EVOLUTION OF DEVELOPMENTAL GENE REGULATION IN THE MYXOBACTERIA

I-Chen Chen

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements for the degree
Doctor of Philosophy
in the Department of Biology,
Indiana University
July 2015

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Gregory J. Velicer, Ph.D.

Patricia L. Foster, Ph.D.

Jeffrey D. Palmer, Ph.D.

Rudolf A. Raff, Ph.D.

Michael J. Wade, Ph.D.

Yuen-Tsu Nicco Yu, Ph.D.

June 10th, 2015

ACKNOWLEDGEMENTS

First of all, I would like to thank my two advisors Greg Velicer and Nicco Yu for their support and relentless encouragement during every stage in my PhD. I will be forever grateful for their guidance and the synergy they provided. In addition, I would like to thank Nicco for helping me with my transitions to Bloomington and Zürich and turning me into a cat person.

I would also like to thank my committee members, Jeff Palmer and Mike Wade, for their guidance, encouragement and critique on my research through my PhD, and Pat Foster and Rudy Raff for kindly joining the committee later on. I am indebted to many graduate students, postdocs and faculty in the Biology Department at Indiana University, and in the Theoretical Biology, Experimental Ecology, Molecular Microbial Ecology groups at ETH Zürich for the supportive and stimulating research environments they provided.

It was a lot of fun working on the myxobacteria in the Velicer lab. I would like to thank my awesome labmates: Lena Mendes-Soares, Susanne Kraemer, Peter Zee, Pauline Manhes, Hansrainer Peitz in Bloomington, and Olaya Rendueles, Sébastien Wielgoss, Francesca Fiegna, Jessica Plucain, Ramith Nair, Kaitlin Schaal and Samay Pande in Zürich. Special thanks go to our lab manager Heike Keller for her steady support over the years.

I was lucky to have some wonderful undergraduate students working with me during my time in Bloomington. In particular, I would like to thank Brad Griesenauer, whose hard work made a significant contribution to the phylogenetic tree in my first chapter.

Of course, graduate school wouldn't be complete without many cool friends. I didn't know anyone when moving to Bloomington and Zürich but was extremely fortunate to meet some great people in both places. I would like to thank Lena Mendes-Soares, Doug Drury, Victoria Jideonwo, Daniela Vergara, Peter Zee, Elise Morton, Paolo Ocampo, Daniel Angst, Helen Alexander, Elin Lilja, Ying-Jui Hsu, Shanshan Hsieh, Yu-Ting Huang, Chien-Yi Wang, Der-Yuan Yu and many others for accepting who I am and all the fun we had together. I hope our friendships continue as long as possible.

Finally, I would like to thank my parents and sister for letting me discover the world on my own. Things wouldn't be possible without them.

I-Chen Chen

EVOLUTION OF DEVELOPMENTAL GENE REGULATION IN THE MYXOBACTERIA

A rapidly growing body of evidence has shown that non-coding small RNAs (sRNAs) regulate a variety of important biological processes across all domains of life, including bacteria. The sRNA Pxr in the model myxobacterial species *Myxococcus xanthus* functions as a developmental gatekeeper that prevents the initiation of fruiting body development until nutrients have been depleted. My dissertation research has focused on the origin and evolution of Pxr and its associated regulatory network in the myxobacteria. Using a combination of phylogenetic and molecular-genetic approaches, I tracked the origin of Pxr and examined its evolution at both sequence and functional levels in the myxobacteria. I showed that Pxr appears to have a single origin at the base of the suborder Cystobacterineae within the Myxococcocales (myxobacteria) order. Homologs of *pxr* are highly conserved and may play a common fundamental role in regulating fruiting body formation across diverse species of myxobacteria. Nevertheless, *pxr* duplications occurred in the genus *Cystobacter* and the specificity of its function may be evolving in these lineages. Further, following from a previous mutagenesis screen to identify genes involved in the Pxr regulatory pathway, I identified four genes that appeared to have important roles in the Pxr pathway. I characterized the evolutionary divergence of these genes across species and functional roles of some of these genes in Pxr synthesis, processing or function in *M. xanthus*. Finally, I discovered new *pxr* duplications in several *Cystobacter* species from whole-genome sequence data that had not been previously identified. Taken together, my research characterizes the evolutionary origin and diversification of a bacterial sRNA, a class of

regulatory elements that has great importance in the function and evolution of bacterial genomes. This work also provides insights into the evolution of developmental gene regulation in prokaryotes.

Gregory J. Velicer, Ph.D.

Patricia L. Foster, Ph.D.

Jeffrey D. Palmer, Ph.D.

Rudolf A. Raff, Ph.D.

Michael J. Wade, Ph.D.

Yuen-Tsu Nicco Yu, Ph.D.

TABLE OF CONTENTS

INTRODUCTION

"No need to hurry. No need to sparkle. No need to be anybody but oneself."

Virginia Woolf (1929)

Adapting to varying environments is essential for survival for all living organisms in nature. Because expressing alternative behaviors or phenotypes is crucial under stressful conditions but costly during normal growth in terms of energy and substrates, many organisms adapt to environmental changes through regulated gene expression. In recent years, extensive studies have established non-coding small RNAs (sRNAs) as important regulators of gene expression in bacteria (Gottesman and Storz, 2011). Many of these sRNAs act in *trans* and control expression of multiple genes. They play central roles in a plethora of biological processes including stress responses, such as iron limitation (Masse and Gottesman, 2002), oxidative stress (Altuvia et al., 1997), glucose-phosphate stress (Vanderpool and Gottesman, 2004) and growth-substrate deprivation (Yu et al., 2010), and have been shown to be implicated in the social traits such as quorum sensing (Lenz et al., 2004; Feng et al., 2015), biofilm formation (Chambers and Sauer, 2013) and fruiting body development (Yu et al., 2010).

The sizes of most known bacterial sRNAs range from ~50 to 300 nucleotides, and they are mostly encoded in intergenic regions and possess their own promoters and ρ-independent terminators (a GC-rich stem-loop that ends with a run of Ts). These sRNAs are post-transcriptional regulators that control gene expression via secondary structures and complementary interactions with mRNA targets, leading to translational repression, degradation

of mRNA targets, or both. Specific proteins are required for their modes of action. For instance, many characterized sRNAs require the RNA chaperone Hfq for their base pairing with mRNAs (De Lay et al., 2013), although a number of sRNAs in certain species, such as a Fur-regulated sRNA in *Bacillus subtilis*, are independent of Hfq (Gaballa et al., 2008). A combination of theoretical and experimental studies shows that, in contrast to protein regulators, sRNAs set up thresholds of and reduce noises in their target gene expression (Levine et al., 2007). Most research on bacterial sRNAs is aimed at elucidating molecular mechanisms of their modes of action and functional roles in model species such as *Escherichia coli*. Little is explored about the evolutionary origins or subsequent diversification of sRNAs.

~~~~

Study organisms: the fruiting myxobacteria

The myxobacteria are unusual in bacteria for their multicellular fruiting body development and social behaviors. These Gram-negative, rod-shaped bacteria belong to the delta-subgroup of proteobacteria and are treated as a monophyletic taxonomic order, the Myxococcales (Shimkets et al., 2006). Myxobacteria are widely found in soil, and also on dung of herbivorous animals, decaying plant materials, or the bark of trees (Dawid, 2000). Recent studies have shown that they appear to be common in and near marine seafloors too (Jiang et al., 2010; Brinkhoff et al., 2012).

The life history of myxobacteria is shown in Figure 1. During vegetative growth, myxobacterial cells glide on surfaces and collectively feed on macromolecules, bacteria or fungi
~~~~

by secreting extracellular digestive enzymes, hence placing them near or at the top of the microbial food chain (Lueders et al., 2006). However, upon nutrient deprivation, many species of myxobacteria initiate a developmental process that builds up multicellular fruiting bodies and only a subset of cells inside fruiting bodies becomes stress-resistant spores. The morphologies of fruiting bodies in different myxobacterial species are diverse in shape, size and color (Shimkets et al., 2006). For example, in the model species *Myxococcus xanthus*, fruiting bodies are hay-stack shaped, whereas *M. stipitatis* fruiting bodies are elevated on individual base stalks. *Cystobacter fuscus* fruiting bodies are clusters of sporangioles embedded in slime sheets, and *Stigmatella aurantiaca* creates several sporangioles attached by tiny stalks to a common supporting base.

Myxobacterial fruiting body development is biologically costly and requires extensive cell–cell signaling and interactions, as well as coordinated changes in gene expression and cell movement (Kroos, 2007; Leonardy et al., 2008). The initiation of development in *M. xanthus* depends on the RelA-mediated stringent response (Singer and Kaiser, 1995). As amino acid levels decrease, the guanosine penta- and tetraphosphate (together (p)ppGpp) synthesized by RelA accumulates in the cells. High levels of (p)ppGpp induce the extracellular A-signal, a cell density signal that estimates the starvation at the population level to ensure an accurate assessment of nutrient status (Singer and Kaiser, 1995). Next, the morphogenetic C-signal helps to pattern group-level cell movement and fruiting body morphogenesis (Kaiser, 2004). A recent study showed that the entry into development in *M. xanthus* is regulated not only by RelA but also by a regulatory sRNA (Yu et al., 2010).

The sRNA Pxr in the model system *M. xanthus*

Pxr is the first regulatory sRNA identified and characterized in *M. xanthus* development. It functions as a developmental gatekeeper that prevents the initiation of fruiting body development when amino acids, peptides and proteins are abundant (Yu et al., 2010). The *pxr* gene is located in the intergenic region between the σ^{54}-dependent response regulator *nla19* and an acetyltransferase gene. Pxr appears to be transcribed from an upstream σ^{54} promoter, terminated at its own ρ-independent terminator at the 3' end and is predicted to form a stable stem-loop structure similar to those of other bacterial sRNAs. There are two specific Pxr sRNA forms: Pxr-L (long) and Pxr-S (short). During vegetative growth, both Pxr-L and Pxr-S are expressed at high levels. When cells are starved for amino acids, the amount of Pxr-S is substantially decreased and development proceeds, suggesting that Pxr-S but not Pxr-L is the main regulator that down-regulates developmentally specific gene expression.

In this thesis, I focus on the origin and evolution of Pxr sRNA and its associated regulatory network across different species in the myxobacteria. Among sequenced genomes, detected homologs of *pxr* were restricted to a subclade of the Myxococcocales order, suggesting a recent origin within this order and providing the opportunity to examine the origins and subsequent divergence of a bacterial sRNA. Overall, I integrate approaches from phylogenetics, comparative genomics and molecular genetics.

~~~~
~~~~

In **Chapter I**, I investigate the phylogenetic distribution of Pxr, in particular, whether Pxr sRNAs share a common ancestry but were lost in some lineages or have independent origins. I first reconstructed the phylogenetic framework of myxobacteria using sequences of five conserved genes and identified the presence or absence of Pxr in different myxobacterial species. My results showed that Pxr is widespread across the suborder Cystobacterineae, suggesting that Pxr recently originated in the lineage basal to this suborder. I also examined the copy number, sequence conservation and gene neighborhood of Pxr.

In **Chapter II**, I study the functional divergence of Pxr in the myxobacteria, specifically, whether Pxr may have coevolved with its targets or not. I introduced *pxr* alleles from different species into an *M. xanthus* deletion mutant lacking *pxr* to test how divergent alleles of *pxr* function in a common genomic background of a standard lab strain. All *pxr* alleles from species with only one copy of this gene controlled development in *M. xanthus* in a manner qualitatively similar to that of the native *M. xanthus* allele. Nevertheless, two paralogs found in the genus *Cystobacter* failed to control development. My results illustrate both that Pxr may play a common fundamental role in developmental gene regulation across diverse species of myxobacteria and that the specificity of its function may be evolving in some lineages.

Thus far, little has been elucidated about the genes involved in the Pxr regulatory function in *M. xanthus*. As reported in **Chapter III**, following a mutagenesis screen to identify genes involved in the Pxr regulatory pathway, I identified four genes that appear to have important roles in the Pxr pathway. These genes are distinct from those described in other bacterial sRNA-based regulation (e.g. the ones in *E. coli*) and their protein sequences in the sequenced

myxobacterial species within the suborder Cystobacterineae where Pxr was distributed are highly similar to the ones in *M. xanthus*. The functional roles of some of these genes in Pxr synthesis, processing or function were investigated and discussed.

Species in the genera *Archangium* and *Cystobacter* in the myxobacteria are not well studied. In the work reported in Chapter I, Pxr sRNA or the *pxr* gene was not detected in one strain of *Archangium gephyra* and two strains of *Cystobacter fuscus*. In **Chapter IV**, I aim to determine whether Pxr is truly lost in these strains or went undetected with previous methods by sequencing the whole-genomes of these strains with the single-molecule sequencing technology recently developed by Pacific Biosciences. The whole-genome sequence data revealed multiple *pxr* duplications that had not been previously identified. Some duplicated alleles are highly divergent from others yet some are identical or almost identical to each other, indicating both old and very recent duplications. The findings here provide future opportunities for examining the role that gene duplication plays in the evolution of bacterial sRNAs.

Together, my research characterizes the evolutionary origin and diversification of a bacterial sRNA, a class of regulatory elements that has great importance in the function and evolution of bacterial genomes. Thus far, targets of Pxr are not identified. It is commonly postulated that sRNAs closely evolve with their targets and the determination of Pxr targets in the future will allow such testing. Also, further work on the identification and characterization of sRNAs involved in fruiting body development in the myxobacteria will provide insights into the evolution of developmental gene regulation in this unique prokaryotic group.

References

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., Storz, G., 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. Cell 90, 43-53.

Brinkhoff, T., Fischer, D., Vollmers, J., Voget, S., Beardsley, C., Thole, S., Mussmann, M., Kunze, B., Wagner-Dobler, I., Daniel, R., Simon, M., 2012. Biogeography and phylogenetic diversity of a cluster of exclusively marine myxobacteria. ISME J. 6, 1260-1272.

Chambers, J.R., Sauer, K., 2013. Small RNAs and their role in biofilm formation. Trends Microbiol. 21, 39-49.

Dawid, W., 2000. Biology and global distribution of myxobacteria in soils. FEMS Microbiol. Rev. 24, 403-427.

De Lay, N., Schu, D.J., Gottesman, S., 2013. Bacterial small RNA-based negative regulation: Hfq and its accomplices. J. Biol. Chem. 288, 7996-8003.

Feng, L., Rutherford, S.T., Papenfort, K., Bagert, J.D., van Kessel, J.C., Tirrell, D.A., Wingreen, N.S., Bassler, B.L., 2015. A Qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. Cell 160, 228-240.

Gaballa, A., Antelmann, H., Aguilar, C., Khakh, S.K., Song, K.B., Smaldone, G.T., Helmann, J.D., 2008. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. Proc. Natl. Acad. Sci. USA 105, 11927-11932.

Gottesman, S., Storz, G., 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb. Perspect. Biol. 3, a003798.

Jiang, D.M., Kato, C., Zhou, X.W., Wu, Z.H., Sato, T., Li, Y.Z., 2010. Phylogeographic separation of marine and soil myxobacteria at high levels of classification. ISME J. 4, 1520-1530.

Kaiser, D., 2004. Signaling in myxobacteria. Annu. Rev. Microbiol. 58, 75-98.

Kroos, L., 2007. The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. Annu. Rev. Genet. 41, 13-39.

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.

Leonardy, S., Bulyha, I., Sogaard-Andersen, L., 2008. Reversing cells and oscillating motility proteins. Mol. Biosyst. 4, 1009-1014.

Levine, E., Zhang, Z., Kuhlman, T., Hwa, T., 2007. Quantitative characteristics of gene regulation by small RNA. PLoS Biol. 5, e229.

Lueders, T., Kindler, R., Miltner, A., Friedrich, M.W., Kaestner, M., 2006. Identification of bacterial micropredators distinctively active in a soil microbial food web. Appl. Environ. Microb. 72, 5342-5348.

Masse, E., Gottesman, S., 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 99, 4620-4625.

Shimkets, L., Dworkin, M., Reichenbach, H., 2006. The Myxobacteria. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackenbrandt, E. (Eds.), The Prokaryotes, Springer, New York, NY.

Singer, M., Kaiser, D., 1995. Ectopic production of guanosine penta- and tetraphosphate can initiate early developmental gene expression in *Myxococcus xanthus*. Gene. Dev. 9, 1633-1644.

Vanderpool, C.K., Gottesman, S., 2004. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. Mol. Microbiol. 54, 1076-1089.

Yu, Y.T., Yuan, X., Velicer, G.J., 2010. Adaptive evolution of an sRNA that controls *Myxococcus* development. Science 328, 993.

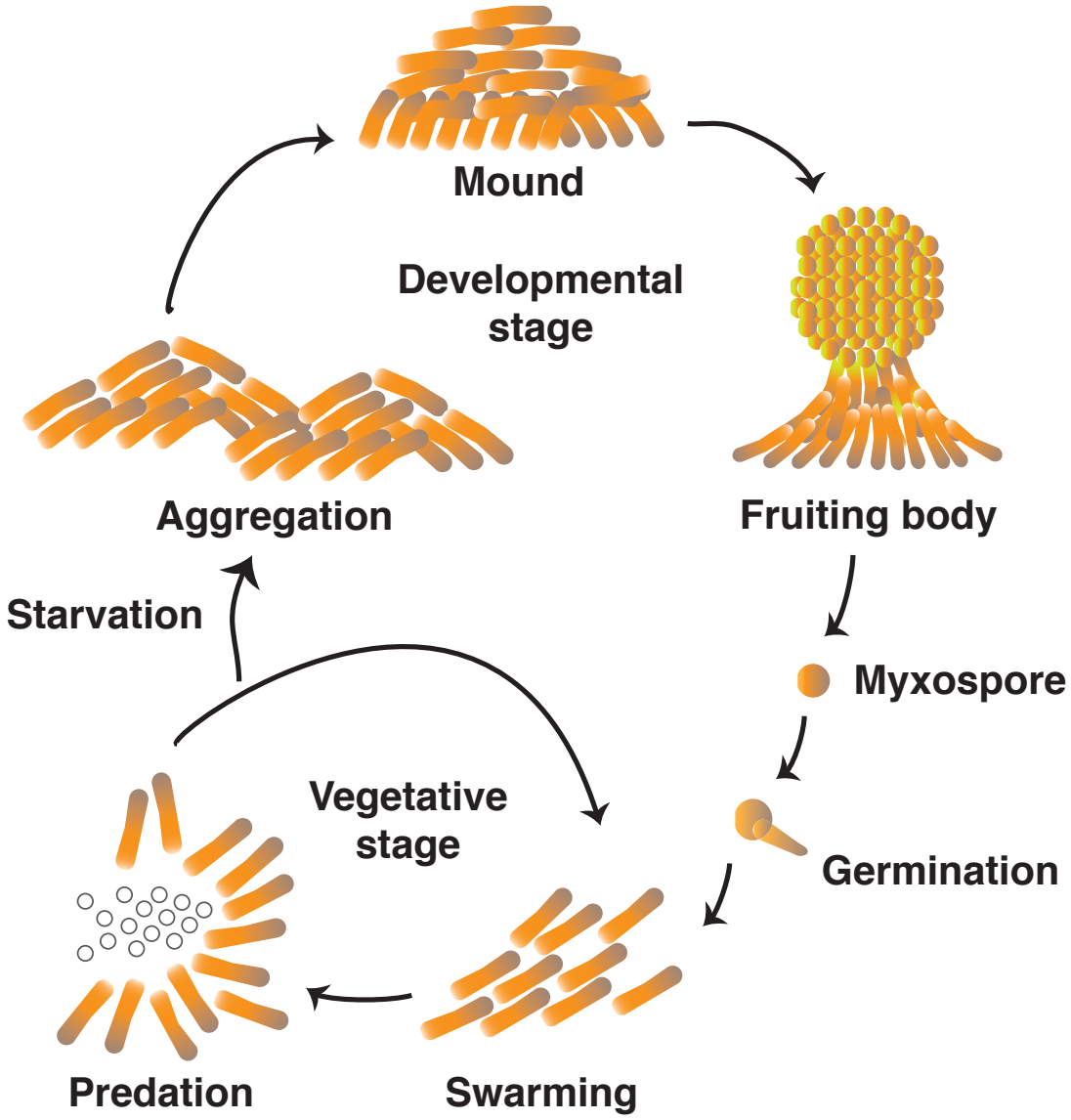


Figure 1. Life history of myxobacteria, shown by the example of *M. xanthus*.

CHAPTER I

A recent evolutionary origin of a bacterial small RNA that controls multicellular fruiting body development

I-Chen Chen[1,2*], Brad Griesenauer[1], Yuen-Tsu Nicco Yu[1,2] and Gregory J. Velicer[1,2]

[1] Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

[2] Institute of Integrative Biology (IBZ), ETH Zurich, CH-8092 Zurich, Switzerland

*E-mail: icchen@indiana.edu

Molecular Phylogenetics and Evolution (2014) Volume 73: 1-9.

Abstract

In animals and plants, non-coding small RNAs regulate the expression of many genes at the post-transcriptional level. Recently, many non-coding small RNAs (sRNAs) have also been found to regulate a variety of important biological processes in bacteria, including social traits, but little is known about the phylogenetic or mechanistic origins of such bacterial sRNAs. Here we propose a phylogenetic origin of the myxobacterial sRNA Pxr, which negatively regulates the initiation of fruiting body development in *Myxococcus xanthus* as a function of nutrient level, and also examine its diversification within the Myxococcocales order. Homologs of *pxr* were found throughout the Cystobacterineae suborder (with a few possible losses) but not outside this clade, suggesting a single origin of the Pxr regulatory system in the basal Cystobacterineae lineage. Rates of *pxr* sequence evolution varied greatly across Cystobacterineae sub-clades in a manner not predicted by overall genome divergence. A single copy of *pxr* was found in most species with 17% of nucleotide positions being polymorphic among them. However three tandem paralogs were present within the genus *Cystobacter* and these alleles together exhibited an elevated rate of divergence. There appears to have been strong selection for maintenance of a predicted stem-loop structure, as polymorphisms accumulated preferentially at loop or bulge regions or as complementary substitutions within predicted stems. All detected *pxr* homologs are located in the intergenic region between the σ^{54}-dependent response regulator *nla19* and a predicted NADH dehydrogenase gene, but other neighboring gene content has diversified.

Keywords: bacterial development, multicellularity, myxobacteria, non-coding small RNAs, post-transcriptional regulation, social evolution

1. Introduction

In recent years, a large and growing body of evidence has shown that non-coding small RNAs regulate a wide range of important biological processes across all domains of life (Chen et al.; Ding and Voinnet, 2007; Stefani and Slack, 2008; Voinnet, 2009; Gottesman and Storz, 2011). The most extensively studied type of such small RNA regulators controls gene expression at the post-transcriptional level by base pairing with mRNA targets and thereby causing the degradation of mRNA, inhibiting translation, or both. In bacteria, regulatory small RNAs (sRNAs) are often associated with stress responses in changing environments, such as oxidative stress, iron limitation, glucose-phosphate stress and growth-substrate deprivation (Altuvia et al., 1997; Masse and Gottesman, 2002; Vanderpool and Gottesman, 2004; Yu et al., 2010), and have been shown to regulate social traits such as biofilm formation (Chambers and Sauer, 2013), quorum sensing (Bejerano-Sagie and Xavier, 2007) and fruiting body development (Yu et al., 2010). Most known bacterial sRNAs are ~100 nucleotides long and are thus longer than their eukaryotic counterparts (e.g. 21-24 nucleotides for small interfering RNAs (siRNAs) (Ding and Voinnet, 2007) and 19-25 nucleotides for microRNAs (miRNAs) (Stefani and Slack, 2008)). Bacterial sRNAs are mostly encoded in intergenic regions, possess their own promoters and rho-independent terminators (a run of Ts), are predicted to fold into stable stem-loop structures and typically have multiple mRNA targets. Many sRNAs in enteric bacteria require the RNA chaperone Hfq for sRNA-based gene regulation, although a number of sRNAs in other species do not require Hfq proteins (Boisset et al., 2007; Gaballa et al., 2008).

Bacterial sRNAs are functionally analogous to eukaryotic miRNAs in their ability to modulate gene expression at the post-transcriptional level (Waters and Storz, 2009; Gottesman and Storz, 2011). Studies of the phylogenetic distributions of miRNAs have indicated independent origins of miRNA families at the bases of animal and plant clades, followed by subsequent expansion and diversification among specific lineages (Berezikov, 2011; Cuperus et al., 2011). Increasing evidence has also uncovered possible genomic sources from which miRNAs can evolve. For instance, in animals, other than random transcripts as *de novo* sources of miRNAs, genetic elements with existing structures such as small nucleolar RNAs, tRNAs, and transposon elements are possible sources from which novel miRNAs can be generated (Piriyapongsa et al., 2007; Ender et al., 2008; Burroughs et al., 2011; Yuan et al., 2011). In addition, introns appear to be common evolutionary pre-cursors of miRNAs as they provide RNA material that is already transcribed and do not require new promoter units (Campo-Paysaa et al., 2011). Not surprisingly, gene duplication events also contribute to the proliferation of miRNA families (Hertel et al., 2006).

Known bacterial sRNAs appear to differ greatly in size and sequence but little is known about their origins or subsequent diversification (Gottesman and Storz, 2011). Investigation of the phylogenetic distributions of bacterial sRNAs therefore should reveal their evolutionary origins, in particular whether known sRNAs share a common ancestry but were lost in some lineages or have independent origins. For example, the sRNA RyhB, which is regulated by the Fur repressor, was first found to down-regulate iron-containing proteins in response to low iron in *Escherichia coli* (Masse and Gottesman, 2002). Other Fur-regulated sRNAs were subsequently identified in *Pseudomonas*, *Neisseria* and *Bacillus subtilis* that do not appear to

share conserved elements with RyhB beyond the Fur binding site (Wilderman et al., 2004; Mellin et al., 2007; Gaballa et al., 2008), suggesting either rapid divergence since their origin or convergent evolution of function. SgrS-like sRNAs associated with glucose-phosphate stress are widely found among enteric bacteria and other Gamma-proteobacteria such as *Aeromonas hydrophila* and *A. salmonicida* (Horler and Vanderpool, 2009), but their phylogenetic origins, like those of Fur-regulated sRNAs, have not been resolved. Several possible evolutionary sources of bacterial sRNAs have been proposed, including tRNAs, mRNAs, gene duplications, and low-level antisense transcription (Gottesman and Storz, 2011), but only gene duplication events have been demonstrated (Lenz et al., 2004; Wilderman et al., 2004).

In this study, we sought to identify the evolutionary origin of the myxobacterial sRNA Pxr. The myxobacteria constitute a monophyletic order (Myxococcales) within the delta-proteobacteria (Shimkets et al., 2006) and are both widespread in terrestrial soils (Dawid, 2000) and appear to be common in and near marine seafloors (Jiang et al., 2010; Brinkhoff et al., 2012). Many myxobacteria species exhibit dramatic social behaviors, including the model species *Myxococcus xanthus*. During vegetative growth, *M. xanthus* cells collectively swarm and prey on other bacteria, fungi, or macromolecules by secreting extracellular digestive enzymes. Most remarkably, upon starvation they initiate a multicellular developmental process that culminates in the formation of spore-bearing fruiting bodies. Only a subset of cells that aggregate into fruiting body mounds become stress-resistant spores, whereas the remainder of cells either undergo cell lysis or remain rod-shaped at the fruiting body periphery (Wireman and Dworkin, 1977; Rosenbluh et al., 1989; O'Connor and Zusman, 1991a; b).

Myxobacterial fruiting body development requires extensive cell-cell signaling and interactions, as well as coordinated changes in gene expression and cell movement (Kroos, 2007; Leonardy et al., 2008). For example, in *M. xanthus* the extracellular A-signal first acts as a quorum sensor to assess starvation at the population level (Singer and Kaiser, 1995). Next, the morphogenetic C-signal helps to regulate group-level cell movement and fruiting body morphogenesis (Kaiser, 2004). The transition from vegetative growth to development in *M. xanthus* was previously shown to depend on the RelA-mediated stringent response (Singer and Kaiser, 1995), but more recently the sRNA Pxr has also been implicated in governing this transition (Yu et al., 2010). Pxr appears to function as a developmental gatekeeper by blocking development when nutrients are abundant. The *pxr* gene is located in the intergenic region between the σ^{54}-dependent response regulator *nla19* and an acetyltransferase gene (predicted genes Mxan_1078 and Mxan_1079, respectively, in the *M. xanthus* genome; (Goldman et al., 2006)). Pxr appears to be transcribed from an upstream σ^{54} promoter and is predicted to form a stable multi-stem-loop structure similar to those of other non-coding regulatory sRNAs (Gottesman, 2005; Yu et al., 2010).

Intriguingly, a *pxr* homolog was previously found only in *Stigmatella aurantiaca* (Yu et al., 2010), a myxobacterium classified within the same suborder as *M. xanthus* (Cystobacterineae), but not in other three sequenced myxobacterial genomes: *Sorangium cellulosum* (suborder Sorangineae), *Haliangium ochraceum* (suborder Nannocystineae), and *Anaeromyxobacter dehalogenan* (a non-fruiting and uniquely anaerobic myxobacterium) (Shimkets et al., 2006). Recent comparative genomic analysis has suggested that the genetic programs for fruiting body formation in *M. xanthus* and *S. aurantiaca* are similar to each other but significantly different

from the programs of *S. cellulosum* and *H. ochraceum* (Huntley et al., 2011).

The fact that *pxr* was found in *M. xanthus* and *S. aurantiaca* but not in other published genomes suggested a recent origin within the myxobacteria and provided the opportunity to examine the origin and subsequent diversification of a bacterial sRNA and its associated regulatory network. Here, we constructed a phylogeny of 28 myxobacterial strains representing 18 named species using sequences of five conserved genes. We detected the presence or absence of Pxr by Northern hybridization in these strains and exploited conservation of the gene neighborhood surrounding *pxr* in *M. xanthus* to sequence *pxr* homologs across the myxobacteria. These data allowed us to identify the likely phylogenetic origin of *pxr* and to characterize its diversification within the myxobacteria.

2. Materials and Methods

2.1. Bacterial strains

Myxobacterial strains investigated in this study were obtained from Hans Reichenbach (Sproer et al., 1999), the German Collection of Microorganisms and Cell Cultures (DSMZ) or had previously sequenced genomes (Table 1.1). We examined 24 strains representing 14 species within the suborder Cystobacterineae and four from outside this suborder (two species in the suborder Nannocystineae, one species in the suborder Sorangineae and one unclassified species). The genomes of seven of the 28 strains examined here have been published (Goldman et al., 2006; Schneiker et al., 2007; Thomas et al., 2008; Ivanova et al., 2010; Huntley et al., 2011; Li et

al., 2011; Huntley et al., 2012). Most of these strains used here were previously classified based on morphological traits such as the size and structure of the vegetative cells, fruiting bodies and myxospores (Sproer et al., 1999; Fudou et al., 2002; Zhang et al., 2005; Shimkets et al., 2006). In Sproer *et al* (1999), morphological and 16S rRNA-based phylogenies were found to be largely congruent at the genus level but not at the species level.

2.2. Culturing and sequencing

Frozen strain stocks were first inoculated on Casitone-Tris (CTT) (Hodgkin and Kaiser, 1977) or yeast (VY/2) (Reichenbach and Dworkin, 1992) hard (1.5%) agar medium at 32 °C with a relative humidity of 90%. After about one week, bacterial cells were transferred to liquid medium and grown until turbid in preparation for DNA and RNA extraction. Genomic DNA was extracted following Wu & Kaiser (1995). The 16S rRNA gene was first PCR-amplified and sequenced in all strains to confirm species' identities. We then sequenced another four conserved loci to attain phylogenetic resolution at the species level: the 23S rRNA gene, *pyrG* (cytidine triphosphate synthetase), *rpoB* (DNA-directed RNA polymerase subunit beta), and *pgm* (phosphoglucomutase). Primer sequences used for PCR amplification and sequencing are listed in Table S1. The trace files of Sanger sequences were checked and assembled using CodonCode Aligner 3.7.1 (CodonCode Corporation, Dedham, MA) and exported to FASTA-formatted files for further analyses.

2.3. Phylogenetic analyses

Sequences of the five conserved loci were aligned with MUSCLE implemented in MEGA version 5.0 (Tamura et al., 2011). The three protein-coding genes (i.e. *pyrG*, *rpoB* and *pgm*) were aligned according to their corresponding protein sequences. Phylogenetic analyses were estimated using both maximum likelihood (ML) and Bayesian inference (BI). We conducted the ML analyses in MEGA 5.0 with sequence alignments of each individual locus and with 5486-nt concatenated sequences of the five loci based on their order in *M. xanthus genome* (16S rRNA - 23S rRNA - *pyrG* - *rpoB* - *pgm*) and estimated the reliabilities of the ML trees with 1,000 bootstrap replicates. The BI analyses of sequence alignments of each individual locus were run in MrBayes 3.1.2 with GTR **+**invgamma model for half a million generations with two parallel independent searches (Ronquist and Huelsenbeck, 2003). The BI analysis of the concatenated sequences was run in a similar fashion but for five million generations instead. In the BI analyses, the sequence datasets were partitioned into gene regions and/or codon positions when possible to allow evolutionary rates to be potentially different across genes and codon positions.

2.4. Identification of Pxr

To detect the presence of Pxr sRNA, the total RNA of each strain was extracted from cultures growing in CTT or VY/2 liquid and small-sized RNAs were enriched with the MirVana miRNA isolation kit (Ambion). Pxr was detected using Northern hybridization with 3'Biotin-TEGpxr oligo probe (5'-ACC GGA AGT GCT GAA GGG GTG GGG GG-3') as previously described (Yu et al., 2010). *pxr* homologs were amplified and sequenced utilizing primers complementary to segments of the gene neighborhood between the σ^{54}-dependent response regulator *nla19* and the downstream NADH dehydrogenase gene that are conserved across the

Cystobacterineae (Table S1). The sequence trace files were processed and aligned as described in the Subsection 2.3. The ML and BI gene trees were reconstructed in MEGA 5.0 and MrBayes 3.1.2. The sequence logo for the *pxr* coding region was generated using WebLogo 3 (Crooks et al., 2004). We inferred the likely ancestral sequence of *pxr* following the procedures in Hall (2011). Secondary structures of the Pxr homologs and ancestor were predicted with Mfold (Zuker, 2003).

2.5. Nucleotide sequence accession numbers

All of the nucleotide sequences analyzed in this study other than the previously sequenced genomes have been deposited in GenBank. The accession numbers are as follows: 16S rRNA gene, KF267719 to KF267739; 23S rRNA gene, KF270694 to KF270714; *pyrG*, KF304709 to KF304729; *rpoB*, KF304730 to KF304750; *pgm*, KF304751 to KF304770; *pxr*, KF304771 to KF304795.

3. Results

3.1. Genetic diversity and phylogeny

Sequences of all five conserved loci were obtained from all strains except for the *pgm* locus in *Nannocystis exedens* subsp. *cinnabaria*. Sequence diversity across the entire order Myxococcales and within the suborder Cystobacterineae is summarized in Table 1.2. In both cases, the 16S and 23S rRNA genes are more conserved than the protein coding genes examined

here. For example, the proportions of polymorphic sites in the 16S and 23S rRNA genes in the suborder Cystobacterineae are 13% and 17% respectively, whereas 33%, 28% and 49% of sites are polymorphic in *pyrG*, *rpoB* and *pgm*, respectively. The Bayesian and ML trees of each locus were shown in Figure S1-5. In general, the topologies of the 16S rRNA, 23S rRNA, *rpoB* and *pgm* trees are consistent with the morphological phylogeny (Sproer et al., 1999; Shimkets et al., 2006) at or above the genus level although there are some differences between gene trees and between Bayesian and ML trees of the same gene sequences, particularly within the suborder Cystobacterineae where the 16S rRNA-based phylogeny alone was less resolved previously (Sproer et al., 1999). The *pyrG* tree demonstrates a different topology, with *N. exedens* subsp. *cinnabaria* from the suborder Nannocystineae and the distantly related *Stigmatella erecta* DSM 16858 clustered with *M. fulvus* and *M. stipitatus* spp. in Cystobacterineae (Fig S3). Nevertheless, the alignment length of *pyrG* (435 nt) is much shorter compared to those of other genes (1017-1523 nt) and the number of polymorphic sites are lower in *pyrG* (195) for estimating phylogeny compared to those in other genes (435-612) (Table 1.2). The bootstrap value for this clustering is also below 50% (Fig S3B).

The Bayesian species phylogeny based on the five conserved loci for all 28 strains is shown in Figure 1.1. The ML analysis also generated the same tree topology. The posterior probabilities and bootstrap values for most branching events are above 90% and 70%, respectively, while a few recent branching events have lower probabilities. Three species, *S. cellulosum* So ce 56 (Suborder: Sorangineae), *H. ochraceum* DSM14365 and *N. exedens* subsp. *cinnabaria* (Suborder: Nannocystineae), represent a subclade that excludes all other strains. This pattern is consistent with the previous 16S rRNA-based and morphological phylogenies of the

myxobacteria (Sproer et al., 1999; Shimkets et al., 2006) and we used these three more distantly related species as outgroup to root the tree. The anaerobic *A. dehalogenan* 2CP-C branches as a singleton outside the suborder Cystobacterineae, but together with the Cystobacterineae forms a combined clade. Within Cystobacterineae, the branching pattern is generally consistent with the morphological phylogeny at the genus level, but multiple isolates with the same species label do not cluster into common clades in several instances, particularly within the genus *Myxococcus*, as seen in Sproer et al. (1999).

3.2. Origin and diversification of Pxr

We detected Pxr RNA and/or the *pxr* gene in 21 out of the 28 strains examined. In seventeen strains with unsequenced genomes, both PCR amplification of the intergenic region between *nla19* and the downstream NADH dehydrogenase gene and Northern hybridization detected *pxr* and Pxr (Fig S6), respectively, and in four sequenced myxobacterial genomes within the Cystobacterineae suborder, BlastN searches also led to the detection of *pxr*. The probe we used to detect Pxr in Northern hybridization contains the first loop and the right side of the first stem, which is conserved between our model species *M. xanthus* and *S. aurantiaca,* the most distantly related species within the Cystobacterineae. Overall, Pxr is widely distributed across the suborder Cystobacterineae but was not detected in *Archangium gephyra* Arg2 or the *Cystobacter fuscus* strains Cbf8 and Cbf10 (Fig 1.1). In those three strains, PCR amplification of the region between *nla19* and the NADH dehydrogenase gene was not successful. An ~400 bp band from *A. gephyra* Arg2 was obtained but the sequencing result matched the reading frame of a 5'-nucleotidase gene, possibly caused by mispriming during PCR. Thus, it appears that either this

gene neighborhood is not conserved in these three species or the primer sequences we used are not conserved in the respective genomes. Neither a BlastN search nor PCR amplification detected *pxr* in the four strains outside the Cystobacterineae suborder, suggesting that Pxr originated in the lineage basal to this suborder.

Among the 17 strains in which Pxr was detected by Northern hybridization, PCR amplification of the intergenic region between *nla19* and the NADH dehydrogenase gene consistently yielded a product ~1.5 kb in size. Sequencing revealed a single copy of *pxr* within these PCR products for the majority of strains, but three tandem paralogs of *pxr* were found in the products from *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 (*pxr-1*, *pxr-2* and *pxr-3* from 5' to 3'). A total of 14 distinct *pxr* alleles were found among the 21 strains, with four of those alleles being present in more than one species and three alleles present in both *C. minor* and *C. violaceus* (Fig 1.2). All of the *pxr* homologs identified were preceded by a predicted σ^{54} promoter sequence.

The Bayesian gene tree of *pxr* (Fig 1.2) is qualitatively congruent with the concatemer-based species phylogeny (Fig 1.1), which suggests that *pxr* inheritance has been solely vertical among these species since its origin. In the Bayesian tree, the *Angiococcus*/*Corallococcus*/*Myxococcus* genera form a *pxr* clade (nine alleles) that contains two multi-strain subclades, one specific to *M. fulvus* and *M. stipitatus* (four alleles) and one containing representatives of *M. flavescens, M. macrosporus, M. virescens* and *M. xanthus* (two alleles). The *Cystobacter* and *Stigmatella* genera represent two distinct *pxr* clades containing four *pxr* alleles and just one allele, respectively. In the ML tree, the branching pattern is overall similar to the one in the Bayesian tree, except that

the *pxr* allele from the genus *Corallococcus* is clustered with the *Cystobacter* alleles, although the bootstrap value supporting this branching pattern is below 50% (Fig S7).

Notably, the three primary *pxr* clades have evolved at strikingly different rates since their divergence (Fig 1.2). The *Cystobacter* cluster evolved much more extensively than the other two clades, whereas the *Stigmatella* cluster evolved little. The elevated rate of *Cystobacter pxr* evolution appears to be specific to *pxr* rather than being due to an elevated rate of overall genome evolution, as the *Cystobacter* clade in the concatemer-based phylogram does not show a similar pattern of greatly elevated divergence from the other two major clades (Fig 1.1). The degree of within-clade *pxr* divergence as reflected by allelic richness ((total # of alleles / # of strains examined) within each clade): *Stigmatella* clade = 0.25 (one allele/four strains); *Angiococcus/Corollococcus/Myxococcus et al*. clade 0.6 (nine alleles/15 strains); *Cystobacter* clade 2.0 (four alleles/two strains), although the strains examined here are not considered to represent random sampling of these clades.

Seventeen percent of sites in the single-copy homologs of *pxr* are polymorphic, comparable to the conservation level of the 23S rRNA gene (Table 1.2). When the tandem copies of *pxr* in *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 are also considered, the proportion of polymorphic sites among all *pxr* alleles increases to 37%. This increased polymorphism is due primarily to divergence among paralogs within each strain rather than among homologous alleles across strains. For example, the pairwise distances between *pxr-1* and *pxr-2*, *pxr-2* and *pxr-3*, and *pxr-1* and *pxr-3* in *C. minor* Cbm 6 are 15%, 17% and 22%, respectively. However, *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 share the same *pxr-1* and *pxr-3* alleles and their *pxr-2* alleles differ by

only two nucleotides. One difference is the presence of an extra cytosine in the first loop in *C.violaceus* Cbvi 34 and the other one is at a polymorphic site (A/G) in the single-stranded region between the second and third stem-loops.

The distribution of polymorphic sites is not random. All of the single-copy *pxr* homologs are predicted to fold into stable triple stem-loop structures (Mfold, Fig S8). For the tandem copies in *Cystobacter*, *pxr-1* and *pxr-3* are predicted to fold into only two long stem-loops, missing the short stem-loop that is present in *pxr-2* and other single-copy *pxr* homologs. The sequence logo for the overall *pxr* homolog alignment reveals greater sequence conservation in the first stem-loop and the middle region of the third stem (i.e. nucleotide positions 67-78 and 98-105) compared to other regions of the Pxr secondary structure (Fig 1.3A). The nucleotide sequences of each *pxr* allele at all sites are depicted in Fig S9.

We inferred the sequence of the most recent common ancestor of extant *pxr* alleles in the clade that includes both *Cystobacter* and *Myxococcus* spp. by rooting the *pxr* tree with the *Stigmatella* allele (the most distantly related lineage in Cystobacterineae) and estimated the ancestral status at the internal node shared by the non-*Stigmatella pxr* homologs (Fig 1.2). The inferred ancestral gene is 111-nt long, also predicted to fold into a stable three stem-loop structure (Mfold, Fig 1.3B) and differs from the *Stigmatella* allele only by the presence of two bases (CU) in the third loop. A comprehensive BlastN search with this inferred ancestral sequence did not yield any significant hits other than previously identified *pxr* homologs.

We characterized *pxr* divergence by tracking the nucleotide changes per site at each internal

node from the estimated ancestral allele to the extant alleles. The different types of changes that have occurred on the predicted stem-loop structure of Pxr are summarized in Figure 1.3C. Most changes (77%) are single nucleotide substitutions and 23% are insertions or deletions. Among the single nucleotide substitutions, 51% occurred in a loop or bulge or in a single-stranded region, 34% of the changes are complementary substitutions at the matching positions on the predicted stems and the remaining 15% are non-complementary substitutions on the predicted stems. The deletion and insertion events are often at loop or bulge regions.

3.3. The pxr gene neighborhood

In all strains examined, the *pxr* locus is flanked immediately upstream by a variant of the σ^{54}-dependent response regulator *nla19* and downstream by a gene encoding NADH dehydrogenase, either directly next to *pxr* or nearby (Fig 1.4). The sequence content between *pxr* and NADH dehydrogenase varies among genera. In all *Myxococcus* strains and *Angiococcus*, *pxr* is flanked immediately downstream by an acetyltransferase gene, whereas in the genus *Stigmatella* it is followed by a hypothetical protein. In the species *C. minor* Cbm 6, *C. violaceus* Cbvi 34 and *C. coralloides* no other coding locus is found between *nla19* and NADH dehydrogenase. The intergenic region in *C. coralloides* DSM2259 is shorter (532 bp) than in other species (e.g. 1328 bp in *M. xanthus* DK1622 and 1530 bp in *S. aurantiaca* DW4/3-1).

4. Discussion

Bacterial sRNAs are widespread, but both the phylogenetic positions and mechanisms of

their evolutionary origins remain largely obscure. Here we have identified a likely single phylogenetic origin of a bacterial sRNA – Pxr – within the fruiting myxobacteria. No evidence of horizontal *pxr* transfer across myxobacterial species is evident and *pxr* was found to exhibit a degree of sequence conservation roughly similar to that of the 23S rRNA gene, which is widely used in construction of bacterial phylogenetic trees (Ludwig and Schleifer, 1994; Hunt et al., 2006). The nucleotide changes that accumulated within *pxr* are often in predicted loops, on bulges, or complementary changes on stems, indicating functional constraint on the stem-loop structure of Pxr. Also, some features of the *pxr* gene neighborhood were found to be conserved across all species in which *pxr* was detected.

Our results also reveal a case of bacterial sRNA duplication. In the *Cystobacter* clade, three tandem copies of *pxr* were found in the intergenic region between *nla19* and the NADH dehydrogenase gene in *C. minor* Cbm 6 and *C. violaceus* Cbvi 34. It is not clear whether the presence of multiple copies of *pxr* in *Cystobacter* has functional significance. The sequence and predicted structural variation among the three paralogs may reflect functional divergence, or these paralogs may retain a common function but multiple copies are employed to tightly regulate the entry into fruiting body development. Under this hypothesis, shared function may have reduced selection for sequence conservation in each individual copy, which in turn might explain increased divergence of the *Cystobacter* alleles. Tandem sRNAs copies have also been documented in *Pseudomonas aeruginosa* (Wilderman et al., 2004). In that case, the two sRNA alleles *prrF1* and *prrF2* both contribute to the regulation of target mRNAs under iron limitation.

Although *p*xr has been retained by most extant species in the Cystobacterineae suborder

since its origin, there are three possible exceptions that may represent evolutionary losses of *p*xr: *A. gephyra* Arg2, *C. fuscus* Cbf8 and Cbf10 (Fig 1.1). Neither Northern hybridization nor attempted PCR from primer sites within the conserved features of the *pxr* gene neighborhood revealed the presence of *pxr* homologs in these strains, suggesting evolutionary loss, although it is possible that highly divergent *pxr* alleles located in a different genomic region went undetected by these methods. Both *A. gephyra* and *C. fuscus* have been reported to form fruiting bodies (Sproer et al., 1999; Dawid, 2000), which raises the intriguing question of how Pxr-like regulation is accomplished (if at all) in these and other species in the suborders Sorangineae and Nannocystineae that appear to lack a *pxr* homolog. Assuming the initiation of fruiting body development in these species is also in response to nutrient deprivation as in *M. xanthus*, it is plausible that such species evolved distinct yet functionally analogous mechanisms for regulating the transition from growth to development.

The restriction of Pxr to the suborder Cystobacterineae in the myxobacteria is similar to the apparently restricted phylogenetic distribution of many other known bacterial sRNAs, suggesting that many bacterial sRNAs are specific to relatively narrow clades and may extensively coevolve with their targets. For example, many known sRNAs, such as SgrS in *E. coli*, are primarily found in the enteric bacteria, whereas PrrF-like sRNAs involved in iron homeostasis in *Pseudomonas aeruginosa* are only found in other *Pseudomonas* spp (Wilderman et al., 2004; Horler and Vanderpool, 2009). In animals, the evolution of microRNAs is often closely associated with the evolution of their targets (Berezikov, 2011). For example, the *miR-310* family is involved in late *Drosophila* development (Leaman et al., 2005). Introducing heterospecific *miR-310s* from *D. pseudoobscura* into *D. melanogaster* destabilizes the transcriptome of *D. melanogaster* and

reduces viability, indicating that *miR-310s* coevolved with their targets and that miRNAs and target genes from different species can interact adversely (Tang et al., 2010). Future identification of Pxr targets will allow similar testing for sRNA-target coevolution.

Fruiting bodies of different myxobacterial species are remarkably diverse in size, shape and color (Dawid, 2000). For example, *M. xanthus* fruiting bodies are hay-stack shaped, *M. stipitatis* fruiting bodies are elevated on a restricted base stalk and the more distant *S. aurantiaca* creates several sporangioles at the top of tall, thick stalks. Significantly, the predicted structure and sequence of Pxr are highly conserved across these morphologically distinct species, suggesting that this regulatory RNA plays a common fundamental role in development rather than mediating species-specific variation. However, even among strains sharing a common *pxr* allele, such variation might arise from distinct interactions between Pxr and other divergent components of the Pxr regulatory network that are not caused by sequence variation in Pxr *per se*. Future work is therefore necessary to test for functional conservation of Pxr and its relationships to the variable genetic programs for fruiting body formation found across different myxobacterial species (Huntley et al., 2011).

Table 1.1 Myxobacterial strains used in this study.

Taxonomic description	Strain	Culture media	Source or reference
Anaeromyxobacter dehalogenans	2CP-C	(Published genome)	(Thomas et al., 2008)
Angiococcus disciformis	And 1	VY/2	(Sproer et al., 1999)
Archangium gephyra	Arg 2	VY/2	(Sproer et al., 1999)
Corallococcus coralloides	DSM 2259	(Published genome)	(Huntley et al., 2012)
Cystobacter fuscus	Cbf 8	VY/2	(Sproer et al., 1999)
Cystobacter fuscus	Cbf 10	VY/2	(Sproer et al., 1999)
Cystobacter minor	Cbm 6	VY/2	(Sproer et al., 1999)
Cystobacter violaceus	Cbvi 34	VY/2	(Sproer et al., 1999)
Haliangium ochraceum	DSM 14365	(Published genome)	(Ivanova et al., 2010)
Myxococcus macrosporus	Ccm 7	VY/2	(Sproer et al., 1999)
Myxococcus macrosporus	Ccm 8	CTT	(Sproer et al., 1999)
Myxococcus fulvus	Mxf 2	VY/2	(Sproer et al., 1999)
Myxococcus fulvus	Mxf 421	CTT	(Sproer et al., 1999)
Myxococcus fulvus	Mxf 428	CTT	(Sproer et al., 1999)
Myxococcus fulvus	HW-1	(Published genome)	(Li et al., 2011)
Myxococcus flavescens	Mxfl 1	CTT	(Sproer et al., 1999)
Myxococcus stipitatus	Mxs 33	VY/2	(Sproer et al., 1999)

Myxococcus stipitatus	Mxs 42	VY/2	(Sproer et al., 1999)
Myxococcus virescens	Mxv 2	CTT	(Sproer et al., 1999)
Myxococcus virescens	Mxv 76	CTT	(Sproer et al., 1999)
Myxococcus xanthus	Mxx 132	CTT	(Sproer et al., 1999)
Myxococcus xanthus	DK1622	(Published genome)	(Goldman et al., 2006)
Nannocystis exedens subsp. *Cinnabaria*	DSM 14641	VY/2	DSMZ
Stigmatella aurantiaca		VY/2	DSMZ
Stigmatella aurantiaca	DW4/3-1	(Published genome)	(Huntley et al., 2012)
Stigmatella erecta	DSM 16858	VY/2	DSMZ
Stigmatella erecta	Pde 3	VY/2	(Sproer et al., 1999)
Sorangium cellulosum	So ce 56	(Published genome)	(Schneiker et al., 2007)

Table 1.2 Sequence diversity in the myxobacteria.

Clade	Gene(s)	No. of strains	Alignment length (nt)	No. of alleles	No. (%) of polymorphic sites
Myxococcales	16S rRNA	28	1523	26	439 (29)
	23S rRNA	28	1387	21	474 (34)
	pyrG	28	435	27	195 (45)
	rpoB	28	1017	28	435 (43)
	pgm	27	1106	27	612 (55)
	Concatemer	28	5468	28	2155 (39)
Cystobacterineae	16S rRNA	25	1512	23	202 (13)
	23S rRNA	25	1366	18	234 (17)
	pyrG	25	435	24	144 (33)
	rpoB	25	1005	25	276 (28)
	pgm	25	1106	25	546 (49)
	pxr[a]	21	113	14	42 (37)
	pxr[b]	19	112	10	19 (17)

a. Including all *p*xr alleles.

b. Excluding *Cystobacter* spp. *pxr* alleles.

References

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., Storz, G., 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. Cell 90, 43-53.

Bejerano-Sagie, M., Xavier, K.B., 2007. The role of small RNAs in quorum sensing. Curr. Opin. Microbiol. 10, 189-198.

Berezikov, E., 2011. Evolution of microRNA diversity and regulation in animals. Nat. Rev. Genet. 12, 846-860.

Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A.C., Benito, Y., Jacquier, A., Gaspin, C., Vandenesch, F., Romby, P., 2007. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. Gene. Dev. 21, 1353-1366.

Brinkhoff, T., Fischer, D., Vollmers, J., Voget, S., Beardsley, C., Thole, S., Mussmann, M., Kunze, B., Wagner-Dobler, I., Daniel, R., Simon, M., 2012. Biogeography and phylogenetic diversity of a cluster of exclusively marine myxobacteria. ISME J. 6, 1260-1272.

Burroughs, A.M., Ando, Y., de Hoon, M.J.L., Tomaru, Y., Suzuki, H., Hayashizaki, Y., Daub, C.O., 2011. Deep-sequencing of human argonaute-associated small RNAs provides insight into miRNA sorting and reveals argonaute association with RNA fragments of diverse origin. RNA Biol. 8, 158-177.

Campo-Paysaa, F., Semon, M., Cameron, R.A., Peterson, K.J., Schubert, M., 2011. MicroRNA complements in deuterostomes: origin and evolution of microRNAs. Evol. Dev. 13, 15-27.

Chambers, J.R., Sauer, K., 2013. Small RNAs and their role in biofilm formation. Trends Microbiol. 21, 39-49.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188-1190.

Cuperus, J.T., Fahlgren, N., Carrington, J.C., 2011. Evolution and functional diversification of *MIRNA* genes. Plant Cell 23, 431-442.

Dawid, W., 2000. Biology and global distribution of myxobacteria in soils. FEMS Microbiol. Rev. 24, 403-427.

Ding, S.W., Voinnet, O., 2007. Antiviral immunity directed by small RNAs. Cell 130, 413-426.

Ender, C., Krek, A., Friedlander, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., Meister, G., 2008. A human snoRNA with microRNA-like functions. Mol. Cell 32, 519-528.

Fudou, R., Jojima, Y., Iizuka, T., Yamanaka, S., 2002. *Haliangium ochraceum* gen. nov., sp. nov. and *Haliangium tepidum* sp. nov.: novel moderately halophilic myxobacteria isolated from coastal saline environments. J. Gen. Appl. Microbiol. 48, 109-116.

Gaballa, A., Antelmann, H., Aguilar, C., Khakh, S.K., Song, K.B., Smaldone, G.T., Helmann, J.D., 2008. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. Proc. Natl. Acad. Sci. USA 105, 11927-11932.

Goldman, B.S., Nierman, W.C., Kaiser, D., Slater, S.C., Durkin, A.S., Eisen, J., Ronning, C.M., Barbazuk, W.B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H.S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S.A., Vaudin, M., Wiegand, R., Kaplan, H.B., 2006. Evolution of sensory complexity recorded in a myxobacterial genome. Proc. Natl. Acad. Sci. USA 103, 15200-15205.

Gottesman, S., 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet. 21, 399-404.

Gottesman, S., Storz, G., 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb. Perspect. Biol. 3, a003798.

Hall, B.G., 2011. Phylogenetic Trees Made Easy: A How-To Manual, 4th Edition. Sinauer Associates, Sunderland, Massachusetts.

Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., Hofacker, I.L., Stadler, P.F., L, S.B.C., 2006. The expansion of the metazoan microRNA repertoire. BMC Genomics 7, 25.

Hodgkin, J., Kaiser, D., 1977. Cell-to-cell stimulation of movement in nonmotile mutants of *Myxococcus*. Proc. Natl. Acad. Sci. USA 74, 2938-2942.

Horler, R.S., Vanderpool, C.K., 2009. Homologs of the small RNA SgrS are broadly distributed in enteric bacteria but have diverged in size and sequence. Nucleic Acids Res. 37, 5465-5476.

Hunt, D.E., Klepac-Ceraj, V., Acinas, S.G., Gautier, C., Bertilsson, S., Polz, M.F., 2006. Evaluation of 23S rRNA PCR primers for use in phylogenetic studies of bacterial diversity. Appl. Environ. Microbiol. 72, 2221-2225.

Huntley, S., Hamann, N., Wegener-Feldbrugge, S., Treuner-Lange, A., Kube, M., Reinhardt, R., Klages, S., Muller, R., Ronning, C.M., Nierman, W.C., Sogaard-Andersen, L., 2011. Comparative genomic analysis of fruiting body formation in Myxococcales. Mol. Biol. Evol. 28, 1083-1097.

Huntley, S., Zhang, Y., Treuner-Lange, A., Kneip, S., Sensen, C.W., Sogaard-Andersen, L., 2012. Complete genome sequence of the fruiting myxobacterium *Corallococcus coralloides* DSM 2259. J. Bacteriol. 194, 3012-3013.

Ivanova, N., Daum, C., Lang, E., Abt, B., Kopitz, M., Saunders, E., Lapidus, A., Lucas, S., Del Rio, T.G., Nolan, M., Tice, H., Copeland, A., Cheng, J.F., Chen, F., Bruce, D., Goodwin, L.,

Pitluck, S., Mavromatis, K., Pati, A., Mikhailova, N., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.J., Jeffries, C.D., Detter, J.C., Brettin, T., Rohde, M., Goker, M., Bristow, J., Markowitz, V., Eisen, J.A., Hugenholtz, P., Kyrpides, N.C., Klenk, H.P., 2010. Complete genome sequence of *Haliangium ochraceum* type strain (SMP-2(T)). Stand. Genom. Sci. 2, 96-106.

Jiang, D.M., Kato, C., Zhou, X.W., Wu, Z.H., Sato, T., Li, Y.Z., 2010. Phylogeographic separation of marine and soil myxobacteria at high levels of classification. ISME J. 4, 1520-1530.

Kaiser, D., 2004. Signaling in myxobacteria. Annu. Rev. Microbiol. 58, 75-98.

Kroos, L., 2007. The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. Annu. Rev. Genet. 41, 13-39.

Leaman, D., Chen, P.Y., Fak, J., Yalcin, A., Pearce, M., Unnerstall, U., Marks, D.S., Sander, C., Tuschl, T., Gaul, U., 2005. Antisense-mediated depletion reveals essential and specific functions of microRNAs in *Drosophila* development. Cell 121, 1097-1108.

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.

Leonardy, S., Bulyha, I., Sogaard-Andersen, L., 2008. Reversing cells and oscillating motility proteins. Mol. Biosyst. 4, 1009-1014.

Li, Z.F., Li, X., Liu, H., Liu, X., Han, K., Wu, Z.H., Hu, W., Li, F.F., Li, Y.Z., 2011. Genome sequence of the halotolerant marine bacterium *Myxococcus fulvus* HW-1. J. Bacteriol. 193, 5015-5016.

Ludwig, W., Schleifer, K.H., 1994. Bacterial phylogeny based on 16S and 23S ribosomal-RNA sequence-analysis. FEMS Microbiol. Rev. 15, 155-173.

Masse, E., Gottesman, S., 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 99, 4620-4625.

Mellin, J.R., Goswami, S., Grogan, S., Tjaden, B., Genco, C.A., 2007. A novel Fur- and iron-regulated small RNA, NrrF, is required for indirect fur-mediated regulation of the *sdhA* and *sdhC* genes in *Neisseria meningitidis*. J. Bacteriol. 189, 3686-3694.

O'Connor, K.A., Zusman, D.R., 1991a. Behavior of peripheral rods and their role in the life cycle of *Myxococcus xanthus*. J. Bacteriol. 173, 3342-3355.

O'Connor, K.A., Zusman, D.R., 1991b. Development in *Myxococcus xanthus* involves differentiation into two cell types, peripheral rods and spores. J. Bacteriol. 173, 3318-3333.

Piriyapongsa, J., Marino-Ramirez, L., Jordan, I.K., 2007. Origin and evolution of human microRNAs from transposable elements. Genetics 176, 1323-1337.

Reichenbach, H., Dworkin, M., 1992. The Myxobacteria. Balows, A., Truper, G.H., Dworkin, M., Harder, W., Schleifer, K.H. (Eds.), The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications, Springer, New York, pp. 3416-3487.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Rosenbluh, A., Nir, R., Sahar, E., Rosenberg, E., 1989. Cell-density-dependent lysis and sporulation of *Myxococcus xanthus* in agarose microbeads. J. Bacteriol. 171, 4923-4929.

Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M.O., Bartels, D., Bekel, T., Beyer, S., Bode, E., Bode, H.B., Bolten, C.J., Choudhuri, J.V., Doss, S., Elnakady, Y.A.,

Frank, B., Gaigalat, L., Goesmann, A., Groeger, C., Gross, F., Jelsbak, L., Jelsbak, L., Kalinowski, J., Kegler, C., Knauber, T., Konietzny, S., Kopp, M., Krause, L., Krug, D., Linke, B., Mahmud, T., Martinez-Arias, R., McHardy, A.C., Merai, M., Meyer, F., Mormann, S., Munoz-Dorado, J., Perez, J., Pradella, S., Rachid, S., Raddatz, G., Rosenau, F., Rueckert, C., Sasse, F., Scharfe, M., Schuster, S.C., Suen, G., Treuner-Lange, A., Velicer, G.J., Vorholter, F.J., Weissman, K.J., DWelch, R., Wenzel, S.C., Whitworth, D.E., Wilhelm, S., Wittmann, C., Blocker, H., Puhler, A., Mueller, R., 2007. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. Nat. Biotechnol. 25, 1281-1289.

Shimkets, L., Dworkin, M., Reichenbach, H., 2006. The Myxobacteria. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackenbrandt, E. (Eds.), The Prokaryotes, Springer, New York, NY.

Singer, M., Kaiser, D., 1995. Ectopic production of guanosine penta- and tetraphosphate can initiate early developmental gene expression in *Myxococcus xanthus*. Gene. Dev. 9, 1633-1644.

Sproer, C., Reichenbach, H., Stackebrandt, E., 1999. The correlation between morphological and phylogenetic classification of myxobacteria. Int. J. Syst. Bacteriol. 49 (Pt 3), 1255-1262.

Stefani, G., Slack, F.J., 2008. Small non-coding RNAs in animal development. Nat. Rev. Mol. Cell Biol. 9, 219-230.

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.

Tang, T., Kumar, S., Shen, Y., Lu, J., Wu, M.L., Shi, S., Li, W.H., Wu, C.I., 2010. Adverse interactions between microRNAs and target genes from different species. Proc. Natl. Acad. Sci. USA 107, 12935-12940.

Thomas, S.H., Wagner, R.D., Arakaki, A.K., Skolnick, J., Kirby, J.R., Shimkets, L.J., Sanford, R.A., Loffler, F.E., 2008. The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. PLoS ONE 3, e2103.

Vanderpool, C.K., Gottesman, S., 2004. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. Mol. Microbiol. 54, 1076-1089.

Voinnet, O., 2009. Origin, biogenesis, and activity of plant microRNAs. Cell 136, 669-687.

Waters, L.S., Storz, G., 2009. Regulatory RNAs in bacteria. Cell 136, 615-628.

Wilderman, P.J., Sowa, N.A., FitzGerald, D.J., FitzGerald, P.C., Gottesman, S., Ochsner, U.A., Vasil, M.L., 2004. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. Proc. Natl. Acad. Sci. USA 101, 9792-9797.

Wireman, J.W., Dworkin, M., 1977. Developmentally induced autolysis during fruiting body formation by *Myxococcus xanthus*. J. Bacteriol. 129, 798-802.

Wu, S.S., Kaiser, D., 1995. Genetic and functional evidence that Type IV pili are required for social gliding motility in *Myxococcus xanthus*. Mol. Microbiol. 18, 547-558.

Yu, Y.T., Yuan, X., Velicer, G.J., 2010. Adaptive evolution of an sRNA that controls *Myxococcus* development. Science 328, 993.

Yuan, Z.D., Sun, X.A., Liu, H.D., Xie, J.M., 2011. MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. PLoS ONE 6, e17666.

Zhang, Y.Q., Li, Y.Z., Wang, B., Wu, Z.H., Zhang, C.Y., Gong, X., Qiu, Z.J., Zhang, Y., 2005. Characteristics and living patterns of marine myxobacterial isolates. Appl. Environ. Microbiol. 71, 3331-3336.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406-3415.

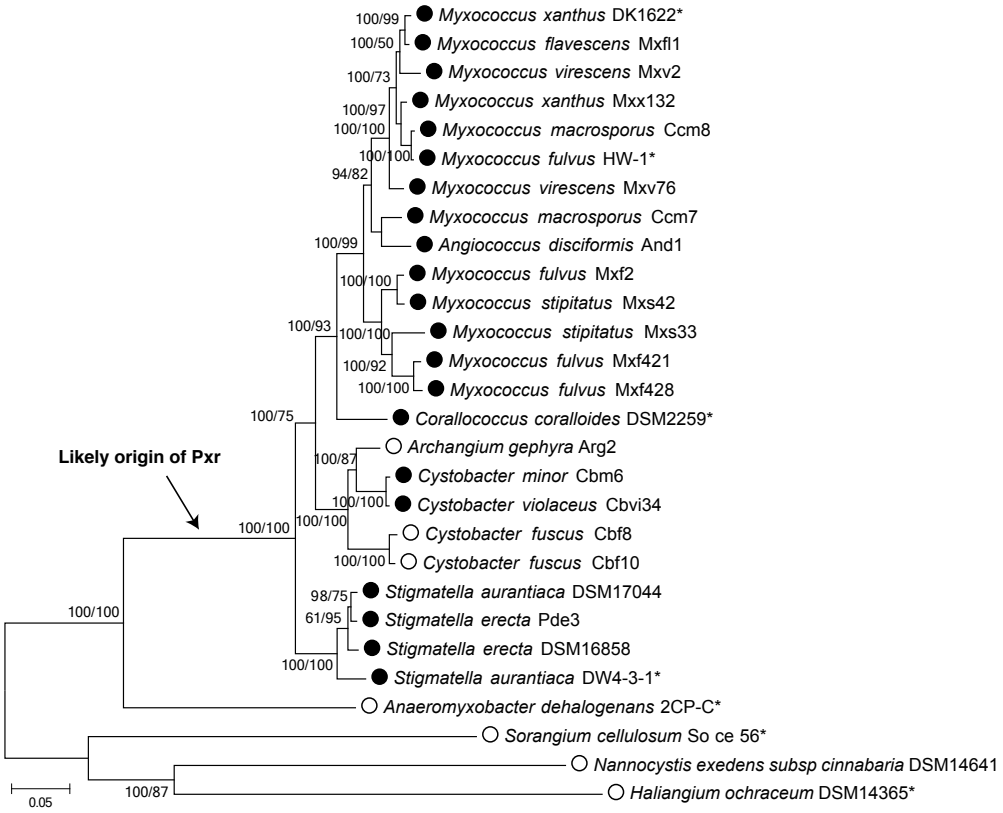


Figure 1.1 Phylogeny of the myxobacteria and Pxr distribution in the suborder Cystobacterineae. The myxobacterial phylogeny was reconstructed by Bayesian inferences based on five conserved loci: 16S rRNA, 23S rRNA, *pyrG*, *rpoB* and *pgm*. Maximum likelihood analysis generated the same topology. Posterior probabilities and bootstrap values (based on 1000 replicates) are shown at the nodes (posterior probabilities shown first). The scale bar shows 0.05 substitutions per site. Sequenced myxobacterial genomes are indicated by asterisks. Solid circles indicate that Pxr was detected in the respective species whereas empty circles indicate that Pxr not detected. Black arrow indicates the branch where Pxr potentially originated.

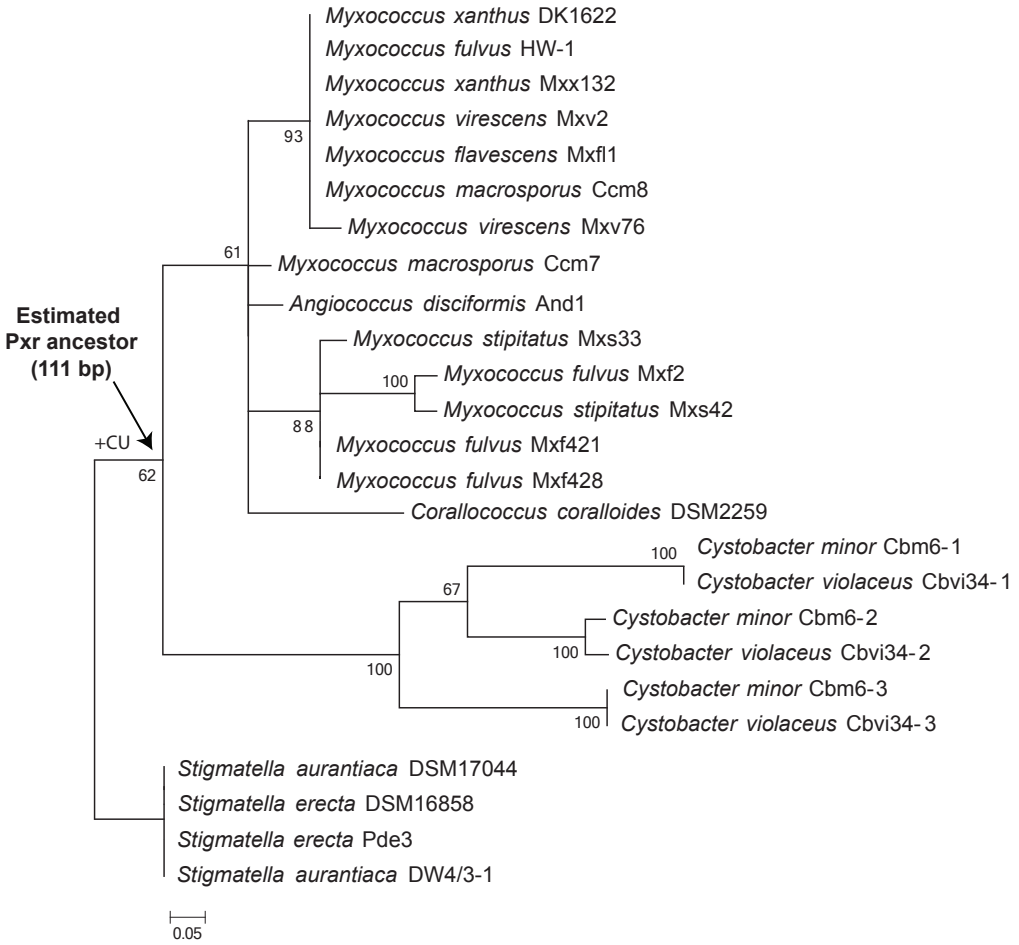


Figure 1.2 Bayesian gene tree of *pxr* based on the 113-nt alignment. Posterior probabilities are shown at the nodes. The *Stigmatella* allele was used as outgroup to root the tree and to construct the Pxr ancestor at the internal node indicated by the black arrow. The inferred ancestor of the large clade including *Myxococcus* and *Cystobacter* spp. only differs from the *Stigmatella* allele by the presence of two bases (CU) in the third loop that are absent in the *Stigmatella* allele.

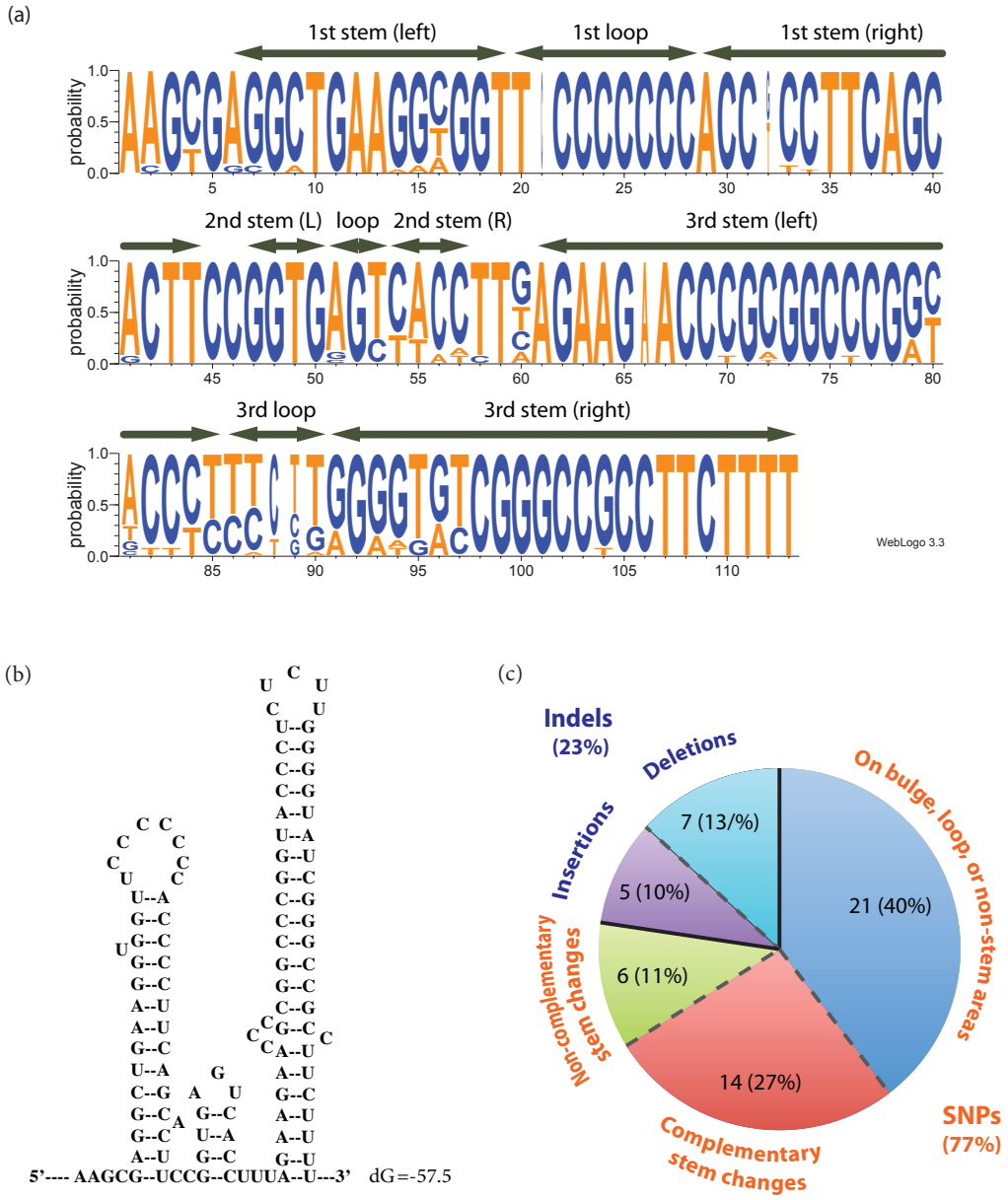


Figure 1.3 (A) WebLogo analysis of the *pxr* homologs. (B) The predicted secondary structure of the inferred Pxr ancestor. (C) Types of nucleotide changes that have occurred on the Pxr stem-loop structure in comparison to the inferred ancestor. Orange text indicates single nucleotide substitutions (SNPs) and dark blue text indicates insertions and deletions (Indels). SNPs are further distinguished to reflect nucleotide changes on different parts of the stem-loop structure.

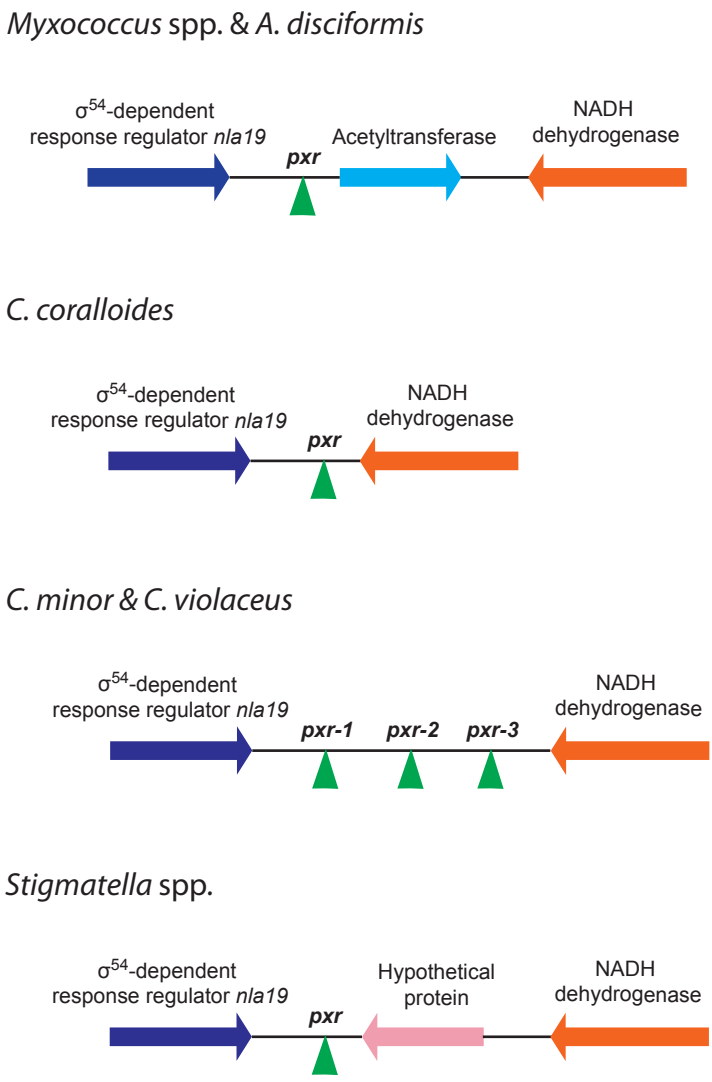


Figure 1.4 *pxr* gene neighborhoods in different myxobacterial species. Gene and intergenic region lengths are not drawn to scale.

Table S1. Primer Sequences

	Forward primer sequence	Reverse primer sequence
16S rRNA	AGAGTTTGATCCTGGCTCAG	AAGGAGGTGATCCAGCC
23S rRNA	CAGGAGGTTGGCTTAGAAGCAGCCA	TTAGATGCTTTCAGCGGTTATCC
pyrG	GAYCCSTACATCAAYGTSGAY	GTGCTGSGTGGGCTTSGTCTT
rpoB	GCGATCAAGGAGCGCATGAG	CCACGGCATGAACGCGAC
pgm	CATCTCSCACGCSATCCTC	AAGCTCTCCGCGTAGATYTTGTAGA
pxr intergenic	AAGATGAAGCGGCTCAACCTCC	TAYTACTACCTGCGCGT

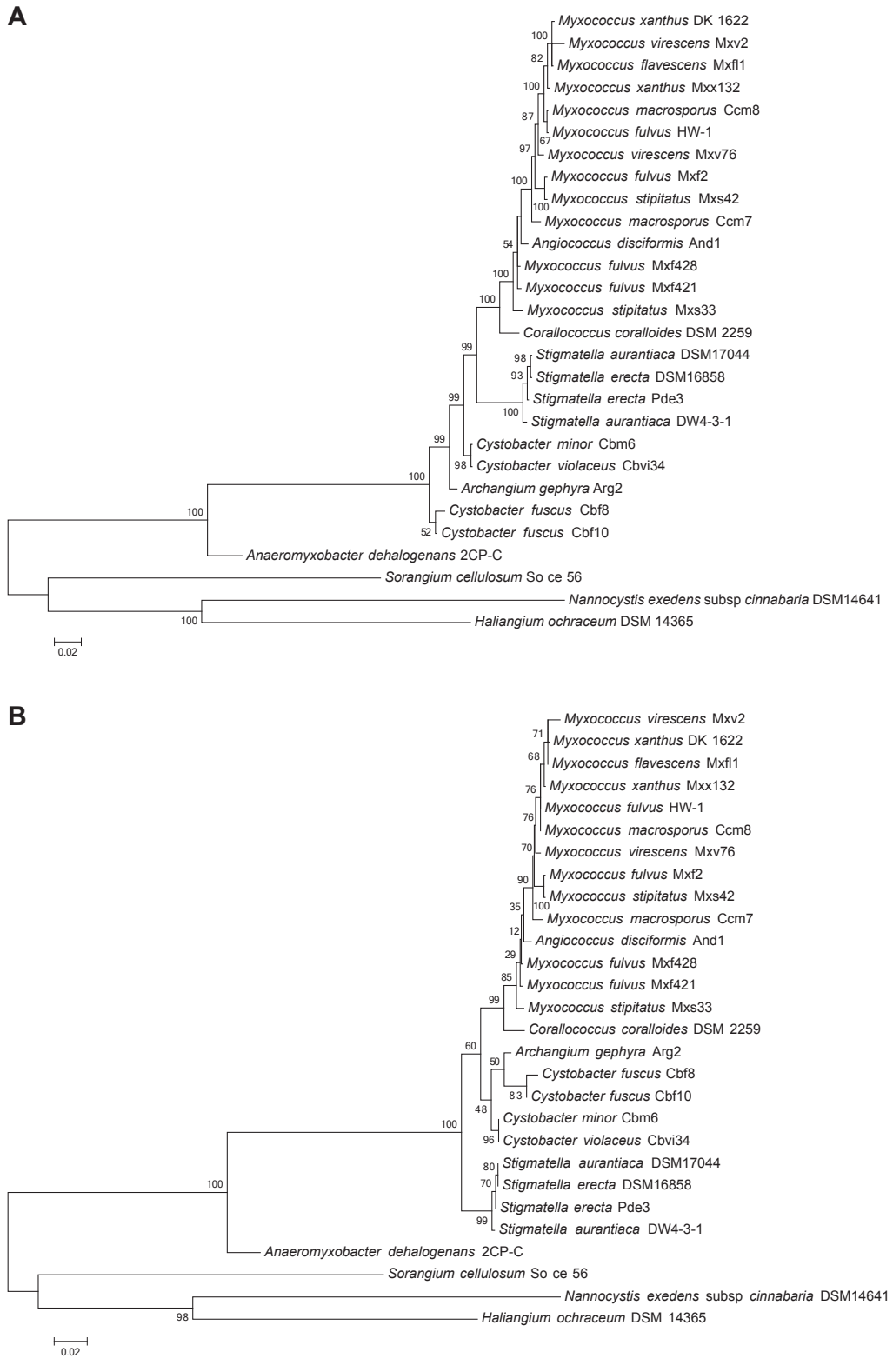


Figure S1. Gene trees reconstructed by Bayesian inference (A) and maximum likelihood analysis (B) of the 16S rRNA gene in the myxobacteria.

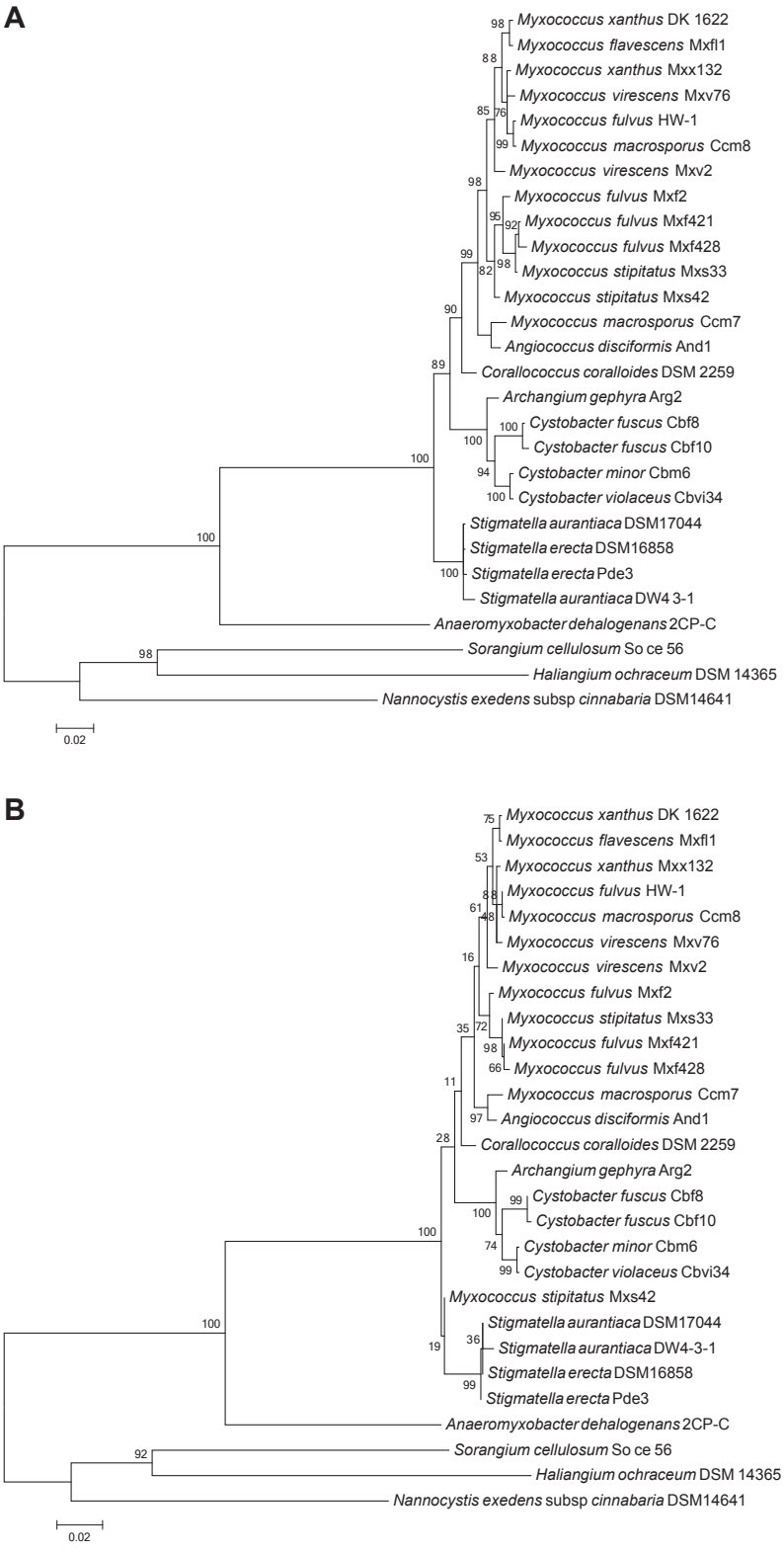


Figure S2. Gene trees reconstructed by Bayesian inference (A) and maximum likelihood analysis (B) of the 23S rRNA gene in the myxobacteria.

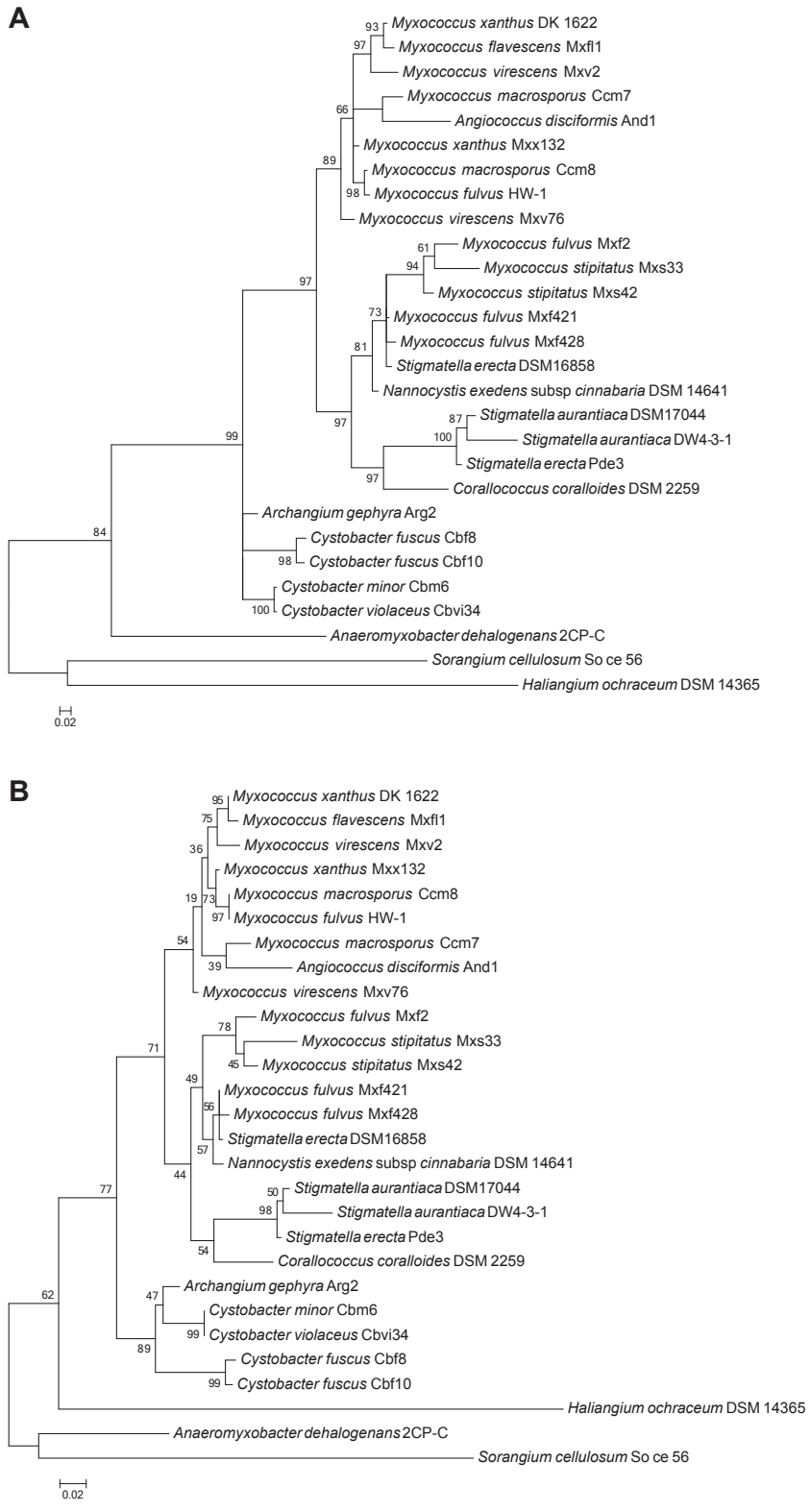


Figure S3. Gene trees reconstructed by Bayesian inference (A) and maximum likelihood analysis (B) of the *pyrG* gene in the myxobacteria.

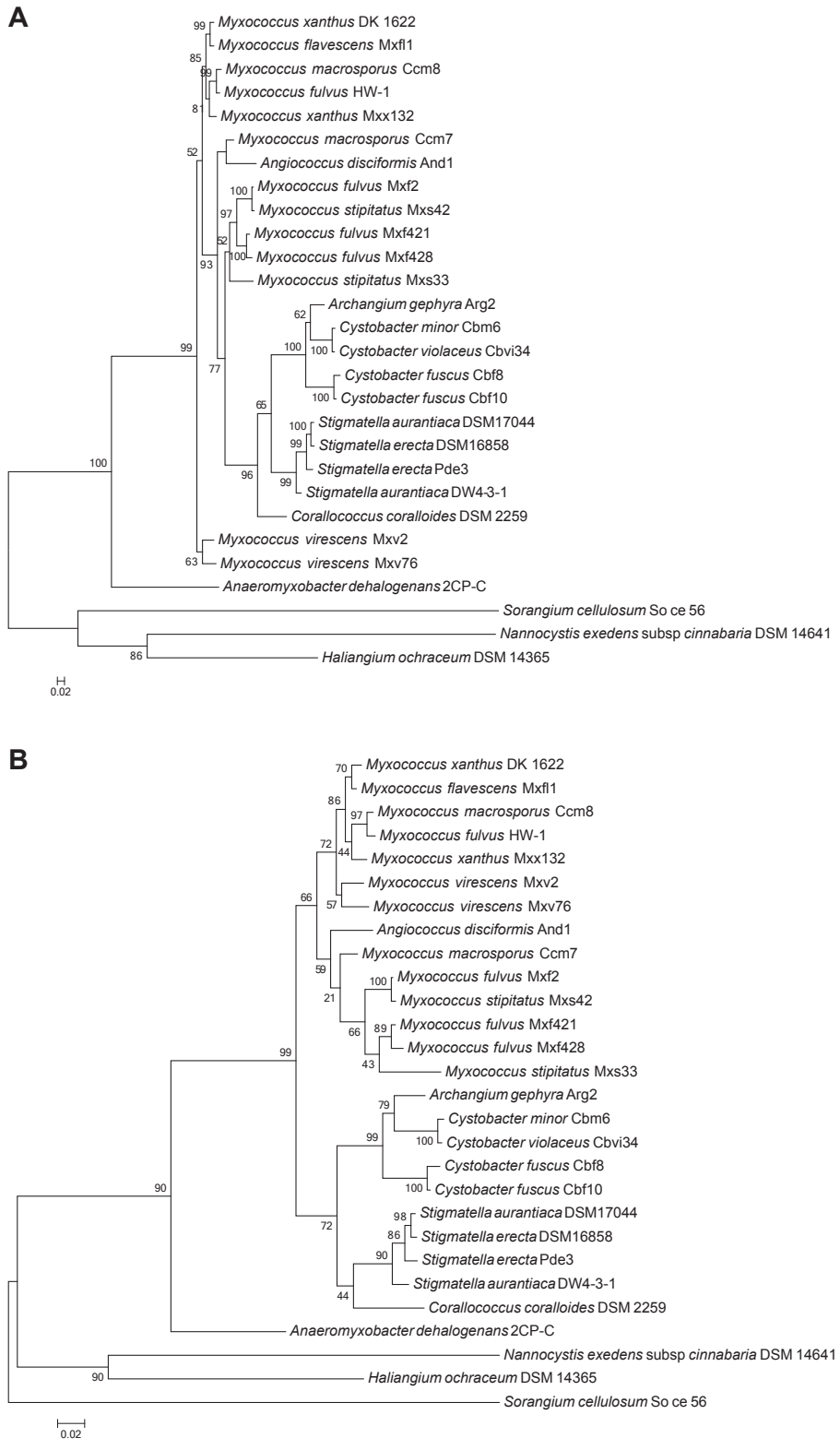


Figure S4. Gene trees reconstructed by Bayesian inference (A) and maximum likelihood analysis (B) of the *rpoB* gene in the myxobacteria.

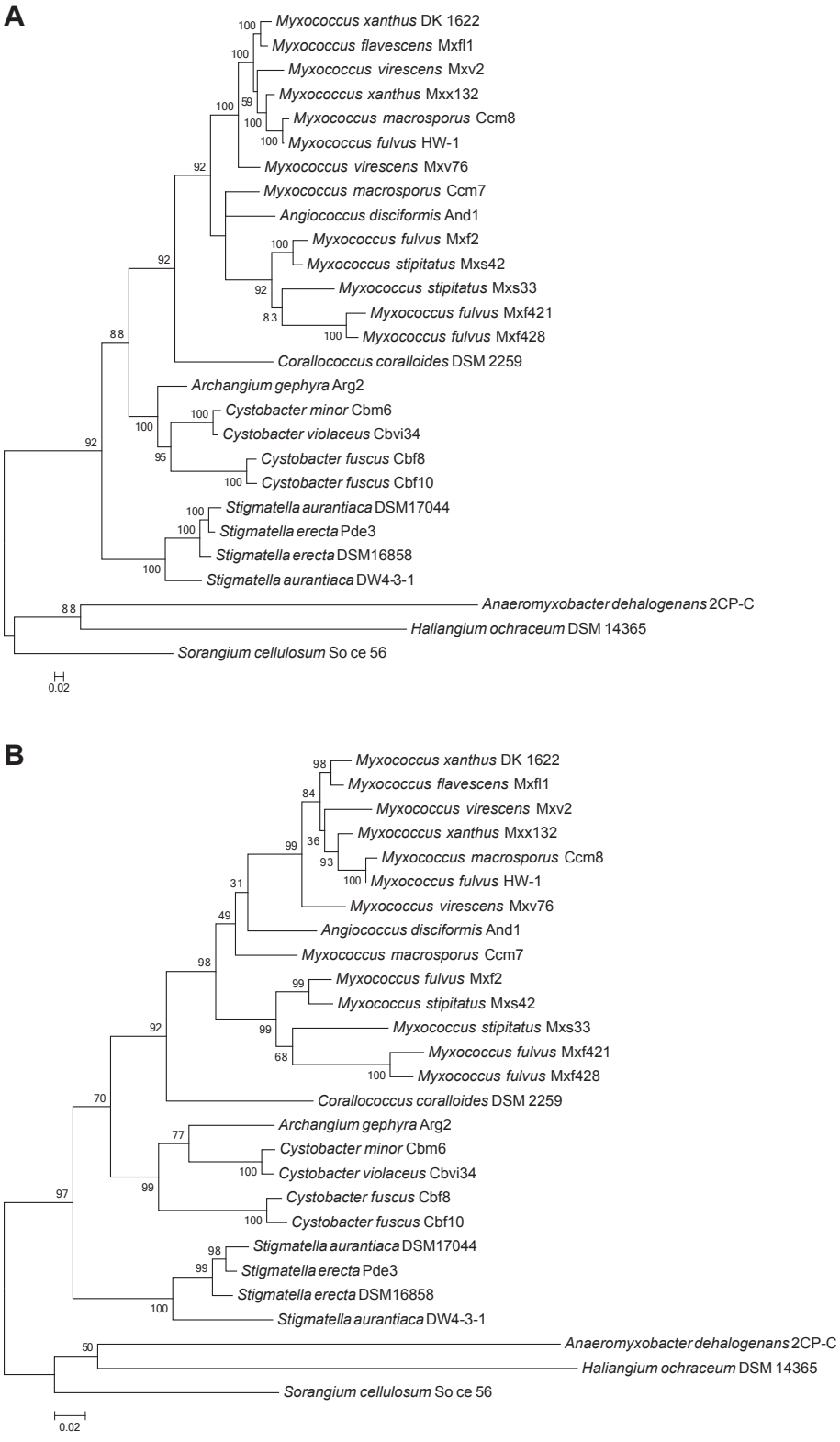


Figure S5. Gene trees reconstructed by Bayesian inference (A) and maximum likelihood analysis (B) of the *pgm* gene in the myxobacteria.

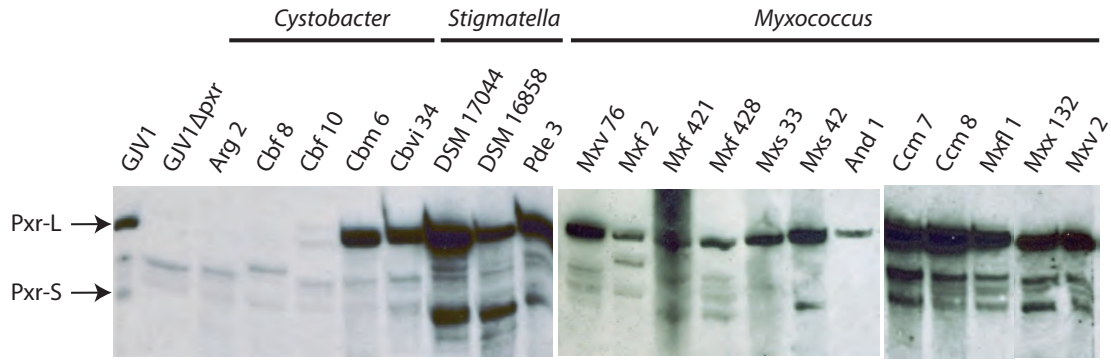


Figure S6. Northern hybridization of sRNAs isolated from 17 myxobacterial strains probed for Pxr expression in three genera as indicated. GJV1 is a lab descendant of the reference *M. xanthus* strain DK1622 (Velicer et al., 2006). Pxr is produced in two forms (Pxr-L (long) and Pxr-S (short)) in *M. xanthus* (Yu et al., 2010) and most other species, but the short form appears to be absent or greatly reduced in some species in this gel image. The apparent absence of the short form from some species might be due to actual non-production of the short form under any conditions. Alternatively, this pattern might be due to differences in culture state (and concomitant differences in gene expression) at the time RNA was isolated. For example, some species grow only as large clumps in liquid culture, which made it difficult to ascertain whether cultures had transitioned from growth phase to stationary phase at the time of RNA isolation.

References

Velicer, G.J., Raddatz, G., Keller, H., Deiss, S., Lanz, C., Dinkelacker, I., and Schuster, S.C., 2006. Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. Proc. Natl. Acad. Sci. USA 103, 8107-8112.

Yu, Y.T., Yuan, X., and Velicer, G.J., 2010. Adaptive evolution of an sRNA that controls *Myxococcus* development. Science 328, 993.

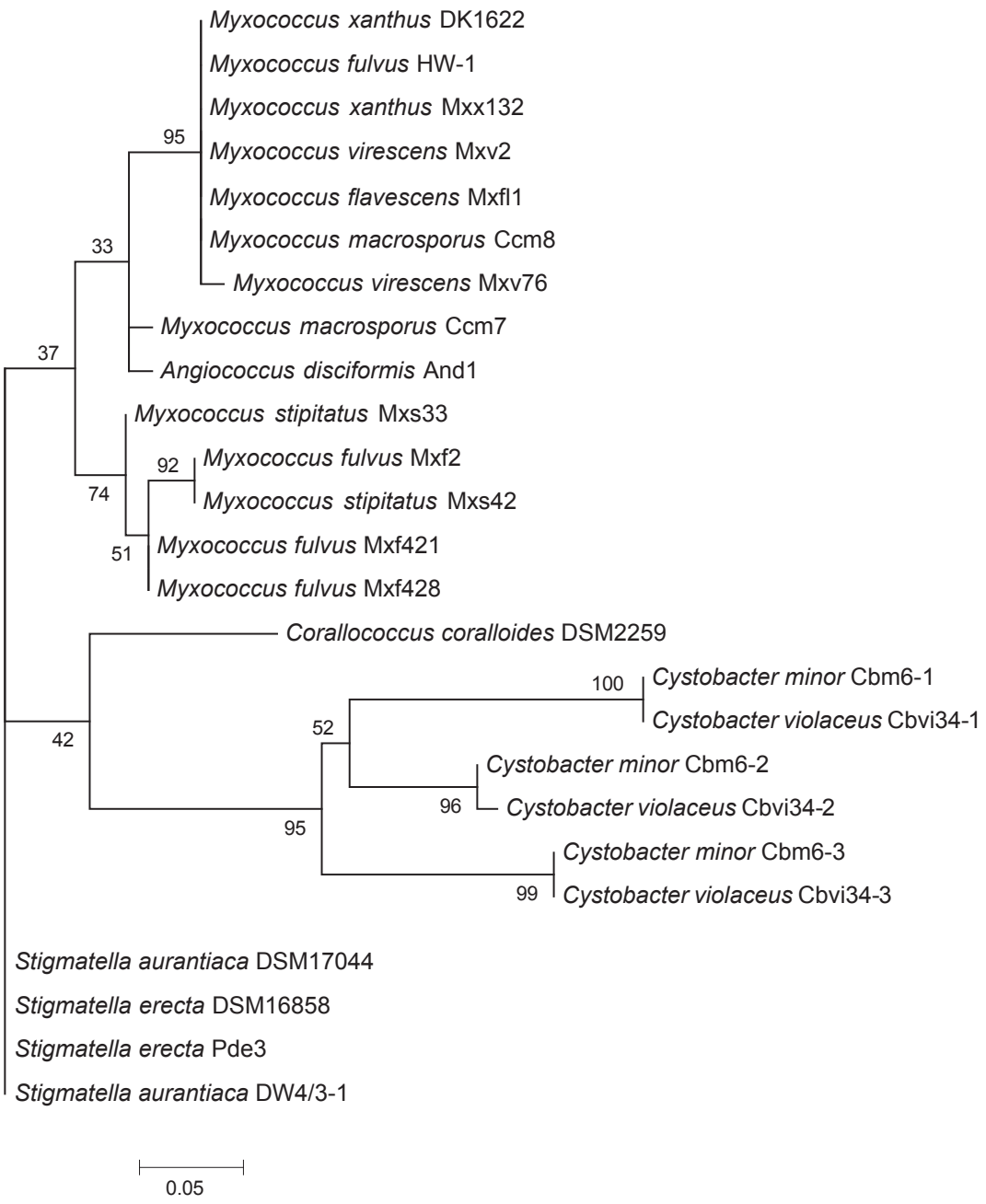


Figure S7. ML gene tree of *pxr* based on the 113-nt sequence alignment. Bootstrap values are indicated at the nodes. The scale bar shows 0.05 substitutions per site.

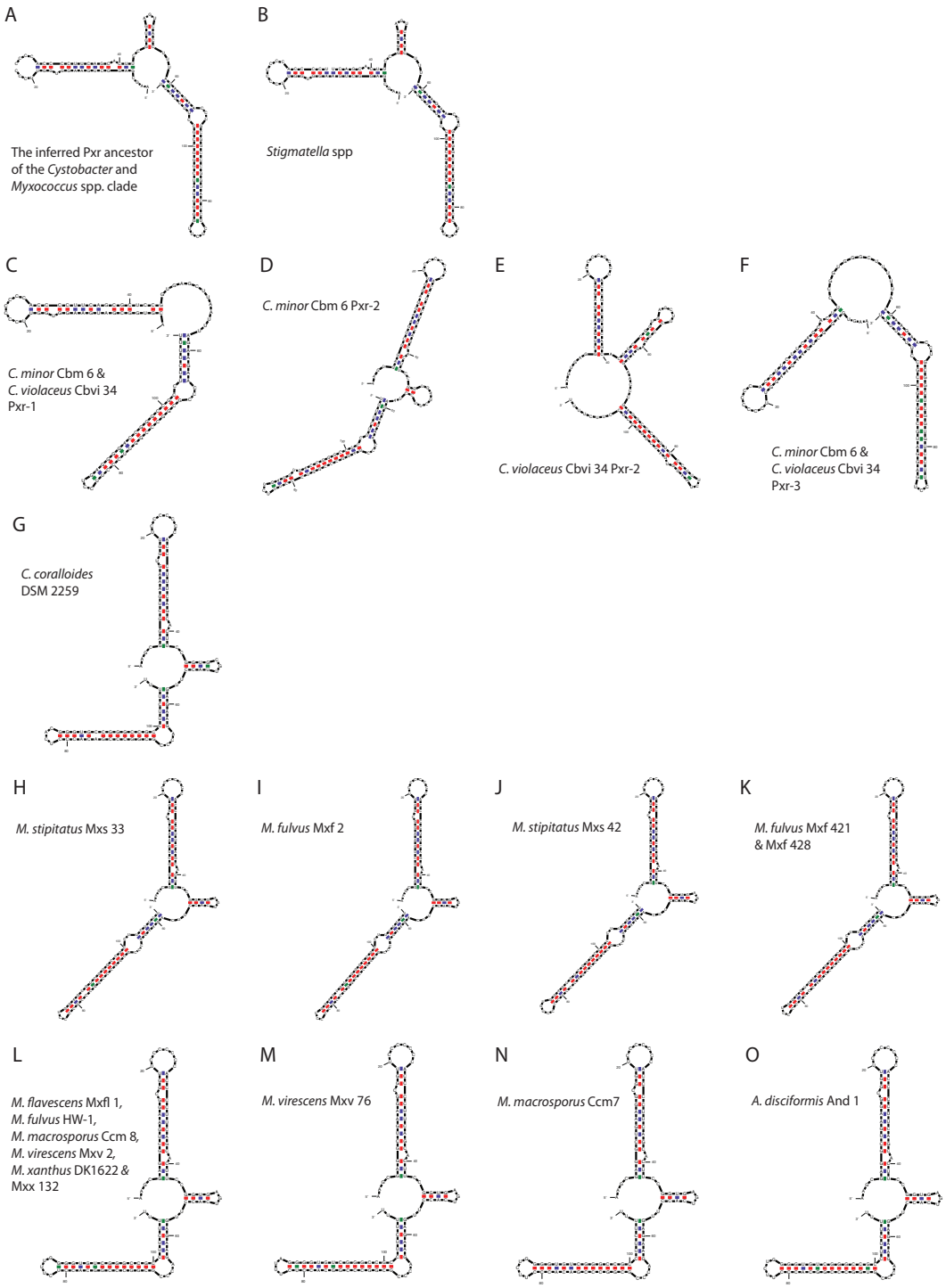


Figure S8. Lowest-energy secondary structures of Pxr homologs based on Mfold. The actual secondary structures are not experimentally known and might be different from the ones shown here.

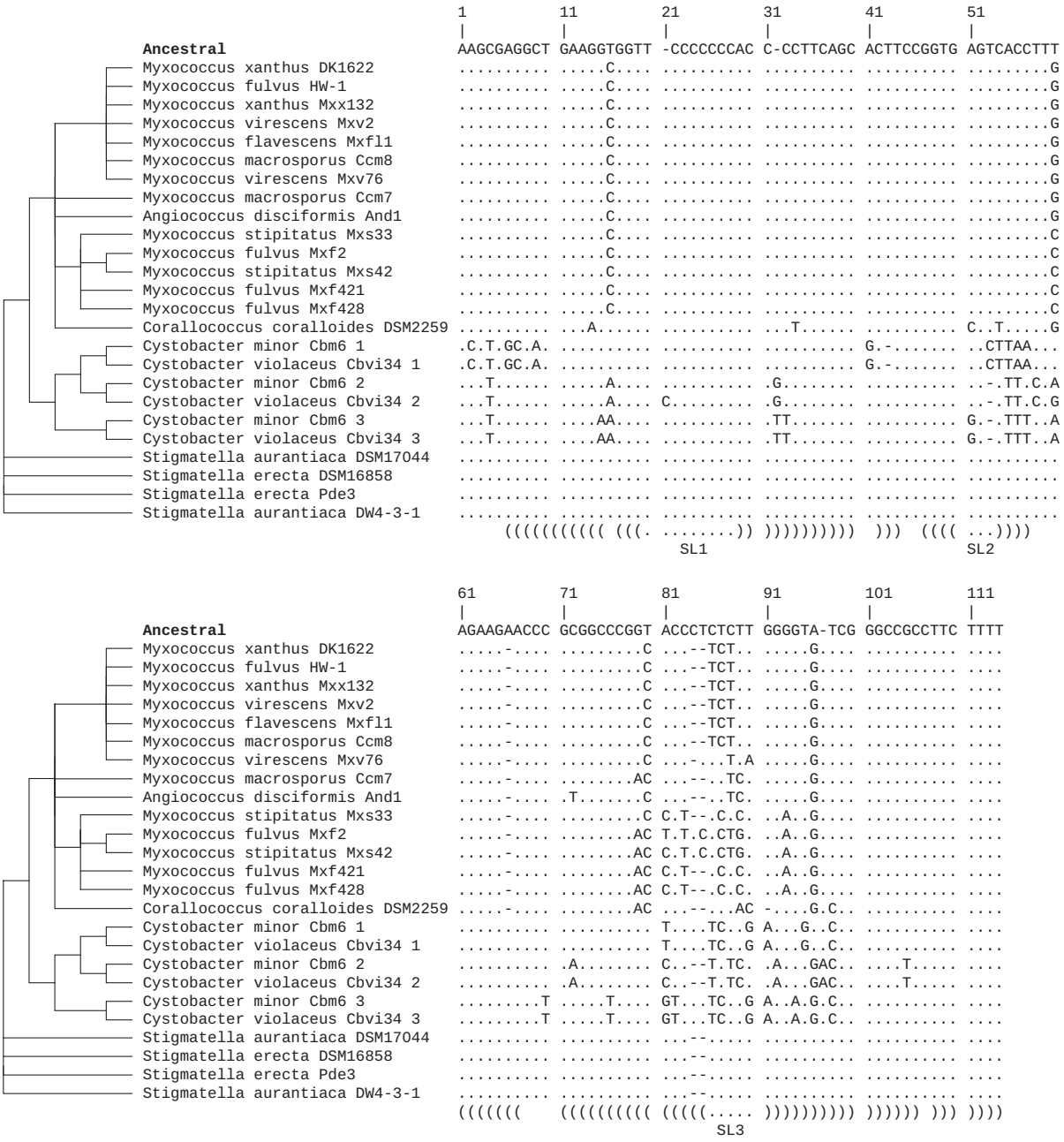


Figure S9. Nucleotide sequences of *pxr* alleles at every site with stem-loops annotated SL1, SL2 and SL3 from 5' to 3'. A dot indicates a perfect match to the inferred ancestor at the internal node shared by the non-*Stigmatella* homologs, whereas letters show nucleotide differences.

CHAPTER II

Functional evolution of Pxr sRNA in the myxobacteria

I-Chen Chen[1,2*], Gregory J. Velicer[1,2] and Yuen-Tsu Nicco Yu[2]

[1] Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

[2] Institute of Integrative Biology (IBZ), ETH Zurich, CH-8092 Zurich, Switzerland

*E-mail: icchen@indiana.edu

Abstract

A rapidly growing body of evidence has shown that non-coding small RNAs (sRNAs) regulate a variety of important biological processes across all life domains, including bacteria. However, little is known about the functional evolution of sRNAs in bacteria, which might occur via changes in structures and stability, interactions with associated regulatory networks or target mRNAs. The sRNA Pxr appears to have a recent evolutionary origin within myxobacteria, and in the model species of myxobacteria, *Myxococcus xanthus*, Pxr functions as a developmental gatekeeper that prevents the initiation of fruiting body development when nutrients are abundant. Here, we introduced the most recent common ancestor of *pxr* and its extant alleles from different species into an *M. xanthus* deletion mutant lacking *pxr* to examine the functional divergence of Pxr within myxobacteria. Our results showed that the regulatory interactions of Pxr observed in *M. xanthus* has been established since this ancestor. Also, all *pxr* alleles from species with only one copy of this gene controlled development in *M. xanthus* in a manner qualitatively similar to that of the native *M. xanthus* allele. Nevertheless, two paralogs found in the genus *Cystobacter* failed to control development. The phenotypic effects of several nucleotide changes conserved and not conserved across *pxr* homologs were also examined. Together, our results illustrate both that Pxr may play a common fundamental role in developmental gene regulation across diverse species of myxobacteria and that the specificity of its function may be evolving in some lineages. Future work will include constructing *pxr* mutants in species other than *M. xanthus* (e.g. *Cystobacter* species) to elucidate the effects of their own *pxr* alleles in their native genomic background.

1. Introduction

In recent years, a rapidly increasing body of studies have shown that small non-coding RNAs are prevalent in regulating gene expression across all domains of life, from bacteria, archaea to eukaryotes (Aravin et al., 2007; Bartel, 2009; Ding, 2010; Gottesman and Storz, 2011; Storz et al., 2011; Babski et al., 2014). One class of the most intensively studied sRNAs are *trans*-encoded sRNAs such as microRNAs (miRNAs) in animals and plants and small regulatory RNAs (sRNAs) in bacteria. These sRNAs modulate gene expression at the post-transcriptional level by base-pairing interactions with mRNA targets, resulting in the degradation of mRNA, inhibiting translation, or both. In bacteria, these small regulators play important roles in a variety of physiological processes, including stress responses (Moller et al., 2002; Masse et al., 2007; Boysen et al., 2010; Durand and Storz, 2010; Yu et al., 2010), outer membrane protein synthesis (Guillier et al., 2006; Vogel and Papenfort, 2006; De Lay and Gottesman, 2009; Overgaard et al., 2009), virulence (Bardill and Hammer, 2012) and social behaviors (Bejerano-Sagie and Xavier, 2007; Yu et al., 2010; Chambers and Sauer, 2013).

Bacterial sRNAs are normally ~100 nucleotides in length, transcribed from the intergenic regions in bacterial genomes with their own promoters and have multiple mRNA targets. They are predicted to fold into stable stem-loop structures including a ρ-independent terminator at the 3' end. The function of an sRNA is determined by at least a seed region of ~6-8 bases that are complementary to a region of their mRNA targets and even a single nucleotide change in this region can abolish sRNA-based regulation (Kawamoto et al., 2006; Hao et al., 2011; Papenfort et al., 2012). Several proteins have been identified to be involved in sRNA activities. For example,

many characterized sRNAs in enteric bacteria require the RNA chaperone Hfq for their interactions with mRNA targets (Vogel and Luisi, 2011), although a number of sRNAs in other species require proteins other than Hfq (Gaballa et al., 2008; Pandey et al., 2011). Ribonucleases also contribute to sRNA activities and levels in various ways. Most base-pairing sRNAs are transcribed as independent units, but some are processed in some manner (Davis and Waldor, 2007; Papenfort et al., 2009; Yu et al., 2010). RNases E and RNase III also facilitate mRNA degradation (Masse et al., 2003; Pfeiffer et al., 2009; Viegas et al., 2011).

Little is known about the evolution of bacterial sRNAs, particularly their functional divergence across species. At the sequence level, the phylogenetic distributions of sRNAs tend to be lineage-specific (Wilderman et al., 2004; Horler and Vanderpool, 2009; Toffano-Nioche et al., 2012; Chen et al., 2014; Peer and Margalit, 2014) and their patterns of presence and absence are primarily due to evolutionary losses rather than lateral gene transfers (Skippington and Ragan, 2012; Chen et al., 2014). For instance, a remarkable fraction of *Escherichia coli* sRNAs has accumulated in the Enterobacteriales order since it split from the rest of the γ-proteobacteria (Peer and Margalit, 2014). One of the most conserved features of sRNA homologous sequences is the base pairing region with targets (Sharma et al., 2007; Horler and Vanderpool, 2009), suggesting strong selective constraints on this region; nevertheless, other sequence regions might also be prominent for function (Hao et al., 2011; Peterman et al., 2014). At the functional level, less is explored about the functional divergence of sRNAs since they originated and whether they coevolved with associated regulatory networks. For example, the SgrS sRNAs in enteric bacteria appear to have coevolved with their associated networks as many of its homologs from different species failed to control target mRNAs in *E. coli* (Wadler and Vanderpool, 2009). Nonetheless,

sgrS is unique in that it also encodes a protein-coding gene *sgrT*, which regulates glucose-phosphate stress via an unknown mechanism (Wadler and Vanderpool, 2007) and might confound the interpretation of the evolution of SgrS.

In this study, we sought to investigate the functional divergence of Pxr sRNA in the myxobacteria. Many species of myxobacteria (Gram-negative, delta-proteobacteria, order Myxococcales), including the best characterized *Myxococcus xanthus*, survive starvation by multicellular development into fruiting bodies and stress-resistant spores (Dawid, 2000). Previous work in *M. xanthus* had established Pxr as a critical developmental gatekeeper that prevents the initiation of fruiting body development until nutrients have been depleted (Yu et al., 2010). Two *pxr*-specific sRNA forms, Pxr-L (long) and Pxr-S (short), are expressed at high amounts during vegetative growth and Pxr-S functions as the primary negative regulator as it is greatly diminished upon starvation, hence allowing development to proceed (Yu et al., 2010). Pxr appears to have a recent evolutionary origin in the lineage basal to the suborder Cystobacterineae inside Myxococcocales (Chen et al., 2014). Most species within this suborder contain a single copy of *pxr* with sequence conservation similar to that of the 23S rRNA gene; nevertheless, tandem paralogs were found in the genus *Cystobacter* and these alleles together evolved much more rapidly (Chen et al., 2014).

Other than identified by traditional genetic selection and screen for mutants defective in developmental gene regulation, Pxr was discovered in an experimental evolution where a developmentally defective *M. xanthus* strain (OC, obligate cheater) evolved into a strain with restored developmental proficiency (PX, phoenix) by deactivating Pxr (Fiegna et al., 2006). OC

is a descendant of the developmentally proficient wild-type strain GJV1 and differs from GJV1 by 14 mutations accumulated during a laboratory evolution in an unstructured environment (Velicer et al. 1998). This cheater strain is defective at development in clonal groups but can exploit GJV1 in chimeric groups to sporulate more efficiently than GJV1. The temporal pattern of Pxr accumulation in OC showed that Pxr-S is constitutively expressed, which in turn down-regulates development even when nutrients are depleted. Accordingly, OC does not form fruiting bodies nor sporulate on nutrient free agar plates. Strain PX emerged from OC after a cheater-induced population crash (Fiegna & Velicer 2003). PX differs from OC by a single C→A mutation in the central position of a seven-base cytosine run in the first loop of Pxr, which confers the phenotype of fruiting body formation and high spore production on nutrient free plates even though its Pxr-S remains present in the cells (Fiegna et al. 2006; Yu et al. 2010).

The facts that OC constitutively expresses Pxr-S and that a single mutation can restore developmental proficiency provide an excellent genetic system to examine the functionalities of *pxr* variants. Alleles of *pxr* proficient at blocking development in the OC background will demonstrate very low spore production, and vice versa. In this study, we first reconstructed the most recent common ancestor of extant *pxr* alleles and introduced it into an OC deletion mutant lacking *pxr* ("OC Δ*pxr*" hereafter) to test whether the regulatory interactions of Pxr observed in *M. xanthus* has been established in this ancestor. Next, we introduced three conserved single-copy *pxr* alleles and four divergent paralogs from different species into OC Δ*pxr* to examine the functional divergence of Pxr. When reconstructing the ancestral allele of *pxr*, we serendipitously obtained several mutant variants with nucleotide changes either conserved or not conserved across *pxr* homologs and we decided to exploit the functional significance of these variants.

These data allow us to examine the functional diversification of Pxr and the coevolution between Pxr and its associated regulatory network within the myxobacteria.

2. Materials and Methods

2.1. Plasmid construction

The ancestral allele of pxr

The likely ancestral sequence of *pxr* was previously inferred based on the sequences of *pxr* homologs identified in *Myxococcus* and *Cystobacter* using the *Stigmatella* allele as outgroup (Chen et al., 2014). The inferred ancestor is hence not ancestral to the *Stigmatella* allele but only differs from it by the presence of two bases (CU) in the third loop. Between the inferred *pxr* ancestor and the wild-type allele in *M. xanthus* GJV1, there are seven base differences (Fig 2.1A). Six of these differences are either nearby or on the third and also last stem-loop of Pxr. Among them three are single base substitutions: a U→G change in the single-stranded region between the second and third stem-loop and a pair of complementary changes on the third stem (U→C combined with A→G). The other three differences are single base deletions: one at the base of the bulge on the left side of the third stem, one nearby the third loop and one in the loop. The difference not around the third stem-loop is a single base change U→C at the bulge on the left side of the first stem.

To reconstruct the ancestral *pxr* allele in the laboratory, we used the genomic DNA of *M. xanthus* GJV1 as template and designed three PCR primers that contain the polymorphisms in

the ancestral allele in order to replace the ones in the *M. xanthus* allele. Two of the primers are forward and reverse sequences at the ancestral *pxr* positions 13-33 and contain the ancestral U on the bulge on the left side of the first stem-loop. The third primer is a 64-nt long primer at the positions 48-111 that contains the rest of six ancestral polymorphisms around the third stem-loop and till the end of the *pxr* coding region. We first PCR-amplified a 462-nt DNA fragment containing a 429-nt fragment preceding *pxr* in conjunction with the nucleotide positions 1-33 in the ancestral *pxr* with the primers GV367 (5' CCC AGG TGG TGG AAG AGG 3') and Node38_Bulge_Rev (5' GGG GGG AAC C**A**C CTT CAG CCT 3') (the bold letter indicates the ancestral nucleotide). Next, we PCR-amplified a 99-nt fragment from the positions 13 to 111 in the *pxr* ancestor using primers Node38_Bulge_For (5' AGG CTG AAG G**T**G GTT CCC CCC 3') and Node38_LongRev (5' AAA AGA AGG CGG CCC GA**T** ACC CCA A**G**A GA**G** GGT **A**CC GGG CCG CGG GT**T** CTT CT**A** AAG GTG ACT C 5'). We subsequently performed a PCR to join the 462-nt and 99-nt fragments together by the overlapping sequences at the positions 13-33 in the *pxr* ancestor and to amplify the joint 540-nt fragment with GV367 and Node38_LongRev. The resulting PCR fragments were cloned into the plasmid pCR®2.1-TOPO® (Invitrogen) that carries a kanamycin-resistance gene. Competent cells of *Escherichia coli* TOP10 were used for plasmid cloning and were grown at 37°C in Luria-Bertani (LB) medium (Sambrook et al., 1989) or on X-Gal/IPTG LB hard (1.5%) agar plates to screen for colonies with PCR fragment inserts. The sequences and orientation of the inserts were confirmed by sequencing.

Alleles of pxr from different myxobacterial species

Seven previously identified alleles of *pxr* were examined in this study. These alleles are located downstream of the σ^{54}-dependent response regulator *nla19* (predicted gene *Mxan_1078* in the *M. xanthus* genome; (Goldman et al., 2006)), appear to be transcribed from an upstream σ^{54} promoter and are predicted to fold into stable multi-stem-loop structures (Chen et al., 2014). Three of them are single-copy *pxr* alleles from phylogenetically diverse species (two from *M. stipitatus* Mxs 33 and 42 and one from *Stigmatella aurantiaca* DSM17044) and four are paralogs found in *Cystobacter minor* Cbm 6 and *C. violaceus* Cbvi 34. Both *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 carry three *pxr* alleles in tandem (*pxr-1*, *pxr-2* and *pxr-3* from 5' to 3') downstream of *nla19*. The divergence among these paralogs within each *Cystobacter* species is greater than the divergence between species. For instance, the pairwise distances between *pxr-1* and *pxr-2*, *pxr-2* and *pxr-3*, and *pxr-1* and *pxr-3* in *C. minor* Cbm are 15%, 17% and 22%, respectively. Nevertheless, the *pxr-1* and *pxr-3* alleles in *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 are identical to each other and their *pxr-2* alleles only differ from each other by two nucleotides. It is likely that those *pxr* alleles first duplicated and diverged in the common ancestor shared by *C. minor* Cbm 6 and *C. violaceus* Cbvi 34. Following the divergence of the two lineages, two more nucleotide differences accumulated in their *pxr-2* alleles. Because *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 share the same *pxr-1* and *pxr-3* alleles, we examined four instead of six alleles in *Cystobacter* in this study.

The *pxr* alleles examined were aligned with MUSCLE implemented in MEGA version 5.0 (Tamura et al., 2011) and their phylogenetic relationships were reconstructed using Bayesian Inference (BI) in MrBayes 3.1.2 for 10,000 generations with two parallel independent searches,

using a GTR + invgamma model (Ronquist and Huelsenbeck, 2003). The phylogenetic relationship of both single-copy *pxr* alleles and the paralogs is shown in Figure 2.3A.

To introduce these different *pxr* alleles into *M. xanthus*, we first PCR-amplified the 429-nt DNA fragment preceding *pxr* in *M. xanthus* for homologous recombination. This region contains the last 167-nt of *nla19* and the 262-nt intergenic region between *nla19* and *pxr*. We also PCR-amplified each *pxr* allele using the genomic DNA extracted from their host strains. Primer sequences used for PCR amplification were listed in Table 2.1. Next, we ligated the 429-nt PCR products with each *pxr* allele using T4 DNA ligase (Thermo Scientific). A subsequent PCR was performed to amplify the ligated fragment using the forward primer complementary to the 5' region of the 429-nt fragment and the reverse primer complementary to the 3' region of each *pxr* allele. The resulting PCR products were cloned into the plasmid pCR®2.1-TOPO® as described in the previous paragraph. The sequences and orientation of the inserts were confirmed by sequencing.

Mutant variants of the ancestral pxr allele

When reconstructing the ancestral allele of *pxr*, we obtained seven variants with serendipitous mutations that are either in the conserved or non-conserved sites across *pxr* homologs and we decided to test their functional significance. Each variant folded in the stem-loop structure is illustrated in Figure 2.5A. For three of the mutant alleles, each of them contains a single-base mutation in the conserved sites: a cytosine deletion in the first loop (the original seven-cytosine stretch was shortened into a six-cytosine stretch), a cytosine insertion between the positions 75 and 76 on the third stem, and a G→A substitution at the position 102 on the third stem, which

weakens the stability of the secondary structure of Pxr. For two of the mutant alleles, each of them contains a single-base mutation in the non-conserved sites: an A→U substitution at the position 49 in the second loop and a uracil deletion at the position 87 in the third loop. We also obtained two variants truncated from the 3' ends from the positions 96 and 103 and differed from the inferred ancestor with the U→C change on the bulge on the left side of the first stem-loop. The cloning procedure was the same as described in the earlier paragraph.

2.2. M. xanthus strain construction

A list of *M. xanthus* strains used in this study is provided in Table 2.2. All strains constructed for this study were derived from the strain OC Δ*pxr* and generated with plasmid integration, which was achieved by electroporation of OC Δ*pxr* competent cells with plasmids that carry the 429-nt homologous region preceding *pxr* in conjunction with different versions of *pxr*. To perform electroporation, overnight cultures of OC Δ*pxr* grown in CTT liquid medium (Hodgkin and Kaiser, 1977) were harvested at exponential-growth phase by centrifugation at room temperature. The cells were washed five times with and resuspended in double distilled water. Seventy-five microliters of the cell suspension and 3 μl of the plasmid DNA were transferred to electroporation cuvette. Electroporation was performed with a Bio-Rad Gene Pulser apparatus set at 400 Ω, 25 μF and 0.65 kV. The cells were then transferred to an Erlenmeyer flask containing 3 ml of CTT liquid and incubated overnight at 32°C with constant shaking at 300 rpm. Samples were diluted into CTT soft (0.5%) agar containing 40 μg kanamycin/ml and incubated at 32°C with 90% rH for a week until colonies became visible. For each constructed strain, two to three clones were isolated whenever possible.

Two control strains "Pxr_{WT+}" and "Pxr_{WT-}" used in this study were obtained from Yuen-Tsu Nicco Yu (personal communication). Both strains are integrated with the plasmids pCR2.1 but differ in that "Pxr_{WT+}" contains the 429-nt intergenic fragment followed by the *M. xanthus* allele of *pxr* while "Pxr_{WT-}" only contains the 429-nt fragment.

2.3. Developmental assay

M. xanthus strains were grown in CTT liquid at 32°C with constant shaking at 300 rpm. To perform the developmental assays, cultures in exponential-growth phase were centrifuged at room temperature and resuspended in TPM liquid (Bretscher and Kaiser, 1978) to a density of 5 x 10^9 cells per ml. For each strain, 50 µl of resuspended cells was spotted on the center of nutrient free TPM hard (1.5%) agar plates and incubated at 32°C with 90% rH. After three days, cells were harvested with a scalpel blade, transferred into 1 ml of double distilled water and heated at 50°C for two hours to select for heat-resistant spores. Samples were then sonicated with a microtip to disperse spores and diluted into CTT soft (0.5%) agar plates. After seven days, visible colonies were counted to estimate the total number of spores produced during development. In cases where no colonies grew at our lowest dilution factor (10^{-1}), a value of ten spores per replicate was entered for data analysis, providing conservatively low estimates of spore production in these cases. In any given assay, two to three clones of identical constructs were tested whenever possible. The same experimental procedure was repeated for at least three times with the same set of strains.

2.4. Expression pattern and stability of Pxr sRNA

We detected the expression of Pxr sRNA by Northern blotting. The total RNA of each strain was extracted from vegetative cultures growing in CTT liquid and small-sized RNAs were enriched with the mirVana miRNA isolation kit (Ambion). RNA concentration was measured with a Nanodrop spectrophotometer and equal amounts of RNA were electrophoresed in 10% SequaGel (National Diagnostics) and electro-transferred onto a BrightStar®-Plus positively charged nylon membrane (Ambion). After UV cross-linking, the membrane was pre-hybridized in 4 ml UltraHyb-Oligo buffer for 30 minutes and subsequently hybridized in the same solution containing 100 pmol 3'Biotin-TEG-*pxr* oligo probe (5'-ACC GGA AGT GCT GAA GGG GTG GGG GG-3') (Sigma) overnight. Pxr sRNA was detected with BrightStar® Biodetect non-isotopic kit (Ambion). We also estimated the stabilities of Pxr variants in one stem-loop structure by computing the self-folding free energy ΔG of their stem-loops using Mfold (Zuker, 2003).

3. Results

We will first describe the results of the *pxr* ancestor and the alleles from different species to examine the functional divergence of Pxr in the myxobacteria since its origin. We will next discuss the results of the serendipitously obtained variants of the *pxr* ancestor to investigate the functional significance of the mutations in these variants. As described in the Materials and Methods, for each constructed strain, we examined two to three clones for their developmental phenotypes whenever possible. Because there were no qualitative differences among clones of identical constructs within and among replicates (with one exception), we used within-replicate

means of independent clones as data points. For one of the constructs with a serendipitously obtained variant (Pxr_{Anc4}, detailed description in the later paragraphs), because the phenotype of its clones tended to vary both within and among replicates, we decided to present our results with among-replicate means of each independent clone. For clarity, we will refer to each *pxr* allele with their myxobacterial strain origin (e.g. "pxr_{Mxs33}" for the *pxr* allele from *M. stipitatus* Mxs 33) or according to their status (e.g. "pxr_{Anc}" refers to the ancestral allele and "pxr_{Anc1}" refers to the first mutant variant listed in Table 2.2). For the two isogenic constructs that differ only in the presence or absence of the *M. xanthus pxr* allele, we will refer to them as "pxr_{WT+}" and "pxr_{WT-}". For the three unmarked control strains GJV1, OC and OC Δ*pxr*, we retain their original designations.

3.1. The ancestral allele of pxr

We examined the proficiency of the Pxr ancestor at controlling development by measuring spore production on nutrient free TPM agar plates. Pxr_{Anc} fully blocked development like Pxr_{WT+} as no spores at lower limit of detection were obtained from the OC Δ*pxr* strain carrying it (Fig 2.1B), indicating that the seven base differences between the ancestor and the *M. xanthus* allele do not affect the developmental phenotype assayed here. The Northern blot showed that pxr_{Anc} was expressed into both Pxr-L and Pxr-S at similar levels and sizes as the *M. xanthus* allele (Fig 2.2), confirming the developmental phenotype we observed.

3.2. Alleles of pxr from different myxobacterial species

We next examined whether the *pxr* alleles from different species of myxobacteria have functionally diverged from the ancestor. Despite continuing efforts to clone accurate sequences of these *pxr* alleles, the sequences we cloned tended to lack the first adenine. Because lacking the first adenine appeared in all the *pxr* alleles we attempted to clone, we decided to proceed with such alleles.

The developmental assays showed that, all the single-copy Pxr homologs (Pxr_{Mxs33}, Pxr_{Mxs42} and the distantly related Pxr_{Sga}) blocked *M. xanthus* development as very low numbers of spores were produced from the OC Δ*pxr* strains carrying them (Fig 2.3B). This also demonstrates that lacking the first adenine in those *pxr* alleles did not appear to affect their functions. In contrast, we observed functional variation among Pxr paralogs in *Cystobacter* (Fig 2.3B). Although both Pxr-2 sRNAs from *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 blocked development like other single-copy homologs, their Pxr-1 and Pxr-3 sRNAs failed to do so as the spore production of the OC Δ*pxr* strains carrying them was as robust as OC Δ*pxr* and the isogenic construct lacking the *M. xanthus* wild-type allele.

Several mechanisms might contribute to the observations that Pxr-1 and Pxr-3 paralogs in *Cystobacter* could not control development in *M. xanthus*. First, the sequences of these *pxr* alleles might affect their auto-regulation, such as increasing or decreasing the amount of transcripts through a positive or negative feedback loop. It is also possible that the *pxr* sequences might affect the stability of the sRNA molecules transcribed from them. If the amount of Pxr-L is reduced or not stable, therefore lowering the amount of Pxr-S, development might not be blocked. Second, the nucleotide changes in the *pxr* sequences might affect how the

endoribonucleases in *M. xanthus* process these alleles. Accordingly, Pxr-L is produced but not Pxr-S, or that Pxr-S is produced in lower amounts or different sizes due to inefficient or inaccurate processing. In addition, other components in the Pxr regulatory pathway, such as RNA chaperones, might also affect the stability of Pxr-L, Pxr-S or both. Third, the nucleotide changes in the *pxr* sequences might be in the seed region and therefore change their interactions with mRNA targets (e.g. weakened hybridization with targets). In this scenario, both Pxr-L and Pxr-S are expressed but development is not down-regulated.

We used Northern blotting to examine the expression patters for each *pxr* allele (Fig 2.4). All three single-copy *pxr* alleles, as well as the *pxr-2* alleles in *C. minor* Cbm 6 and *C. violaceus* Cbvi 34, were expressed into Pxr-L and Pxr-S at similar amounts and sizes as the *M. xanthus* allele, supporting the developmental phenotypes we observed from the strains carrying them. For the *pxr-1* and *pxr-3* alleles that did not block development, their Pxr-L was produced in similar sizes as the *M. xanthus* allele but at much lower amounts, and the Pxr-S from *pxr-1* was not detected and the one from *pxr-3* barely detected.

We calculated RNA structural stabilities by self-folding free energy ΔG (the stabler the RNA structure, the lower the ΔG) to explore the possibility that the *pxr* sequences of *pxr-1* and *pxr-3* in *Cystobacter* affect the stability of sRNA molecules transcribed from them. Compared to the stability of the *M. xanthus* Pxr sRNA (ΔG_{WT} = -58.3), the ones of the Pxr ancestor and all the naturally occurring variants do not appear to be less stable (ranging from -58.9 to -52.2), including the Pxr-1 and Pxr-3 sRNAs in *Cystobacter* (ΔG = -58.9 and -54, respectively).

Therefore, it is unlikely that the reduced expression of Pxr-1 and Pxr-3 sRNAs in the *M. xanthus* cells is caused by their structural stabilities alone.

3.3. Mutant variants of the ancestral pxr allele

In contrast to the results of the inferred ancestor and *pxr* alleles from different species where development was either fully blocked or not, the effects of serendipitous mutations on Pxr function are more variable. Both Pxr_{Anc1} (a cytosine insertion on the third stem) and Pxr_{Anc5} (an A→U substitution in the second loop) controlled fruiting body development in a manner qualitatively similar to Pxr_{WT+} and Pxr_{Anc} (Fig 2.5B). Pxr_{Anc2} (a uracil deletion in the third loop) also controlled development although the results were slightly noisy. Pxr_{Anc4} (a G→A substitution on the third stem that affects the stem structure and creates an internal loop inside the stem) represented an interesting case as its ability to control development tended to very among replicates and this was consistent with the two individual clones we tested. The averaged spore productions of the two clones are 3.51 and 2.95 on the logarithmic scale and the standard deviations are 1.77 and 1.60, respectively (for Pxr_{Anc}, its averaged spore production is 1.79 and its standard deviation 0.49; for Pxr_{WT+} here, no spores were detected in all replicates). This is unlikely due to the conditions that varied among replicates as the phenotypes of the two clones tended to vary within replicates too. Finally, Pxr_{Anc3} (a cytosine deletion in the first loop), Pxr_{Anc6} (the U→C change on the bulge on the left side of the first stem and the 9-bases truncation at 3' end) and Pxr_{Anc7} (the same U→C change and the 16-bases truncation at 3' end) were all deficient at controlling development in a manner similar to OC Δ*pxr* and Pxr_{WT-} strains. We reason that the deficiencies of Pxr_{Anc6} and Pxr_{Anc7} were caused by the truncations at 3' ends rather than the

U→C change on the bulge on the left side of the first stem as the single-copy Pxr_{Mxs33} and Pxr_{Mxs42} from *M. stipitatus* also contain the U→C change but still fully blocked development. Alternatively, but more unlikely, such deficiencies might be due to the interactions between the U→C change and the truncations at 3' ends.

We also examined how these serendipitous mutations affect the expression patterns of Pxr. For the variants with the mutations that did not appear to have phenotypic effects (Pxr_{Anc1}, Pxr_{Anc2} and Pxr_{Anc5}), both Pxr-L and Pxr-S were detected at similar levels and sizes as the *M. xanthus* allele (Fig 2.6). Pxr_{Anc4}, the allele that generated increased phenotypic variation, was also expressed into stable forms of Pxr-L and Pxr-S. For the alleles that were deficient at controlling development, various expression patterns were observed. Pxr_{Anc3}, the allele with a cytosine deletion in the first loop, was expressed into stable forms of Pxr-L and Pxr-S. Nonetheless, for Pxr_{Anc6}, the amount of Pxr-L was greatly reduced and Pxr-S barely detected; for Pxr_{Anc7}, neither Pxr-L nor Pxr-S was detected.

The mutant variants that resulted in lower regulatory efficiencies (i.e. higher spore production) tend to produce less stable stem-loop structures. Compared to the calculated self-folding free energies of Pxr_{WT+} and Pxr_{Anc} (ΔG = -58.3 and -57.5), the ones of Pxr_{Anc4}, Pxr_{Anc3} and Pxr_{Anc6} are slightly higher (ΔG = -51.1, -51.1 and -49.9, respectively) and the one of Pxr_{Anc7} is much higher (ΔG = -32.1). On the other hand, for the mutant alleles that remain high regulatory efficiencies (i.e. low spore production), their sequences fold into more stable stem-loop structures and their ΔG is in the range between -57.5 and -53.7.

4. Discussion

Bacterial sRNAs are widespread and regulate a diversity of important physiological processes but how their functions diverged across species remains less studied. It is well conceived in evolutionary biology that genes from the same species coevolve together and may not be compatible with genes from other species (Dobzhansky, 1937; Muller, 1942). Hybrid dysfunction caused by such mechanism is now supported by abundant evidence in animals and plants (Coyne and Orr, 2004). Under similar conceptual framework, sRNAs from one species may not be able to control target mRNAs from another species. Here, we showed that the regulatory interactions of the Pxr sRNA observed in *M. xanthus* has been established since its likely origin within the myxobacteria and that such interactions appear to be widely conserved across different species. Remarkably, the *pxr* alleles we introduced into the *M. xanthus pxr* mutant are from species diversified in fruiting body size, shape and color; however, many of their Pxr sRNAs controlled *M. xanthus* development in a manner similar to the one of *M. xanthus*, indicating strong selective constraints on both the processing and interactions with targets in different species. Nevertheless, our results did not exclude the possibility that regulons of Pxr have expanded in species other than *M. xanthus*.

Our results also showed that the specificity of Pxr function might be evolving in some lineages. Both the Pxr-1 and Pxr-3 sRNAs from *C. minor* Cbm6 and *C. violaceus* Cbvi34 failed to controlled development in *M. xanthus* and the their transcripts were present in much lower amounts than were transcripts from the *M. xanthus* allele and other homologs. Although the structural stabilities of all the naturally occurring *pxr* alleles examined were comparable to the

one in *M. xanthus*, the sequences of *pxr-1* and *pxr-3* in *Cystobacter* were predicted to fold into only two long stem-loops, missing the short stem-loop that is present in Pxr-2 and other single-copy Pxr homologs. It is plausible that either the sequences or folded structures of *pxr-1* and *pxr-3* affect their auto-regulation, or that these alleles have coevolved with accessory proteins that modulate their stability. For example, in *E. coli*, the RNA chaperone Hfq protects sRNAs from degradation by ribonucleases (Masse et al., 2003). Nevertheless, Hfq preferentially binds to AU-rich sequences (Valentin-Hansen et al., 2004; Brennan and Link, 2007) and no homologs of Hfq were detected in the sequenced myxobacterial genomes that are generally GC-rich. It is also known that in *E. coli* polynucleotide phosphorylase (PNPase) increases the stability of sRNAs (De Lay and Gottesman, 2011). Further identification of Pxr accessory proteins in the myxobacteria will allow testing for the coevolution between Pxr and its associated regulatory network.

Known sRNA gene duplications in bacteria include OmrA and OmrB responsible for high osmolarity in *E. coli* (Guillier and Gottesman, 2006), PrrF1 and PrrF2 for iron limitation in *Pseudomonas aeruginosa* (Wilderman et al., 2004) and four copies of Qrr for quorum sensing in *Vibrio cholera* (Lenz et al., 2004). In these cases, the duplicated sRNAs share similar targets and are redundant in function, although there appear to be subtle differences in their expression profiles. Having multiple copies of an sRNA can have regulatory implications, allowing for tight regulation on mRNA targets or differential regulation of the copies. Evolution of gene duplications has long been recognized as a major source of genes with new functions (reviewed in Taylor and Raes, 2004). New genes can evolve from a redundant copy of a duplicated parental gene and then acquire mutations providing a new beneficial function. Our results that the

sequence variation among *pxr* paralogs in *Cystobacter* is functionally significant might present a further step after the initial gene duplication event. Future elucidation on the effects of each *pxr* allele in the *Cystobacter* species will provide insights into the functional diversification of Pxr sRNAs in this subclade.

Finally, we examined the effects of serendipitous mutations in the conserved and non-conserved positions across *pxr* homologs. The seven-base cytosine run in the first loop of Pxr is well conserved in all the *pxr* homologs identified except the *pxr-2* allele in *C. violaceus* Cbvi34, which contains one extra cytosine but still blocks *M. xanthus* development. Here we showed that one cytosine deletion that shortens the seven-cytosine stretch into a six-cytosine stretch abolished Pxr function. This evidence, along with the C→A mutation that mediates the evolutionary transition from OC to PX, strongly indicates that the first loop plays a pivotal role in Pxr function and awaits further investigation.

The intrinsic termination site of ρ-independent RNAs (including sRNAs) contains a stable, GC-rich hairpin followed by a string of uridine residues at the 3' end, and the termination efficiency depends both on sequence specificity and secondary structure (reviewed in Taylor and Raes, 2004). Recent studies revealed that Hfq-dependent sRNAs also use their 3' terminator to recruit Hfq for their modes of action (Cheng et al., 1991; Wilson and von Hippel, 1995). Our results show that the truncations of *pxr* alleles at the positions 96 and 103 resulted in great reduction or no expression of Pxr sRNAs, presumably due to reduced transcription termination efficiency. We also found a G→A mutation on the 3' stem that weakens the stability of the stem structure and resulted in increased variation in regulatory efficiency. Notably, the expression

levels of Pxr-L and Pxr-S from this mutant allele were comparable to the wild-type allele, suggesting that this nucleotide is required for Pxr to consistently regulate development, rather than for transcription termination. An apparent following step is to test whether the compensatory mutation on the other side of the stem restores the regulatory efficiency or not. This will reveal whether it's the structure or the sequence *per se* that contributed to the phenotype. It would also be interesting to replace the terminator of Pxr with ρ-independent terminators from other genes to examine whether the terminator of Pxr serves as a termination site alone or is associated with Pxr processing or regulatory function (e.g. as a processing site by ribonucleases or a site to recruit RNA chaperones to modulate it stability and interactions with mRNA targets).

Table 2.1 Primer Sequences

Amplified fragment	Forward primer sequence	Reverse primer sequence
429-nt fragment preceding *pxr*	CCC AGG TGG TGG AAG AGG	CGA TGT GTC CCG CGC ATT CCT A
pxr_{Mxs33}	AAG CGA GGC TGA AGG CGG TT	ATG CCC AAA AGA AGG CGG CCC GAC AC
pxr_{Mxs42}	AAG CGA GGC TGA AGG CGG TT	CAG CAC AAA AGA AGG CGG CCC GAC AC
$pxr_{Cbvi34\text{-}1}$	ACG TGG CGA TGA AGG TGG TTC CCC CCC ACC C	GGC CCA AAA AGA AGG CGG CCC GGT CCC CTC AGG
$pxr_{Cbm6\text{-}2}$	AAG TGA GGC TGA AGG AGG TTC CCC CCC ACC G	GGC GCA AAA AGA AGG AGG CCC GGT CAC CTC AGA
$pxr_{Cbvi34\text{-}2}$	AAG TGA GGC TGA AGG AGG TTC CCC CCC CAC CG	GGC GCA AAA AGA AGG AGG CCC GGT CAC CTC AGA
$pxr_{Cbvi34\text{-}3}$	AAG TGA GGC TGA AGA AGG TTC CCC CCC ACC T	GCC CAA AAA AGA AGG CGG CCC GGC ATC CTC AGG
pxr_{Sga}	AAG CGA GGC TGA AGG TGG TT	GGG CTA GAA AAG AAG GCG GCC CGA TA

Table 2.2 List of *M. xanthus* strains. "*intergenic*" refers to a 429-nt fragment that contains the last 167-nt of *nla19* (*Mxan_1078)* and the 262-nt intergenic region between *nla19* and *pxr* coding region.

Strain	Genotype	*pxr* allele description	Reference/source
GJV1	Derivative isolate of *Myxococcus xanthus* DK1622	Wild-type allele of *pxr*	(Fiegna *et al*, 2006; Goldman *et al*, 2006)
GVB207.3	Evolutionary descendant of GJV1 (herein "OC")	Wild-type allele of *pxr*	(Velicer *et al*, 1998; Fiegna *et al*, 2006)
GJV207	GVB207.3 Δ*pxr* (herein "OC Δ*pxr*")	In-frame deletion of *pxr*	(Yu *et al*, 2010)
NY01	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{WT}$, kan^R	Wild-type allele of *pxr*	Yuen-Tsu Nicco Yu
NY02	GJV207 *intergenic*::*pCR2.1_intergenic*, kan^R	*Mxan_1078-pxr* intergenic region only	Yuen-Tsu Nicco Yu
KC01	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Mxs33}$, kan^R	*Myxococcus stipitatus* Mxs 33 *pxr* allele	This study
KC02	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Mxs42}$, kan^R	*Myxococcus stipitatus* Mxs 42 *pxr* allele	This study
KC03	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Cbvi34\text{-}1}$, kan^R	*Cystobacter violaceus* Cbvi34 *pxr*-1 allele	This study
KC04	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Cbm6\text{-}2}$, kan^R	*Cystobacter minor* Cbm6 *pxr*-2 allele	This study
KC05	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Cbvi34\text{-}2}$, kan^R	*Cystobacter violaceus* Cbvi34 *pxr*-2 allele	This study
KC06	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Cbvi34\text{-}3}$, kan^R	*Cystobacter violaceus* Cbvi34 *pxr*-3 allele	This study
KC07	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Sga}$, kan^R	*Stigmatella aurantiaca* DSM17044 *pxr* allele	This study
KC08	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc}$, kan^R	The inferred *pxr* ancestral allele shared by *Myxococcus* and *Cystobacter* spp.	This study
KC09	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc1}$, kan^R	*pxr*$_{Anc}$ with a cytosine insertion between position 75 and 76 on the third stem	This study
KC10	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc2}$, kan^R	*pxr*$_{Anc}$ with a uracil deletion at position 87 in the third loop	This study
KC11	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc3}$, kan^R	*pxr*$_{Anc}$ with a cytosine deletion in the first loop	This study
KC12	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc4}$, kan^R	*pxr*$_{Anc}$ with point mutation G→A at position 102 on the third stem, weakens stem stability, C:A base pairs	This study
KC13	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc5}$, kan^R	*pxr*$_{Anc}$ with point mutation A→G at position 49 in the second loop	This study

KC14	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc6}$, kan^{R}	*pxr*$_{Anc}$ with point mutation U→C at position 16 (bulge), truncated at 3' end from positions 103 to 111	This study
KC15	GJV207 *intergenic*::*pCR2.1_intergenic-pxr*$_{Anc7}$, kan^{R}	*pxr*$_{Anc}$ with point mutation U→C at position 16 (bulge), truncated at 3' end from positions 96 to 111	This study

References

Aravin, A.A., Hannon, G.J., Brennecke, J., 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318, 761-764.

Babski, J., Maier, L.K., Heyer, R., Jaschinski, K., Prasse, D., Jager, D., Randau, L., Schmitz, R.A., Marchfelder, A., Soppa, J., 2014. Small regulatory RNAs in Archaea. RNA Biol. 11, 484-493.

Bardill, J.P., Hammer, B.K., 2012. Non-coding sRNAs regulate virulence in the bacterial pathogen *Vibrio cholerae*. RNA Biol. 9, 392-401.

Bartel, D.P., 2009. MicroRNAs: Target Recognition and Regulatory Functions. Cell 136, 215-233.

Bejerano-Sagie, M., Xavier, K.B., 2007. The role of small RNAs in quorum sensing. Curr. Opin. Microbiol. 10, 189-198.

Boysen, A., Moller-Jensen, J., Kallipolitis, B., Valentin-Hansen, P., Overgaard, M., 2010. Translational regulation of gene expression by an anaerobically induced small non-coding RNA in *Escherichia coli*. J. Biol. Chem. 285, 10690-10702.

Brennan, R.G., Link, T.M., 2007. Hfq structure, function and ligand binding. Curr. Opin. Microbiol. 10, 125-133.

Bretscher, A.P., Kaiser, D., 1978. Nutrition of *Myxococcus xanthus*, a fruiting myxobacterium. J. Bacteriol. 133, 763-768.

Chambers, J.R., Sauer, K., 2013. Small RNAs and their role in biofilm formation. Trends Microbiol. 21, 39-49.

Chen, I.C., Griesenauer, B., Yu, Y.T., Velicer, G.J., 2014. A recent evolutionary origin of a bacterial small RNA that controls multicellular fruiting body development. Mol. Phylogenet. Evol. 73, 1-9.

Cheng, S.W., Lynch, E.C., Leason, K.R., Court, D.L., Shapiro, B.A., Friedman, D.I., 1991. Functional importance of sequence in the stem-loop of a transcription terminator. Science 254, 1205-1207.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer, Sunderland, Massachusetts.

Davis, B.M., Waldor, M.K., 2007. RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. Mol. Microbiol. 65, 373-385.

Dawid, W., 2000. Biology and global distribution of myxobacteria in soils. FEMS Microbiol. Rev. 24, 403-427.

De Lay, N., Gottesman, S., 2009. The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. J. Bacteriol. 191, 461-476.

De Lay, N., Gottesman, S., 2011. Role of polynucleotide phosphorylase in sRNA function in *Escherichia coli*. RNA 17, 1172-1189.

Ding, S.W., 2010. RNA-based antiviral immunity. Nat. Rev. Immunol. 10, 632-644.

Dobzhansky, T., 1937. Further Data on the Variation of the Y Chromosome in *Drosophila Pseudoobscura*. Genetics 22, 340-346.

Durand, S., Storz, G., 2010. Reprogramming of anaerobic metabolism by the FnrS small RNA. Mol. Microbiol. 75, 1215-1231.

Fiegna, F., Yu, Y.T., Kadam, S.V., Velicer, G.J., 2006. Evolution of an obligate social cheater to a superior cooperator. Nature 441, 310-314.

Gaballa, A., Antelmann, H., Aguilar, C., Khakh, S.K., Song, K.B., Smaldone, G.T., Helmann, J.D., 2008. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. Proc. Natl. Acad. Sci. USA 105, 11927-11932.

Goldman, B.S., Nierman, W.C., Kaiser, D., Slater, S.C., Durkin, A.S., Eisen, J.A., Ronning, C.M., Barbazuk, W.B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H.S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S.A., Vaudin, M., Wiegand, R., Kaplan, H.B., 2006. Evolution of sensory complexity recorded in a myxobacterial genome. Proc. Natl. Acad. Sci. USA 103, 15200-15205.

Gottesman, S., Storz, G., 2011. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. Cold Spring Harb. Perspect. Biol. 3, a003798.

Guillier, M., Gottesman, S., 2006. Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. Mol. Microbiol. 59, 231-247.

Guillier, M., Gottesman, S., Storz, G., 2006. Modulating the outer membrane with small RNAs. Gene. Dev. 20, 2338-2348.

Hao, Y., Zhang, Z.J., Erickson, D.W., Huang, M., Huang, Y., Li, J., Hwa, T., Shi, H., 2011. Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. Proc. Natl. Acad. Sci. USA 108, 12473-12478.

Hodgkin, J., Kaiser, D., 1977. Cell-to-cell stimulation of movement in nonmotile mutants of *Myxococcus*. Proc. Natl. Acad. Sci. USA 74, 2938-2942.

Horler, R.S., Vanderpool, C.K., 2009. Homologs of the small RNA SgrS are broadly distributed in enteric bacteria but have diverged in size and sequence. Nucleic Acids Res. 37, 5465-5476.

Kawamoto, H., Koide, Y., Morita, T., Aiba, H., 2006. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Mol. Microbiol. 61, 1013-1022.

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.

Masse, E., Escorcia, F.E., Gottesman, S., 2003. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Gene. Dev. 17, 2374-2383.

Masse, E., Salvail, H., Desnoyers, G., Arguin, M., 2007. Small RNAs controlling iron metabolism. Curr. Opin. Microbiol. 10, 140-145.

Moller, T., Franch, T., Udesen, C., Gerdes, K., Valentin-Hansen, P., 2002. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. Gene. Dev. 16, 1696-1706.

Muller, H.J., 1942. Isolating mechanisms, evolution and temperature. Biol. Symp. 6, 71-125.

Overgaard, M., Johansen, J., Moller-Jensen, J., Valentin-Hansen, P., 2009. Switching off small RNA regulation with trap-mRNA. Mol. Microbiol. 73, 790-800.

Pandey, S.P., Minesinger, B.K., Kumar, J., Walker, G.C., 2011. A highly conserved protein of unknown function in *Sinorhizobium meliloti* affects sRNA regulation similar to Hfq. Nucleic Acids Res. 39, 4691-4708.

Papenfort, K., Podkaminski, D., Hinton, J.C., Vogel, J., 2012. The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. Proc. Natl. Acad. Sci. USA 109, E757-764.

Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J.C., Vogel, J., 2009. Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Mol. Microbiol. 74, 139-158.

Peer, A., Margalit, H., 2014. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. RNA 20, 994-1003.

Peterman, N., Lavi-Itzkovitz, A., Levine, E., 2014. Large-scale mapping of sequence-function relations in small regulatory RNAs reveals plasticity and modularity. Nucleic Acids Res. 42, 12177-12188.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C., Vogel, J., 2009. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat. Struct. Mol. Biol. 16, 840-846.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sharma, C.M., Darfeuille, F., Plantinga, T.H., Vogel, J., 2007. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Gene. Dev. 21, 2804-2817.

Skippington, E., Ragan, M.A., 2012. Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella* genomes. Genome Biol. Evol. 4, 330-345.

Storz, G., Vogel, J., Wassarman, K.M., 2011. Regulation by Small RNAs in Bacteria: Expanding Frontiers. Mol. Cell 43, 880-891.

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.

Taylor, J.S., Raes, J., 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38, 615-643.

Toffano-Nioche, C., Nguyen, A.N., Kuchly, C., Ott, A., Gautheret, D., Bouloc, P., Jacq, A., 2012. Transcriptomic profiling of the oyster pathogen *Vibrio splendidus* opens a window on the evolutionary dynamics of the small RNA repertoire in the *Vibrio* genus. RNA 18, 2201-2219.

Valentin-Hansen, P., Eriksen, M., Udesen, C., 2004. The bacterial Sm-like protein Hfq: a key player in RNA transactions. Mol. Microbiol. 51, 1525-1533.

Viegas, S.C., Silva, I.J., Saramago, M., Domingues, S., Arraiano, C.M., 2011. Regulation of the small regulatory RNA MicA by ribonuclease III: a target-dependent pathway. Nucleic Acids Res. 39, 2918-2930.

Vogel, J., Luisi, B.F., 2011. Hfq and its constellation of RNA. Nat. Rev. Microbiol. 9, 578-589.

Vogel, J., Papenfort, K., 2006. Small non-coding RNAs and the bacterial outer membrane. Curr. Opin. Microbiol. 9, 605-611.

Wadler, C.S., Vanderpool, C.K., 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc. Natl. Acad. Sci. USA 104, 20454-20459.

Wadler, C.S., Vanderpool, C.K., 2009. Characterization of homologs of the small RNA SgrS reveals diversity in function. Nucleic Acids Res. 37, 5477-5485.

Wilderman, P.J., Sowa, N.A., FitzGerald, D.J., FitzGerald, P.C., Gottesman, S., Ochsner, U.A., Vasil, M.L., 2004. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. Proc. Natl. Acad. Sci. USA 101, 9792-9797.

Wilson, K.S., von Hippel, P.H., 1995. Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proc. Natl. Acad. Sci. USA 92, 8793-8797.

Yu, Y.T., Yuan, X., Velicer, G.J., 2010. Adaptive evolution of an sRNA that controls *Myxococcus* development. Science 328, 993.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406-3415.

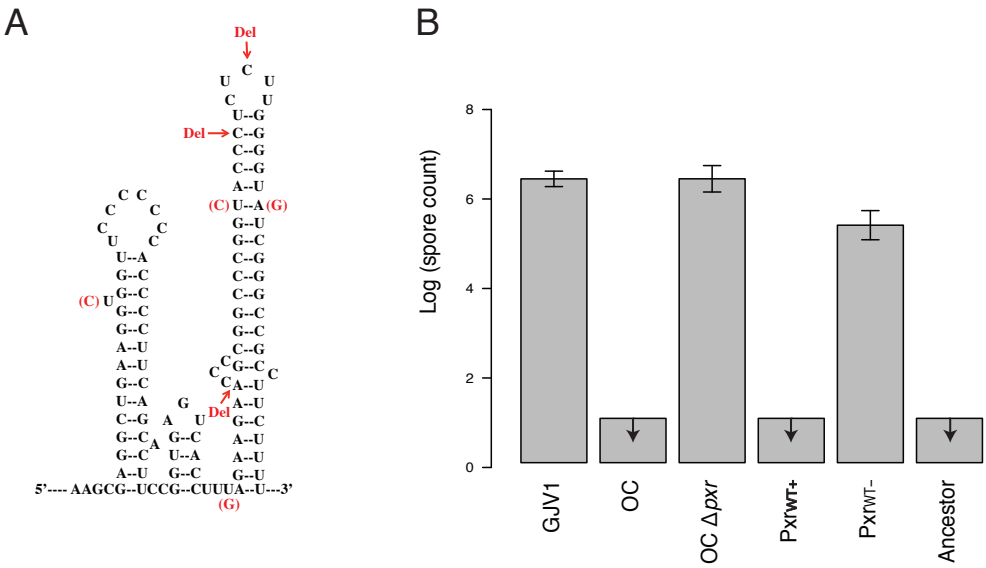


Figure 2.1 Pxr ancestor blocked *M. xanthus* development. (A) The predicted secondary structure of the Pxr ancestor, which has a calculated self-folding free energy of -57.5 kcal/mol. Base symbols in parentheses illustrate the differences between the ancestor and the *M. xanthus* Pxr sRNA. (B) Spore production of the OC Δ*pxr* strain carrying the *pxr* ancestral alleles and the control strains. Arrows indicate that no spores were observed at the lower limit of detection. Error bars represent standard deviations.

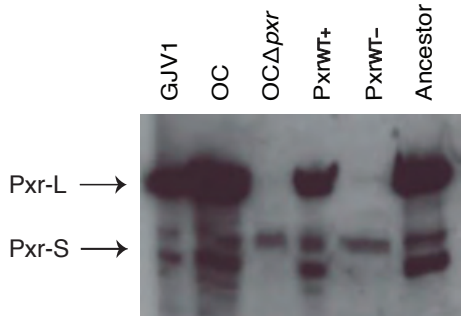


Figure 2.2 Northern blot showing the expression of Pxr-L and Pxr-S produced by the OC Δ*pxr* strain carrying the *pxr* ancestral allele and the control strains.

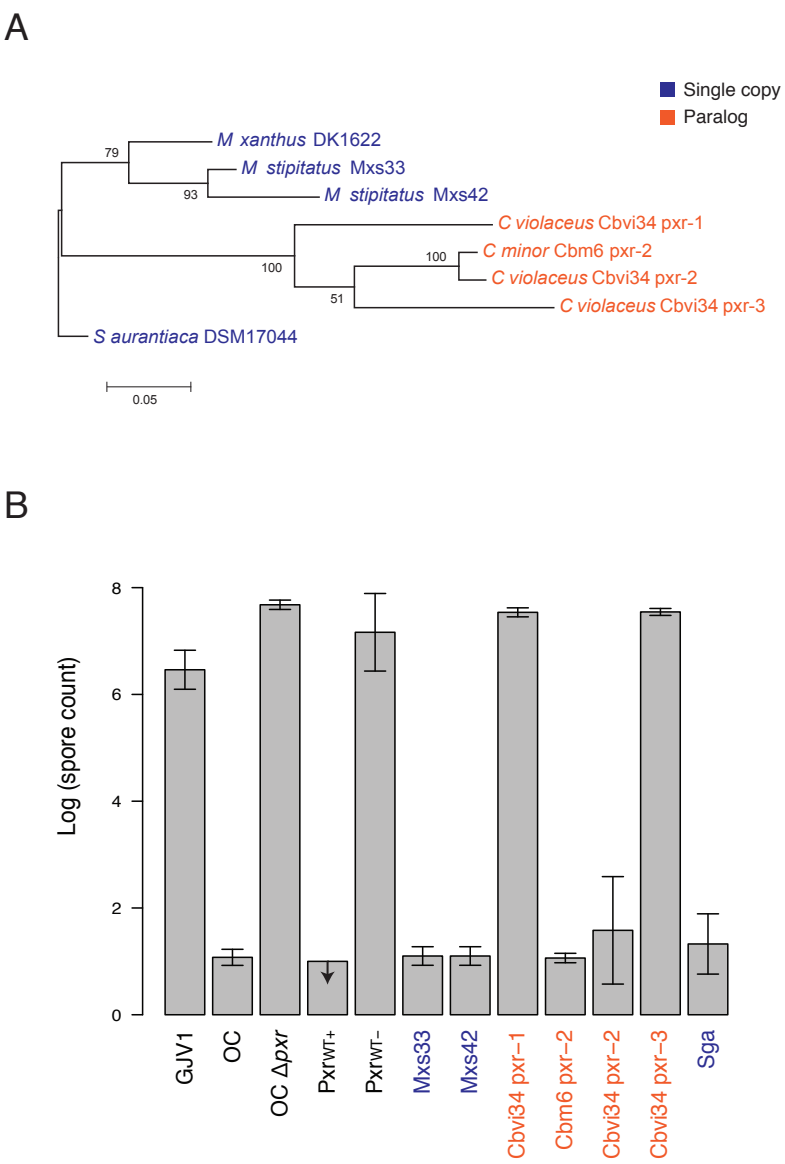


Figure 2.3 Single-copy Pxr homologs from different species blocked *M. xanthus* development but there appears to be functional variation among the paralogs in *Cystobacter*. Dark blue marks single-copy *pxr*/Pxr homologs and orange marks *pxr*/Pxr paralogs. (A) Bayesian gene tree of the *pxr* alleles examined in this study. Posterior probabilities are shown at the nodes. The scale bar shows 0.05 nucleotide substitutions per site. (B) Spore production of the OC Δ*pxr* strains

integrated with *pxr* alleles from different species. Arrows indicate that no spores were produced at the lower limit of detection. Error bars represent standard deviations.

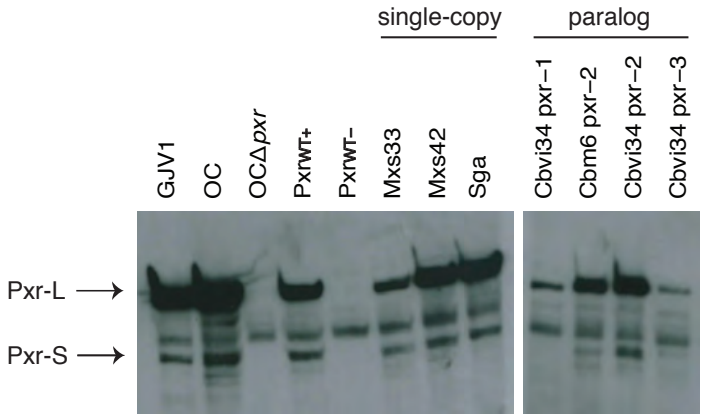


Figure 2.4 Northern blot showing the expression of Pxr-L and Pxr-S produced by the OC Δ*pxr* strains integrated with *pxr* alleles from different species of myxobacteria and the control strains.

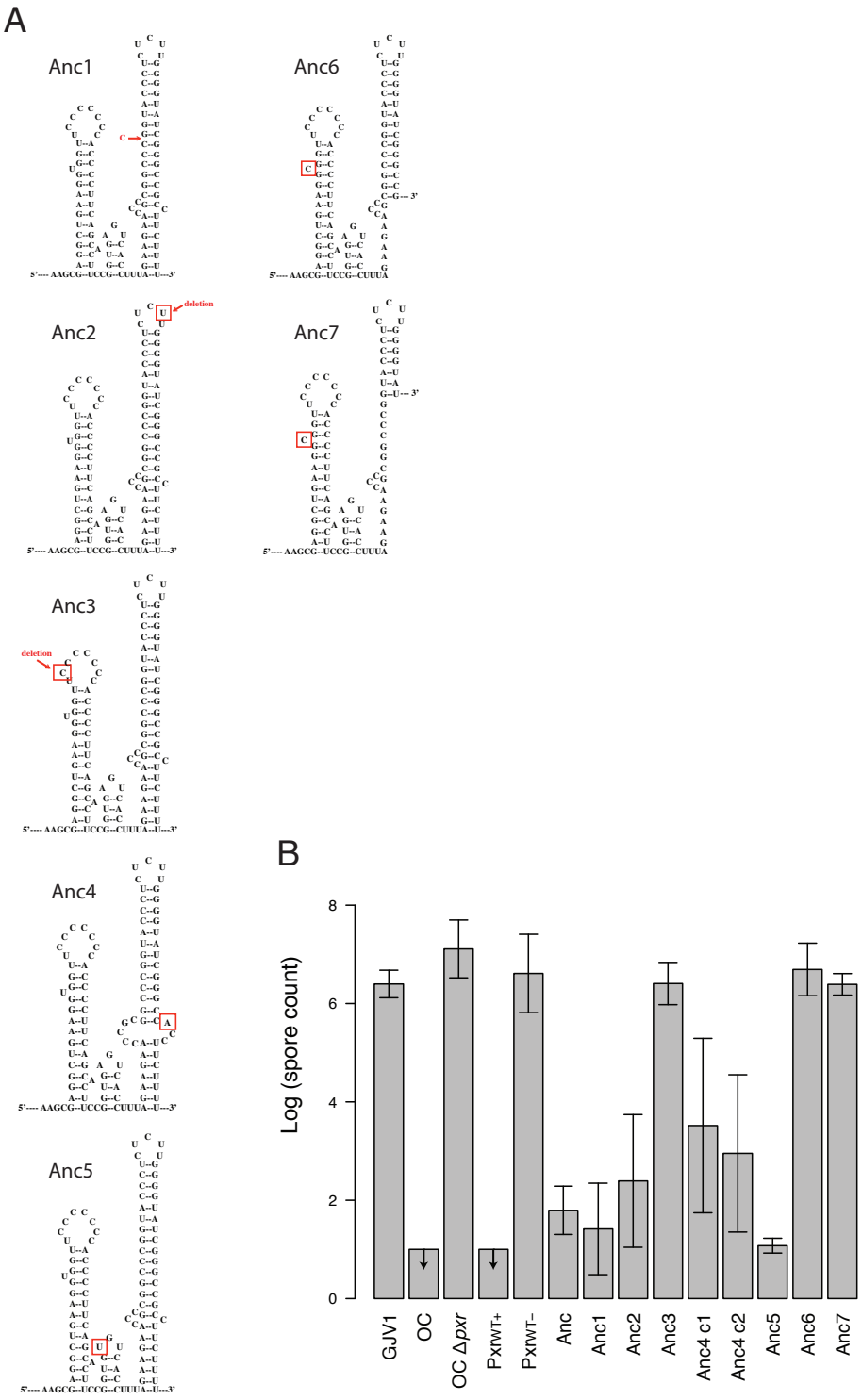


Figure 2.5 (A) Variants of the Pxr ancestor shown as RNA sequences folded into the secondary structures. (B) Spore production of the OC Δ*pxr* strains integrated with mutated *pxr* ancestral

alleles. Arrows indicate that no spores were observed at the lower limit of detection. Error bars represent standard deviations.

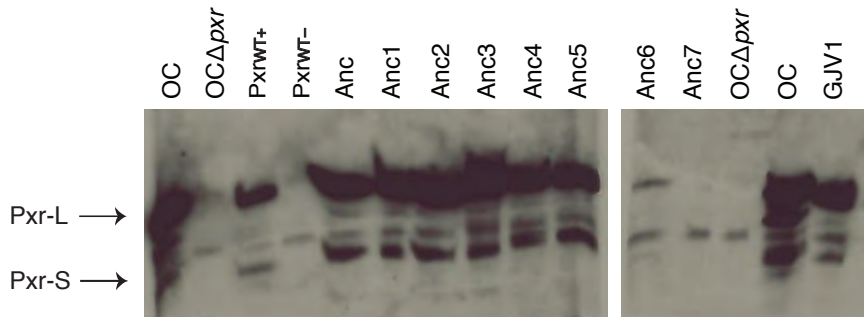


Figure 2.6 Northern blots showing the expression of Pxr-L and Pxr-S produced by the OC Δ*pxr* strains integrated with mutated *pxr* ancestral alleles.

CHAPTER III

Identification of genes involved in the Pxr sRNA regulatory pathway in *Myxococcus xanthus* with transposon mutagenesis

I-Chen Chen, Gregory J. Velicer and Yuen-Tsu Nicco Yu

Abstract

Small non-coding RNAs (sRNAs) have been identified as important regulators of gene expression in a variety of important biological processes in bacteria. These *trans*-encoded regulators are small in size (~80 to 100 nucleotides), are predicted to form hairpin structures and regulate expression of a wide variety of genes. Nevertheless, little is known about the evolution of sRNA regulatory pathways in bacteria. The sRNA Pxr in the bacterium *Myxococcus xanthus* functions as a gatekeeper that prevents the initiation of fruiting body development when nutrients are abundant. Pxr was discovered as the result of adaptive evolution of a developmentally defective strain (Obligate Cheater, OC) to a developmentally proficient strain (Phoenix, PX). In OC, Pxr is constitutively expressed and blocks development regardless of nutrient level. In PX, a single mutation deactivates Pxr and allows development to proceed. Here, we conducted insertion mutagenesis in the OC genetic background with a transposon derived from the eukaryotic *Himar1* system. We screened for insertions that mediated the conversion of OC to mutants with PX-like phenotypes to identify genes involved in the Pxr regulatory pathway. Insertions were found in four genes: *rnd*, *rnhA*, *stkA* and an uncharacterized gene, distinct from those described in other bacterial sRNA-based regulation (e.g. the ones in *Escherichia coli*). We examined the evolutionary divergence of these genes across species and the functional roles of some of these genes. Taken together, our results show that the Pxr regulatory pathway evolved not only by co-opting existing genes but by incorporating a novel gene exclusive to the myxobacterial subclade. Precisely how these genes function in the Pxr regulatory pathway remains to be investigated.

1. Introduction

Tremendous efforts in the past decade have established small non-coding RNAs as critical regulators that modulate gene expression of a variety of biological processes across all domains of life. For example, microRNAs (miRNAs) are an abundant class of small (21-22 nucleotides) non-coding RNAs involved in developmental gene regulation in animals and plants (Jones-Rhoades et al., 2006; Stefani and Slack, 2008) and have been found to play an essential role in aging and cancer (Kato and Slack, 2008; Garzon et al., 2009). In defense against invading nucleic acids such as viruses and transposons, eukaryotic small interfering RNAs (siRNAs) provide hosts immunity via complementary interactions with these foreign elements (Waterhouse et al., 2001; Li and Ding, 2005). In bacteria, small non-coding RNAs (sRNAs, 80-100 nucleotides) are functionally analogous to miRNAs and also regulate a range of important physiological processes, including many stress responses (Gottesman and Storz, 2011; Storz et al., 2011). In general, small RNAs regulate gene expression at the post-transcriptional level and specific proteins are required for their identification, processing and interactions with mRNA targets.

In eukaryotes, miRNAs and siRNAs regulate gene expression through a widely conserved mechanism, namely RNA interference (RNAi) machinery (Fire et al., 1998). The key components of RNAi machinery include the RNaseIII-like endonuclease Dicer, Argonaute-Piwi (Ago-Piwi) family protein and RNA-dependent RNA polymerase (RdRP). Dicer is the primary RNA recognition and processing protein that processes the precursors of miRNAs and siRNAs into short double-stranded RNAs (dsRNAs) (Jaskiewicz and Filipowicz, 2008). The short

dsRNAs are incorporated into a multi-protein complex named the RNA-induced silencing complex (RISC), of which members of the Ago-Piwi protein are core components (Pham et al., 2004; Tomari et al., 2004). The short RNA duplex is then unwound to a passenger strand and a guiding strand, and the latter directs the RISC to target and degrade complementary mRNA sequences. RdRP is involved mainly in additional siRNA amplification (Pak and Fire, 2007; Sijen et al., 2007). Phylogenetic studies have shown that Dicer-like, Ago-Piwi and RdRP proteins are present in at least some members of four of the five supergroups of eukaryotes, indicating that the last common ancestor of eukaryotes (LECA) possessed the biochemical capacity for the RNAi mechanism. Nevertheless, RNAi machinery appears to have been lost several times in independent lineages (e.g. *Saccharomyces cerevisiae*, *Leishmania major* and *Plasmodium falciparum*) (Cerutti and Casas-Mollano, 2006; Shabalina and Koonin, 2008).

In prokaryotes, however, less is understood about the evolution of sRNA regulatory pathways, for example, whether core components are conserved across different lineages or specialized in specific lineages. Many characterized sRNAs require the RNA chaperone Hfq for their activities (Gottesman and Storz, 2011) as Hfq can induce conformational changes of folded sRNAs and facilitate their base paring with target mRNAs (Brennan and Link, 2007). However, the YbeY protein in the rhizobial bacterium *Sinorhizobium meliloti*, which is highly conserved in most bacteria and shares structural similarities with the eukaryotic Argonaute proteins, modulates the expression of sRNAs and target mRNAs in a fashion similar to Hfq (Pandey et al., 2011). Also, in the bacterium *Bacillus subtilis*, FsrA sRNA depends partially on three Fur-regulated small proteins and is independent of an Hfq-like homolog encoded in its genome (Gaballa et al., 2008). Unlike eukaryotic miRNAs that are derived from precursors, most bacterial sRNAs are presented

to their target mRNAs without processing, although ArcZ sRNA in *Escherichia coli* and *Salmonella enterica* serovar Typhimurium and MicX sRNA in *Vibrio cholera* are processed from long transcripts into active or stable short transcripts by RNase E (Argaman et al., 2001; Davis and Waldor, 2007; Papenfort et al., 2009). In addition to processing, RNases E and RNase III are involved in mRNA degradation (Masse et al., 2003; Pfeiffer et al., 2009; Viegas et al., 2011). Until now, nonetheless, the phylogenetic patterns of these key components involved in sRNA pathways have not been extensively explored.

The free-living soil bacterium *Myxococcus xanthus* is a model system for multicellular fruiting body development and social behaviors. During vegetative growth, *M. xanthus* cells glide over surfaces and consume other bacteria, fungi or macromolecules by collectively secreting extracellular digestive enzymes. Remarkably, upon starvation they initiate a multicellular developmental program that culminates in the formation of spore-bearing fruiting bodies. A previous study showed that Pxr sRNA serves as a checkpoint that prevents the initiation of fruiting body development when nutrients are abundant in the environment (Yu et al., 2010). The *pxr* gene contains a putative σ^{54} promoter, folds into a stable triple stem-loop structure, and is located in the intergenic region between the σ^{54}-dependent response regulator *nla19* and an acetyltransferase gene (predicted genes *Mxan_1078 and Mxan_1079*, respectively, in the *M. xanthus* DK1622 genome (Goldman et al., 2006)) (Yu et al., 2010). There are two *pxr*-specific sRNA forms, Pxr-L (long) and Pxr-S (short), and Pxr-S appears to be the primary regulator as it is expressed in high amounts during vegetative growth but predominantly diminished upon starvation (Yu et al., 2010). Thus far, little has been known about the genes involved in the Pxr regulatory pathway, and no homolog of the RNA chaperone Hfq required for

many sRNA-based gene regulations in bacteria is present in the *M. xanthus* genome. The presence of Pxr-L and Pxr-S suggests that a ribonuclease functions to shorten the long form into the active short form.

Notably, Pxr was discovered as the results of adaptive evolution by a developmentally defective *M. xanthus* strain (OC, obligate cheater) to a strain with restored developmental proficiency (PX, phoenix) (Fiegna et al., 2006; Yu et al., 2010). OC is a descendant of the developmentally proficient wild type GJV1 after 1000 generations of evolution in rich liquid medium and accumulated 14 mutations during this period (Velicer et al., 1998; Velicer et al., 2006). This cheater strain is defective at fruiting body development in clonal groups but can exploit GJV1 in chimeric groups by sporulating more efficiently than its competitor. PX evolved from OC through a spontaneous C→A mutation in the first loop of Pxr (Yu et al., 2010). Temporal expression patterns showed that Pxr-S is constitutively expressed in the OC genetic background, suggesting that OC is defective at modulating Pxr-S levels in response to starvation. The C→A mutation deactivates Pxr function and leads to fruiting body formation and high spore production in PX despite the fact that Pxr-S levels remain high in the cells. The genetic background of OC hence provides an excellent opportunity to identify genes involved in the Pxr regulatory pathway. We reasoned that mutations in other loci of the *M. xanthus* genome that are essential for Pxr regulatory function should also lead to PX-like phenotypes.

The aims of this study were to identify the core components involved in the regulatory pathway of Pxr sRNA in *M. xanthus,* generate hypotheses about their functional roles and analyze their phylogenetic distribution and relatedness. Previously, a mutagenesis screen with a

Himar1-based transposon was conducted and mutants were screened for insertions that caused a change from the defective OC developmental phenotype to a sporulation proficient, PX-like phenotype. *Himar1* is a member of the *mariner* superfamily of transposon elements in eukaryotes and was originally isolated from the horn fly *Haematobia irritans* (Hartl et al., 1997; Rubin et al., 1999). Studies have shown that the transposition of this transposon occurs in bacteria including both *E. coli* and *Mycobacterium smegmatis*, as well as the focal species *M. xanthus* (Rubin et al., 1999; Youderian et al., 2003). The spectra of transposon insertions appear to have little site specificity beyond the requirement for the dinucleotide TA. Here, I investigated the insertion loci of transposon-derived mutants that restore the developmental proficiency in OC and characterized the Pxr-related effects of some of the candidate genes. This approach allowed us to identify genes essential for the Pxr regulatory function.

2. Materials and methods

2.1. Bacterial strains and plasmids

A list of *M. xanthus* strains and plasmids used in this study is provided in Table 3.1. *M. xanthus* control strains include the wild-type strain GJV1, the developmentally defective strain OC and the developmentally proficient strains PX and OC Δ*pxr*. Strain OC was electroporated with the plasmid pMiniHimar that contains the R6K origin of replication and the kanamycin resistance gene originally derived from transposon Tn*5*, flanked by *Himar1*-inverted repeats (Heidi Kaplan, personal communications). Twenty-one PX-like transposon mutants were obtained. Two of the mutants, OC::TnE1 and OC::TnF4 (Table 3.1), were originally screened for

a proficient PX developmental phenotype on CF (*c*lone *f*ruiting) 1.5% agar (10 mM Tris-HCl, 8 mM $MgSO_4$, 0.2mg/ml $(NH_4)_2SO_4$, 0.15 mg/ml Bacto Casitone, 2 mg/ml Na citrate, 1 mg/ml Na pyruvate, 1 mM KH_2PO_4-K_2HPO_4 and 1.5% agar) topped with TPM 0.7% agar (10 mM Tris-HCl, 8 mM $MgSO_4$, 1 mM KH_2PO_4-K_2HPO_4 and 0.7% agar) by Yuen-Tsu Nicco Yu. The remaining 19 mutants were screened on CF 1.5% agar (1 mg/ml Bacto Casitone) by the previous undergraduate students Brandon Satinsky and Carolyn Rhodebeck.

2.2. Backcrossing transposon mutations into the parental strain OC

The 21 transposon insertions were backcrossed into the parental strain OC to confirm that the respective PX-like phenotypes were linked with a transposon insertion instead of spontaneous mutations or multiple transposon insertions. To backcross insertions into OC, genomic DNA from each transposon mutant was isolated from vegetative cultures following Wu and Kaiser (1995). The competent OC cells were prepared by growing cultures to an optical density between 0.4 and 0.6 in CTT liquid (10 mM Tris-HCl, 8 mM $MgSO_4$, 10 mg/ml Bacto Casitone and 1 mM KH_2PO_4-K_2HPO_4) at 32°C with constant shaking at 300 rpm, harvested by centrifugation, washed and resuspended with double distilled water. A 75 µl aliquot of competent cells was mixed immediately with 1 µl of genomic DNA and subjected to electroporation conditions of 400 Ω, 0.65 KV and 25 µFD. The cells in the electroporation cuvette were transferred immediately to 3 ml CTT liquid and incubated overnight at 32°C and 300 rpm before plating on CF (0.15 mg/ml Bacto Casitone) 1.5% agar topped with TPM 1.5% agar supplemented with 40 µg/ml kanamycin. After 10 days of incubation at 32°C with 90% rH, developmental phenotypes (i.e. the presence or absence of fruiting body formation) of individual colonies were examined

under a microscope. The plates with fruiting colonies were incubated at 50°C for 2 h to kill any living non-spore cells that remained near fruiting bodies. The fruiting bodies from individual colonies were then collected and inoculated into CTT liquid supplemented with 40 μg/ml kanamycin. The freezer stocks of backcrossed mutants were stored in 80% glycerol at -80°C.

2.3. Identifying insertion loci in the M. xanthus genome

The method of identifying insertion loci is illustrated in Figure 3.1. Genomic DNA of each backcrossed mutant was isolated following Wu and Kaiser (1995). Aliquots of DNA were digested with SacII, ligated with T4 ligase and PCR amplified with the primers pMiniHimar and himar 615 (Table 3.1) that hybridized the ends of transposon insertion. The PCR products were then sequenced and searched against *M. xanthus* DK 1622 genome using BLASTn to determine the flanking DNA sequences.

2.4. Developmental assays

To initiate development, *M. xanthus* cells were grown to OD values between 0.4 and 0.6 in CTT liquid at 32°C with constant shaking at 300 rpm (supplemented with 40 μg kanamycin/ml for backcrossed mutants). Cells were harvested by centrifugation and resuspended in TPM liquid to a density of 5 x 10^9 cells per ml. For each strain, 50 μl of resuspended cells were spotted on the center of TPM and CF (0.15 mg/ml Bacto Casitone) 1.5% agar plates and incubated at 32°C with 90% rH. After three days on TPM plates and five days on CF plates, samples were examined under a microscope for fruiting body phenotypes. Cells were then harvested with a

scalpel blade, transferred into 1 ml of double distilled water and heated at 50°C for two hours to select for heat-resistant spores. Samples were sonicated with a microtip to disperse spores and plated into CTT 0.5% agar in a series of dilution. After seven days, visible colonies were counted to estimate the total number of spores produced during development. In cases where no colonies grew at the lowest dilution factor (10^{-1}), a value of ten spores per replicate was entered for data analysis, providing a high-end estimate of spore production in these cases. The same experimental procedure was repeated three times with the same set of strains.

2.5. Phylogenetic analyses

To investigate whether the identified gene candidates may have coevolved with the Pxr sRNA in the myxobacteria, we examined the conservation pattern of these loci in the 16 sequenced genomes of the delta-proteobacteria, eight of which are myxobacterial genomes. Homologs of *pxr* were found in five species within the suborder Cystobacterineae but not in three other genomes outside this suborder. The remaining eight genomes are phylogenetically diverse delta-proteobacteria. To reconstruct the phylogeny in the delta-proteobacteria, DNA sequences of their 16S rRNA genes were downloaded from Genbank and aligned with MUSCLE implemented in MEGA version 5.0 (Tamura et al., 2011). The phylogeny was estimated using the maximum likelihood (ML) analysis in MEGA 5.0. The reliability of the ML tree was evaluated with 1000 bootstrap replicates. To examine the phylogenetic conservation of the candidate genes, the protein sequence of each gene in *M. xanthus* was used as query sequence to search against each delta-proteobacterial genome in the BLASTp analysis (Altschul et al., 1997). The obtained bit score was divided by the one when searching against its own *M. xanthus* genome. The resulting

percentages range from 100% (when compared with its own sequence), <30% (no overall sequence similarity but motifs with similar functions were found) to 0% (no significant hits in the BLASTp). As comparison, we also estimated the protein sequence conservation of the three conserved loci used in the phylogenetic analyses in Chapter I (*pyrG*, *rpoB* and *pgm*).

2.6. Inactivating the downstream genes of the identified loci

Disruption of genes downstream of the identified loci that appear to be located in multi-gene operons in the OC background was performed by homologous recombination leading to merodiploid mutants using internal fragments of those genes. The internal PCR fragments were amplified using primers listed in Table 3.1. The resulting PCR products were cloned into the pCR®2.1-TOPO® cloning vector (Invitrogen) to create inactivation plasmids. Disruption mutants were generated by electroporation of OC competent cells with plasmids and selection on CTT 0.5% agar supplemented with 40 µg kanamycin/ml. For each disruption mutant, two to three clones were examined for their developmental phenotypes to reassure that their phenotypes were linked with plasmid interruptions rather than spontaneous mutations.

2.7. Northern blot analysis of Pxr sRNA in the insertion mutants

To examine the roles of the insertion loci in the Pxr regulatory pathway, the expression of Pxr sRNAs in the insertion mutants was detected by Northern blotting. The total RNA of each strain was extracted from vegetative cultures growing in CTT liquid and cells incubated on TPM plates at 32°C with 90% rH. The cells on plates were washed with TPM liquid and collected by

centrifugation after one, three and six hours from the onset of development. Small-sized RNAs were enriched with the mirVana miRNA isolation kit (Ambion). RNA concentration was measured with a Nanodrop spectrophotometer and equal amounts of RNA were electrophoresed in 10% SequaGel (National Diagnostics) and electro-transferred onto a BrightStar®-Plus positively charged nylon membrane (Ambion). After UV cross-linking, the membrane was pre-hybridized in 4 ml UltraHyb-Oligo buffer for 30 min, and subsequently hybridized in the same solution containing 100 pmol 3'Biotin-TEG-*pxr* oligo probe (5'-ACC GGA AGT GCT GAA GGG GTG GGG GG-3') (Sigma) overnight. Pxr sRNA was detected with BrightStar® Biodetect non-isotopic kit (Ambion).

2.8. mRNA transcript levels of identified loci by RT-PCR

Reverse transcription (RT)-PCR was carried out to examine the expression patterns of the genes essential for Pxr function in the wild-type strain GJV1. The total RNA was extracted with the mirVana miRNA isolation kit (Ambion) from vegetative cultures growing in CTT liquid and cells incubated on TPM plates throughout nine hours of development. The cells on plates were washed with TPM liquid and harvested by centrifugation after one, three, six and nine hours from the onset of development. The RNA preparations were treated with TURBO DNA-*free* kit (Ambion) to remove contaminating DNA. RT-PCR was performed using RETROscript reverse transcription kit (Ambion) for reverse transcription and Advantage GC 2 PCR kit (Clontech) for PCR. The primers used for RT-PCR were listed in Table 3.1. The PCR products were resolved by staining with SYBR Safe DNA Gel Stain (Thermo) and electrophoresis on 1% agarose gel.

3. Results

Nine out of 21 original transposon mutants demonstrated PX-like phenotype on TPM agar, CF agar or both after backcrossed into the parental strain OC (Table 3.1) and we proceeded to identify the inserted loci in these nine mutants (two mutants were obtained from the CF/TPM screen and seven from the CF screen). For the two mutants originally screened on CF/TPM plates, OC::TnE1 formed fruiting bodies on both TPM and CF plates and OC::TnF4 formed fruiting bodies only on TPM but not on CF plates, although occasionally a few fruiting bodies were spotted. For the seven mutants originally screened on CF plates, three of them (OC::TnA26.1, OC::TnB22.2 and OC::TnD28.1) formed fruiting bodies both on TPM and CF plates and another three (OC::TnD41.1, OC::TnD75.1 and OC::TnE29.1) also formed fruiting bodies on TPM plates but only occasionally on CF plates. In contrast, the mutant OC::TnF77.2 only formed fruiting bodies on CF but not TPM plates. We also observed two mutants, OC::TnF45.1 and OC::TnF72.1, that appeared to form fruiting bodies or aggregated strongly on CF but not TPM plates and such phenotype was consistent between the original and backcrossed strains. Nevertheless, these two strains did not yield viable spores during development like the PX strain and we decided not to proceed with further characterizations. For the rest of the ten original mutants, our results suggest that their original PX-like fruiting body formation was caused by spontaneous mutations rather than transposon interruptions.

3.1. Mapping the insertion loci of OC::Tn mutants

For the nine transposon insertions that gave rise to fruiting body formation, they fall into four genes: *rnd* (*Mxan_5981*), *rnhA* (*Mxan_2265*), *stkA* (*Mxan_3474*) and a hypothetical protein (*HP*; *Mxan_5793*) (Fig 3.2). We mapped *stkA* in four mutants (OC::TnF4, OC::TnD41.1, OC::TnD75.1 and OC::TnE29.1) and *HP* in three mutants (OC::TnA26.1, OC::TnB22.2 and OC::TnD28.1); *rnd* and *rnhA* were each mapped in OC::TnE1 and OC::TnF77.2, respectively. In the mutants OC::TnA26.1 and OC::TnB22.2, the *Himar1* minitransposons were inserted in the intergenic region between *Mxan_5792* and *Mxan_5793* and it is likely that the insertions abolish the expression of the downstream *HP* (*Mxan_5793*). This is consistent with the phenotype of the mutant OC::TnD28.1 whose insertion is located within the reading frame of *HP*.

3.2. Developmental proficiency in OC::Tn mutants

To quantify the developmental proficiency caused by the transposon insertions in each candidate locus, we examined both fruiting body formation and spore production of their representative mutants, i.e. OC::TnE1 for *rnd*, OC::TnF4 for *stkA*, OC::TnA26.1 for *HP* and OC::TnF77.2 for *rnhA*. For the insertion loci *stkA* and *HP*, we only examined the phenotypes of OC::TnF4 and OC::TnA26.1 as similar developmental phenotypes were observed from other mutants with the same insertion loci.

On TPM plates, transposon insertions in *rnd*, *stkA* and *HP* gave rise to dark fruiting bodies and robust spore production almost as high as the PX strain (Fig 3.3). However, the insertion in *rnhA* did not yield evident fruiting bodies and its spore production was much lower compared to other mutants. In contrast, for the developmental proficiency on CF plates, insertions in *rnd* and

HP gave rise to dark fruiting bodies but low numbers of spores in the case of *rnd* or no spores at all (at the lower limit of detection) for *HP*. The insertion in *stkA* rarely gave rise to fruiting bodies and no viable spores were detected. Nevertheless, the insertion in *rnhA* gave rise to mature fruiting bodies and spore production as high as the PX strain.

3.3. Identified loci are highly conserved within the myxobacteria

We next compared the conservation patterns of the four candidate genes with the three conserved loci across the delta-proteobacteria (Fig 3.4). For the three conserved "control" loci, their protein sequences showed a gradual decrease in similarity from *M. xanthus* to other species of myxobacteria and delta-proteobacteria. In contrast, the loci that are candidates for involvement in the Pxr regulatory pathway showed a dramatically different conservation pattern. The protein sequences of *rnd*, *stkA* and *HP* are highly conserved (with bit score percentages > 75%) within the suborder Cystobacterineae in the myxobacteria where Pxr sRNA was found. Nevertheless, outside this suborder, both in the distantly related myxobacterial and other delta-proteobacterial genomes, their bit score percentages were all below 30%, except for *stkA* in *Anaeromyxobacter dehalogenan* (69%), *Sorangium cellulosum* (36%) and *Haliangium ochraceum* (35%) in the order Myxococcales. Intriguingly, no homologs of *HP* were found outside Cystobacterineae. For the fourth identified locus *rnhA*, its protein sequence was highly conserved within the genus *Myxococcus* (also with bit score percentages > 75%); however, outside *Myxococcus*, the bit score percentages were below 30% except in the species *Bdellovibrio bacteriovorus* (47%).

3.4. Inactivating the downstream genes of HP and *rnhA*

We decided to further characterize the Pxr-related effects of *HP* and *rnhA* as their potential roles in the Pxr pathway are unclear. Because these two genes appear to be located in multi-gene operons, we first examined whether their PX-like phenotypes are linked with transposon insertions in the two loci *per se* or rather due to polar effects on their downstream genes. If knockouts of downstream genes were to exhibit PX-like phenotypes, this would demonstrate that the phenotypes were due to polar effects on the downstream genes. If they were to display OC-like phenotypes, then the identified loci would be implicated. *HP* is upstream of another hypothetical protein (*Mxan_5794*) and an exonuclease gene (*Mxan_5795*). The knockout of the exonuclease gene displayed OC-like phenotype on TPM agar, indicating that PX-like phenotype in the transposon mutants OC::TnA26.1, OC::TnB22.2 and OC::TnD28.1 was not due to polar effects on this exonuclease gene (Fig 3.5A). However, for the construction of disruption mutant in the hypothetical protein (*Mxan_5794*), the transformation was not successful and we thus can not exclude the possibility that the phenotype caused by disrupting *HP* is mediated by a polar effect upon another hypothetical protein.

The locus of *rnhA* is oriented upstream of *mutA* (*Mxan_2264*), *mutB* (*Mxan_2263*), a gene encoding arginine/ornithine transport system ATPase (*Mxan_2262*) and a methylmalonyl-CoA mutase family protein (*Mxan_2261*). The *mutA* knockout displayed OC-like phenotype on CF agar, indicating that PX-like phenotype in the mutant OC::TnF77.2 was due to interruption of *rnhA per se*, not polar effects on *mutA* or other downstream genes (Fig 3.5B).

3.5. Pxr sRNA expression patterns in HP and rnhA mutants

We next examined how insertions in *HP* and *rnhA* affect the expression of Pxr-L and Pxr-S in vegetative growth and throughout six hours of development. In the parental strain OC, both Pxr-L and Pxr-S were constitutively expressed both the in vegetative stage and throughout six hours of development. In the mutant OC::TnA26.1 with insertion in *HP*, the expression of Pxr-L and Pxr-S remained high, like in OC, during both vegetative growth and development on TPM plates (Fig 3.6A). In the mutant OC::TnF77.2 with an insertion in *rnhA*, nonetheless, accumulation of Pxr-S was reduced both in vegetative growth and throughout six hours of development on CF plates (Fig 3.6B).

3.6. mRNA transcript levels of HP and rnhA in the wild-type GJV1

We further examined the mRNA transcript levels of *HP* and *rnhA* in the wild-type GJV1 using RT-PCR analysis to test their correlation with Pxr sRNA expression. The expression of both genes showed a decreasing trend from vegetative growth throughout nine hours of development on TPM plates, positively correlated with the expression of Pxr-S, but at different amounts and decreasing rate (Fig 3.7). The expression level of *HP* was high in vegetative cells, started to decrease until three hours after the onset of development and remained at similar levels thereafter. The expression of *rnhA* was low in the vegetative cells compared to that of *HP* and further decreased one hour from the onset of development. The expression was not detected after three hours of development.

4. Discussion

Taken together, insertion mutagenesis with the *Himar-1* transposon reveals four candidate genes in the Pxr sRNA regulatory pathway in *M. xanthus*. Three of them are annotated genes: *rnd*, *rnhA* and *stkA*, and the fourth identified gene is a previously uncharacterized gene (*HP*). None of the genes identified overlap with the ones required for other bacterial sRNA-based gene regulations, indicating that the genetic composition of sRNA regulatory pathways in bacteria can be highly variable. The fact that we identified both *stkA* and the uncharacterized gene in multiple mutants suggests that this mutagenesis approach may have saturated the genes involved in the Pxr regulatory pathway.

The evolutionary patterns of the candidate genes exhibited a high degree of conservation within the suborder Cystobacterineae where Pxr was found but not outside this suborder, which is significantly different from the patterns of the three conserved "control" loci. The protein sequences of the three annotated genes showed limited similarities with homologs outside Cystobacterineae, suggesting that the specificity of their functions may have evolved in this subclade. Further, the protein sequences of *rnhA* are specifically conserved within the genus *Myxococcus*, implying the incorporation of this genetic element for specific functions during a later stage in the evolution of the Pxr pathway. Intriguingly, the homologs of the uncharacterized gene were only detected within Cystobacterineae but not outside this suborder, representing a case of a newly evolved gene. Many species in the myxobacteria contain large genomes, with sizes of almost 10 Mb or even larger (Huntley et al., 2011). Until now, little is explored about the

contribution of novel genes to the large genome sizes and life history of the myxobacteria. Our finding here provides an interesting case that awaits further investigations.

Below we postulate the functional roles for each candidate gene (Table 3.2). Two of the identified genes are *rnd* and *stkA.* The gene product of *rnd* is ribonuclease D, or RNase D, an exoribonuclease that catalyzes the removal of extra nucleotides at the 3' end of a precursor tRNA to yield the mature tRNA (Cudny and Deutscher, 1980; Cudny et al., 1981). It is plausible that RNase D in *M. xanthus* is involved in processing Pxr-L to produce Pxr-S and this is supported by the fact that the transposon mutant OC::TnE1 (with insertion in *rnd*) only produces Pxr-L but not Pxr-S (Yuen-Tsu Nicco Yu, personal communication). Resolving the sequence of Pxr-S in the future will help to identify the recognition site for RNase D and reveal the structure and possible regulatory sequences in Pxr-S that might interact with other identified candidates.

The *stkA* locus encodes StkA, a homolog of the chaperone DnaK and Hsp70 family that facilitate the folding of newly synthesized proteins and the refolding of proteins that were denatured or misfolded by heat shock in many organisms (Genevaux et al., 2007). In *M. xanthus*, StkA appears to function as a negative regulator of extracellular polysaccharide (Moak et al., 2015) and was not found to be induced by heat shock (Otani et al., 2001). Our findings here suggest a new functional role of StkA in the Pxr-mediated developmental gene regulation. Preliminary results by Yuen-Tsu Nicco Yu has shown that, in contrast to the parental strain OC where Pxr-S is present during development, in the mutant OC::TnF4 (with insertion in *stkA*) Pxr-S is not stable upon starvation, indicating that StkA is involved in stabilizing Pxr-S in the cells.

The other two identified genes are *rnhA* and the uncharacterized gene, and their roles in the Pxr regulatory pathway remain obscure. The *rnhA* locus encodes Ribonuclease H, an endoribonuclease that cleaves the RNA strand in a DNA:RNA hybrid duplex (Vournakis et al., 1975; Donis-Keller, 1979; Gubler, 1987). During DNA replication, RNase H is involved in removing RNA primers, hence allowing the completion of DNA synthesis. If RNase H in *M. xanthus* evolved the ability to recognize a RNA:RNA duplex, it is possible that RNase H might cleave the mRNA in a sRNA:mRNA duplex, degrading the mRNA and recycling the sRNA back into the regulatory network. The pure speculation here suggests that RNase H might play a role in modulating the amounts of Pxr-S in the cells by releasing Pxr from the sRNA:mRNA duplex. This could explain the lower amounts of Pxr-S in the *rnhA* mutant OC::TnF77.2.

It is also unclear why the mutant OC::TnF77.2 only formed fruiting bodies and sporulated efficiently on CF but not TPM plates (CF differs from TPM by containing 0.02% $[NH_4]_2SO_4$, 0.1% pyruvate, 0.2% citrate and 0.015-0.1% Bacto Casitone). Because RNase H is also responsible for cellular function such as DNA replication, it is plausible that other ribonucleases complement its activities in OC::TnF77.2 but require substrates present in CF plates. The results here also suggest that Pxr-based gene regulation might be environment-dependent.

Finally, little is known about the uncharacterized gene identified here. The temporal expression pattern of its mutant showed that the amounts of both Pxr-L and Pxr-S remained stable in the cells, indicating that this gene is likely involved in the process after the maturation of Pxr-S, such as facilitating the base pairing between Pxr-S and targets. Further investigations are needed to better understand its role in the Pxr regulatory pathway.

Table 3.1 List of *M. xanthus* strains and plasmids.

Control strains

Strain name	Genotype	Reference/source
GJV1	Derivative isolate of *Myxococcus xanthus* DK1622	(Fiegna et al., 2006; Goldman et al., 2006)
GVB207.3	Evolutionary descendant of GJV1 (herein "OC")	(Velicer et al., 1998; Fiegna et al., 2006)
GJV81	Evolutionary descendant of GVB207.3 Kan^R (herein "PX")	(Fiegna et al., 2006)
GJV207	GVB207.3 Δ*pxr* (herein "OC Δ*pxr*")	(Yu et al., 2010)

Transposon mutants in the OC background and selected on CF/TPM

Strain name	Original developmental phenotype	Backcrossed developmental/ sporulational phenotype TPM	Loci mapped	Annotation	Reference/source
OC::TnE1	+	+/+	Mxan_5981	*rnd*	This study
OC::TnF4	+	+/+	Mxan_3474	*stkA*	This study

Transposon mutants in the OC background and selected on CF

Strain name	Original developmental phenotype	Backcrossed developmental/ sporulational phenotype TPM	CF	Loci mapped	Annotation	Reference/source
OC::TnA26.1	+	+/+	+/−	Mxan_5793	Hypothetical protein	This study
OC::TnA26.2	+	−	−			This study
OC::TnB3.2	+	−	−			This study
OC::TnB9.1	+	−	−			This study
OC::TnB17.2	+	−	−			This study
OC::TnB22.2	+	+/nd[a]	+/nd	Mxan_5793	Hypothetical protein	This study
OC::TnD1.1	+	−	−			This study
OC::TnD28.1	+	+/nd	+/nd	Mxan_5793	Hypothetical protein	This study
OC::TnD41.1	+	+/nd	+/nd	Mxan_3474	*stkA*	This study
OC::TnD75.1	+	+/nd	nd/nd	Mxan_3474	*stkA*	This study
OC::TnE6.1	+	−	−			This study
OC::TnE9.1	+	−	−			This study
OC::TnE26.1	+	−	−			This study
OC::TnE29.1	+	+/nd	−/nd	Mxan_3474	*stkA*	This study
OC::TnF45.1	+	−/−	+/−	Mxan_2903	Putative lipoprotein	
OC::TnF64.1	+	−	−			This study
OC::TnF72.1	+	−/nd	+/nd	Mxan_2903	Putative lipoprotein	This study
OC::TnF77.2	+	−/−	+/+	Mxan_2265	*rnhA*	This study
OC::TnF78.1	+	−	−			This study

Disruption mutants in the OC background

Strain name	Description	Reference/source
KC16	OC *mxan_5795::pCR2.1_5795*, kan^R	This study
KC17	OC *mxan_2264::pCR2.1_2264*, kan^R	This study

Plasmids

Plasmid	Relevant features	Reference/source
pMiniHimar	miniHimar transposon mutagenesis vector	Heidi Kaplan
pCR2.1	Cloning vector	Invitrogen
pCR2.1_2264	pCR2.1 with *Mxan_2264* internal fragment insert	This study
pCR2.1_5794	pCR2.1 with *Mxan_5794* internal fragment insert	This study
pCR2.1_5795	pCR2.1 with *Mxan_5795* internal fragment insert	This study

Primers

Primer	Relevant features	Sequence 5'- 3'
pMiniHimar	Used to sequence *M. xanthus* DNA flanking	CAT TTA ATA CTA GCG ACG CCA TCT
himar 615	transposon insertions	TCG GGT ATC GCT CTT GAA GGG
Mxan_2264 238f	Used to construct *Mxan_2264* disruption	CTG GTG TGC CAG GAG TAC AG
Mxan_2264 657r	mutant in OC	CGT GGA CAC GAG CAG CAC
Mxan_5794 -5f	Used to construct *Mxan_5794* disruption	CAT CCA TGC GTC AGT TCA TC
Mxan_5794 398r	mutant in OC	AAA ACG CCG GAA ATC ATC TT
Mxan_5795 89f	Used to construct *Mxan_5795* disruption	AGC TGG GCT GCA TCT TCT T
Mxan_5795 489r	mutant in OC	CAC CAG CAC CGA GTA CAC C
Mxan_2265 15f	Used for RT-PCR analysis of Mxan_2265 in	GAC CCT TGT CTT CGC TGA TG
Mxan_2265 221r	GJV1	GAG TCC GTG TGG ATG TGG AT
Mxan_5793 542f	Used for RT-PCR analysis of Mxan_5793 in	GCT TCG ACC AGG TGT ACG AG
Mxan_5793 755r	GJV1	ACT TCT GCC TTG TCG CTG AT

[a] nd: not detected.

Table 3.2 Functional-role hypotheses for the candidate genes

Candidate locus	Previous annotation	Empirical result	Functional-role hypothesis
rnd (*Mxan_5981*)	Processing tRNA at the 3' end	Only Pxr-L detected not Pxr-S	Processing Pxr-L into Pxr-S
stkA (*Mxan_3474*)	DnaK family protein	Pxr-S not stable	Stabilizing Pxr-S
rnhA (*Mxan_2265*)	Cleaving RNA in DNA:RNA hybrid	Reduced Pxr-S	?
HP (*Mxan_5793*)	Hypothetical protein	Both Pxr-L and Pxr-S detected	Interactions with targets

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., Altuvia, S., 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr. Biol. 11, 941-950.

Brennan, R.G., Link, T.M., 2007. Hfq structure, function and ligand binding. Curr. Opin. Microbiol. 10, 125-133.

Cerutti, H., Casas-Mollano, J.A., 2006. On the origin and functions of RNA-mediated silencing: from protists to man. Curr. Genet. 50, 81-99.

Cudny, H., Deutscher, M.P., 1980. Apparent involvement of ribonuclease D in the 3' processing of tRNA precursors. Proc. Natl. Acad. Sci. USA 77, 837-841.

Cudny, H., Zaniewski, R., Deutscher, M.P., 1981. *Escherichia coli* RNase D. Purification and structural characterization of a putative processing nuclease. J Biol. Chem. 256, 5627-5632.

Davis, B.M., Waldor, M.K., 2007. RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. Mol. Microbiol. 65, 373-385.

Donis-Keller, H., 1979. Site specific enzymatic cleavage of RNA. Nucleic Acids Res. 7, 179-192.

Fiegna, F., Yu, Y.T., Kadam, S.V., Velicer, G.J., 2006. Evolution of an obligate social cheater to a superior cooperator. Nature 441, 310-314.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391, 806-811.

Gaballa, A., Antelmann, H., Aguilar, C., Khakh, S.K., Song, K.-B., Smaldone, G.T., Helmann, J.D., 2008. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. Proc. Natl. Acad. Sci. USA 105, 11927-11932.

Garzon, R., Calin, G.A., Croce, C.M., 2009. MicroRNAs in Cancer. Annu. Rev. Med. 60, 167-179.

Genevaux, P., Georgopoulos, C., Kelley, W.L., 2007. The Hsp70 chaperone machines of *Escherichia coli*: a paradigm for the repartition of chaperone functions. Mol. Microbiol. 66, 840-857.

Goldman, B.S., Nierman, W.C., Kaiser, D., Slater, S.C., Durkin, A.S., Eisen, J.A., Ronning, C.M., Barbazuk, W.B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H.S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S.A., Vaudin, M., Wiegand, R., Kaplan, H.B., 2006. Evolution of sensory complexity recorded in a myxobacterial genome. Proc. Natl. Acad. Sci. USA 103, 15200-15205.

Gottesman, S., Storz, G., 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb. Perspect. Biol. 3, a003798.

Gubler, U., 1987. Second-strand cDNA synthesis: mRNA fragments as primers. Methods Enzymol. 152, 330-335.

Hartl, D.L., Lohe, A.R., Lozovskaya, E.R., 1997. Modern thoughts on an ancyent marinere: function, evolution, regulation. Annu. Rev. Genet. 31, 337-358.

Huntley, S., Hamann, N., Wegener-Feldbrugge, S., Treuner-Lange, A., Kube, M., Reinhardt, R., Klages, S., Muller, R., Ronning, C.M., Nierman, W.C., Sogaard-Andersen, L., 2011. Comparative genomic analysis of fruiting body formation in Myxococcales. Mol. Biol. Evol. 28, 1083-1097.

Jaskiewicz, L., Filipowicz, W., 2008. Role of Dicer in posttranscriptional RNA silencing. Curr. Top. Microbiol. 320, 77-97.

Jones-Rhoades, M.W., Bartel, D.P., Bartel, B., 2006. MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. 57, 19-53.

Kato, M., Slack, F.J., 2008. microRNAs: small molecules with big roles - *C. elegans* to human cancer. Biol. Cell 100, 71-81.

Li, H.W., Ding, S.W., 2005. Antiviral silencing in animals. FEBS Lett. 579, 5965-5973.

Masse, E., Escorcia, F.E., Gottesman, S., 2003. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Gene. Dev. 17, 2374-2383.

Moak, P.L., Black, W.P., Wallace, R.A., Li, Z., Yang, Z., 2015. The Hsp70-like StkA functions between T4P and Dif signaling proteins as a negative regulator of exopolysaccharide in *Myxococcus xanthus*. PeerJ 3, e747.

Otani, M., Tabata, J., Ueki, T., Sano, K., Inouye, S., 2001. Heat-shock-induced proteins from *Myxococcus xanthus*. J. Bacteriol. 183, 6282-6287.

Pak, J., Fire, A., 2007. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. Science 315, 241-244.

Pandey, S.P., Minesinger, B.K., Kumar, J., Walker, G.C., 2011. A highly conserved protein of unknown function in *Sinorhizobium meliloti* affects sRNA regulation similar to Hfq. Nucleic Acids Res. 39, 4691-4708.

Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J.C., Vogel, J., 2009. Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Mol. Microbiol. 74, 139-158.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C., Vogel, J., 2009. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat. Struct. Mol. Biol. 16, 840-846.

Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., Sontheimer, E.J., 2004. A Dicer-2-dependent 80S complex cleaves targeted mRNAs during RNAi in *Drosophila*. Cell 117, 83-94.

Rubin, E.J., Akerley, B.J., Novik, V.N., Lampe, D.J., Husson, R.N., Mekalanos, J.J., 1999. *In vivo* transposition of *mariner*-based elements in enteric bacteria and mycobacteria. Proc. Natl. Acad. Sci. USA 96, 1645-1650.

Shabalina, S.A., Koonin, E.V., 2008. Origins and evolution of eukaryotic RNA interference. Trends Ecol. Evol. 23, 578-587.

Sijen, T., Steiner, F.A., Thijssen, K.L., Plasterk, R.H., 2007. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science 315, 244-247.

Stefani, G., Slack, F.J., 2008. Small non-coding RNAs in animal development. Nat. Rev. Mol. Cell Biol. 9, 219-230.

Storz, G., Vogel, J., Wassarman, K.M., 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell 43, 880-891.

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.

Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., Zamore, P.D., 2004. RISC assembly defects in the *Drosophila* RNAi mutant armitage. Cell 116, 831-841.

Velicer, G.J., Kroos, L., Lenski, R.E., 1998. Loss of social behaviors by *Myxococcus xanthus* during evolution in an unstructured habitat. Proc. Natl. Acad. Sci. USA 95, 12376-12380.

Velicer, G.J., Raddatz, G., Keller, H., Deiss, S., Lanz, C., Dinkelacker, I., Schuster, S.C., 2006. Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. Proc. Natl. Acad. Sci. USA 103, 8107-8112.

Viegas, S.C., Silva, I.J., Saramago, M., Domingues, S., Arraiano, C.M., 2011. Regulation of the small regulatory RNA MicA by ribonuclease III: a target-dependent pathway. Nucleic Acids Res. 39, 2918-2930.

Vournakis, J.N., Efstratiadis, A., Kafatos, F.C., 1975. Electrophoretic patterns of deadenylylated chorion and globin mRNAs. Proc. Natl. Acad. Sci. USA 72, 2959-2963.

Waterhouse, P.M., Wang, M.B., Lough, T., 2001. Gene silencing as an adaptive defence against viruses. Nature 411, 834-842.

Wu, S.S., Kaiser, D., 1995. Genetic and functional evidence that Type IV pili are required for social gliding motility in *Myxococcus xanthus*. Mol. Microbiol. 18, 547-558.

Youderian, P., Burke, N., White, D.J., Hartzell, P.L., 2003. Identification of genes required for adventurous gliding motility in *Myxococcus xanthus* with the transposable element *mariner*. Mol. Microbiol. 49, 555-570.

Yu, Y.T., Yuan, X., Velicer, G.J., 2010. Adaptive evolution of an sRNA that controls *Myxococcus* development. Science 328, 993.

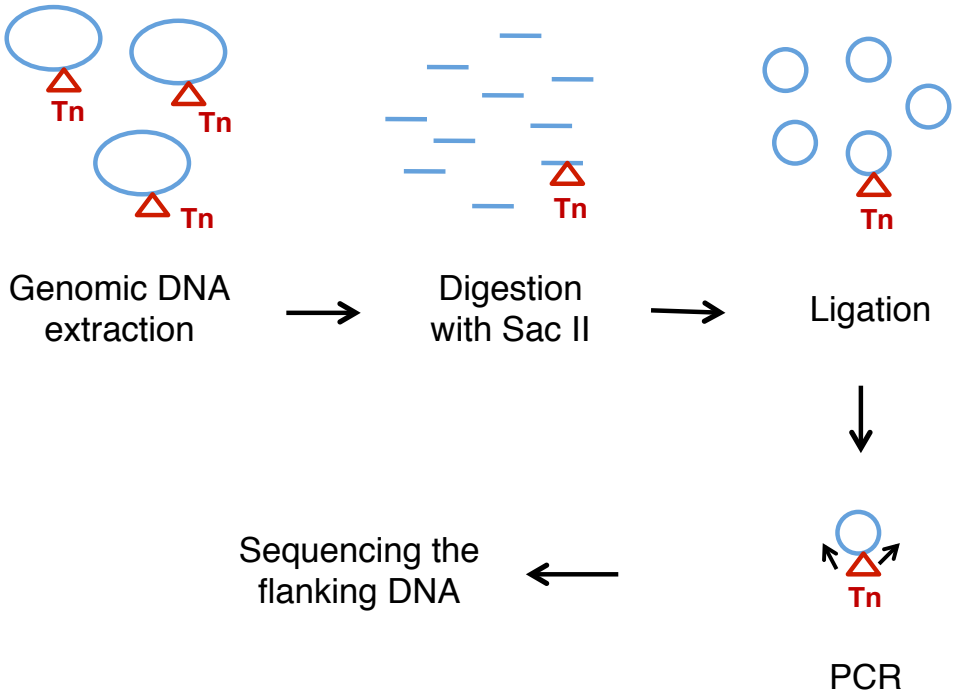


Figure 3.1 Illustration of the mapping method to identify insertion loci in transposon mutants. The big blue circles in the "Genomic DNA extraction" step represent chromosomal DNA of mutants; the small blue circles in the following steps indicate ligated DNA fragments. "Tn" refers to *Himar1*-derived transposon. The small arrows in the "PCR" step indicate the primers that hybridize the ends of transposon insertion.

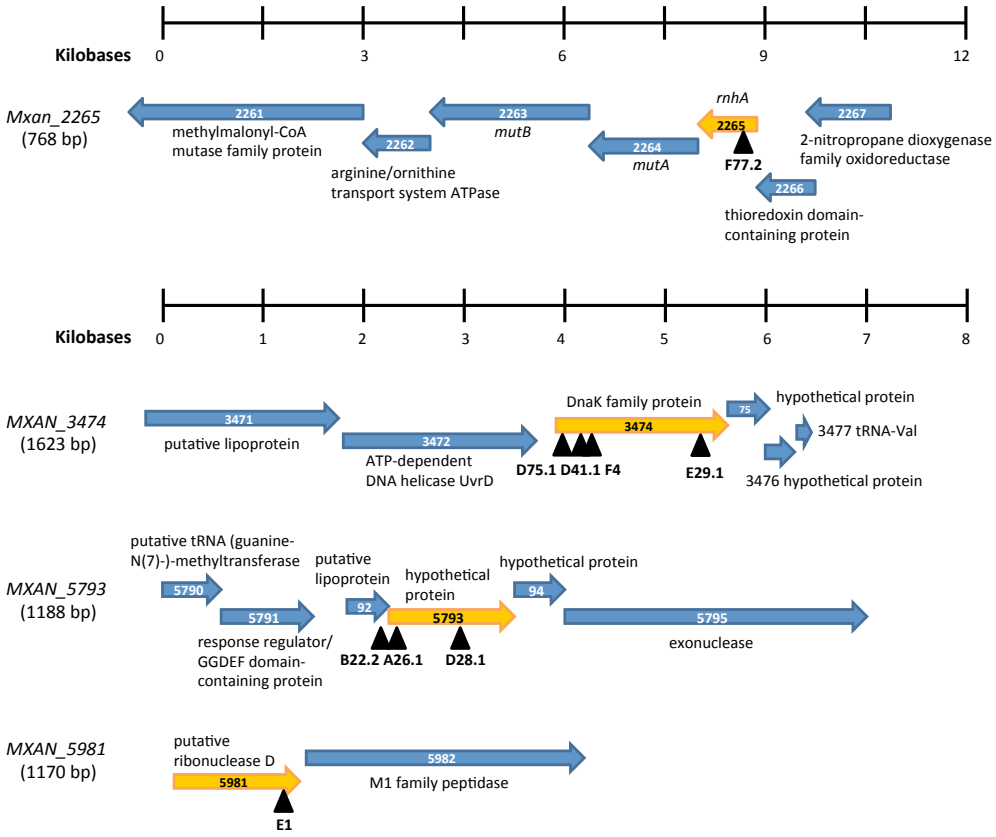


Figure 3.2 Transposon insertions in four genetic loci that restored developmental proficiency to the *M. xanthus* obligate cheater OC. The predicted loci and sequence lengths are shown in the left hand column. Transposon insertions are indicated as black triangles with mutant strain names below. Each arrow represents a gene and its orientation and length were drawn in proportion to the kilobase scales above. The annotated function of each gene is indicated either below or above the corresponding arrow.

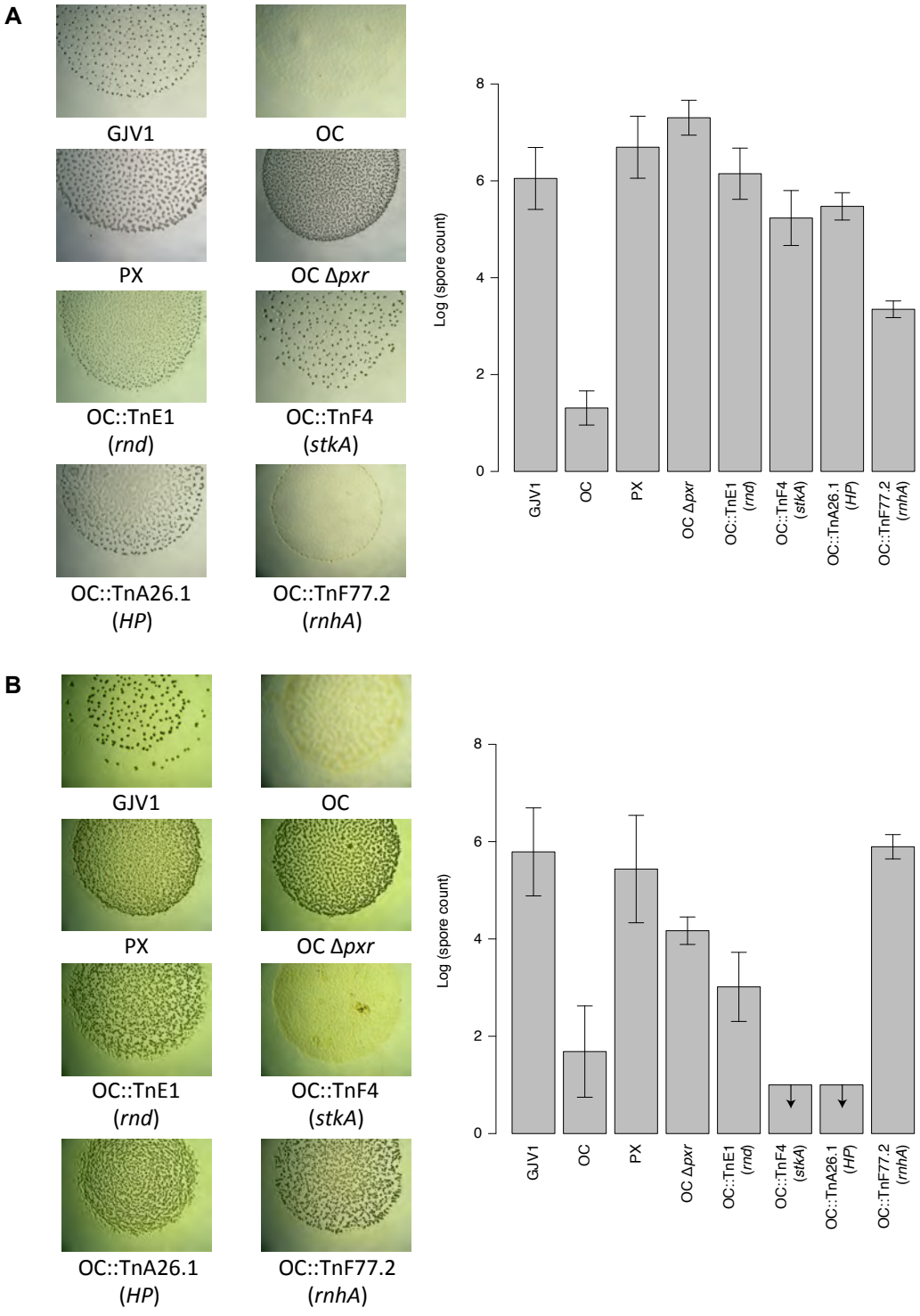


Figure 3.3 Transposon insertions in four loci restored developmental proficiency in OC. (A) Developmental phenotypes and spore production of representative mutants of each insertion locus on TPM hard agar. (B) Developmental phenotype and spore production of representative

mutants of each insertion locus on CF hard agar. Error bars indicate standard deviations. Arrows indicate that no spores were produced at the lower limit of detection.

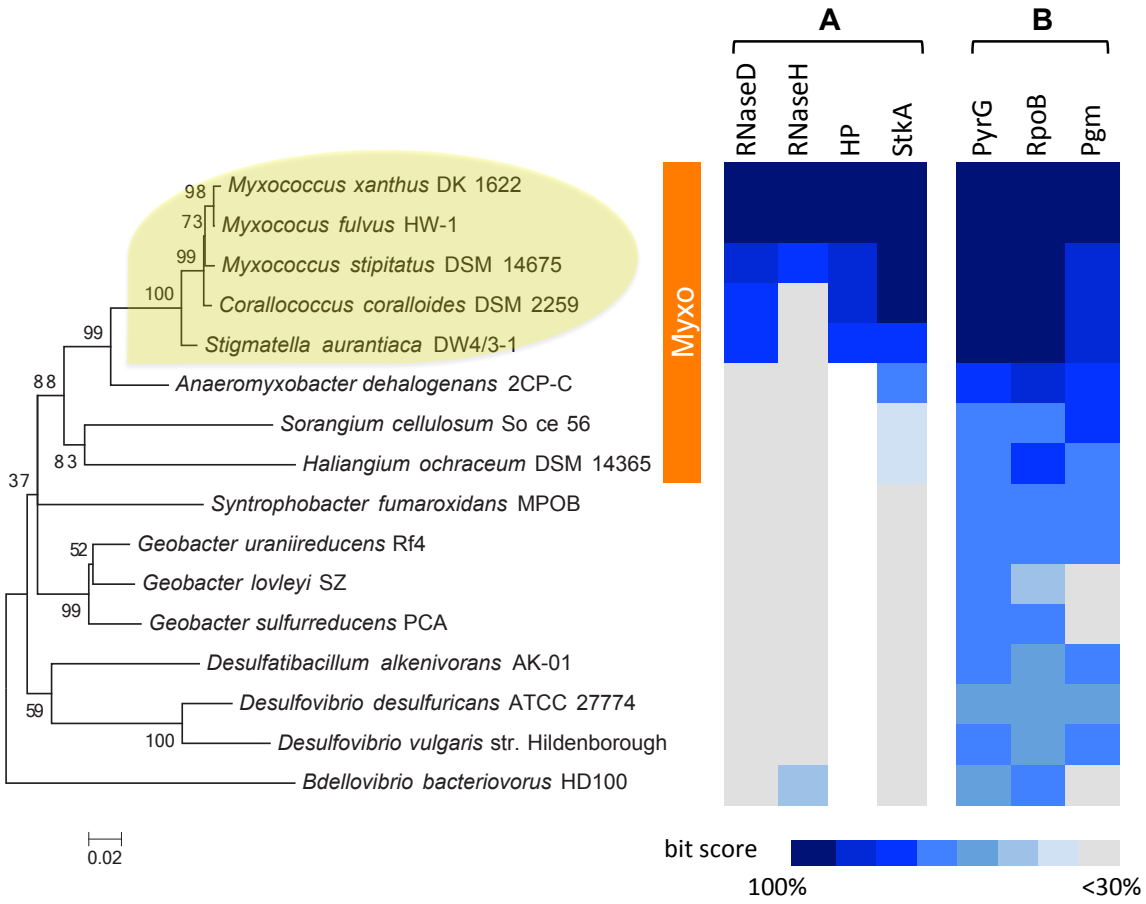
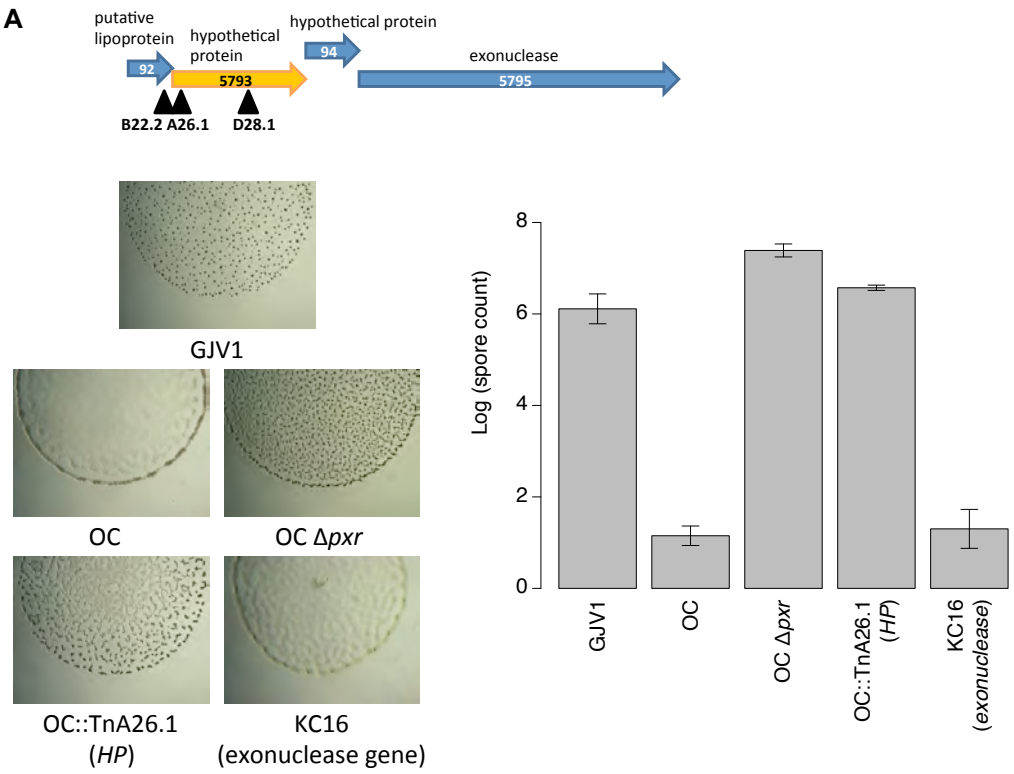


Figure 3.4 Phylogenetic of presence/absence and sequence similarity patterns of the four insertion loci in the delta-proteobacteria. The maximum likelihood phylogeny of the delta-proteobacteria based on the 16S rRNA gene is shown on the left. The bootstrap values (1000 replicates) are shown next to the branches. The scale bar shows 0.02 substitutions per site. The orange bar indicates the order Myxococcales and the light yellow oval marks the suborder Cystobacterineae where Pxr sRNA was found. The gene products in the group A are from the identified loci in this study; the ones in the group B are from the conserved loci used in the phylogenetic analyses in Chapter I. For each *M. xanthus* DK1622 query protein, the sequence conservation against each sequenced genome in the δ-proteobacteria is indicated by bit scoring system in the BlastP analysis. Values range from 100% similarity (dark blue), < 30% similarity but motifs with similar functions were found (grey) to 0% (white, no significant hits).


A
putative lipoprotein
hypothetical protein
hypothetical protein
exonuclease
92
5793
94
5795
B22.2 A26.1
D28.1
GJV1
OC
OC Δpxr
OC::TnA26.1 (HP)
KC16 (exonuclease gene)
Log (spore count)
0
2
4
6
8
GJV1
OC
OC Δpxr
OC::TnA26.1 (HP)
KC16 (exonuclease)

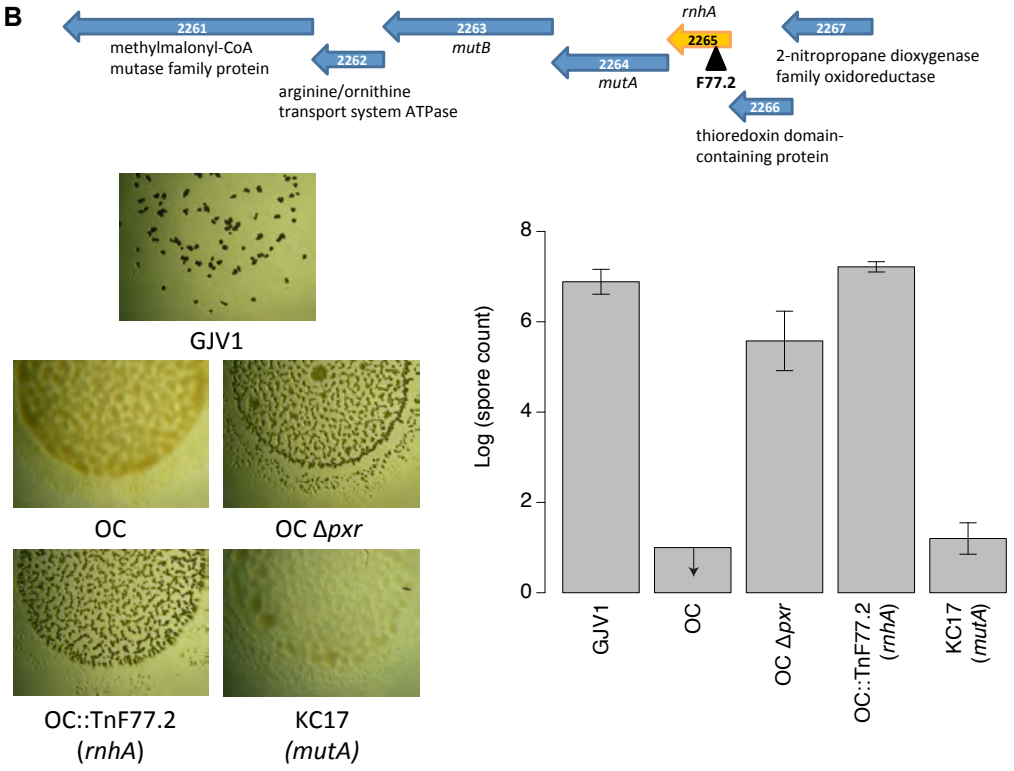

B
2261
methylmalonyl-CoA mutase family protein
2262
arginine/ornithine transport system ATPase
2263
mutB
2264
mutA
rnhA
2265
F77.2
2267
2-nitropropane dioxygenase family oxidoreductase
2266
thioredoxin domain-containing protein
GJV1
OC
OC Δpxr
OC::TnF77.2 (rnhA)
KC17 (mutA)
Log (spore count)
0
2
4
6
8
GJV1
OC
OC Δpxr
OC::TnF77.2 (rnhA)
KC17 (mutA)

Figure 3.5 PX-like phenotypes are caused by transposon interruptions in *HP* (*Mxan_5793*) and *rnhA* (*Mxan_2265*), not polar effects on the downstream genes. (A) The transposon interruption in *HP*, but not the plasmid integration in the downstream hypothetical protein (*Mxan_5795*), led to fruiting body formation and robust spore production on TPM hard agar. (B) The transposon interruption in *rnhA*, but not the plasmid integration in the downstream *mutA* gene (*Mxan_2264*), led to fruiting body formation and high spore production on CF hard agar. Error bars indicate standard deviations. Arrows indicate that no spores were produced at the lower limit of detection.

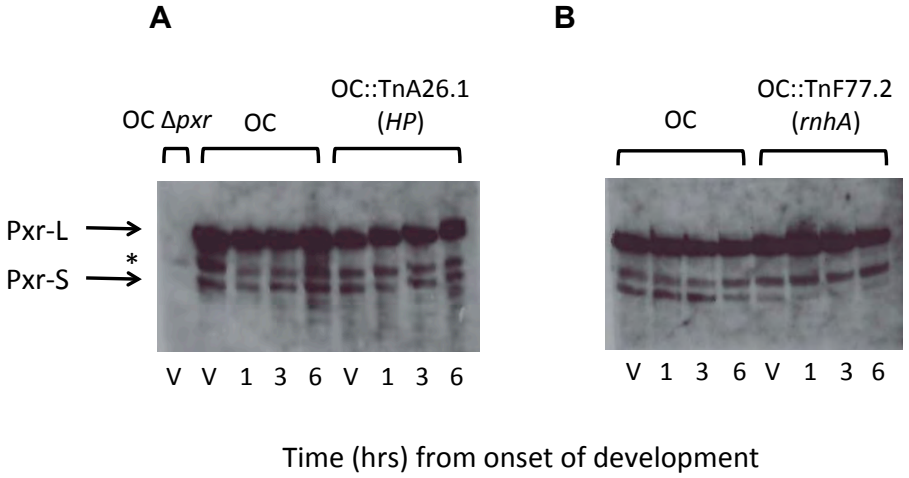


Figure 3.6 (A) Expression of Pxr-L and Pxr-S remains high in both vegetative growth and throughout 6 hours of starvation on TPM hard agar in the transposon mutant OC::TnA26.1 (with insertion in *HP*). (B) Accumulation of Pxr-S is reduced in the mutant OC::TnF77.2 (with insertion in *rnhA*) in both vegetative growth and throughout 6 hours of development on CF hard agar. The asterisk indicates the non-specific binding of the probe.

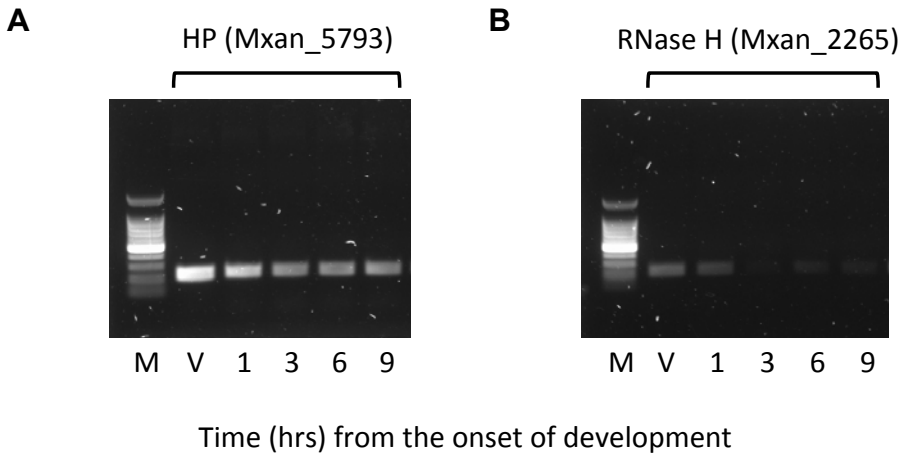


Figure 3.7 Reverse transcription PCR (RT-PCR) expression analyses of HP (A) and RNase H (B) in the wild-type GJV1 from vegetative growth throughout 9 hours of development on TPM plates. "M" refers to 100 bp marker. "V" refers to vegetatively growing cells harvested at the exponential phase. The numbers indicate the hours from the onset of development.

CHAPTER IV

Gene duplications of *pxr* in the genera *Archangium* and *Cystobacter* in the myxobacteria

I-Chen Chen[1,2], Sébastien Wielgoss[1] and Gregory J. Velicer[1,2]

[1] Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

[2] Institute of Integrative Biology (IBZ), ETH Zurich, CH-8092 Zurich, Switzerland

Abstract

Gene duplication is one of the major mechanisms of evolving new gene functions and its consequences have been demonstrated in many eukaryotic systems. Nevertheless, bacteria tend to remove non-beneficial genes and maintain relatively compact genome contents. Non-coding small RNAs (sRNAs) are a class of regulatory elements that have great importance in the function and evolution of bacterial genomes. Thus far, little is known about the roles that gene duplication may play in the evolution of bacterial sRNAs. Here, we used the Pacific Biosciences sequencing technology to sequence five myxobacterial strains in the genera *Archangium* and *Cystobacter,* in which the sRNA gene *pxr* was not previously detected (*Archangium* and *Cystobacter*) or had been previous found in tandem duplicates (*Cystobacter*). The whole-genome sequence data of these strains revealed new *pxr* duplications that had not been previously identified. Each sequenced strain contains five to seven copies of *pxr* in tandem in the conserved gene neighborhood between the σ^{54}-dependent response regulator *nla19* and a predicted NADH dehydrogenase gene. When copies within species are compared, some of them are highly divergent from others yet some are identical or nearly identical to each other, suggesting both old gene duplications, and recent gene duplications or concerted evolution of multiple copies. Future studies will be needed to elucidate the functional significance of maintaining multiple *pxr* copies and the sequence divergence among different *pxr* duplicates to shed light on the consequences of *pxr* duplications in the evolution of *Archangium* and *Cystobacter* genomes.

1. Introduction

Previously, our investigation of the phylogenetic distribution of the sRNA Pxr showed that, although Pxr is widely distributed in the suborder Cystobacterineae within the myxobacteria, it was not detected in the strains *Archangium gephyra* Arg 2, and *Cystobacter fuscus* Cbf 8 and Cbf 10 (Chen et al., 2014). Neither Northern hybridization nor attempted PCR using primers within the conserved regions of the gene neighborhood surrounding *pxr* revealed the presence of its homologs, suggesting evolutionary losses of *pxr* in these three strains. Nevertheless, absence of Pxr in Northern hybridization might be due to culture status at the time of RNA collection. For example, these species might express Pxr under only some environmental conditions, making it difficult to discern whether they possess Pxr or not with RNA collected from liquid cultures in the laboratory. Also, although all the detected *pxr* homologs are located in the intergenic region between the σ^{54}-dependent response regulator *nla19* and a predicted NADH dehydrogenase gene, this gene neighborhood might not be conserved in these three strains or the primer sequences used are not conserved in the respective genomes. Furthermore, three rapidly evolving tandem paralogs of *pxr* were found in the closely related strains *C. minor* Cbm6 and *C. virescens* Cbvi34, suggesting that highly divergent *pxr* variants that might be located in a different genomic region in Arg 2, Cbf 8 and Cbf 10 could go undetected by the methods used.

Next-generation sequencing (NGS) is widely used to address a range of questions in microbial research, such as identifying genome-wide polymorphisms among closely related bacterial strains, detecting mutations in laboratory evolved populations, or analyzing mutants isolated in genetic screens when a sequenced reference genome is available (MacLean et al., 2009;

Deatherage and Barrick, 2014). For *de novo* genome assembly, however, despite the low cost and increasing throughput provided by NGS technologies, many genomes are heavily fragmented into hundreds of contigs, which require costly and time-consuming manual gap-closing. Also, current NGS technologies generate relatively short reads (e.g. 150 – 300 bp for Illumina MiSeq and HiSeq), which poses a great challenge to assembly algorithms and makes resolving repeated sequences difficult (Phillippy et al., 2008; Koren et al., 2013). The PacBio RS sequencing technology recently released by Pacific Biosciences is a single-molecule, real-time DNA sequencer that provides solutions to the problems mentioned above. The current PacBio RS sequencer produces reads that are on average 9.5 Kb long, exceeding the size of the longest repeat in most bacteria and archaea (the rRNA operon, 5 to 7 Kb) (Treangen et al., 2009) and allowing closure of many microbial genomes (Koren et al., 2013). Moreover, it does not require amplification of source DNA before sequencing; despite the error rate of ~15% in single-pass sequence reads, the errors are randomly distributed and consensus sequences can achieve high accuracy (Carneiro et al., 2012; Koren et al., 2012).

In this study, we sequenced the genomes of *A. gephyra* Arg 2, *C. fuscus* Cbf 8 and Cbf 10 using a PacBio RS II sequencer to determine whether *pxr* is truly lost in these strains or escaped detection by previous methods. As controls, we also sequenced the genomes of *C. minor* Cbm6 and *C. virescens* Cbvi34 in which three tandem paralogs had been previously identified. Thus far, no complete genomes of myxobacterial strains closely related to *A. gephyra* Arg 2, *C. fuscus* Cbf 8 and Cbf 10, or *C. minor* Cbm6 and *C. virescens* Cbvi34 are available. Therefore it is not feasible to sequence these genomes with the Illumina sequencer and map the short reads to a reference genome. In contrast, the current PacBio sequencer is likely to return closed genomes

without gaps, or nearly closed genomes with a few gaps. Further, its long reads will allow identification of recent *pxr* duplications should such events happen in the genomes of our interest.

2. Materials and Methods

2.1. Bacterial strains and sequencing

Four strains in the genus *Cystobacter* (family Cystobacteraceae) and one strain in the genus *Archangium* (family Archangiaceae) obtained from Hans Reichenbach (Sproer et al., 1999) were examined in this study: *C. minor* Cbm 6, *C. violaceus* Cbvi 34, *C. fuscus* Cbf 8 and Cbf 10, and *A. gephyra* Arg 2. Although *A. gephyra* Arg 2 was morphologically classified in the family Archangiaceae, it formed a cluster with the other four *Cystobacter* strains based on the multilocus species phylogeny in the previous study (Chen et al., 2014). Frozen stocks of each strain were inoculated on CY hard (1.5%) agar (Shimkets et al., 2006) at 32 °C with a relative humidity of 90%. Cell cultures of the four *Cystobacter* strains were transferred to CAS liquid medium (Shimkets et al., 2006) at 32°C with constant shaking at 300 rpm to the mid-log phase in preparation for DNA extraction. The strain *A. gephyra* Arg 2 grew poorly in CAS liquid and the cells of this strain were collected directly from CY hard agar plates as they formed a sheet on the surface. DNA of each *Cystobacter* strain was extracted using Genomic-tip 200/G kit (Qiagen) following the manufacturer's protocol for Gram-negative bacteria. For *A. gephyra* Arg 2, we followed the protocol for tissue instead to break down the extracellular matrix Arg 2 formed on

agar surface. The genome sequences of these five strains were sequenced using the PacBio RS II system to a 100-fold coverage and assembled at the Functional Genomics Center Zurich (FGCZ).

2.2. Identification of pxr homologs and phylogenetic analyses

Previously identified *pxr* paralogs in *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 were used as query sequences in the BLASTn analysis implemented in CLC Genomics Workbench 7 (http://www.clcbio.com/) against sequenced genomes of *A. gephyra* Arg 2, *C. fuscus* Cbf 8 and Cbf 10 to detect the presence of *pxr* and also against genomes of *C. minor* Cbm 6 and *C. violaceus* Cbvi 34 to confirm the presence of tandem *pxr* paralogs. Both previously and newly identified *pxr* homologs were then aligned with MUSCLE implemented in MEGA version 5.0 (Tamura et al., 2011). Phylogenetic analyses were estimated using both maximum likelihood (ML) and Bayesian inference (BI). Here we used both the σ^{54} promoter region and the *pxr* stem-loop coding region for alignment to cover more polymorphic sites for better resolution in the gene trees, as some *pxr* duplicates were very similar to each other. The ML tree was estimated in MEGA 5.0 with 1,000 bootstrap replicates. The BI tree was reconstructed in MrBayes 3.2.5 using GTR **+**invgamma model with two parallel independent runs (Ronquist et al., 2012). The analysis was terminated when the average standard deviation in clade probabilities between the two runs was below 0.01. The likely ancestral sequence of *pxr* was inferred following the procedures in Hall (2011).

3. Results

3.1. Closed genomes of Archangium and Cystobacter strains

Using the third generation PacBio RS II sequencer, we obtained single circular contigs, i.e. closed genomes, for four of the five strains sequenced, i.e. *A. gephyra* Arg 2, *C. violaceus* Cbvi 34, *C. fuscus* Cbf 8 and 10, with sizes of 12.9 Mb, 13.3 Mb, 12.2 Mb and 11.8 Mb, respectively (Table 4.1). For the strain *C. minor* Cbm 6, we obtained two circular contigs with sizes of 13.2 Mb and 104.5 Kb. The former contig likely represents the chromosomal sequence of *C. minor* Cbm 6 itself and the latter might represent an associated plasmid. The genome sizes we found in the *Archangium-Cystobacter* clade are on average bigger than the previously sequenced genomes in the suborder Cystobacterineae, i.e. *Myxococcus xanthus* (Goldman et al., 2006), *M. fulvus* (Li et al., 2011), *M. stipitatus* (Huntley et al., 2013), *Corallococcus coralloides* (Huntley et al., 2012) and *Stigmatella aurantiaca* (Huntley et al., 2011), with sizes of 9.1 Mb, 9.0 Mb, 10.4 Mb, 10.0 Mb and 10.3 Mb, respectively.

3.2. pxr is present in all five strains examined

The whole-genome data disclosed previously undetected *pxr* copies in all five sequenced strains (Fig 4.1). In the three strains in which we did not detect *pxr* previously, we identified five copies of *pxr* in *A. gephyra* Arg 2, six copies in *C. fuscus* Cbf 8, and seven copies in *C. fuscus* Cbf 10, all arranged in tandem, oriented in the same direction and located in the conserved gene neighborhood between the σ^{54} dependent response regulator *nla19* and a predicted NADH dehydrogenase gene, as are single-copy *pxr* homologs in other species (Chen *et al.* 2014). In *A. gephyra* Arg 2, its *pxr* cluster is followed immediately downstream by an acetyltransferase gene.

Hereafter, we designate each *pxr* copy based on the order they appear in the intergenic region in each strain (e.g. the first copy of *pxr* in the intergenic region in *A. gephyra* Arg 2 is designated as *A. gephyra* Arg 2 *pxr-1*). In the strains *C. minor* Cbm 6 and *C. violaceus* Cbvi 34, in which three tandem copies were found before, the whole genome sequences uncovered two more copies after the first three in the same conserved gene neighborhood and we henceforth name these two copies *pxr-4* and *pxr-5* (prefaced by the respective strain name) in these two strains. In *C. minor* Cbm 6, the coding sequences of *pxr-2* and *pxr-4* are identical to each other except a polymorphic site (A/G) in the single-stranded region between the second and third stem-loops of Pxr, and its *pxr-3* and *pxr-5* are also identical. In *C. violaceus* Cbvi 34, its *pxr-2* and *pxr-4* share the same sequence, as do Cbvi 34 *pxr-3* and *pxr-5*. It is likely that when sequencing the PCR product of this intergenic region before, the reads produced by Sanger sequencing only read from the 5' end at *nla19* towards the end of *pxr-3*, and from the 3' end at the predicted NADH dehydrogenase gene towards the beginning of *pxr-5*. When aligning these reads together, *pxr-3* was aligned with *pxr-5* as they are identical to each other, causing *pxr-4* and *pxr-5* (or *pxr-3*) to remain unidentified.

In total, 28 copies of *pxr* were found in these five strains. Each *pxr* copy identified contains a predicted σ^{54} promoter sequence in the upstream region; however, for *pxr-3* and *pxr-4* in *C. minor* Cbm 6, there is a 5-nt deletion in the σ^{54} promoter region (Fig 4.2). Also, 42% of sites across these *pxr* copies are polymorphic (when including both the predicted σ^{54} promoter region and the stem-loop coding region of Pxr). When previously identified single-copy *pxr* homologs are considered, the proportion of polymorphic sites increases to 50%. The sequence alignment of all the *pxr* homologs examined here is illustrated in Figure 4.2.

3.3. Duplications of pxr in the Archangium-Cystobacter clade

The Bayesian gene tree of *pxr* was rooted with the *pxr* homolog from the most distantly related *Stigmatella aurantiaca* DW4/3-1 (Fig 4.3). The duplicated *pxr* homologs in the *Archangium*-*Cystobacter* clade are well separated from the single-copy *pxr* homologs in other myxobacterial species. Within the *Archangium*-*Cystobacter* clade, *pxr* homologs are clustered in groups and each group is well separated by long branches. The branching pattern of the ML tree is in general similar to the one of the Bayesian tree, although there are some differences towards the tips of the tree (Fig 4.4). For example, the relationships among homologs in *A. disciformis*, *M. flavescens*, *M. macrosporus*, *M. virescens* and *M. xanthus* are slightly different in the ML tree. Also, in the *Archangium*-*Cystobacter* clade, *pxr-1* in *C. fuscus* Cbf 8 and 10 are more closely related to *pxr-1* in *A. gephyra* Arg2. Nevertheless, the bootstrap vales for these groupings are low (below 50%).

We herein arbitrarily assign four groups (Groups A-D) in the Bayesian tree for illustration (Fig 4.3). Group A includes the first copies of *pxr* present in the intergenic region across all five *Archangium* and *Cystobacter* strains. The sequences of these copies have diverged from each other and the branching pattern among them is qualitatively congruent with the species phylogeny, suggesting that they share a common ancestor together before the strains diverged.

Group B contains *pxr-2* through *pxr-5* in *C. minor* Cbm 6 and *pxr-2* through *pxr-5* in *C. violaceus* Cbvi 34. We found that *pxr-2* and *pxr-4* in *C. minor* Cbm 6 are strongly clustered with

pxr-2 and *pxr-4* in *C. violaceus* Cbvi 34, and that *pxr-3* and *pxr-5* in *C. minor* Cbm 6 are strongly clustered with *pxr-3* and *pxr-5* in *C. violaceus* Cbvi 34, indicating that the *pxr-2*-*pxr-3*/*pxr-4*-*pxr-5* duplication occurred prior to the divergence of these two species.

Group C consists of *pxr-2* through *pxr-6* in *C. fuscus* Cbf 8 and *pxr-2* through *pxr-7* in Cbf 10 and these duplicates exhibit a high degree of sequence similarity. In *C. fuscus* Cbf 8, its *pxr* copies differ from each other by a few mutations; in *C. fuscus* Cbf 10, *pxr-2* through *pxr-7* are entirely identical to each other. It is likely that the ancestor shared by these two strains possessed five copies already (other than *pxr-1*). After the strains diverged, *C. fuscus* Cbf 10 acquired one more copy by mechanisms such as unequal crossing-over. The high sequence similarity of these copies also suggests that they may undergo homogenization by gene conversion, but the rate at which homogenization occurs may differ slightly between *C. fuscus* Cbf 8 and 10. Alternatively, but less likely, because *pxr-5* and *pxr-6* in *C. fuscus* Cbf 8 are more closely clustered with *pxr-2* through *pxr-7* in *C. fuscus* Cbf 10, it is possible that the ancestor of *pxr-5* and *pxr-6* in Cbf 8 was horizontally transferred to *C. fuscus* Cbf 10, followed by rapid duplications in *C. fuscus* Cbf 10.

In Group D, *pxr-2*, *pxr-3*, *pxr-4* and *pxr-5* in *A. gephyra* Arg 2 are clustered together, suggesting recent amplifications of *pxr* in this species. The alternative can be concerted evolution of these duplicated *pxr* sequences through gene conversion. Among these four copies, *pxr-2* is more divergent from others. It is likely that *pxr-2* accumulated more mutations and homogenization by gene conversion is not efficient enough anymore, therefore this copy diverged more rapidly.

4. Discussion

In summary, we discovered new *pxr* duplications from whole-genome sequence data of *Archangium* and *Cystobacter* species that had not been previously identified by other methods. We not only detected the presence of *pxr* in the three strains where *pxr* appeared to be absent with the methods used before, but also found more duplications in the two strains where *pxr* paralogs were found already. There are five to seven *pxr* copies per genome, encoded in tandem and oriented in the same direction in the same gene neighborhood conserved across species. Some *pxr* homologs from different species are more similar to each other, indicating old gene duplications in their common ancestors. Some homologs are identical or nearly identical to each other, suggesting recent gene duplications or concerted evolution by gene conversion.

Gene duplication is one of the major mechanisms for evolving new gene functions, and its consequences have been well characterized in many gene families in eukaryotes (Taylor and Raes, 2004). However, bacterial genomes are usually compact and there is a pervasive bias towards removing non-beneficial genes (Lawrence et al., 2001; Mira et al., 2001). Thus far, several examples of two or more non-coding small RNAs in the same bacteria have been documented, some with redundant regulatory functions and some without redundant regulatory functions (Caswell et al., 2014). The prevalence of multiple copies of *pxr* found in the *Archangium* and *Cystobacter* strains implies that it is unlikely that those are recent duplications without functional significance that haven't been purged by selection. The results here suggest the possibility that bacterial sRNA paralogs might evolve distinct beneficial functions and thus be maintained by selection. For example, *pxr-1* present in all five strains may diverge

functionally. Multiple similar *pxr* copies in the same strain here may reflect a requirement for cells to tightly regulate expression of certain genes. Further investigations on both individual and combined effects of *pxr* duplicates identified here will be needed to elucidate the evolutionary consequences of *pxr* duplications in this subclade in the myxobacteria.

Finally, the PacBio sequencing technology we applied satisfyingly resulted in five complete genomes out of the five strains sequenced. Nonetheless, PacBio is error-prone with single base insertions or deletions (indels), and the whole-genome sequence data of the two natural isolates of the model myxobacterium *M. xanthus* generated by Illumina and Pacbio technologies have shown that the sequences generated by Pacbio can contain up to 1000 single base indels per genome (Sébastien Wielgoss, personal communication). Although it is unlikely that such instances will affect the overall results presented here (the *pxr* sequence lengths per genome we examined ranges from 775 to 1085 nt, and the probability of having indels in these sequences is less than 0.1 nt), indels can cause frame shifts when interpreting protein-coding regions and greatly bias genome annotations. Ongoing work includes sequencing the five strains with the Illumina sequencer to correct potential indels in the PacBio sequence data. The finished genomes of these strains will allow future genomic studies on the evolution of Pxr regulatory network as well as the consequences of *pxr* duplications in these genomes.

Table 4.1 Total assembly sizes for five sequenced genomes

Taxonomic description	Strain	Assembly bp	No. of contigs
Archangium gephyra	Arg 2	12,972,443	1
Cystobacter minor	Cbm 6	13,221,510	2
		104,518[a]	
Cystobacter violaceus	Cbvi 34	13,326,447	1
Cystobacter fuscus	Cbf 8	12,220,439	1
Cystobacter fuscus	Cbf 10	11,846,083	1

[a]This is a circular contig itself.

References

Carneiro, M.O., Russ, C., Ross, M.G., Gabriel, S.B., Nusbaum, C., DePristo, M.A., 2012. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics 13, 375.

Caswell, C.C., Oglesby-Sherrouse, A.G., Murphy, E.R., 2014. Sibling rivalry: related bacterial small RNAs and their redundant and non-redundant roles. Front. Cell. Infect. Microbial. 4, 151.

Chen, I.C., Griesenauer, B., Yu, Y.T., Velicer, G.J., 2014. A recent evolutionary origin of a bacterial small RNA that controls multicellular fruiting body development. Mol. Phylogenet. Evol. 73, 1-9.

Deatherage, D.E., Barrick, J.E., 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Method. Mol. Biol. 1151, 165-188.

Goldman, B.S., Nierman, W.C., Kaiser, D., Slater, S.C., Durkin, A.S., Eisen, J.A., Ronning, C.M., Barbazuk, W.B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H.S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S.A., Vaudin, M., Wiegand, R., Kaplan, H.B., 2006. Evolution of sensory complexity recorded in a myxobacterial genome. Proc. Natl. Acad. Sci. USA 103, 15200-15205.

Hall, B.G., 2011. Phylogenetic Trees Made Easy: A How-To Manual, 4th Edition. Sinauer Associates, Sunderland, Massachusetts.

Huntley, S., Hamann, N., Wegener-Feldbrugge, S., Treuner-Lange, A., Kube, M., Reinhardt, R., Klages, S., Muller, R., Ronning, C.M., Nierman, W.C., Sogaard-Andersen, L., 2011.

Comparative genomic analysis of fruiting body formation in Myxococcales. Mol. Biol. Evol. 28, 1083-1097.

Huntley, S., Kneip, S., Treuner-Lange, A., Sogaard-Andersen, L., 2013. Complete genome sequence of *Myxococcus stipitatus* strain DSM 14675, a fruiting myxobacterium. Genome Announcements 1, e0010013.

Huntley, S., Zhang, Y., Treuner-Lange, A., Kneip, S., Sensen, C.W., Sogaard-Andersen, L., 2012. Complete genome sequence of the fruiting myxobacterium *Corallococcus coralloides* DSM 2259. J. Bacteriol. 194, 3012-3013.

Koren, S., Harhay, G.P., Smith, T.P., Bono, J.L., Harhay, D.M., McVey, S.D., Radune, D., Bergman, N.H., Phillippy, A.M., 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol. 14, R101.

Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., Adam, M.P., 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. Nat. Biotechnol. 30, 693-700.

Lawrence, J.G., Hendrix, R.W., Casjens, S., 2001. Where are the pseudogenes in bacterial genomes? Trends Microbiol. 9, 535-540.

Li, Z.F., Li, X., Liu, H., Liu, X., Han, K., Wu, Z.H., Hu, W., Li, F.F., Li, Y.Z., 2011. Genome sequence of the halotolerant marine bacterium *Myxococcus fulvus* HW-1. J. Bacteriol. 193, 5015-5016.

MacLean, D., Jones, J.D., Studholme, D.J., 2009. Application of 'next-generation' sequencing technologies to microbial genetics. Nat. Rev. Microbiol. 7, 287-296.

Mira, A., Ochman, H., Moran, N.A., 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet. 17, 589-596.

Phillippy, A.M., Schatz, M.C., Pop, M., 2008. Genome assembly forensics: finding the elusive mis-assembly. Genome Biol. 9, R55.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539-542.

Shimkets, L., Dworkin, M., Reichenbach, H., 2006. The Myxobacteria. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, Stackenbrandt, E. (Eds.), The Prokaryotes, Springer, New York.

Sproer, C., Reichenbach, H., Stackebrandt, E., 1999. The correlation between morphological and phylogenetic classification of myxobacteria. Int. J. Syst. Bacteriol. 49 Pt 3, 1255-1262.

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.

Taylor, J.S., Raes, J., 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38, 615-643.

Treangen, T.J., Abraham, A.L., Touchon, M., Rocha, E.P., 2009. Genesis, effects and fates of repeats in prokaryotic genomes. FEMS Microbiol. Rev. 33, 539-571.

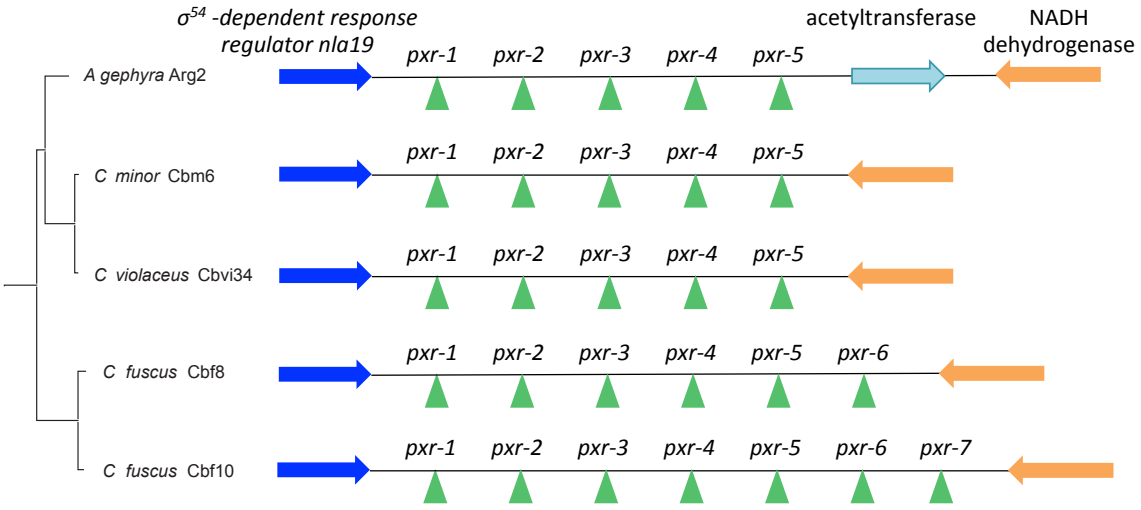
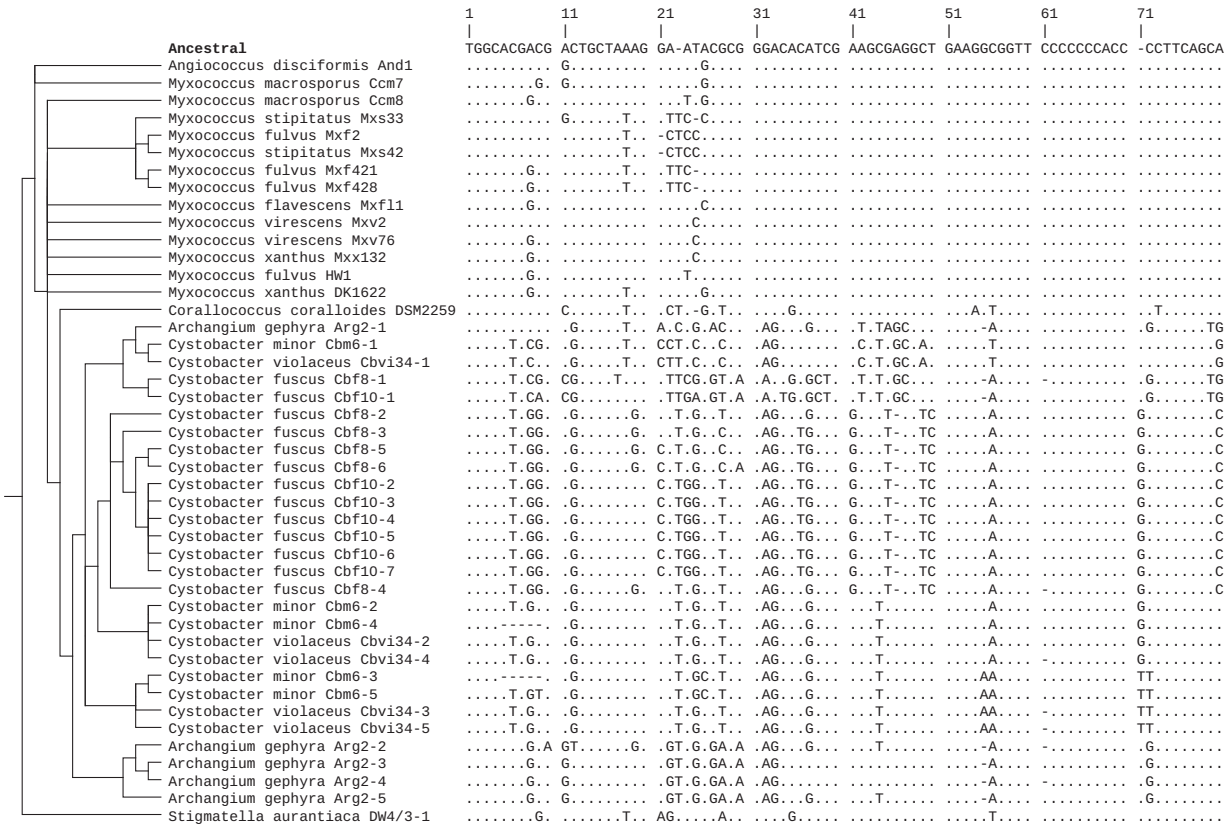


Figure 4.1 Duplications of *pxr* between the σ^{54} dependent response regulator *nla19* and the NADH dehydrogenase gene in the *Cystobacter* clade. The species phylogeny is shown on the left. The numerical designation for each *pxr* copy represents the order they appear in the intergenic region in each species, not an indication of the order they evolved. Genes and intergenic regions are not drawn to scale.


1 11 21 31 41 51 61 71
Ancestral TGGCACGACG ACTGCTAAAG GA-ATACGCG GGACACATCG AAGCGAGGCT GAAGGCGGTT CCCCCCCACC -CCTTCAGCA
Angiococcus disciformis And1 ..... G..... .....G.... ....
Myxococcus macrosporus Ccm7 .....G. G..... .....G.... ....
Myxococcus macrosporus Ccm8 .....G.. ....T.G.... ....
Myxococcus stipitatus Mxs33 ..... G.....T.. .TTC-C.... ....
Myxococcus fulvus Mxf2 ....T.. -CTCC..... ....
Myxococcus stipitatus Mxs42 ....T.. -CTCC..... ....
Myxococcus fulvus Mxf421 .....G.. .....T.. .TTC-..... ....
Myxococcus fulvus Mxf428 .....G.. .....T.. .TTC-..... ....
Myxococcus flavescens Mxfl1 .....G.. ....C.... ....
Myxococcus virescens Mxv2 ....C..... ....
Myxococcus virescens Mxv76 .....G.. ....C..... ....
Myxococcus xanthus Mxx132 .....G.. ....C..... ....
Myxococcus fulvus HW1 .....G.. ....T..... ....
Myxococcus xanthus DK1622 .....G.. .....T.. .....G.... ....
Corallococcus coralloides DSM2259 ..... C.....T.. .CT.-G.T.. ....G..... ....A.T.... ....T.....
Archangium gephyra Arg2-1 ..... .G.....T.. A.C.G.AC.. .AG...G... .T.TAGC... ....-A.... ..... .G.....TG
Cystobacter minor Cbm6-1 .....T.CG. .G.....T.. CCT.C..C.. .AG..... .C.T.GC.A. .....T.... ....G
Cystobacter violaceus Cbvi34-1 .....T.C.. .G.....T.. CTT.C..C.. .AG..... .C.T.GC.A. .....T.... ....G
Cystobacter fuscus Cbf8-1 .....T.CG. CG....T... .TTCG.GT.A .A..G.GCT. .T.T.GC... ....-A.... -..... .G.....TG
Cystobacter fuscus Cbf10-1 .....T.CA. CG..... .TTGA.GT.A .A.TG.GCT. .T.T.GC... ....-A.... ..... .G.....TG
Cystobacter fuscus Cbf8-2 .....T.GG. .G.....G. ..T.G..T.. .AG...G... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf8-3 .....T.GG. .G.....G. ..T.G..C.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf8-5 .....T.GG. .G.....G. C.T.G..C.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf8-6 .....T.GG. .G.....G. C.T.G..C.A .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf10-2 .....T.GG. .G..... C.TGG..T.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf10-3 .....T.GG. .G..... C.TGG..T.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf10-4 .....T.GG. .G..... C.TGG..T.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf10-5 .....T.GG. .G..... C.TGG..T.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf10-6 .....T.GG. .G..... C.TGG..T.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf10-7 .....T.GG. .G..... C.TGG..T.. .AG..TG... G...T-..TC .....A.... ..... G.....C
Cystobacter fuscus Cbf8-4 .....T.GG. .G.....G. ..T.G..T.. .AG...G... G...T-..TC .....A.... -..... G.....C
Cystobacter minor Cbm6-2 .....T.G.. .G..... ..T.G..T.. .AG...G... ...T..... .....A.... ..... G.....
Cystobacter minor Cbm6-4 .....-----. .G..... ..T.G..T.. .AG...G... ...T..... .....A.... ..... G.....
Cystobacter violaceus Cbvi34-2 .....T.G.. .G..... ..T.G..T.. .AG...G... ...T..... .....A.... ..... G.....
Cystobacter violaceus Cbvi34-4 .....T.G.. .G..... ..T.G..T.. .AG...G... ...T..... .....A.... -..... G.....
Cystobacter minor Cbm6-3 .....-----. .G..... ..T.GC.T.. .AG...G... ...T..... ....AA.... ..... TT.....
Cystobacter minor Cbm6-5 .....T.GT. .G..... ..T.GC.T.. .AG...G... ...T..... ....AA.... ..... TT.....
Cystobacter violaceus Cbvi34-3 .....T.G.. .G..... ..T.G..T.. .AG...G... ...T..... ....AA.... -..... TT.....
Cystobacter violaceus Cbvi34-5 .....T.G.. .G..... ..T.G..T.. .AG...G... ...T..... ....AA.... -..... TT.....
Archangium gephyra Arg2-2 .....G.A GT.....G. .GT.G.GA.A .AG...G... ...T..... ....-A.... -..... .G.....
Archangium gephyra Arg2-3 .....G.. G..... .GT.G.GA.A .AG..... ....-A.... ..... .G.....
Archangium gephyra Arg2-4 .....G.. G..... .GT.G.GA.A .AG..... ....-A.... -..... .G.....
Archangium gephyra Arg2-5 .....G.. G..... .GT.G.GA.A .AG...G... ...T..... ....-A.... ..... .G.....
Stigmatella aurantiaca DW4/3-1 .....G. .....T.. AG.....A.. ....G..... ....T.... ....

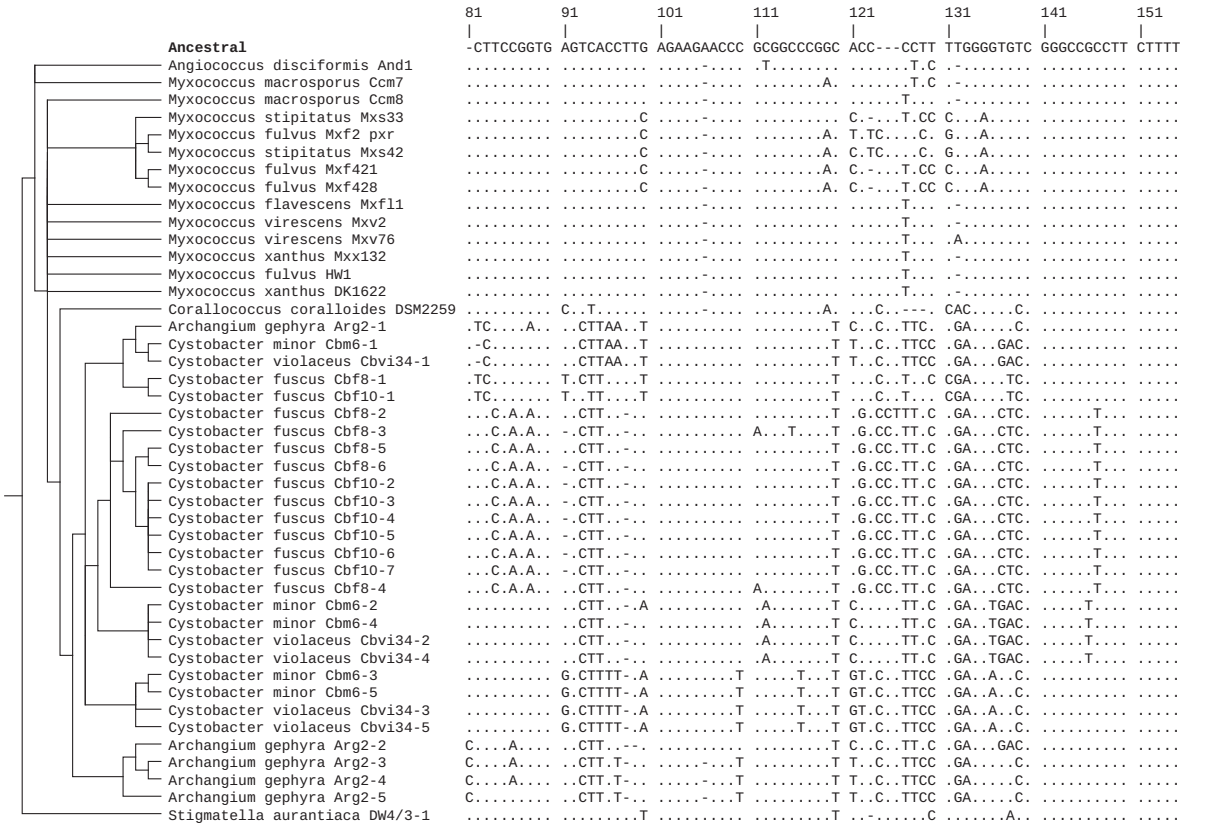

81 91 101 111 121 131 141 151
Ancestral -CTTCCGGTG AGTCACCTTG AGAAGAACCC GCGGCCCGGC ACC---CCTT TTGGGGTGTC GGGCCGCCTT CTTTT
Angiococcus disciformis And1 ....-.... .T..... .....T.C .-..... ....
Myxococcus macrosporus Ccm7 ....-.... .....A. .....T.C .-..... ....
Myxococcus macrosporus Ccm8 ....-.... ....T... .-..... ....
Myxococcus stipitatus Mxs33 ....C .....-.... ..... C.-...T.CC C...A..... ....
Myxococcus fulvus Mxf2 pxr ....C .....-.... .....A. T.TC....C. G...A..... ....
Myxococcus stipitatus Mxs42 ....C .....-.... .....A. C.TC....C. G...A..... ....
Myxococcus fulvus Mxf421 ....C .....-.... .....A. C.-...T.CC C...A..... ....
Myxococcus fulvus Mxf428 ....C .....-.... .....A. C.-...T.CC C...A..... ....
Myxococcus flavescens Mxfl1 ....-.... ....T... .-..... ....
Myxococcus virescens Mxv2 ....-.... ....T... .-..... ....
Myxococcus virescens Mxv76 ....-.... ....T... .A..... ....
Myxococcus xanthus Mxx132 ....-.... ....T... .-..... ....
Myxococcus fulvus HW1 ....-.... ....T... .-..... ....
Myxococcus xanthus DK1622 ....-.... ....T... .-..... ....
Corallococcus coralloides DSM2259 ..... C..T..... .....-.... ....A. ...C..---. CAC.....C. ....
Archangium gephyra Arg2-1 .TC....A.. ..CTTAA..T ....T C..C..TTC. .GA.....C. ....
Cystobacter minor Cbm6-1 .-C..... ..CTTAA..T ....T T..C..TTCC .GA...GAC. ....
Cystobacter violaceus Cbvi34-1 .-C..... ..CTTAA..T ....T T..C..TTCC .GA...GAC. ....
Cystobacter fuscus Cbf8-1 .TC..... T.CTT....T ....T ...C..T..C CGA....TC. ....
Cystobacter fuscus Cbf10-1 .TC..... T..TT....T ....T ...C..T... CGA....TC. ....
Cystobacter fuscus Cbf8-2 ...C.A.A.. ..CTT..-.. ....T .G.CCTTT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf8-3 ...C.A.A.. -.CTT..-.. .... A...T....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf8-5 ...C.A.A.. ..CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf8-6 ...C.A.A.. -.CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf10-2 ...C.A.A.. -.CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf10-3 ...C.A.A.. -.CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf10-4 ...C.A.A.. -.CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf10-5 ...C.A.A.. -.CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf10-6 ...C.A.A.. -.CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf10-7 ...C.A.A.. -.CTT..-.. ....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter fuscus Cbf8-4 ...C.A.A.. ..CTT..-.. .... A.....T .G.CC.TT.C .GA...CTC. ......T... .....
Cystobacter minor Cbm6-2 ....CTT..-.A .... .A.....T C.....TT.C .GA..TGAC. .....T.... .....
Cystobacter minor Cbm6-4 ....CTT..-.. .... .A.....T C.....TT.C .GA..TGAC. .....T.... .....
Cystobacter violaceus Cbvi34-2 ....CTT..-.. .... .A.....T C.....TT.C .GA..TGAC. .....T.... .....
Cystobacter violaceus Cbvi34-4 ....CTT..-.. .... .A.....T C.....TT.C .GA..TGAC. .....T.... .....
Cystobacter minor Cbm6-3 ..... G.CTTTT-.A ....T...T GT.C..TTCC .GA..A..C. ....
Cystobacter minor Cbm6-5 ..... G.CTTTT-.A ....T...T GT.C..TTCC .GA..A..C. ....
Cystobacter violaceus Cbvi34-3 ..... G.CTTTT-.A ....T...T GT.C..TTCC .GA..A..C. ....
Cystobacter violaceus Cbvi34-5 ..... G.CTTTT-.A ....T...T GT.C..TTCC .GA..A..C. ....
Archangium gephyra Arg2-2 C....A.... ..CTT..--. ....T C..C..TT.C .GA...GAC. ....
Archangium gephyra Arg2-3 C....A.... ..CTT.T-.. .....-...T ....T T..C..TTCC .GA.....C. ....
Archangium gephyra Arg2-4 C....A.... ..CTT.T-.. .....-...T ....T T..C..TTCC .GA.....C. ....
Archangium gephyra Arg2-5 C..... ..CTT.T-.. .....-...T ....T T..C..TTCC .GA.....C. ....
Stigmatella aurantiaca DW4/3-1 ....T ....T ..-.....C .....A.. ....

Figure 4.2 Nucleotide sequences of *pxr* homologs from the σ^{54} promoter region to the *pxr* stem-loop coding region at all sites. A dot indicates a perfect match to the inferred ancestor at the internal node shared by the non-*Stigmatella* homologs, whereas letters show nucleotide differences.

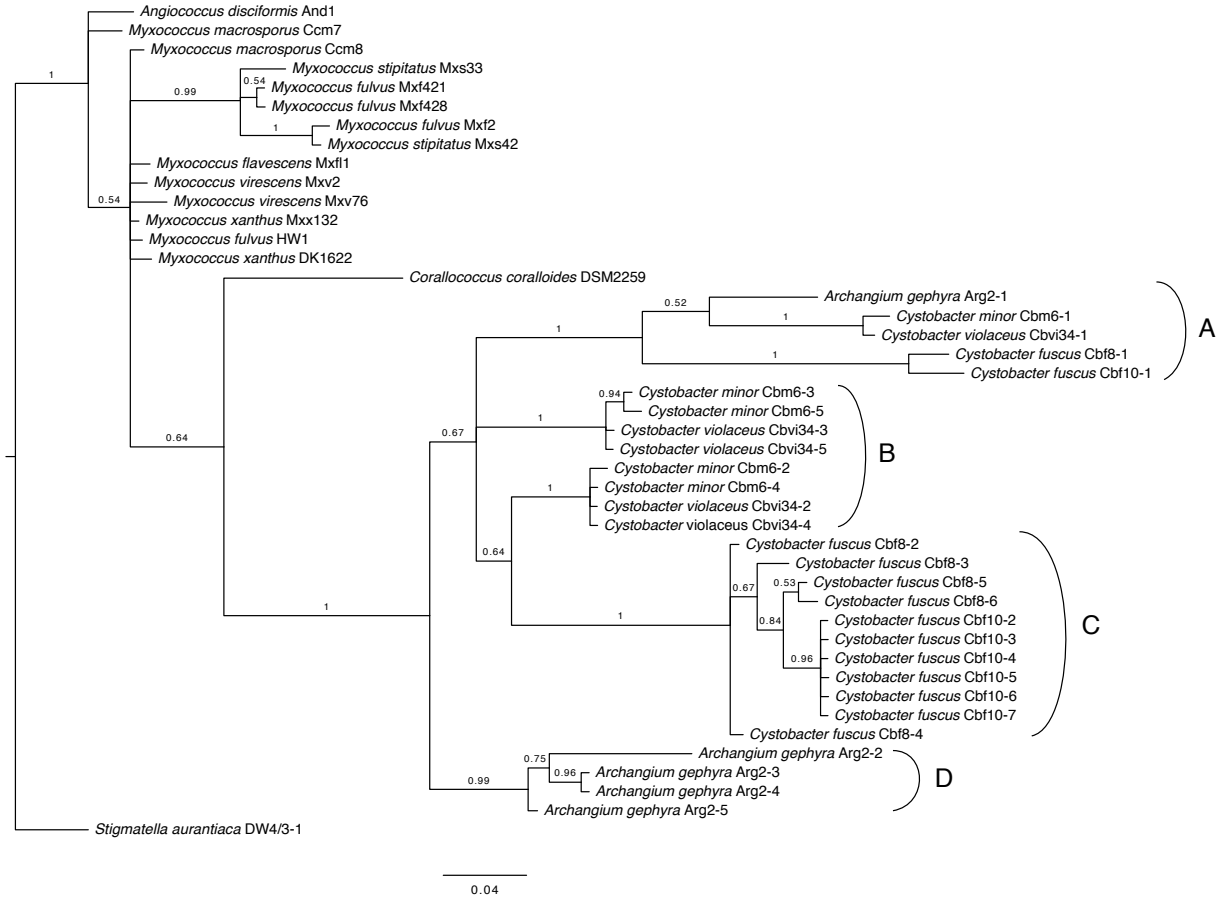


Figure 4.3 Bayesian gene tree of *pxr* including both the predicted σ^{54} promoter region and the stem-loop coding region of Pxr. The arbitrary groups A-D are assigned for illustration (see text). The tree is rooted with the homolog from the most distantly related species *Stigmatella aurantiaca*. Posterior probabilities are shown next to the nodes. The scale bar indicates 0.04 substitutions per site.

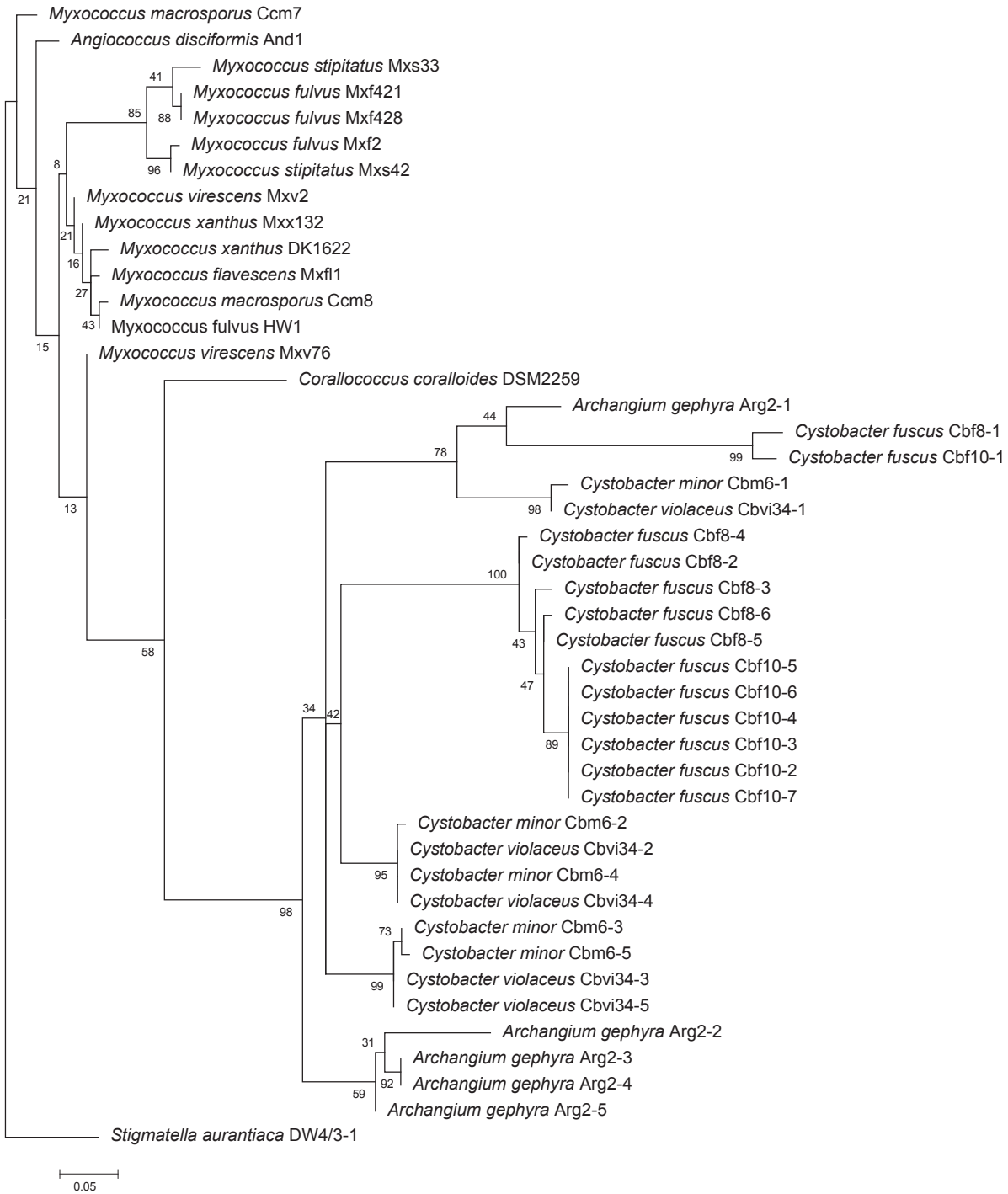


Figure 4.4 ML gene tree of *pxr* based on both previously and newly identified homologs. Bootstrap values are indicated at the nodes. The scale bar shows 0.05 substitutions per site.

I-CHEN (KIMBERLY) CHEN

Indiana University, Department of Biology, 1001 E. 3rd Street, Bloomington, IN 47405, USA

icchen@indiana.edu

Current address:

Institute of Integrative Biology (IBZ), ETH Zurich, CH-8092, Zurich, Switzerland

+41 44 633 8273 ◆ kimberly.chen@env.ethz.ch

EDUCATION AND EMPLOYMENT

2012 - 2015	Ph.D., Department of Biology, Indiana University, USA Program in Evolution, Ecology and Behavior Advisor: Gregory Velicer
2008 - 2015	Visiting Scientist, ETH Zurich, Switzerland Evolutionary Biology Group (Gregory Velicer)
2006 - 2007	M.S., Division of Biology, Imperial College London, UK Program in Ecology, Evolution and Conservation Advisor: Timothy Barraclough
2003 - 2006	Research Assistant Academia Sinica, Biodiversity Research Center, Taipei, Taiwan
1999 - 2003	B.S., National Taiwan University, Taipei, Taiwan Forest Resources Conservation

PUBLICATIONS

Chen, I.-C.K.,Velicer, G.J., and Yu, Y.-T.N. Functional evolution of the small RNA Pxr that controls multicellular fruiting body development in the myxobacteria. *In prep*.

Chen, I.-C.K., Griesenauer, B., Yu, Y.-T.N., and Velicer, G.J. (2014). A recent evolutionary origin of a bacterial small RNA that controls multicellular fruiting body development. Molecular Phylogenetics and Evolution *73*, 1-9.

Yu, Y.-T.N., **Chen, I.-C.K.**, and Velicer G.J. The role of a DnaK homolog in sRNA-mediated multicellular development in the myxobacteria. *In prep.*

Mendes-Soares, H., **Chen, I.-C.K.**, Fitzpatrick, K., and Velicer, G.J. (2014) Chimeric load among sympatric social bacteria increases with genotype richness. Proceedings of the Royal Society B: Biological Sciences *281*(1781), 20140285.

Swanstrom, J., **Chen, K.**, Castillo, K., Barraclough, T., and Fontaneto, D. (2010) Testing for evidence of inefficient selection in bdelloid rotifers: do sample size and habitat differences matter? Hydrobiologia *662*(1): 19-25.

Fontaneto, D., Barraclough, T., **Chen, K.,** Ricci, C., and Herniou, E. (2008). Molecular evidence for broad-scale distributions in bdelloid rotifers: everything is not everywhere but most things are very widespread. Molecular Ecology, *17*(13): 3136-3146.

FELLOWSHIPS AND AWARDS

2011 Floyd Summer Fellowship in Microbiology, Indiana University
2010 Floyd Summer Fellowship in Microbiology, Indiana University
2010 Indiana University Fellowship (1-semester tuition and stipend)

LAB TECHNIQUES AND COMPUTING SKILLS

DNA/RNA/plasmid extraction, Southern/Northern blots, PCR, Reverse-Transcription PCR, DNA sequencing, DNA cloning, allele exchange, genetic knockouts, bacterial transformation, bacterial culture, phenotypic assays
Microbial genome assembly and annotation, phylogenetic analyses (e.g. MEGA, MrBayes and PAML), R, Python

TEACHING AND MENTORSHIP

2011 - 2012 Mentor to undergraduate student Brad Griesenauer (IU biotechnology major), project on reconstructing molecular species phylogeny in the myxobacteria.
2008 - 2011 Associate Instructor, Department of Biology, Indiana University
Courses taught: Evolution and Diversity, Biological Mechanisms and Biology Lab

ACTIVITIES

2012 Organizer of the book club "The ecology of adaptive radiation" by Dolph Schluter, Institute of Integrative Biology (IBZ), ETH Zurich

PROFESSIONAL PRESENTATIONS

2014 15th International Symposium on Microbial Ecology (Seoul, Korea)
2013 Gordon Research Conference - Microbial Population Biology (New Hampshire, USA)
2013 Society for Molecular Biology & Evolution Annual Meeting (Chicago, USA)
2012 Evolution (Ottawa, Canada)
2012 Indiana University Women in Science Research Conference (Indiana, USA)
2011 Annual Conference on the Biology of Myxobacteria (New York, USA)
2011 Indiana University EEB Brown Bag Seminar (Indiana, USA)