '*Ca*. Photodesmus katoptron' as well.

Within *Vibrio harveyi* three different signal molecules operate via a complex sensor-kinase phosphorylation cascade to activate *lux* operon expression by relieving inhibition of LuxR expression by *luxO* (Tu et al., 2008; Waters and Bassler, 2005). In '*Ca*. Photodesmus katoptron,' however, genes coding for proteins involved in synthesis of these signals, LuxS, LuxM, and CqsA, and for the components of the sensor-kinase phosphorylation cascade, LuxO, LuxU, LuxP, LuxQ, LuxN, and CqsS, are absent. Therefore, luminescence in '*Ca*. Photodesmus katoptron' is either expressed constitutively or is regulated in a manner different from *V. harveyi*. The apparent absence from the genome of '*Ca*. Photodesmus katoptron' of regulatory genes other than a *V. harveyi*-type *luxR* may be consistent with an obligate association, one in which luminescence is always required and in which genes that turn off luminescence would not be needed and could be lost with no negative effect. The regulatory genes therefore might have been lost from the lineage leading to anomalopid bacteria. Alternatively, the *V. harveyi*-type *luxR* gene might have arisen in the common ancestor of anomalopid symbionts and *Vibrio* before *Vibrio* evolved upstream quorum sensing genes.

Acknowledgements

This work was coauthored with Paul V. Dunlap and published in Molecular Phylogenetics and Evolution in 2011. We thank the staff of the University of Michigan Sequencing Core for DNA sequencing and advice and Jeffrey R. de Wet, University of Michigan Bioinformatics Analysis Core, for carrying out sequence alignments. LiveAquaria, Quality Marine, and Fish Doctors Aquarium provided assistance in obtaining fish. Support for T. Hendry was provided by departmental Block Grant awards from the University of Michigan Department of Ecology and Evolutionary Biology and by Rackham Graduate Student Grants from the University of Michigan Rackham Graduate School.

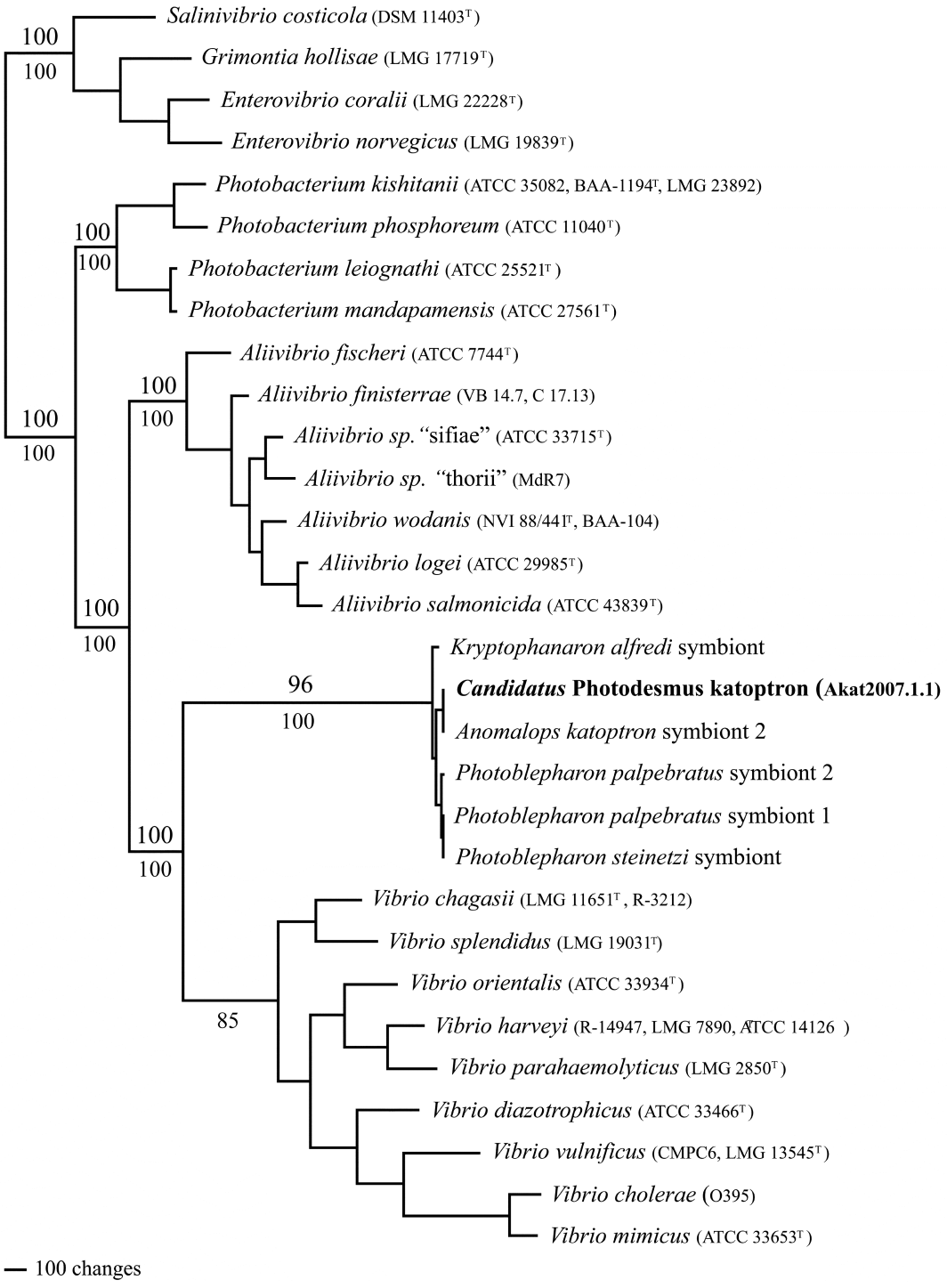

Figure 2.1. Maximum likelihood trees from housekeeping genes (16S rRNA gene, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, and *topA*). Maximum Likelihood bootstrap numbers are shown above branches and Bayesian posterior probabilities are shown below. Strain designations follow taxa names.

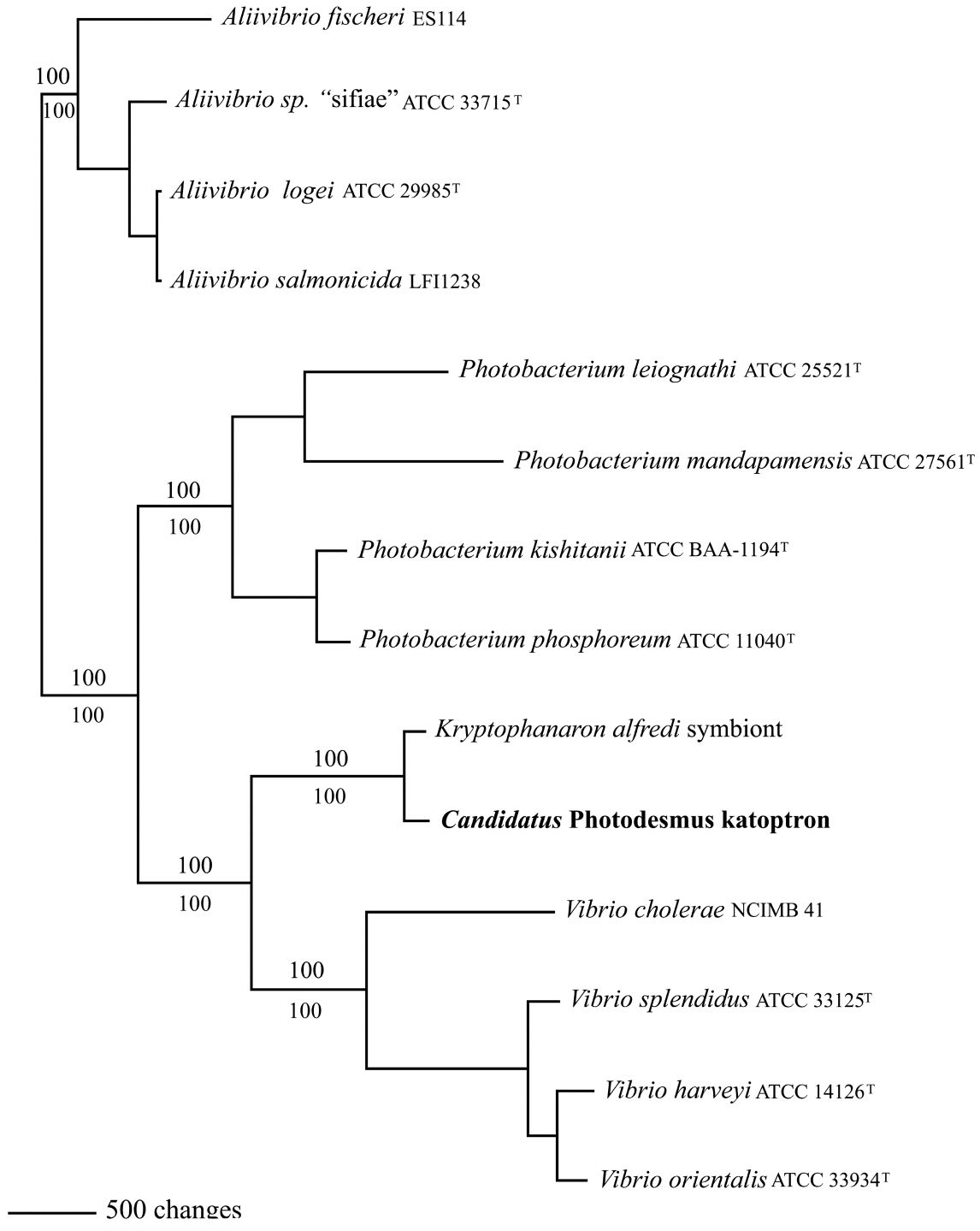

Figure 2.2. Maximum likelihood trees from the *lux* operon. Maximum Likelihood bootstrap numbers are shown above branches and Bayesian posterior probabilities are shown below. Strain designations follow taxa names.

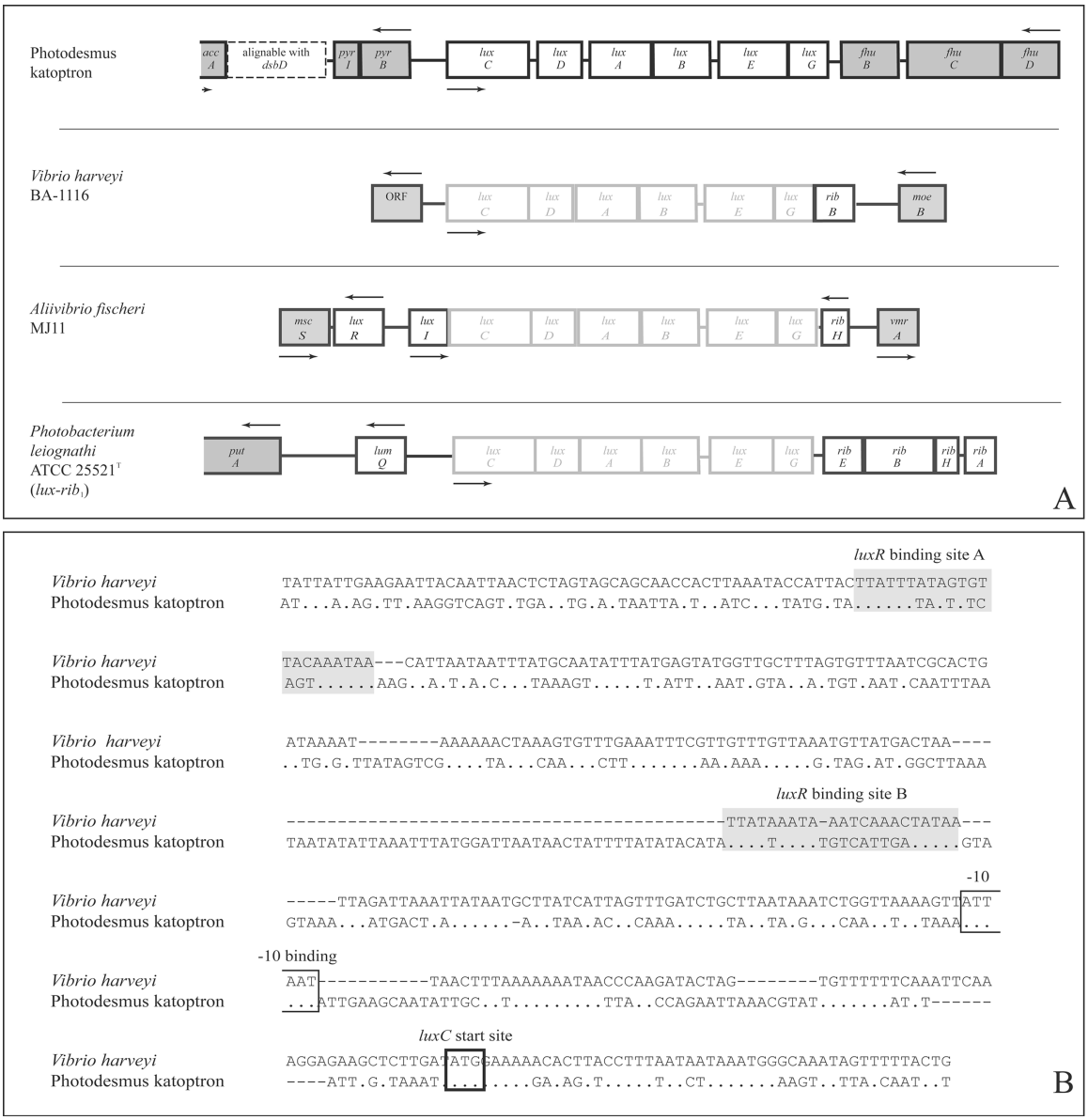

Figure 2.3. A. The structure of the *lux* operon and flanking regions in the *A. katopron* symbiont and representative members of *Vibrionaceae*. Genes and intergenic spacers are drawn to scale. Arrows indicate the direction of transcription. Faded sections are shared and homologous with the *A. katopron* symbiont sequence. Data compiled from Genbank and Ast et al., 2007; Lee et al., 2008; Lin et al., 1995; Lin et al., 1996. B. Alignment of the regulatory region upstream of the *lux* operon in the *A. katopron* symbiont and *Vibrio harveyi* (Lee et al., 2008). Shared bases are denoted by (.). Shown are both *luxR* binding sites, the putative transcription start site (-10 binding) and the coding start site of *luxC*.

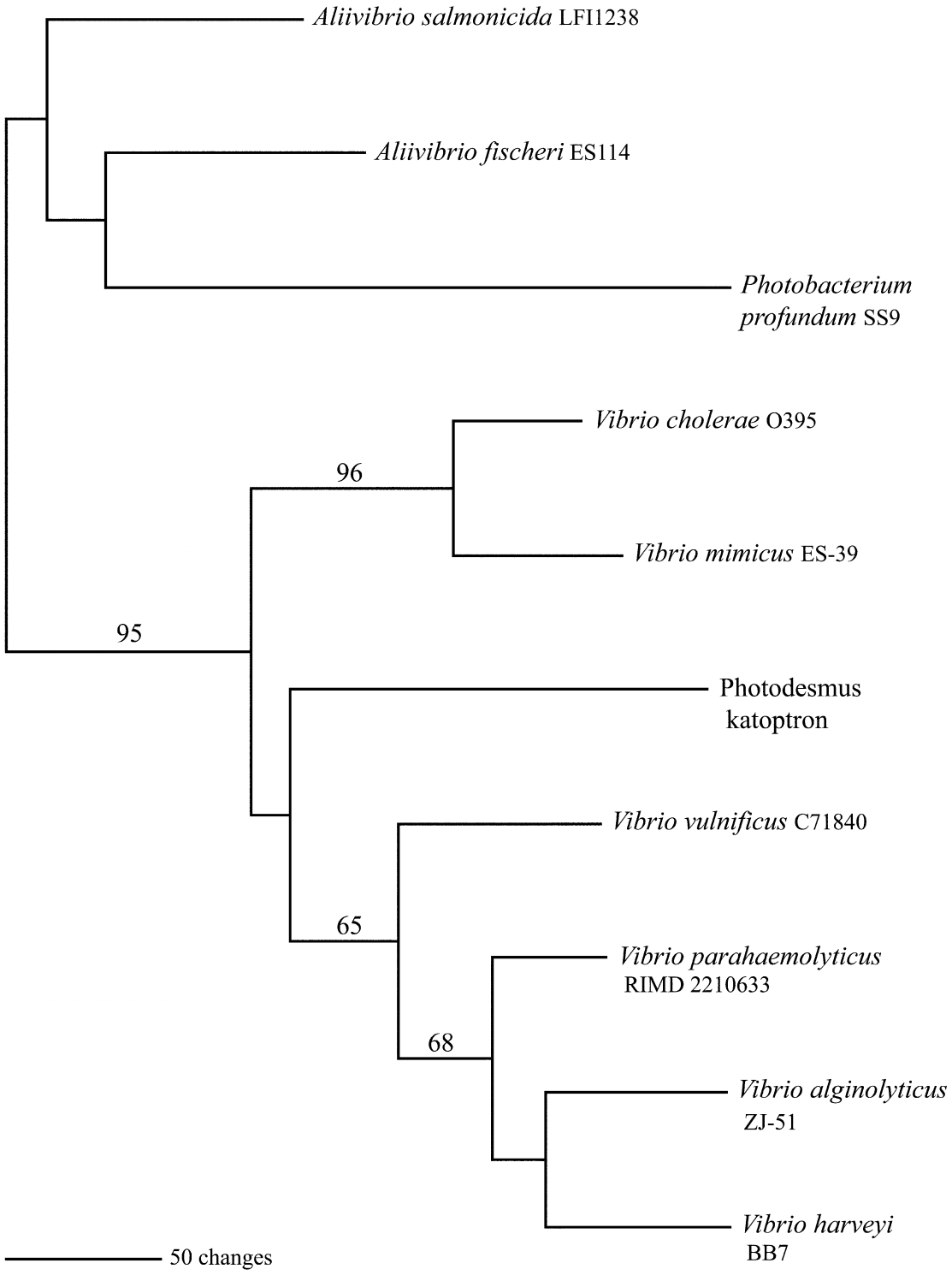

Figure 2.4. Maximum likelihood tree of '*Ca*. Photodesmus katoptron' *luxR* and *Vibrionaceae* homologs. Bootstrap values greater than 60 and strain designations are shown. Analysis was done in Garli (Zwickl, 2006) with the GTR + I + Γ model and program calculated values. 500 generation and 100 bootstrap replicates were done.

Table 2.1. GenBank accession numbers and strain designations for all taxa in the housekeeping gene analysis.

	gyrB	*rpoA*	16S	*gapA*	*topA*	*recA*	*atpA*	*pyrH*
Anomalopid symbionts								
Anomalops katoptron symbiont (this study)	Akat2007.1.1 HQ333494	Akat2007.1.1 HQ333498	Akat2007.1.1 HQ333493	Akat2007.1.1 HQ333496	Akat2007.1.1 HQ333492	Akat2007.1.1 HQ333497	Akat2007.1.1 HQ333485	Akat2007.1.1 HQ333495
Anomalops katoptron symbiont 1	------	------	NA Z19081	------	------	------	------	------
Kryptophanaron alfredi symbiont	------	------	NA Z19003	------	------	------	------	------
Photoblepharon palpebratus symbiont 1	------	------	NA Z19085	------	------	------	------	------
Photoblepharon palpebratus symbiont 2	------	------	NA Z19079	------	------	------	------	------
Photoblepharon steinetzi symbiont	------	------	NA Z19080	------	------	------	------	------
Vibrionaceae								
Aliivibrio fischeri	ATCC 7744^T AY455874	ATCC 7744^T EF415578	ATCC 7744^T AY341436	ATCC 7744^T EF415488	LMG 4414^T DQ907482	ATCC 7744^T EF415542	LMG 4414^T EF601271	ATCC 7744^T EF415528
Aliivibrio finisterrae	------	VB 14.7 EU931116	C 17.13 EU541616	------	------	VB. 14.7 EU931115	VB 14.7 EU931117	C 17.3 EU889129
Aliivibrio logei	ATCC 29985^T EF380255	LMG 19806^T AJ842643	NCIMB 2252^T AJ437616	ATCC 29985^T EU185847	LMG 19806^T DQ907494	ATCC 29985^T AJ842457	ATCC 29985^T EF601226	ATCC 29985^T EF380234
Aliivibrio salmonicida	ATCC 43839^T EF380256	ATCC 43839^T EF380249	NCIMB 2262^T X70643	ATCC 43839^T EU185817	NCIMB 2262^T DQ907517	ATCC 43839^T EF380243	NCIMB 2262^T EU871967	ATCC 43839^T EU118245

	gyrB	*rpoA*	16S	*gapA*	*topA*	*recA*	*atpA*	*pyrH*
Aliivibrio sp. "sifiae"	ATCC 33715 EU185886	ATCC 33715 EU185960	ATCC 33715 EU185828	ATCC 33715 EU185857	ATCC 33715 HQ333490	ATCC 33715 EU185937	ATCC 33715 HQ333483	ATCC 33715 E185914
Aliivibrio sp. "thorii"	MdR7 EU185897	MdR7 EU185971	MdR7 EU185839	MdR7 EU185868	MdR7 HQ333491	MdR7 EU185948	MdR7 HQ333484	MdR7 EU185897
Aliivibrio wodanis	BAA-104 EU185883	BAA-104 EU185957	NVI 88/441^T AJ132227	BAA-104 DQ907314	NCIMB 13582^T DQ907523	BAA-104 EU185934	K16 EF601361	BAA-104 EU185913
Enterovibrio coralii	LMG 22228^T AB298198	LMG 22228^T AJ842530	LMG 22228^T AJ842343	LMG 22228^T DQ907268	LMG 22228^T EF114217	LMG 22228^T AJ842347	LMG 22228^T EF601349	------
Enterovibrio norvegicus	LMG 19839^T AB298199	------	LMG 19839^T AJ316208	LMG 19839^T DQ907269	LMG 19839^T EF114217	LMG 19839^T AJ842348	------	------
Grimontia hollisae	LMG 17719^T AB298259	LMG 17719^T AJ842535	LMG 17719^T AJ514909	LMG 17719^T DQ907317	LMG 17719^T EF114215	LMG 17719^T AJ842351	LMG 17719^T EF601247	------
Photobacterium kishitanii	ATCC 35082 EF415510	BAA-1194 EF415588	pjapo.1.1 AY341439	pjapo.1.1 EF415500	pjapo.1.1 AB453692	pjapo.2.1 DQ648396	pjapo.1.1 HQ333486	LMG 23892 EF415538
Photobacterium leiognathi	ATCC 25521^T AY455879	ATCC 25521^T EF415581	ATCC 25521^T X74686	ATCC 25521^T EF415492	NCIMB 2193^T DQ907463	ATCC 25521^T EF415546	ATCC 25521^T HQ333487	ATCC 25521^T EF380238
Photobacterium manadapamensis	ATCC 27561^T AY455883	ATCC 27561^T EF415583	ATCC 27561^T AY341441	ATCC 27561^T EF415494	ATCC 27561^T AB453693	ATCC 27561^T EF415548	ATCC 27561^T HQ333488	ATCC 27561^T EF415532
Photobacterium phosphoreum	ATCC 11040^T AY455875	ATCC 11040^T EF415585	ATCC 11040^T D25310	ATCC 11040^T EF415496	ATCC 14401^T DQ907465	ATCC 11040^T EF415550	ATCC 11040^T HQ333489	ATCC 11040^T EF380239
Salinivibrio costicola	DSM 11403^T AB298255	LMG 11651^T AJ842552	NCIMB 701^T X95527	LMG 11651^T DQ907316	LMG 11651^T DQ907465	LMG 11651^T AJ842367	------	------
Vibrio chagasii	LMG 21353^T AM162568	LMG 21353^T AJ842572	R-3212 AJ316199	LMG 21353^T DQ481611	------	LMG 21353^T AJ842385	LMG 21353^T EF601280	LMG 21353^T EU118252

	gyrB	*rpoA*	16S	*gapA*	*topA*	*recA*	*atpA*	*pyrH*
Vibrio cholerae	O395 NC_009457	O395 NC_009457	O395 NC_009457	O395 NC_009457	O395 NC_009457	O395 NC_009457	O395 NC_009457	O395 NC_009457
Vibrio diazotrophicus	ATCC 33466^T AB014951	LMG 7893^T AJ842598	ATCC 33466^T X56577	LMG 7893^T DQ907280	LMG 7893^T DQ907480	LMG 7893^T AJ842411	LMG 7893^T EF601270	------
Vibrio harveyi	ATCC 14126^T DQ648280	R-14947 AJ842630	ATCC 14126^T X74706	LMG 4044^T EF596145	LMG 4044^T EF596501	ATCC 14126^T DQ648369	LMG 7890 EF601233	ATCC 14126^T EU118238
Vibrio mimicus	ATCC 33653^T EF380259	ATCC 33653^T EF643486	ATCC 33653^T X74713	LMG 7896^T DQ907292	LMG 7896^T DQ907498	ATCC 33653^T EF643485	LMG 7896^T EF601243	ATCC 33653^T EU118242
Vibrio orientalis	ATCC 33934^T EF380260	LMG 7897^T AJ842672	ATCC 33934^T X74719	LMG 7897^T DQ907299	LMG 7897^T DQ907507	LMG 7897^T AJ842485	LMG 7897^T EF601341	ATCC 33934^T EU118243
Vibrio parahaemolyticus	ATCC 17802^T AF007287	LMG 2850^T AJ842677	ATCC 17802^T X74720	LMG 2850^T DQ449618	LMG 2850^T DQ907509	LMG 2850^T AJ842490	LMG 2850^T EF601274	ATCC 17802^T EU118240
Vibrio splendidus	ATCC 33125^T EF380261	LMG 19031^T AJ842725	ATCC 33125^T X74724	LMG 19031^T DQ481622	LMG 19031^T DQ481661	LMG 19031^T AJ842511	LMG 19031^T EF601244	ATCC 33125^T EU118241
Vibrio vulnificus	ATCC 27562^T AY705491	LMG 13545^T AJ842737	ATCC 27562^T X74726	LMG 13545^T DQ907313	LMG 13545^T DQ907522	LMG 13545^T AJ842523	CMPC6 AE016795	ATCC 43382 EU118244

Table 2.2. GenBank accession numbers and strain designations for taxa in the *lux* gene analysis.

	Strain	*luxC*	*luxD*	*luxA*	*luxB*	*1uxE*	*luxG*
Anomalops katoptron symbiont	metagenomic	HQ333499	HQ333499	HQ333499	HQ333499	HQ333499	HQ333499
Kryptophanaron alfredi symbiont		------	------	M36597	M36597	------	------
Aliivibrio fischeri	(ES114)	CP000021	CP000021	CP000021	CP000021	CP000021	CP000021
Ailivibrio logei	(ATCC 29985^{T})	------	------	EF576941	EF576941	------	------
Aliivibrio salmonicida	(LFI1238)	FM178380	FM178380	FM178380	FM178380	FM178380	FM178380
Aliivibrio sp. "sifiae"	(ATCC 33715)	------	------	EU185990	EU185990	EU185990	------
Photobacterium kishitanii	(pjapo.1.1)	DQ988874	DQ988874	DQ988874	DQ988874	DQ988874	DQ988874
Photobacterium leiognathi	(ATCC 25521^{T})	M63594	M63594	M63594	M63594	M63594	M63594
Photobacterium manadapamensis	(ATCC 27561^{T})	DQ988878	DQ988878	DQ988878	DQ988878	DQ988878	DQ988878
Photobacterium phosphoreum	(ATCC 11040^{T})	DQ988873	DQ988873	DQ988873	DQ988873	DQ988873	DQ988873
Vibrio cholerae	(NCIMB 41)	AB115761	AB115761	AB115761	AB115761	AB115761	AB115761
Vibrio harveyi	(ATCC 14126^{T})	EU122288	EU122288	EU122288	EU122288	EU122288	EU122288
Vibrio orientalis	(ATCC 33934^{T})	EU122287	EU122287	EU122287	EU122287	EU122287	EU122287
Vibrio splendidus	(ATCC 33125^{T})	------	------	EF536342	EF536342	------	------

References

Abbot, P., Moran, N.A., 2002. Extremely low levels of genetic polymorphism in endosymbionts (Buchnera) of aphids (*Pemphigus*). Mol. Ecol. 11, 2649-2660.

Achtman, M., 2008. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu. Rev. Microbiol. 62, 53-70.

Achtman, M., Morelli, G., Zhu, P.X., Wirth, T., Diehl, I., Kusecek, B., Vogler, A.J., Wagner, D.M., Allender, C.J., Easterday, W.R., Chenal-Francisque, V., Worsham, P., Thomson, N.R., Parkhill, J., Lindler, L.E., Carniel, E., Keim, P., 2004. Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc. Natl. Acad. Sci. U. S. A. 101, 17837-17842.

Ast, J.C., Dunlap, P.V., 2004. Phylogenetic analysis of the lux operon distinguishes two evolutionarily distinct clades of *Photobacterium leiognathi*. Arch. Microbiol. 181, 352-361.

Ast, J.C., Dunlap, P.V., 2005. Phylogenetic resolution and habitat specificity of members of the *Photobacterium phosphoreum* species group. Environmental Microbiology 7, 1641-1654.

Ast, J.C., Urbanczyk, H., Dunlap, P.V., 2007. Natural merodiploidy of the lux-fib operon of *Photobacterium leiognathi* from coastal waters of Honshu, Japan. J. Bacteriol. 189, 6148-6158.

Ast, J.C., Urbanczyk, H., Dunlap, P.V., 2009. Multi-gene analysis reveals previously unrecognized phylogenetic diversity in *Aliivibrio*. Syst. Appl. Microbiol. 32, 379-386.

Balbi, K.J., Feil, E.J., 2007. The rise and fall of deleterious mutation. Res. Microbiol. 158, 779-786.

Baldwin, C.C., Johnson, G.D., Paxton, J.R., 1997. *Protoblepharon rosenblatti*, a new genus and species of flashlight fish (Beryciformes: Anomalopidae) from the tropical South Pacific, with comments on anomalopid phylogeny. Proceedings of the Biological Society of Washington 110, 373-383.

Bassot, J.-M., 1966. On the comparative morphology of some luminous organs. In: Johnson, F.H., Haneda, Y. (Eds.), Bioluminescence in progress. Princeton University, Princeton, NJ.

Chatterjee, J., Miyamoto, C.M., Meighen, E.A., 1996. Autoregulation of luxR: The Vibrio harveyi lux-operon activator functions as a repressor. Mol. Microbiol. 20, 415-425.

Chatterjee, J., Miyamoto, C.M., Zouzoulas, A., Lang, B.F., Skouris, N., Meighen, E.A., 2002. MetR and CRP bind to the *Vibrio harveyi lux* promoters and regulate luminescence. Mol. Microbiol. 46, 101-111.

Di Meo, C.A., Wilbur, A.E., Holben, W.E., Feldman, R.A., Vrijenhoek, R.C., Cary, S.C., 2000. Genetic variation among endosymbionts of widely distributed vestimentiferan tubeworms. Appl. Environ. Microbiol. 66, 651-658.

Dunlap, P.V., 2009. Bioluminescence, microbial. In: Schaechter, M. (Ed.), Encyclopedia of microbiology. Elsevier, Oxford, pp. 45-61.

Dunlap, P.V., Ast, J.C., Kimura, S., Fukui, A., Yoshino, T., Endo, H., 2007. Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbioses. Cladistics 23, 507-532.

Fidopiastis, P.M., Miyamoto, C.M., Jobling, M.G., Meighen, E.A., Ruby, E.G., 2002. LitR, a new transcriptional activator in *Vibrio fischeri*, regulates luminescence and symbiotic light organ colonization. Mol. Microbiol. 45, 131-143.

Funk, D.J., Wernegreen, J.J., Moran, N.A., 2001. Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-buchnera association. Genetics 157, 477-489.

Harmer, T.L., Rotjan, R.D., Nussbaumer, A.D., Bright, M., Ng, A.W., DeChaine, E.G., Cavanaugh, C.M., 2008. Free-living tube worm endosymbionts found at deep-sea vents. Appl. Environ. Microbiol. 74, 3895-3898.

Harvey, E.N., 1922. The production of light by the fishes *Photoblepharon* and *Anomalops*. Carnegie Institute of Washington, Washington, D.C.

Hasegawa, H., Hase, C.C., 2009. TetR-Type Transcriptional Regulator VtpR Functions as a Global Regulator in *Vibrio tubiashii*. Appl. Environ. Microbiol. 75, 7602-7609.

Hastings, J.W., Nealson, K., 1981. The symbiotic luminous bacteria. In: Starr, M.P., Stolp, H., Truper, H.G., Balows, A., Schlegel, H.G. (Eds.), The Prokaryotes: a Handbook on Habitats, Isolation, and Identification of Bacteria. Springer-Verlag, Berlin, pp. 1332-1345.

Haygood, M.G., 1990. Relationships of the luminous bacterial symbiont of the Caribbean flashlight fish, *Kryptophanaron alfredi* (family Anomalopidae) to other luminous bacteria based on bacterial luciferase (*luxA*) genes. Arch. Microbiol. 154, 496-503.

Haygood, M.G., 1993. Light organ symbioses in fishes. Crit. Rev. Microbiol. 19, 191-216.

Haygood, M.G., Distel, D.L., 1993. Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. Nature 363, 154-156.

Haygood, M.G., Tebo, B.M., Nealson, K.H., 1984. Luminous bacteria of a monocentrid fish (*Monocentris japonicus*) and 2 anomalopid fishes (*Photoblepharon palpebratus* and *Kryptophanaron alfredi*) - populations sizes and growth within the light organs, and rates of release into the seawater. Marine Biology 78, 249-254.

Henke, J.M., Bassler, B.L., 2004a. Quorum sensing regulates type III secretion in Vibrio harveyi and *Vibrio parahaemolyticus*. J. Bacteriol. 186, 3794-3805.

Henke, J.M., Bassler, B.L., 2004b. Three parallel quorum-sensing systems regulate gene expression in Vibrio harveyi. J. Bacteriol. 186, 6902-6914.

Herring, P.J., Morin, J.G., 1978. Bioluminescence in fishes. In: Herring, P.J. (Ed.), Bioluminescence in Action. Academic Press, London.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Jaques, S., McCarter, L.L., 2006. Three new regulators of swarming in *Vibrio parahaemolyticus*. J. Bacteriol. 188, 2625-2635.

Jobling, M.G., Holmes, R.K., 1997. Characterization of hapR, a positive regulator of the *Vibrio cholerae* HA protease gene hap, and its identification as a functional homologue of the *Vibrio harveyi* luxR gene. Mol. Microbiol. 26, 1023-1034.

Johnson, G.D., Rosenblatt, R.H., 1988. Mechanisms of light organ occlusion in flashlight fishes, family Anomalopidae (Teleostei, Beryciformes), and the evolution of the group. Zoological Journal of the Linnean Society 94, 65-96.

Johnson, G.D., Seeto, J., Rosenblatt, R.H., 2001. *Parmops echinatus*, a new species of flashlight fish (Beryciformes : Anomalopidae) from Fiji. Proceedings of the Biological Society of Washington 114, 497-500.

Kaeding, A.J., Ast, J.C., Pearce, M.M., Urbanczyk, H., Kimura, S., Endo, H., Nakamura, M., Dunlap, P.V., 2007. Phylogenetic diversity and cosymbiosis in the bioluminescent symbioses of "*Photobacterium mandapamensis*". Appl. Environ. Microbiol. 73, 3173-3182.

Kasai, S., Okada, K., Hoshino, A., Iida, T., Honda, T., 2007. Lateral transfer of the lux gene cluster. J. Biochem. (Tokyo). 141, 231-237.

Kessel, M., 1977. The ultrastructure of the relationship between the luminous organ of the teleost fish *Photoblepharon palpebratus* and its symbiotic bacteria. Cytobiologie 15, 145-158.

Lee, D.H., Jeong, H.S., Jeong, H.G., Kim, K.M., Kim, H., Choi, S.H., 2008. A consensus sequence for binding of SmcR, a *Vibrio vulnificus* LuxR homologue, and genome-wide identification of the SmcR regulon. J. Biol. Chem. 283, 23610-23618.

Leisman, G., Cohn, D.H., Nealson, K.H., 1980. Bacterial origin of luminescence in marine animals. Science 208, 1271-1273.

Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., Bassler, B.L., 2004. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell 118, 69-82.

Lin, J.W., Chao, Y.F., Weng, S.F., 1996. Nucleotide sequence and functional analysis of the *luxE* gene encoding acyl-protein synthetase of the lux operon from *Photobacterium leiognathi*. Biochem. Biophys. Res. Commun. 228, 764-773.

Lin, J.W., Yu, K.Y., Chao, Y.F., Weng, S.F., 1995. The *lumQ* gene is linked to the *lumP* gene and the lux operon in *Photobacterium leiognathi*. Biochem. Biophys. Res. Commun. 217, 684-695.

Maddison, W.P., Maddison, D.R., 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Mass.

McCosker, J.E., Rosenblatt, R.H., 1987. Notes on the biology, taxonomy, and distribution of flashlight fishes (Beryciformes, Anomalopidae). Japanese Journal of Ichthyology 34, 157-164.

Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresee, J.S., Shapiro, C., Griffin, P.M., Tauxe, R.V., 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. 5, 607-625.

Miller, M.B., Bassler, B.L., 2001. Quorum sensing in bacteria. Annu. Rev. Microbiol. 55, 165-199.

Mira, A., Moran, N.A., 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microb. Ecol. 44, 137-143.

Miyamoto, C.M., Dunlap, P.V., Ruby, E.G., Meighen, E.A., 2003. LuxO controls *luxR* expression in *Vibrio harveyi*: evidence for a common regulatory mechanism in *Vibrio*. Mol. Microbiol. 48, 537-548.

Miyamoto, C.M., Sun, W.Q., Meighen, E.A., 1998. The LuxR regulator protein controls synthesis of polyhydroxybutyrate in *Vibrio harveyi*. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1384, 356-364.

Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873-2878.

Morin, J.G., Harrington, A., Nealson, K., Krieger, N., Baldwin, T.O., Hastings, J.W., 1975. Light for all reasons - versatility in behavioral repertoire of flashlight fish. Science 190, 74-76.

Nyholm, S.V., McFall-Ngai, M.J., 2004. The winnowing: Establishing the squid-*Vibrio* symbiosis. Nature Reviews Microbiology 2, 632-642.

Pagel, M., 1997. Inferring evolutionary processes from phylogenies. Zoologica Scripta 26, 331-348.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Price, T., 1997. Correlated evolution and independent contrasts. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 352, 519-529.

Reen, F.J., Almagro-Moreno, S., Ussery, D., Boyd, E.F., 2006. The genomic code: inferring *Vibrionaceae* niche specialization. Nature Reviews Microbiology 4, 697-704.

Rosenblatt, R.H., Johnson, G.D., 1991. *Parmops coruscans*, a new genus and species of flashlight fish (Beryciformes, Anomalopidae) from the south Pacific. Proc. Biol. Soc. Wash. 104, 328-334.

Spaulding, A.W., von Dohlen, C.D., 1998. Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera : Sternorrhyncha). Mol. Biol. Evol. 15, 1506-1513.

Sreevatsan, S., Escalante, P., Pan, X., Gillies, D.A., Siddiqui, S., Khalaf, C.N., Kreiswirth, B.N., Bifani, P., Adams, L.G., Ficht, T., Perumaalla, V.S., Cave, M.D., VanEmbden, J.D.A., Musser, J.M., 1996. Identification of a polymorphic nucleotide in oxyR specific for *Mycobacterium bovis*. J. Clin. Microbiol. 34, 2007-2010.

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599-607.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

Thompson, F.L., Gevers, D., Thompson, C.C., Dawyndt, P., Naser, S., Hoste, B., Munn, C.B., Swings, J., 2005. Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. Appl. Environ. Microbiol. 71, 5107-5115.

Thompson, F.L., Iida, T., Swings, J., 2004. Biodiversity of vibrios. Microbiol. Mol. Biol. Rev. 68, 403-+.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. ClustalW - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matric choice. Nucleic Acids Res. 22, 4673-4680.

Tu, K.C., Waters, C.M., Svenningsen, S.L., Bassler, B.L., 2008. A small-RNA-mediated negative feedback loop controls quorum-sensing dynamics in *Vibrio harveyi*. Mol. Microbiol. 70, 896-907.

Ulitzur, S., Dunlap, P.V., 1995. Regulatory circuitry controlling luminescence autoinduction in *Vibrio fischeri*. Photochem. Photobiol. 62, 625-632.

Urbanczyk, H., Ast, J.C., Kaeding, A.J., Oliver, J.D., Dunlap, P.V., 2008. Phylogenetic analysis of the incidence of lux gene horizontal transfer in *Vibrionaceae*. J. Bacteriol. 190, 3494-3504.

Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., Kachur, S.M., Leadem-Dougherty, R.R., Rhoton, S.D., Zinser, G., Farlow, J., Coker, P.R., Smith, K.L., Wang, B.X., Kenefic, L.J., Fraser-Liggett, C.M., Wagner, D.M., Keim, P., 2007. Global Genetic Population Structure of *Bacillus anthracis*. Plos One 2.

Visick, K.L., Ruby, E.G., 2006. *Vibrio fischeri* and its host: it takes two to tango. Curr. Opin. Microbiol. 9, 632-638.

Wada, M., Azuma, N., Mizuno, N., Kurokura, H., 1999. Transfer of symbiotic luminous bacteria from parental *Leiognathus nuchalis* to their offspring. Marine Biology 135, 683-687.

Waters, C.M., Bassler, B.L., 2005. Quorum sensing: Cell-to-cell communication in bacteria. Annu. Rev. Cell Dev. Biol. 21, 319-346.

Wernegreen, J.J., 2002. Genome evolution in bacterial endosymbionts of insects. Nature Reviews Genetics 3, 850-861.

Wernegreen, J.J., Moran, N.A., 1999. Evidence for genetic drift in endosymbionts (Buchnera): Analyses of protein-coding genes. Mol. Biol. Evol. 16, 83-97.

Wolfe, C.J., Haygood, M.G., 1991. Restriction fragment length polymorphism analysis reveals high levels of genetic divergence among the light organ symbionts of flashlight fish. Biol. Bull. 181, 135-143.

Woolfit, M., Bromham, L., 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20, 1545-1555.

Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821-829.

Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas, Austin, Texas.

CHAPTER III

Genome reduction and host dependence in a luminous symbiont.

Abstract

Genome reduction by extensive gene loss has occurred independently in numerous host-restricted bacterial lineages and has lead to considerable interest in how this pattern arises. Vertical transmission between host generations is thought to play a critical role in the process of genome reduction (McCutcheon and Moran, 2012). In contrast to vertically transmitted lineages, environmentally acquired, facultatively symbiotic bacteria inhabit multiple environments and often have relatively large genomes reflecting adaptation to diverse habitats. Here we demonstrate obligate host dependence and genome reduction in the extracellular, environmentally acquired luminous symbiont of an anomalopid flashlight fish. Consistent with patterns of genome reduction in host-restricted symbionts, the genome of the anomalopid symbiont '*Candidatus* Photodesmus katoptron' (Hendry and Dunlap, 2011) is one fifth the size of relatives' genomes and lacks most genes necessary for amino acid synthesis and for metabolism of energy sources other than glucose. This gene loss and resulting inability to synthesize essential nutrients demonstrates the first example of an obligate relationship between a luminous symbiont and its host. '*Candidatus* Photodesmus katoptron' is the first known symbiont to evolve obligate dependence and genome reduction in the absence of an intracellular

phase or vertical transmission. Therefore, direct vertical transmission and intracellular lifestyles are not necessary for the evolution of obligate dependence and genome reduction in symbionts.

Introduction

Host-restricted bacterial lineages, such as some insect endosymbionts, demonstrate that obligate dependence on a host can have significant evolutionary consequences for symbionts. Many insect endosymbionts have severely reduced genomes, increased evolutionary rates, and high AT substitution biases compared to relatives (Woolfit and Bromham, 2003; Wernegreen and Funk, 2004; McCutcheon and Moran, 2012). High levels of genetic drift in small, non-recombining symbiont populations cause these patterns. Restriction to the host environment decreases selection for the maintenance of genes needed only outside the host. Additionally, population bottlenecks during vertical transmission between host generations decrease the effective population size of symbiont lineages, thus amplifying the effect of drift (Moran, 1996; Wernegreen and Moran, 1999; Mira and Moran, 2002). In contrast, environmentally acquired symbionts, such as facultatively symbiotic luminous bacteria of the family *Vibrionaceae*, live in a variety of habitats and maintain large population sizes (Dunlap et al., 2007). The genomes of luminous symbionts are very similar in size and gene content to free-living, non-symbiotic relatives. Here we report that the unculturable luminous symbiont, '*Ca.* Photodesmus katoptron,' is an exception to this pattern and has a much smaller genome than relatives.

Results and discussion

The *de novo* assembled '*Ca.* Photodesmus katoptron' genome sequence is greatly reduced in size although its structure and content are homologous with relatives. The 1Mb '*Ca.* Photodesmus katoptron' genome is approximately one fifth the size of other *Vibrionaceae* genomes (mean: 4.8Mb, range: 3.9M - 6.1Mb, Fig. 3.1). As with other *Vibrionaceae* genomes, the '*Ca.* Photodesmus katoptron' genome appears to consist of two chromosomes and one plasmid (Fig. 3.2). The extent of homology with relatives is much higher for chromosome I than for chromosome II or the plasmid. This pattern is similar to comparisons between other *Vibrionaceae* species (Reen et al., 2006) and is probably exaggerated in '*Ca*. Photodesmus katoptron' because of the symbiont's high level of genetic divergence (Hendry and Dunlap, 2011). The '*Ca*. Photodesmus katoptron' genome is predicted to contain 916 genes, most of which are orthologous with close relatives, whereas sequenced *Vibrionaceae* genomes range from 3839 to 6237 genes (Fig. 3.1). The high read coverage depth (Methods) and apparent completeness (Appendix 1) of the '*Ca*. Photodesmus katoptron' genome indicate that this low number of genes is not due to insufficient sequencing. It is also unlikely that the '*Ca*. Photodesmus katoptron' genome represents that ancestral genome size of the family, as phylogenetic analyses do not resolve anomalopid symbionts as basal within the group (Fig. 3.3) (Hendry and Dunlap, 2011). The most parsimonious explanation for the small '*Ca*. Photodesmus katoptron' genome is that it has been reduced in size by gene loss. Such gene loss is commonly found in obligate symbionts and is consistent with obligate dependence of the bacterium on the host. Additionally, the AT content of the '*Ca.*

Photodesmus katoptron' genome (69%), is higher than the range typically found in *Vibrionaceae* (55-59%), another similarity to obligate endosymbionts.

We compared '*Ca.* Photodesmus katoptron' gene content to strains of two representative *Vibrionaceae* species, *Aliivibrio fischeri*, a facultatively symbiotic luminous species and *Vibrio campbellii* (previously classified as a strain of *V. harveyi*) (Lin et al., 2010), a non-symbiotic luminous species. In comparison to these relatives, '*Ca.* Photodesmus katoptron' lacks genes from all functional categories analyzed (Fig. 3.4). Categories with the smallest amount of loss were those likely to be necessary regardless of symbiotic state, such as genes involved in transcription, translation, and DNA synthesis (Fig. 3.4). In contrast, metabolic categories of genes, including amino acid synthesis and energy metabolism genes, are some of the most reduced. Loss of metabolic genes is common in obligate symbionts, as hosts provide nutrients to the bacteria, which allows the genes involved in the synthesis of these nutrients to be lost (McCutcheon and Moran, 2012). Based on which pathways have lost genes, it is possible to infer which necessary nutrients are provided to the bacteria by the fish.

Genes required for amino acid synthesis show one of the most dramatic reductions compared to free living relatives (Fig. 3.4, Appendix 2), indicating that '*Ca.* Photodesmus katoptron' cannot synthesize a full complement of amino acids and must acquire them from its host. Only three amino acid synthesis pathways appear to be complete in '*Ca.* Photodesmus katoptron' (Fig. 3.5). Of the remaining amino acid synthesis genes, 13 of 18 are found in pathways that produce intermediate compounds used in other synthetic pathways for which '*Ca.* Photodesmus katoptron' has maintained genes (Fig. 3.5), suggesting they have likely been maintained for reasons other than

amino acid synthesis. This gene loss contrasts with other *Vibrionaceae*, which can synthesize a full complement of amino acids, and it is striking compared to obligate and intracellular bacteria, which can typically synthesize three or more amino acids, even if they do not provide amino acids to the host (Yu et al., 2009). The inability to synthesize amino acids implies that the fish host must be providing these nutrients to its symbiont.

Energy metabolism genes are reduced in '*Ca.* Photodesmus katoptron' compared to relatives (Fig. 3.4). '*Ca.* Photodesmus katoptron' has maintained complete pathways for glycolysis and the citric acid cycle, the genes for all necessary components of ATP synthase, and numerous genes involved in electron transport (Appendix 1, Appendix 2). Compared to *A. fischeri* and *V. campbellii*, '*Ca.* Photodesmus katoptron,' however, has lost many genes necessary to catabolize amino acids and many carbohydrates besides glucose (Appendix 2). Additionally, the '*Ca.* Photodesmus katoptron' genome contains only one set of phosphotransferase system (PTS) genes, which are used to transport sugar sources (Appendix 2). Based on homology these PTS genes are specific to glucose. Typically members of *Vibrionaceae* are metabolically diverse and can utilize many carbon/energy sources (Reen et al., 2006). The loss of energy metabolism genes and lack of other sugar transport genes suggests that '*Ca.* Photodesmus katoptron' can only utilize glucose as a carbon/energy source, indicating that the host is providing glucose to the symbiont. The dependence of '*Ca.* Photodesmus katoptron' on glucose and amino acids is strong evidence for an obligate relationship with the host; habitats other than the host light organ are unlikely to provide a consistent supply of these nutrients. This evidence of nutrient dependence in '*Ca.* Photodesmus katoptron' provides the first indication of what nutrients symbiotically luminous fish provide to their symbiotic bacteria. These nutrients

are different and more energy rich than those that are provided to *A. fischeri* by the squid *E. scolopes*, the only other bioluminescent symbiosis for which data are available (Wier et al., 2010). The extent of nutrient provisioning by the host has implications for the evolution of dependence; obligate dependence is more likely to evolve when the host environment is more nutrient dense than other possible environments. We speculate that high nutrient provisioning of '*Ca.* Photodesmus katoptron' by the host may have allowed for initial gene loss and initiation of obligate dependence.

The gene content of the '*Ca.* Photodesmus katoptron' genome is more similar to oligotrophic (adapted to stable, low nutrient conditions) or intracellular bacteria than to copiotrophic (adapted to variable, high nutrient conditions) relatives (Fig. 3.6). We compared the gene content of '*Ca.* Photodesmus katoptron' with the genomes of various species representing different ecological lifestyles: obligate endosymbionts, facultative copiotrophic symbionts, non-symbiotic copiotrophs, and non-symbiotic oligotrophs. Regardless of phylogenetic relationships, taxa grouped together based on their characterization as obligate symbionts, oligotrophs, or copiotrophs. '*Ca.* Photodesmus katoptron' was more similar in gene content to unrelated obligate symbionts or free-living oligotrophs than to close relatives, both symbiotic and not. Specifically, the '*Ca.* Photodesmus katoptron' genome displays characteristics linked to slow growth in a stable environment. These traits include a low number of rRNA operons, a small number of transport proteins, and low numbers of regulatory and signal transduction genes (Lauro et al., 2009) (Table 3.1). These signatures of adaptation to a stable environment are predicted to occur if '*Ca.* Photodesmus katoptron' exists predominantly in the host light organ rather than in multiple habitats. Further evidence that '*Ca.* Photodesmus katoptron'

may grow exclusively within the protected light organ environment comes from the complete lack of phage associated DNA, such as prophages or CRISPR sequences, which are found in the genomes of other *Vibrionaceae* species. A lack of phage associated DNA is more typical of intracellular than extracellular bacteria (Table 3.1) (McCutcheon and Moran, 2012) and suggests that '*Ca.* Photodesmus katoptron' rarely encounters bacteriophages. Additional support for restriction to a stable environment comes from the types of cell surface proteins found in '*Ca.* Photodesmus katoptron,' which also has implications for the way the symbiont interacts with the fish host. Both facultative and obligate symbionts, including *A. fischeri*, use surface binding pilin proteins and secretion proteins for transfer to new hosts[14-16] (Dale et al., 2001; Nyholm et al., 2004; Ruby et al., 2005). '*Ca.* Photodesmus katoptron', however, has lost all pilin genes found in relatives as well as many secretion systems (Appendix 2). The lack of pilin genes implies that not only is '*Ca.* Photodesmus katoptron' interacting with a limited number of surface types and organisms, but also that it might use novel means of colonizing new host generations.

Certain groups of genes likely to be used outside of the host, those needed for chemotaxis, motility, and for production of the cell envelope, remain mostly intact in '*Ca.* Photodesmus katoptron' (Appendix 2). Chemotaxis and motility genes are unlikely to be used within the light organ due to the high density of bacterial cells (Kessel, 1977). Additionally, '*Ca.* Photodesmus katoptron' has retained many genes involved in DNA repair and recombination, which are frequently lost in endosymbionts (Dale et al., 2003). The maintenance of genes needed outside the host is consistent with the bacteria persisting at least briefly outside the host. Haygood *et al.* previously showed that symbionts are regularly released from pores on anomalopid light organ surfaces and

remain luminous for several hours (Haygood et al., 1984). However, the high likelihood that they cannot grow outside of the host indicates that chemotaxis might be used to search for uncolonized light organs rather than for other environments. The exact mechanisms of how anomalopid symbionts are transmitted between host generations are not known, but evidence demonstrates that transfer is environmental rather than vertical (Haygood, 1993). An intriguing possibility is that the anomalopid behavior of aggregating in caves during the day (Morin et al., 1975) may allow their symbionts to build up higher density in the water and infect larval fish without having to disperse or persist for extended periods in open water (Meyer-Rochow, 1976). This scenario would ensure that symbionts regularly encounter new hosts and it could lessen the selection on symbiont genes needed in non-host environments.

This study represents the first example of an obligate luminous symbiont and one of only three extracellular symbionts shown to have the same extent of evolutionary change common to intracellular obligate symbionts (Kikuchi et al., 2009; Hosokawa et al., 2010). Because they have many characteristics similar to well-studied endosymbionts, including genome reduction, gene content, high AT content, high evolutionary rate[2], and reduced within-host polymorphism (Hendry and Dunlap, 2011), the evolutionarily distinct anomalopid symbionts can provide an independent test of why these patterns are found in obligate symbionts. This study indicates that vertical transmission or an intracellular life cycle phase are not necessary for obligate dependence and genome reduction to occur. Additionally, our findings suggest the hypothesis that host behavior and high nutrient provisioning of the symbiont by the host contributed to the genome reduction in this system.

Methods

Sequencing and assembly

DNA for genome sequencing came from five light organs of four fish specimens of *Anomalops katoptron* caught near the island of Cebu in the Philippines. DNA was extracted as in Hendry & Dunlap (Hendry and Dunlap, 2011) and combined for sequencing. Sequencing was performed by the DNA Sequencing Core at the University of Michigan on half a plate of a Roche 454FLX Genome Sequencer. Reads were assembled in MIRA3 (Chevreux et al., 2002) into 29 contigs with at least 70X coverage and an average coverage of 165X. At this coverage cut-off all contigs represented '*Ca*. Photodesmus katoptron' sequences rather than host or contaminant sequences (based on BLAST searches and nucleotide signatures). Six low coverage contigs were determined by BLAST to be slightly divergent versions of larger contigs with frame shift sequencing errors and were thrown out. Seventeen contigs were then assigned by BLAST alignments into two large scaffolds and one small. These scaffolds were compared to chromosomes and plasmids of the related species *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116) using the Artemis Comparison Tool (Carver et al., 2005) (Fig. 3.2). Nucleotide sequences were compared using an e value cutoff of 1.0. This cutoff is high enough to find significant homology between genera. The extent of homology did not increase when predicted amino acid sequence was compared. The largest scaffold was found to have similar levels of sequence identity with chromosome I from other members of *Vibrionaceae* as is found between *Vibrionaceae* species, and so is assumed to be homologous. The same pattern was seen for the smallest scaffold and *Vibrio* plasmids.

The remaining large scaffold is assumed to be homologous with *Vibrionaceae* chromosomes II, which typically have low sequence identity between species (Reen et al., 2006). Five small contigs consisting of rRNA and tRNA genes and one contig containing an rRNA gene and several protein coding genes could not be assigned to scaffolds (Table 3.2). This is likely due to conversion between rRNA operons, which also made it impossible to determine the exact number of rRNA genes or to assemble many of them into scaffolds. The high level of sequencing coverage as well as the number of intact conserved pathways and genes (Appendix 1) (Raes et al., 2007) indicates that the entire genome of '*Ca.* Photodesmus katoptron' is represented by these sequences. To ensure that no low coverage contigs were missed, contigs of greater than 2500 bases in length but less than 70X coverage were checked by BLAST and found to be contaminant sequences of host DNA or distantly related bacterial DNA. As quality control, all contigs (both greater than and less than 70X coverage) were checked for similarity of genomic nucleotide signatures to each other versus relatives (*Escherichia coli* K12, *Vibrio vulnificus* CMCP6, *Photobacterium profundum* SS9, *V. campbellii* ATCC BAA-1116, *Aliivibrio fischeri* MJ11, *Vibrio splendidus* LGP32, and *Vibrio cholerae* MJ 1236) (Dick et al., 2009). All greater than 70X contigs mapped to the *'Ca. Photodesmus katoptron'* genome except rRNA and plasmid sequences, which typically have different nucleotide signatures than other areas of a genome. The '*Candidatus* Photodesmus katoptron' genome has been deposited in GenBank under the accession PRJNA80863.

Annotation and comparative analyses

The '*Ca*. Photodesmus katoptron' genome was annotated by the IGS Annotation Engine at the University of Maryland. The annotation was further checked manually. Because 454 sequencing generates many insertion/deletion errors, genes were occasionally split into multiple ORFs; these were all manually corrected. Genes were assigned into TIGRfams categories of genes by the IGS Annotation Engine and these assignments were also checked manually. As many genes as possible in the categories of unclassified, unknown function, and hypothetical proteins were placed in functional categories based on the TIGRfam database. Additionally, all genes already placed in functional categories were checked for accuracy of placement against the TIGRfam database. Genes of unknown function account for 21% of the '*Ca.* Photodesmus katoptron' genome.

For comparisons of gene content, genes classified into TIGRfams were obtained from genomes in the Integrated Microbial Genomes (IMG) database of the DOE Joint Genome Institute. For *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116) each gene in every category was checked for homology to genes in '*Ca.* Photodesmus *katoptron'* categories. When a gene was found to be present in the '*Ca.* Photodesmus katoptron' genome and not in the TIGRfam genes of *A. fischeri* or *V. campbellii*, the genomes of these species were checked for the gene by BLAST. It is therefore possible that the *A. fischeri* or *V. campbellii* genomes actually contain more possible TIGRfam genes than shown here, but it is not possible a gene is present in '*Ca.* Photodesmus katoptron' but missed in the relatives analyzed.

For hierarchical clustering analysis genes in TIGRfams were obtained from IMG for the following species: *A. fischeri* ES114, '*Candidatus* Baumannia cicadellicola' Hc, '*Candidatus* Blochmannia pennsylvanicus' BPEN, *Buchnera aphidicola* APS, *E. coli* K12 DH1, '*Candidatus* Pelagibacter ubique' SAR11 HTCC1062, *P. profundum* SS9, *Roseobacter denitrificans* OCh 114, '*Candidatus* Ruthia magnifica' Cm, *V. cholerae* O395, *V. campbellii* ATCC BAA-1116, *Wigglesworthia glossinidia*, and *Yersinia pestis* KIM 10. Genes present in the '*Ca.* Photodesmus katoptron' genome were then added to this matrix. Genes present in only one taxon were eliminated. Genes were coded by copy number as absent (0), present with 1-5 copies (1), present with 6-11 copies (2), or present with >11 copies (3). These characters were considered ordered. An unrooted dendogram was constructed using the neighbor joining method in PAUP. Other hierarchical clustering methods were also tried, including k-means and average linkage clustering and topology shown in Fig. 3.6 was robust.

Acknowledgements

This work will be coauthored with Jeffrey R. de Wet and Paul V. Dunlap when published. Specimens were obtained with assistance from LiveAquaria.com, Quality Marine Wholesale Suppliers, and Fish Doctors Aquarium of Ypsilanti, Michigan. Genome annotation was performed by the Annotation Engine at the Institute for Genome Sciences, University of Maryland School of Medicine. We thank Katherine Dougan for assistance manually checking gene annotations. Analyzing sequence reads for genomic nucleotide signatures was done by Sunit Jain and we thank him and Greg J. Dick for assistance. Computational assistance for hierarchical clustering analysis was provided by Susanna M. Messinger at the University of Michigan. We thank Greg J. Dick at the University of Michigan for discussions about the paper. This project was funded by Department of Ecology and Evolutionary Biology Block Grants from the University of Michigan and by Rackham Graduate Student Research Grants from the University of Michigan Rackham Graduate School.

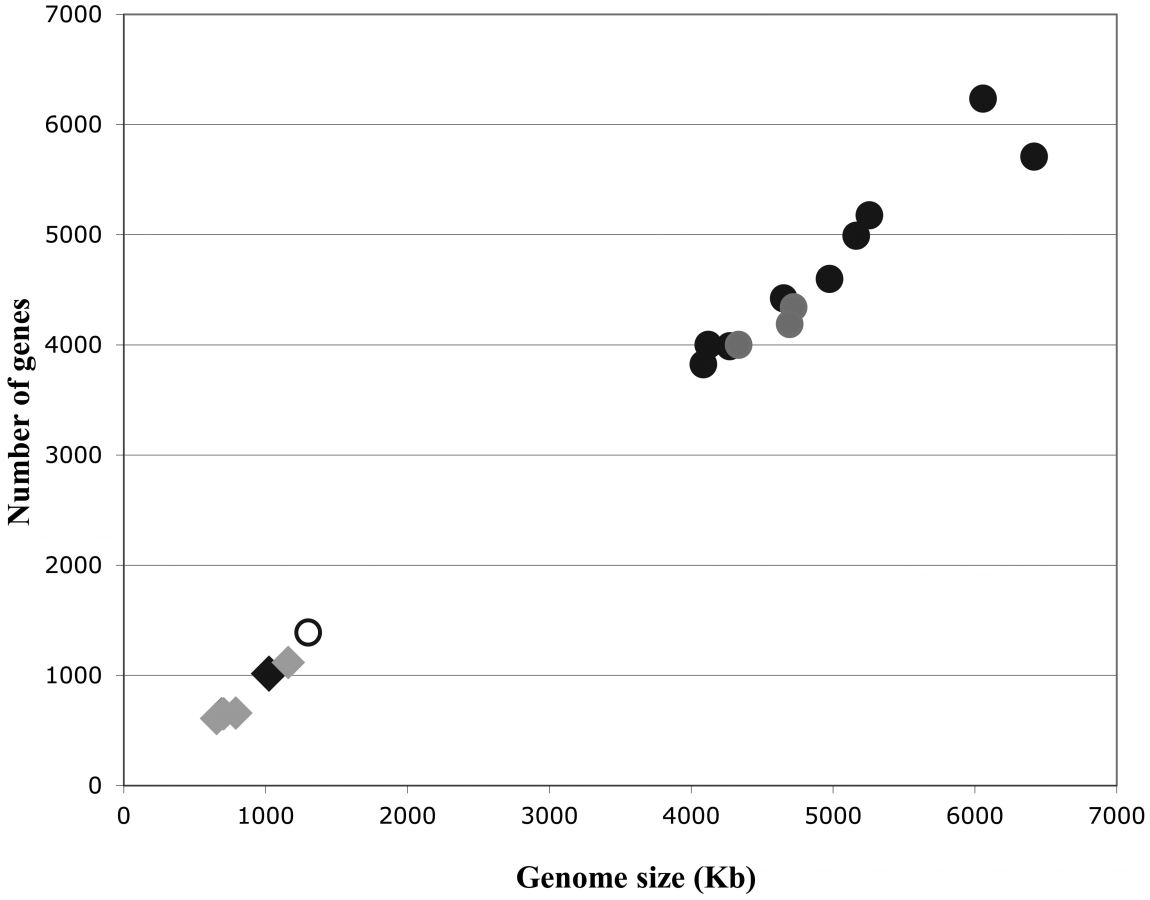

Fig. 3.1. Graph of genome size versus number of genes. '*Candidatus* Photodesmus katoptron' (black diamond) has a genome similar in size to intracellular obligate symbionts ('*Candidatus* Baumannia cicadellicola' Hc, '*Candidatus* Blochmannia pennsylvanicus' BPEN, *Buchnera aphidicola* APS, '*Candidatus* Ruthia magnifica' Cm, and *Wigglesworthia glossinida*), shown in grey diamonds. '*Candidatus* Pelagibacter ubique' SAR11 HTCC1062, often considered to have the smallest genome of a free-living bacterium, is shown in the white circle. Free-living representatives of *Vibrionaceae* (*Aliivibrio fischeri* ES114, *Aliivibrio salmonicida* LF11238, *Photobacterium profundum* SS9, *Vibrio anguillarum* 775, *Vibrio campbellii* ATCC BAA-1116, *Vibrio cholera* O395, *Vibrio parahaemolyticus* RIMD 2210633, *Vibrio splendidus* LGP32, and *Vibrio vulnificus* YJ016) are show in black circles. Other free-living bacteria (*Escherichia coli* K12 DH1, *Roseobacter denitrificans* OCh 114, and *Yersinia pestis* KIM 10) are shown in grey circles.

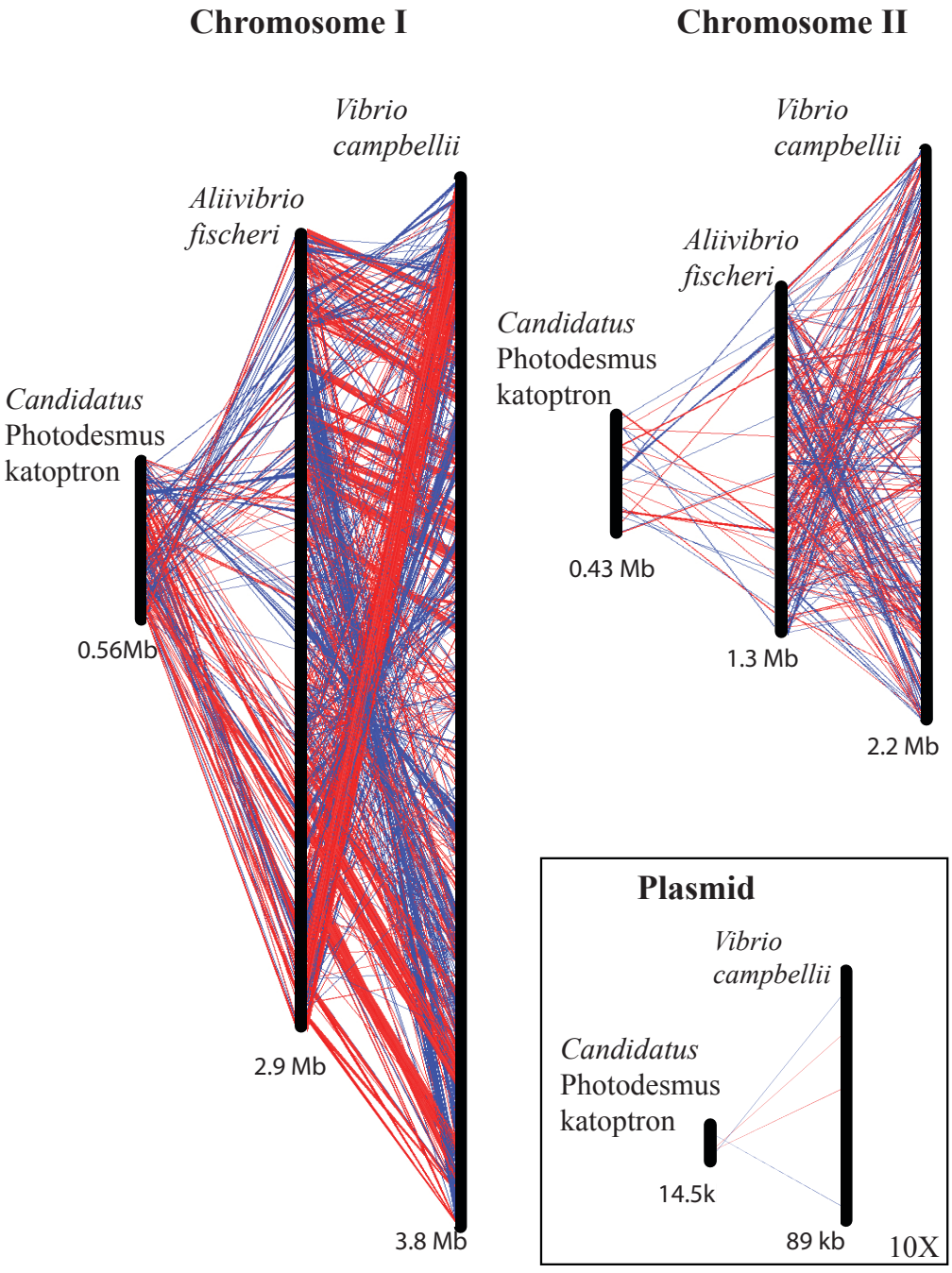

Fig. 3.2. Chromosome comparisons between '*Ca.* Photodesmus katoptron' and *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116). Chromosomes are drawn to scale with alignments as determined by the Artemis Comparison Tool. Each connecting line represents an area of high nucleotide sequence similarity; red lines indicate alignment of two positive strands and blue lines represent similarity between positive and negative strands. Homology was determined with an e value cut off of 1.0. Plasmid representations for '*Ca.* Photodesmus katoptron' and *V. campbellii* (ATCC BAA-1116) are shown on a different scale.

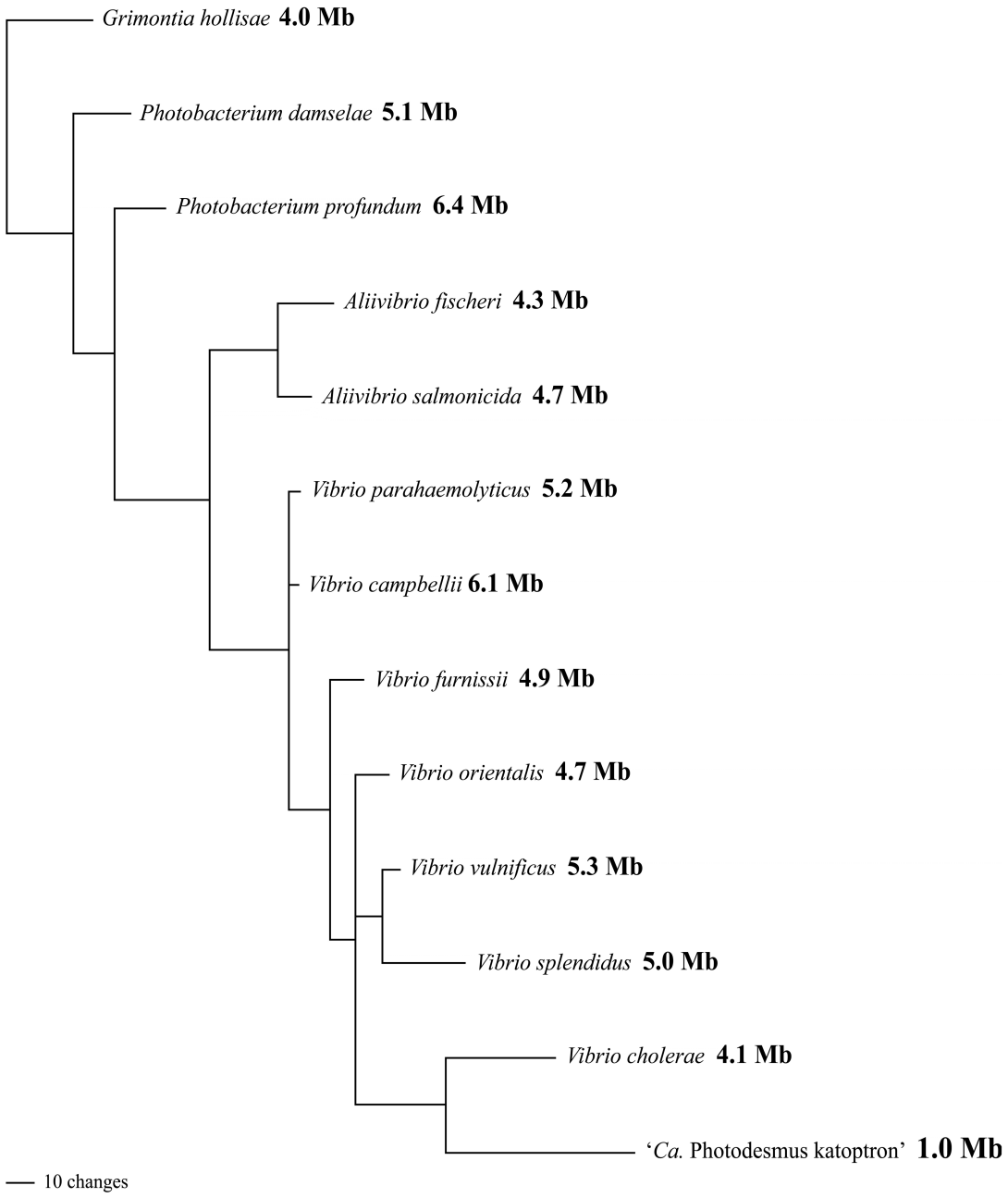

Fig. 3.3. Phylogenetic tree with genome sizes of '*Ca*. Photodesmus katoptron' and relatives. Sequences of the 16S rRNA gene and genome sizes for *Vibrionaceae* strains with complete genomes (one strain per species) were obtained from GenBank. Sequences were aligned in ClustalX[2], manually refined in MacClade[3] and then analyzed using maximum parsimony in PAUP[3]. The strains used were *Aliivibrio fischeri* ES114, *Aliivibrio salmonicida* LFI1238, *Grimontia hollisae* CIP 101886, *Photobacterium damselae* CIP 102761, *Photobacterium profundum* SS9, *Vibrio campbellii* ATCC BAA-1116, *Vibrio cholerae* O395, *Vibrio furnissii* NCTC 11218, *Vibrio orientalis* ATCC 33934, *Vibrio parahaemolyticus* RIMD 2210633, *Vibrio splendidus* LGP32, and *Vibrio vulnificus* YJ016.

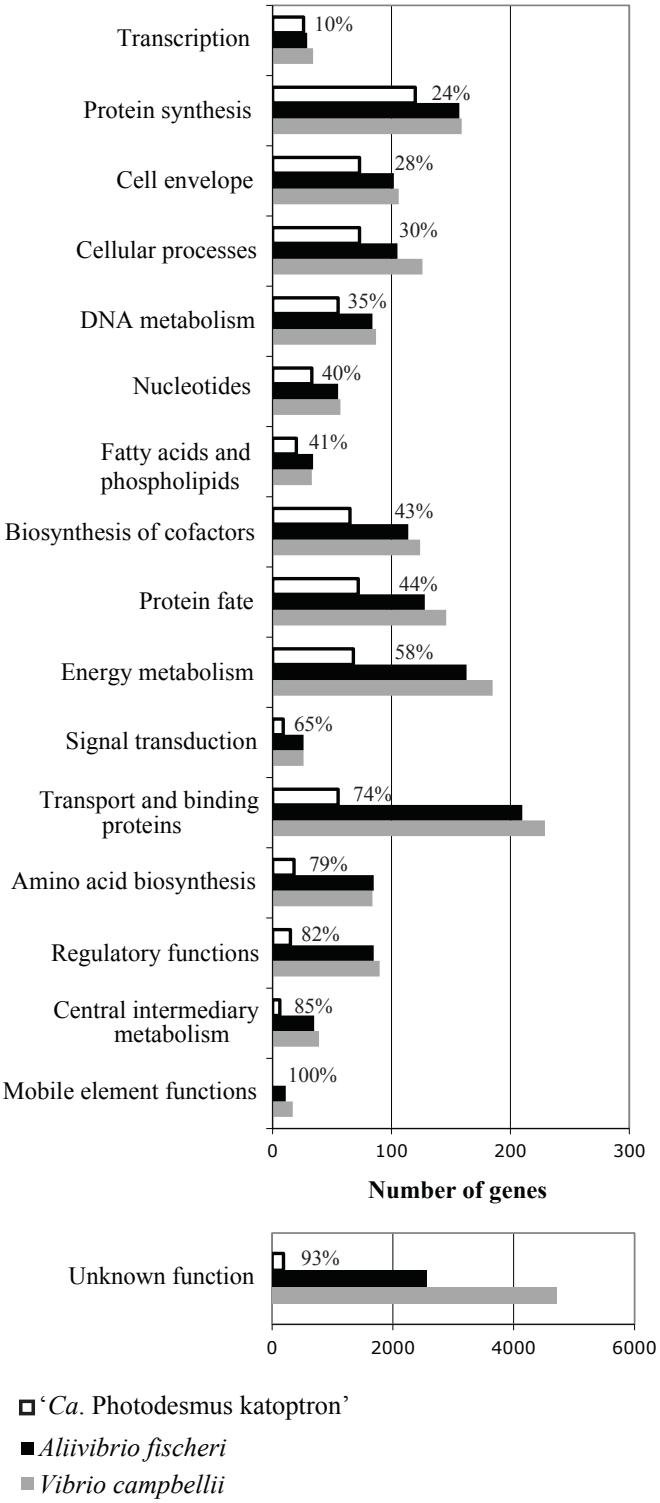

Fig. 3.4. Genes in TIGRfam functional categories for '*Ca.* Photodesmus katoptron' and *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116). Percentages show the amount of reduction in the '*Ca.* Photodesmus katoptron' genome compared to *A. fischeri*. Genes of unknown function are shown on a separate scale.

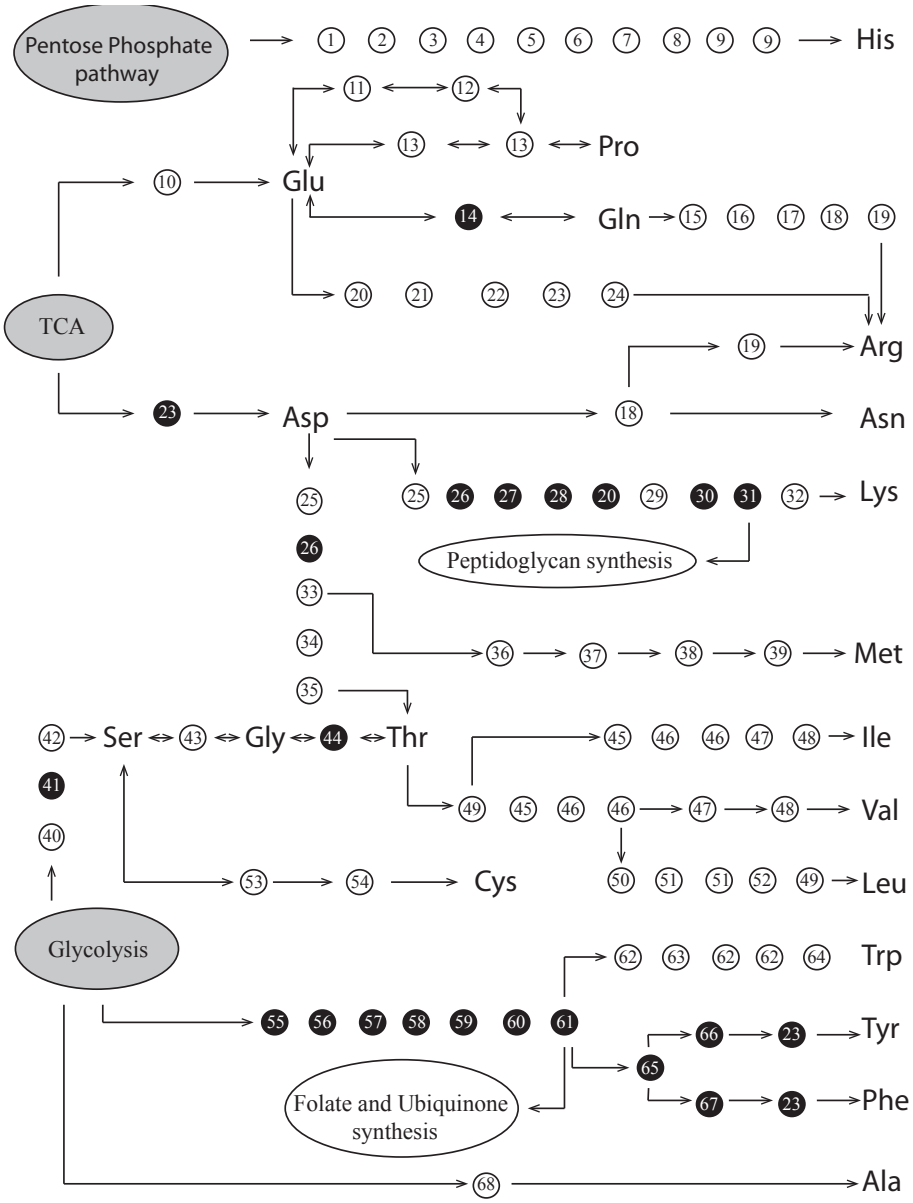

Fig. 3.5. Amino acid synthesis pathways showing gene loss in '*Ca.* Photodesmus katoptron.' Pathways for the synthesis of the 20 standard amino acids encoded by the universal genetic code were taken from the KEGG pathway database for taxa *A. fischeri* (ES114) and *V. campbellii* (ATCC BAA-1116). Each circle represents an enzymatic step. Numbers within circles correspond to KEGG pathway database EC numbers and can be found matched to gene names in the Appendix 2. Darkened circles represent enzymatic transformations that '*Ca.* Photodesmus katoptron' is theoretically capable of carrying out based on gene presence, whereas white circles indicate missing enzymatic steps. Grey ovals show pathways that contribute starting products for amino acid synthesis. White ovals show other pathways that utilize amino acid synthesis genes. Pathways for which '*Ca.* Photodesmus katoptron' has maintained a majority of genes (synthesis of lysine, tyrosine, and phenylalanine) are also used in the synthesis of peptidoglycan, folate, and ubiquinone.

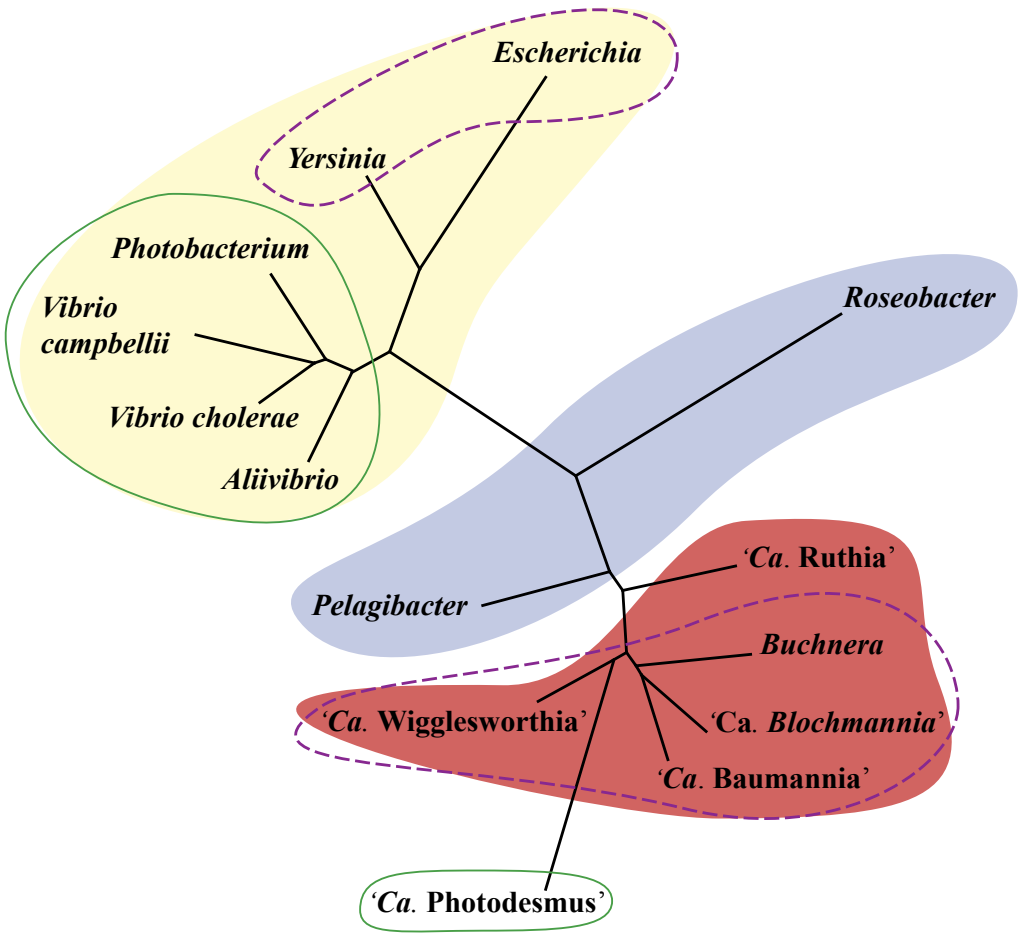

Fig. 3.6. Genomic hierarchical clustering dendogram showing that '*Ca.* Photodesmus katoptron' is more similar in gene content to distantly related obligate symbionts than it is to free-living close relatives. The dendogram is based on gene presence/absence for representative species. Shaded areas show bacteria with similar lifestyles (red: obligate intracellular symbionts; blue: free-living oligotrophs; and yellow: free-living copiotrophs, symbiotic and not). Outlines show phylogenetic relationships of taxonomically classified *Gammaproteobacteria* (purple dashed: family *Enterobacteriaceae*; green solid: family *Vibrionaceae*). The symbiont *Candidatus* Ruthia magnifica is an unclassified member of *Gammaproteobacteria*. Oligotrophic taxa, *Roseobacter* and *Pelagibacter*, are *Alphaproteobacteria* classified in separate families and distantly related to other taxa analyzed.

Table 3.1. Genome characteristics that vary with lifestyle compared to '*Ca.* Photodesmus katoptron'. Adapted from Lauro *et al*.[1] with values for intracellular bacteria based on data described in Fig. 3.6.

	Copiotrophs	Oligotrophs	Intracellular	Photodesmus
Genome size	Large	Small	Very small	Very small
rRNA operon copies	Many (7-15)[a]	Few (1)	Few (1)	Few (2-5)
Regulatory genes	Many	Few	Few	Few
Prophages	Many	Few	0	0
CRISPRs	Many	Few	0	0
Transport proteins	Many	Few	Few	Few

[a] Numbers of rRNA operons found in sequenced members of *Vibrionaceae*.

Table 3.2. Contig and assembly information showing size (bp) and read coverage depth of contigs by category.

Contig fate	#	Min. Size	Max. Size	Mean Size	Mean Cov.
Chromosome I	6	5891	351602	86738	240
Chromosome II	8	4362	153534	58629	137
Plasmid	3	3673	7885	5639	450
Redundant	6	1300	3429	1942	101
Unassigned	6	1018	5078	2171	157

References

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., Parkhill, J., 2005. ACT: the Artemis comparison tool. Bioinformatics 21, 3422-3423.

Chevreux, B., Pfisterer, T., Suhai, S., 2002. Automatic assembly and editing of genomic sequences. In: Suhai, S. (Ed.), *Genomics and Proteomics - Functional and Computational Aspects.* Kluwer Academic Publishers, New York, NY, 51-65.

Dale, C., Young, S. A., Haydon, D. T., Welburn, S. C., 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc. Natl. Acad. Sci. U. S. A. 98, 1883-1888.

Dale, C., Wang, B., Moran, N., Ochman, H., 2003. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. Mol. Biol. Evol. 20, 1188-1194.

Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P., Banfield, J. F., 2009. Community-wide analysis of microbial genome sequence signatures. Genome Biol. 10.

Dunlap, P. V., Ast, J. C., Kimura, S., Fukui, A., Yoshino T., Endo, H., 2007. Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbiosis. Cladistics. 23, 507-532.

Hendry, T. A., Dunlap, P. V., 2011 The uncultured luminous symbiont of *Anomalops katoptron* (Beryciformes: Anomalopidae) represents a new bacterial genus. Mol. Phylogenet. Evol. 61, 834-843.

Haygood, M. G., Tebo, B. M., Nealson, K. H., 1984. Luminous bacteria of a monocentrid fish (*Monocentris japonicus*) and 2 anomalopid fishes (*Photoblepharon palpebratus* and *Kryptophanaron alfredi*) - populations sizes and growth within the light organs, and rates of release into the seawater. Mar. Biol. 78, 249-254.

Haygood, M. G., 1993. Light organ symbioses in fishes. Crit. Rev. Microbiol. 19, 191-216.

Hosokawa, T., Kikuchi, Y., Nikoh, N., Meng, X.-Y., Hironaka, M., Fukatsu, T., 2010. Phylogenetic position and peculiar genetic traits of a midgut bacterial symbiont of the stinkbug *Parastrachia japonensis*. Appl. Environ. Microbiol. 76, 4130–4135.

Kessel, M., 1977. The ultrastructure of the relationship between the luminous organ of the teleost fish *Photoblepharon palpebratus* and its symbiotic bacteria. Cytobiologie. 15, 145-158.

Kikuchi, Y., Hosokawa, T., Nikoh, N., Meng, X.-Y., Kamagata, Y., Fukatsu, T., 2009. Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. BMC Biol. 7.

Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., DeMaere, M. Z., Ting, L., Ertan, H., Johnson, J., Ferriera, S., Lapidus, A., Anderson, I., Kyrpides, N., Munk, A. C., Detter, C., Han, C. S., Brown, M. V., Robb, F. T., Kjelleberg, S., Cavicchioli, R., 2009. The genomic basis of trophic strategy in marine bacteria. Proc. Natl. Acad. Sci. U. S. A. 106, 15527-15533.

Lin, B., Wang, Z., Malanoski, A. P., O'Grady, E. A., Wimpee, C. F., Vuddhakul, V., Alves Jr., N., Thompson, F. L., Gomez-Gil, B., Vora, G. J., 2010. Comparative genomic analyses identify the *Vibrio harveyi* genome sequenced strains BAA-1116 and HY01 as *Vibrio campbellii*. Environ. Microbiol. Rep. 2, 81-89.

Maddison, D. R., W. P. Maddison, 2002. MacClade 4: Analysis of phylogeny and character evolution. Version 4.0. Sinauer, Sunderland, Massachusetts.

McCutcheon, J. P., Moran, N. A., 2012, Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10, 13-26.

Meyer-Rochow, V. B., 1976. Some observations on spawning and fecundity in luminescence fish *Photoblepharon palpebratus*. Mar. Biol. 37, 325-328.

Mira, A., Moran, N. A., 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microb. Ecol. 44, 137-143.

Moran, N. A., 1996 Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873-2878.

Morin, J. G., Harrington, A., Nealson, K., Krieger, N., Baldwin, T.O., Hastings, J.W., 1975. Light for all reasons - versatility in behavioral repertoire of flashlight fish. Science 190, 74-76.

Nyholm, S. V., McFall-Ngai, M. J., 2004. The winnowing: Establishing the squid-*Vibrio* symbiosis. Nat. Rev. Microbiol. 2, 632-642.

Raes, J., Korbel, J. O., Lercher, M. L., von Mering, C., Bork, P. , 2007. Prediction of effective genome size in metagenomic samples. Genome Biol. 8.

Reen, F. J., Almagro-Moreno, S., Ussery, D., Boyd, E. F., 2006. The genomic code: inferring *Vibrionaceae* niche specialization. Nat. Rev. Microbiol. 4, 697-704.

Ruby, E. G., Urbanowski, M., Campbell, J., Dunn, A., Faini, M., Gunsalus, R., Lostroh, P., Lupp, C., McCann, J., Millikan, D., Schaefer, A., Stabb, E., Steve, A., Visick, K., Whistler, C., Greenberg, E. P., 2005. Complete genome sequence of *Vibrio fischeri:* A symbiotic bacterium with pathogenic congeners. Proc. Natl. Acad. Sci. U. S. A. 102, 3004-3009.

Swofford, D.L., 1998. PAUP* 4.0b10: Phylogenetic analysis using parsimony. Sinauer, Sunderland, Massachusetts.

Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. ClustalW - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matric choice. Nucleic Acids Res. 22, 4673-4680.

Wernegreen, J. J., Moran, N. A., 1999. Evidence for genetic drift in endosymbionts (Buchnera): Analyses of protein-coding genes. Mol. Biol. Evol. 16, 83-97.

Wernegreen, J. J., Funk, D. J., 2004. Mutation exposed: A neutral explanation for extreme base composition of an endosymbiont genome. J. Mol. Evol. 59, 849-858.

Wier, A. M., Nyholm, S. V., Mandel, M. J., Massengo-Tiasse, R. P., Schaefer, A. L., Koroleva, I., Splinter-BonDurant, S. S., Brown, B., Manzella, L., Snir, E., Almabrazi, H., Scheetz., T. E., de Fatima Bonaldo, M., Casavant, T. L., Soares, M. B., Cronan, J. E., Reed, J. L., Ruby, E. G., McFall-Ngai, M. J., 2010. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proc. Natl. Acad. Sci. U. S. A. 107, 2259-2264.

Woolfit, M., Bromham, L., 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20, 1545-1555.
Yu, X. J., Walker, D. H., Liu, Y., Zhang, L. H., 2009. Amino acid biosynthesis deficiency in bacteria associated with human and animal hosts. Infect. Genet. Evol. 9, 514-517.

CHAPTER IV

The obligate luminous symbionts of flashlight fish are specific to host genera.

Abstract

The luminous bacterial symbionts of anomalopid flashlight fish, which are obligately dependent on their hosts for growth, share several evolutionary patterns with unrelated obligate symbionts. It is not known, however, if their associations with hosts are species-specific, a pattern common to obligate intracellular symbionts. Unlike most obligate symbionts, anomalopid symbionts are extracellular and environmentally acquired, which may decrease the likelihood of evolving specific interactions between symbionts and hosts. To test for species specificity we performed phylogenetic analysis on the symbionts of multiple anomalopid host species. These analyses resolve the symbionts of hosts in the genus *Photoblepharon* as a new species for which we propose the name '*Candidatus* Photodesmus blepharus.' The analyses also indicate that anomalopid symbiont species are specific to host genera. Specificity is found even when host genera overlap in geographic range, suggesting that the observed specificity is not the result of geographic isolation. The observation of specificity is consistent with the obligate dependence of anomalopid symbionts on their hosts but is not typical for environmentally acquired symbionts.

Introduction

Multiple lineages of marine fish and squid engage in mutualistic symbioses with luminous bacteria of the *Gammaproteobacteria* family *Vibrionaceae* (Dunlap et al., 2007). Most of these associations involve facultatively symbiotic bacteria that maintain large free-living populations in diverse habitats (Dunlap et al., 2007; Dunlap et al., 2009). The symbionts of anomalopid flashlight fish are an exception to this pattern; the anomalopid symbiont '*Candidatus* Photodesmus katoptron' is obligately dependent on the host for growth and is the only known extracellular, environmentally acquired symbiont demonstrating evolutionary patterns similar to intracellular obligate symbionts (Chapter III). These patterns include genome reduction, high AT nucleotide content, and a high evolutionary rate (Wernegreen, 2002; Woolfit and Bromham, 2003; Hendry and Dunlap, 2011; Chapter III). Because only one species of anomalopid symbiont has been well studied, it is not known if anomalopid symbionts show species-specificity with hosts, which is also common to obligate symbionts (Clark et al., 2002; Sachs et al., 2011). Facultative luminous symbionts typically do not show specificity with hosts. Instead, both symbionts and hosts form associations with multiple species (Dunlap et al., 2007; Kaeding et al., 2007). Like facultative luminous symbionts, anomalopid symbionts are extracellular and environmentally acquired while most other obligate symbionts are intracellular and vertically transmitted (Wernegreen, 2002; Sachs et al., 2011a, Sachs et al., 2011b). Specificity is not necessary for an obligate symbiont and may be less likely to occur in environmentally acquired symbionts than vertically transmitted symbionts. To test for species-specificity between anomalopid fish and their symbionts, we performed phylogenetic analyses on the symbionts of multiple anomalopid species.

Anomalopid flashlight fish are nocturnal marine fish found globally in shallow to deep-water tropical reef habitats. Behavioral observations of some species have found that the fish occupy cave shelters by day and at night they leave to feed on zooplankton, using light from subocular light organs. The light is also used for communication and predator avoidance (Morin et al., 1975; Herring and Morin, 1978). There are eight anomalopid fish species and the symbionts of four species, *Anomalops katoptron*, *Kryptophanaron alfredi*, *Photoblepharon palpebratus*, and *Photoblepharon steinetzi*, are included in analyses here. These species represent a wide geographic range; *A. katoptron* is found in the eastern Indian Ocean and co-occurs with *P. palpebratus* in the south Pacific Ocean, *P. steinetzi* occurs in the Red Sea and western Indian Ocean, and *K. alfredi* is found in the Caribbean (McCosker and Rosenblatt, 1987; Johnson and Rosenblatt, 1988; Rosenblatt and Johnson, 1991; Baldwin et al., 1997; Johnson et al., 2001).

Anomalopid symbionts are extracellular and environmentally acquired. They are densely packed within the anomalopid light organ and are regularly released into the surrounding seawater through pores on the light organ surface (Kessel, 1977; Haygood et al., 1984). From there they are presumed to persist long enough to colonize new hosts as larval fish (Haygood et al., 1984; Haygood, 1993; Chapter III). Since they are not vertically transmitted anomalopid symbionts need not be host specific. To test for specificity in anomalopid symbionts, we performed phylogenetic analyses on the symbionts of multiple anomalopid species to determine if different host species harbor different symbionts. Because anomalopid species are difficult to acquire and their symbionts are not culturable (Herring and Morin, 1978; Haygood, 1993), the symbiont of

only one host species, *A. katoptron*, has been well studied. In the current study, we have added sequences of protein coding genes for symbionts of *P. palpebratus* and *P. steinetzi* to determine their relationships to each other and to the *A. katoptron* symbiont.

Results and discussion

Phylogenetic support for 'Candidatus Photodesmus'

Previous work has demonstrated that anomalopid symbionts represent a divergent genus, '*Candidatus* Photodesmus' (Greek: photo = light, desmus = servant) within the family *Vibrionaceae* (Hendry and Dunlap, 2011). With the addition of more taxa, analyses still recover a monophyletic clade for '*Ca*. Photodesmus.' The housekeeping gene Maximum Likelihood (ML) tree (Fig. 4.1) places '*Ca*. Photodesmus' as sister to the genus *Vibrio*, consistent with earlier findings (Hendry and Dunlap, 2011). Of note is the fact that very low support is found for the clade *Vibrio* as currently configured. The Bayesian (BA) housekeeping gene analysis differed slightly, in that the BA tree (not shown) could not resolve the relationship between the anomalopid symbiont clade and the *Vibrio* clade, instead finding a polytomy. This ambiguity, and the low support for the clade *Vibrio* in the ML tree, suggests that the genus *Vibrio* may be a paraphyletic group. Both ML and BA analyses of *lux* genes resolved identical topologies with '*Ca*. Photodesmus' as divergent from other *Vibrionaceae* genera and sister to *Vibrio*. (Fig. 4.2). The monophyly and high level of divergence of anomalopid symbionts is consistent with their obligate host dependence, since obligate symbionts often evolve at a faster rate than free-living relatives (Moran, 1996; Woolfit and Bromham, 2003). The long branch leading to the '*Ca*. Photodesmus' clade in Fig. 4.1 suggests that an obligate interaction

evolved early on in the evolutionary history of the clade and that all anomalopid symbionts are therefore obligately dependent on their hosts.

Different anomalopid genera host different symbiont species

Previous studies demonstrating monophyly of anomalopid symbionts have used only 16S sequence for the symbionts of all fish hosts except *Anomalops katoptron* (Haygood and Distel, 1993; Hendry and Dunlap, 2011), making it difficult to determine if the symbionts of different host fish are unique species. Here we demonstrate that different species of the genera *Photoblepharon* host the same symbiont species but other fish genera host different bacterial species. In all analyses the symbionts of the hosts *P. palpebratus* and *P. steinetzi* were closely related with strong support. Consistent with this, the 16S rRNA gene sequences of symbiont isolates from these two hosts show 99.6% identity, indicating that they are likely the same species. These analyses use symbiont sequences from only one *P. steinetzi* individual. Additional sequences could uncover more divergence between the symbionts of the two *Photoblepharon* species, possibly distinguishing the symbionts as separate species. The *Photoblepharon* symbionts were resolved as highly divergent from the *Anomalops katoptron* symbiont '*Ca.* Photodesmus katoptron.' The 16S rRNA gene sequences of the *P. palpebratus* and *P. steinetzi* symbionts are only 94.8% identical to '*Ca*. Photodesmus katoptron.' The low 16S similarity and long branches that separate the *Photoblepharon* symbionts from '*Ca*. Photodesmus katoptron' in all analyses support the creation of a new species designation for the *Photoblepharon* symbionts. We propose the name '*Candidatus* Photodesmus blepharus' (Greek: blephar = eyelid) after the host genus, which is so named for the lid-

like structure individuals raise over the light organ to control light emission. Only 16S sequence data is available for symbionts from the fourth species of host included here, *Kryptophanaron alfredi*. However, the low 16S identity between *K. alfredi* symbionts and '*Ca*. Photodesmus katoptron,' 94.3%, as well as the long branches separating the *K. alfredi* symbiont from other symbionts, indicate that the fish genus *Kryptophanaron* likely also hosts a distinct species of symbiont. These results demonstrate that different genera of fish host specific species of symbionts. It is likely that the species *P. palpebratus* and *P. steinetzi* do not have unique symbiont species because the hosts have diverged recently and not enough time has passed for significant symbiont divergence to take place.

Facultative luminous symbionts show very little specificity with hosts; all known species can colonize multiple unrelated groups of hosts and host species can frequently maintain multiple bacterial species, sometimes simultaneously within an individual (Dunlap et al., 2007; Kaeding et al., 2007). Facultatively symbiotic strains do not cluster phylogenetically with other strains isolated from the same host, nor do they display the level of divergence between host genera found here (Fig. 4.3) (Dunlap et al., 2007). A possible explanation for the specificity observed in anomalopid symbionts could be geographic isolation and divergence of hosts and symbionts. However, geographic divergence is unlikely in this case because host species that co-occur have different symbionts. The fish species *A. katoptron* and *P. palpebratus* co-occur for much of their range (south Pacific Ocean, Philippines to Vanuatu) and are often collected at the same time and location (Wolfe and Haygood, 1991). In spite of this proximity, *A. katoptron* and *P. palpebratus* collected at the same location harbor divergent symbiont isolates that

group with the species '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' respectively (Fig. 4.3). It is therefore likely that the specificity observed has a genetic basis, though it is possible that *A. katoptron* and *P. palpebratus* larval fish develop separately from the opposite species and the symbiont of each species does not disperse far enough to reach larval fish of the opposite species. If there is a genetic basis for the specificity, it may be the result of selection or genetic drift and imposed by either the host or the symbiont.

Implications for the evolution the symbiotic interaction

Obligate intracellular symbionts often show specificity with hosts due to vertical transmission and codivergence (Clark et al., 2000; Sachs et al., 2011a; Sachs et al., 2011b). Anomalopid symbionts are environmentally acquired rather than transferred by direct vertical transmission (Haygood et al., 1984; Haygood, 1993; Chapter III), so codivergence is possible though not inevitable within the group. The number of host-symbiont pairs included here is too small to fully test for codivergence but the topology resolved for symbionts does not mirror the phylogeny of the host fish. The current host phylogeny, based on morphological characters, places the genera *Kryptophanaron* and *Photoblepharon* as more closely related to each other than they are to *Anomalops*, while the symbiont phylogeny resolves *Anomalops* and *Photoblepharon* symbionts as most closely related (Fig. 4.4). These non-congruent phylogenies contradict codivergence between hosts and symbionts. Several possible explanations exist for why this pattern could arise in spite of codivergence. It is possible that the host phylogeny is incorrect. It is also possible that multiple evolutions of obligate dependence or host shifts have

occurred (Haygood and Distel, 1993), though these scenarios are less parsimonious than others. The fact that old world symbionts ('*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus') are more closely related to each other than they are to the new world symbiont (the *K. alfredi* symbiont) suggests that specificity may have evolved more recently than the origin of the symbiosis (Fig. 4.4). It is most likely that all anomalopid symbionts are obligate and that obligate dependence arose in the common ancestor of the hosts. However, the ancestral obligate symbiont would not necessarily have codiverged with the host. It is possible that all symbionts maintained the ability to colonize multiple host species until after the separation of old world and new world hosts and that symbionts have subsequently codiverged with hosts.

Conclusions

Specificity of anomalopid symbionts to their hosts is consistent with their obligate host dependence, but not typical for environmentally acquired symbionts. The observation of symbiont specificity in anomalopid hosts that co-occur geographically suggests that there is a genetic cause of specificity in anomalopid symbionts. '*Candidatus* Photodesmus' symbiont species are highly divergent and show signatures of high levels of genetic drift (Hendry and Dunlap, 2011, Chapter III). It is possible that genes needed to colonize certain host species would be lost by chance in some lineages, making them specific to a subset of hosts. Alternatively, gene loss in symbiont lineages might make them inferior symbionts to some host species, imposing selection on the host or bacterium to prevent colonization. Gene content or gene expression comparisons of multiple symbiont species could be used to find a possible functional basis for specificity.

In addition to demonstrating specificity, the results presented here suggest that codivergence between symbionts and hosts may have begun after some host speciation had taken place rather than with the origin of the symbiosis. To fully test for codivergence, a molecular analysis of anomalopids to confirm the phylogenetic relationships between hosts and the addition of more symbiont sequences is needed.

Methods

Phylogenetic analysis

New sequences used in this study were obtained from whole genome Illumina sequencing of the *P. palpebratus* symbiont DNA and PCR amplification of the *P. steinetzi* symbiont DNA. For the *P. palpebratus* symbiont, four specimens (Ppalp.1-Ppalp.4) were collected from coastal waters in the Republic of Vanuatu in 2011 and DNA was extracted as in Hendry and Dunlap (2011). DNA from one light organ of each specimen was combined for sequencing. Very little polymorphism exists within the symbiont of a host species (Hendry and Dunlap, 2011; T.A. Hendry and P.V. Dunlap, unpublished data), so sequences obtained from the combined samples should not be significantly different than if they had come from an individual. Illumina reads were assembled in Mira3 (Chevreux et al., 2002) by staff of the University of Michigan CCDU Bioinformatics Core. DNA from the *P. steinetzi* symbiont came from the sample described in Wolfe and Haygood (1991); the fish (Pstein.1) was obtained from the Coral World aquarium in Eilat, Israel in 1987 and was likely collected from around Dahab on the Sinai peninsula. Previous work has found that both light organs of an individual contain monoclonal bacteria of the same genotype, so DNA from the *P. steinetzi* sample

can be considered one strain. PCR amplification of *P. steinetzi* symbiont loci followed Hendry and Dunlap (2011). Sequences obtained in this study from the *P. palpebratus* symbiont (the 16S rRNA gene, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, *topA*, and *luxCDABEG*) and from the *P. steinetzi* symbiont (the 16S rRNA gene, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, *topA*, and *luxCDAEG*) were deposited in GenBank. Phylogenetic analysis methods follow Hendry and Dunlap (2011). For housekeeping genes a concatenated gene matrix was analyzed using Maximum Likelihood in Garli (Zwickl, 2006) under the GTR + I + Γ model. The *lux* operon was analyzed as one locus with non-coding regions removed using the GTR + I + Γ model. For both matrices, 1000 bootstrap replicates were performed, with each run for 1000 generations. For Bayesian analyses, each gene was analyzed in MRBAYES v3.1.2 (Huelsenbeck and Ronquist, 2001) using the GTR + I + Γ model over 100,000 generations sampling every 100 generations.

Comparison with facultative luminous symbionts

We compared divergence between anomalopid symbionts from different host genera to divergence between facultative symbionts from different host genera. A Maximum Likelihood tree was generated in MEGA5 (Tamura et al., 2011) using 16S sequences taken from GenBank or generated for this study for the following strains: '*Ca*. P. katoptron' = Akat2007.1.1 (Hendry and Dunlap, 2011), A.katoptron symbiont (GenBank accession Z19081), Akat.4-Akat.7 (Chapter III), Akat.8-Akat.15 (this study, 8 *A. katoptron* specimens collected in Vanuatu and processed as described for *P. palpebratus*); *Kryptophanaron alfredi* symbiont = Kryptophanaron alfredi symbiont (GenBank accession Z19003); '*Ca*. P. blepharus' = P.palpebratus symbiont (accession

Z19085), P.steinetzi symbiont (accession Z19080), Ppalp.1-Ppalp.4 (this study); *A. fischeri* = ET101, ET301, ET401, CG101, MJ101, EM17, EB12.

Acknowledgements

This work will be coauthored with Paul V. Dunlap when published. We thank Margo G. Haygood for the generous gift of *P. steinetzi* symbiont DNA. We also thank the staff of the University of Michigan Sequencing Core for DNA sequencing and advice and Jeffrey R. de Wet, of the University of Michigan Bioinformatics Analysis Core, for carrying out sequence assembly. Grant Norton of Sustainable Reef Suppliers, Brad Remmer of Sea Dwelling Creatures and Fish Doctors Aquarium provided assistance in obtaining *P. palpebratus* specimens. Departmental Block Grant awards from the University of Michigan Department of Ecology and Evolutionary Biology funded this work.

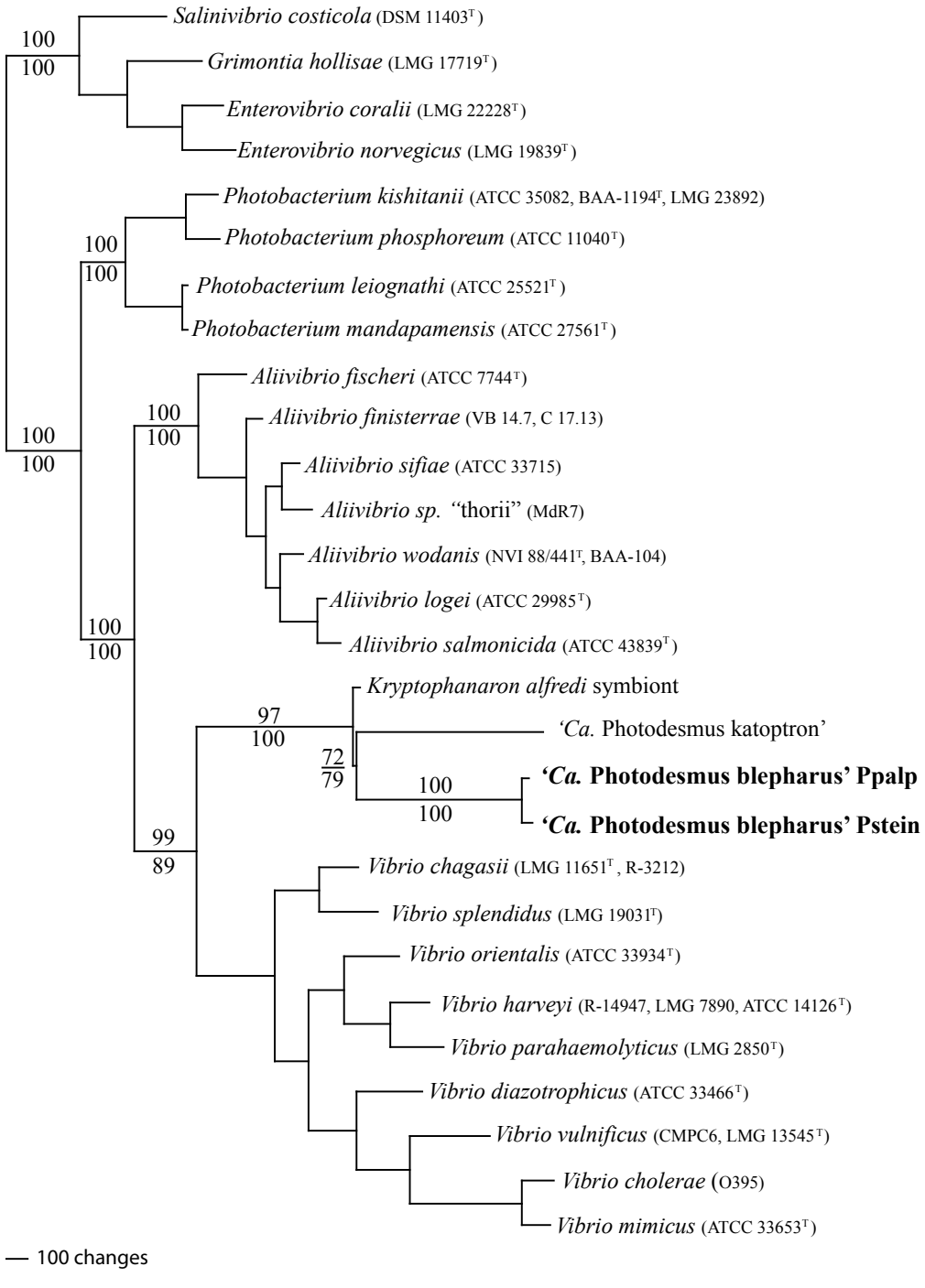
Salinivibrio costicola (DSM 11403^T)
Grimontia hollisae (LMG 17719^T)
Enterovibrio coralii (LMG 22228^T)
Enterovibrio norvegicus (LMG 19839^T)
Photobacterium kishitanii (ATCC 35082, BAA-1194^T, LMG 23892)
Photobacterium phosphoreum (ATCC 11040^T)
Photobacterium leiognathi (ATCC 25521^T)
Photobacterium mandapamensis (ATCC 27561^T)
Aliivibrio fischeri (ATCC 7744^T)
Aliivibrio finisterrae (VB 14.7, C 17.13)
Aliivibrio sifiae (ATCC 33715)
Aliivibrio sp. "thorii" (MdR7)
Aliivibrio wodanis (NVI 88/441^T, BAA-104)
Aliivibrio logei (ATCC 29985^T)
Aliivibrio salmonicida (ATCC 43839^T)
Kryptophanaron alfredi symbiont
'Ca. Photodesmus katoptron'
'*Ca.* Photodesmus blepharus' Ppalp
'*Ca.* Photodesmus blepharus' Pstein
Vibrio chagasii (LMG 11651^T, R-3212)
Vibrio splendidus (LMG 19031^T)
Vibrio orientalis (ATCC 33934^T)
Vibrio harveyi (R-14947, LMG 7890, ATCC 14126^T)
Vibrio parahaemolyticus (LMG 2850^T)
Vibrio diazotrophicus (ATCC 33466^T)
Vibrio vulnificus (CMPC6, LMG 13545^T)
Vibrio cholerae (O395)
Vibrio mimicus (ATCC 33653^T)
100 changes

Fig. 4.1. Maximum likelihood trees from housekeeping genes (16S rRNA gene, *atpA*, *gapA*, *gyrB*, *pyrH*, *recA*, *rpoA*, and *topA*). Maximum Likelihood bootstrap numbers are shown above branches and Bayesian posterior probabilities are shown below. Strain designations follow taxa names. Taxa with new sequences are shown in bold. 'Ppalp' refers to symbionts isolated from *P. palpebratus*, and 'Pstein' indicated symbiont isolates from *P. steinetzi*. Accession numbers for sequences taken from GenBank can be found in Hendry and Dunlap (2011).

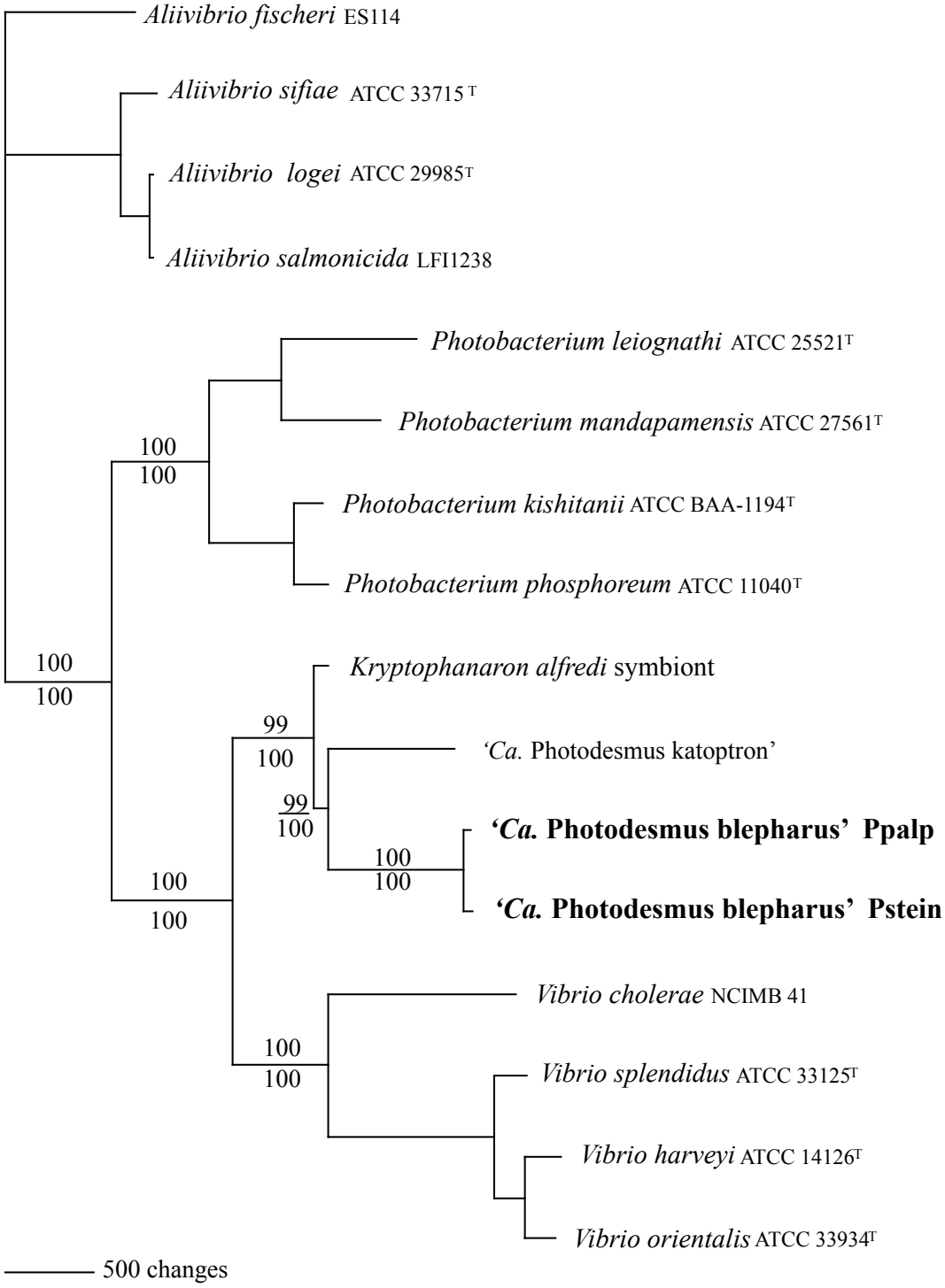

Fig. 4.2. Maximum likelihood trees from *lux* operon (*luxCDABEG*). Maximum Likelihood bootstrap numbers are shown above branches and Bayesian posterior probabilities are shown below. Strain designations follow taxa names. Taxa with new sequences are shown in bold. 'Ppalp' refers to symbionts isolated from *P. palpebratus*, and 'Pstein' indicated symbiont isolates from *P. steinetzi*. Accession numbers for sequences taken from GenBank can be found in Hendry and Dunlap (2011).

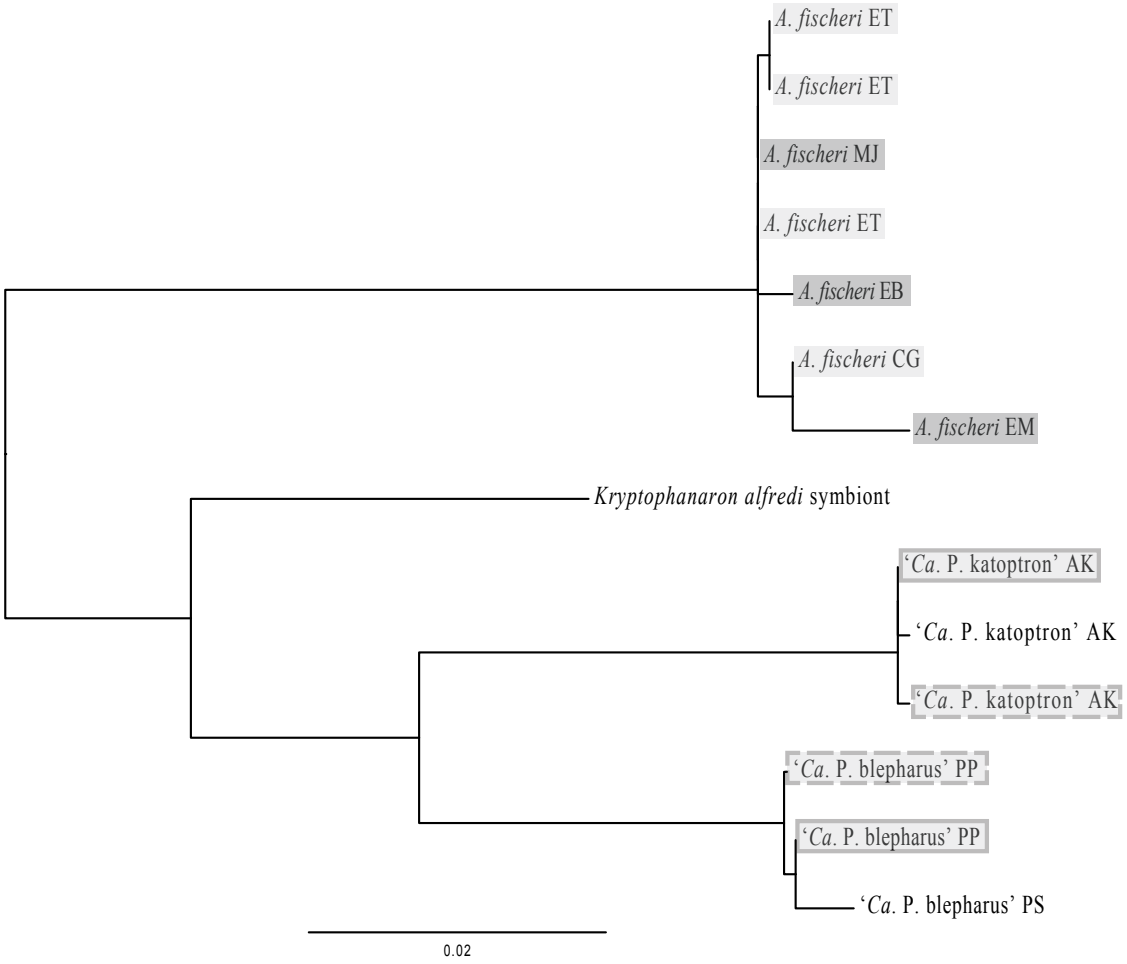

Fig. 4.3. Divergence between strains from different host genera for the symbionts *Aliivibrio fischeri* and '*Ca*. Photodesmus spp.' '*Ca*. Photodesmus' species show much more divergence between host genera than *A. fischeri* does. Abbreviations show the host species each strain was isolated from (ET = *Euprymna tasmanica*, MJ = *Monocentris japonica*, EB = *Euprymna berryi*, CG = *Cleidopus gloriamaris*, EM = *Euprymna morsei*, AK = *Anomalops katoptron*, PP = *Photoblepharon palpebratus*, PS = *Photoblepharon steinetzi*). Similarly shaded squares around strains indicate that hosts were collected at the same time and place.

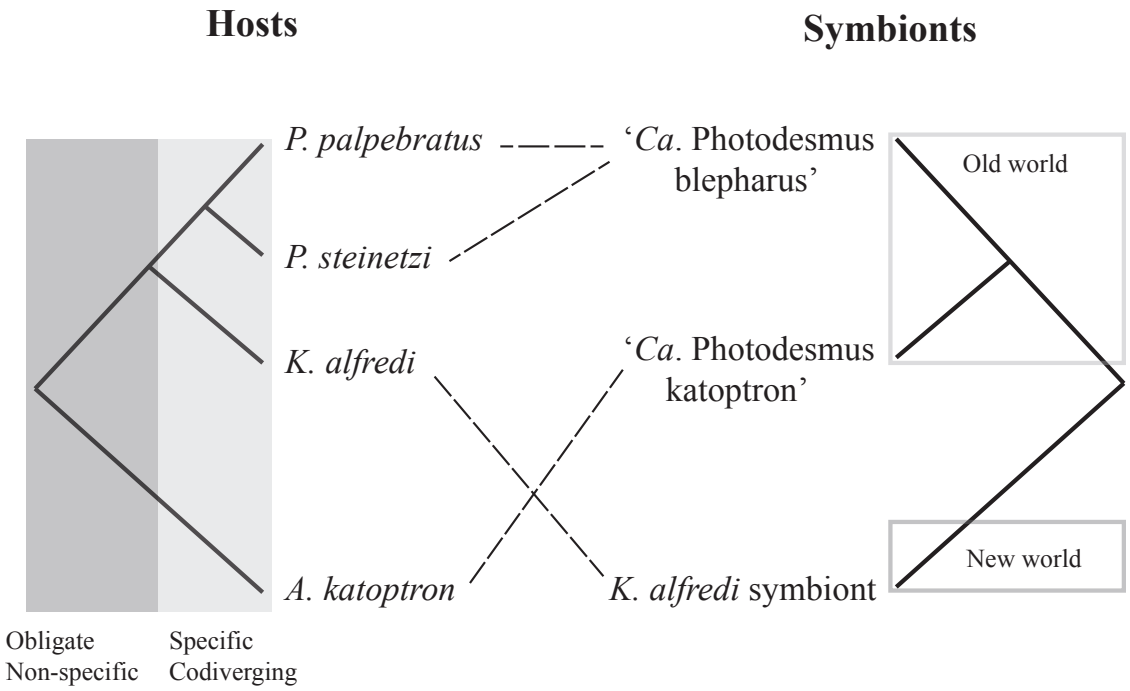

Fig. 4.4. Cladograms comparing anomalopid host phylogeny (Baldwin et al., 1997; Johnson et al., 2001) and anomalopid symbiont phylogeny. Dashed lines demonstrate host-symbiont relationships. Shown is a possible scenario for explaining the lack of congruence between phylogenies, in which obligate dependence in anomalopid symbionts evolved before specificity.

References

Baldwin, C.C., Johnson, G.D., Paxton, J.R., 1997. *Protoblepharon rosenblatti*, a new genus and species of flashlight fish (Beryciformes: Anomalopidae) from the tropical South Pacific, with comments on anomalopid phylogeny. Proc. Biol. Soc. Wash. 110, 373-383.

Chevreux, B., Pfisterer, T., Suhai, S., 2002. Automatic assembly and editing of genomic sequences. In: Suhai, S. (Ed.), *Genomics and Proteomics - Functional and Computational Aspects.* Kluwer Academic Publishers, New York, NY, 51-65.

Clark, M.A., Moran, N.A., Baumann, P. and Wernegreen, J.J., 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution 54, 517-525.

Dunlap, P.V., 2009. Microbial bioluminescence. In *Encyclopedia of microbiology*. Schaechter, M. (ed). Oxford: Elsevier, pp. 45-61.

Dunlap, P.V., Ast, J.C., Kimura, S., Fukui, A., Yoshino, T., Endo, H., 2007. Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbioses. Cladistics 23, 507-532.

Haygood, M.G., 1993. Light organ symbioses in fishes. Crit. Rev. Microbiol. 19, 191-216.

Haygood, M.G., Distel, D.L., 1993. Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. Nature 363, 154-156.

Haygood, M.G., Tebo, B.M., Nealson, K.H., 1984. Luminous bacteria of a monocentrid fish (*Monocentris japonicus*) and 2 anomalopid fishes (*Photoblepharon palpebratus* and *Kryptophanaron alfredi*) - population sizes and growth within the light organs, and rates of release into the seawater. Mar. Biol. 78, 249-254.

Herring, P.J., Morin, J.G., 1978. Bioluminescence in fishes. In *Bioluminescence in action.* Herring, P.J. (ed). London: Academic Press, pp. 273-329.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Johnson, G.D., Rosenblatt, R.H., 1988. Mechanisms of light organ occlusion in flashlight fishes, family *Anomalopidae* (Teleostei, Beryciformes), and the evolution of the group. Zool. J. Linn. Soc. 94, 65-96.

Johnson, G.D., Seeto, J., Rosenblatt, R.H., 2001. *Parmops echinatus*, a new species of flashlight fish (Beryciformes : Anomalopidae) from Fiji. Proc. Biol. Soc. Wash. 114, 497-500.

Kaeding, A.J., Ast, J.C., Pearce, M.M., Urbanczyk, H., Kimura, S., Endo, H., Nakamura, M., Dunlap, P.V., 2007. Phylogenetic diversity and cosymbiosis in the

bioluminescent symbioses of "*Photobacterium mandapamensis*". Appl. Environ. Microbiol. 73, 3173-3182.
Kessel, M., 1977. The ultrastructure of the relationship between the luminous organ of the teleost fish *Photoblepharon palpebratus* and its symbiotic bacteria. Cytobiologie 15, 145-158.
Maddison, W.P., Maddison, D.R., 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Mass.
McCosker, J.E., Rosenblatt, R.H., 1987. Notes on the biology, taxonomy, and distribution of flashlight fishes (Beryciformes, Anomalopidae). Jpn J. Ichthyol. 34, 157-164.
Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873-2878.
Morin, J.G., Harrington, A., Nealson, K., Krieger, N., Baldwin, T.O., Hastings, J.W., 1975. Light for all reasons - versatility in behavioral repertoire of flashlight fish. Science 190, 74-76.
Rosenblatt, R.H., Johnson, G.D., 1991. *Parmops coruscans*, a new genus and species of flashlight fish (Beryciformes, Anomalopidae) from the south Pacific. Proc. Biol. Soc. Wash. 104, 328-334.
Sachs, J.L., Essenberg, C.J., Turcotte, M.M., 2011a. New paradigms for the evolution of beneficial infections. Trends Ecol. Evol. 26, 202-209.
Sachs, J.L., Skophammer, R.G., Regus, J.U., 2011b. Evolutionary transitions in bacterial symbiosis. Proc. Natl. Acad. Sci. U. S. A. 108, 10800-10807.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.
Wernegreen, J.J., 2002. Genome evolution in bacterial endosymbionts of insects. Nature Rev. Genet. 3, 850-861.
Wolfe, C.J., Haygood, M.G., 1991. Restriction fragment length polymorphism analysis reveals high levels of genetic divergence among the light organ symbionts of flashlight fish. Biol. Bull. 181, 135-143.
Woolfit, M., Bromham, L., 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20, 1545-1555.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas, Austin, Texas.

CHAPTER V

Extremely low genetic diversity and population dynamics in an obligate luminous symbiont.

Abstract

Host interactions are predicted to have large impacts on the population dynamics and genetic diversity of bacteria, though the precise nature of the effects of host associations is unclear. A variety of host-associated bacteria have very low genetic diversity. In the case of *Buchnera aphidicola*, this is due to high levels of drift between host generations and frequent bottlenecks in host populations. To test the broadness of this pattern, we investigated the genetic diversity of two species of anomalopid flashlight fish symbionts. These symbionts are closely related to facultatively host-associated bacteria that are typically very diverse, but like *B. aphidicola* the flashlight fish symbionts are obligately dependent on their hosts for growth. Similarly to *Buchnera*, anomalopid symbionts have extremely low genetic polymorphism, with values between ten and 100 times lower than in free-living relatives. Our data show that '*Ca*. P. kaptoptron' is nearly monomorphic when sampled across a wide geographic range. Furthermore, both anomalopid symbiont species have an excess of rare alleles and one species has a high number of nonsynonymous substitutions, as is predicted to occur due to drift. This findings are similar to *B. aphidicola* and indicate that low polymorphism may be broadly found in obligate symbionts. However, one anomalopid symbiont species

has very low nonsynonymous substitution rates, in keeping with purifying selection rather than drift. Additionally, unlike *B. aphidicola*, the anomalopid symbiont may have population structure across its geographic range. These differences between anomalopid symbionts and *B. aphidicola* suggest that different factors may be influencing diversity patterns in each system.

Introduction

Numerous bacterial species form symbiotic associations with animal hosts and these associations have the potential to affect bacterial population dynamics and thus affect the course of evolution in bacterial lineages (Sachs et al., 2011). Host association can act on bacterial population dynamics by restricting growth to within host populations, which may lead to selection on symbiotic bacteria or affect their genetics through drift. Additionally, the population dynamics of hosts will also likely affect their symbiont populations. Facultatively symbiotic bacteria that are not host-restricted are sometimes highly diverse (Dunlap et al., 2004; Preheim et al., 2011), whereas some obligately host-associated bacteria have very low levels of genetic diversity due to drift and host population dynamics (Abbot and Moran, 2002; Funk et al., 2001). In most cases the relative importance of these factors, selection, drift, or host dynamics, are unknown. One way to address this issue is to sample a range of host-associated bacteria that vary in host population dynamics and their level of host restriction. Here we report genetic polymorphism patterns from two species of luminous symbionts that have similarities to both obligate and facultative symbionts. Anomalopid flashlight fish symbionts, of the genus '*Candidatus* Photodesmus,' are obligately dependent on their hosts for growth

(Chapter III). However, they differ from many obligate symbionts because they are environmentally acquired and have a vertebrate host, which is likely to have very different population dynamics than the hosts of most known obligate symbionts (Chapter III; Haygood, 1993). We compared genetic polymorphism in anomalopid symbionts to free-living, host-associated relatives and other obligate symbionts to identify which factors, host restriction, drift, selection, or host dynamics, might have the greatest affect bacterial diversity patterns.

Both theory and some empirical studies suggest that associations with hosts may limit bacterial diversity through selection and adaptation to the host environment (Sachs et al., 2011; Wollenberg and Ruby, 2012). Limited diversity and host adaptation are not the norm in the widespread host-associated bacteria of the marine family *Vibrionaceae*. Many facultatively host-associated *Vibrionaceae* species display little host specialization and large amounts of genetic polymorphism (Dunlap et al., 2007; Dunlap et al., 2008; Dunlap et al., 2004; Kaeding et al., 2007; Preheim et al., 2011; Thompson et al., 2005). Much of this polymorphism is likely to be neutral (Thompson et al., 2005), indicating that forces purging genetic diversity are rare in free-living *Vibrionaceae* (Preheim et al., 2011). However, these studies have mainly focused on species that are transiently associated with a host rather than those that sometimes form closer symbioses with hosts, such as luminous symbionts. It is possible that in long-term interactions more bacterial adaptation and loss of diversity may occur. Some findings consistent with this idea have been found in specific loci of the squid symbiont *Aliivibrio fischeri* (Wollenberg and Ruby, 2012). Other luminous symbionts, however, display high levels of genetic diversity genome wide (Dunlap et al., 2004, Dunlap et al., 2008). It is not known if the

cause of diversity is that facultative symbionts are not host restricted and therefore maintain large population sizes or that that they are adapting to many different environments.

In contrast to facultatively host-associated bacteria, some obligately host-associated bacteria maintain strikingly little genetic polymorphism even across a wide geographic range. This is true of the aphid endosymbiont *Buchnera aphidicola* when it is isolated from within a single host species (Abbot and Moran, 2002; Funk et al., 2001). However, this lack of diversity results from high genetic drift rather than selection for specific genotypes. Because *B. aphidicola* only occurs within the host and undergoes population bottlenecks during transfer between host generations, it has small effective population sizes (Mira and Moran, 2002). Selection is therefore weak in the species and drift is strong (Fry and Wernegreen, 2005; Moran, 1996; Wernegreen and Funk, 2004; Wernegreen and Moran, 1999; Woolfit and Bromham, 2003). Within a host species, bottlenecks between host generations purge many mutations so that little genetic variation exists and what does exist is rich in nonsynonymous and rare substitutions. As a result, a common *B. aphidicola* genotype dominates even across a wide geographic range (Abbot and Moran, 2002; Funk et al., 2001). The suggested hypothesis for why there is no local divergence of *B. aphidicola* comes from aspects of the aphid host ecology. Aphids themselves undergo seasonal population bottlenecks, which should purge symbiont diversity. Additionally, they disperse very widely, which should homogenize symbiont population structure (Abbot and Moran, 2002; Funk et al., 2001). These data indicate that both the interaction with the host and host ecology play important roles in shaping diversity patterns within *B. aphidicola*, although it is not known how broadly this

applies. For instance, some obligate pathogens also display very low intraspecific diversity (Achtman, 2008), but this is not likely to be caused by host dynamics.

Anomalopid flashlight fish symbionts of the genus '*Ca*. Photodesmus' have similarities to both free-living luminous symbionts and obligate endosymbionts. Like other luminous symbionts, '*Ca.* Photodesmus' species are in the family *Vibrionaceae* (Hendry and Dunlap, 2011), and are extracellular (Haygood, 1993; Kessel, 1977) and environmentally acquired (Chapter III; Haygood, 1993). In spite of this, they are also obligately dependent on their hosts for growth and have extremely reduced genomes (Chapter III; Chapter VI), indicating that much like *B. aphidicola* and other obligate endosymbionts they might be highly subject to drift. If this is the case, anomalopid symbionts should show similar polymorphism trends to *B. aphidicola*. However, their hosts, anomalopid flashlight fish, are unlikely to display the same patterns of regular population bottlenecks and widespread dispersal that *B. aphidicola* hosts do. '*Candidatus* Photodesmus' can therefore be used to disentangle the affects of obligate host dependence and host ecology on symbiont diversity. To do this, we characterized patterns of genetic polymorphism across whole genomes from two species of anomalopid symbionts and compared them to *B. aphidicola*. Also for comparison, we calculated genetic diversity values for a variety of facultatively host-associated *Vibrionaceae* members, both luminous symbionts and not. If the same factors affecting genetic diversity in *B. aphidicola* are at work in the anomalopid symbionts, we predict that '*Ca*. Photodesmus' polymorphism patterns should be more similar to *B. aphidicola* than related free-living species.

Materials and Methods

Samples and DNA sequencing

Samples of '*Ca*. Photodesmus blepharus' Ppalp came from four specimens of *Photoblepharon palpebratus* collected in the Republic of Vanuatu in April 2011. Whole light organs were used for DNA extraction as in Hendry and Dunlap (2011). Equal amounts of DNA from one light organ of each fish were combined for sequencing. For '*Ca*. Photodesmus katoptron,' the same procedure was followed for eight specimens of *Anomalops katoptron* also collected in the Republic of Vanuatu in April 2012. Libraries of each symbiont were sequenced in one lane of an Illumina HiSeq 2000 sequencer at the University of Michigan DNA Sequencing Core. Additionally, five light organs from four specimens of *A. katoptron* collected near Cebu City in the Philippines in 2008 were used for sequencing on half a plate of a Roche Genome Sequencer FLX 454 at the University of Michigan DNA Sequencing Core. Sequences from the Christmas Island '*Ca*. Photodesmus katoptron' came from one light organ of one specimen collected near Christmas Island in 2007. These sequences were amplified with either PCR or cloning with a Lucigen CloneSmart ChimeraFree cloning kit and sequenced with Sanger sequencing.

Analysis of anomalopid symbionts

To detect polymorphic sites, reads from each high-throughput library, '*Ca*. Photodesmus katoptron' Christmas Island, '*Ca.* Photodesmus katoptron' Vanuatu, and '*Ca*. Photodesmus blepharus' Ppalp Vanuatu, were aligned against the assembled genome for each species. For '*Ca*. Photodesmus katoptron,' the Christmas Island the genome with

GenBank accession PRJNA80863 was used. For both Vanuatu libraries, genomes were assembled by the University of Michigan Bioinformatics Core using Mira3 (Chevreux et al., 2002). Glimmer 3.02 (Delcher et al., 2007) was used to predict orfs within both assemblies and orfs were annotated by blast to the Swiss-Prot and UniRef 90 databases (December 2011 releases). Contigs from both Vanuatu genomes were assembled by comparison to the previously published '*Ca*. Photodesmus katoptron' genome and were highly similar in structure. Reads were aligned against chromosome I, chromosome II, and plasmid sequences using Bowtie 0.12.7 (Li and Durbin, 2009). Bowtie was run in –v mode so that each read was allowed three mismatches with the consensus to maximize detection of polymorphic reads. rRNA operon sequences, which could not be assembled into full genome contigs, were not included in the analysis.

Samtools 0.1.17 was used to detect polymorphisms in aligned reads. Whole genome Samtools calls were used for sites that were polymorphic at greater than 25% of reads out of 100 reads. For sites that were polymorphic at less than 25% of reads, Samtools calls were not always accurate. For these sites, polymorphisms were detected manually in Samtools. For each library, 400 kilobases spread over both chromosomes and the plasmid were searched for sites polymorphic at 4-25% of reads. This is slightly less than half of the total 1Mb genome. Again for each library, 40kb of sequence on each chromosome was used to calculate the rate of likely sequencing errors (sites polymorphic at less that 4% of reads). A coverage depth of fifty reads was used so that every position was covered by the same number of reads.

Based on the number of samples represented in each sequencing library we were able to make predictions for the percentage of reads that would be divergent at a site if

the polymorphism was found in one light organ or multiple. We refer to the number of divergent reads at a given site as the 'polymorphism index.' The distribution of polymorphism index values around the predicted values was used to determine cut offs for determining the number of light organs in which a polymorphism was found (Fig. 5.3 and Fig. 5.4). We also used the distribution of polymorphism index values to infer which sites might represent intra-light organ polymorphism. A large number of polymorphic sites were divergent in less than 4% of reads and we assume these represent mainly sequencing errors. Errors were similarly common in each species. Polymorphic sites with an index value greater than 4% but less than the distribution for one light organ were classified as intra-light organ polymorphism. Intra-light organ polymorphisms were less frequent than sequencing errors. Numbers of polymorphisms were corrected to be out of 1 kilobase to be directly comparable. For calculating nonsynonymous versus synonymous substitutions, annotated genome assemblies of both Vanuatu samples were used. Fifty polymorphisms, spread over both chromosomes, were included for each species. These polymorphic sites represented an even distribution of different polymorphism index values.

Analysis of other species

Sequences for determining diversity in free-living *Vibrionaceae* were taken from GenBank (Appendix 3). Species were chosen to represent ecological diversity and genes were chosen for the amount of sequence available. Additionally, only genes with relatively low dN/dS ratios were used to limit analysis to neutral diversity. A possible confound in the comparison of these species to the anomalopid symbionts is that the

anomalopid symbionts were collected at one time and place, whereas their relatives have wide-ranging populations. To control for this, we used two data sets for each relative species, one that contained just strains collected in one location at approximately the same time (local), and one that contained strains collected from many locations over a broad time range (global). For luminous symbionts, strains were defined as 'local' if they were collected from the same host at the same location within a week of each other. For non-symbionts 'local' strains were defined as being collected from the same location within four months of each other. For 'global' data sets, no more than three strains from the same location were included.

Sequences for each gene in each species were aligned in ClustalX (Thompson et al., 1994) and analyzed in MEGA5 (Tamura et al., 2011). A maximum likelihood tree was constructed for each alignment to confirm the phylogenetic identity of each strain. When known, sequences from the species' closest relative were included and any sequences grouping with the relative were thrown out. When the species' closest relative was not known, any highly divergent sequences not grouping with the main clade were thrown out. Sequences were grouped as either 'local' or 'global' and the mean number of pairwise nucleotide differences was calculated using complete deletion within each group. Standard errors were calculated from 500 bootstrap replicates. Diversity estimates used in Figures 5.1 and 5.2 are averages across all genes analyzed for a species. Numbers of nonsynonymous and synonymous substitutions we also calculated in MEGA5 for all coding regions. For calculation of the number of shared polymorphisms from *A. fischeri* and *P. kishitanii*, *luxA* sequences were used for all local strains.

For *B. aphidicola* diversity estimates, data was taken from the five loci used in Funk et al. (2001) and six loci used in Abbot and Moran (2002). 'Local' values came from samples collected at the same time and place and 'global' values represent all samples. Numbers of nucleotide changes were averaged between the two species used in these studies and presented as polymorphisms per kilobase to be comparable. Nonsynonymous and synonymous substitution values were also taken from these data. For genetically monomorphic pathogens, reported polymorphism values were taken from Achtman et al. (2004), Barker et al. (2004), and Pearson et al. (2004).

Geographic divergence analysis

Twenty loci, sixteen coding genes, four rRNA and tRNA genes, and five noncoding spacer regions, were used to calculate divergence between the three '*Ca*. Photodesmus katoptron' populations. They were used to calculate the mean p-distance between the three populations in MEGA5 and to construct a maximum parsimony tree in PAUP. These genes are presented in the Table 5.5. For *A. fischeri* and *P. kishitanii*, strains collected from the same host from Appendix 3 were used to calculate mean p-distance. For *A. fischeri luxAB* and the spacer between them were used, and for *P. kishitanii* the genes *gyrB*, *luxA*, and *recA* were used.

Results

Genetic polymorphism within anomalopid symbionts

Genetic polymorphism in each anomalopid symbiont species is extremely low, with an average of 0.26 polymorphic sites per kilobase in '*Ca*. Photodesmus katoptron'

and 0.45 polymorphic sites per kilobase in '*Ca*. Photodesmus blepharus' Ppalp (Table 5.1). The pattern and levels of polymorphism differed slightly between the two symbiont species. Higher intra-light organ polymorphism levels were found on the '*Ca*. Photodesmus blepharus' Ppalp chromosomes than in '*Ca*. Photodesmus katoptron.' Higher intra-light organ polymorphism may be due to an increased mutation rate in '*Ca*. Photodesmus blepharus' Ppalp. Alternatively, it may be that '*Ca*. Photodesmus blepharus' Ppalp maintains bigger population sizes in the host. This is likely because the '*Ca*. Photodesmus blepharus' Ppalp host, *Photoblepharon palpebratus*, has larger light organs than the '*Ca*. Photodesmus katoptron' host, *Anomalops katoptron*. In both species low percentages of polymorphic sites were found in more than one light organ.

Anomalopid symbiont diversity compared to relatives and other host-associated bacteria

Rates of polymorphism in the '*Ca*. Photodesmus' species are much lower than in facultatively host-associated relatives (Fig. 5.1). We compared anomalopid symbiont polymorphism rates with those of six relatives, *A. fischeri*, *Photobacterium kishitanii*, *Photobacterium leiognathi*, *Photobacterium mandapamensis*, *Vibrio cholerae*, and *Vibrio harveyi*. All of these species are free-living and sometimes host-associated, and four, *A. fischeri*, *P. kishitanii*, *P. leiognathi*, and *P. mandapamensis*, are free-living luminous symbionts (Table 5.2, Appendix 3). Even the lowest polymorphism value in a free-living species, local polymorphism for *V. harveyi*, is over ten times higher than levels seen in the anomalopid symbionts. The highest rate found in relatives, *A. fischeri* global polymorphism, is over 150 times higher than polymorphism in anomalopid symbionts. Furthermore, the polymorphism values calculated for relatives are likely to be

underestimates. The genes used here are mostly conserved genes, which are likely to maintain only neutral variation. Additionally, rarefaction curves for genotypes of the free-living species suggest that at least double the number of strains would be needed to fully sample the population diversity (data not shown). The anomalopid symbiont values represent all variation, both neutral and not. Also, anomalopid symbiont polymorphism values are unlikely to change significantly with greater sampling, based on the large geographic scale over which low polymorphism is found (discussed below). The difference in diversity between anomalopid symbionts and free-living relatives demonstrates the large effect that host restriction can have on bacterial populations. These data also discount the hypothesis that all luminous symbionts might have low diversity because of adaptation to their long-term hosts. In fact, the luminous symbionts *A. fischeri*, *P. leoignathi*, and *P mandapamensis* have the highest polymorphism values reported here.

We compared anomalopid symbiont polymorphism rates to published rates from other host-restricted bacteria that have been reported to have very low diversity (Fig. 5.2). Anomalopid symbionts have lower polymorphism rates than the aphid symbiont *B. aphidicola* isolated from single a host species (Abbot and Moran, 2002; Funk et al., 2001) and similar levels to the genetically monomorphic, host restricted, pathogens *Yersinia pestis*, *Bacillus anthracis*, and *Mycobacterium tuberculosis* (Achtman, 2008; Achtman et al., 2004; Sreevatsan et al., 1997; Van Ert et al., 2007). The observation of extremely low genetic diversity in bacteria that are phylogenetically and ecologically distinct from those formerly reported, suggests that the pattern of low polymorphism in host-restricted bacteria may be fairly broad. To determine what factors might be causing

low polymorphism in the anomalopid symbionts, selection or high levels of drift, we looked at what types of substitutions are found in anomalopid symbionts.

Patterns in anomalopid symbiont substitutions

'*Candidatus* Photodesmus blepharus' Ppalp displays fairly high levels of nonsynonymous substitutions compared to synonymous substitutions, whereas '*Ca*. Photodesmus katoptron' does not (Table 5.3). The rate of nonsynonymous substitutions in '*Ca*. Photodesmus blepharus' Ppalp was higher than in the conserved genes from free-living relatives and similar to values reported for *B. aphidicola*. '*Candidatus* Photodesmus katoptron,' however, had nonsynonymous substitution values that were more similar to free-living relatives than to *B. aphidicola*. *Buchnera aphidicola* strains fix high numbers of nonsynonymous substitutions compared to synonymous substitutions (Abbot and Moran, 2002; Funk et al., 2001). This trend is predicted to occur if genetic drift dominates over selection. The nonsynonymous substitution values from '*Ca*. Photodesmus blepharus' Ppalp are consistent with the hypothesis that a high level of drift, not adaptation, is responsible for purging diversity in this species. However, the pattern found in '*Ca*. Photodesmus katoptron' is suggestive of purifying selection rather than drift. There are two possible explanations for the difference in patterns between the two species. First, it is possible that '*Ca*. Photodesmus katoptron' undergoes a less severe bottleneck during transfer between host generations and is therefore less subject to drift. Second, '*Ca*. Photodesmus katoptron' has lost more genes than '*Ca*. Photodesmus blepharus' Ppalp (Chapter VI), thus making the function of the remaining genes more

important and possibly increasing selection against nonsynonymous mutations (Allen et al., 2009).

The distribution of polymorphism index values found for both anomalopid symbionts (Fig. 5.3 and Fig. 5.4) indicate that very few polymorphisms are found in more than one light organ (Table 5.1). Excess numbers of rare alleles are predicted to exist if most substitutions are lost because the cells carrying them are not transferred to a new host generation. In this case, most of the variation seen will be in the form of recently evolved substitutions that are not shared by many cells. Our data do not allow us to assess the actual number of rare alleles in each anomalopid symbiont species directly since we cannot assign substitutions to particular alleles. However, comparing alleles between light organs should give similar values, assuming that light organs are colonized with a small number of cells, which is likely (McCann et al., 2003). For each species we find that only a small number of substitutions (1.9-20%, with a mean of 10.7%) are found in more than one light organ. For comparison, within the *lux* genes from *A. fischeri* and *P. kishitanii* strains analyzed here, 55.5% and 81.0% of substitutions are shared between more than one host, even when hosts were collected from distant locations. Furthermore, no substitutions were found to be shared between three different populations of '*Ca*. Photodesmus katoptron' (discussed below). These findings indicate that the anomalopid symbionts have more rare alleles than relatives, which is consistent with genetic drift during symbiont population bottlenecks between host generations. The high number of rare alleles suggests that, like *B. aphidicola*, the anomalopid symbionts lack genetic diversity because of drift caused by their obligate association with the host. It is not clear from this data if population dynamics of the host also affect anomalopid symbionts. To

test this, we assessed anomalopid symbiont diversity across the geographic range of a host.

Geographic patterns in anomalopid symbiont polymorphism

Genetic polymorphism values in '*Ca*. Photodesmus katoptron' are low through out a large portion of its geographic range. The host of this symbiont species, *A. katoptron*, is found in suitable habitats throughout the eastern Indian Ocean and the south Pacific Ocean (Morin et al., 1975; Johnson et al., 1988). We analyzed whole genome 454 sequencing data from the symbionts of four specimens collected near Cebu City in the Philippines. The polymorphism rates calculated from the Cebu City data were lower than those calculated for '*Ca*. Photodesmus katoptron' from Vanuatu (Table 5.1). Rates of polymorphism were 0.02, 0.04, and 0.07 polymorphisms per kilobase for chromosome I, chromosome II and the plasmid respectively. We believe that these rates are misleadingly low because the lower read depth and higher error rate (2% versus 1%) in the 454 reads made accurate polymorphism calling difficult. However, these values indicate that the Cebu City population of '*Ca*. Photodesmus katoptron' is unlikely to harbor significantly more diversity than the Vanuatu population.

'*Candidatus* Photodesmus katoptron' sequences from three locations widely distributed across its range, Christmas Island off Indonesia, near Cebu City in the Philippines, and Vanuatu, display very little divergence from each other (Fig. 5.5). We compared the sequences of 16 coding genes, four rRNA and tRNA genes, and five noncoding spacer regions for symbiont samples from each location (Table 5.5). Sequences were extremely similar across this broad geographic range. Twelve loci were

100% identical in all three samples. The Christmas Island sample contained one divergent site and the Vanuatu sample was divergent at nine sites distributed over eight loci. We note that the extent of divergence between samples correlated with the geographic distance between them (Fig. 5.5). To compare these divergence values with free-living symbionts, we calculated polymorphism rates between pairs of strains isolated from the same host for two free-living symbionts, *A. fischeri* and *Photobacterium kishitanii*. Even when collected simultaneously, free-living symbiont strains are much more divergent than '*Ca*. Photodesmus katoptron' collected over a wide range (Table 5.4). The low level of divergence between '*Ca*. Photodesmus katoptron' populations indicates that the symbiont exhibits low polymorphism throughout its host's geographic range. However, the fact that each symbiont population maintains unique polymorphisms indicates that there might be population structure over the geographic range. This hypothesis is further supported by the fact that the more geographically distant population is the most divergent.

Discussion

Genetic drift in anomalopid symbionts

Host associations influence obligate symbionts by causing repeated population bottlenecks and limiting effective population size in symbiotic bacteria. Anomalopid symbionts are obligately dependent on their hosts and have numerous similarities to other obligate symbionts, including genome reduction, increased AT nucleotide content, and an increased evolutionary rate (Chapter II; Chapter III; McCutcheon and Moran, 2012; Wernegreen, 2002). These characteristics are all indicative of high levels of drift. We

compared diversity patterns in anomalopid symbionts to patterns in other obligate symbionts to see if further drift related similarities existed. Like the obligate symbiont *B. aphidicola*, anomalopid symbionts have low levels of intra-specific polymorphism and an excess of rare alleles. One anomalopid symbiont species, '*Ca*. Photodesmus blepharus' Ppalp, also has high number of nonsynonymous substitutions, though this pattern is not found in the '*Ca*. Photodemus katoptron.' Taken together, these findings suggest that anomalopid symbionts are subject to high levels of drift. As with other obligate symbionts, drift is likely caused by bottlenecks between host generations and small population sizes due to host restriction. A low mutation rate alone could not cause this pattern because it would not lead to an excess of rare or nonsynonymous substitutions. Another possible explanation for low polymorphism in anomalopid symbionts, recent selective sweeps, would be unlikely to occur independently in both species at the same time. The confirmation of low polymorphism and high drift in anomalopid symbionts suggests that drift may be generally important in causing low intraspecific diversity in host-restricted bacteria.

The method by which bacteria colonize new host generations is likely to have a profound influence on their population dynamics. Anomalopid symbionts are environmentally acquired rather than vertically transmitted like most obligate symbionts (Chapter III; Haygood et al., 1993). It is likely that a relatively small number of cells initially colonize the light organs of hosts and then increase exponentially in number, creating a population bottleneck and expansion. The size of the colonizing population will determine the extent of the bottleneck. The number of colonizing cells reported for other bioluminescent symbioses is much smaller than the number of *B. aphidicola* cells

transferred between host generations (McCann et al., 2003; Mira et al., 2002). A smaller number of colonizing cells could account for why anomalopid symbionts have lower intra-specific diversity levels than *B. aphidicola*. However, another possible cause could be the presence of DNA repair mechanisms. *Buchnera aphidicola* has lost most genes involved in DNA repair and it is not known how much the inability to repair DNA accounts for their increased mutation rate (Sharples, 2009; Tamas et al., 2002). The anomalopid symbiont '*Ca*. Photodesmus katoptron' has maintained nearly all DNA repair genes found in relatives (Chapter III) and might therefore have a lower mutation rate than *B. aphidicola*. Given the significant difference between *B. aphidicola* and anomalopid symbionts in polymorphism, it is possible that the inability to repair DNA accounts for a larger number of mutations than previously thought.

One finding is inconsistent with high levels of drift, which is the low number of nonsynonymous mutations in '*Ca*. Photodesmus katoptron.' Since '*Ca*. Photodesmus katoptron' has very low levels of genetic diversity and an excess of rare alleles, it is likely to be experiencing high levels of genetic drift. We propose that the genome of '*Ca*. Photodesmus katoptron' may be so reduced that selection on the remaining genes has become sufficiently strong to overcome the effect of drift in fixing nonsynonymous mutations. A decrease in the effect of drift over time in obligate symbionts has been reported in other lineages (Allen et al., 2009). This hypothesis is testable by looking at substitution patterns at a finer scale, rather than genome wide. At loci that should not be under selection, such as intergenic spacers or pseudogenes, rates of substitution should be higher than in coding regions, since mutations are presumably being purged from coding loci by purifying selection.

Understanding the cause of genetic monomorphism

A variety of host-associated bacteria with very low genetic diversity have been reported. They are mainly intracellular and include the obligate symbiont *B. aphidicola* and the pathogens *Bacillus anthracis* (anthrax), *Mycobacterium tuberculosis* (tuberculosis), and *Yersinia pestis* (plague) (Achtman, 2008; Baker et al., 2004; Pearson et al., 2004; Sreevatsan et al., 1997; Van Ert et al., 2007). Anomalopid symbionts have similar levels of diversity to these genetically monomorphic pathogens. In each of these cases a hypothesized cause of low diversity is relatively recent origins or population expansion of pathogenic strains. However, these theories have not been tested and are not consistent with the generation time or likely population size of these pathogens. A recent origin or expansion are unlikely explanations for low diversity in anomalopid symbionts because the host family is fairly old (Carroll, 1988). Likely causes for genetic monomorphism in host-associated bacteria therefore need more investigation.

In *B. aphidicola*, genetic drift plus population bottlenecks and high dispersal in the aphid host are thought to lead to low polymorphism and little geographic structure in symbionts (Abbot and Moran, 2002; Funk et al., 2001). Anomalopid host populations are unlikely to have the same dynamics as aphids. Fish are much longer lived and do not have seasonal population bottlenecks and expansions as aphids do. Furthermore flashlight fish are extremely photophobic, only leaving their caves on dark nights (Morin et al., 1975), so they are unlikely to disperse long distances as adults. It is not known if larvae disperse in this group, though some evidence suggests that they do not (Chapter III; Meyer-Rochow, 1976). The divergence that is found between populations of the anomalopid symbiont '*Ca*. Photodesmus katoptron' correlates with geographic distance,

suggesting that some geographic structure may exist in this species. We conclude that host bottlenecks and dispersal are therefore not responsible for the widespread lack of diversity in anomalopid symbionts. We suggest that a low mutation rate may play an important role, based on the difference in DNA repair genes between *B. aphidicola* and anomalopid symbionts.

The affect of host dependence on bacteria

Theory predicts that long-term host associations will select for certain symbiont genotypes and thus eliminate diversity (Sachs et al, 2011). Some evidence of selection on luminous symbionts has been found (Wollenberg et a., 2011). Here however, we find that facultative luminous symbionts are highly diverse even when collected from the same host species and same location, which is inconsistent with adaptation to the host. Facultative luminous symbionts inhabit a variety of diverse habitats in addition to being luminous symbionts (Dunlap et al., 2007; Dunlap, 2009). It is possible that the addition of an extra habitat, the light organ, creates diversity by causing divergence and adaptation. This suggests that rather than thinking of bacteria-host interactions in terms of how long-term or dependent they are, we should consider the breadth of niches occupied by bacteria, as this is likely to be highly influential on bacterial diversity and evolution. This framework is also consistent with findings from obligate symbionts.

Acknowledgements

This work will be coauthored with Paul V. Dunlap when published. Specimens were obtained with assistance from Quality Marine Wholesale Suppliers, Grant Norton of Sustainable Reef Suppliers, Brad Remmer of Sea Dwelling Creatures and Fish Doctors Aquarium of Ypsilanti, Michigan. We thank the University of Michigan DNA Sequencing Core for help and advice with sequencing. We also thank Jeffrey R. de Wet, of the University of Michigan Bioinformatics Analysis Core, for carrying out sequence assembly and annotation. Departmental Block Grant awards from the University of Michigan Department of Ecology and Evolutionary Biology funded this work.

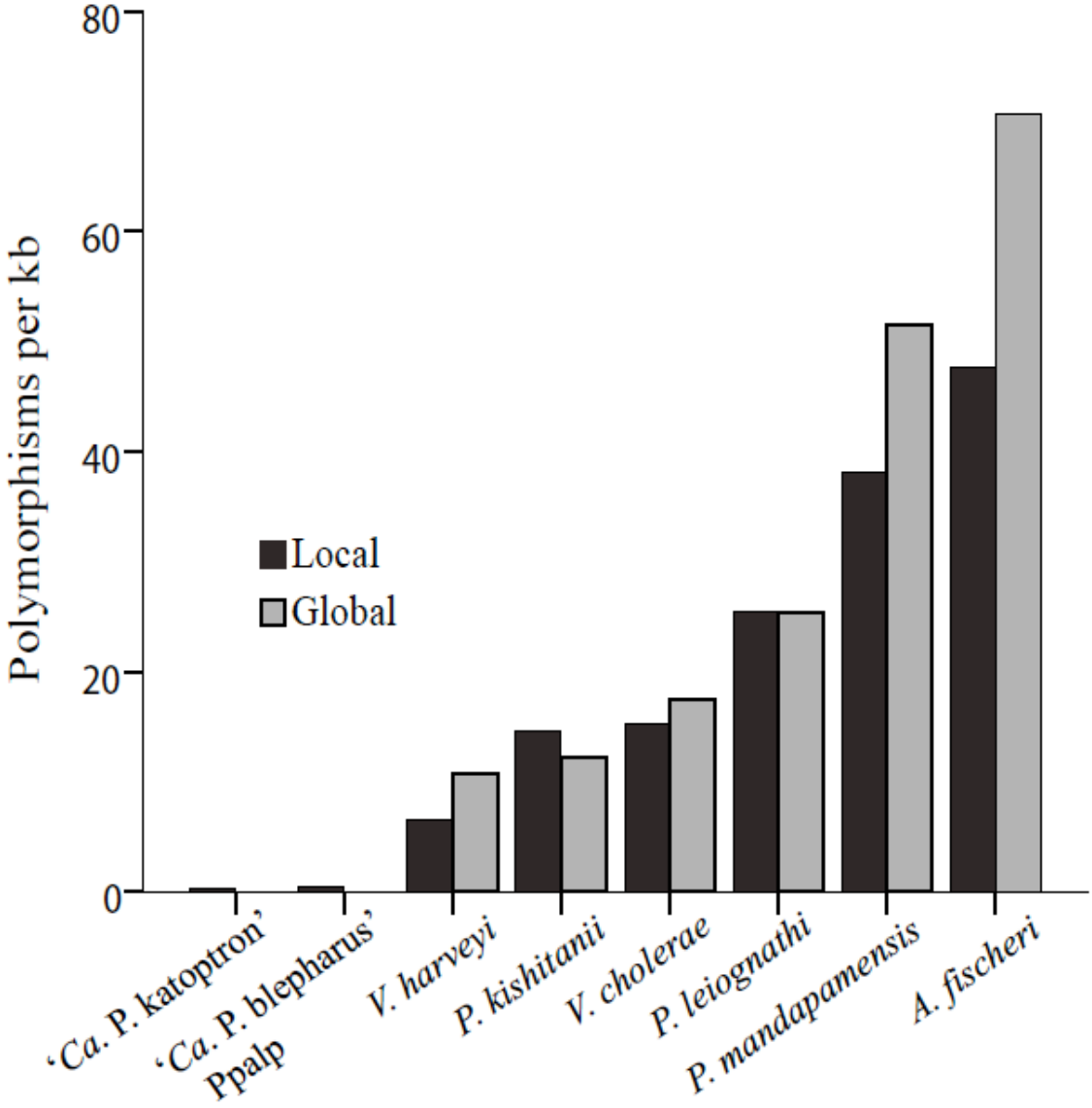

Fig. 5.1. Diversity estimates for the anomalopid symbionts and various facultatively host-associated relatives. 'Local' values of polymorphisms per kilobase are shown for the anomalopid symbionts '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp. Diversity estimates for facultatively host-associated bacteria are pairwise numbers of nucleotide differences per kilobase. *Photobacterium kishitanii*, *P. leiognathi*, *P. mandapamensis*, and *A. fischeri* are luminous symbionts. All values are mean values from multiple genes.

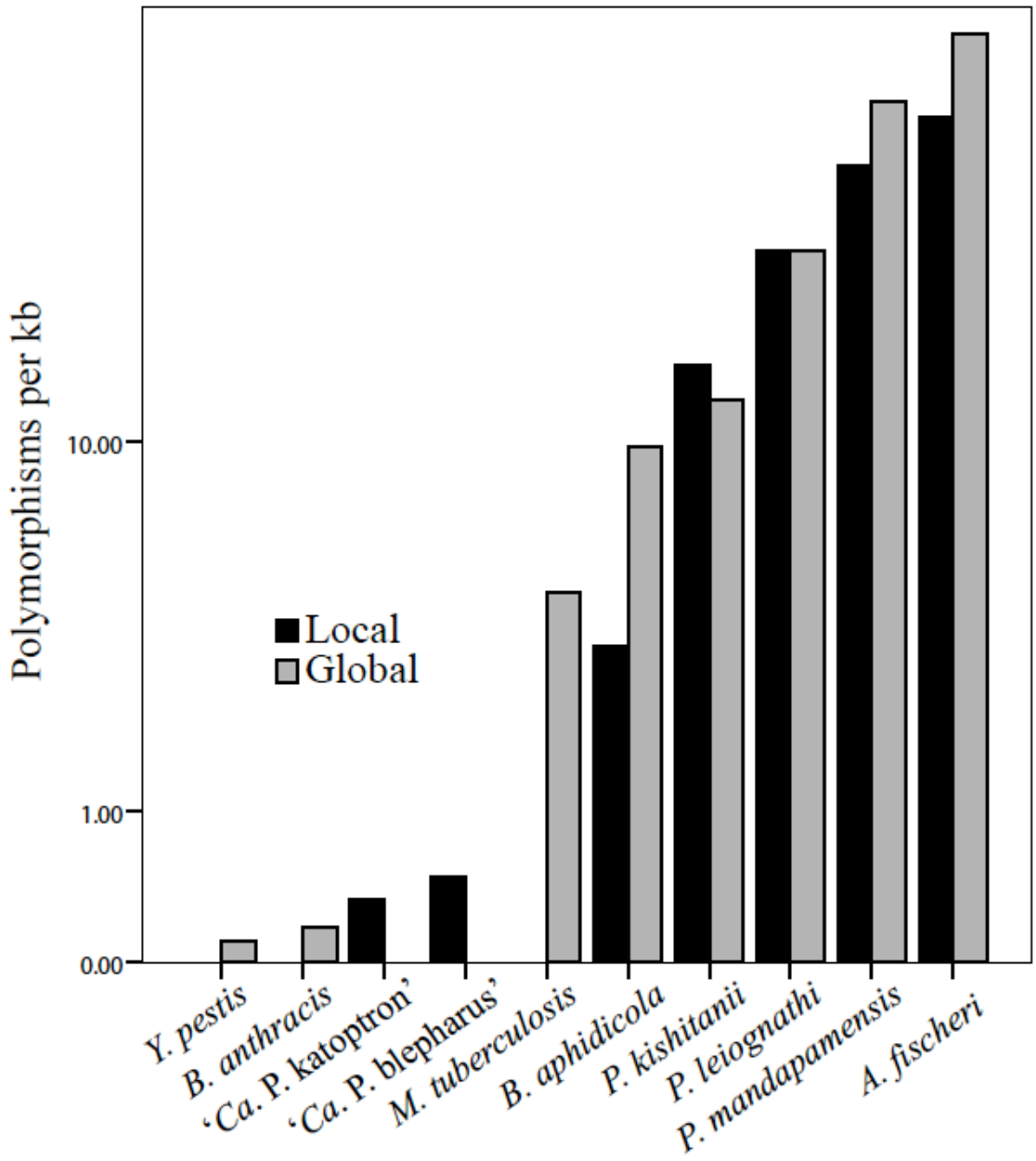

Fig. 5.2. Diversity estimates for anomalopid symbionts, free-living luminous symbionts, and genetically monomorphic obligately host-associated bacteria. Diversity estimates for facultatively luminous symbionts, *Photobacterium kishitanii*, *P. leiognathi*, *P. mandapamensis*, and *A. fischeri*, are pairwise numbers of nucleotide differences per kilobase. All other estimates are numbers of polymorphisms per kilobase. Only 'local' diversity values are shown for the anomalopid symbionts '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp. Only 'global' diversity values are shown for the genetically monomorphic pathogens *Yersinia pestis*, *Bacillus anthracis*, and *Mycobacterium tuberculosis*. *Buchnera aphidicola* is an obligate symbiont shown to have very low intra-species diversity. All values are mean values from multiple genes. The y-axis is shown on a logarithmic scale.

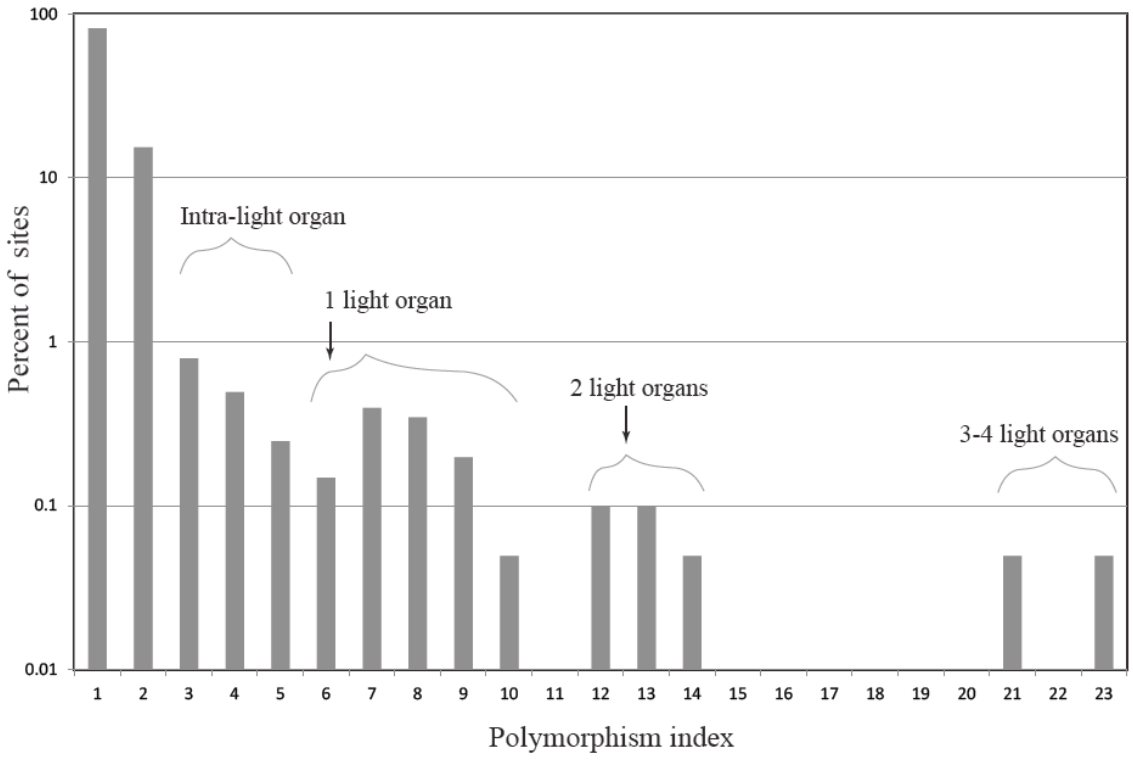

Fig. 5.3. Percentage of sites polymorphic for a given polymorphism index (number polymorphic reads at a site) in '*Ca*. Photodesmus katoptron.' Arrows indicate the predicted number of reads for polymorphisms unique to one, two, or three to four light organs. Brackets delineate cut offs used to determine intra-light organ polymorphisms, single light organ polymorphisms, and multi light organ polymorphisms. The y-axis is shown on a logarithmic scale.

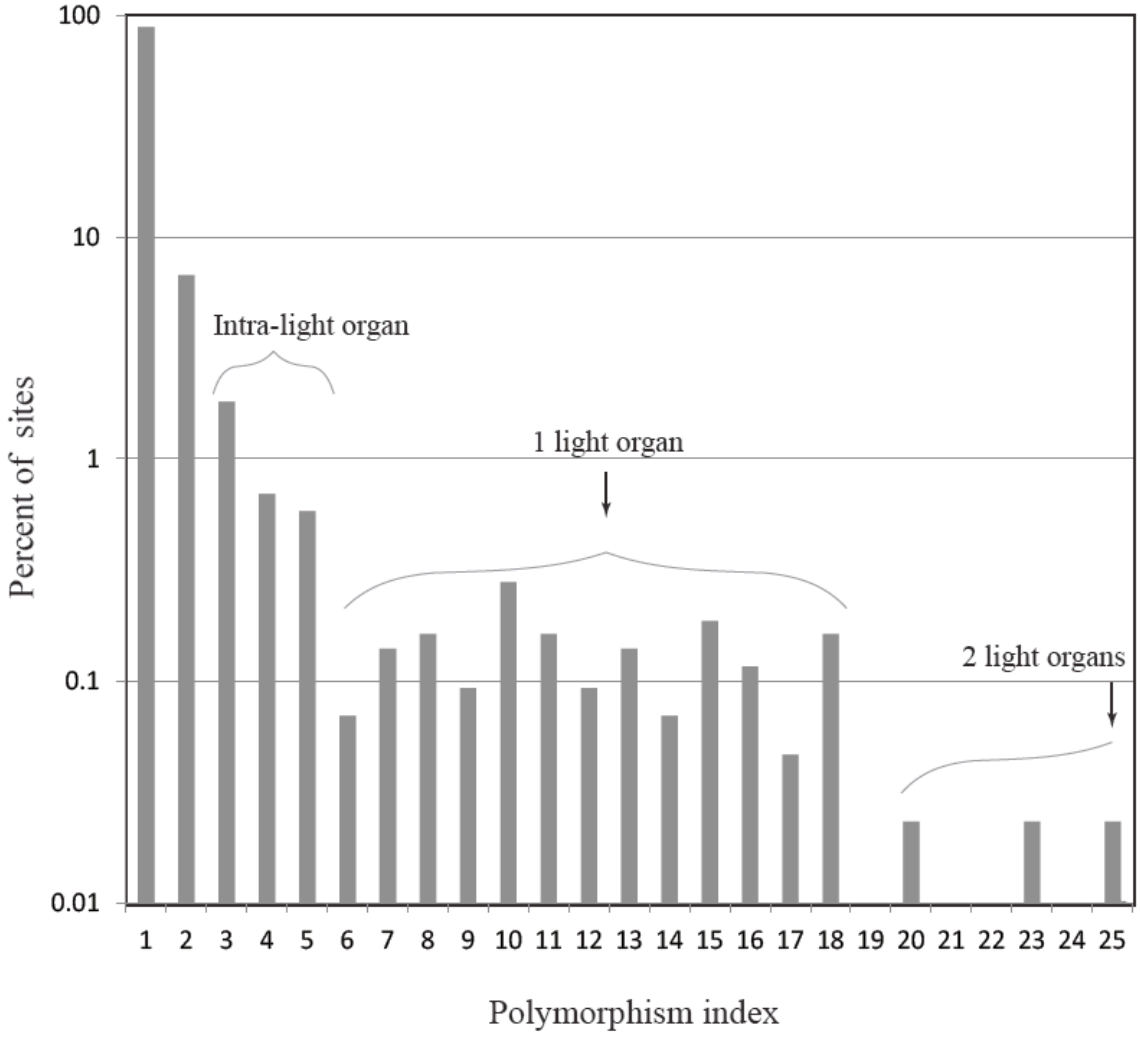

Fig. 5.4. Percentage of sites polymorphic for a given polymorphism index (number polymorphic reads at a site) in '*Ca*. Photodesmus blepharus' Ppalp. Arrows indicate the predicted number of reads for polymorphisms unique to one, two, or three to four light organs. Brackets delineate cut offs used to determine intra-light organ polymorphisms, single light organ polymorphisms, and multi light organ polymorphisms. The y-axis is shown on a logarithmic scale.

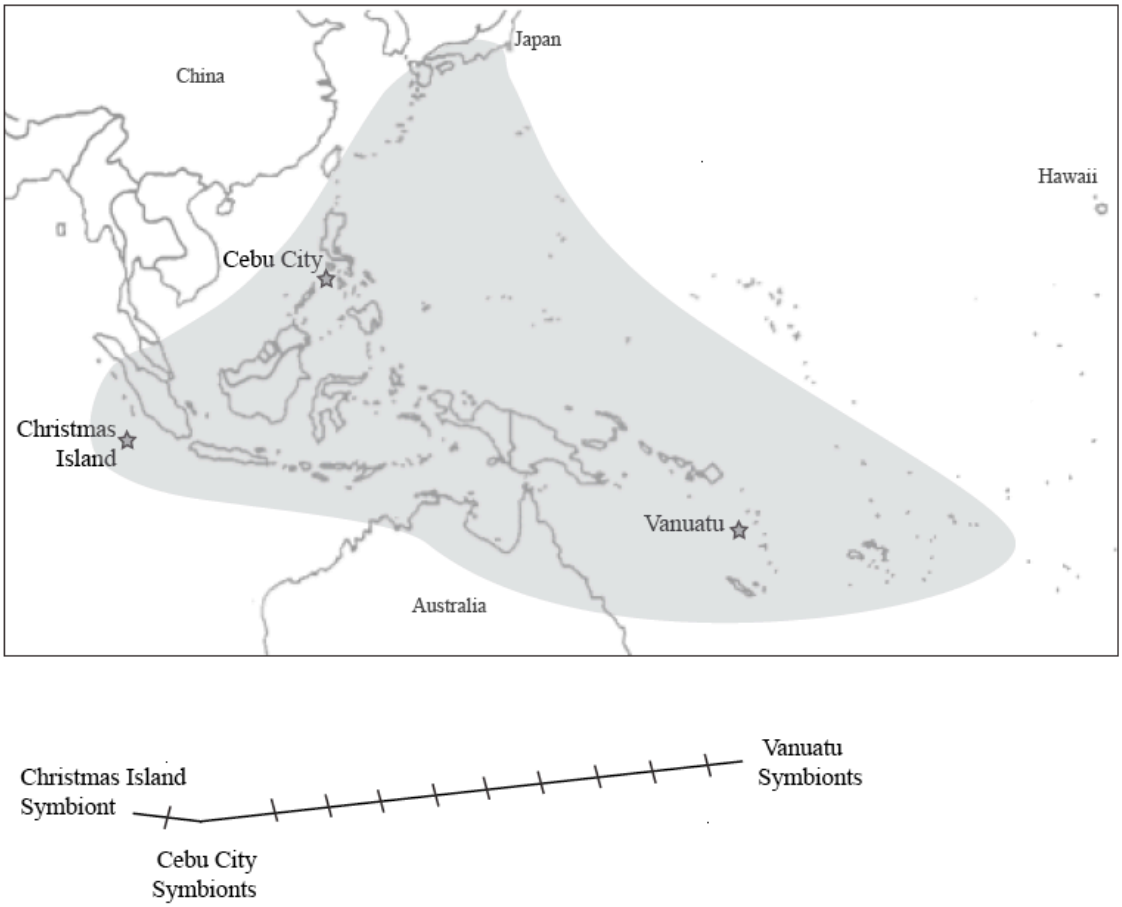

Fig. 5.5. '*Candidatus* Photodesmus katoptron' host distribution, sample sites and genetic divergence. Geographic range of the host *A. katoptron* is shown in grey. Sample sites for '*Ca*. Photodesmus katoptron' are labeled with stars. Genetic divergence between symbiont populations is shown with an unrooted phylogram generated by maximum parsimony. Tick marks on branches indicate the number of changes.

Table 5.1. Polymorphism values for anomalopid symbionts. Single nucleotide polymorphisms per kilobase are shown for chromosome I, chromosome II, and the plasmid for each species. Total polymorphism values as well as the intra and inter-light organ fractions are shown. The percentage of SNPs found in more than one light organ are also shown.

Sample	Molecule	Specimens	Total SNPs	Intra L.O.	Inter L.O.	2 L.O.	3-4 L.O.
'*Ca*. Photodesmus katoptron'	Chr I	8	0.20	0.09	0.12	9.1%	0
	Chr II		0.23	0.12	0.11	10.6%	4.3%
	Plasmid		0.34	0.14	0.21	0	0
'*Ca*. Photodesmus blepharus' Ppalp	Chr I	4	0.57	0.40	0.17	1.9%	0
	Chr II		0.52	0.36	0.17	8.3%	0
	Plasmid		0.27	0.05	0.22	20%	0

Table 5.2. Polymorphism values for free-living, host-associated *Vibrionaceae* species. Pairwise numbers of nucleotide differences per kilobase are shown along with standard errors (S.E.) for each species and gene. 'Local' values are from strains collected at the same time and place and 'global' values are from strains collected over a variety of locations and times.

Species	Gene	Length(bp)	# of Local strains	Local	S.E.	# of Global strains	Global	S.E.
Aliivibrio fischeri	*recA*	690/531[a]	20	9.8	2.4	14	3	1
	luxA	1784/661[a]	32	70.1[b]	3.5	21	138.4	7.9
Photobacterium kishitanii	*gyrB*	1213	27	16.7	2.6	48	13.4	2.1
	recA	785	12	16.3	3.3	36	13.4	2.7
	luxA	681	28	10.7	2.8	46	10	2.3
Photobacterium leiognathi	*gyrB*	791	NA			30	16.9	2.4
	luxA	1991	NA			30	33.9	2.5
Photobacterium mandapamensis	*gyrB*	1209/1217[a]	13	16.9	2.1	11	18.8	2.1
	luxA	2657/617[a]	22	58.3	2.6	19	55.3	5.8
Vibrio cholerae	*gyrB*	1119/1300[a]	45	13.7[b]	2.0	29	16.3	1.7
	recA	690/994[a]	45	16.6[b]	2.8	55	18.6	2.7
Vibrio harveyi	*gyrB*	595	24	2.4	1.0	28	5.4	1.7
	recA	682	22	12.8	2.8	22	16.2	3.5
	luxA	492	24	4.3	1.3	NA		

[a] local length/global length

[b] Two groups of local sequences analyzed separately and then averaged.

Table 5.3. Rates of nonsynonymous and synonymous substitutions for anomalopid symbionts, facultatively host-associated relatives and *B. aphidicola*.

Species	locus	# N sites/kb	# S sites/kb	N/S
'*Ca*. Photodesmus katoptron'	Chr I	0.02	0.11	0.17
	Chr II	0.02	0.17	0.25
'*Ca*. Photodesmus blepharus' Ppalp	Chr I	0.21	0.23	0.91
	Chr II	0.11	0.18	0.61
Vibrio harveyi	*gyrB*	0.17	3.70	0.05
	recA	0	15.04	0
	luxA	0.73	3.67	0.20
Photobacterium kishitanii	*gyrB*	0.55	14.51	0.04
	recA	1.60	13.31	0.12
	luxA	1.70	8.66	0.20
Vibrio cholerae	*gyrB*	0.72	14.30	0.05
	recA	0.24	17.34	0.01
Photobacterium leiognathi	*gyrB*	0.17	17.72	0.01
	luxA	9.49	19.19	0.49
Photobacterium mandapamensis	*gyrB*	2.85	29.84	0.10
	luxA	11.09	27.24	0.41
Aliivibrio fischeri	*recA*	0	6.35	0
	luxA	39.38	74.94	0.53
Buchnera aphidicola Ua				0.92
Buchnera aphidicola Po				0.87

Table 5.4. Genetic divergence between symbiont populations for anomalopid symbionts and free-living luminous symbionts. Mean uncorrected p-distances between the three '*Ca*. Photodesmus katoptron' populations and between pairs of *P. kishitanii* and *A. fischeri* strains from the same host individual are shown.

Species	# of pairs	Total bp	# loci used	mean p-distance
'*Ca*. Photodesmus katoptron'	NA	20952	25	0.000
Photobacterium kishitanii	14	2664	3	0.013
Aliivibrio fischeri	33	1760	3	0.046

Table 5.5. Genes used to investigate divergence between populations of '*Ca*. Photodesmus katoptron.' Numbers of polymorphisms in each population are shown.

Gene	Christmas Island	Cebu City	Vanuatu
16S rRNA	0	0	**2**
23S rRNA	0	0	0
alanine tRNA	0	0	0
isoleucine tRNA	0	0	0
ITS 1 (16S-Ala tRNA)	0	0	0
ITS 2 (Ala tRNA-Ile tRNA)	0	0	0
ITS 3 (Ile tRNA-16S)	0	0	0
atpA	0	0	0
dihydroorotate dehydrogenase	0	0	**1**
gapA	0	0	**1**
glycyl tRNA synthase	0	0	0
gyrA	0	0	0
gyrB	0	0	**1**
luxABE and spacers	0	0	**2**
protease IV	**1**	0	0
pyrH	0	0	0
recA	0	0	**1**
rpoA	0	0	0
Rnf reducing system gene	0	0	0
sulfite reductase	0	0	0
topA	0	0	**1**

References

Abbot, P., Moran, N.A., 2002. Extremely low levels of genetic polymorphism in endosymbionts (Buchnera) of aphids (Pemphigus). Mol. Ecol. 11, 2649-2660.

Achtman, M., 2008. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu. Rev. Microbiol. 62, 53-70.

Achtman, M., Morelli, G., Zhu, P.X., Wirth, T., Diehl, I., Kusecek, B., Vogler, A.J., Wagner, D.M., Allender, C.J., Easterday, W.R., Chenal-Francisque, V., Worsham, P., Thomson, N.R., Parkhill, J., Lindler, L.E., Carniel, E., Keim, P., 2004. Microevolution and history of the plague bacillus, Yersinia pestis. Proc. Natl. Acad. Sci. U. S. A. 101, 17837-17842.

Allen, J.M., Light, J.E., Perotti, M.A., Braig, H.R., Reed, D.L., 2009. Mutational Meltdown in Primary Endosymbionts: Selection Limits Muller's Ratchet. Plos One 4.

Baker L., Brown T., Maiden M. C., Drobniewski F., 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg. Infect. Dis. 10, 1568–77

Carroll, R., 1988. Vertebrate paleontology and evolution. W.H. Freeman, New York, NY.

Chevreux, B., Pfisterer, T., Suhai, S., 2002. Automatic assembly and editing of genomic sequences. In: Suhai, S. (Ed.), *Genomics and Proteomics - Functional and Computational Aspects.* Kluwer Academic Publishers*,* New York, NY, 51-65.

Delcher, A. L., Bratke, K. A., Powers, E. C., Salzberg, S. L., 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics 23, 673-679.

Dunlap, P.V., 2009. Bioluminescence, microbial. In: Schaechter, M. (Ed.), Encyclopedia of microbiology. Elsevier, Oxford, pp. 45-61.

Dunlap, P.V., Ast, J.C., Kimura, S., Fukui, A., Yoshino, T., Endo, H., 2007. Phylogenetic analysis of host-symbiont specificity and codivergence in bioluminescent symbioses. Cladistics 23, 507-532.

Dunlap, P.V., Davis, K.M., Tomiyama, S., Fujino, M., Fukui, A., 2008. Developmental and Microbiological Analysis of the Inception of Bioluminescent Symbiosis in the Marine Fish Nuchequula nuchalis (Perciformes: Leiognathidae). Appl. Environ. Microbiol. 74, 7471-7481.

Dunlap, P.V., Jiemjit, A., Ast, J.C., Pearce, M.M., Marques, R.R., Lavilla-Pitogo, C.R., 2004. Genomic polymorphism in symbiotic populations of Photobacterium leiognathi. Environmental Microbiology 6, 145-158.

Fry, A.J., Wernegreen, J.J., 2005. The roles of positive and negative selection in the molecular evolution of insect endosymbionts. Gene 355, 1-10.

Funk, D.J., Wernegreen, J.J., Moran, N.A., 2001. Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-buchnera association. Genetics 157, 477-489.

Haygood, M.G., 1993. Light organ symbioses in fishes. Crit. Rev. Microbiol. 19, 191-216.

Hendry, T.A., Dunlap, P.V., 2011. The uncultured luminous symbiont of Anomalops katoptron (Beryciformes: Anomalopidae) represents a new bacterial genus. Mol. Phylogenet. Evol.

Johnson, G.D., Rosenblatt, R.H., 1988. Mechanisms of light organ occlusion in flashlight fishes, family Anomalopidae (Teleostei, Beryciformes), and the evolution of the group. Zoological Journal of the Linnean Society 94, 65-96.

Kaeding, A.J., Ast, J.C., Pearce, M.M., Urbanczyk, H., Kimura, S., Endo, H., Nakamura, M., Dunlap, P.V., 2007. Phylogenetic diversity and cosymbiosis in the bioluminescent symbioses of "Photobacterium mandapamensis". Appl. Environ. Microbiol. 73, 3173-3182.

Kessel, M., 1977. The ultrastructure of the relationship between the luminous organ of the teleost fish *Photoblepharon palpebratus* and its symbiotic bacteria. Cytobiologie 15, 145-158.

Li H., Durbin R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-60.

McCann, J., Stabb, E.V., Millikan, D.S., Ruby, E.G., 2003. Population dynamics of Vibrio fischeri during infection of Euprymna scolopes. Appl. Environ. Microbiol. 69, 5928-5934.

McCutcheon, J. P., Moran, N. A., 2012, Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10, 13-26.

Meyer-Rochow, V. B., 1976. Some observations on spawning and fecundity in luminescence fish *Photoblepharon palpebratus*. Mar. Biol. 37, 325-328.

Mira, A., Moran, N.A., 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microb. Ecol. 44, 137-143.

Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873-2878.

Morin, J.G., Harrington, A., Nealson, K., Krieger, N., Baldwin, T.O., Hastings, J.W., 1975. Light for all reasons - versatility in behavioral repertoire of flashlight fish. Science 190, 74-76.

Pearson T., Busch J. D., Ravel J., Read T. D., Rhoton S. D., U'ren, J. M., Simonson, T. S., Kachur, S. M., Leadem, R. R., Cardon, M. L., Van Ert, M. N., Huynh, L. Y., Fraser, C. M., Keim, P., 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. Proc. Natl. Acad. Sci. U. S. A. 101:13536–41.

Preheim, S.P., Boucher, Y., Wildschutte, H., David, L.A., Veneziano, D., Alm, E.J., Polz, M.F., 2011. Metapopulation structure of Vibrionaceae among coastal marine invertebrates. Environmental Microbiology 13, 265-275.

Sachs, J.L., Essenberg, C.J., Turcotte, M.M., 2011. New paradigms for the evolution of beneficial infections. Trends in Ecology & Evolution 26, 202-209.

Sharples, G. J., 2009. For absent friends: life without recombination in mutualistic γ-proteobacteria. Trends in Microbiology 17, 233-242.

Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M., 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. U. S. A. 94, 9869-9874.

Tamas, I., Klasson, L., Canback, B., Naslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandstrom, J.P., Moran, N.A., Andersson, S.G.E., 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 296, 2376-2379.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. ClustalW - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matric choice. Nucleic Acids Res. 22, 4673-4680.

Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J., Sarma-Rupavtarm, R., Distel, D.L., Polz, M.F., 2005. Genotypic diversity within a natural coastal bacterioplankton population. Science 307, 1311-1313.

Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T.S., U'Ren, J.M., Kachur, S.M., Leadem-Dougherty, R.R., Rhoton, S.D., Zinser, G., Farlow, J., Coker, P.R., Smith, K.L., Wang, B.X., Kenefic, L.J., Fraser-Liggett, C.M., Wagner, D.M., Keim, P., 2007. Global Genetic Population Structure of Bacillus anthracis. Plos One 2.

Wernegreen, J.J., 2002. Genome evolution in bacterial endosymbionts of insects. Nature Rev. Genet. 3, 850-861.

Wernegreen, J.J., Funk, D.J., 2004. Mutation exposed: A neutral explanation for extreme base composition of an endosymbiont genome. J. Mol. Evol. 59, 849-858.

Wernegreen, J.J., Moran, N.A., 1999. Evidence for genetic drift in endosymbionts (Buchnera): Analyses of protein-coding genes. Mol. Biol. Evol. 16, 83-97.

Wollenberg, M.S., Ruby, E.G., 2012. Phylogeny and fitness of Vibrio fischeri from the light organs of Euprymna scolopes in two Oahu, Hawaii populations. Isme Journal 6, 352-362.

Woolfit, M., Bromham, L., 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20, 1545-1555.

CHAPTER VI

Genomic divergence between obligate luminous symbiont species

Abstract

Genome reduction is a common phenomenon in obligate bacterial symbionts, and understanding how the process of genome reduction occurs is important for understanding symbiont evolution. Genomic comparisons of related species of obligate symbionts have uncovered several common trends in the process of genome reduction, including high levels of genetic drift early in reduction and genome stasis later in reduction, but it is not known how broad these trends are. Here we compare the sequenced genomes of two species of anomalopid flashlight fish symbionts to test the generality of high drift and genome stasis in reduced genomes of obligate symbionts. We find the two anomalopid symbiont species to be highly divergent at the nucleotide and amino acid level, as would be expected under high genetic drift. However, the gene content and gene order is largely conserved between the two species, suggesting that their genomes may be relatively static. This study indicates that the process of genome reduction may be quite similar in evolutionarily and ecologically distinct symbionts. Additionally, we find that the most divergent genes between the two anomalopid symbiont species contain a high number of chemotaxis and motility genes, suggesting

that these pathways may be important in host specificity and ecological divergence between the symbionts. We also discuss our plans to further test the results of this work.

Introduction

Genome reduction is a dynamic process, characterized by high levels of genetic drift in early stages and extreme genome stasis in late stages (Burke and Moran, 2011; McCutcheon and Moran, 2012; Moran and Mira, 2001; Tamas et al., 2002; Toh et al., 2006). These patterns are found in multiple symbiont lineages that share numerous characteristics thought to cause genetic drift and genome stasis, including being intracellular, vertically transmitted, and non-recombining. Recent genomic analyses of the luminous symbionts of anomalopid flashlight fish have demonstrated that these bacterial symbionts have reduced genomes. Unlike other symbionts with reduced genomes, anomalopid symbionts are extracellular, environmentally acquired, and have the potential for recombination. Because of these characteristics, genome reduction in anomalopid symbionts may have proceeded differently than in intracellular symbionts. Elucidating the evolutionary processes that have led to genome reduction in anomalopid symbionts, then, provides a test of the generality of theories of genome reduction. To investigate the process of genome reduction in anomalopid flashlight fish symbionts, we compared the genomes of two different species of anomalopid symbionts with reduced genomes. We tested the anomalopid symbionts for patterns common to intracellular symbiont lineages: a history of strong genetic drift and genome stasis.

For intracellular bacteria, the initial restriction to host cells causes elevated levels of genetic drift by greatly decreasing the symbiont's effective population size.

Additionally, restriction to the host environment relaxes selection on genes needed only outside the host (McCutcheon and Moran, 2012). High levels of genetic drift cause increased fixation of genomic changes, including nucleotide substitutions, gene loss, and genome rearrangements in young intracellular bacterial lineages (McCutcheon and Moran, 2012; Moran et al., 2009; Moran and Mira, 2001; Tamas et al., 2002). Typically, gene loss occurs by deletion of multiple genes simultaneously, even if some genes are likely important for cell function, indicating that the gene loss is not due to selection (Moran et al., 2009). Excessively high numbers of nucleotide substitutions in obligate symbionts are nonsynonymous, more than are likely to have been fixed by positive selection, and therefore likely deleterious (Abbot and Moran, 2002; Funk et al., 2001; Moran, 1996; Wernegreen and Moran, 1999).

Over long periods of time the consequences of genetic drift - high rates of nucleotide substitutions and gene loss - are predicted to be harmful for symbiotic bacteria (Moran, 1996). Therefore, when genomes become highly reduced selection should limit the rate of change, causing genome stasis. Consistent with this, the rate of both nucleotide change and gene loss slow with time in symbiotic lineages (Allen et al., 2009; McCutcheon and Moran, 2012; Tamas et al., 2002). Additionally, stasis may increase over time because reduced genomes may have less capacity for genome rearrangements, due to the loss of mobile genetic elements, repeated sequences, and DNA recombination genes(McCutcheon and Moran, 2012; Tamas et al., 2002). Some of the oldest intracellular symbiont lineages display striking genome stasis, or a lack of genomic change, between species. For example, some *Buchnera aphidicola* genomes have undergone no genomic rearrangements in the last 50 million years (Tamas et al., 2002).

The luminous symbionts of anomalopid flashlight fish, like intracellular symbionts, are obligately dependent on their hosts and have reduced genomes (Chapter III). Since anomalopid symbionts are potentially relatively old (the order Beryciformes dates to 70 mya (Carroll, 1988), they may have undergone sufficient genomic reduction to cause genome stasis. However, the '*Ca*. Photodesmus katoptron' genome is on the larger end of reduced genomes, so gene loss could be ongoing. Additionally, anomalopid symbionts differ from the intracellular symbionts displaying genome stasis because they have DNA recombination genes and some repetitive DNA sequences (rRNA operons), creating the potential for genomic rearrangements. These characteristics might make genome stasis between species less likely. To test for genome stasis, we compared gene order between two anomalopid symbiont species. Previous work had focused on the genome of one anomalopid symbiont species, '*Candidatus* Photodesmus katoptron.' Here we sequenced the genome of a second species, '*Candidatus* Photodesmus blepharus' Ppalp from the host *Photoblepharon palpebratus* (Chapter IV).

Even if the genomes of anomalopid symbionts are relatively static, gene content would likely not be identical between different species because the symbionts have different ecologies. The two anomalopid symbiont species analyzed here have specific interactions with different host species, *Anomalops katoptron* and *P. palpebratus*, in spite of the fact that the hosts overlap in geographic range and could theoretically be colonized by either symbiont species. The genomes of these symbionts species could provide some explanation for the host specificity. We investigated which genes differ between the two species to form hypotheses for how host specificity might be controlled. We also discuss our plans for future work to test these hypotheses.

Since anomalopid symbionts are extracellular and environmentally acquired (Chapter III, Haygood, 1993), they may be less subject to genetic drift than intracellular symbionts and may not show the same genomic patterns described above. High evolutionary rates and high nonsynonymous substitution rates have been found in anomalopid symbiont species, suggesting that the bacteria may be influenced by genetic drift. However, these patterns vary depending on the species of anomalopid symbiont and are not always consistent with high drift (Chapter V). To clarify this issue, we tested anomalopid symbionts for patterns caused by high genetic drift. We compared the two symbiont genomes for higher than expected divergence, as would indicate an evolutionary history of high drift. We predict that if anomalopid symbionts have been subject to high levels of genetic drift they should show genome wide high rates of nucleotide divergence. In contrast, selection would not affect divergence at all loci in the genome. We also assessed how many genes had likely been lost in each species by multi-gene deletion, as would indicate drift. Additionally, we discuss our future plans to more definitively test for a high substitution rate and drift between the two genomes.

Results and discussion

Genome stasis

We find the genomes of '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp to be very similar in gene content, with 841 protein coding genes (96% and 88% of coding genes, respectively) shared (Table 1). At 1.1 Mb the '*Ca*. Photodesmus blepharus' Ppalp genome is slightly larger than the 1 Mb '*Ca*. Photodesmus katoptron' genome, but is still significantly smaller than the genomes of free-living

relatives (Chapter III). This demonstrates that like '*Ca*. Photodesmus katoptron,' '*Ca*. Photodesmus blepharus' Ppalp is likely obligately dependent on the host for growth. Consistent with this, '*Ca*. Photodesmus blepharus' Ppalp is lacking most amino acid synthesis and energy metabolism genes, making growth outside the host unlikely. '*Candidatus* Photodesmus blepharus' Ppalp has 82 distinct genes not shared with '*Ca*. Photodesmus katoptron, while '*Ca.* Photodesmus katoptron' genome only has 35 distinct genes. In both cases, most unique genes are also found in other *Vibrionaceae* species and therefore likely represent genes found the common ancestor of the anomalopid symbionts and lost in one lineage or the other (Appendix 4). An exception to this is plasmid genes, which are mostly unique in each species and often lack *Vibrionaceae* homologs. The high number of genes shared by the two species suggests that most of the gene loss had already taken place in the common ancestor of the two symbionts and that the rate of gene loss slowed after divergence of the two species. However, it is possible that each species independently lost the same genes. A decreasing rate of gene loss over time is also found in other obligate symbionts and is consistent with high levels of genetic drift in early stages of genome reduction.

The genomes of the two species also show a high degree of similarity in gene order. We cannot determine absolute synteny between the species since the genomes are not fully assembled and so we do not know the order of contigs. Figure 1A shows the contigs of greater than 10,000 bp in length for each symbiont species arranged to maximize synteny. In this arrangement, five 'Ca. Photodesmus katoptron' contigs (33% of the genome) display complete synteny with 'Ca. Photodesmus blepharus' Ppalp contigs (Fig. 1A). Four more contigs (63.5 % of the genome) contain only seven

differences in gene order from ‘Ca. Photodesmus blepharus’ Ppalp, five inversions and two gene relocations. It is possible that the orthologous contigs shown in Fig. 1A are actually inverted from each other and that the genomes are less similar than they appear. It is also difficult to compare the level of synteny observed between the anomalopid symbionts to that seen in within a genus for other members of *Vibrionaceae* (Fig. 1B). Other members of *Vibrionaceae* typically have areas of high conservation of gene order on one chromosome but not the other (Reen et al., 2006). Since the anomalopid symbiont contigs represent genes from both chromosome I and chromosome II, it suggests that gene order between the two symbionts is more conserved than expected from the whole genome. While the possible high level of synteny between the two symbiont genomes is suggestive of genome stasis, as is seen with relatively old lineages of obligate symbionts, we can not confirm this hypothesis with these data.

Evidence for genetic drift

While the ‘*Ca*. Photodesmus katoptron’ and ‘*Ca*. Photodesmus blepharus’ Ppalp genomes are very similar in gene content, they are highly divergent at both the nucleotide and amino acid level. The average nucleotide identity (ANI) between the two symbionts is 74% and the average amino acid identity (AAI) is 67%. The ANI is well below the species designation cutoff of 95%, confirming the phylogenetic based designation of ‘*Ca*. Photodesmus katoptron’ and ‘*Ca*. Photodesmus blepharus’ Ppalp as separate species (Konstantinidis and Tiedje, 2005a). Values of AAI do not correlate well with taxonomy, but the value of 67% is within the range observed for highly divergent strains of *B. aphidicola* from different host species (Konstantinidis and Tiedje, 2005b). The high level

of divergence is consistent with the high rate of nucleotide change observed in anomalopid symbionts previously (Hendry and Dunlap, 2011) and indicative of high levels of drift.

We assume that genes that are only found in one symbiont species, but commonly found in other *Vibrionaceae*, have been lost by the other symbiont species. We also consider pseudogenes to have been 'lost' from a genome. To investigate the nature of gene loss, we determined what proportion of lost genes had been lost by multi-gene deletion versus pseudogenization in each species. Genes that are absent from one species but co-localized in the other species were designated as having been lost by multi-gene deletion. For both species, the majority of lost genes were lost as single genes (Table 2). These genes may have been lost as deletions of single genes or by pseudogenization and subsequent deletion. For '*Ca*. Photodesmus katoptron,' a higher proportion of genes were lost by deletion of multiple genes than were lost by pseudogenization. However, for '*Ca*. Photodesmus blepharus' Ppalp, the opposite trend was found. An increase in loss of genes through multi-gene deletions is predicted to result from high levels of genetic drift (Moran and Mira, 2001). It is possible that '*Ca*. Photodesmus katoptron' has been subject to higher levels of drift and has therefore lost more genes by deletion. This could account for '*Ca*. Photodesmus katoptron' having a smaller genome.

Ecological divergence

'*Candidatus* Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp are predicted to have very similar ecologies except that they are specific to different host species. To determine a possible genetic basis for host specificity, we investigated which

genes are unique to each species. We found no trends within the unique genes of each species, they appear to be representative of the types of genes found in both genomes (Appendix 4). We also looked within genes that are shared by both species for high levels of interspecies divergence, which might indicate different functions. Most shared genes are equally similar between the anomalopid symbionts as they are to relatives (Fig. 2). A small number are more divergent between the symbionts than they are from relatives and a slightly larger portion are more conserved between the anomalopid symbionts than expected by comparison to relatives. We expect that genes that are highly divergent between the anomalopid symbionts might represent host differences, and genes that are highly conserved within the anomalopid symbionts might be related to their unique symbiotic habitat. In general these two groups of genes are representative of the types of genes found in the symbiont genomes, with the exception of chemotaxis and motility genes. While chemotaxis and motility genes make up only 5% of the anomalopid symbiont genomes, they account for 15.5% and 11.8% of the conserved and divergent genes respectively. There are also unique chemotaxis and motility genes in the 'Ca. Photodesmus blepharus' Ppalp genome, but not at a higher than expected rate. The high number of conserved and divergent chemotaxis and motility genes is intriguing because these genes are likely to be used differently in anomalopid symbionts compared to relatives and to be related to host colonization.

Chemotaxis and motility genes that are conserved in anomalopid symbionts may differ in function from genes in other *Vibrionaceae*. Since anomalopid symbionts are unlikely to persist for long periods outside of the host, they likely use chemotaxis and motility genes only to find new hosts for colonization (Chapter III). In contrast, most

Vibrionaceae live in many diverse habitats and therefore likely need chemotaxis and motility functions to respond to a variety of chemical stimuli. Genes that are divergent between the anomalopid symbionts might relate to the host specific associations of the symbionts. Chemotaxis and motility genes may be used to detect specific host species. The luminous symbiont *Aliivibrio fischeri* has been shown to be attracted to chemical secretions of the developing host light organ (DeLoney-Marino et al., 2003), but no examples are known of luminous symbionts using chemical cues to distinguish between hosts.

Conclusions

Overall, these results suggest a similar pattern for the process of genome reduction in anomalopid symbionts as is seen on other obligate symbionts. Specifically, a history of high levels of genetic drift and possible genome stasis between species is observed. The similarities to relatively old obligate symbiont lineages such as *B. aphidicola*, which is 150 million years old (Moran and Mira, 2001), indicates that anomalopid symbionts have been obligate for most of their symbiosis with the host. These analyses also suggest the possibility that chemotaxis and motility genes are important for determining host range in anomalopid symbionts.

Future work

Further tests of genetic drift

Previous work and the data presented here have found that anomalopid symbionts are evolving faster than relatives and that the symbionts of different host species are

highly divergent (Chapter IV, Hendry and Dunlap, 2011). This trend is found widely in obligate symbionts (Clark et al., 2000; Moran, 1996; Wernegreen and Moran, 1999; Woolfit and Bromham, 2003). In order to quantify rates of divergence, we will perform relative rate tests on a variety of genetic loci found in both anomalopid symbionts and close relatives. We predict that most loci, regardless of function, will show higher than expected divergence between anomalopid symbionts and relatives and between anomalopid symbionts from different host species. If high divergence is found across the genome, it will support the hypothesis that anomalopid symbionts are experiencing high levels of drift, which leads to increased fixation of substitutions. We will use both Tajima's relative rate test (Tajima, 1993) and a tree-based maximum likelihood test (Woolfit and Bromham, 2003) of rate differences between taxa. We will include both coding and noncoding regions from throughout the anomalopid genomes. We will also choose loci from the categories used in this study of being conserved between anomalopid species, divergent, or neutral compared to relatives.

Another indication of high levels of genetic drift is the types of substitutions that become fixed. If substitutions are being fixed by chance and not being removed by purifying selection, we expect to see nearly equal numbers of nonsynonymous and synonymous substitutions. This pattern is common in other obligate symbionts (Abbot and Moran, 2002; Funk et al., 2001; Moran, 1996; Wernegreen and Moran, 1999). In previous work looking at intraspecific substitutions, we found that '*Ca*. Photodesmus blepharus' Ppalp, but not '*Ca*. Photodesmus katoptron,' had high levels of nonsynonymous substitutions (Chapter V). We will test for high rates of interspecific nonsynonymous substitutions using the same protein coding genes used for testing

relative rates. We will compare rates of nonynonymous to synonymous substitutions (d_N/d_S) both between the anomalopid symbionts and between each symbiont species and close relatives. We will also calculate d_N/d_S values between relative species, and these values will be used to determine significance of d_N/d_S values in anomalopid symbionts.

One factor that can skew d_N/d_S is codon usage bias, especially in taxa with AT biased genomes, such as anomalopid symbionts (Wernegreen and Moran, 1999). Preliminary analyses have not detected significant codon usage bias in anomalopid symbionts (Hendry, unpublished). To confirm this, we will test all coding genes used for codon usage bias using a χ^2-based test (Abbot and Moran, 2002). If codon usage bias is found, this will need to be accounted for in calculating d_N/d_S values. If higher than expected rates of nonsynonymous substitutions are found in multiple anomalopid symbiont loci, this will be strong evidence for high levels of genetic drift in anomalopid symbionts.

Testing for selection on chemotaxis and motility genes

We will use patterns of nonsynonymous substitutions in chemotaxis and motility genes to test for signals of selection in the anomalopid symbionts. Since anomalopid symbionts likely undergo repeated population bottlenecks and fix a high number of nonsynonymous substitutions due to drift, many commonly employed tests of selection or neutrality will not be accurate (Li, 2011). However, in the anomalopid symbionts chemotaxis and motility genes fall into three categories, conserved within anomalopid symbionts, divergent between anomalopid symbionts, and neutral with respect to relatives. We propose that these categories correspond to genes that have a unique

function in anomalopid symbionts, have different functions in different anomalopid symbionts, and have the same function as in relatives, respectively. Genes in these categories should show different substitution signatures, even with high levels of drift. Our predictions are as follows:

1) Conserved genes: A higher proportion of nonsynonymous substitutions should be shared between the two anomalopid symbionts than synonymous substitutions.
2) Divergent genes: A higher proportion of synonymous substitutions should be shared between the two anomalopid symbionts than nonsynonymous substitutions.
3) Neutral genes: Nonsynonymous and synonymous substitutions should be shared equally often between the two species. The proportion of shared substitutions in these genes can be used for comparison to determine significance of any differences found in genes from the categories above.

We will also calculate the proportion of shared nonsynonymous and synonymous substitutions in these genes between pairs of relatives. We expect that patterns should differ in relatives compared to anomalopid symbionts. If these predictions hold true, this indicates that chemotaxis and motility genes have been important in the ecological divergence of anomalopid symbionts.

Methods

Symbiont DNA was obtained from anomalopid flashlight fish collected in coastal waters in the Republic of Vanuatu in 2011 and DNA was extracted as in Hendry and Dunlap (2011). For the *P. palpebratus* symbiont, four specimens (Ppalp.1-Ppalp.4) were collected and DNA from one light organ of each specimen was combined for sequencing. For the A. katoptron symbiont, eight specimens (Akat.10-Akat.18) were combined. Very little polymorphism exists within the symbiont of a host species (Hendry and Dunlap, 2011; T.A. Hendry and P.V. Dunlap, unpublished data), so sequences obtained from the combined samples should not be significantly different than if they had come from an individual. Sequencing was done at the University of Michigan DNA Sequencing core on one lane of an Illumina HiSeq 2000 run. Illumina reads were assembled in Mira3 (Chevreux et al., 2002) by staff of the University of Michigan CCDU Bioinformatics Core. Glimmer 3.02 (Delcher et al, 2007) was used to predict orfs within both assemblies and orfs were annotated by BLAST comparison to the Swiss-Prot and UniRef 90 databases (December 2011 releases). Predicted orfs were thrown out if they were less than 100 amino acids in length and had less than 30% identity to database protein sequences. Predicted orfs were categorized as pseudogenes if they were above this threshold but contained premature stop codons (a stop codon in the first 66% of the gene) or frame shift mutations.

Due to the small length of Illumina reads, the genomes could be not fully assembled. Both assemblies consist of several greater than 10,000 bp contigs containing protein coding genes (nine contigs in ‘Ca. Photodesmus katoptron’ and 19 in ‘Ca. Photodesmus blepharus’ Ppalp) and smaller contigs containing tRNA and rRNA genes

and intergenic spacers. Additionally, in both assemblies complete circular plasmid sequences were recovered. The high read depth (minimum coverage = 70x, average coverage = 181x for 'Ca. Photodesmus katoptron' and 197x for 'Ca. Photodesmus blepharus' Ppalp') indicates that each genome is completely sequenced. Furthermore, this assembly of the 'Ca. Photodesmus katoptron' genome has nearly the same number of genes as the previously reported complete assembly (916 versus 915), the difference coming from different numbers of rRNA genes. Therefore, all comparisons of gene content and genome wide divergence are valid, though the genomes are not fully assembled. The genome assemblies of both species will be submitted to GenBank.

For comparisons of synteny, the Artemis Comparison Tool was used (Carver et al., 2005). Comparisons were based on predicted amino acid sequence using an e value cut off of 1.0. Calculations of average nucleotide identity (ANI) and average amino acid identity (AAI) used all genes shared between the two symbionts species with an identity cut off of 60% for ANI and 30% for AAI. For both calculations a 70% of query length match cut off was used (Konstantinidis and Tiedje, 2005a, b). For comparisons of intra symbiont amino acid identity to identity with relatives, the amino acid sequence of all shared genes was used. BLAST calculations of identity were weighted by the percent of the query match length. Reciprocal searches were performed and the results were averaged. The majority of shared genes had a 1:1 ratio of intra symbiont identity to identity with relatives. Genes were considered highly divergent or highly conserved if they differed from this ratio by more than one standard deviation.

Acknowledgements

This work will be coauthored with Jeffrey R. de Wet, Katherine E. Dougan, and Paul V. Dunlap for publication. We thank the staff of the University of Michigan Sequencing Core for DNA sequencing and advice and Jeffrey R. de Wet, of the University of Michigan Bioinformatics Analysis Core, for carrying out sequence assembly and annotation. We also thank Katherine Dougan for assistance with gene curation. Grant Norton of Sustainable Reef Suppliers, Brad Remmer of Sea Dwelling Creatures and Fish Doctors Aquarium provided assistance in obtaining specimens. Departmental Block Grant awards from the University of Michigan Department of Ecology and Evolutionary Biology funded this work.

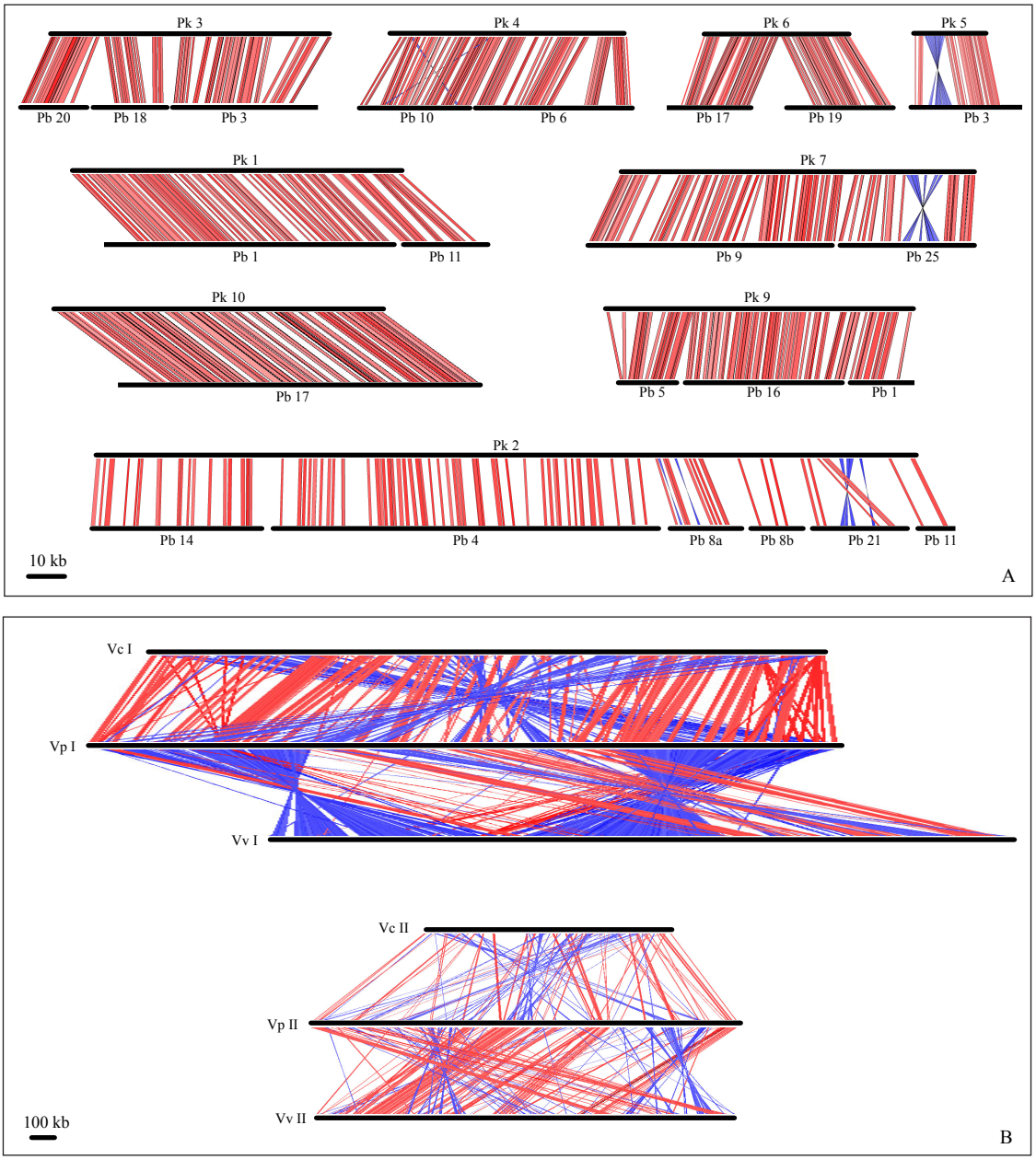

Figure 6.1. A. Alignments of syntenic contigs from 'Ca. Photodesmus katoptron' (Pk) and 'Ca. Photodesmus blepharus' Ppalp (Pp). Lines connecting contigs show BLAST amino acid similarities with an e value cut off of 1.0. Red lines are syntenic in the same direction and blue lines represent inversions. rRNA operons are not included. All contigs are shown to scale. B. Alignments of chromosome I and chromosome II from *Vibrio cholerae* N16961, *Vibrio parahaemolyticus* RIMD 2210633, and *Vibrio vulnificus* CMCP6. Lines connecting contigs show BLAST amino acid similarities with an e value cut off of 1.0. Red lines are syntenic in the same direction and blue lines represent inversions.

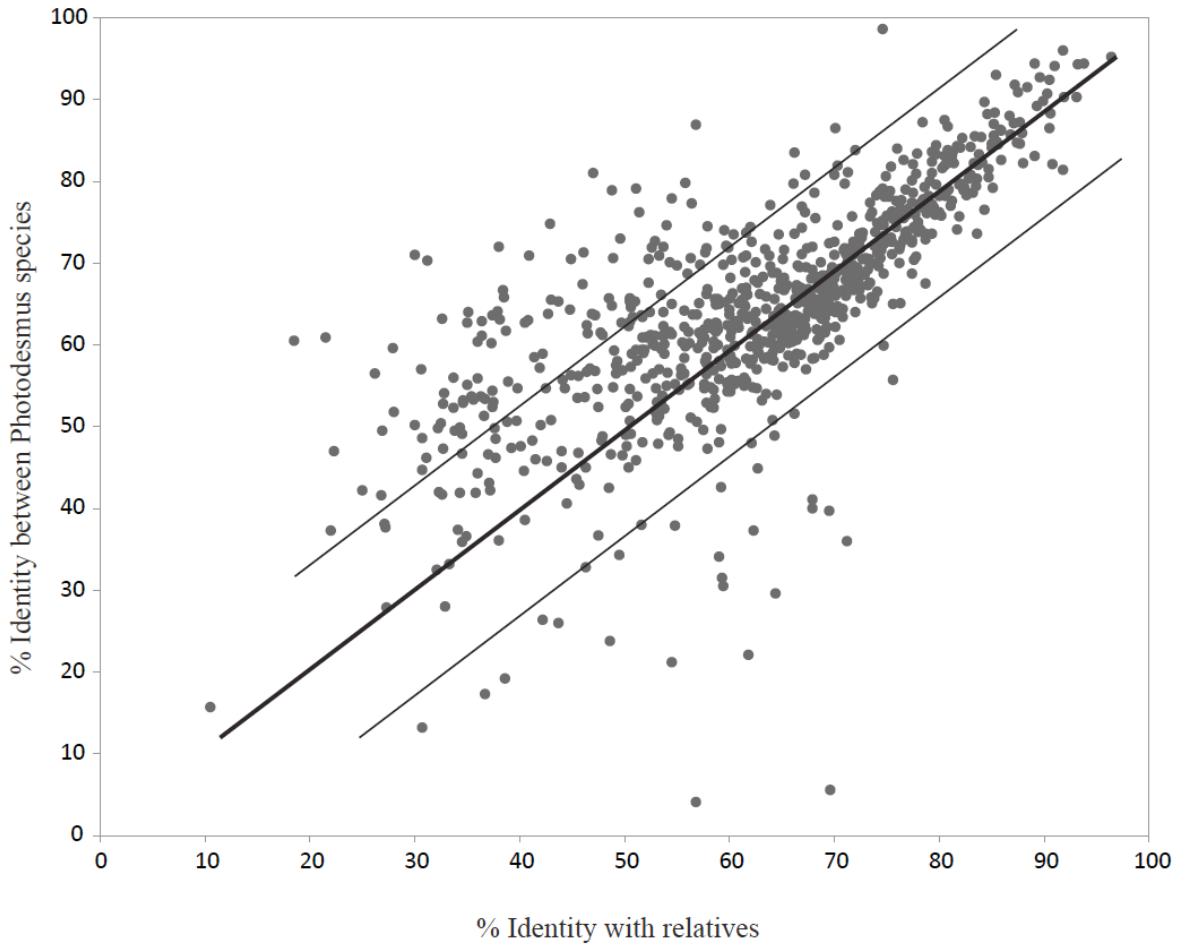

Figure 6.2. Genes shared by '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp arranged by percent identity between the anomalopid symbionts compared to percent identity to relatives. Percent identities based on BLAST alignments were scaled by the percent of the query that matched the subject. The between symbiont identities are the mean of reciprocal comparisons and the identities with relatives are the average of each symbiont's homolog compared to the best match in the Swiss-prot and Uniref90 databases. The thick line shows a 1:1 ratio of identities values, demonstrating that most genes fall along this line. The thin lines are 1 standard deviation away from the 1:1 ratio. Genes below and above the standard deviation lines were considered highly divergent and highly conserved, respectively, among the anomalopid symbionts.

Table 6.1. Genome content information for the '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus' Ppalp genomes.

	'*Ca*. P. katoptron'	'*Ca*. P. blepharus' Ppalp
Genome size	1 Mb	1.1 Mb
Total genes	914	968
Protein coding genes	877	923
rRNA genes	15	16
tRNA genes	23	29
Unique genes	35	82
Pseudogenes	10	20

Table 6.2. Comparison of how genes unique to one symbiont species were lost in the other species. Absolute numbers and percentages are shown.

	'*Ca*. P. katoptron'	'*Ca*. P. blepharus' Ppalp
Lost as pseudogenes	10 (10.9%)	20 (36.4%)
Lost in multi-gene deletion	28 (30.4%)	2 (3.6%)
Single lost genes	54 (58.7%)	33 (60%)

References

Abbot, P., Moran, N.A., 2002. Extremely low levels of genetic polymorphism in endosymbionts (*Buchnera*) of aphids (*Pemphigus*). Mol. Ecol. 11, 2649-2660.

Allen, J.M., Light, J.E., Perotti, M.A., Braig, H.R., Reed, D.L., 2009. Mutational Meltdown in Primary Endosymbionts: Selection Limits Muller's Ratchet. Plos One 4.

Burke, G.R., Moran, N.A., 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. Genome Biology and Evolution 3, 195-208.

Carroll, R., 1988. Vertebrate paleontology and evolution. W.H. Freeman, New York, NY.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., Parkhill, J., 2005. ACT: the Artemis comparison tool. Bioinformatics 21, 3422-3423.

Chevreux, B., Pfisterer, T., Suhai, S., 2002. Automatic assembly and editing of genomic sequences. In: Suhai, S. (Ed.), *Genomics and Proteomics - Functional and Computational Aspects.* Kluwer Academic Publishers, New York, NY, 51-65.

Clark, M.A., Moran, N.A., Baumann, P., Wernegreen, J.J., 2000. Cospeciation between bacterial endosymbionts (Buchnera) and a recent radiation of aphids (Uroleucon) and pitfalls of testing for phylogenetic congruence. Evolution 54, 517-525.

Delcher, A. L., Bratke, K. A., Powers, E. C., Salzberg, S. L., 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics 23, 673-679.

DeLoney-Marino, C.R., Wolfe, A.J., Visick, K.L., 2003. Chemoattraction of Vibrio fischeri to serine, nucleosides, and N-acetylneuraminic acid, a component of squid light-organ mucus. Appl. Environ. Microbiol. 69, 7527-7530.

Funk, D.J., Wernegreen, J.J., Moran, N.A., 2001. Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-buchnera association. Genetics 157, 477-489.

Hendry, T.A., Dunlap, P.V., 2011. The uncultured luminous symbiont of Anomalops katoptron (Beryciformes: Anomalopidae) represents a new bacterial genus. Mol. Phylogenet. Evol.

Konstantinidis, K.T., Tiedje, J.M., 2005a. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 2567-2572.

Konstantinidis, K.T., Tiedje, J.M., 2005b. Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187, 6258-6264.

Li, H.P., 2011. A new test for detecting recent positive selection that is free from the confounding impacts of demography. Mol. Biol. Evol. 28, 365-375.

McCutcheon, J.P., Moran, N.A., 2012. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10, 13-26.

Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873-2878.

Moran, N.A., McLaughlin, H.J., Sorek, R., 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 323, 379-382.

Moran, N.A., Mira, A., 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biology 2.

Reen, F.J., Almagro-Moreno, S., Ussery, D., Boyd, E.F., 2006. The genomic code: inferring *Vibrionaceae* niche specialization. Nature Reviews Microbiology 4, 697-704.

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599-607.

Tamas, I., Klasson, L., Canback, B., Naslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandstrom, J.P., Moran, N.A., Andersson, S.G.E., 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 296, 2376-2379.

Toh, H., Weiss, B.L., Perkin, S.A.H., Yamashita, A., Oshima, K., Hattori, M., Aksoy, S., 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. 16, 149-156.

Wernegreen, J.J., Moran, N.A., 1999. Evidence for genetic drift in endosymbionts (Buchnera): Analyses of protein-coding genes. Mol. Biol. Evol. 16, 83-97.

Woolfit, M., Bromham, L., 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. 20, 1545-1555.

CHAPTER VII

Conclusion

The goals of this dissertation were to determine if anomalopid symbionts are obligately dependent on hosts for growth and to explore the evolutionary patterns found in anomalopid symbionts, when compared to other symbionts with different lifestyles. I chose to ask these questions about the symbionts of anomalopid flashlight fish because they are unusual compared to other luminous symbionts; they cannot be cultured outside the host and are therefore thought to be obligately dependent on their hosts for growth. Through the course of this work I found that anomalopid symbionts are indeed obligately dependent on their hosts and that this interaction has had profound consequences on their evolution. I found that anomalopid symbionts have multiple characteristics in common with other obligate symbionts rather than related facultative symbionts. These findings highlight the importance of host interactions in determining the evolutionary history of symbionts.

The phylogenetic analyses in Chapter II confirmed that anomalopid symbionts are members of *Vibrionaceae* and are most closely related to the genus *Vibrio*. I proposed the genus name '*Candidatus* Photodesmus' for the divergent clade within *Vibrionaceae* formed by anomalopid symbionts. I also proposed the species name '*Candidatus* Photodesmus katoptron' for the symbionts of the host fish *Anomalops katoptron*. An

intriguing finding of this chapter was that anomalopid symbionts are evolving at a faster rate than relatives, just as obligate symbionts do. These analyses allowed me to perform phylogenetically informed comparisons throughout the rest of the dissertation.

Chapter III presents strong evidence that anomalopid symbionts are obligately dependent on their hosts for growth. The genome of '*Ca*. Photodesmus katoptron' is approximately one fifth the size of free-living relatives' genomes. The extensive gene loss in the anomalopid symbiont is particularly high in the categories of amino acid synthesis genes and energy metabolism genes. From the absence of genes I inferred that the anomalopid symbiont can synthesize at most three amino acids and can only use glucose as a carbon/energy source. This is strong support for '*Ca*. Photodesmus katoptron' being obligate dependent on the host since other habitats are unlikely to consistently supply both a full complement of amino acids and glucose. Gene content in other functional categories is also consistent with the symbiont living primarily in the light organ environment and is highly similar to the gene content of other obligate symbionts. I also found that the anomalopid symbiont has maintained chemotaxis and motility genes, consistent with their persisting for a period outside the host, between host generations. After finding this evidence that anomalopid symbionts are obligate, I sought to test for other patterns found in obligate symbionts.

In Chapter IV I expanded my analyses from one species of anomalopid symbiont to multiple, to determine if anomalopid symbiont species are specific to host species, as is found in other obligate symbionts. I found that anomalopid symbionts are specific to host genera, but that species of host within a genus share the same symbiont species. I hypothesized that the observed specificity has a genetic basis and may be caused by

recent codivergence. I also proposed the name '*Candidatus* Photodesmus blepharus' for symbionts of the fish genus *Photoblepharon*. Demonstrating that the symbionts of *A. katoptron* and *P. palpebratus* fish are different species provides a basis for comparisons that I perform in Chapters V and VI; I investigate genetic diversity in both symbiont species and divergence between the two species.

In Chapter V I assessed genetic diversity in two anomalopid symbiont species, in part to compare them to other obligate symbionts, which have very low intraspecific diversity, and in part to test for demographic patterns such as frequent population bottlenecks. I found that while '*Ca*. Photodesmus' species are genetically monomorphic, the causes of polymorphism patterns in anomalopid symbionts are potentially more complicated than hypothesized. Consistent with the symbionts undergoing population bottlenecks with transfers between host generations, anomalopid symbionts have very low levels of genetic diversity and a high number of rare substitutions. As predicted, high numbers of nonsynonymous substitutions are found in '*Ca*. Photodesmus blepharus' Ppalp, but this was not the case in '*Ca*. Photodesmus katoptron.' This finding suggests that genes in '*Ca*. Photodesmus katoptron' are under purifying selection rather than being subject to high levels of drift, possibly because this species has a more reduced genome. In Chapter V I also investigated the divergence of one anomalopid symbiont, '*Ca*. Photodesmus katoptron' across a wide geographic scale. I found very little divergence across the range, but the divergence that was found correlated with geographic distance. I propose that unlike in the aphid symbiont *B. aphidicola*, low genetic diversity across a wide geographic range in the anomalopid symbiont is not caused by wide host dispersal, but possibly by efficient DNA repair and low mutation rates, compared to *B. aphidicola*.

In Chapter VI I compared gene content and genetic divergence between the two anomalopid symbiont species. I found that the two species are very similar in gene content and contain mostly ancestral genes, suggesting that most of the genomic reduction in anomalopid symbionts took place before the divergence of '*Ca*. Photodesmus katoptron' and '*Ca*. Photodesmus blepharus.' Additionally, the two symbiont's genomes appear to be relatively static in gene order, though our ability to test this is limited, and highly divergent at both the nucleotide and amino acid level. All of these findings are consistent with our understanding of how genome reduction progressed in other obligate symbionts, but I plan to further test these results. Based on the comparison of the two genomes I hypothesized that chemotaxis and motility genes may have played an important role in the ecological divergence of the two species. I also outline our plans to test this hypothesis.

Open questions

Why are anomalopid symbionts obligate?

A major question that arises from this work is why anomalopid symbionts, and not other extracellular, environmentally acquired luminous symbionts, became obligately dependent on their hosts. The answer to this question appears straightforward for intracellular, vertically transmitted symbionts; the restriction to the host environment and guaranteed transfer to new hosts causes gene loss and consequently host dependence. In Chapter III I suggested that the anomalopid behavior of schooling in groups and the possibility of anomalopid larvae developing near adults allows for symbiont transmission that is similar to vertical transmission, though not necessarily direct from parent to

offspring. However, the symbiont of the squid *Euprymna scolopes*, *Aliivibrio fischeri*, becomes locally enriched in the area in which hatchling squid acquire symbionts (Lee and Ruby, 1994), yet *A. fischeri* has not become dependent on the host. Therefore, local transmission of symbionts does not seem to be sufficient to cause the evolution of obligate dependence. I proposed that high nutrient provisioning by 'Ca. Photodesmus' species might lead to obligate dependence. Our results indicate that anomalopids provide a variety of nutrients and a high energy carbon/energy source (glucose) to their symbionts, whereas *E. scolopes* provides less energy rich carbon/energy sources (chitin and glycerol) (Wier et al., 2010). It is possible that the anomalopid light organ provided higher fitness for the symbionts than other possible environments, whereas the relative benefit of the symbiotic environment for *A. fischeri* is lower. I suggest that it may be best to consider symbioses in terms of the relative benefits of the symbiotic habitat over other possible habitats, rather than simply as facultative versus obligate. Though these relative differences might be hard to quantify in nature, it is a worthwhile question to pursue.

Are anomalopids and their symbionts codiverging?

Another question that remains unanswered by this work is whether anomalopid symbionts have been codiverging with their hosts. In Chapter IV I presented evidence that anomalopids and their symbionts may not have been codiverging throughout their evolutionary history. I also presented the hypothesis that one species of obligate symbiont may have been able to colonize multiple species of hosts early in the evolutionary history of the symbiosis, and that specificity evolved after the divergence of new world and old world hosts. This hypothesis highlights the differences between anomalopid symbionts

and obligate intracellular symbionts. Because they are not physically tied to their hosts, the evolutionary divergence of anomalopid symbionts may not mirror that of the host. A future area of research on anomalopid symbiont would be to test for codivergence by constructing a molecular host phylogeny and including more species of both hosts and symbionts.

With a better understanding of how much host and symbiont phylogenies reflect each other, we can better quantify host-symbiont specificity at different points in their evolution and form hypotheses about how specificity developed. In the case of obligate intracellular symbionts, specificity is caused by vertical transmission and codivergence, rather than selection for specificity (Clark et al., 2000; Takiya et al., 2006). Extracellular symbionts like anomalopid symbionts could also become specific for neutral reasons, due to geographic isolation and drift. Mutualism theory predicts that there may be a benefit for specificity because it ensures that individuals acquire symbiotic partners that provide the highest benefits, and thus specificity might be selected for (Sachs et al., 2011). Discovering which genes, if any, lead to host specificity and determining if they have been under positive selection can test if specificity is caused by drift or selection in anomalopid symbionts. I propose a plan to do this in Chapter VI.

What causes genetic monomorphism?

This work adds anomalopid symbionts to the list of host associated bacteria that have dramatically low levels of genetic diversity (Abbot and Moran, 2002; Achtman, 2008; Funk et al., 2001). In many of these cases it is not know what causes this pattern. Of course there may be different specific causes in every case of genetic monomorphism.

In the case of *B. aphidicola*, symbiont population bottlenecks and wide host dispersal appear to lead to low intraspecific diversity (Abbot and Moran, 2002; Funk et al., 2001). For anomalopid symbionts these may not fully explain diversity patterns. On the whole, many lines of evidence presented in this dissertation support anomalopid symbionts being subject to high levels of drift and therefore likely population bottlenecks, as in *B. aphidicola*. However, the low rate of nonsynonymous substitutions on '*Ca*. Photodesmus katoptron' suggests that drift may not be strong compared to selection in this species and therefore drift may not be the cause of low polymorphism. Furthermore, widespread dispersal of the fish hosts seems an unlikely explanation for the wide geographic range of low diversity observed. A possible future direction of this work would be to test for host dispersal using population genetics, which I was not able to do here due to a lack of informative genetic markers. I suggest that low mutation rates in bacteria may be an important factor causing low genetic diversity. Our understanding of bacterial mutation rate may be biased by considering mostly culturable strains, and it is possible that many bacteria simply do not fix many mutations. Again, this highlights the importance using studies from bacteria in nature to inform our understanding of their evolution and ecology.

Significance

Since it has become feasible to examine unculturable host associated microbial communities, research has shown that symbiotic bacteria are critically important for animal health and evolution (Dethlefsen et al., 2007; Douglas, 2011; Robinson et al., 2010). How these bacteria evolve should then influence both host health and evolution,

but this is understudied. This dissertation highlights the considerable impact that host interactions can have on bacterial evolution. It also shows that we do not yet have enough understanding of how host interactions affect bacteria to predict the evolutionary outcomes of host interactions on bacteria. For instance, this work demonstrates that bacteria can undergo genome reduction and become obligate in the absence of an intracellular phase or vertical transmission. This suggests that as high-throughput sequencing technology continues to be used to study unculturable, host-associated bacteria, more examples like anomalopid symbionts will likely be found. Since the majority of host associated bacteria are not culturable, they can typically only be studied indirectly via genetic and genomic characteristics, as I have done here with anomalopid symbionts. For observed genomic trends in unknown species to have meaning, they must be compared to species with known ecologies for which similar data is available. Therefore, the more we discover about how host interactions affect the evolutionary history of bacteria, the more predictions we can make when exploring the vast numbers of largely unknown host associated bacteria. The symbionts of anomalopid flashlight fish make a good model for this work because, although they are somewhat difficult to acquire, they can be isolated in nearly pure culture in large numbers. We can then use them to identify how extracellular obligate symbionts may evolve in host interactions. Such findings have implications for understanding bacterial evolution and ecology, genome evolution, and the evolution of host associations themselves.

References

Abbot, P., Moran, N.A., 2002. Extremely low levels of genetic polymorphism in endosymbionts (Buchnera) of aphids (Pemphigus). Mol. Ecol. 11, 2649-2660.

Achtman, M., 2008. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu. Rev. Microbiol. 62, 53-70.

Clark, M.A., Moran, N.A., Baumann, P., Wernegreen, J.J., 2000. Cospeciation between bacterial endosymbionts (Buchnera) and a recent radiation of aphids (Uroleucon) and pitfalls of testing for phylogenetic congruence. Evolution 54, 517-525.

Dethlefsen, L., McFall-Ngai, M., Relman, D. A., 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature 449, 811-818.

Douglas, A. E., 2011. Lessons from studying insect symbioses. Cell Host Microbe 10, 359-367.Funk, D.J., Wernegreen, J.J., Moran, N.A., 2001. Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-buchnera association. Genetics 157, 477-489.

Lee, K.H., Ruby, E.G., 1994. Effect of the squid host on abundance and distribution of symbiotic *Vibrio fischeri* in nature. Appl. Environ. Microbiol. 60, 1565-1571.

Robinson, C. J., Bohannan, B. J. M., Young, V. B., 2010. From structure to function: the ecology of host-associated microbial communities. Microbiol. Mol. Biol. R. 74, 453.

Sachs, J.L., Essenberg, C.J., Turcotte, M.M., 2011. New paradigms for the evolution of beneficial infections. Trends in Ecology & Evolution 26, 202-209.

Takiya, D.M., Tran, P.L., Dietrich, C.H., Moran, N.A., 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta : Hemiptera : Cicadellidae) and their dual bacterial symbionts. Mol. Ecol. 15, 4175-4191.

Wier, A.M., Nyholm, S.V., Mandel, M.J., Massengo-Tiasse, R.P., Schaefer, A.L., Koroleva, I., Splinter-BonDurant, S., Brown, B., Manzella, L., Snir, E., Almabrazi, H., Scheetz, T.E., Bonaldo, M.D., Casavant, T.L., Soares, M.B., Cronan, J.E., Reed, J.L., Ruby, E.G., McFall-Ngai, M.J., 2010. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proc. Natl. Acad. Sci. U. S. A. 107, 2259-2264.

APPENDICES

Appendix 1

RNA polymerase gene	Present in P. katoptron?
DNA-directed RNA polymerase, alpha subunit (rpoA)	Y
DNA-directed RNA polymerase, beta subunit (rpoB)	Y
DNA-directed RNA polymerase, omega subunit (rpoZ)	Y
RNA polymerase sigma 70 (rpoD)	Y
RNA polymerase sigma 38 (rpoS)	Y
DNA-directed RNA polymerase, beta prime subunit (rpoC)	Y
RNA polymerase sigma 32 (rpoH)	N
RNA polymerase sigma 24 (rpoE)	N
RNA polymerase sigma 54 (rpoN)	N
RNA polymerase sigma 28 (sigD)	N
Total	6/10

Conserved marker gene	Present in P. katoptron?
COG0012 Predicted GTPase, probable translation factor	Y
COG0016 Phenylalanyl-tRNA synthetase alpha subunit	Y
COG0048 Ribosomal protein S12	Y
COG0049 Ribosomal protein S7	Y
COG0052 Ribosomal protein S2	Y
COG0080 Ribosomal protein L11	Y
COG0081 Ribosomal protein L1	Y
COG0085 DNA-directed RNA polymerase, beta subunit/140 kD subunit	Y
COG0087 Ribosomal protein L3	Y
COG0088 Ribosomal protein L4	Y
COG0090 Ribosomal protein L2	Y
COG0091 Ribosomal protein L22	Y
COG0092 Ribosomal protein S3	Y
COG0093 Ribosomal protein L14	Y
COG0094 Ribosomal protein L5	Y
COG0096 Ribosomal protein S8	Y
COG0097 Ribosomal protein L6P/L9E	Y
COG0098 Ribosomal protein S5	Y
COG0099 Ribosomal protein S13	Y
COG0100 Ribosomal protein S11	Y
COG0102 Ribosomal protein L13	Y
COG0103 Ribosomal protein S9	Y
COG0124 Histidyl-tRNA synthetase	Y
COG0185 Ribosomal protein S19	Y
COG0186 Ribosomal protein S17	Y
COG0197 Ribosomal protein L16/L10E	Y

Conserved marker gene	Present in P. katoptron?
COG0200 Ribosomal protein L15	Y
COG0201 Preprotein translocase subunit SecY	Y
COG0256 Ribosomal protein L18	Y
COG0495 Leucyl-tRNA synthetase	Y
COG0522 Ribosomal protein S4 and related proteins	Y
COG0525 Valyl-tRNA synthetase	Y
COG0533 Metal-dependent proteases with possible chaperone activity	Y
COG0541 Signal recognition particle GTPase	Y
COG0184 Ribosomal protein S15P/S13E	N
Total	34/35

ATP Synthase gene	Present in P. katoptron?
ATP synthase F1 epsilon subunit	Y
ATP synthase F1 beta subunit	Y
ATP synthase F1 gamma subunit	Y
ATP synthase, F1 alpha subunit	Y
ATP synthase F0 B subunit	Y
ATP synthase, F0 C subunit	Y
ATP synthase, F0 A subunit	Y
ATP synthase F1 delta subunit	N

Citric acid cycle gene	Present in P. katoptron?
2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)	Y
2-oxoglutarate dehydrogenase E2 component (EC 2.3.1.61)	Y
aconitase 2 (EC 4.2.1.3)	Y
citrate synthase (EC 2.3.3.1)	Y

Citric acid cycle gene	Present in P. katoptron?
Fumarase (EC:4.2.1.2)	Y
isocitrate dehydrogenase, NADP-dependent (EC:1.1.1.42)	Y
malate dehydrogenase (EC 1.1.1.37)	Y
succinate dehydrogenase subunit A flavoprotein (EC 1.3.99.1)	Y
succinate dehydrogenase subunit B, iron-sulfur unit (EC 1.3.99.1)	Y
succinate dehydrogenase subunit D membrane protein	Y
succinyl-CoA synthetase (ADP-forming) alpha subunit (EC 6.2.1.5)	Y
succinyl-CoA synthetase (ADP-forming) beta subunit (EC 6.2.1.5)	Y
aconitase 1 (EC 4.2.1.3)	N

Glycolysis gene	Present in P. katoptron?
6-phosphofructokinase (EC 2.7.1.11)	Y
enolase (EC 4.2.1.11)	Y
fructose-bisphosphate aldolase (EC 4.1.2.13)	Y
glucose-6-phosphate isomerase (EC 5.3.1.9)	Y
glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	Y
phosphoglucomutase, alpha-D-glucose phosphate-specific (EC:5.4.2.2)	Y
phosphoglycerate kinase (EC 2.7.2.3)	Y
phosphoglycerate mutase (EC 5.4.2.1)	Y
pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase (EC:2.3.1.12)	Y
pyruvate dehydrogenase E1 component, alpha subunit (EC:1.2.4.1)	Y
pyruvate kinase (EC 2.7.1.40)	Y
pyruvate dehydrogenase E1 component, beta subunit (EC:1.2.4.1)	N

Appendix 2

P. katoptron plasmid genes	**present on *Vibrio* plasmids?**	**present in *Vibrio* genomes?**
plasmid replication (REP) gene	Y	na
revolvase	Y	na
Methyl-accepting chemotaxis protein	N	Y
glutamate-1-semialdehyde-2,1-aminomutase	N	Y
7-cyano-7-deazaguanine reductase	N	Y
rarD protein	N	Y
Recombination-dependent growth factor C (rdgC)	N	Y
3-methyl-2-oxobutanoate hydroxymethyltransferase	N	N
Enhancing lycopene biosynthesis protein 2	N	N
uvrD helicase	N	N
band 7 protein	N	N
hypothetical protein	N	N
hypothetical protein	N	N
hypothetical protein	N	N

Amino acid synthesis gene name	(Fig. 3.5) role	*V. campbellii*	*A. fischeri*	P. katoptron
ATP phosphoribosyltransferase (homohexameric) (EC 2.4.2.17)	1	1	1	0
1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase (EC 5.3.1.16)	4	1	1	0
imidazole glycerol phosphate synthase subunit hisH (EC 2.4.2.-)	5	2	2	0
histidinol phosphate aminotransferase apoenzyme (EC 2.6.1.9)	7	1	1	0
histidinol dehydrogenase (EC 1.1.1.23)	9	2	2	0
glutamate synthase (EC 1.4.1.13)	10	3	3	0
glutamate 5-kinase (EC 2.7.2.11)	11	1	1	0
glutamate-5-semialdehyde dehydrogenase (EC 1.2.1.41)	12	1	1	0
pyrroline-5-carboxylate reductase (EC 1.5.1.2)	13	1	1	0
L-glutamine synthetase (EC 6.3.1.2)	14	1	1	1
glutaminase (EC 3.5.1.2)	15	1	1	0
carbamate kinase (EC 2.7.2.2)	16	1	1	0
ornithine carbamoyltransferase (EC 2.1.3.3)	17	1	1	0
argininosuccinate synthase (EC 6.3.4.5)	18	2	2	0
N-acetylglutamate synthase (EC 2.3.1.1) / argininosuccinate lyase (EC 4.3.2.1)	19	2	2	0
2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase (EC 2.3.1.117)	20	1	1	1
N-acetylglutamate kinase (EC 2.7.2.8)	21	1	1	0
N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38)	22	1	1	0
aspartate aminotransferase (EC 2.6.1.1)	23	1	1	1
acetylornithine deacetylase (EC 3.5.1.16)	24	1	1	0
aspartate semialdehyde dehydrogenase (EC 1.2.1.11)	26	2	2	1
dihydrodipicolinate synthase (EC 4.2.1.52)	27	1	1	1
dihydrodipicolinate reductase (EC 1.3.1.26)	28	1	1	1
acetylornithine/N-succinyldiaminopimelate aminotransferase (EC 2.6.1.17)	29	1	1	0
succinyldiaminopimelate desuccinylase (EC 3.5.1.18)	30	1	1	1
diaminopimelate epimerase (EC 5.1.1.7)	31	1	1	1
diaminopimelate decarboxylase (EC 4.1.1.20)	32	1	1	0

Amino acid synthesis gene name	role	V. campbellii	A. fischeri	P. katoptron
homoserine kinase (EC 2.7.1.39)	34	1	1	0
L-threonine synthase (EC 4.2.3.1)	35	1	1	0
homoserine O-succinyltransferase (EC 2.3.1.46)	36	1	1	0
O-succinylhomoserine (thiol)-lyase (EC:2.5.1.48)	37	1	1	0
cystathionine beta-lyase (EC 4.4.1.8)	38	1	1	0
methionine synthase (B12-independent) (EC 2.1.1.14)	39	2	2	0
D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	40	1	1	0
phosphoserine aminotransferase apoenzyme (EC 2.6.1.52)	41	1	1	1
phosphoserine phosphatase (EC 3.1.3.3)	42	1	1	0
glycine hydroxymethyltransferase (EC 2.1.2.1)	43	2	2	0
threonine aldolase (EC 4.1.2.5)	44	1	1	1
acetolactate synthase, large subunit (EC 2.2.1.6)	45	3	3	0
ketol-acid reductoisomerase (EC 1.1.1.86)	46	1	1	0
dihydroxyacid dehydratase (EC 4.2.1.9)	47	1	1	0
branched chain amino acid aminotransferase apoenzyme (EC 2.6.1.42)	48	1	1	0
L-threonine ammonia-lyase (EC 4.3.1.19)	49	1	1	0
2-isopropylmalate synthase (EC 2.3.3.13)	50	1	1	0
3-isopropylmalate dehydratase, small subunit (EC 4.2.1.33)	51	2	2	0
3-isopropylmalate dehydrogenase (EC 1.1.1.85)	52	1	1	0
serine O-acetyltransferase (EC 2.3.1.30)	53	1	1	0
cysteine synthase (EC 2.5.1.47)	54	1	1	0
3-deoxy-D-arabinoheptulosonate-7-phosphate synthase (EC 2.5.1.54)	55	3	3	1
3-dehydroquinate synthase (EC 4.2.3.4)	56	1	1	1
3-dehydroquinate dehydratase (EC 4.2.1.10)	57	1	1	1
shikimate dehydrogenase (EC 1.1.1.25)	58	1	1	1
shikimate kinase (EC 2.7.1.71)	59	1	1	1
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)	60	1	1	1
chorismate synthase (EC 4.2.3.5)	61	1	1	1
anthranilate synthase / indole-3-glycerol phosphate synthase / phosphoribosylanthranilate isomerase	62	2	3	0
(EC 4.1.3.27) (EC 5.3.1.24) (EC 4.1.1.48)				

Amino acid synthesis gene name	**role**	***V. campbellii***	***A. fischeri***	**P. katoptron**
anthranilate phosphoribosyltransferase (EC 2.4.2.18)	63	1	1	0
tryptophan synthase, beta chain (EC 4.2.1.20)	64	2	2	0
alanine dehydrogenase (EC 1.4.1.1)	68	1	1	0
phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) / phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31)	2, 3	1	1	0
aspartate kinase (EC 2.7.2.4) / homoserine dehydrogenase (EC 1.1.1.3)	25, 33	5	5	0
prephenate dehydrogenase (EC 1.3.1.12)(EC 4.2.1.51) / chorismate mutase (EC 5.4.99.5)	65, 66, 67	2	2	1
imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) / histidinol-phosphate phosphatase (EC 3.1.3.15)	6, 8	1	1	0
Total		84	85	18

Energy metabolism gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
ATP synthase F1 beta subunit (atpD)	2	1	1	ATP synthesis
ATP synthase F1 epsilon subunit (atpC)	2	1	1	ATP synthesis
ATP synthase F1 gamma subunit	2	1	1	ATP synthesis
ATP synthase, F0 B subunit	2	1	1	ATP synthesis
ATP synthase, F0 C subunit (atpE)	2	1	1	ATP synthesis
ATP synthase, F0 I subunit	1	1	1	ATP synthesis
ATP synthase, F0 subunit A	2	1	1	ATP synthesis
ATP synthase, F1 alpha subunit (atpA)	2	1	1	ATP synthesis
ATPase, flagellum specific (fliI)	3	1	1	ATP synthesis
4Fe-4S binding domain	1	1	1	electron transport
c-type cytochrome biogenesis protein CcmE	1	1	1	electron transport
c-type cytochrome biogenesis protein CcmF	1	1	1	electron transport
c-type cytochrome biogenesis protein CcmH	1	2	1	electron transport
c-type cytochrome biogenesis protein CcmI	1	2	1	electron transport
Cytochrome b subunit of the bc complex	1	1	1	electron transport
cytochrome c oxidase, cbb3-type, subunit I (EC:1.9.3.1)	2	2	1	electron transport
cytochrome c oxidase, cbb3-type, subunit II (EC:1.9.3.1)	2	2	1	electron transport
cytochrome c oxidase, cbb3-type, subunit III (EC:1.9.3.1)	2	2	1	electron transport

Energy metabolism gene name	*V. campbellii*	*A. fischeri*	P. katoptron	role
cytochrome c1	1	1	1	electron transport
cytochrome c4	1	1	1	electron transport
cytochrome c5	1	1	1	electron transport
cytochrome oxidase maturation protein, cbb3-type	1	1	1	electron transport
electron transport complex, RnfABCDGE type, A subunit	1	1	1	electron transport
electron transport complex, RnfABCDGE type, B subunit	1	1	1	electron transport
electron transport complex, RnfABCDGE type, C subunit	1	1	1	electron transport
electron transport complex, RnfABCDGE type, D subunit	1	1	1	electron transport
electron transport complex, RnfABCDGE type, E subunit	1	0	1	electron transport
electron transport complex, RnfABCDGE type, G subunit	1	1	1	electron transport
flavodoxin	2	2	1	electron transport
glutaredoxin-related protein, monothiol glutaredoxin	1	1	1	electron transport
NAD(P) transhydrogenase, alpha subunit (EC:1.6.1.2)	1	1	1	electron transport
succinate dehydrogenase subunit C cytochrome b556	1	1	1	electron transport
Thiol:disulfide interchange protein (dsbD)	1	2	1	electron transport
thioredoxin	1	1	1	electron transport
thioredoxin (trxB)	2	2	1	electron transport
ubiquinol-cytochrome c reductase, iron-sulfur subunit (EC:1.10.2.2) (petA)	1	1	1	electron transport
UDP-galactose 4-epimerase (EC 5.1.3.2) (galE)	2	2	1	galactose metabolism
6-phosphofructokinase (EC 2.7.1.11)	1	1	1	glycolysis
carbon storage regulator (csrA)	1	1	1	glycolysis
enolase (EC 4.2.1.11) (eno)	2	1	1	glycolysis
fructose-bisphosphate aldolase (EC 4.1.2.13) (fbaA)	2	2	1	glycolysis
glucose-6-phosphate isomerase	1	1	1	glycolysis
glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	3	3	1	glycolysis
phosphoglucomutase, alpha-D-glucose phosphate-specific (EC:5.4.2.2)	1	1	1	glycolysis
phosphoglycerate kinase (EC 2.7.2.3)	1	1	1	glycolysis
phosphoglycerate mutase (EC 5.4.2.1)	1	1	1	glycolysis
pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase (EC:2.3.1.12)	1	1	1	glycolysis
pyruvate dehydrogenase E1 component, alpha subunit (EC:1.2.4.1)	1	1	1	glycolysis

Energy metabolism gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
pyruvate kinase (EC 2.7.1.40)	2	2	1	glycolysis
triosephosphate isomerase (EC 5.3.1.1)	1	1	1	glycolysis
arabinose 5-phosphate isomerase (EC:5.3.1.13)	1	2	1	pentose phosphate pathway
ribose-5-phosphate isomerase (EC 5.3.1.6) (rpiA)	1	1	1	pentose phosphate pathway
ribulose-5-phosphate 3-epimerase (EC 5.1.3.1) (rpe)	2	2	1	pentose phosphate pathway
transaldolase (EC 2.2.1.2) (tal)	1	1	1	pentose phosphate pathway
transketolase (EC 2.2.1.1) (tkt)	2	2	1	pentose phosphate pathway
L-serine ammonia-lyase (EC 4.3.1.17)	3	2	1	serine metabolism
2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (sucA)	1	1	1	TCA
2-oxoglutarate dehydrogenase E2 component (EC 2.3.1.61)	1	1	1	TCA
aconitase 2 (EC 4.2.1.3)	1	0	1	TCA
citrate synthase (EC 2.3.3.1) (gltA)	1	1	1	TCA
Fumarase (EC:4.2.1.2)	1	0	1	TCA
isocitrate dehydrogenase, NADP-dependent (EC:1.1.1.42)	1	1	1	TCA
malate dehydrogenase (EC 1.1.1.37)	1	1	1	TCA
succinate dehydrogenase subunit A flavoprotein (sdhA)	2	2	1	TCA
succinate dehydrogenase subunit B, iron-sulfur unit	2	2	1	TCA
succinate dehydrogenase subunit D membrane protein	1	1	1	TCA
succinyl-CoA synthetase (ADP-forming) alpha subunit (EC 6.2.1.5) (sucD)	1	1	1	TCA
succinyl-CoA synthetase (ADP-forming) beta subunit (EC 6.2.1.5) (sucC)	1	1	1	TCA
acetylornithine aminotransferase apoenzyme (EC 2.6.1.11)	1	1	0	amino acid metabolism
arginine succinyltransferase (EC 2.3.1.109)	1	1	0	amino acid metabolism
asparaginase (EC 3.5.1.1)	1	1	0	amino acid metabolism
aspartate ammonia-lyase (EC:4.3.1.1)	1	1	0	amino acid metabolism
D-serine ammonia-lyase (EC 4.3.1.18)	1	0	0	amino acid metabolism
formiminoglutamase (EC 3.5.3.8)	1	1	0	amino acid metabolism

Energy metabolism gene name	*V. campbellii*	*A. fischeri*	P. katoptron	role
glycine cleavage system H protein	1	1	0	amino acid metabolism
glycine dehydrogenase (decarboxylating) alpha/beta subunit	1	1	0	amino acid metabolism
histidine ammonia-lyase (EC 4.3.1.3)	1	1	0	amino acid metabolism
imidazolonepropionase (EC 3.5.2.7)	1	1	0	amino acid metabolism
L-alanine dehydrogenase (EC 1.4.1.1)	1	1	0	amino acid metabolism
L-glutaminase (EC 3.5.1.2)	1	1	0	amino acid metabolism
L-threonine 3-dehydrogenase (EC 1.1.1.103)	1	1	0	amino acid metabolism
maleylacetoacetate isomerase (EC 5.2.1.2)	1	0	0	amino acid metabolism
methylmalonate-semialdehyde dehydrogenase [acylating] (EC 1.2.1.27)	1	0	0	amino acid metabolism
Phenylalanine 4-hydroxylase (EC 1.14.16.1)	1	1	0	amino acid metabolism
tryptophanase (EC 4.1.99.1)	1	0	0	amino acid metabolism
urocanate hydratase (EC 4.2.1.49)	1	1	0	amino acid metabolism
2-amino-3-ketobutyrate coenzyme A ligase (EC 2.3.1.29)	1	1	0	amino acid metabolism
3-hydroxyisobutyrate dehydrogenase (EC 1.1.1.31)	1	0	0	amino acid metabolism
4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)	1	0	0	amino acid metabolism
aspartate racemase (EC:5.1.1.13)	0	2	0	amino acid metabolism
carbamate kinase (EC 2.7.2.2)	0	1	0	amino acid metabolism
glycine cleavage system T protein (EC:2.1.2.10)	1	1	0	amino acid metabolism
succinylarginine dihydrolase (EC 3.5.3.23)	0	1	0	amino acid metabolism
arginine deiminase (EC 3.5.3.6)	1	1	0	amino acid metabolism
delta-1-pyrroline-5-carboxylate dehydrogenase (EC 1.5.1.12) / L-proline dehydrogenase (EC 1.5.99.8)	1	1	0	amino acid metabolism
succinylglutamate desuccinylase (EC 3.5.1.-)	1	1	0	amino acid metabolism
succinylglutamic semialdehyde dehydrogenase (EC 1.2.1.-)	1	1	0	amino acid metabolism
Uncharacterized protein required for formate dehydrogenase activity	1	1	0	anaerobic metabolism
formate dehydrogenase gamma subunit (EC 1.2.1.2)	1	2	0	anaerobic metabolism
pyruvate formate-lyase 1-activating enzyme (EC:1.97.1.4)	1	1	0	anaerobic metabolism
ATP synthase F1, delta subunit	2	1	0	ATP synthesis
L-lysine 2,3-aminomutase (EC 5.4.3.2)	1	1	0	electron transport
NADPH-glutathione reductase (EC 1.8.1.7)	1	1	0	electron transport

Energy metabolism gene name	*V. campbellii*	*A. fischeri*	P. katoptron	role
azurin	1	0	0	electron transport
cyd operon protein YbgE	1	1	0	electron transport
cyd operon protein YbgT	1	1	0	electron transport
periplasmic nitrate reductase maturation protein NapF	1	1	0	electron transport
periplasmic nitrate reductase subunit NapA apoprotein	1	1	0	electron transport
trimethylamine N-oxide reductase system, TorE protein	1	1	0	electron transport
cytochrome bd quinol oxidase subunit 2 apoprotein (EC 1.10.3.-)	1	1	0	electron transport
cytochrome bo3 quinol oxidase subunit 1 apoprotein (EC 1.10.3.-)	1	0	0	electron transport
cytochrome bo3 quinol oxidase subunit 2 (EC 1.10.3.-)	1	0	0	electron transport
cytochrome bo3 quinol oxidase subunit 3 (EC 1.10.3.-)	1	0	0	electron transport
cytochrome bo3 quinol oxidase subunit 4 (EC 1.10.3.-)	1	0	0	electron transport
cytochrome c oxidase, subunit I (EC:1.9.3.1)	1	1	0	electron transport
cytochrome c oxidase, subunit II (EC:1.9.3.1)	1	1	0	electron transport
flavocytochrome c	1	0	0	electron transport
Glutaredoxin, Thiolredoxin fold, GrxB	1	1	0	electron transport
respiratory nitrite reductase (cytochrome; ammonia-forming) precursor (EC 1.7.2.2) NrfA	1	1	0	electron transport
respiratory nitrite reductase specific cytochrome c biogenesis protein NrfE	1	1	0	electron transport
respiratory nitrite reductase specific cytochrome c biogenesis protein NrfF	1	0	0	electron transport
respiratory nitrite reductase specific menaquinol--cytochrome-c reductase complex subunit NrfB	1	1	0	electron transport
respiratory nitrite reductase specific menaquinol--cytochrome-c reductase complex subunit NrfC	1	1	0	electron transport
respiratory nitrite reductase specific menaquinol--cytochrome-c reductase complex subunit NrfD	1	1	0	electron transport
iron-sulfur cluster binding protein, putative	1	1	0	electron transport
NAD(P)H:quinone oxidoreductase, type IV	1	1	0	electron transport
acetate kinase (EC 2.7.2.1)	1	2	0	fermentation
formate acetyltransferase 1 (EC:2.3.1.54)	2	2	0	fermentation
butyrate kinase (EC 2.7.2.7)	1	0	0	fermentation
galactokinase (EC 2.7.1.6)	1	1	0	galactose metabolism

Energy metabolism gene name	*V. campbellii*	*A. fischeri*	P. katoptron	role
UDP-glucose-hexose-1-phosphate uridylyltransferase (EC 2.7.7.12) / UTP-hexose-1-phosphate uridylyltransferase (EC 2.7.7.10)	1	2	0	galactose metabolism
glycerol kinase (EC 2.7.1.30)	1	1	0	glycerol metabolism
phosphoglycolate phosphatase	1	1	0	glycolate metabolism
erythrose 4-phosphate dehydrogenase (EC 1.2.1.72)	1	1	0	glycolysis
phosphoenolpyruvate carboxykinase (ATP) (EC:4.1.1.49)	1	1	0	glycolysis
phosphoenolpyruvate synthase (EC 2.7.9.2)	1	0	0	glycolysis
sugar fermentation stimulation protein	1	1	0	maltose metabolism
mannose-6-phosphate isomerase, type 1 (EC 5.3.1.8)	2	1	0	mannose metabolism
2-keto-3-deoxy-phosphogluconate aldolase (EC 4.1.2.14)	1	2	0	pentose phosphate pathway
6-phosphogluconate dehydratase (EC 4.2.1.12)	1	1	0	pentose phosphate pathway
6-phosphogluconate dehydrogenase (decarboxylating) (EC 1.1.1.44)	1	1	0	pentose phosphate pathway
6-phosphogluconolactonase (EC 3.1.1.31)	1	0	0	pentose phosphate pathway
deoxyribose-phosphate aldolase (EC:4.1.2.4)	1	1	0	pentose phosphate pathway
glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49)	1	1	0	pentose phosphate pathway
2-deoxy-D-gluconate 3-dehydrogenase	0	1	0	polysaccharide metabolism
glycogen synthase (ADP-glucose)	1	1	0	polysaccharide metabolism
glucose-1-phosphate adenylyltransferase (EC:2.7.7.27)	2	1	0	polysaccharide metabolism
4-alpha-glucanotransferase (EC:2.4.1.25)	1	1	0	polysaccharide metabolism
alpha-1,4-glucan:alpha-1,4-glucan 6-glycosyltransferase (EC:2.4.1.18)	1	0	0	polysaccharide metabolism
alpha-1,6-glucosidases, pullulanase-type	1	0	0	polysaccharide metabolism

Energy metabolism gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
glycogen debranching enzyme	1	3	0	polysaccharide metabolism
glycogen/starch/alpha-glucan phosphorylases (EC:2.4.1.1)	1	1	0	polysaccharide metabolism
ribokinase (EC:2.7.1.15)	1	1	0	sugar metabolism
aconitase 1 (EC 4.2.1.3)	1	0	0	TCA
Citrate lyase synthetase (EC:6.2.1.22)	1	1	0	TCA
isocitrate lyase (EC 4.1.3.1)	1	1	0	TCA
malate synthase A (EC:2.3.3.9)	1	1	0	TCA
glycerate kinase (EC:2.7.1.31)	1	1	0	unknown
diaminopropionate ammonia-lyase (EC:4.3.1.15)	1	0	0	unknown
Total	185	163	68	

Transport and binding gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
PTS system IIB component/PTS system IIC component	9	5	2	sugar transport
sodium/alanine symporter agcS	1	1	1	amino acid uptake
arginine/ornithine symporter	0	0	1	amino acid uptake
branched-chain amino acid uptake carrier	3	3	1	amino acid uptake
Multiple antibiotic resistance protein marC	4	2	1	antibiotic transporter
Multidrug resistance protein norM	1	1	1	antibiotic transporter
putative efflux protein, MATE family	9	7	1	antibiotic transporter
protein translocase subunit secA	1	1	1	cell division
Cell division and transport-associated protein TolA	1	2	1	cell division/colicin export
Cell division and transport-associated protein TolQ	1	1	1	cell division/colicin export
Cell division and transport-associated protein TolR	1	1	1	cell division/colicin export
Cell division and transport-associated protein TolB	1	1	1	cell division/colicin export
ABC transporter, ATP-binding protein CydD	1	1	1	cytochrome synthesis
ABC transporter, permease CydC	1	1	1	cytochrome synthesis
heme exporter protein CcmA	1	1	1	cytochrome synthesis
heme exporter protein CcmB	1	1	1	cytochrome synthesis

Transport and binding gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
heme exporter protein CcmC	1	1	1	cytochrome synthesis
copper-(or silver)-translocating P-type ATPase	3	2	1	cytochrome synthesis
RND family efflux transporter, MFP subunit	11	12	1	efflux pump
Cation efflux protein	1	1	1	heavy metal efflux
magnesium and cobalt efflux protein CorC	1	1	1	heavy metal efflux
iron uptake regulator	1	1	1	iron uptake
ferric iron ABC transporter, ATPase	1	1	1	iron uptake
ferric iron ABC transporter, transmembrane	1	1	1	iron uptake
ferric iron ABC transporter, permease	1	1	1	iron uptake
Lipopolysaccharide ABC transporter, ATP-binding protein LptB	1	1	1	Lipid A biosynthesis
Lipopolysaccharide ABC transporter, LptC	1	1	1	Lipid A biosynthesis
Lipopolysaccharide ABC transporter, periplasmic protein LptB	1	1	1	Lipid A biosynthesis
lipid A export permease MsbA	1	1	1	Lipid A biosynthesis
capsular exopolysaccharide family	2	1	1	lipopolysaccharide transport
Mg transporter MgtE	2	1	1	magnesium uptake
ABC-type transport system, multidrug efflux pump, ATPase component	2	2	1	organic toxins
choline/carnitine/betaine transport	4	4	1	osmolyte uptake
mechanosensitive ion channel mscS	1	1	1	osmotic balance
sodium or potassium/proton antiporter, NhaB	1	1	1	pH regulation
sodium/proton antiporter, NhaD	1	1	1	pH regulation
sodium or potassium/proton antiporter, NhaC	1	1	1	pH regulation
phosphate transporter	1	1	1	phosphate transport
Kef-type potassium/proton antiporter, CPA2 family	2	1	1	potassium efflux
potassium ion channel protein	1	1	1	potassium uptake
NADH:ubiquinone oxidoreductase, Na(+)-translocating, A subunit	1	1	1	sodium translocation
NADH:ubiquinone oxidoreductase, Na(+)-translocating, B subunit	2	1	1	sodium translocation
NADH:ubiquinone oxidoreductase, Na(+)-translocating, C subunit	1	1	1	sodium translocation
NADH:ubiquinone oxidoreductase, Na(+)-translocating, D subunit	1	1	1	sodium translocation
NADH:ubiquinone oxidoreductase, Na(+)-translocating, E subunit	1	0	1	sodium translocation

Transport and binding gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
NADH:ubiquinone oxidoreductase, Na(+)-translocating, F subunit	1	1	1	sodium translocation
sodium/calcium antiporter	1	1	1	sodium/calcium exchange
PTS system IIA component	5	3	1	sugar transport
Cation/multidrug efflux pump	2	9	1	toxin efflux
ABC-type multidrug transport system, ATPase and permease components	1	1	1	unknown
hypothetical ttg2B	1	1	1	unknown
ABC zinc transporter znuA	1	1	1	zinc transport
ABC zinc transporter znuB	1	1	1	zinc transport
ABC zinc transporter znuC	1	1	1	zinc transport
Putative threonine efflux protein	3	1	0	amino acid efflux
amino acid ABC transporter membrane protein, PAAT family	2	1	0	amino acid uptake
amino acid ABC transporter substrate-binding protein, PAAT family	3	2	0	amino acid uptake
amino acid ABC transporter membrane protein 1, PAAT family	4	3	0	amino acid uptake
amino acid ABC transporter membrane protein 2, PAAT family	4	3	0	amino acid uptake
Amino acid permeases, specific	5	3	0	amino acid uptake
amino acid/peptide transporter (Peptide:H+ symporter), bacterial	1	6	0	amino acid uptake
cadaverine:lysine antiporter	1	1	0	amino acid uptake
Carboxylate/Amino Acid/Amine Transporter	1	1	0	amino acid uptake
Predicted branched-chain amino acid permease (azaleucine resistance)	1	0	0	amino acid uptake
putrescine:ornithine antiporter, APA family	1	2	0	amino acid uptake
sodium--glutamate symport carrier (gltS)	2	2	0	amino acid uptake
ammonium transporter	1	0	0	ammonium transport
anion transporter	2	2	0	anion transporter
drug resistance transporter, EmrB/QacA subfamily	6	5	0	antibiotic transporter
arsenical-resistance protein	1	1	0	arsenic efflux
Outer membrane cobalamin receptor protein	1	1	0	b12 transport
chromate transporter, chromate ion transporter (CHR) family	1	1	0	chromate transport
The Hydrophobe/Amphiphile Efflux-1 (HAE1) Family	3	4	0	efflux pump
formate/nitrite transporter	2	2	0	formate transport
gluconate transporter	2	0	0	gluconate transport

Transport and binding gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
glycerol-3-phosphate transporter	1	2	0	glycerol transport
heavy metal efflux pump (cobalt-zinc-cadmium)	1	1	0	heavy metal efflux
2,3-dihydroxybenzoate-AMP ligase	1	0	0	iron transport
bacterioferritin	1	0	0	iron uptake
outer membrane transport energization protein TonB	3	3	0	iron uptake
TonB-dependent siderophore receptor	3	4	0	iron uptake
L-lactate permease	1	0	0	lactate uptake
Lysine efflux permease	1	0	0	lysine exporter
cation diffusion facilitator family transporter	3	1	0	metal efflux
NRAMP metal ion transporters	1	0	0	metal transport
molybdate ABC transporter, permease protein	1	1	0	molybdenum uptake
molybdenum ABC transporter, ATP-binding protein	1	1	0	molybdenum uptake
molybdenum ABC transporter, periplasmic molybdate-binding protein	1	1	0	molybdenum uptake
nicotinamide mononucleotide transporter PnuC	1	1	0	nicotinamide transport
nitrate transport ATP-binding subunits C and D	1	0	0	nitrate transport
Nucleoside permease	3	5	0	nucleoside transport
Purine-cytosine permease and related proteins	1	1	0	nucleoside transport
thiamine ABC transporter, ATP-binding protein	1	1	0	nucleoside transport
thiamine ABC transporter, periplasmic binding protein	1	1	0	nucleoside transport
thiamine ABC transporter, permease protein	1	1	0	nucleoside transport
choline ABC transporter, ATP-binding protein	1	0	0	osmolyte uptake
choline ABC transporter, periplasmic binding protein	1	0	0	osmolyte uptake
glycine betaine/L-proline transport ATP binding subunit	2	1	0	osmolyte uptake
Predicted solute sodium symporter	1	1	0	osmotic balance
oligopeptide/dipeptide ABC transporter, ATP-binding protein	5	7	0	peptide transport
sodium or potassium/proton antiporter, NhaA	1	1	0	pH regulation
phosphate ABC transporter ATP-binding protein, PhoT family	2	2	0	phosphate transport
phosphate ABC transporter membrane protein 1, PhoT family	1	1	0	phosphate transport
phosphate ABC transporter membrane protein 2, PhoT family	1	1	0	phosphate transport
phosphate ABC transporter substrate-binding protein, PhoT family	2	3	0	phosphate transport

Transport and binding gene name	*V. campbellii*	*A. fischeri*	P. katoptron	role
phosphate ABC transporter, permease protein PstA	1	1	0	phosphate transport
potassium uptake protein, TrkH family	3	3	0	potassium uptake
sodium/proline symporter	2	2	0	proline symport
oxaloacetate decarboxylase alpha subunit	1	1	0	sodium transport
sodium ion-translocating decarboxylase, beta subunit	1	1	0	sodium transport
spermidine/putrescine ABC transporter ATP-binding subunit	1	1	0	spermidine/putrescine uptake
diguanylate phosphodiesterase	3	2	0	sugar transport
PTS system glucosamine dimer-specific IIB component, Lac family	1	1	0	sugar transport
PTS system glucosamine dimer-specific IIC component, Lac family	1	1	0	sugar transport
PTS system N-acetylglucosamine-specific IIB component, IIC component	1	2	0	sugar transport
Sulfate permease and related transporters (MFS superfamily)	1	1	0	sulfate transport
putative 2-aminoethylphosphonate ABC transport system, 1-aminoethylphosphonate-	1	1	0	unknown
binding protein component				
putative 2-aminoethylphosphonate ABC transport system, ATP-binding protein	1	1	0	unknown
component (EC:3.6.3.30)				
putative 2-aminoethylphosphonate ABC transport system, permease protein	1	1	0	unknown
anaerobic c4-dicarboxylate membrane transporter family protein	2	4	0	unknown
MIP family channel proteins	2	2	0	unknown
Predicted permease	2	2	0	unknown
Predicted symporter	1	1	0	unknown
rarD protein	2	2	0	unknown
TRAP transporter solute receptor, TAXI family	3	2	0	unknown
TRAP transporter, 4TM/12TM fusion protein	2	2	0	unknown
TRAP transporter, DctM subunit	3	0	0	unknown
uracil-xanthine permease	2	2	0	uracil uptake
Total	229	210	55	

Cell envelope gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
prepilin-type N-terminal cleavage/methylation domain	14	17	2	pilus
o-antigen polymerase	5	5	2	surface polysaccharide
ADP-heptose:LPS heptosyltransferase	0	1	2	surface polysaccharide
Glycosyltransferase	2	2	2	surface polysaccharide
putative lipoprotein	2	2	2	unknown
Periplasmic protein SypC involved in polysaccharide export	1	1	1	biofilm formation
2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase	1	1	1	murein and peptioglycan
alanine racemase	1	1	1	murein and peptioglycan
Cell division protein FtsI/penicillin-binding protein 2	1	1	1	murein and peptioglycan
cell elongation-specific peptidoglycan biosynthesis regulator RodA	2	2	1	murein and peptioglycan
D-alanine--D-alanine ligase	1	1	1	murein and peptioglycan
D-alanyl-D-alanine carboxypeptidase, serine-type, PBP4 family	1	1	1	murein and peptioglycan
Diaminopimelate epimerase	1	1	1	murein and peptioglycan
Dihydrodipicolinate reductase	1	1	1	murein and peptioglycan
dihydrodipicolinate synthase	1	1	1	murein and peptioglycan
glutamate racemase	1	1	1	murein and peptioglycan
lipopolysaccharide heptosyltransferase II	1	1	1	murein and peptioglycan
lytic murein transglycosylase	1	1	1	murein and peptioglycan
N-acetylmuramoyl-L-alanine amidase	1	1	1	murein and peptioglycan
penicillin-binding protein, 1A family	1	1	1	murein and peptioglycan
phosphoglucosamine mutase	1	1	1	murein and peptioglycan
Phospho-N-acetylmuramoyl-pentapeptide-transferase	1	1	1	murein and peptioglycan
rod shape-determining protein MreB	1	1	1	murein and peptioglycan
rod shape-determining protein MreC	1	1	1	murein and peptioglycan
rod shape-determining protein MreD	1	1	1	murein and peptioglycan
S-adenosyl-methyltransferase MraW	1	1	1	murein and peptioglycan
succinyl-diaminopimelate desuccinylase	1	1	1	murein and peptioglycan
UDP-N-acetylglucosamine 1-carboxyvinyltransferase	1	1	1	murein and peptioglycan
UDP-N-acetylglucosamine pyrophosphorylase	1	1	1	murein and peptioglycan

Cell envelope gene name	*V. campbellii*	*A. fischeri*	P. katoptron	role
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-	1	1	1	murein and peptioglycan
undecaprenol N-acetylglucosamine transferase				
UDP-N-acetylmuramate dehydrogenase	1	1	1	murein and peptioglycan
UDP-N-acetylmuramate--L-alanine ligase	1	1	1	murein and peptioglycan
UDP-N-acetylmuramoylalanine--D-glutamate ligase	1	1	1	murein and peptioglycan
UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	1	1	1	murein and peptioglycan
UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	1	1	1	murein and peptioglycan
Undecaprenyl-diphosphatase	1	1	1	murein and peptioglycan
penicillin-binding protein 1B	1	1	1	murein and peptioglycan
outer membrane associated protein	1	1	1	outer membrane maintenance
3-deoxy-8-phosphooctulonate synthase	1	1	1	surface polysaccharide
3-deoxy-D-manno-octulosonate cytidylyltransferase	1	1	1	surface polysaccharide
3-deoxy-D-manno-octulosonic acid kinase	1	1	1	surface polysaccharide
acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	1	1	1	surface polysaccharide
ADP-glyceromanno-heptose 6-epimerase precursor	1	1	1	surface polysaccharide
D-alpha,beta-D-heptose 1,7-bisphosphate phosphatase	1	1	1	surface polysaccharide
D-beta-D-heptose 1-phosphate adenylyltransferase	1	1	1	surface polysaccharide
DnaA-interacting protein DiaA	1	1	1	surface polysaccharide
Glycosyl hydrolase family 92.	1	1	1	surface polysaccharide
3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	0	1	1	surface polysaccharide
lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase	1	1	1	surface polysaccharide
lipid A biosynthesis lauroyl (or palmitoleoyl) acyltransferase	3	2	1	surface polysaccharide
lipid A export permease/ATP-binding protein MsbA	1	1	1	surface polysaccharide
lipid-A-disaccharide kinase	1	1	1	surface polysaccharide
lipid-A-disaccharide synthase	1	1	1	surface polysaccharide
lipopolysaccharide transport periplasmic protein LptA	1	1	1	surface polysaccharide
monofunctional biosynthetic peptidoglycan transglycosylase	1	1	1	surface polysaccharide
peptidoglycan glycosyltransferase	1	1	1	surface polysaccharide

Cell envelope gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
peptidoglycan glycosyltransferase	1	1	1	surface polysaccharide
phosphoheptose isomerase	1	1	1	surface polysaccharide
UDP-2,3-diacylglucosamine hydrolase	1	1	1	surface polysaccharide
UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	1	1	1	surface polysaccharide
UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	1	1	1	surface polysaccharide
UDP-glucose pyrophosphorylase	1	1	1	surface polysaccharide
Polysaccharide biosynthesis chain length regulator SypO	1	1	1	surface polysaccharide
Undecaprenyl pyrophosphate synthetase	1	1	1	surface polysaccharide
undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphatetransferase	1	1	1	surface polysaccharide
lipopolysaccharide biosynthesis glycosyltransferase	0	0	1	surface polysaccharide
Glycosyltransferase WavL	0	0	1	surface polysaccharide
Membrane proteins related to metalloendopeptidases, lipoprotein NlpD	1	1	1	surface polysaccharide
rare lipoprotein A	2	1	1	unknown
glutamine--fructose-6-phosphate transaminase	1	1	0	murein and peptioglycan
N-acetylmuramic acid 6-phosphate etherase	1	1	0	murein and peptioglycan
UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	1	1	0	murein and peptioglycan
pilus (MSHA type) biogenesis protein MshL	1	1	0	pilus
pilus retraction protein PilT	2	2	0	pilus
alpha-1,2-mannosidase, putative	6	0	0	surface polysaccharide
mannose-1-phosphate guanylyltransferase	1	0	0	surface polysaccharide
UDP-N-acetylglucosamine 2-epimerase	2	2	0	surface polysaccharide
lipoprotein, YaeC family	1	1	0	unknown
outer membrane lipoprotein, Slp family	1	1	0	unknown
uncharacterized proteobacterial lipoprotein	1	1	0	unknown
Total	106	102	74	

Cellular processes gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
TsaA	1	1	1	antioxidant
Superoxide dismutase	2	1	1	antioxidant
Thiol peroxidase, Bcp-type	2	2	1	antioxidant
cell division protein FtsA	1	1	1	cell division
cell division protein FtsL	1	1	1	cell division
cell division protein FtsZ	1	1	1	cell division
cell division protein ZipA	1	1	1	cell division
cell division-specific peptidoglycan biosynthesis regulator FtsW	1	1	1	cell division
cell elongation-specific peptidoglycan biosynthesis regulator RodA	2	2	1	cell division
FtsK	1	1	1	cell division
FtsQ	1	1	1	cell division
sporulation related protein	1	1	1	cell division
ZapA	1	1	1	cell division
cell cycle protein	1	1	1	cell division
GTP-binding Era	1	1	1	cell division
membrane protease FtsH catalytic subunit	1	1	1	cell division
septum site-determining protein MinC	1	1	1	cell division
CheA	1	1	1	chemotaxis
CheR	1	1	1	chemotaxis
CheW	1	1	1	chemotaxis
CheY	1	1	1	chemotaxis
CheZ	1	1	1	chemotaxis
PomA	1	1	1	chemotaxis
cold-shock DNA-binding protein, CspD	1	1	1	cold shock
cold shock protein cspG	1	0	1	cold shock
Toluene tolerance protein Ttg2	1	1	1	detoxification
FlgI	1	1	1	flagellar function
FlaG	1	1	1	flagellar function
flagella basal body P-ring formation protein FlgA	2	1	1	flagellar function

Cellular processes gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
flagellar basal-body M-ring protein/flagellar hook-basal body protein (fliF)	2	1	1	flagellar function
flagellar basal-body P-ring protein	1	1	1	flagellar function
flagellar basal-body rod protein FlgB	2	1	1	flagellar function
flagellar basal-body rod protein FlgC	2	1	1	flagellar function
flagellar basal-body rod protein FlgF, proteobacterial	2	1	1	flagellar function
flagellar basal-body rod protein FlgG, Gram-negative bacteria	2	1	1	flagellar function
flagellar biosynthesis protein FlhA	2	1	1	flagellar function
flagellar biosynthetic protein FlhB	2	1	1	flagellar function
flagellar biosynthetic protein FliP	2	1	1	flagellar function
flagellar biosynthetic protein FliQ	2	1	1	flagellar function
flagellar biosynthetic protein FliR	2	1	1	flagellar function
flagellar biosynthetic protein FliS	2	1	1	flagellar function
flagellar hook capping protein	1	1	1	flagellar function
flagellar hook-associated protein 3	2	1	1	flagellar function
flagellar hook-associated protein FlgK	2	1	1	flagellar function
flagellar hook-basal body complex protein (FliE)	2	1	1	flagellar function
flagellar motor switch protein FliG	1	1	1	flagellar function
flagellar motor switch protein FliM	1	1	1	flagellar function
flagellar motor switch protein FliN	2	1	1	flagellar function
flagellar protein export ATPase FliI (EC:3.6.3.14)	2	1	1	flagellar function
flagellin	1	1	1	flagellar function
FlgE	1	1	1	flagellar function
FlgH	1	1	1	flagellar function
FlgJ	1	1	1	flagellar function
FlhF	1	1	1	flagellar function
FlhG	1	1	1	flagellar function
FliH	1	1	1	flagellar function
FliJ	1	1	1	flagellar function
flagellar-specific RNA polymerase sigma factor FliA	0	0	1	flagellar function

Cellular processes gene name	*V. campbellii*	*A. fischeri*	P. katoptron	role
FliO	1	1	1	flagellar function
flagellar hook-length control protein, FliK	1	1	1	flagellar function
flagellar hook-associated protein 2	1	1	1	flagellar function
type-a flagellin	1	1	1	flagellar function
MotX	1	1	1	flagellar function
MotY	1	1	1	flagellar function
RNA polymerase, sigma 32 subunit, RpoH	1	1	1	heat shock
RNA polymerase, sigma 38 subunit, RpoS	1	1	1	heat shock
organic solvent tolerance, OstA	1	1	1	organic solvent tolerance
(p)ppGpp synthetase, RelA/SpoT family	2	2	1	stress response
bis(5'-nucleosyl)-tetraphosphatase (symmetrical)	1	1	1	stress response
stringent starvation protein A	1	1	1	stress response
stringent starvation protein B	1	1	1	stress response
Predicted transcriptional regulator, BolA	1	1	1	stress response
colicin production protein	1	1	1	toxin production
S-(hydroxymethyl)glutathione dehydrogenase/class III alcohol dehydrogenase	1	2	0	alcohol tolerance
(EC:1.1.1.1,EC:1.1.1.284)				
alkyl hydroperoxide reductase, F subunit	1	1	0	antioxidant
catalase/peroxidase HPI	1	1	0	antioxidant
Predicted redox protein, regulator of disulfide bond formation	1	1	0	antioxidant
septum site-determining protein MinD	1	1	0	cell division
cell division ATP-binding protein FtsE	1	1	0	cell division
cell division protein FtsN	1	1	0	cell division
cell division protein FtsX	1	1	0	cell division
cell division topological specificity factor MinE	1	1	0	cell division
intracellular septation protein A	1	1	0	cell division
MAF protein	2	2	0	cell division
arsenate reductase	1	1	0	detoxification

Cellular processes gene name	***V. campbellii***	***A. fischeri***	**P. katoptron**	**role**
arsenical-resistance protein	1	1	0	detoxification
conserved hypothetical protein TIGR02293	1	0	0	detoxification
drug resistance transporter, EmrB/QacA subfamily	1	2	0	detoxification
heavy metal efflux pump (cobalt-zinc-cadmium)	1	1	0	detoxification
cysteine protease domain, YopT-type	1	0	0	protease, toxin
GTP-binding protein TypA/BipA	1	1	0	stress response
methionine-R-sulfoxide reductase (EC:1.8.4.12)	4	4	0	stress response
phage shock protein A	1	1	0	stress response
phage shock protein B	1	1	0	stress response
phage shock protein C	1	1	0	stress response
phage shock protein G	1	1	0	stress response
prevent-host-death family protein	3	1	0	stress response
redox-sensitive transcriptional activator SoxR	1	0	0	stress response
competence protein ComEA helix-hairpin-helix repeat region	1	2	0	transformation
DNA protecting protein DprA	1	1	0	transformation
Total	126	105	73	

Appendix 3

Species	Gene	GenBank accession numbers 'Local'	'Global'
Aliivibrio fischeri	*recA*	EU907948-EU907950, EU907952, JF509771, JF509772 JF509758-JF509762, JF50976, JF509764-JF509767, HQ595314, HQ595313, HQ595307	EU907962-EU907964, EU907947, EU907945, EU907956, EU907958, EU907960, EU907942, EU907944, JF509781, HQ595310, HQ595306 HQ595308
	luxA	FJ803908, FJ803906, FJ803904, FJ803902, FJ803900, FJ803898, FJ803896, FJ803894, FJ803892, FJ803890, FJ803888, FJ803886, FJ803884, FJ803882, FJ803870, FJ803868, FJ803866, FJ803864, FJ803862, FJ803860, FJ803858, FJ803856, FJ803854, FJ803852, FJ803850, FJ803848, FJ803846, FJ803844, FJ803842, FJ803840, FJ803838, FJ803836	DQ648325-DQ648328, JF509903, JF509905, JF509899, EU185976, JF509936, DQ026810, EU185974, EU185977, DQ026815, JF509942, DQ26816, JF509938, JF509901, JF509910, JF509907, JF509904, JF509909
Photobacterium kishitanii	*gyrB*	AY849449, AY849461, AY849464, AY849450, AY642172-AY642181, AY849457-AY849462, DQ648293-DQ648302, AY849453-AY849455,	DQ648287-DQ648292, AY849463, AY849452, AY849481, AY849472-AY849477, AY849451, DQ648303-DQ648305, DQ648307, DQ648308, AY849434-AY849436, EF415508-EF415510, EF415524, EF415512-EF415519, EF415527, EF415525, AY455876, AY642171, DQ648306
	recA	DQ648391, DQ648382-DQ648387, DQ648400, DQ648390, DQ648401	DQ648376-DQ648381, EF415562, EF415575, EF415562-EF415567, EF415568, EF415577, EF415555-EF415559, EF415574, DQ648389, DQ648388, EF415552, EF415569, DQ648402, DQ648392-DQ648398, DQ648399, DQ648403, EF415553

Species	Gene	GenBank accession numbers 'Local'	'Global'
Photobacterium kishitanii	*luxA*	DQ648337-DQ648339, AY849508, AY642227 AY849514-AY849516, AY642198-AY642201 DQ648342, DQ648350, AY642193, AY642183 AY642184, AY642215, AY642224, AY642226, AY642504-AY642506, AY642509-AY642513	DQ648331-DQ648336, AY849483, DQ648340, DQ648341, DQ6483351, DQ6483352, AY849484, DQ648343-DQ648349, AY849482, AY849528, DQ988874, AY849523-AY849525, EF415626 EF415613-EF415615, EF415618-EF415620, EF415628-EF415631, AY849500-AY849503, EF415625, AY849485, AY341064, AY642182, EF441349
Photobacterium leiognathi	*gyrB*		DQ648315-DQ648323, AB298193, AF136385, DQ652566, DQ371341-DQ371348, AY455882, AY455879, EU118211, AY455880, EF372600, DQ499006, DQ648309-DQ648312
	luxA		DQ371368-DQ371375, DQ648362-DQ648368, DQ648353-DQ648360, DQ652567, AY456750 M63594, AY341069, EF536338, AY341070, EF372601
Photobacterium mandapamensis	*gyrB*	EU118202, DQ790881, DQ790883, DQ371359, DQ371345, EF372600, AY455884, AY455887 AY455886, DQ371366, DQ371367	DQ371361-DQ371365, DQ371350- DQ371355, AY455883, DQ371341
	luxA	DQ790859, DQ790852-DQ790858 , DQ790850, DQ371388-DQ371392, DQ371377-DQ371381, DQ790849	DQ371387, DQ371386, EU200993, DQ790860 EU200991, EU200989, AY456752, AY456751 EF372602, DQ371394, DQ988878, EU200994 DQ790861, DQ790863-DQ790865, DQ371393 EU200997, EU200999

Species	Gene	GenBank accession numbers	
		'Local'	'Global'
Vibrio cholerae	*gyrB*	HM009714-HM009718, HM009655, HM009706, HM009690, HM009680, HM009672-HM009674, HM009668, HM009664-HM009666, HM009662, HM009652, HM009646, HM009629, HM009622, HM009618, HM009608, HM009606, HM009602, HM009581-HM009586, HM009601, HM009572, HM009721, HM009693, HM009685, HM009661, HM009649, HM009643, HM009641, HM009639, HM009635-HM009637, HM009621, HM009609	HM042663, AF501897, DQ316971-DQ316976, DQ021082, FJ970860, JF939044, HM009720, HM009694, AB298207, DQ021080, DQ021095, DQ021093, DQ021088-DQ021091, DQ021087, DQ021083, DQ021085, DQ021081, FJ970861, EU101412, EU101414
	recA	HM010380, HM010327, HM010350, HM010342, HM010343, HM010308, HM010259-HM010365, HM010373-HM010375, HM010337-HM010339, HM010333, HM010331, HM010329, HM010325, HM010321, HM010287, HM010358, HM010235, HM010384, HM010283, HM010367-HM010371, HM010352, HM010356, HM010297, HM010346, HM010314-HM010316, HM010379, HM010377, HM010307	HM003867, FJ645930, GQ214728, FJ645929, GU299673, AJ842391, EU118222, AF301103, AF301108, HM003855, HM003853, HM003859, HM003861, HM003869, HM003856-HM003858 HM003846, FJ479702-FJ479705, FJ224374, HM042659, HM042655, HM042656, HM042658, GU299670, GU299666-GU299668, GU299664, AJ842389, AJ842387, U10162, AF540664, AF540661, AF540663, EU118228, EF990327 EF990307, EF990304, EF643503, DQ513162, EF643499-EF643501, EF643504, DQ513171, DQ513169, DQ513167, DQ513165, DQ513158, DQ513156
Vibrio harveyi	*gyrB*	EF596227, EF596225, EF596229, EF596231, EF596193, EF596214, EF596224, EF596187, EF596200-EF596205, EF596195-EF596199, EF596208-EF596210, EF596175, EF596168	EU130501, HM224411, EF596212, EF596194, EF596233, EF596188, EF596169, EF596221, EF596178-EF596184, EF596172, EF596170, EF596217, EF596211, EF596189-EF596192, DQ499007, EF596215, EF596219, EF596207, EF596213

Species	Gene	GenBank accession numbers	
		'Local'	'Global'
Vibrio harveyi	*recA*	EF596440, EF596429, EF596437, EF596438, EF596410-EF596420, EF596441, EF596402 EF596423- EF596425, EF596390, EF596383	EF596434, EF596432, EF596430, EF596428, EF596426, EF596422, EF596406-EF596407 EF596393-EF596396, EF596384, EF596449, EF596405, EF596399, EF596398, EF596387, AJ842437, EF596397, EF596427, EF596443
	luxA	EU201022, EU201015, EU201029, EU201030, EU201033-EU201035, EU201023, EU201021, EU201003, EU201010-EU201014, EU201025, EU201005-EU201007, EU201017-EU201019	

Appendix 4

Unique *'Ca.* Photodesmus katoptron' genes	role	present in *Vibrio*
glutamine synthetase	amino acid synthesis	Y
MadN protein	amino acid transport	Y
ParA	cell division	Y
3-oxoacyl-[acyl-carrier-protein] synt	fatty acid synthesis	Y
4-amino-4-deoxychorismate lyase	folic acid synthesis	Y
ribonuclease G	mRNA modificaiton	Y
UDP-N-acetylenolpyruvoylglucosamine	murein synthesis	Y
Adenine phosphoribosyltransferase	nucleotide synthesis	Y
oligoendopeptidase F	protein fate	Y
Arginine deiminase 2	protein modification	Y
Ribosome biogenesis GTPase A	ribosomal protein	Y
YrbG ion exchange	transport and binding	Y
Copper-translocating P-type ATPase	transport and binding	Y
kef-type potassium channel	transport and binding	Y
MarC	transport and binding	Y
tRNA-dihydrouridine synthase	tRNA modification	Y
S-adenosylmethionine:tRNA ribosyltransferase-isomerase	tRNA modification	Y
hypothetical protein	unknown	Y
dTDP-glucose 4,6-dehydratase	unknown	Y
Flavin reductase	unknown	Y
membrane insertion efficiency	unknown	Y
hypothetical protein	unknown	Y

Unique 'Ca. Photodesmus katoptron' genes	role	present in Vibrio
Para-aminobenzoate synthase	unknown	Y
DNA polymerase III	unknown	Y
hypothetical protein	unknown	Y
Para-aminobenzoate synthase	unknown	Y
Flavin reductase	unknown	Y
CAAX amino terminal protease	unknown	Y
STAS anti sigma factor b	unknown	Y
RarD (plasmid)	unknown	Y
RdgC (plasmid)	unknown	Y
resolvase (plasmid)	unknown	Y
f-box DNA helicase (plasmid)	unknown	N
nuclease (plasmid)	unknown	N

Unique *'Ca.* Photodesmus blepharus' Ppalp genes	role	present in *Vibrio*
2-amino-3-ketobutyrate coenzyme A ligase	amino acid synthesis	Y
L-threonine 3-dehydrogenase	amino acid synthesis	Y
Pyrroline-5-carboxylate reductase	amino acid synthesis	Y
glutamate decarboxylase	amino acid transport	Y
MTA/SAH nucleosidase	amino acid transport	Y
microcin immunity protein MccF	antibiotic immunity	Y
bolA	cell division	Y
Chromosome partition protein Smc	cell division	Y
chromosome partitioning	cell division	Y
chromosome partitioning	cell division	Y
ParA	cell division	Y
ParA like	cell division	Y
ParA like	cell division	Y
galactosyl-transferase	cell wall	Y
lipid A biosynthesis	cell wall	Y
O-antigen export system ATP-binding protein RfbB	cell wall	N
O-antigen export system permease protein RfbA	cell wall	N
dsbB	disulfide bonds	Y
RecQ	DNA helicase	Y
acetyltransferase	energy	Y
ADP-ribose pyrophosphatase	energy	Y
Bifunctional protein hldE	energy	Y
Glucose-1-phosphate adenylyltransferase	energy metabolism	Y
Maltodextrin phosphorylase	energy metabolism	Y
UDP-N-acetylglucosamine 2-epimerase	LPS synthsis	Y
UDP-N-acetylglucosamine 4,6-dehydratase	LPS synthsis	Y

Unique 'Ca. Photodesmus blepharus' Ppalp genes	role	present in Vibrio
Tyrosine recombinase XerC	moblie element insertion	Y
Chemotaxis protein methyltransferase	motility	Y
chemotaxis-specific methylesterase	motility	Y
flagellar biosynthesis sigma factor	motility	Y
flagellar regulatory protein FleQ	motility	Y
fleS flagellar sensor histidine kinase	motility	Y
Negative regulator of flagellin synthesis	motility	Y
phage-related tail component	phage	Y
aminopeptidase B	protein fate	Y
GacS	regulatory	Y
hypothetical protein	regulatory	Y
kinase	regulatory	Y
luxS	regulatory	Y
NAD-dependent deacetylase	regulatory	Y
tetR family	regulatory	Y
dTDP-4-dehydrorhamnose 3,5-epimerase	rhamnose pathway	Y
dTDP-4-dehydrorhamnose reductase	rhamnose pathway	Y
50S ribosomal protein L3	ribosomal protein	Y
Ribosome biogenesis GTPase	ribosome synthesis	Y
hfq	small rnas	Y
glycogen synthase	starch synthesis	Y
hsp	stress tolerance	Y
Universal stress protein E	stress tolerance	Y
Peptide chain release factor 3	translation	Y
1,4-alpha-glucan branching enzyme GlgB	transport	Y
potassium transport	transport	Y

Unique 'Ca. Photodesmus blepharus' Ppalp genes	role	present in Vibrio
SapB	transport	Y
SapC	transport	Y
SapD	transport	Y
SapF	transport	Y
Serine/threonine transporter sstT	transport	Y
Thiamin transport system permease protein ThiP	transport	Y
Thiamine import ATP-binding protein ThiQ	transport	Y
Thiamine-binding periplasmic protein	transport	Y
beta-ketoacyl synthase	unknown	Y
Diacylglycerol kinase	unknown	Y
DNA transformation protein TfoX	unknown	Y
DNA-binding protein HU-alpha	unknown	Y
DNA-binding protein HU-beta	unknown	Y
glycosyltransferase family protein	unknown	N
haloacid dehalogenase	unknown	N
hypothetical protein	unknown	N
hypothetical protein	unknown	N
hypothetical protein	unknown	N
hypothetical protein	unknown	Y
hypothetical protein	unknown	Y
hypothetical protein	unknown	Y
hypothetical protein	unknown	Y
hypothetical protein	unknown	Y
hypothetical protein	unknown	Y
hypothetical protein	unknown	Y
hypothetical protein	unknown	Y

Unique 'Ca. Photodesmus blepharus' Ppalp genes	role	present in Vibrio
hypothetical protein	unknown	Y
kinase	unknown	N
Phosphoenolpyruvate carboxykinase	unknown	Y
SyrB	unknown	Y