Distinct biofilm regulatory strategies confer host specificity in *Vibrio fischeri*

by

Katherine M. Bultman

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 11/20/2023

The dissertation is approved by the following members of the Final Oral Committee:

Mark J. Mandel, Associate Professor, Medical Microbiology and Immunology

John-Demian Sauer, Associate Professor, Medical Microbiology and Immunology

Garret Suen, Professor, Bacteriology

Tu Anh Huyhn, Assistant Professor, Food Science

Caitlin Pepperell, Associate Professor, Medical Microbiology and Immunology

ABSTRACT

During the initial stages of colonization of both plant and animal hosts, bacteria assemble into biofilm structures or suspended aggregates. To regulate biofilm formation, bacteria use a variety of methods including secondary messengers and two component signaling. The specific regulatory mechanisms for biofilm formation can be a key determinant of specificity in host-microbe interactions. *Vibrio fischeri* produces a host-specific exopolysaccharide, termed the symbiosis polysaccharide (Syp), that is required for aggregate formation during the initial stages of light organ colonization. in the well-studied Hawaiian isolate ES114, previous work had suggested that the ability of *V. fischeri* to colonize the squid host was dependent on the presence of the positive Syp regulator RscS. However, two Mediterranean squid isolates, SA1 and SR5, were found to lack RscS but maintain the ability to colonize their native squid hosts in addition to the Hawaiian bobtail squid (*Euprymna scolopes*). In this dissertation, I elucidate the distinct biofilm regulatory strategies across *V. fischeri* isolates and how the diversity in the ancestral isolate SR5 leads to biofilm formation and squid colonization independent of RscS. Further characterization of the regulatory strategy in SR5 revealed novel regulators within the biofilm pathway. This work expands our knowledge of the diversity in the regulatory mechanisms of biofilm formation for the establishment of beneficial host-microbe interactions and lays the groundwork for future studies on the evolution of biofilm regulation for host colonization specificity.

ACKNOWLEDGMENTS

I am so appreciative to all the people who have gotten me to where I am today. I would not be the person and scientist I am without you. A thank you to:

My undergraduate mentors Drs. Bart and Beth DeStasio. You encouraged me to follow my own path and provided the support when I needed it. Your belief in me meant the world.

My undergraduate research advisor Dr. Kimberly Dickson. You gave me my start as a researcher and started my career path. I can never thank you enough for giving me the opportunity to work with you and become more of a colleague than a student.

My technician advisor Dr. Robert Cramer. You took a chance on me when I joined your lab. I learned so many valuable skills and provided me with an invaluable scientific foundation. Thank you for treating me like everyone else and pushing me to be the best I could be.

My advisor Dr. Mark Mandel. I could not have asked for a better mentor throughout this process. You always believed in me and pushed me when I needed it the most. Thank you for making me the scientist I am today and for allowing me to explore career paths that have led me to where I'm going next.

Past and present members of the Mandel lab especially Denise, Ruth, Natalia, and Avery. Thank you for creating a supportive environment that not only made it fun to be in the lab every day but also a place for scientific brainstorming. Denise, I have looked up to you from the start. Thank you for always bouncing ideas with me and your insightful comments. Ruth, my lab twin, I am so thankful to have someone to go on this journey with. Your support throughout the years has been invaluable.

My undergraduate mentee Andrew. It has been a pleasure watching you grow in confidence as you become an independent researcher. I could not have done this without your help especially in the last few months.

My parents, Mike and Karen. Thank you for your undying support throughout the years. I love you both.

My loving husband Josh. I literally could not have done this without you. Thank you for always believing in me and reminding me when I needed it the most that I am capable and could do this. Your support means the world to me, more than you could ever know. I am so thankful for your dedication to watching our son so I could accomplish this monumental task. I love you from the bottom of my heart.

To my son Landon. Don't let anything stop you from chasing your dreams. I love you to the moon and back.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: The *Vibrio*-squid symbiosis as a model to study mechanisms of host colonization specificity.

Katherine M. Bultman

ABSTRACT

Plant and animal hosts must acquire horizontally-transmitted beneficial symbionts from their surrounding environment. To find their partners, both the host and bacteria employ specific mechanisms that lead to the establishment of correct associations. Host colonization for many horizontally-acquired symbionts relies on two-component regulation of biofilm formation for the lifestyle switch needed to associate with a host. In *Vibrio fischeri*, a multikinase network regulates the production of a host specific biofilm as a necessary step in the winnowing process leading to the colonization of the Hawaiian bobtail squid light organ. Here I summarize current literature of biofilm formation by two-component regulation and the use of the *Vibrio*-squid system as a model for studying the specific molecular mechanisms involved. Further, I discuss the diversity of colonization behaviors that lead to host specificity.

INTRODUCTION

Bacteria live in complex communities, either free living in the environment or associated with a plant or animal host. This constant interaction between two or more organisms has been termed ‘symbiosis’ (1, 2). These interactions are generally stable and last the lifetime of the organisms involved (2). When the term symbiosis was initially introduced, its definition encompassed a range of interactions with varying degrees of cost and benefit for those involved. Across the spectrum there are three main categories; parasitism or pathogenicity, commensalism, and mutualism (2). In parasitism, one partner benefits while the other is harmed, while in commensal relationships the second partner is not harmed but also doesn’t benefit (2). Mutualism results in the best of both worlds as both partners benefit from the interaction (2). Much of the current research on plant and animal host-microbe interactions is focused on the parasitic or pathogenic relationships. However, the study of mutualistic relationships between bacteria and their hosts can reveal important aspects of host biology as these organisms evolved in a microbe-rich

environment (3–5). For the host, these beneficial interactions are linked to host development, nutrient acquisition, pathogen defense, and other necessary roles (3, 6–9). Disruption of this vital partnership often results in disease for the host (7, 10, 11). Therefore, it is necessary to understand how beneficial relationships are established and maintained throughout the life of the host.

The establishment of beneficial relationships between bacteria and their hosts occurs through one of two modes of transmission. For obligate relationships where the partners cannot survive without the other, symbionts are passed directly from host parent to offspring through reproduction or embryogenesis, and this is termed vertical transmission (12, 13). The second mode of transmission, termed horizontal transmission, results from the acquisition of symbionts from the surrounding environment (12). Host juveniles emerge without their symbionts, or aposymbiotically, and must recruit their specific partners from the millions of environmental free-living bacteria. Therefore, the colonization process involves tactics by both host and bacterial partners for the establishment of this very specific relationship. For example, hosts secrete mucus and produce cilia to increase the surface area for bacterial attachment while bacteria often rely on biofilm-mediated aggregation during the colonization process (12, 14–19). Furthermore, bacteria have to survive a selective winnowing process as they move to where they will be housed long term (12, 20, 21). My work is focused on how diversity within the colonization mechanisms utilized by a horizontally acquired beneficial symbiont leads to host specificity.

Host specificity

The acquisition and establishment of specific beneficial relationships requires specific strategies employed by both the host and bacteria. Hosts utilize various tactics such as secreting

chemoattractants or antimicrobials, creating a bottleneck using a physical barrier, or the physical transfer from one host to another (22–26). On the other hand, bacteria employ multiple behaviors such as motility, surface attachment, and evasion of host defenses (27). Across both plants and animals, hosts have evolved specialized structures Fronk and Sachs have termed 'symbiotic organs' to facilitate the maintenance and in some cases transmission of beneficial microbes (26). These organs provide a space for microbes to interact with host cells and aid in the selection process of specific symbionts. This filtering process employed by the host plays a key role in driving the specific genotypic and phenotypic diversity of their symbionts which is critical for maintaining host specificity (26).

Two-component signaling in response to environmental signals

Bacteria must respond to environmental signals, particularly during host colonization, and formulate the appropriate behavioral response. One way in which bacteria control these behaviors is through two-component signaling (TCS). These systems are comprised of a pair of proteins, typically a histidine kinase (HK) and response regulator (RR), and found across all domains of life - archaea, bacteria, and eukarya (28–34). With the large diversity of signals that bacteria must respond to, individual species typically contain dozens to hundreds of unique two-component systems each regulating their own unique response (35, 36). The basic mechanism of these systems is to receive a signal and then mediate a response within the cell, typically through a change in gene expression (28–33). TCS has been found to mediate a variety of behavioral responses including physical and chemical stress, quorum sensing, sporulation, and expression of virulence factors such as colonization, toxin production, and antimicrobial resistance (28). Studying the mechanisms and complexities of TCS is necessary for

understanding how bacteria respond to complex environmental signals and coordinate the appropriate response.

Canonical TCS includes a membrane bound histidine kinase (HK) and a cytoplasmic response regulator (RR) (28, 31, 33, 37–39). The HK senses an environmental signal and translates that input into the desired output through the phosphorylation of its cognate RR leading to a cellular response (30, 31, 36). The most basic HKs contain a sensory domain, catalytic domain (CA), dimerization and histidine phosphotransferase domain (DHp) (29, 31, 39). To initiate a response, HKs utilize a variety of stimulus specific sensory domains that are either periplasmic, cytoplasmic, or transmembrane domains (28, 29, 31). The signal from the sensory domains for many HKs is transmitted to the catalytic domains through additional extensions of the transmembrane helices (31). In roughly 30% of all HKs, this is a HAMP domain (alpha-helix domain present in histidine kinases, adenylyl cyclases, methyl-accepting proteins and phosphatases) that assists in a conformational change to transduce the signal from the sensory domains to the catalytic core (31, 36). Upon receiving and transmitting a signal through intracellular signal domains, ATP binds to the CA domain of HK homodimers and autophosphorylates the conserved His residue in the DHp domain (30, 31). Autophosphorylation typically occurs *in trans* within the homodimer, but at least some HKs utilize a *cis* mechanism (29). Once phosphorylated, the DHp domain transfers the phosphoryl group to the conserved Asp residue in the receiver (REC) domain of the cognate RR (28–30). HKs can also act as a phosphatase to remove phosphoryl groups from their RR to control the phosphorylation state and downstream cellular response (29). In response to sensory input, HKs modulate either autokinase activity, phosphatase activity, or both to control the phosphorylation level of their RR (29).

The RR protein is the key element to turn the sensory input into a specific cellular response (31, 37). These proteins typically contain a REC domain and an effector domain (30). Once phosphorylated, RRs form homodimer or homomultimers that are required for output generation (31, 36). RRs contain a diverse range of effector domains that allow for the execution of many different cellular responses (31). Approximately 70% of all RRs are thought to be transcriptional regulators as they contain a DNA-binding domain. Others contain enzymatic domains, such as adenylate cyclases, that are involved in further signal transduction, or in rare cases RNA-binding domains (31, 36). The activation of these RR results in a cellular response to the input signal.

HKs must be able to signal through their cognate RR in order to establish a cellular response upon receiving a signal. Within cells, there is minimal cross-talk between different TCS pathways (40). The specificity in HK-RR interactions avoids crosstalk with other systems. For the necessary signaling to occur, HKs must be able to distinguish between cognate and non-cognate partners (41). A study by Skerker *et al.* identified patterns of amino acid coevolution and specificity residues in HKs and RRs within cognate pairs (42). They demonstrated that mutation of these subsets of coevolving residues was sufficient to switch the specificity of the kinase EnvZ (42). In addition, RRs can be nonspecifically phosphorylated by other HKs or cellular pools of acetyl-phosphate, therefore phosphatase activity by the cognate HK can reduce this cross-talk (35, 41). Furthermore, temporal or spatial restriction of pathways aids in maintaining specificity in two-component pathways (35, 41). All of these different mechanisms to reduce cross-talk suggest strong evolutionary pressure to maintain specificity between HKs and RRs.

Bacteria must adapt as they transition into new niches. To respond to new niche-specific signals, bacteria can acquire TCS pathways through horizontal gene transfer (HGT) or gene

duplication events (43). As the majority of new TCS emerge through gene duplication events (35, 41), a recent study by Nocedal and Laub determined specificity residues that quickly mutated after gene duplication to maintain the protein interaction specificity (44). Some mutations strengthened the interactions between the HK and RR while others weakened the interactions between the paralogous systems (44). This is the result of strong selection against cross-talk to establish relative stasis of these signaling pathways (44). Conversely, mutation of TCSs that are acquired through HGT generally isn't necessary as they typically preserve their function and maintain their specificity between the HK and RR (43). The acquisition of new TCS pathways and partners allows bacteria to respond to new signals and expand their niche specificity.

One main modification to canonical TCSs is the formation and utilization of hybrid histidine kinases (HHK). These HHKs are classified through a tethered REC domain and in some cases either a tethered or separate histidine phosphotransfer (HPt) domain (28, 33, 39). It has been found that nearly 25% of all HK are hybrids and are more dominant in eukaryotic species although they are still found in prokaryotes (30, 36, 39). Within HHKs, the flow of phosphoryl groups follows the canonical His-Asp phosphotransfer between the DHp and REC domains, but there is additional relay from the tethered REC domain to a conserved His in the HPt domain and then onto the Asp in the REC domain of the RR (28, 39). The formation of HHK employs a more sophisticated phosphorelay and tighter regulation of bacterial behaviors (39). Compared to canonical TCS pairs, HHKs demonstrate reduced co-evolution and specificity between the kinase and REC domains (45, 46). The spatial tethering and thus increased concentrations of the REC domain results in almost exclusive phosphotransfer between the DHp and REC domain reducing signaling cross-talk (45). Therefore there is less selective pressure for these proteins to diversify and maintain specificity residues (45).

Further complexities within TCS include the coordination of HKs in multikinase networks (47). Within these systems, there is coordination of multiple environmental stimuli for more precise control of bacterial behaviors including the regulation of major lifestyle changes (47). There are multiple types of multikinase networks that regulate bacterial behaviors. The first type is phosphorylation of the same downstream regulator by multiple HKs (47). Examples of this include quorum sensing in *Vibrio* species, sporulation in *Bacillus subtilis*, and control of virulence in Pseudomonads (47–50). The second type involves the interaction of HKs with each other to regulate a downstream RR (47). Commonly, this interaction involves phosphotransfer from one HK to another to either positively or negatively regulate behaviors (47). The best characterized example of this is the GacS/RetS/LadS coordination in *Pseudomonas aeruginosa* for the lifestyle switch between a planktonic and biofilm state (47). Finally, there are networks where the RR can affect HK signaling (47). In *Caulobacter crescentus*, the phosphorylated RR DivK is the ligand for the HK DivL. The phosphorylation state of DivK determines if DivL can interact with another HK CckA and switches the function between kinase and phosphatase activity to regulate the cell cycle (47). Overall, the diversity and complexity of these signaling pathways allows bacteria to fine-tune responses to a vast range of environmental signals leading to the coordination of behaviors necessary for niche adaptation and specificity. Their study provides insights into the mechanisms that bacteria use to sense and respond to their ever changing environment.

Biofilm production for successful colonization of the host and the use of TCS

As part of the colonization process, many symbiotic bacteria must transition to or from a free-living planktonic state to living within a biofilm. Generally, biofilms are characterized as aggregates of bacterial cells living in close association encased in a exopolysaccharide matrix

(51, 52). This matrix can be composed of multiple different polysaccharides such as cellulose, proteins, lipids, and eDNA (51–53). Formation of biofilms allows bacteria to adhere to a multitude of surfaces including particles within the water column and living tissue (51, 52). In some cases, formation of the exopolysaccharide matrix promotes cohesion of bacterial cells together compared to adhesion to a surface (54). The transition to a biofilm state is coordinated in multiple ways. For some bacteria, secondary messengers such as c-di-GMP initiate the lifestyle switch (55–57). Others rely on cell density and quorum sensing or exogenous environmental signals like calcium to initiate a change in gene expression for biofilm production (52, 58–62). Here, I focus on how biofilm is regulated through TCS in both pathogenic and beneficial bacteria.

Quorum sensing (QS) is utilized by bacteria for the coordination of bacterial behaviors including the transition to and from a biofilm state (58). This cell-density dependent switch has largely been characterized in *Vibrio* species and was originally found to control bioluminescence (63). Commonly, extracellular homoserine lactone signaling molecules are sensed by histidine kinases which coordinate a multikinase network for phosphotransfer and downstream gene expression changes (49, 63). In the pathogenic species *Vibrio salmonicida*, QS negatively regulates biofilm formation and attenuates virulence suggesting the coordination of biofilm formation under low cell density conditions (64, 65). A similar pattern is seen in *Vibrio cholerae*, where a main QS regulator negatively regulates production of the *Vibrio* polysaccharide (VPS) (57). *Xanthomonas* species utilize QS to coordinate microcolony formation through biofilm production during the infection of their plant host xylem (58). The formation of a biofilm is a necessary part of the root colonization process for beneficial *Rhizobia* species. In *Azorhizobium caulinodans*, QS and high cell density was found to increase biofilm formation with multiple LuxR-type receptors negatively impacting biofilm formation when mutated (59). These examples

demonstrate the impact of cell density on biofilm formation regulation through two-component QS pathways across bacterial colonizers.

Biofilm formation by bacteria is also coordinated through canonical TCS outside of QS and multi-component networks during host colonization (66). In *Vibrio cholerae*, the RvvAB system negatively regulates the transcription of VPS leading to altered biofilm formation (67). Disruption of this pathway led to attenuation of intestinal colonization and infection (67). In addition, the response regulator VpsR promotes biofilm formation through the transcription of the *vps* locus (57). The FixJL system in *Burkholderia* negatively impacts biofilm formation and various mutations within the FixJ HK determines better survival within macrophages or the clinical decline of patients with cystic fibrosis (CF) (68). Regulation of biofilm production in bacteria may also utilize the coordination of multiple histidine kinases to form multi-component networks (66). This coordination can facilitate cross regulation of systems for biofilm control (66). Multiple histidine kinases were found to impact exopolysaccharide production, including Pel, in *Pseudomonas aeruginosa* (69). Chronic host colonization, particularly in the lungs of CF patients, relies on the switch from planktonic acute infection to a chronic biofilm state (70, 71). Tight regulation of this switch in multiple *Pseudomonas* species is controlled through a tripartite GacS/RetS/LadS multikinase network (70, 71). Biofilm production during host colonization for the beneficial symbiont *Vibrio fischeri* is also controlled through a multikinase network (62, 72–74). Much of what we know about TCS regulation of biofilm formation comes from pathogenic bacteria. Therefore, the continued study of these pathways in beneficial partners will provide a greater understanding of biofilm regulation and its impact on host colonization more broadly.

The *Vibrio*-Squid System

V. fischeri is a horizontally acquired symbiont of the Hawaiian bobtail squid (*Euprymna scolopes*) light organ. The acquisition from the environment is a multi-step process. Long-term study of this system has led to a better understanding of the specific molecular mechanisms for the establishment and maintenance of symbiotic relationships (75). Coordination of precise regulatory mechanisms are utilized at each stage and are discussed below.

Establishing the Symbiosis: Recruitment from the seawater

Within the Hawaiian coastal water column, there are approximately 1 million bacterioplankton per mL of seawater. *V. fischeri* makes up less than 0.1% of this total except in regions where there is an abundant squid population (20, 76–78). Here, *V. fischeri* is 24-30 times more abundant in both the water column and sediment (78). Once hatched, aposymbiotic hatchlings must recruit bacteria from the seawater to establish its symbiosis (20, 79). The squid vent water through its mantle cavity and over the gills and light organ as it swims. Each hatchling is able to vent approximately 2 µL of water through its body per second (20, 80). Given the low concentration of *V. fischeri*, the squid requires specific mechanisms to actively recruit the bacteria from the water.

The juvenile light organ contains 2 ciliated appendages on each side of the internal light organ (81). These cilia beat and promote flow of bacteria-sized particles towards the pores of the light organ (15, 20, 80, 82). In addition, these cilia secrete mucus upon exposure to bacterial peptidoglycan and generate a mucus field outside the light organ to further trap the bacteria (15, 82). Within the mucus field, *V. fischeri* make up a large portion of the bacteria that are adhered

(83). It is here that *V. fischeri* first encounters the host immune system and begins to evade hemocyte phagocytosis (84). These initial selective processes begin the establishment of the symbiosis between *V. fischeri* and the squid host.

Initiation: Aggregation

A necessary step of the colonization process is the aggregation of *V. fischeri* cells within the host mucus. *V. fischeri* spend approximately 4-6 hrs in the mucus field before beginning to migrate into the light organ (15). It is during this time that the bacteria aggregate through the production of an host-specific exopolysaccharide termed the symbiosis polysaccharide (Syp) (15, 19). Production of the Syp biofilm is required for colonization to occur, as strains that are unable to form this polysaccharide cannot colonize the squid host (15, 19). The production of Syp is tightly controlled by a TCS that activates σ^{54}-dependent transcription of an 18-gene locus that is responsible for the regulation, generation, and export of the polysaccharide (19), (54, 85, 86). Additional pressures are placed on *V. fischeri* during the aggregation stage. Nitric oxide (NO) is produced by the cilia and ducts of the light organ and is present within the mucus field (87). In response to NO, *V. fischeri* has been shown to up or down-regulate genes to reduce oxidative stress (88). Furthermore, *V. fischeri* encodes two proteins, NsrR and HnoX, that sense NO and mediate transcriptional responses (88, 89). It has been suggested that this mechanism may be the reason why *V. fischeri* dominates within the aggregates (20, 87). A study by the Visick lab demonstrated that HnoX affects biofilm formation through inhibition of the positive regulator HahK (90). It is thought that this negative regulation by NO aids in the dispersal of *V. fischeri* from the aggregates to continue the colonization process (90). Discussion of the regulation of Syp production is discussed in detail below.

Migration: Motility and chemotaxis

Upon dispersal from the aggregates, the bacteria must migrate through the pores of the light organ and through the ducts to the deep crypt space. Migration through these spaces has been found to be dependent on both motility and chemotaxis (23, 91–94). *V. fischeri* use flagellar motility to swim towards a host-derived chitin oligosaccharide that leads the bacteria from the cilia to the ducts (23, 92–96). Once inside the crypts however, *V. fischeri* do not maintain their flagella (97). However, they are rebuilt shortly after expulsion into the surrounding seawater at dawn (97).

Persistence: Crypt colonization and luminescence

Entry of *V. fischeri* into the deep crypt spaces causes a variety of host responses to occur. Upon reaching the crypt space, the structural bottleneck between the antechamber and crypt constricts thereby restricting the bacteria to the crypt space and limiting the population diversity to those that have already colonized (25). Apoptosis and regression of the ciliated appendages upon crypt colonization further restricts recruitment of bacteria from the environment (81, 84, 98–100). Colonization of the crypts initiates a 24 hr diel cycle that *V. fischeri* and the squid host will undergo for the remainder of the host life cycle. Every day at dawn, 95% of the bacteria are expelled from the light organ (78, 101, 102). Throughout the day, the bacterial population regrows to reach optimal cell density for luminescence (103, 104). During early evening, the bacteria undergo a metabolic switch initiated by the squid host to preserve oxygen for luciferase activity (105–107). The squid utilize the light produced by the bacteria to match the moonlight in a form of camouflage known as counterillumination (101, 108–110). Throughout the night, the epithelial barrier of the crypt breaks down and at dawn both *V. fischeri* and host cells are expelled (102).

This multi-step colonization process demonstrates multiple relevant phenotypes that are necessary for the establishment and maintenance of symbiotic relationships. The use of the *Vibrio*-squid model provides a unique opportunity to identify the molecular mechanisms at each stage with the one-symbiont, one-host relationship.

Host-specific biofilm regulation in *V. fischeri:* The *syp* regulon

Of particular interest to this work is the production of the host-specific exopolysaccharide Syp and its regulatory pathway. This biofilm and subsequent bacterial cell aggregation is a required step within the initiation stage of the colonization process as mutants that are unable to produce this biofilm are unable to colonize the squid host (19, 85). The *syp* locus was first identified through a transposon screen looking for strains that did not colonize the squid host (19). Across the 18-gene locus, four mutants were found in genes predicted to be involved in exopolysaccharide biosynthesis (19, 54). The molecular mechanisms for Syp biofilm regulation has been well studied in the Hawaiian isolate ES114. My work focuses on how this regulation differs across isolates leading to reproducible squid colonization. What is currently known about the key regulators involved in Syp production is discussed below (**Fig. 1.1**).

Two-component regulation for biofilm formation

A predicted orphan hybrid histidine kinase was identified through a transposon screen as a mutant that failed to colonize squid but did not affect other processes involved in colonization such as motility and luminescence (111). Therefore, this kinase was termed the regulator of symbiotic colonization sensor, or *rscS* (111). A mutated allele of *rscS* that had increased

activity, termed *rscS1,* expressed from a multicopy plasmid was found to increase *syp* expression and also produce robust wrinkled colony formation which was not previously seen by *V. fischeri* in culture (85). Through the use of this allele, it was demonstrated that this increase in biofilm formation by RscS1 was dependent on the downstream response regulator SypG leading to faster colonization, larger aggregates, and an increase in competitive fitness (85, 112). In contrast, deletion of this regulator resulted in a lack of aggregate formation (85, 112).

As a hybrid sensor kinase, RscS contains a DHp, REC, and HPt domain with conserved phosphorelay residues at H412, D709, and H867 (111, 113). Both H412 and D709 residues are necessary for the promotion of *syp* transcription, wrinkled colony formation and pellicle formation (113). However, the H867 residue within the HPt domain does not completely abolish these phenotypes suggesting that the DHp and REC domains of RscS are vital for its function within the *syp* regulatory pathway.

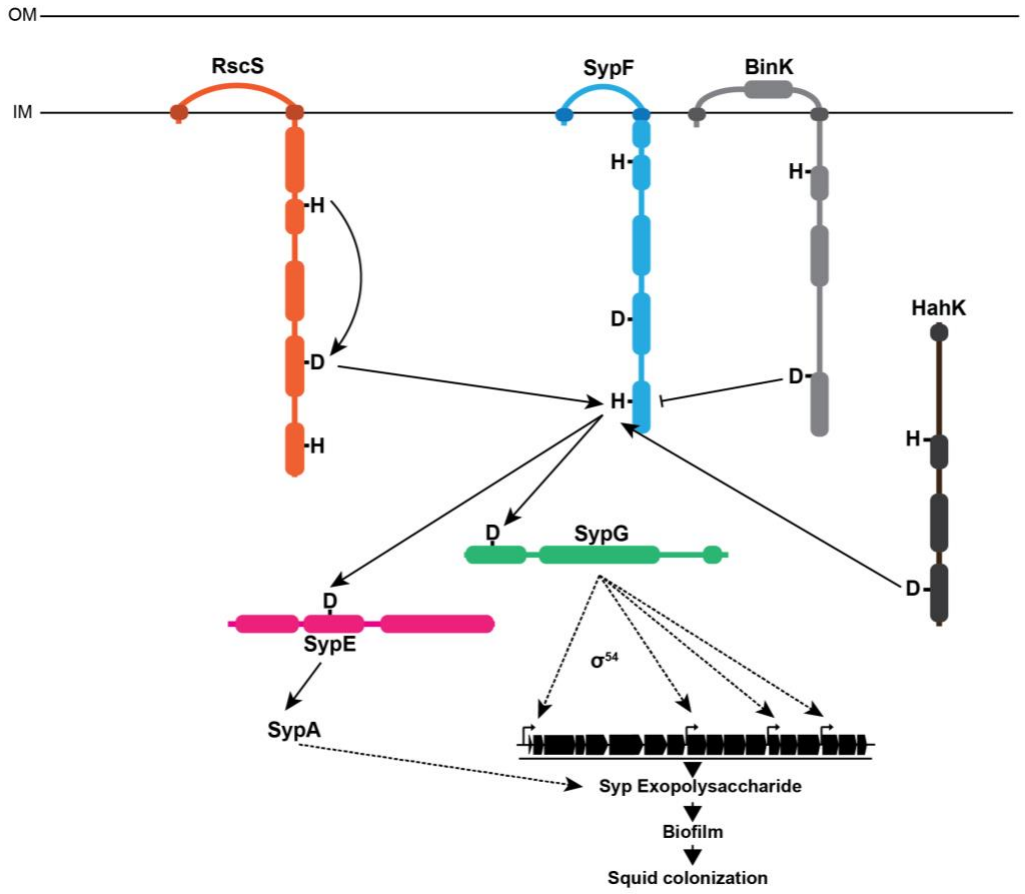


Figure 1.1. The Syp regulation pathway in *V. fischeri* ES114.
Model of Syp phosphorelay between known syp regulators. Solid arrows indicate phosphoryl group transfer and dashed arrows indicate activation through σ^{54} for SypG and unknown mechanisms for SypA. Phosphorylation between SypE and SypA is a Serine/Threonine phosphorylation compared to Histidine/Aspartic acid phosphorylation of the other regulators.

The hybrid sensor kinase SypF is encoded within the first operon of the *syp* locus and has two response regulators, SypE and SypG encoded on either side (19). Similar to RscS, SypF contains DHp, REC, and HPt domains (19). Work investigating the phospho-receiving residues in SypF demonstrated that the H250, D549, and H705 are all necessary for inducing biofilm formation suggesting kinase activity (72). SypF functions upstream of the response regulator SypG to control *syp* transcription and biofilm formation (72, 114). As both RscS and SypF were shown to act upstream of SypG, it was suggested that *syp* may be regulated by a multikinase network. The function of RscS in promoting *syp* transcription and biofilm formation was found to

be dependent on the presence of SypF, suggesting that RscS functions upstream of this regulator (72). Further analysis of the relationship between RscS and SypF showed that only the H705 residue in the HPt domain of SypF was required for biofilm formation (72). A chimera containing RscS except for the HPt domain was constructed with the HPt domain of SypF. The expression of this chimera on a multicopy plasmid in a Δ*sypF* background produced wrinkled colony formation and was able to colonize squid in a Δ*rscS* Δ*sypF* strain (72). These data provided evidence of a phosphorelay between the RscS REC domain and the SypF HPt domain for biofilm production through SypG.

In addition to its role as an important positive regulator, SypF has recently been shown to also negatively regulate biofilm production. A frameshift mutation in *sypF* led to this variant promoting biofilm formation within a strain lacking the negative regulators SypE and BinK and was termed *sypF2* (74). Further investigation found that the H250Q and D549A variants in this background also permitted biofilm formation while the H705Q HPt domain variant did not (74). It was therefore concluded that SypF also functions as a biofilm negative regulator. The current model suggests that as a negative regulator, SypF is acting as a phosphatase with the DHp and REC domains pulling phosphates from the HPt domain. When these domains are mutated, the HPt domain is able to keep its phosphates and pass them onto SypG leading to *syp* transcription and biofilm formation (74). It is currently unknown how SypF switches functions from a kinase to a phosphatase and further work is necessary to elucidate how this switch may impact biofilm formation during the colonization process.

As part of the two-component phosphorelay with SypF, phosphorylation of the conserved D53 residue within the N-terminal REC domain of SypG leads to the increased transcription of the *syp* locus in an *rpoN*-dependent manner and robust biofilm formation (19, 115). This established the role of SypG as a σ^{54}-dependent regulator of *syp* transcription. SypG was found to bind to

four distinct promoter regions before *sypA, sypI, sypM* and *sypP* that contain a conserved 22 bp enhancer sequence (19). Binding to this enhancer sequence at all four promoters is necessary for *syp* transcription and biofilm formation. Overall, SypG has been characterized as a response regulator with a REC domain, HTH DNA-binding domain, and nucleotide binding domain that interacts with σ^{54} (115). Binding to its enhancer sequence leads to the activation of σ^{54} open complex formation and subsequent *syp* transcription (116).

Located directly upstream of *sypF*, a second putative RR *sypE* is encoded within the *syp* locus (19). SypE was found to act antagonistically in SypG-induced biofilms as overexpression or phosphomimetic *sypG* is unable to produce robust wrinkled colony formation unless *sypE* was deleted (117, 118). However, the presence of *sypE* is necessary for RscS-induced biofilms (117, 118). Further analysis of the *syp* regulon found that SypF phosphorylates both SypG and SypE *in vitro* through the HPt domain (72). SypE contains a central REC domain located between an N-terminal serine kinase and a C-terminal serine/threonine phosphatase (118, 119). Analysis of the domain function found that the phosphatase activity from the C-terminal domain positively affects biofilm while the N-terminal kinase domain inhibits it (119). As the REC domain is situated between these domains of opposing functions, it was hypothesized that the phosphorylation state modulates the activity of SypE (119). A D192A REC domain variant is unable to be phosphorylated and led to the inhibition of biofilm formation by SypG (120). This result led to the conclusion that phosphorylation of SypE by SypF leads to the switch from inhibiting to promoting biofilm production (119).

The opposing functions of the kinase and phosphatase domains in SypE suggested a “partner switching” model where an anti-sigma factor with a phosphorylatable serine in a STAS (sulphate transporter and anti-sigma antagonist) domain is controlled by an upstream serine/threonine kinase or phosphatase (118, 121). The first protein encoded in the *syp* locus, SypA contains a

conserved serine in a STAS domain and is predicted to be an anti-sigma factor suggesting that this protein partners with SypE (118–120). *In vivo* co-immunoprecipitation assays in combination with *in vitro* phosphorylation assays demonstrated that SypE and SypA interact with the N-terminal domain of SypE acting as a kinase to phosphorylate SypA while the C-terminal domain acts as a phosphatase to dephosphorylate SypA (120). SypE controls biofilm formation through the phosphorylation state of the conserved S56 in the SypA STAS domain (120). A phospho-blind S56A SypA variant promoted biofilm formation while a phosphomimetic S56D variant inhibited biofilm formation (120). This led to the conclusion that phosphorylation of SypA by SypE rendered it inactive resulting in the inability to promote biofilm and squid colonization (120). While SypA has been suggested to operate downstream of *syp* transcription, the specific mechanism remains unknown (118, 120).

Other two-component regulators involved in *syp* regulation

The negative regulator BinK was identified through a transposon insertion sequencing (INSeq) experiment identifying genes influencing squid colonization (122). Mutants of *binK* were found to be overrepresented in the output compared to the input suggesting that disruption of this gene confers an advantage during the colonization process (73, 122). Further analysis demonstrated that a Δ*binK* strain not only had a competitive advantage over wild-type, but also produced larger aggregates outside the pores of the light organ (73). Deletion of *binK* also results in increased *syp* transcription and faster wrinkled colony formation. In contrast, overexpression of *binK* leads to a loss of biofilm formation which can be overcome with the overexpression of *sypG (73)*. Furthermore, the deletion of *binK* rescues both biofilm formation and squid colonization in a strain lacking the positive regulator RscS (123). Double deletion of *rscS* and *binK* results in the same increase in *syp* transcription and aggregate formation as the *binK* deletion in addition to restoring colonization to wild-type levels compared to an *rscS* deletion

alone (123). This result argued for a paradigm shift where the RscS pathway is necessary but constitutive, and that derepression of BinK upon receiving a host signal leads to aggregate formation.

From a pool of HKs that were labeled as colonization factors in the INSeq screen, HahK was identified as a biofilm regulator in the presence of calcium (62, 73, 122). HahK is annotated as a sensor kinase, and was found to have a significant decrease in colonization when disrupted (73, 122). This gene is found in an operon with the nitric oxide (NO) sensor HnoX and was named HnoX-associated histidine kinase (62). Predicted to be cytoplasmic due to a lack of transmembrane domains, HahK contains a DHp, CA, and REC domain (62). HahK has been shown to be a positive regulator of biofilm in the absence of *binK* and under calcium inducing conditions acting through the HPt domain of SypF (62). Induction of biofilm formation for a Δ*binK* strain on calcium is dependent on *hahK*, however the exact mechanism remains unknown. As the sensor kinase encoded downstream of HnoX in other bacteria is inhibited by HnoX (124), it was asked if these two proteins interact in *V. fischeri* (90). A previous study demonstrated that HnoX mutants had a competitive advantage during squid colonization (125). Wrinkled colony and *syp* transcription assays showed that the HnoX inhibition of biofilm formation was dependent on HahK (90). As HnoX senses NO, it was determined that the presence of NO inhibited biofilm formation in a HnoX and HahK-dependent manner with the deletion of *hnoX* resulting in larger aggregates during colonization (90). Further analysis showed that biofilm formation was dependent on the conserved His and Asp residues in HahK (90). These data suggest that during the colonization process, NO activates HnoX to inhibit HahK from phosphorylating the HPt domain of SypF for biofilm formation.

Diversity in the establishment of *V. fischeri* colonization

The apparent simplicity of the vibrio-squid symbiosis hides the diversity present in both the symbionts and the hosts. While initial work in the vibrio-squid system (81) suggested a uniform pattern in which *V. fischeri* strains colonize *E. scolopes* and other species do not, natural sampling began to reveal a more complicated pattern. Studies revealed that the colonization ability of different *V. fischeri* strains varied over at least three orders of magnitude and that different strains exhibited substantial differences in their competitive colonization ability (76, 126–128).

Despite an understanding that there existed diversity in the ability of *V. fischeri* to colonize squid, it was still surprising when it was revealed that natural adult *E. scolopes*, which contain six crypts per light organ, are typically colonized by an average of 6-8 isolates per animal (129, 130). Experimental studies using strains derived from ES114 that had been tagged with different antibiotic resistance cassettes demonstrated that multi-strain colonization of animals in laboratory conditions is readily observed at an inoculum concentration of 3,400 CFU/animal (95). A similar experiment was conducted in which red- and green-fluorescent derivatives of the same strain were competed leading to a direct observation of dual-colonized crypts. Development of transposon sequencing in *V. fischeri* led to the observation that over 80 strains can colonize an individual animal, although the high inoculum concentrations and presence of mutant strains in the pool suggest that this may not be representative of natural colonization dynamics (122).

In some cases, the presence of a single regulatory protein dramatically alters the ability of bacteria to colonize a host. To begin to identify the genetic basis for the differential ability to

colonize *E. scolopes*, a comparative genomic study of North Pacific *V. fischeri* identified the biofilm regulator RscS as present in squid symbionts and divergent or absent in most fish symbionts (131). Sequencing the genome of fish symbiotic strain MJ11 revealed that the biofilm regulator, RscS, which is critical for squid colonization, was not encoded in the MJ11 genome. However, the locus of eighteen *syp* target genes is present and conserved in MJ11 (19, 131). In the canonical squid symbiont, ES114, RscS is required for robust host colonization (72, 85, 112). Introduction of RscS into MJ11 is sufficient to activate biofilm phenotypes *in vitro* and to enable the fish symbiont to colonize squid (131). This pointed to the presence or absence of RscS as a key factor in whether natural *V. fischeri* are capable of forming the symbiotic biofilm required to colonize squid.

Phylogenetic analyses revealed that the ancestral group, including MJ11 and consisting largely of fish symbionts, did not encode the biofilm regulator RscS. In contrast, a derived group that includes ES114 and North Pacific squid symbionts had *rscS* DNA. These results pointed to a single horizontal transfer event in which *rscS* was acquired, which proceeded to hijack regulation of the *syp* biofilm genes that are conserved across the species (131, 132). But there was a twist: some fish symbionts in that derived group had divergent *rscS* genes, and the encoded proteins are not sufficient to enable squid colonization (131). It is intriguing to speculate that this RscS allele responds to distinct environmental signals, activating biofilm formation in non-squid hosts or in the ocean. However, given the prevalence of fish symbionts that do not encode the regulator at all, RscS does not seem to be required for colonization of fish hosts.

Although the ancestral group mainly consists of fish symbionts, there are two isolates within this group, SA1 and SR5, that were isolated from Mediterranean bobtail squid species (128, 131). Like MJ11, both of these isolates do not encode RscS but are able to colonize not only their

native squid host but also *E. scolopes* (128, 133). As it was thought that RscS was the crucial factor for squid colonization, these isolates presented a paradox. How do these isolates colonize squid in the absence of this positive regulator? My work aims to answer this question using SR5 as a representative ancestral isolate.

Analysis of natural squid symbiotic isolates has revealed a diversity of symbiosis-related phenotypes that extend beyond biofilm formation. Hawaiian squid symbionts examined typically fall into one of two phenotypic bins: those that largely exhibited higher motility, lower luminescence, and yellow colony pigmentation; or lower motility, higher luminescence, and whiter pigmentation (129). ES114 was typical of the former, whereas a tight clade of closely-related strains were in the latter group. The authors termed these strains as “Group A” *V. fischeri*. Analysis of the colonization ability of Group A strains revealed that they typically outcompete other *E. scolopes* symbionts, though the basis for this phenotype is not yet understood (132, 134).

Additional studies of natural isolates determined that a subset of Hawaiian strains could establish the symbiosis more efficiently than competing strains. Initial experiments demonstrated that this subset of isolates would dominate during colonization, and were therefore termed “Dominant” (D-type) strains (135). D-type strains identified from the Hawaiian population overlap with strains that are phylogenetically Group A. Consistent with this overlap, the dominant phenotype seen in Group A strain MB11B1 is not dependent on *rscS*, and aggregate size in Group A strain MB13B2 is only modestly reduced in a Δ*rscS* derivative *(132, 136)*. There is evidence that the dominant phenotype results from a kinetic benefit in which D-type strains form larger aggregates and form aggregates more quickly than non-dominant S-type strains. Providing an S-type strain prior to a D-type strain--i.e., giving the S-type strain a “head start”-- mitigates the dominance phenotype (130). Until recently, the dominant phenotype

of *V. fischeri* was observed solely in Hawaiian isolates, but a recent study identifies the phenotype in isolates from Pacific *Euprymna morsei* squid (137). Unlike the Hawaiian D-type strains, those from *E. morsei* are not restricted to Group A, suggesting that the dominant phenotype can be achieved independent from the Group A clade.

Bacterial mechanisms for host specificity

The specificity between hosts and bacteria rely on gene content and/or expression by bacteria (138). This includes specific gene content, such as the Nod factors in *Rhizobia*, allelic variation of genes involved in bacterial behaviors, or variation in gene expression (131, 138–145). Often, host specificity is determined by the presence of specific genomic islands containing genes such as effectors, immunity proteins, and secretion systems (138, 140). Across *Bartonella* species, expression of the T4SS from the rat-specific *Bartonella tribocorum* in cat or human-specific species increased the ability to infect rat erythrocytes (138, 146). In addition, both single genes as well as SNPs have been found to change the host range of both pathogenic and beneficial bacteria (131, 138, 140, 142, 144, 147, 148). For the beneficial symbiont *Xenorhabdus nematophila,* the presence of *nilABC* cassette is both necessary and sufficient for host-specific colonization of *Steinernema carpocapsae* nematodes (141). In *Salmonella enterica subsp. Typhimurium,* the FimH variants present in bovine isolates mediates the best bacterial binding to bovine intestinal epithelial cells, and the same occurs with human isolates (142, 148). Variation of the *nodD* gene across *Rhizobium* species results in host specificity of red and white clover (139). The diversity in host specificity tactics demonstrates the need for not only studying species specific methods, but also intraspecies variation that leads to the colonization of multiple hosts.

In this work, I focus on how bacteria establish beneficial relationships with specific hosts. Biofilm formation is a key determinant of host colonization and specificity for bacterial species across the symbiotic spectrum (54, 70, 136, 149–151). Studies on the regulation of biofilm formation across and within species demonstrates the diversity in this process including the use of secondary messengers, quorum sensing, and two-component signaling (56, 66, 74, 150, 152). In the naturally simplified *Vibrio*-squid model system, biofilm formation is an essential step in establishing colonization of the juvenile light organ. Natural isolates of *V. fischeri* from various hosts have revealed different regulatory strategies for this biofilm formation leading to differences in host specificity (131). Here, I investigate the natural diversity in the molecular mechanisms of squid colonization focusing specifically on regulation of the *syp* biofilm. While key work has been done in ES114 to identify the key regulators within that strain, the overall diversity of *V. fischeri* isolates and the lack of requiring the positive regulator *rscS* suggests variation in how *syp* is regulated across the species. I have found that there are three evolutionary groups within *V. fischeri* that demonstrate different regulatory strategies for *syp* while also maintaining conservation of a functional negative regulator BinK. Within the ancestral-like Group C isolate SR5 that lacks *rscS*, the hybrid sensor kinase SypF promotes biofilm formation and squid colonization in an *rscS*-independent manner. This phenotype is dependent on the REC domain suggesting a divergence in function of that domain compared to ES114. However, the HPt domain of SypF alone is sufficient for both biofilm formation and squid colonization in SR5. Two hybrid sensor kinases, GacS and HahK, were found to promote biofilm formation in this isolate although the specific mechanisms in which this occurs remains unclear. My work further reveals the diversity of biofilm regulation across *V. fischeri* and demonstrates the necessity of investigating intraspecies variation of processes that are vital for host colonization and specificity.

Chapter 2: Natural strain variation reveals diverse biofilm regulation in squid-colonizing *Vibrio fischeri*

Ella R. Rotman†, Katherine M. Bultman†, John F. Brooks II, Mattias C. Gyllborg, Hector L. Burgos, Michael S. Wollenberg, Mark J. Mandel

† Authors contributed equally

A version of this chapter is published as:

Rotman ER, Bultman KM, Brooks JF 2nd, Gyllborg MC, Burgos HL, Wollenberg MS, Mandel MJ. Natural Strain Variation Reveals Diverse Biofilm Regulation in Squid-Colonizing *Vibrio fischeri*. J Bacteriol. 2019 Apr 9;201(9):e00033-19. doi: 10.1128/JB.00033-19.

Author contributions: ERR, KMB, MJM designed experiments, ERR and KMB performed experiments, ERR and HLB constructed strains, KMB and MJM wrote the manuscript, HLB wrote SR5 Δ*binK* strain methods.

ABSTRACT

The mutualistic symbiont *Vibrio fischeri* builds a symbiotic biofilm during colonization of squid hosts. Regulation of the exopolysaccharide component, termed Syp, has been examined in strain ES114, where production is controlled by a phosphorelay that includes the inner membrane hybrid histidine kinase RscS. Most strains that lack RscS or encode divergent RscS proteins cannot colonize a squid host unless RscS from a squid symbiont is heterologously expressed. In this study, we examine *V. fischeri* isolates worldwide to understand the landscape of biofilm regulation during beneficial colonization. We provide a detailed study of three distinct evolutionary groups of *V. fischeri* and find that while the RscS-Syp biofilm pathway is required in one of the groups, two other groups of squid symbionts require Syp independent of RscS. Mediterranean squid symbionts, including *V. fischeri* SR5, colonize without an RscS homolog encoded in their genome. Additionally, Group A *V. fischeri* strains, which form a tightly-related clade of Hawaii isolates, have a frameshift in *rscS* and do not require the gene for squid colonization or competitive fitness. These same strains have a frameshift in *sypE*, and we provide evidence that this Group A *sypE* allele leads to an upregulation in biofilm activity. This work thus describes the central importance of Syp biofilm in colonization of diverse isolates, and demonstrates that significant evolutionary transitions correspond to regulatory changes in the *syp* pathway.

IMPORTANCE

Biofilms are surface-associated, matrix-encased bacterial aggregates that exhibit enhanced protection to antimicrobial agents. Previous work has established the importance of biofilm formation by a strain of luminous *Vibrio fischeri* bacteria as the bacteria colonize their host, the Hawaiian bobtail squid. In this study, expansion of this work to many natural isolates revealed

that biofilm genes are universally required, yet there has been a shuffling of the regulators of those genes. This work provides evidence that even when bacterial behaviors are conserved, dynamic regulation of those behaviors can underlie evolution of the host colonization phenotype. Furthermore, this work emphasizes the importance of investigating natural diversity as we seek to understand molecular mechanisms in bacteria.

INTRODUCTION

A fundamental question in studying host-associated bacterial communities is understanding how specific microbial taxa assemble reproducibly in their host. Key insights into these processes were first obtained by studying plant-associated microbes, and the discovery and characterization of Nod factors in Rhizobia was valuable to understand how partner choice between microbe and host could be mediated at the molecular level (153, 154). There are complex communities in humans and other vertebrate animals, yet metagenomic and imaging analyses of these communities have revealed striking reproducibility in the taxa present and in the spatial arrangement of those taxa (155–157). Invertebrate animal microbiomes provide appealing systems in which to study microbiome assembly in an animal host: the number of taxa are relatively small, and examination and manipulation of these organisms have yielded abundant information about processes underlying host colonization (158). For this work we focused on the binary symbiosis between *Vibrio fischeri* and bobtail squids, including the Hawaiian bobtail squid, *Euprymna scolopes*. Bobtail squid have an organ for the symbiont termed the light organ, and passage of specific molecules between the newly-hatched host and the symbiont leads to light organ colonization specifically by planktonic *V. fischeri* and not by other bacteria (20, 23, 159). The colonization process involves initiation, accommodation, and persistence steps, resulting in light organ crypt colonization by *V. fischeri*. Upon colonization of the squid light organ, bacteria accumulate to high density and produce light. The bacterial light

is modulated by the host to camouflage the moonlight shadow produced by the nighttime foraging squid in a cloaking process termed counter-illumination (110, 160). A diel rhythm leads to a daily clearing of 90-95% of the bacteria from the crypts and regrowth of the remaining cells (103). However, the initial colonization process, including biofilm-based aggregation on the host ciliated appendages, occurs only in newly-hatched squid. This work examines regulation of biofilm formation in diverse squid-colonizing *V. fischeri* strains.

In the well-studied *V. fischeri* strain ES114, biofilm formation is required to gain entry into the squid host. RscS is a hybrid histidine kinase that regulates *V. fischeri* biofilm formation through a phosphorelay involving the hybrid histidine kinase SypF and the response regulator and σ^{54}-dependent activator SypG (19, 85, 112). This pathway regulates transcription of the symbiosis polysaccharide (Syp) locus, which encodes regulatory proteins (SypA, SypE, SypF, and SypG), glycosyltransferases, factors involved in polysaccharide export, and other biofilm-associated factors (19, 54). The products of the ES114 *syp* locus direct synthesis and export of a biofilm exopolysaccharide that is critical for colonization. Additional pathways have been identified to influence biofilm regulation in ES114, including the SypE-SypA pathway and inhibition of biofilm formation by BinK and HahK (62, 73, 90, 161, 162).

V. fischeri biofilm regulation is connected to host colonization specificity. In the Pacific Ocean, the presence of *rscS* DNA is strongly correlated to the ability to colonize squid (131). As one example, while the fish symbiont MJ11 encodes a complete *syp* locus, it lacks RscS and does not robustly colonize squid. Heterologous expression of ES114 RscS in MJ11 activates the biofilm pathway and is sufficient to enable squid colonization (131). Similarly, addition of ES114 RscS to *mjapo*.8.1--a fish symbiont that encodes a divergent RscS that is not functional for squid colonization--allows the strain to colonize squid (131). RscS has also been shown to be necessary for squid colonization in certain strains. In addition to ES114, interruption of *rscS* in

V. fischeri strains KB1A97 and MJ12 renders them unable to colonize squid. Previous phylogenetic analysis revealed that ancestral *V. fischeri* do not encode *rscS*, and that it was acquired once during the organism's evolution, likely allowing for an expansion in host range. From this analysis, it was concluded that strains with *rscS* can colonize squid, with the only exception being the fish symbionts that harbor the divergent RscS, including *mjapo*.8.1 (131).

There are similar *Vibrio*-squid associations worldwide, yet only *V. fischeri* and the closely-related *Vibrio logei* have been isolated from light organs (76, 128, 163, 164). Our 2009 study revealed that although most symbionts have *rscS* DNA, there are Mediterranean *V. fischeri* (e.g., SR5) that do not have *rscS* yet can colonize squid (128, 131, 133). This unexpected finding prompted the current work to examine whether strains such as SR5 colonize with the known biofilm pathway or with a novel pathway. Here, we show that all *V. fischeri* strains tested require the *syp* locus to colonize a squid host, and we identify two groups of isolates that colonize with novel regulation. Given the exquisite specificity by which *V. fischeri* bacteria colonize squid hosts, this work reinforces the importance of biofilm formation and reveals different regulatory modes across the evolutionary tree.

RESULTS

Most *V. fischeri* strains synthesize biofilm in response to RscS overexpression.

Biofilm formation is required for squid colonization, and overexpression of the biofilm regulator RscS in strain ES114 stimulates a colony biofilm on agar plates (85). Our previous work demonstrated that *V. fischeri* strain MJ11 synthesizes a colony biofilm under similar inducing conditions, which is notable because MJ11 does not encode RscS in its chromosome (131).

While the ancestral strain MJ11 did not encode RscS, it had what seemed to be an intact *syp* locus, and overexpression of the heterologous RscS from ES114 was sufficient to enable robust squid colonization (131). We examined a phylogenetic tree of *V. fischeri* isolates (**Fig. 2.1**), and in this study we expand our analysis of RscS-Syp biofilm regulation in a wider group of *V. fischeri* strains.

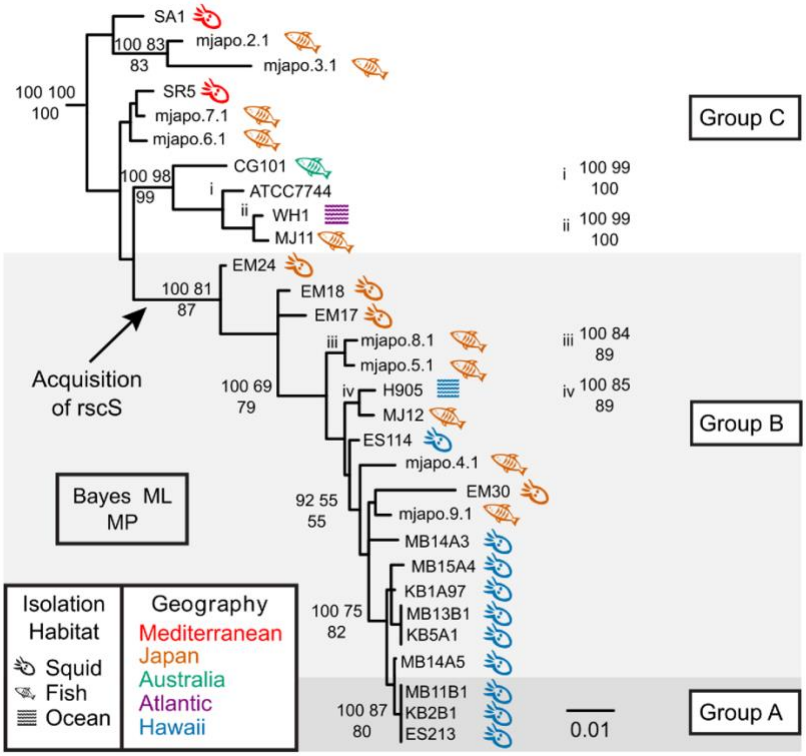


Figure 2.1. *Vibrio fischeri* phylogeny, highlighting the source of each strain. Bayesian phylogram (50% majority-rule consensus) inferred with a SYM+I+Γ model of evolution for the concatenated gene fragments *recA, mdh,* and *katA.* In this reconstruction, the root connected to a clade containing the four non-*V. fischeri* outgroup taxa. Statistical support is represented at nodes by the following three numbers: upper left, Bayesian posterior probability (of approximately 37,500 non-discarded samples) multiplied by 100; upper right, percentage of 1000 bootstrap Maximum Likelihood pseudo-replicates; bottom middle center, percentage of 1000 bootstrap Maximum Parsimony pseudo-replicates. Statistical support values are listed only at nodes where more than 2 methods generated support values ≥ 50%. Strains sharing identical sequences for a given locus fragment are listed next to a vertical bar at a leaf; because of a lack of space, some support values have been listed either immediately to the right of their associated nodes and are marked with italicized lower-case Roman numerals in the phylogram. The isolation habitat and geography of each strain are indicated by symbol and color, respectively. The black bar represents 0.01 substitutions/site.

Initially, we asked whether responsiveness to RscS overexpression would yield a similar colony biofilm in this diverse group of strains. We took the same approach as our previous study and introduced plasmid pKG11, which overexpressed ES114 RscS, into strains across the evolutionary tree (131, 165). We observed that almost all strains tested, including those that lack *rscS*, were responsive to overexpression of ES114 RscS (**Fig. 2.2**). The morphology of the colony biofilms differed across isolates; but in most cases colony biofilm was evident at 24 h and prominent at 48 h. All of the strains exhibited some wrinkled colony morphology at 48 h with the exception of CG101, which was isolated from the pineapplefish *Cleidopus gloriamaris (76)*. These results demonstrated that most *V. fischeri* strains can produce biofilm in response to RscS overexpression, and this includes strains that presumably have not encountered *rscS* in their evolutionary history.

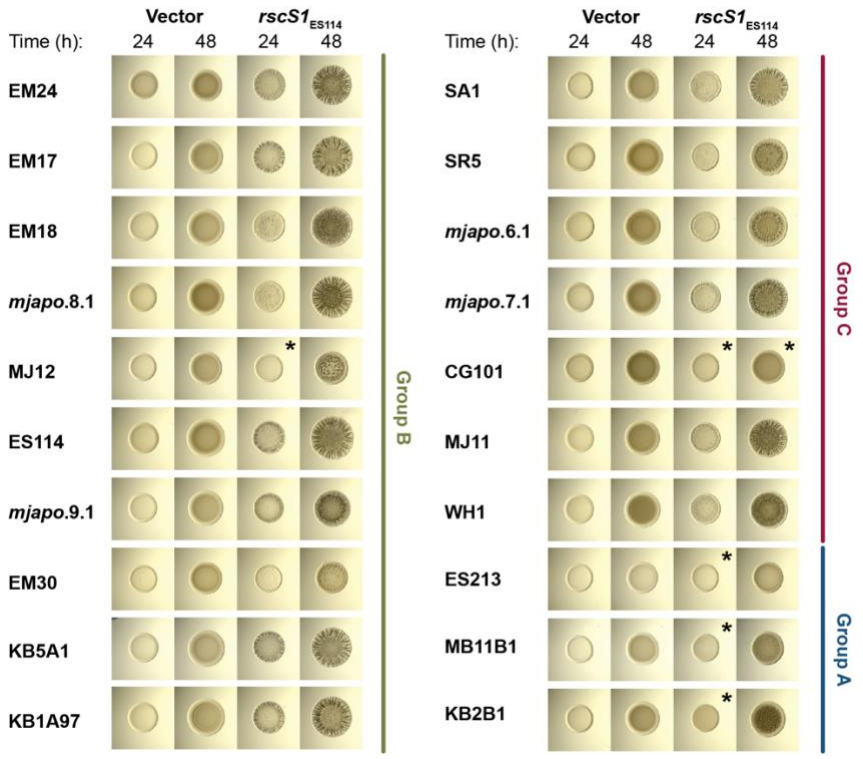


Figure 2.2. Most *V. fischeri* strains tested form colony biofilm in response to RscS overexpression.
Spot assays of the indicated *V. fischeri* strains carrying pKV69 (vector) or pKG11 (*rscS1*; overexpressing ES114 *rscS*) after 24 and 48 h. Strains are MJM1268, MJM1269, MJM1246, MJM1247, MJM1266, MJM1267, MJM1219, MJM1221, MJM1238, MJM1239, MJM1104, MJM1106, MJM1276, MJM1277, MJM1270, MJM1271, MJM1258, MJM1259, MJM1254, MJM1255, MJM1242, MJM1243, MJM1240, MJM1241, MJM1272, MJM1273, MJM1274, MJM1275, MJM1278, MJM1279, MJM1109, MJM1111, MJM1280, MJM1281, MJM1260, MJM1261, MJM1244, MJM1245, MJM1256, and MJM1257. Different phenotypes were observed in the isolates examined; in most cases we observed wrinkled colonies, but in some cases we observed only a subtle pocked pattern (EM30), and in other cases we did not observe any change in colony morphology compared to the vector control (noted by *). The black bar is 5 mm in length.

One unexpected observation was that there was a subset of *rscS*-encoding strains that were reproducibly delayed in their colony biofilm, and had only a mild wrinkled colony phenotype at 48 h (strains MB11B1, ES213, KB2B1; **Fig. 2.2**). We considered whether this was due to differential growth of the strains, but resuspension of spots and dilution plating to determine CFU/spot demonstrated no significant growth difference between these strains and ES114 under these conditions. The strains are closely-related (**Fig. 2.1**) and a previous study had noted that this group shared a number of phenotypic characteristics, e.g. reduced motility in soft

agar (129). Those authors termed this tight clade as “Group A” *V. fischeri* (134). Our results in Figure 2 argue that Group A strains do not respond to RscS in the same manner as other *V. fischeri* strains, which prompted us to investigate the evolution of the RscS-Syp signaling pathway. We have maintained the Group A nomenclature here, and furthermore we introduce the nomenclature of Group B (a paraphyletic group of strains that contain *rscS*; this group includes the common ancestor of all *rscS*-containing strains) and Group C (a paraphyletic group of strains that contains the common ancestor of all *V. fischeri* - these strains do not contain *rscS*), as shown in Figure 1.

Ancestral Group C squid isolates colonize *E. scolopes* independent of RscS and dependent on Syp.

Group C strains generally cannot colonize squid, yet there are Mediterranean squid isolates that appear in this group (**Fig. 2.1**; (131)). The best-studied of these strains, SR5, was isolated from *Sepiola robusta*, is highly luminous, and colonizes the Hawaiian bobtail squid *E. scolopes* (128). Nonetheless, this strain lacks *rscS* (133). We first asked whether the strain can colonize in our laboratory conditions, and we confirmed that it colonizes robustly, consistent with the result result previously published by Fidopiastis et al. (128) (**Fig. 2.3**). Next, we asked whether it uses the Syp biofilm to colonize. To address this question, we deleted the 18 kb *syp* locus (i.e., *sypA* through *sypR*) in strains SR5 and ES114. Deletion of *rscS* or the *syp* locus in ES114 led to a substantial defect in colonization, consistent with a known role for these factors (**Fig. 2.3**). Similarly, deletion of the *syp* locus in SR5, a strain that does not encode *rscS*, led to a dramatic reduction in colonization (**Fig. 2.3**). Therefore, even though strain SR5 does not encode *rscS*, it can colonize squid, and it requires the *syp* locus to colonize normally.

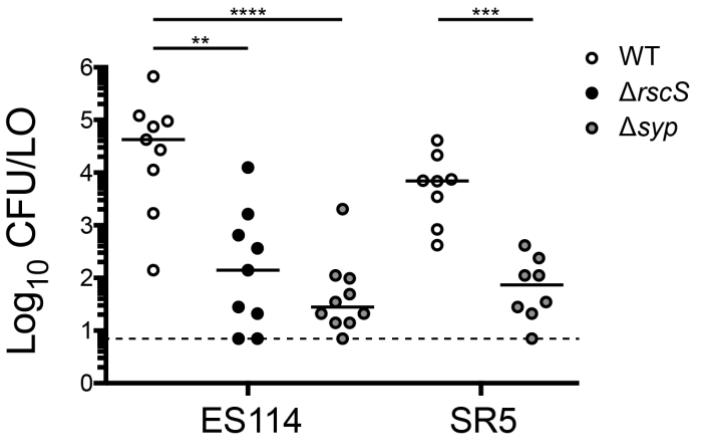


Figure 2.3. Squid colonization in Group C strain SR5, which does not encode RscS, is dependent on the *syp* polysaccharide locus.
Single-strain colonization experiments were conducted and circles represent individual animals. The limit of detection for this assay, represented by the dashed line, is 7 CFU/LO, and the horizontal bars represent the median of each set. Hatchling squid were inoculated with 1.5-3.2 × 10^3 CFU/ml bacteria, washed at 3 h and 24 h, and assayed at 48 h. Each dot represents an individual squid. Strains are: MJM1100, MJM3010, MJM3062, MJM1125, and MJM3501. Statistical comparisons by the Mann-Whitney test, ** p<0.01, *** p<0.001, **** p<0.0001.

RscS is dispensable for colonization in Group A strains.

We noted in the wrinkled colony biofilm assays shown in Figure 2 that Group A strains exhibited a more modest response to overexpression of RscS. Sequencing of the native *rscS* gene in these strains revealed a predicted -1 frameshift (ΔA1141) between the PAS domain and the histidine kinase CA domain. Whereas ES114 and other Group B strains have nine adenines at this position, the Group A strains have eight, leading to a frameshift and then truncation at an amber stop codon, raising the possibility that Group A strains have a divergent biofilm signaling pathway (**Fig. 2.4A**). Given the importance of RscS in the Group B strains including ES114, we considered the possibility that this apparent frameshift encoded a functional protein, either through ribosomal frameshifting or through the production of two polypeptides that together provided RscS function; there is precedent for both of these concepts in the literature (166, 167). We first introduced a comparable frameshift into a plasmid-borne overexpression allele of ES114 *rscS*, and this allele did not function with the deletion of the single adenine (**Fig. 2.4B**).

This result suggested to us that the frameshift in the Group A strains may not be functional. Therefore, we proceeded to delete *rscS* in two Group A strains (MB11B1, ES213) and two Group B strains (ES114, MB15A4). The Group B strains required RscS for squid colonization (**Fig. 2.5A**). However, the Group A strains exhibited no deficit in the absence of *rscS* (**Fig. 2.5A**). We next attempted a more sensitive assay in which a Group A strain was competed against MB15A4. Previous studies have demonstrated that in many cases Group A strains outcompete Group B strains (134, 135). We competed Group A strain MB11B1 against Group B strain MB15A4 and observed a significant competitive advantage for the Group A strain, as was observed previously (134). Deletion of *rscS* in the Group A strain did not affect competitive fitness, demonstrating that MB11B1 can outcompete a Group B strain even if MB11B1 lacks RscS (**Fig. 2.5B**).

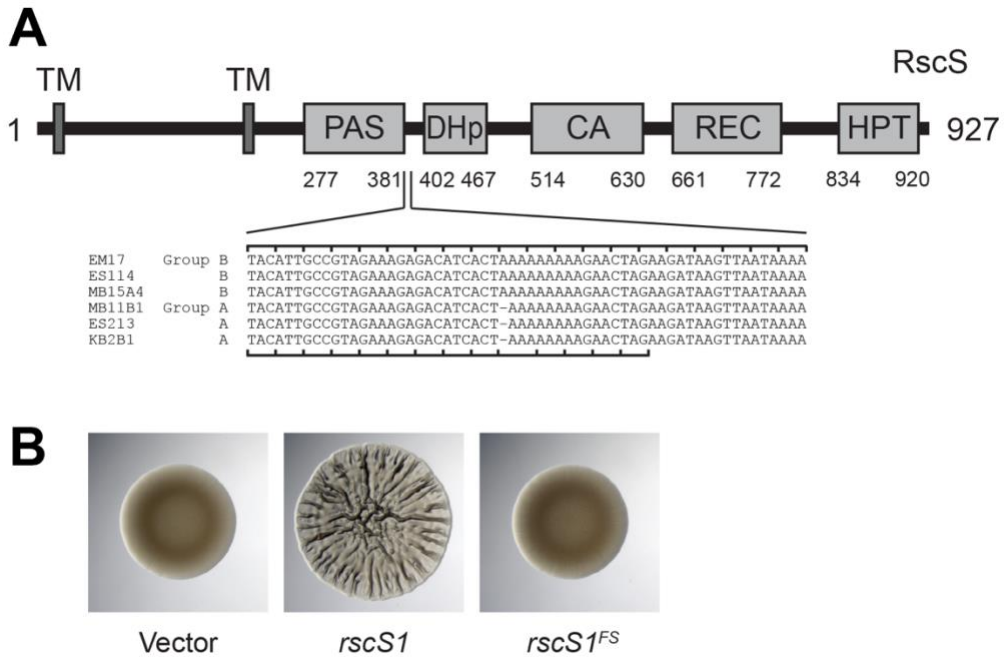


Figure 2.4. Group A strains have a frameshift in *rscS*.
A) ES114 RscS protein domains. Nucleotides 1114-1173 in ES114 RscS (AF319618) and their homologous sequences in the other Group B and Group A strains are listed. The -1 frameshift is present in the Group A *rscS* alleles. The ES114 reading frame is noted on the top of the alignment and the Group A reading frame on the bottom, which is predicted to end at the amber stop codon. **B)** Deletion of nucleotide A1141 in ES114 to mimic this frameshift in pKG11 renders

it unable to induce a colony biofilm in a spot assay at 48 h. Strains are MJM1104, MJM1106, and MJM2226.

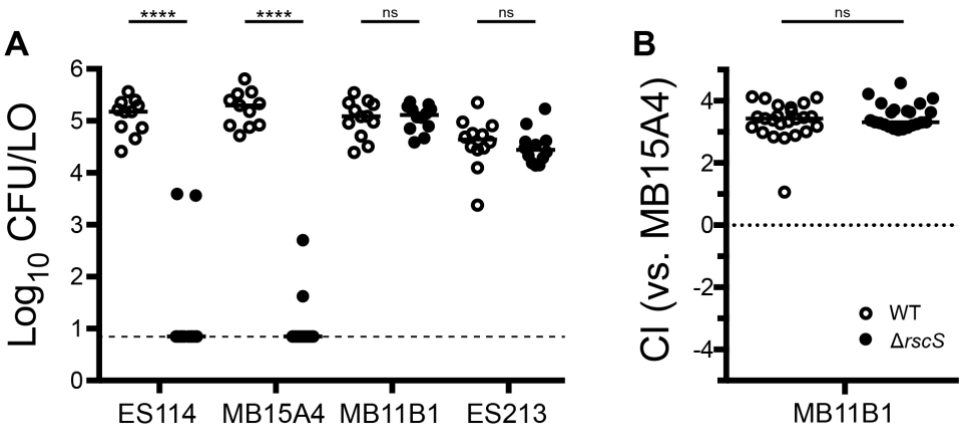


Figure 2.5. Group A strains MB11B1 and ES213 do not require RscS for squid colonization.
Wild-type (WT) and Δ*rscS* derivatives of the indicated strains were assayed in **A)** a single-strain colonization assay and **B)** competitive colonization against Group B strain MB15A4. Hatchling squid were inoculated at 3.5-14 × 10^3 CFU/ml bacteria, washed at 3 h and 24 h, and assayed at 48 h. Each dot represents an individual squid. (A) Strains: MJM1100, MJM3010, MJM2114, MJM3042, MJM1130, MJM3046, MJM1117, and MJM3017. The limit of detection is represented by the dashed line, and the horizontal bars represent the median of each set. In both panels, open dots are wild type and filled dots are Δ*rscS*. (B) The competitive index (CI) is defined in the methods and is shown on a Log_{10} scale. Strains: MJM1130 and MJM3046, each competed against MJM2114. Values greater than 1 indicate more MB11B1. Statistical comparisons by the Mann-Whitney test, ns not significant, **** $p<0.0001$.

The *syp* locus is broadly required for squid colonization.

Given that Group A strains seemed to represent a tight phylogenetic group in which RscS was not required for colonization or competitive fitness, we next asked whether this group requires the Syp biofilm for colonization. We proceeded to delete the entire *syp* locus in two Group A and two Group B strains and to conduct single-strain colonization analysis. In each strain assayed, the *syp* locus was required for full colonization, and we observed a 2-4 log reduction in CFU per animal in the absence of the *syp* genes, pointing to a critical role for Syp biofilm in these strains (**Fig. 2.6**). In Group A strains in particular, no colonization was detected in the absence of the *syp* locus.

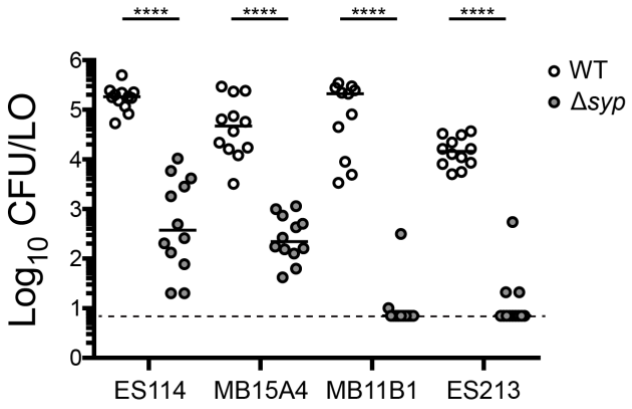


Figure 2.6. Group B and Group A strains require the *syp* locus for robust squid colonization.
Wild type (WT) and Δ*syp* derivatives of the indicated strains were assayed in a single strain colonization assay. Hatchling squid were inoculated with 6.7-32 × 10^2 CFU/ml bacteria (ES114 and MB15A4 backgrounds) or 5.2-8.9 × 10^2 CFU/ml bacteria (MB11B1 and ES213 backgrounds), washed at 3 h and 24 h, and assayed at 48 h. Each dot represents an individual squid. The limit of detection is represented by the dashed line and the horizontal bars represent the median of each set. Strains are MJM1100, MJM3062, MJM2114, MJM3071, MJM1130, MJM3065, MJM1117, and MJM3068. Statistical comparisons by the Mann-Whitney test, **** $p<0.0001$.

Group A strains encode an alternate allele of SypE.

It seemed curious to us that Group A strains do not encode a functional RscS and do not require *rscS* for colonization, yet in many cases Group A strains can outcompete Group B strains (e.g. MB11B1 in **Fig. 2.5B**; and Refs. (134, 135)). We reasoned that if the Syp biofilm had a different regulatory architecture in Group A strains--e.g., constitutively activated or activated by a different regulatory protein--then this could explain the Syp regulation independent of RscS. Genome sequencing of SR5 and MB11B1 did not identify a unique histidine kinase that was likely to directly substitute for RscS (133, 135). Given that the *syp* locus encodes biofilm regulatory proteins, we examined *syp* conservation. We used TBLASTN with the ES114 Syp proteins as queries to determine amino acid conservation in the other *V. fischeri* Group A strain MB11B1, Group C strain SR5, and the *Vibrio vulnificus* type strain ATCC

27562 (168, 169). As shown in Figure 2.7, ES114 SypE, a response regulator and serine kinase/phosphatase that is a negative regulator of the Syp biofilm (119, 161), exhibited the lowest level of conservation among *syp* locus products. *V. vulnificus* does not encode a SypE ortholog (152), as the syntenic (but not homologous) RbdE encodes a predicted ABC transporter substrate-binding protein. The closest hit for SypE was AOT11_RS12130 (9% identity), compared to 7% identity for the RbdE. Due to the reduced conservation at both the strain and species levels, we analyzed *V. fischeri* MB11B1 SypE in greater detail. Examination of the *sypE* coding sequence revealed an apparent -1 frameshift mutation in which the position 33 (guanine in ES114 and adenine in other Group B and C strains examined) is absent in Group A strains (**Fig. 2.7B**). We therefore considered the hypothesis that SypE is nonfunctional in Group A, and that these strains can colonize because they are lacking a functional copy of this negative regulator that is itself regulated by RscS.

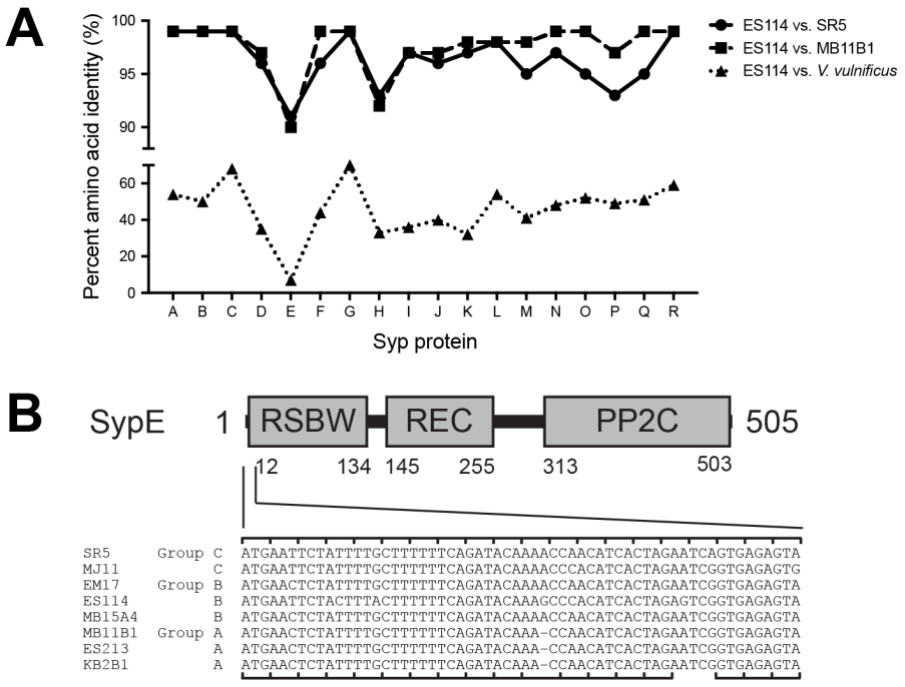


Figure 2.7. Group A strains have a frameshift in *sypE*.
A) Amino acid identity in the Syp locus. Results show the identity from TBLASTN query using the *V. fischeri* ES114 protein sequences as queries against genes in the homologous loci in *V. fischeri* strains or *V. vulnificus* ATCC 27562. The identity for SypE against *V. vulnificus* is plotted for the syntenous RbdE, although this is not the highest TBLASTN hit, as described in the text. **B)** ES114 SypE protein domains. Nucleotides 1-60 in ES114 *sypE* and their homologous sequences in the other Group C, B, and A strains are listed. A -1 frameshift is present in the Group A *sypE* alleles. The ES114 reading frame is noted on the top of the alignment and the Group A reading frame on the bottom, which is predicted to end at the amber stop codon. A possible GTG start codon for the resumption of translation in the ES114 reading frame is present at the position corresponding to the 18th codon in ES114 *sypE*.

To test this hypothesis, we relied on knowledge of the biofilm regulatory pathway from ES114, in which overexpression of SypG produces a wrinkled colony phenotype, but only in strains lacking SypE activity (117, 120). Therefore, we introduced the SypG-overexpressing plasmid pEAH73 into strains as a measure of whether the SypE pathway was intact. In the ES114 strain background, we observed cohesive wrinkled colony formation at 48 h in an ES114 Δ*sypE* strain, but not in the wild-type parent (**Fig. 2.8A**). If the *sypE* frameshift observed in MB11B1 led to a loss of function, then introduction of that frameshift into ES114 would lead to a strain that is equivalent to the Δ*sypE* strain. We constructed this strain and upon SypG overexpression we observed wrinkled colony formation. Surprisingly, the biofilm phenotype was observed earlier

(i.e., by 24 h) and leads to more defined colony biofilm architecture at 48 h. While the lack of SypE leads to increased and more rapid biofilm formation, in this assay we observed an even greater increase as a result of the frameshift in *sypE* (**Fig. 2.8A**).

We proceeded to conduct a similar assay in the MB11B1 strain background. The colony biofilm phenotypes were muted compared to the ES114 background, but the pattern observed is the same. Strains lacking the additional nucleotide at position 33 (i.e., the native MB11B1 allele) exhibited the strongest cohesion, whereas strains with the nucleotide to mimic ES114 *sypE* (i.e., added back in MB11B1 *sypE*(nt::33G)) were not cohesive (**Fig. 2.8B**). These results argue that a novel allele of *sypE* is found in Group A strains and this allele results in more substantial biofilm formation than in a Δ*sypE* strain.

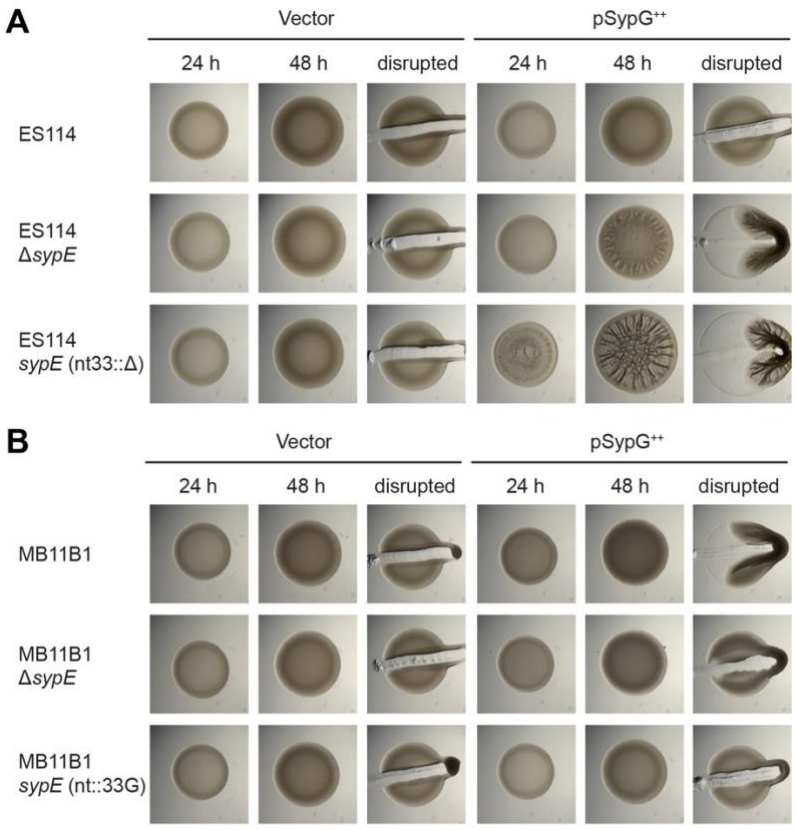


Figure 2.8. The MB11B1 *sypE* frameshift leads to an enhanced biofilm phenotype upon SypG overexpression.
Spot assays of strains carrying the pKV69 vector or pEAH73 SypG overexpression plasmid. **A)** ES114 strain background. Strains lacking SypE produce a wrinkled colony phenotype upon SypG overexpression. Deletion of nucleotide 33 in *sypE* to mimic the Group A frameshift led to earlier wrinkling and a more pronounced colony biofilm at 48 h. Strains: MJM1104, MJM3455, MJM3418, MJM3419, MJM3364, and MJM3365. **B)** Group A strain MB11B1, which naturally carries a -1 frameshift in *sypE*, exhibits a cohesive phenotype at 48 h with overexpression of SypG. Deletion of *sypE* reduces this phenotype, and repairing the frameshift by addition of a guanosine at nucleotide 33 further reduces the cohesiveness of the spot. Strains: MJM3370, MJM3371, MJM3411, MJM3412, MJM3398, and MJM3399.

Our finding that the MB11B1 *sypE* allele promotes biofilm formation bolstered the model that this allele contributes to the ability of MB11B1 to colonize squid independent of RscS. To test this model, we introduced the frameshift into ES114 or "corrected" the frameshift in MB11B1. We then conducted single-strain colonization assays, and in each case the *sypE* allele alone was not sufficient to alter the overall colonization behavior of the strain (**Fig. 2.9**). Therefore, these data suggest that the frameshift in the MB11B1 *sypE* is not sufficient to explain its ability

to colonize independent of RscS, and therefore other regions of SypE and/or other loci in the MB11B1 genome contribute to its ability to colonize independent of RscS.

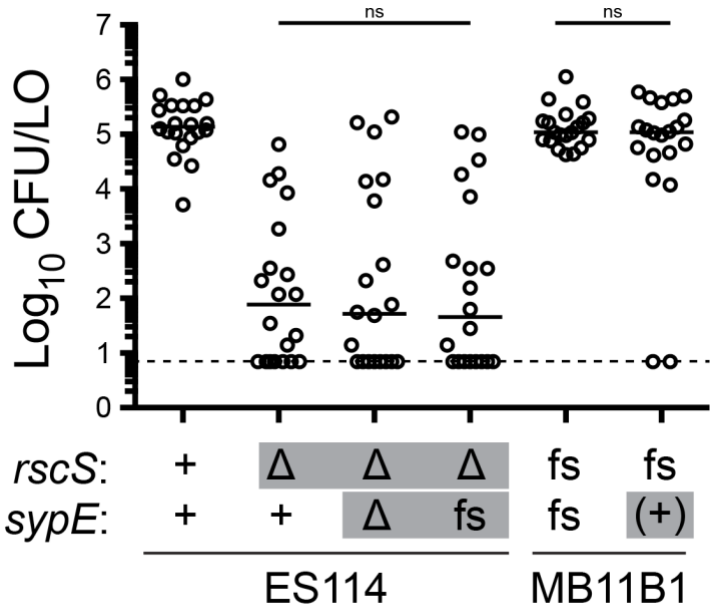


Figure 2.9. The *sypE* -1 frameshift allele is not sufficient to affect colonization ability. The indicated strains were assayed in a single-strain colonization assay. Gray boxes denote alleles distinct from their wild-type background. Frameshift "fs" refers to alleles--relative to an ES114 reference--that lack *rscS* nucleotide A1141, or that lack *sypE* nucleotide G33. The wild-type MB11B1 strain contains natural frameshifts in these loci, and the ES114 nt33::ΔG allele was constructed. Addition back of the nucleotide in MB11B1 *sypE* is denoted as "(+)". Hatchling squid were inoculated with 6.8-8.4 × 10^2 CFU/ml bacteria (MB11B1 background) or 4.0-5.4 × 10^3 CFU/ml bacteria (ES114 background), washed at 3 h and 24 h, and assayed at 48 h. Each dot represents an individual squid. The limit of detection is represented by the dashed line and the horizontal bars represent the median of each set. Strains are MJM1100, MJM3010, MJM4323, MJM3394, MJM1130, and MJM3397. Statistical comparisons by the Mann-Whitney test, ns not significant.

BinK is active in Group A, B, and C strains.

We recently described the histidine kinase, BinK, which negatively regulates *syp* transcription and Syp biofilm formation (73). In ES114, overexpression of BinK impairs the ability of *V. fischeri* to colonize. We therefore reasoned that if BinK could function in Group A strains and acted similarly to repress Syp biofilm, then overexpression of BinK would reduce colonization of

these strains. We introduced the pBinK plasmid (i.e., ES114 *binK* (73)) and asked whether multicopy *binK* would affect colonization. In strain MB11B1, BinK overexpression led to a dramatic reduction in colonization (**Fig. 2.10A**). Therefore, there is a clear effect for BinK overexpression on the colonization of the Group A strain MB11B1.

We attempted to ask the same question in Group C strain SR5, but the pES213-origin plasmids were not retained during squid colonization. Therefore, we instead asked whether deletion of the BinK, a negative regulator of ES114 colonization, has a comparable effect in SR5 (73). We deleted *binK* and observed a 2.4-fold competitive advantage during squid competition (**Fig. 2.10B**), arguing that BinK in this Group C strain is active and performs an inhibitory function similar to that in ES114.

We next examined the colony biofilm phenotype for strains lacking BinK. MB11B1 Δ*binK* exhibited a mild colony biofilm phenotype at 48 h, as evidenced by the cohesiveness of the spot when disrupted with a toothpick (**Fig. 2.10C**). The colonies also exhibited an opaque phenotype. In a minority of experimental replicates, wrinkled colony morphology was evident at 48 h, but in all samples wrinkled colony morphology was visible at 7 d (data not shown). The SR5 Δ*binK* strain also exhibited slightly elevated biofilm morphology at 48 h, though the cells were not as cohesive as those of MB11B1 Δ*binK* (**Fig. 2.10C**). Together, the results in Figure 10 argue that BinK, a factor that has been characterized as a negative regulator of Syp biofilm, plays similar roles in Group A and Group C strains and has a widely-conserved function across the *V. fischeri* evolutionary tree.

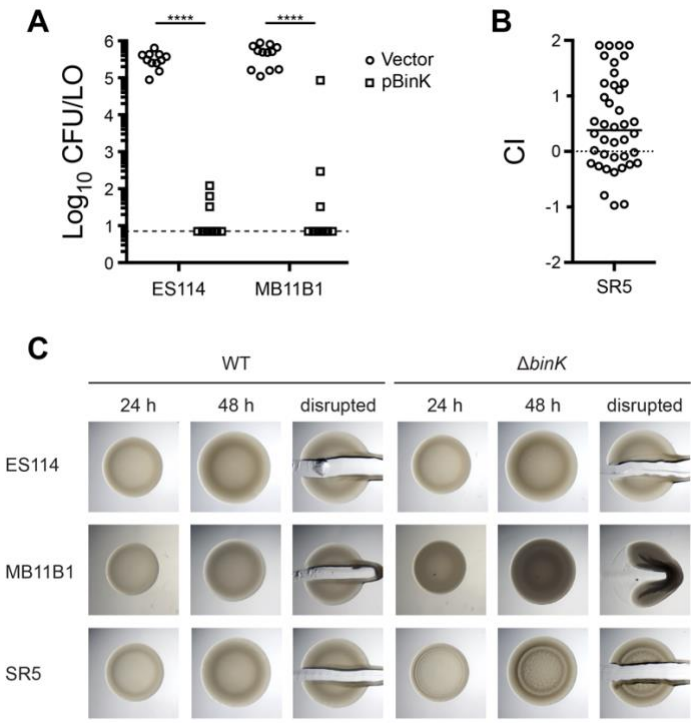


Figure 2.10. BinK is active in Groups A, B, and C.
A) Overexpression of pBinK inhibits colonization in Group A strain MB11B1. Hatchling squid were inoculated with 3.6-6.8 × 10^3 CFU/ml bacteria, washed at 3 h and 24 h, and assayed at 48 h. Each dot represents an individual squid. The limit of detection is represented by the dashed line and the horizontal bars represent the median of each set. The vector control is pVSV104. Strains are MJM1782, MJM2386, MJM2997, and MJM2998. **B)** Deletion of *binK* confers a colonization defect in Group C strain SR5. Strains are MJM1125 and MJM3571; mean inoculum of 7.2 × 10^3 CFU/ml; median competitive index (CI) was 0.38 (i.e., 2.4-fold advantage for the mutant). **C)** Deletion of the native *binK* in MB11B1 yielded opaque and cohesive spots, which are stronger phenotypes than we observe in ES114. Strains are MJM1100, MJM2251, MJM1130, MJM3084, MJM2997, and MJM2998. Statistical comparisons by the Mann-Whitney test, **** p<0.0001.

DISCUSSION

This study examines regulation of a beneficial biofilm that is critical to host colonization specificity in *V. fischeri.* The Syp biofilm was discovered thirteen years ago and has been characterized extensively for its role in facilitating squid colonization by *V. fischeri.* This work establishes that the *syp* locus is required broadly across squid symbionts, and it uncovers three

groups of *V. fischeri* that use different regulatory programs upstream of the *syp* locus. A simplified phylogenetic tree showing key features of squid symbionts in these three groups is shown in Figure 2.11.

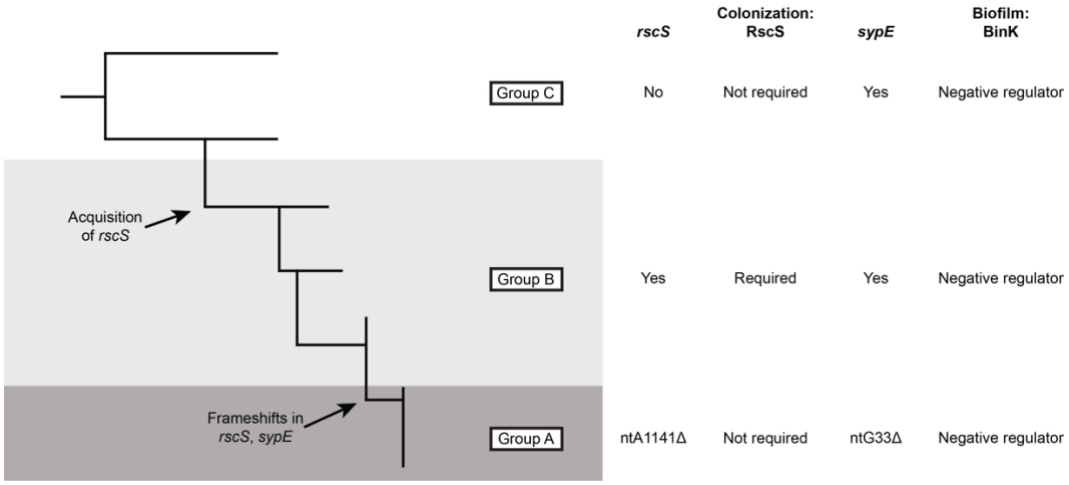


Figure 2.11. Summary model of distinct modes of biofilm formation in squid-colonizing *V. fischeri*.
Phylogenetic tree is simplified from Figure 1, and illustrates key features of squid symbionts in the three groups. Shown are divergent aspects (RscS, SypE) and conserved regulation (BinK). In all groups, the *syp* exopolysaccharide locus is required for squid colonization.

There are three nested evolutionary groups of *V. fischeri* that have been described separately in the literature and here we formalize the nomenclature of Groups C, B, and A. Group A is a monophyletic group, as are Groups AB and ABC (**Fig. 2.1**). This work provides evidence that squid symbionts in each group have a distinct biofilm regulatory architecture. Most *V. fischeri* isolates that have been examined from the ancestral Group C cannot colonize squid; however, those that can colonize do so without the canonical biofilm regulator RscS. We show that the known targets of RscS regulation—genes in the *syp* biofilm locus—are nonetheless required for squid colonization by this group. Group B strains include the well-characterized ES114 strain, which requires RscS and the *syp* locus to colonize squid. Group A strains differ phenotypically and behaviorally from the sister Group B strains (134), and we demonstrate that these strains

have altered biofilm regulation. Group A strains have a frameshift in *rscS* that renders it nonfunctional, and a 1 bp deletion in *sypE*, and we provide evidence that the *sypE* allele promotes biofilm development in the absence of RscS. Additionally, we note that the *sypE* frameshift is not present in SR5, arguing for distinct modes of biofilm regulation in Groups A, B, and C.

At the same time, this study provides evidence that some aspects of biofilm regulation are conserved in diverse squid symbionts, such as the effects of the strong biofilm negative regulator BinK. Published data indicate that evolved BinK alleles can alter colonization of H905 (Group B) and MJ11 (Group C), and that a deletion of MJ11 *binK* leads to enhanced colonization (162). Our experiments in Figure 2.10 show a clear effect for BinK in all three phylogenetic groups. We also observed responsiveness to RscS overexpression in all squid symbionts examined (**Fig. 2.2**). CG101 was the only *V. fischeri* strain examined that did not exhibit a colony biofilm in response to RscS overexpression. CG101 was isolated from the Australian fish *Cleidopus gloriamaris*; based on these findings, we suspect that the strain does not have an intact *syp* locus or otherwise has divergent biofilm regulation.

It remains a formal possibility that the entire *syp* locus is not required in Group A or Group C, but instead that only one or a subset of genes in the locus are needed. Aggregation in squid mucus has been observed for the Group A strain MB13B2, and this aggregation is dependent on *sypQ* (136). In our data we note that Group A strains were completely unable to colonize in the absence of the *syp* locus, unlike the tested Group B & C strains that exhibited reduced colonization in their respective mutants (**Figs. 2.3, 2.6**). Therefore, the simplest explanation is that the *syp* locus is required in divergent strains in a manner similar to how it is used in ES114. We think that the ability to completely delete the *syp* locus is a clean way to ask whether the

locus is required for specific phenotypes, and our strains are likely to be useful tools in probing Syp protein function in diverse *V. fischeri* isolates.

It is intriguing to speculate as to how the two frameshifts in the Group A strains arose, and why the nonfunctional RscS is tolerated in this group. One possible scenario is that the Group A strains acquired a new regulatory input into the Syp pathway, and that the presence of this new regulator bypassed the requirement for RscS. We note that comparative genomic analysis of Hawaiian D (dominant)-type strains--which largely overlap with Group A--revealed an additional 250 kb of genomic DNA compared to other isolates, yielding a large cache of genes that could play a role in this pathway (135). A related possibility is that *rscS*-independent colonization results from altered regulation of the *syp* locus, either due to changes in regulators (e.g. SypF) or sites that are conserved with Group B. An additional possibility is that the *sypE* frameshift arose, enabling Group A strains to colonize independent of *rscS*. Given that correction of this frameshift in MB11B1 does not significantly affect colonization ability (**Fig. 2.9**), this sequence of events seems less likely, and we expect that another regulator in MB11B1 is required for the RscS-independent colonization phenotype. There is evidence that under some conditions LuxU can regulate the *syp* biofilm (170), and as this protein is conserved in *V. fischeri* it may play an important role in Group A or Group C.

Results from two experimental conditions suggest that the Group A strains may have an elevated baseline level of biofilm formation. Our data indicate that in the absence of BinK or upon SypG overexpression, MB11B1 colonies exhibit strong cohesion under conditions in which ES114 does not (**Figs. 2.8, 2.10**). Furthermore, we note that the Group A strain MB11B1, when lacking BinK, also exhibits a darker, or more opaque, colony phenotype (**Fig. 2.10**). This phenotype has been observed in some ES114 mutants (54) but not in the corresponding ES114 Δ*binK* strain (**Fig. 2.10**). The entire colonization lifecycle likely requires a balance between

biofilm formation/cohesion and biofilm dispersal, and these data argue that Group A strains may be more strongly tilted toward the biofilm-producing state. There is evidence that strains lacking BinK exhibit a colonization advantage in the laboratory (73, 162), suggesting that this strategy of more readily forming biofilms may provide a fitness advantage in nature. At the same time, the biofilm negative regulator BinK is conserved among *V. fischeri* strains examined (including MB11B1; **Fig. 2.10**), arguing that there is a benefit to reducing biofilm formation under some conditions.

Our study provides hints as to the role of SypE in MB11B1 and other Group A strains. In ES114, the C-terminus is a PP2C serine kinase domain, whereas the N-terminus of SypE is an RsbW serine phosphatase domain. SypE acts to phosphorylate and dephosphorylate SypA Ser-56, with the unphosphorylated SypA being the active form to promote biofilm development (161). The balance between SypE kinase and phosphatase is modulated by a central two-component receiver domain (161). Our data that the MB11B1 *sypE* allele promotes biofilm formation suggest that the protein is tilted toward the phosphatase activity. In MB11B1, the frameshift early in *sypE* suggests that there is a different start codon and therefore a later start codon. An alternate GTG start codon in MB11B1 occurs corresponding to codon 18 in ES114 *sypE* (**Fig. 2.7**), and this is likely the earliest start for the MB11B1 polypeptide. We attempted to directly identify the SypE N-terminus by mass spectrometry, yet we could not identify the protein from either strain. Additional study is required to elucidate how MB11B1 SypE acts to promote biofilm formation.

V. fischeri strains are valuable symbionts in which to probe the molecular basis to host colonization specificity in animals (76, 131, 164). A paradigm has emerged in which biofilm formation through the RscS-Syp pathway is required for squid colonization but not for fish colonization. This study affirms a role of the Syp biofilm, but at the same time points out

divergent (RscS-independent) regulation in Group C and Group A isolates. In another well-studied example of symbiotic specificity, Rhizobial Nod factors are key to generating specificity with the plant host, yet strains have been identified that do not use this canonical pathway (171, 172). Future work will elaborate on these RscS-independent pathways to determine how non-canonical squid colonization occurs in diverse natural isolates.

MATERIALS & METHODS

Bacterial strains and growth conditions. *V. fischeri* and *E. coli* strains used in this study can be found in **Table 2.1**. *E. coli* strains, used for cloning and conjugation, were grown in Luria-Bertani (LB) medium (25 g Difco LB Broth [BD] per liter). *V. fischeri* strains were grown in Luria-Bertani salt (LBS) medium (25 g Difco LB Broth [BD], 10 g NaCl, and 50 ml 1 M Tris buffer pH 7.0, per liter). Growth media were solidified by adding 15 g Bacto agar (BD) per liter. When necessary, antibiotics (Gold Biotechnology) were added at the following concentrations: tetracycline, 5 µg/ml for *V. fischeri*; erythromycin, 5 µg/ml for *V. fischeri*; kanamycin, 50 µg/ml for *E. coli* and 100 µg/ml for *V. fischeri*; and chloramphenicol, 25 µg/ml for *E. coli*, 2.5 -5 µg/ml for Group B *V. fischeri*, and 1 - 2.5 µg/ml for Group A *V. fischeri*. The two MB11B1 / pKV69 strains listed reflect two separate constructions of this strain, though we have not identified any differences between them.

Phylogenetic analysis. Phylogenetic reconstructions assuming a tree-like topology were created with three methods: maximum parsimony (MP); maximum likelihood (ML); and Bayesian inference (Bayes) as previously described (131, 134). Briefly, MP reconstructions were performed by treating gaps as missing, searching heuristically using random addition, tree-bisection reconnection with a maximum of 8 for swaps, and swapping on best only with 1000 repetitions. For ML and Bayesian analyses, likelihood scores of 1500+ potential evolutionary

models were evaluated using both the corrected and uncorrected Akaike Information Criterion, the Bayesian Information Criterion, and Decision Theory (Performance Based Selection) as implemented by jModelTest2.1 (173). For all information criteria, the most optimal evolutionary model was a symmetric model with a proportion of invariable sites and a gamma distribution of rate heterogeneity (SYM+I+Γ).

ML reconstruction was implemented via PAUP*4.0a163 (174) by treating gaps as missing, searching heuristically using random addition, tree-bisection reconnection for swaps, and swapping on best only with 1000 repetitions. Bayesian inference was done by invoking the 'nst=6' and 'rates=invgamma' and 'statefreqpr=fixed(equal)' settings in the software package MrBayes3.2.6 (175). The Metropolis-coupled Markov chain Monte Carlo (MCMCMC) algorithm used to estimate the posterior probability distribution for the sequences was set up with 'temp=0.2' and one incrementally 'heated' chain with three 'cold' chains; these four chains were replicated two times per analysis to establish convergence of the Markov chains (i.e., 'stationarity' as defined by (176) and interpreted previously in (134)). For this work, stationarity was achieved after approximately 50,000 samples (5,000,000 generations) were collected, with 25% discarded. The ~37,500 samples included were used to construct a 50% majority-rule consensus tree from the sample distribution generated by MCMCMC and assess clades' posterior probabilities. For ML and MP analyses, the statistical confidence in the topology of each reconstruction was assessed using 1000 bootstrap replicates. Phylogenetic trees were visualized with FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree); the final tree was edited for publication with Inkscape 0.91 (http://inkscape.org/) and GIMP 2.8.22 (http://www.gimp.org/).

DNA synthesis and sequencing. Each of the primers listed in **Table 2.3** was synthesized by Integrated DNA Technologies (Coralville, IA). Full inserts from all cloned constructs were verified by Sanger DNA sequencing through ACGT, Inc via the Northwestern University

Feinberg School of Medicine NUSeq Core Facility; or the University of Wisconsin-Madison Biotechnology Center. Sequence data was analyzed with SeqMan Pro (DNAStar software), SnapGene (GSL Biotech), and Benchling.

Construction of gene deletions. Deletions in *V. fischeri* strains ES114 and MB11B1 were made according to the lab's gene deletion protocol: doi:10.5281/zenodo.1470836. In brief, 1.6 kb upstream and 1.6 kb downstream of the targeted gene or locus were cloned into linearized plasmid pEVS79 (amplified with primers pEVS79_rev_690/pEVS79_for_691) using Gibson Assembly (NEBuilder HiFi DNA Assembly cloning kit) with the primer combinations listed in Table S1. The Gibson mix, linking together the upstream and downstream flanking regions, was transformed into *E. coli* on plates containing X-gal, with several white colonies selected for further screening by PCR using primers flanking the upstream/downstream junction (Tables 3 and S1). The resulting plasmid candidate was confirmed by sequencing and conjugated into the *V. fischeri* recipient by tri-parental mating with helper plasmid pEVS104, selecting for the chloramphenicol resistance of the plasmid backbone. *V. fischeri* colonies were first screened for single recombination into the chromosome by maintaining antibiotic resistance in the absence of selection and then screened for double recombination by the loss of both the antibiotic resistant cassette and the gene/locus of interest. Constructs were verified by PCR (Table 3) and sequencing.

Deletion of SR5 *binK* was conducted using Splicing by Overlap Extension PCR (SOE-PCR) and natural transformation (method modified from (177)). Oligos binK-F1 and binK-R1-LUH, and oligos binK-F2-RUH and binK-R2 were used in a PCR with MJM1125 (SR5) genomic DNA as the template to amplify DNA fragments containing ~1 kb of sequence upstream and downstream relative to *binK*, respectively. Using SOE-PCR, these fragments were fused on either side to a third DNA fragment containing an Erm^R cassette, which was amplified using

pHB1 as template and oligos HB41 and HB42. We then used natural transformation with pLostfoX (178) to insert this mutagenic DNA into MJM1125, where the flanking sequences guide the Erm^R cassette to replace *binK*, generating the desired gene deletion. Candidate SR5 Δ*binK* mutants were selected after growth on LBS-Erm5 plates. Oligos binK-F1 and binK-R2, and HB8 and binK-FO were used to screen candidates for the correct deletion scar by PCR, and oligos KMB_036 and KMB_037 were used to confirm the absence of *binK* in the genome. The deletion was verified by Sanger sequencing with primers HB8, HB9, HB42, and HB146. The base plasmid pHB1 contains an erythromycin resistance cassette flanked by FRT sites, and was constructed using oligos HB23 and HB39 with gBlock gHB1 (sequence in Supplementary File S1; Integrated DNA Technologies, Inc.) as template to amplify the Erm^R cassette flanked by HindIII and BamHI sites, which was then cloned into the corresponding site in pUC19.

For most constructs, the deleted genetic material was between the start codon and last six amino acids (179), with two exceptions: the Δ*sypE* in MJM1130 included the ATG that is two amino acids upstream of the predicted start codon, but not the canonical start codon; and the Δ*binK* alleles in MJM1117, MJM1130, and MJM2114, which were constructed to be equivalent to MJM2251 (Δ*binK* in ES114) (73). The Δ*binK* alleles in these strains include the start codon, the next six codons, two codons resulting from ATCGAT (ClaI site), and the last three codons for a predicted 12 amino acid peptide.

Construction of *sypE* alleles. To create *sypE*(ntG33Δ) in MJM1100 and *sypE*(nt33::G) in MJM1130, the single point mutation was created by amplifying the gene in two halves, with the N-terminal portion consisting of approximately 300 bp upstream of *sypE* up through nucleotide 33 and the C-terminal portion consisting of nucleotide 33 and the remaining *sypE* gene. The overlap between the two halves contained the single nucleotide polymorphism in the primers that connected them. The altered *sypE* alleles were initially cloned into plasmid pEVS107

(linearized with primers pEVS107_3837/pEVS107_3838) using Gibson Assembly and then the entire altered *sypE* allele was subcloned into pEVS79 with Gibson Assembly (Table S1). After double recombination of the vector into *V. fischeri*, candidate colonies for the altered *sypE* in MJM1100 were screened with primers ES114_indel_for/ES114_indel_rev. The primer set anneals more strongly to the wildtype *sypE* sequence than to *sypE*(ntG33::Δ). Candidates in the MJM1100 background with a fainter PCR band were sequenced and confirmed to have the *sypE*(ntG33::Δ) allele. For MJM1130, the primer set MB11B1_indel_for/MB11B1_indel_rev anneals more strongly to the *sypE*(nt33::G) allele than to the naturally occurring *sypE* allele and candidates in MJM1130 that contained a more robust PCR band were selected for sequencing to be confirmed as being *sypE*(nt33::G).

Construction of pKG11 *rscS1*(ntA1141::Δ). Plasmid pKG11 encodes an overexpression allele of RscS, termed *rscS1 (85, 165)*. *rscS* nucleotide A1141 was deleted on the plasmid using the Stratagene Quikchange II Site-Directed Mutagenesis Kit with primers rscS_del1F and rscS_del1R. The resulting plasmid, pMJM33, was sequenced with primers MJM-154F and MJM-306R to confirm the single base pair deletion.

Squid colonization. Hatchling *E. scolopes* were colonized by exposure to approximately 3 x 10^3 CFU/ml (ranging from 5.2 x 10^2 - 1.4 x 10^4 CFU/ml; as specified in figure legends) of each strain in a total volume of 40 ml of FSIO (filter-sterilized Instant Ocean) for 3 hours. Squid were then transferred to 100 ml of FSIO to stop the inoculation and then transferred to 40 ml FSIO for an additional 45 hours with a water change at 24 hours post inoculation. For Figure 10A, kanamycin was added to the FSIO to keep selective pressure on the plasmid. After 48 hours of colonization, the squid were euthanized and surface sterilized by storage at -80 ºC, according to standard practices (180). For determination of CFU per light organ, hatchlings were thawed, homogenized, and 50 µl of homogenate dilutions was plated onto LBS plates. Bacterial colonies

from each plate were counted and recorded. Mock treated, uncolonized hatchlings ("apo-symbiotic") were used to determine the limit of detection in the assay. The competitive index (CI) was calculated from the relative CFU of each sample in the output (light organ) versus the input (inoculum) as follows: Log_{10} ((Test strain[light organ] / Control strain[light organ]) / (Test strain[inoculum] / Control strain[inoculum])). For competitions of natural isolates, the Group A strain (or its Δ*rscS* derivative) was the test strain and the Group B strain was the control strain. Colony color was used to enumerate colonies from each--white for Group A strains MB11B1 and ES213; yellow for Group B strains ES114 and MB15A4--along with PCR verification of selected colonies. For competition between SR5 and SR5 Δ*binK*, 100 colonies per squid were patched onto LBS-Erm5 and LBS.

Colony biofilm assays. Bacterial strains were grown in LBS media (Fig. 10C) or LBS-Cam2.5 media (Figs. 2, 8) for approximately 17 hours, then 10 µl (Fig. 2) or 8 µl (Fig. 8, 10C) was spotted onto LBS plates (Fig. 10C) or LBS-Tet5 plates (Figs. 2, 8). Spots were allowed to dry and the plates incubated at 25 °C for 48 hours. Images of the spots were taken at 24 and 48 h post-spotting using a Leica M60 microscope and Leica DFC295 camera. After 48 h of growth, the spots were disrupted using a flat toothpick and imaged similarly.

Analysis of DNA and protein sequences *in silico*. Amino acid sequences for *V. fischeri* ES114 *syp* genes were obtained from RefSeq accession NC_006841.2. Local TBLASTN queries were performed for each protein against nucleotide databases for the following strains, each of which were derived from the RefSeq cds_from_genomic.fna file: *V. fischeri* SR5 (GCA_000241785.1), *V. fischeri* MB11B1 (GCA_001640385.1) and *V. vulnificus* ATCC27562 (GCA_002224265.1). Percent amino acid identity was calculated as the identity in the BLAST query divided by the length of the amino acid sequence in ES114. Domain information is from the PFAM database (181).

ACKNOWLEDGMENTS

The authors thank Elizabeth Bacon, Jacklyn Duple, Cheeneng Moua, Lynn Naughton, Olivia Sauls, and Denise Tarnowski for assistance with experiments.

TABLES

Table 2.1. Bacterial strains.

Strain	Genotype	Source/Reference
V. fischeri		
MJM1059	MJ11	(76, 182)
MJM1100	ES114	(101)
MJM1104	ES114 (MJM1100) / pKV69	This study
MJM1106	ES114 (MJM1100) / pKG11	This study
MJM1109	MJ11 (MJM1059) / pKV69	This study
MJM1111	MJ11 (MJM1059) / pKG11	This study
MJM1114	MJ12	(182)
MJM1115	CG101	(76)
MJM1117	ES213	(183)
MJM1119	EM18	(76, 182)

MJM1120	EM24	(127, 182)
MJM1121	EM30	(182)
MJM1122	WH1	(184)
MJM1125	SR5	(128)
MJM1126	SA1	(128)
MJM1127	KB1A97	(129)
MJM1128	KB2B1	(129)
MJM1129	KB5A1	(129)
MJM1130	MB11B1	(129)
MJM1136	EM17	(127)
MJM1147	*mjapo.*6.1	(131)
MJM1149	*mjapo*.7.1	(131)
MJM1151	*mjapo*.8.1	(131)
MJM1153	*mjapo*.9.1	(131)
MJM1219	*mjapo*.8.1 / pKV69	This study
MJM1221	*mjapo*.8.1 / pKG11	This study
MJM1238	MJ12 (MJM1114) / pKV69	This study

MJM1239	MJ12 (MJM1114) / pKG11	This study
MJM1240	SR5 (MJM1125) / pKV69	This study
MJM1241	SR5 (MJM1125) / pKG11	This study
MJM1242	SA1 (MJM1126) / pKV69	This study
MJM1243	SA1 (MJM1126) / pKG11	This study
MJM1244	MB11B1 (MJM1130) / pKV69	This study
MJM1245	MB11B1 (MJM1130) / pKG11	This study
MJM1246	EM17 (MJM1136) / pKV69	This study
MJM1247	EM17 (MJM1136) / pKG11	This study
MJM1254	KB1A97 (MJM1127) / pKV69	This study
MJM1255	KB1A97 (MJM1127) / pKG11	This study
MJM1256	KB2B1 (MJM1128) / pKV69	This study
MJM1257	KB2B1 (MJM1128) / pKG11	This study
MJM1258	KB5A1 (MJM1129) / pKV69	This study
MJM1259	KB5A1 (MJM1129) / pKG11	This study
MJM1260	ES213 (MJM1117) / pKV69	This study
MJM1261	ES213 (MJM1117) / pKG11	This study

MJM1266	EM18 (MJM1119) / pKV69	This study
MJM1267	EM18 (MJM1119) / pKG11	This study
MJM1268	EM24 (MJM1120) / pKV69	This study
MJM1269	EM24 (MJM1120) / pKG11	This study
MJM1270	EM30 (MJM1121) / pKV69	This study
MJM1271	EM30 (MJM1121) / pKG11	This study
MJM1272	*mjapo*.6.1 (MJM1147) / pKV69	This study
MJM1273	*mjapo*.6.1 (MJM1147) / pKG11	This study
MJM1274	*mjapo*.7.1 (MJM1149) / pKV69	This study
MJM1275	*mjapo*.7.1 (MJM1149) / pKG11	This study
MJM1276	*mjapo*.9.1 (MJM1151) / pKV69	This study
MJM1277	*mjapo*.9.1 (MJM1151) / pKG11	This study
MJM1278	CG101 (MJM1115) / pKV69	This study
MJM1279	CG101 (MJM1115) / pKG11	This study
MJM1280	WH1 (MJM1122) / pKV69	This study
MJM1281	WH1 (MJM1122) / pKG11	This study
MJM1782	ES114 (MJM1100) pVSV104	(73)

MJM2114	MB15A4	(129)
MJM2226	ES114 (MJM1100) / pMJM33	This study
MJM2251	ES114 (MJM1100) Δ*binK*	(73)
MJM2386	ES114 (MJM1100) / pBinK	This study
MJM2997	MB11B1 (MJM1130) / pVSV104	This study
MJM2998	MB11B1 (MJM1130) / pBinK	This study
MJM2999	ES213 (MJM1117) / pVSV104	This study
MJM3000	ES213 (MJM1117) / pBinK	This study
MJM3010	ES114 (MJM1100) Δ*rscS*	This study
MJM3017	ES213 (MJM1117) Δ*rscS*	This study
MJM3042	MB15A4 (MJM2114) Δ*rscS*	This study
MJM3046	MB11B1 (MJM1130) Δ*rscS*	This study
MJM3062	ES114 (MJM1100) Δ*syp*	This study
MJM3065	MB11B1 (MJM1130) Δ*syp*	This study
MJM3068	ES213 (MJM1117) Δ*syp*	This study
MJM3071	MB15A4 (MJM2114) Δ*syp*	This study
MJM3084	MB11B1 (MJM1130) Δ*binK*	This study

MJM3354	ES114 (MJM1100) *sypE*(ntG33Δ)	This study
MJM3364	ES114 (MJM1100) *sypE*(ntG33Δ) / pKV69	This study
MJM3365	ES114 (MJM1100) *sypE*(ntG33Δ) / pEAH73	This study
MJM3370	MB11B1 (MJM1130) / pKV69	This study
MJM3371	MB11B1 (MJM1130) / pEAH73	This study
MJM3394	ES114 (MJM1100) Δ*rscS* *sypE*(ntG33Δ)	This study
MJM3397	MB11B1 (MJM1130) *sypE*(nt33::G)	This study
MJM3398	MB11B1 (MJM1130) *sypE*(nt33::G) / pKV69	This study
MJM3399	MB11B1 (MJM1130) *sypE*(nt33::G) / pEAH73	This study
MJM3410	MB11B1 (MJM1130) Δ*sypE*	This study
MJM3411	MB11B1 (MJM1130) Δ*sypE* / pKV69	This study
MJM3412	MB11B1 (MJM1130) Δ*sypE* / pEAH73	This study
MJM3417	ES114 (MJM1100) Δ*sypE*	This study
MJM3418	ES114 (MJM1100) Δ*sypE* / pKV69	This study
MJM3419	ES114 (MJM1100) Δ*sypE* / pEAH73	This study
MJM3423	ES114 (MJM1100) Δ*rscS* Δ*sypE*	This study
MJM3455	ES114 (MJM1100) / pEAH73	This study

MJM3501	SR5 (MJM1125) Δ*syp*	This study
MJM3751	SR5 (MJM1125) ΔbinK::*erm*	This study
E. coli		
MJM534	CC118 λpir / pEVS104	(185)
MJM537	DH5α λpir	Lab stock
MJM570	DH5α / pEVS79	(185)
MJM580	DH5α λpir / pVSV104	(186)
MJM581	DH5α / pKV69	(111)
MJM583	DH5α / pKG11	(85)
MJM639	XL1-Blue / pMJM33	This study
MJM658	BW23474 / pEVS107	(95)
MJM2384	DH5α λpir / pBinK	(73)
MJM2540	KV5264 / pEAH73	(117)
MJM3008	DH5α / pEVS79-Δ*rscS*[MJM1100]	This study
MJM3014	DH5α λpir / pEVS79-Δ*rscS*[MJM1117]	This study
MJM3039	DH5α λpir / pEVS79-Δ*rscS*[MJM2114]	This study
MJM3043	DH5α λpir / pEVS79-Δ*rscS*[MJM1130]	This study

MJM3060	NEB5α / pEVS79-Δ*syp*[MJM1100]	This study
MJM3063	NEB5α / pEVS79-Δ*syp*[MJM1130]	This study
MJM3066	DH5α λpir / pEVS79-Δ*syp[*MJM1117]	This study
MJM3069	DH5α λpir / pEVS79-Δ*syp*[MJM2114]	This study
MJM3082	NEB5α / pEVS79-Δ*binK*[MJM1130]	This study
MJM3287	NEB5α / pHB1	This study
MJM3338	DH5α λpir / pEVS107-*sypE*[MJM1130](nt33::G)	This study
MJM3340	DH5α λpir / pEVS107-*sypE*[MJM1100](ntG33Δ)	This study
MJM3351	NEB5α / pEVS79-*sypE*[MJM1130](nt33::G)	This study
MJM3352	NEB5α / pEVS79-*sypE*[MJM1100](ntG33Δ)	This study
MJM3409	NEB5α / pEVS79-Δ*sypE*[MJM1130]	This study
MJM3416	NEB5α / pEVS79-Δ*sypE*[MJM1100]	This study

Table 2.2. Plasmids.

Plasmid	**Relevant genotype**	**Source/Reference**
pEVS79	Vector backbone (Cam^R) for deletion construction	(185)
pKV69	Vector backbone (Cam^R/Tet^R)	(111)

pKG11	pKV69 carrying *rscS1*	(85)
pMJM33	pKG11 *rscS1*(ntA1141::Δ)	This study
pEVS104	Conjugation helper plasmid (Kan^R)	(185)
pEVS107	Mini-Tn7 mobilizable vector (Erm^R/Kan^R)	(95)
pEAH73	pKV69 carrying *sypG* from ES114	(117)
pVSV104	Complementation vector (Kan^R)	(186)
pBinK	pVSV104 carrying *binK* from MJM1100	(73)
pHB1	pUC19 FRT-*erm*-FRT	This study
pEVS79-Δ*rscS*[MJM1100]	pEVS79 carrying 1.6 kb US/1.6 kb DS of *rscS* from MJM1100	This study
pEVS79-Δ*rscS*[MJM1117]	pEVS79 carrying 1.6 kb US/1.6 kb DS of *rscS* from MJM1117	This study
pEVS79-Δ*rscS*[MJM2114]	pEVS79 carrying 1.6 kb US/1.6 kb DS of *rscS* from MJM2114	This study
DH5α λpir / pEVS79-Δ*rscS*[MJM1130]	pEVS79 carrying 1.6 kb US/1.6 kb DS of *rscS* from MJM1130	This study
pEVS79-Δ*syp*[MJM1100]	pEVS79 carrying 1.6 kb US of *sypA*/1.6 kb DS of *sypR* from MJM1100	This study
pEVS79-Δ*syp*[MJM1130]	pEVS79 carrying 1.6 kb US of *sypA*/1.6 kb DS of *sypR* from MJM1130	This study
pEVS79-Δ*syp[*MJM1117]	pEVS79 carrying 1.6 kb US of *sypA*/1.6 kb DS of *sypR* from MJM1117	This study
pEVS79-Δ*syp*[MJM2114]	pEVS79 carrying 1.6 kb US of *sypA*/1.6 kb DS of *sypR* from MJM2114	This study

pEVS79-Δ*binK*[MJM1130]	pEVS79 carrying 1.6 kb US/1.6 kb DS of *binK* from MJM1130	This study
pEVS107-*sypE*[MJM1130](nt33::G)	pEVS107 carrying the *sypE*(nt33::G) allele from MJM1130	This study
pEVS107-*sypE*[MJM1100](ntG33Δ)	pEVS107 carrying the *sypE*(ntG33Δ) allele from MJM1100	This study
pEVS79-*sypE*[MJM1130](nt33::G)	pEVS79 carrying the *sypE*(nt33::G) allele from MJM1130	This study
pEVS79-*sypE*[MJM1100](ntG33Δ)	pEVS79 carrying the *sypE*(ntG33Δ) allele from MJM1100	This study
pEVS79-Δ*sypE*[MJM1130]	pEVS79 carrying 1.6 kb US/1.6 kb DS of *sypE* from MJM1130	This study
pEVS79-Δ*sypE*[MJM1100]	pEVS79 carrying 1.6 kb US/1.6 kb DS of *sypE* from MJM1100	This study

Table 2.3. DNA oligonucleotides for PCR amplification and sequencing.

Primer name	**Sequence (5' to 3')**
DAT_015F	ACCAAGAAGCAGTACGACGATTAT
ES114_DS_ver	GGATGTTTTAGATGTTGCGG
ES114_indel_for	TTACTTTTTTCAGATACAAAGCCC
ES114_indel_rev	GTTGTTCTGATAGTGCGTGA
ES114_US_ver	ATCAACTCAAGAAACTCCCC
for_ver_sypE	CCGGCTCAAACTATTGCAG
Gib_ES114_binK_DS_for	attaatcgatGCGTATACATAAATAATGATTCATATATAC

Gib_ES114_binK_DS_rev	gcaggaattcgatatcaagcTTTCAATACTGTGTTTTTATGC
Gib_ES114_binK_US_for	gaggtcgacggtatcgataaGAGCCTTTTAAATCCCCTAAC
Gib_ES114_binK_US_rev	atgtatacgcATCGATTAATGACATATTATTATTCATAAAAAAC
Gib_ES114_rscS_DS_for	taatgcaatgGAGAAGTATGAAACACAATAAAC
Gib_ES114_rscS_DS_rev	gcaggaattcgatatcaagcAAAAATACATTGTTGCACTTG
Gib_ES114_rscS_US_for	gaggtcgacggtatcgataaGACGTCTAAAACTGAATCG
Gib_ES114_rscS_US_rev	catacttctcCATTGCATTAGCTCCTATAAAATAG
Gib_ES114_syp_DS_for	gcttattatgATATTTGCTCGAGGCCAATAAAAAC
Gib_ES114_syp_DS_rev	gcaggaattcgatatcaagcTGGTGAATGTAGGATCCAC
Gib_ES114_syp_US_for	gaggtcgacggtatcgataaCAACCGTAGCGCCAAATG
Gib_ES114_syp_US_rev	gagcaaatatCATAATAAGCTCCTAGGGAATAATC
Gib_ES114_sypE_C_for	cagatacaaaCCCACATCACTAGAGTCG
Gib_ES114_sypE_C_rev	ctagtggccaggtacctcgaAATTAAGCTTCCATCTTCAC
Gib_ES114_sypE_DS_for	tgtaatcatgCTGTTAATTGAGAATCAATAAAAAG
Gib_ES114_sypE_DS_rev	caactctttttccgaaggtaTTGAGTAACCGGCATAATTTAG
Gib_ES114_sypE_N_for	tagagggccctaggcgcgccTGTTTCACAACTCAATACC
Gib_ES114_sypE_N_rev	gtgatgtgggTTTGTATCTGAAAAAAGTAAAGTAG

Gib_ES114_sypE_US_for	gaggtcgacggtatcgataaTGGTCAGATGAAATGTCATTTTTAG
Gib_ES114_sypE_US_rev	caattaacagCATGATTACACCACTGTTG
Gib_ES213_rscS_US_rev	catacttctcCATTGTATTAGCTCCTATAAAATAG
Gib_MB11B1_syp_DS_for	gcttattatgATATTTGCTCGAGGTCAATAAAAG
Gib_MB11B1_syp_US_for	gaggtcgacggtatcgataaGCACACTGATAACTAAATTATTAC
Gib_MB11B1_syp_US_rev	gagcaaatatCATAATAAGCTCCTAGGG
Gib_MB11B1_sypE_C_for	cagatacaaaGCCAACATCACTAGAATC
Gib_MB11B1_sypE_C_rev	ctagtggccaggtacctcgaTCAACAATTAAGCTTCCATC
Gib_MB11B1_sypE_DS_for	cagtggtatgCTGTTAATTGAAAACCAATAGC
Gib_MB11B1_sypE_DS_rev	gcaggaattcgatatcaagcATTTAGGATGTTTTTAATAACAATTTG
Gib_MB11B1_sypE_N_for	tagagggccctaggcgcgccAGTTTCACAACTCAATACTAATAATATTC
Gib_MB11B1_sypE_N_rev	tgatgttggcTTTGTATCTGAAAAAAGCAAAATAG
Gib_MB11B1_sypE_US_for	gaggtcgacggtatcgataaGAATGGTCAGATGAAATGTC
Gib_MB11B1_sypE_US_rev	caattaacagCATACCACTGTTGATAAAAATC
Gib_pEVS79_ES_sypE_for	gaggtcgacggtatcgataaTGTTTCACAACTCAATACC
Gib_pEVS79_ES_sypE_rev	gcaggaattcgatatcaagcAATTAAGCTTCCATCTTCAC
Gib_pEVS79_MB_sypE_for	gaggtcgacggtatcgataaAGTTTCACAACTCAATACTAATAATATTC

Gib_pEVS79_MB_sypE_rev	gcaggaattcgatatcaagcTCAACAATTAAGCTTCCATC
Gib_SR5_syp_DS_for	gcttattatgATATTTGCTCGAGGACAATAAAAAG
Gib_SR5_syp_DS_rev	gcaggaattcgatatcaagcTGGTGAGTGTAGAATCCATTC
Gib_SR5_syp_US_for	gaggtcgacggtatcgataaAACCGTAGCGCCAAATGG
Gib_SR5_syp_US_rev	gagcaaatatCATAATAAGCTCCTAGGGAATAATCC
HB8	ACAAAATTTTAAGATACTGCACTATCAACACACTCTTAAG
HB9	GGGAGGAAATAATCTAGAATGCGAGAGTAGG
HB23	TTGGAGAGCCAGCTGCGTTCGCTAA
HB39	TAGGAAGCTTACGAGACGAGCTTCTTATATATGCTTCGCCAGGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCTTAGAAGCAAACTTAAGAGTGTG
HB41	CGATCTTGTGGGTAGAGACATCCAGGTCAAGTCCAGCCCCGCTCTAGTTTGGGAATCAAGTGCATGAGCGCTGAAG
HB42	ACGAGACGAGCTTCTTATATATGCTTCGCCAG
HB146	CGATCTTGTGGGTAGAGACATC
binK-F1	GAAATTACCATGGAGCCAACAGCAAGAC
binK-R1-LUH	ctggcgaagcatatataagaagctcgtctcgtCATAAAAAACCTAGCGCTTTATTTGTAGATATAATTATTAACTATAATCGC
binK-F2-RUH	gacttgacctggatgtctctacccacaagatcgCGCTCATTGTATCTATAGAGTATGTACTGAGTTACG
binK-R2	GGCATCATTATGGCAACCATTAAAGACG

binK-FO	CCGTTAATACTGGATTATTCGCTTGAATTTGAACG
KMB_036	CCACAATAGCAGAATACAAATTCGCTG
KMB_037	CTCAAAATGACAGTCAGAGTATCGTAGGC
JFB_287	ATGGAGTTTCTACGTCAACCAGAA
JFB_287_MB11B1	ATGGAGTTTTTACGTCAACCAGAG
JFB_288	TGTTATAACGATTACATGGCAGCG
JFB_365	GGAAAGAGAATGATTAAG
M13for	GTAAAACGACGGCCAG
M13rev	CAGGAAACAGCTATGAC
MB11B1_indel_for	GCTTTTTTCAGATACAAAGCCA
MB11B1_indel_rev	ATACCTGATGGAAACGACCT
MJM-154F	TAAAAAGGGAATTAATCCGC
MJM-306R	AACTCTAACCAAGAAGCA
pEVS107_3837	GGCGCGCCTAGGGCCCTC
pEVS107_3838	TCGAGGTACCTGGCCACTAG
pEVS79_for_691	GCTTGATATCGAATTCCTG
pEVS79_rev_690	TTATCGATACCGTCGACC

rev_ver_sypE	TTCACCATGAGTGCCAAATC
rscS_del1F	CTTATCTTCTAGTTCTTTTTTTTAGTGATGTCTCTTTC TACGGC
rscS_del1R	GCCGTAGAAAGAGACATCACTAAAAAAAAGAACTAG AAGATAAG
rscS_ver_1	GTAATTCAGTAATGCTACC
rscS_ver_2	GTCGCACCGTCAGGTATA
rscS_ver_3	AAGAAATTATTCGCTACC
rscS_ver_4	AGTTAGTAGGCCATTACG
SR5_syp_ver_for	TAGGCGTATCAAAAACCACCT
SR5_syp_ver_rev	TCAGGAATGTCGATGGCAG
Syp_ver_DS_rev	ATCGAGCATATTTTGCCAATC
Syp_ver_US_for	ACCTATCAACTCTTAAGTCGATTC
syp4F	TGAGGATCCCATCGTGCCATA
syp4R	AGCTCCTTTGCAATGTTTGCTT
syp5F	TATTAGGCCGTTTCCACCAGG
syp5F-B	TATTAGGTCGTTTCCATCAGG
sypA_out	AACAGGAATTGCGTTTTCAA
US_syp_flank_for	ACCACTGTGATAACTTGCAC

US_syp_flank_rev	ATGAGGCATAACCTGTTCCA

For Gibson assembly primers, capital letters indicate homology to the template. All primers were designed for this study except MJM-154F, MJM-306R (131); JFB_287, JFB_288, and JFB_365 (73); and M13 for, M13 rev.

Chapter 3: The hybrid histidine kinase SypF confers strain specific biofilm formation and host-colonization in *Vibrio fischeri*

Katherine M. Bultman, Andrew C. Luy, Mark J. Mandel

A version of this chapter has been submitted for publication.

Author contributions: KMB and MJM designed experiments, KMB and ACL constructed strains, KMB and ACL conducted experiments, KMB and MJM wrote the manuscript.

ABSTRACT

Microbial symbioses are interwoven into the natural history of animals, playing critical functions in development, immunity, nutrition, and defense. Animals need to acquire environmentally-transmitted symbionts each generation, and carefully-regulated behaviors on both the animal

and microbial sides of the symbiosis ensure reproducibility and specificity in these relationships. Despite a growing appreciation for key roles played by animal microbiomes, the mechanisms by which such specificity is regulated is not well understood. Previously, we identified symbiotic biofilm regulation through hybrid histidine kinase RscS as necessary and sufficient for bobtail squid light organ colonization specificity in luminescent *Vibrio fischeri* bacteria in the North Pacific Ocean. Here, we identify the mechanism by which *V. fischeri* isolate SR5 from the Mediterranean Sea colonizes bobtail squids. In SR5 the same symbiotic biofilm is required. However, RscS is absent in this isolate, which instead encodes a divergent variant of the histidine kinase SypF. SR5 SypF is necessary and sufficient for squid colonization in strains that lack RscS. We determined that the two-component REC domain of SypF, differing in only three amino acids, encodes the specificity. We observed a higher level of biofilm activation in SR5 SypF, suggesting that this variant may be kinase-dominant compared to the orthologous protein in Hawaiian symbionts that requires RscS for efficient host colonization. This study thus reveals a mechanism in which evolution in a single histidine kinase domain has facilitated a dramatic change in host specificity.

SIGNIFICANCE STATEMENT

The molecular mechanisms by which bacteria can evolve broader host ranges and invade new niches are not well understood. This study describes evolution of a single bacterial domain that has determined whether an animal symbiont is able to colonize its host. The outcome of those changes is an upregulated biofilm phenotype. Bacteria in nature are often found in biofilm-associated aggregates and colonies, and this work combines studies on natural isolates with functional work in culture and in the animal host to determine the consequences of this diversification. By understanding the mechanisms by which symbiotic host range is regulated,

this work demonstrates the different evolutionary strategies employed by an animal symbiont for host specificity and host range expansion.

INTRODUCTION

Acquisition of environmentally-transmitted bacterial symbionts requires recruitment and retention of specific microbes from a diverse milieu. This “winnowing” process, as coined by Nyholm and McFall-Ngai (20), results from a series of successive steps in which increasing specificity is imposed by the host, and increasingly limited bacterial species (and even strains) are able to complete the gauntlet of steps required to initiate and persist for successful colonization. Host restrictions include induction of immune responses, production of antimicrobials, and regulation of nutrients to the symbiont (152, 187, 188). To persist, symbionts often must evade immune responses, manipulate the host with effectors, and produce crucial colonization factors (131, 138, 142, 189). Genetic differences between species, and even among strains of the same species, can lead to different outcomes in colonization (27, 135, 136, 143, 189).

A key bacterial behavior that interfaces directly with host immune and selective processes is biofilm formation (57, 58, 190–194). To control the transition from a planktonic state to a sessile and biofilm state, bacteria use a variety of mechanisms. For some, the use of the secondary messenger cyclic-di-GMP drives this life-style switch (55, 56). Two-component signaling is often employed for stimulating biofilm production during host colonization. *Vibrio cholerae* and *Vibrio parahaemolyticus* both utilize the two-component regulators of the quorum sensing pathway for biofilm production (57). In *Pseudomonas aeruginosa*, a tripartite of histidine kinases modulate the switch between a planktonic acute infection and a biofilm chronic infection (70, 71). Other

Vibrios utilize two-component regulation to directly regulate the production of exopolysaccharide loci leading to biofilm formation (57, 152).

A system in which biofilm formation is particularly relevant for host colonization and host-microbe specificity is colonization of the Hawaiian bobtail squid (*Euprymna scolopes*) light organ by marine *Vibrio fischeri* bacteria. Symbiotic biofilm formation is required for squid colonization (19, 132), and acquisition of the biofilm regulator RscS was key to expansion of the *V. fischeri* host range in the North Pacific Ocean (131). During the initial stages of colonization, *V. fischeri* are recruited from the seawater into the ciliated mucus field surrounding the pores to the internal light organ (15, 80). In the host mucus, *V. fischeri* produce the symbiosis exopolysaccharide (Syp) for bacterial cell aggregation that is required for the colonization process (19, 85, 112, 132). The bacteria then disperse from the biofilm and enter the host light organ, where additional steps ensure that only luminescent *V. fischeri* form the long-term symbiotic population in the host (90, 101, 195). At night, bacterial bioluminescence provides counterillumination to camouflage the squid host as it hunts for prey (20, 109, 110, 127). However, to reach the host crypts where they produce the symbiotic light, during each host generation the bacteria must produce biofilm to even initiate the symbiosis and enter the host light organ.

Regulation of the Syp exopolysaccharide has been studied most extensively in the Hawaiian bobtail squid symbiont ES114. The 18 *syp* biofilm genes are encoded in a four-operon locus that is transcribed by $E\sigma^{54}$ aided by the enhancer-binding protein SypG (19, 115, 116). Three membrane-bound hybrid histidine kinases are key to regulating SypG activity: RscS, SypF, and BinK. RscS autophosphorylates at H412 in its dimerization and histidine phosphotransferase (DHp) domain, transfers the phosphoryl group to D709 in its receiver domain, and then transfers the group to H705 in the SypF histidine phosphotransfer (HPt) domain (72). SypF then transfers the phosphoryl group to D53 in the receiver (REC) domain of the downstream response-

regulator SypG, with phospho-SypG stimulating $E\sigma^{54}$-dependent biofilm gene transcription (72, 114). BinK acts as an inhibitor, likely to dephosphorylate SypG via SypF (123).

The activating histidine kinase RscS has been studied as a key determinant for host specificity. Loss of RscS in strain ES114 significantly reduces host colonization. (111, 131, 132). However, RscS is not encoded by all characterized strains, even when the Syp exopolysaccharide is encoded and required, suggesting divergent regulation of Syp production during host colonization (131–133). Most characterized *rscS*$^{-}$ strains were isolated from a fish host, *Monocentrus japonica*, and are unable to colonize squid (131). However, introduction of ES114 *rscS* into Japanese fish isolates confers the ability to colonize squid, providing evidence that *rscS* is necessary for squid colonization and was sufficient to expand the symbiont's host range (131). *V. fischeri* isolates from the Mediterranean bobtail squids *Sepiola robusta* and *Sepiola affinis* also lack *rscS* (128, 131, 132). These isolates are able to colonize the heterologous Hawaiian host, *E. scolopes*, which enables mechanistic studies, and where it has been shown that colonization requires the *syp* locus (132).

Therefore, the current model is that all *V. fischeri* isolates rely on a conserved biofilm machinery for initiating host colonization, yet regulatory diversification has contributed to different upstream control of those biofilm genes in a Mediterranean Sea symbiont versus North Pacific Ocean symbionts. In this study, we define a mechanism by which the Mediterranean squid symbiont *V. fischeri* SR5 has evolved to colonize squid independent of the regulator RscS. We demonstrate that this mechanism is necessary for SR5 colonization and is sufficient to enable colonization by the Hawaiian *V. fischeri* ES114 strain if we delete its *rscS* gene.

RESULTS

SR5 SypF is necessary and sufficient for RscS-independent squid colonization

While all tested strains require the *syp* EPS genes for squid colonization, the Mediterranean symbiont SR5 lacks *rscS* while Hawaiian symbiont ES114 contains *rscS* and requires the gene to colonize the squid host (**Fig. 3.S1**) (132). We therefore sought to determine how Syp biofilm regulation differs between the two strains to enable RscS-independent biofilm activation and colonization in SR5. Previously, we demonstrated that the most divergent gene product encoded from the 18-gene *syp* locus was SypE, a response regulator that acts in a serine kinase/phosphatase pathway (119, 120, 132, 161). SypE is encoded in the *sypEFG* regulatory cassette within the *syp* locus (54, 132). Given the key regulatory roles for the *sypEFG* gene products in symbiotic biofilm regulation, we asked whether the variation in these regulators was responsible for the RscS-independent colonization by SR5. We swapped the *sypEFG* cassette at its native locus between ES114 and SR5. An ES114 strain containing *sypEFG* from SR5 colonized to the same levels as wild-type ES114. In contrast, there was a significant reduction in colonization by an SR5 strain upon exchange of the *sypEFG* cassette from ES114 (**Fig. 3.1A**). Since colonization by ES114 relies on the presence of *rscS*, we next asked if *sypEFG* from SR5 could rescue colonization of an ES114 strain lacking *rscS*. Exchange of the *sypEFG* cassette from SR5 into an ES114 Δ*rscS* strain led to colonization levels similar to wild-type ES114 (**Fig. 3.1B**). These data demonstrate that *sypEFG* from SR5 are both necessary and sufficient to enable RscS-independent colonization.

Since the initial allele swap included all three regulators, we next sought to determine if a single gene was responsible for the colonization phenotype. We exchanged the open reading frame sequence for either *sypE*, *sypF,* or *sypG* from ES114 into its corresponding native locus in SR5.

Colonization with each strain showed that only when *sypF* from ES114 was swapped did we observe a reduction in colonization similar to swapping the entire *sypEFG* cassette (**Fig. 3.1C**). Correspondingly, the SR5 *sypF* allele was able to facilitate colonization of an ES114 Δ*rscS* strain when it was substituted for the ES114 *sypF* allele (**Fig. 3.1D**). Together, these data reveal that *sypF* from SR5 is necessary and sufficient to confer RscS-independent colonization and suggest a key role for SypF in mediating strain-specific colonization behavior.

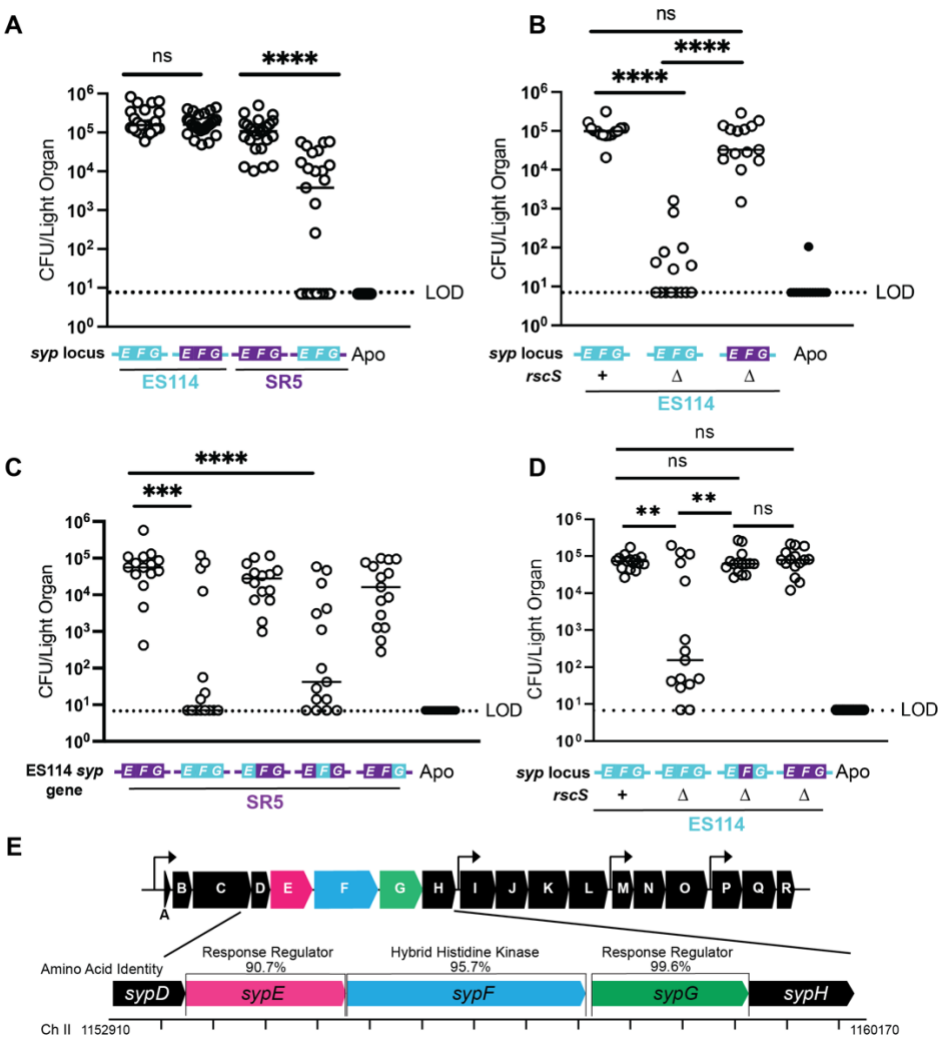


Figure 3.1. The hybrid histidine kinase SypF confers strain-specific colonization.
A-D) Single strain colonization experiments with circles representing individual animals, *N* = 15. Horizontal bars represent the median for each set and the dashed line indicates the limit of detection of 7 CFU/light organ. Hatchling squid were inoculated with **A)** 3.7 x 10^3 - 3 x 10^4 CFU/mL bacteria, **B)** 1.6 x 10^4 - 2.5 x 10^4 CFU/mL bacteria, **C)** 7.3 x 10^3 - 2.5 x 10^4 CFU/mL bacteria and **D)** 4.5 x 10^3 - 1.2 x 10^4 CFU/mL bacteria. Statistical comparisons for all

colonization experiments were done with the Mann-Whitney test. *P*, **** < 0.0001, *P*, *** = 0.0003 *P*, **** < 0.0001, *P*, ** < 0.01 **E)** 18-gene *syp* locus with focus on *sypEFG* including percent amino acid identity of encoded proteins. Numbers indicate chromosomal location of *syp* genes in GenBank accession CP000021.2 with hatch marks indicating every 500 bp.

The SR5 SypF variant is more active to promote symbiotic biofilm than the ES114 variant.

We considered whether the SR5 SypF allele reflected a more active form relative to its ES114 ortholog or whether the protein was being expressed at higher levels. For these studies, we conducted colony biofilm assays in a background lacking the biofilm negative regulator *binK* to provide a greater dynamic range. It has been shown that ES114 Δ*binK* or ES114 Δ*binK* Δ*rscS* strains exhibit a wrinkled colony morphology under calcium inducing conditions that is indicative of Syp production (62). Interestingly, an ES114 Δ*rscS* Δ*binK* strain carrying the SR5 *sypF* allele exhibited robust colony biofilm formation as early as 48 h even without added calcium (**Fig. 3.2A**). Phosphorylation of SypG via SypF (and RscS) has previously been shown to initiate *syp* transcription and biofilm formation (72, 114). We therefore measured *syp* transcription levels using a *sypA'-gfp*$^{+}$ transcriptional reporter in the strain background described above. The SR5 *sypF* allele displayed over 3-fold elevated reporter activity compared to the ES114 *sypF* allele (**Fig. 3.2B**). Therefore, the SR5 SypF variant is able to activate transcription downstream in the absence of RscS, and to a higher level than a corresponding strain carrying the ES114 SypF variant.

Our experiments (**Fig. 3.1CD;Fig. 3.2AB**) varied only the coding sequences of *sypF* and did not alter any flanking sequences. Nonetheless, it remained possible that posttranscriptional control of the *sypF* alleles could impact levels of the two variants, so we asked whether the additional biofilm activation observed from the SR5 allele was due to elevated SypF protein levels. We raised a polyclonal antibody against a conserved peptide in the HPt domains of both proteins

and performed western blot analysis of the ES114 Δ*rscS* Δ*binK* background strain carrying either the ES114 or the SR5 *sypF* allele. We observed approximately two-fold less signal for SypF on the western blot in the strain carrying the SR5 SypF variant compared to the one carrying the ES114 SypF variant (**Fig. 3.2C**), even though the former strain encodes the SypF variant that displays the elevated biofilm activation (**Fig. 3.2B**). Therefore, the increase in biofilm production observed from SR5 *sypF* is not due to an increase in SypF protein levels.

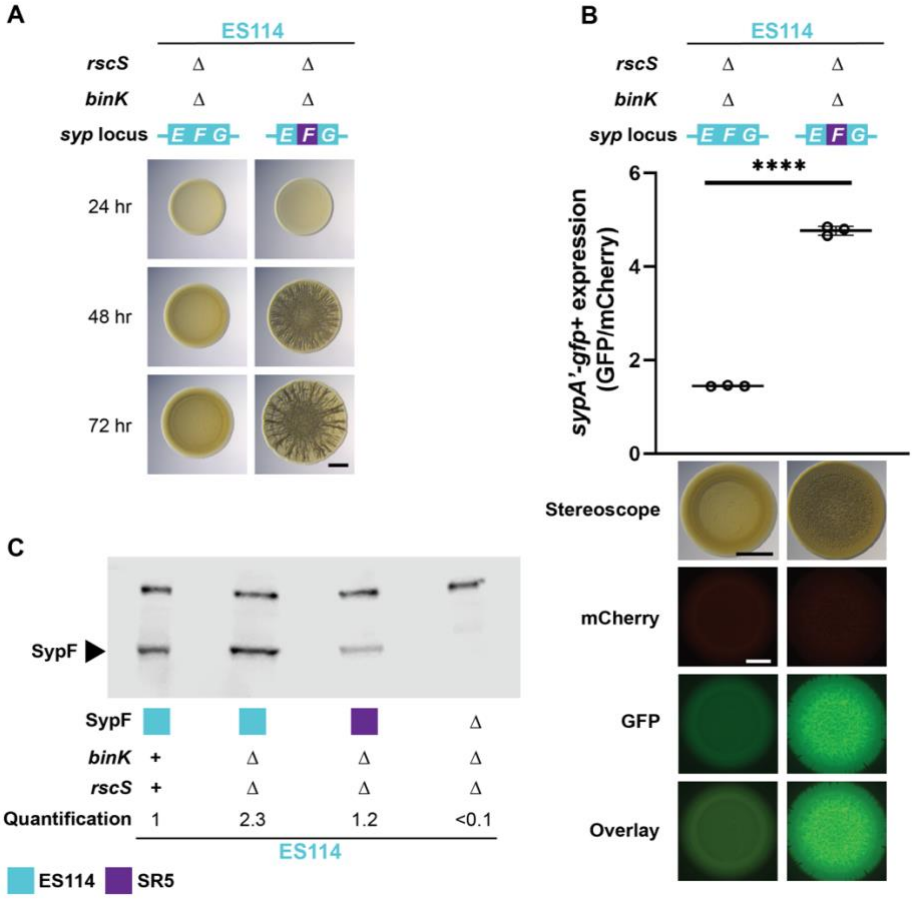


Figure 3.2. The SypF variant from SR5 promotes colony biofilm formation and biofilm gene expression.
A) In background ES114 Δ*rscS* Δ*binK*, elevated biofilm formation is observed when *sypF* is swapped from SR5. Biofilm spot assays of strains grown at 20 ˚C on TBS for 72 h. Scale bar = 2 mm. **B)** Presence of SR5 *sypF* results in increased *sypA'-gfp*+ expression. Strains were grown at 20℃ on TBS for 72 h. Images were taken on a Zeiss Axio Zoom fluorescence microscope and Leica brightfield stereoscope. Fluorescence scale bar = 1 mM; stereoscope scale bar = 2 mm. Graph represents quantification of fluorescence intensity. GFP intensity of the entire spot

was normalized to mCherry intensity. Each spot represents the average of 3 technical replicates and error bars represent standard deviation, *N* = 3. Statistical significance determined by Student's T-test, *P,* **** = <0.0001 **C)** Western blot against SypF using peptide antibody raised against the HPt domain shows no increase in expression with SR5 SypF. Strains were grown in TBS liquid culture at 20℃ for 18 h. Mean gray values of SypF bands were normalized to the total protein from the stain-free gel, then normalized to ES114 value.

The SypF REC domain is sufficient to confer strain-specific regulation

As a membrane-bound hybrid histidine kinase, SypF contains a dimerization and histidine phosphotransferase (DHp) domain, a catalytic and ATP-binding (CA) domain, a receiver (REC) domain, and a histidine phosphotransfer (HPt) domain in the cytoplasm (**Fig. 3.3A**) (72, 196). In ES114, the SypF HPt domain is necessary for biofilm regulation in the presence of RscS with the DHp and REC domains acting as negative regulators (72, 74). We therefore generated chimeric proteins between ES114 and SR5 SypF to isolate the RscS-independent colonization behavior by SR5 SypF (**Fig. 3.3A**). Chimeras were constructed using SR5 SypF as the backbone and domains were sequentially exchanged from ES114 SypF and expressed from the native chromosomal locus. Colony biofilm assays were conducted in an ES114 Δ*rscS* Δ*binK* background as described above and revealed that SypF variants containing the SR5 REC domain conferred the elevated biofilm phenotype of wrinkled colony formation, whereas strains carrying the ES114 REC domain had the smooth colony phenotype (**Fig. 3.3A**). Squid colonization assays for the SypF chimeras were conducted in an ES114 Δ*rscS* background. Again, we observed that the origin of the REC domain was determinative of the outcome: SR5 SypF REC facilitated host colonization, while colonization was reduced by over two orders of magnitude for strains in which SypF contained the ES114 REC domain (**Fig. 3.3B**). These results clearly illustrated that the SR5 SypF REC domain is necessary for the activated biofilm development and the RscS-independent squid colonization observed for the SR5 SypF variant.

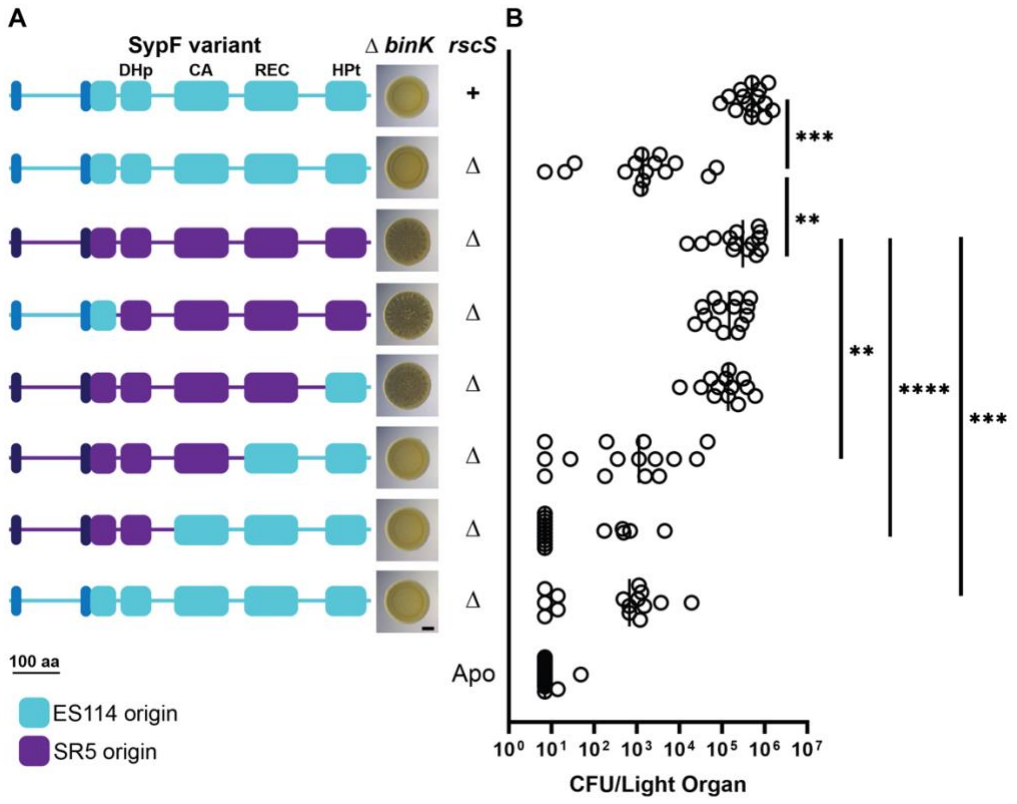


Figure 3.3. The REC domain of SR5 SypF is required for increased biofilm formation and squid colonization.
A) Colony biofilm formation is observed when the SR5 SypF REC domain is encoded as part of the SypF protein in the ES114 Δ*rscS* Δ*binK* background (SypF variants as shown). Biofilm spot assay of strains grown at 20˚C on TBS for 72 h. Scale bar = 2 mm. **B)** Squid colonization is observed when the SR5 SypF REC domain is encoded as part of the SypF protein in the ES114 Δ*rscS* background (SypF variants as shown). Single strain colonization with each circle representing an individual animal, *N* = 15. Horizontal bars represent the median for each set and the dashed line indicates the limit of detection of 7 CFU/light organ. Hatchling squid were inoculated with 2.4 x 10^3 - 1 x 10^4 CFU/mL. Statistical significance determined by a Kruskal-Wallis test. **, $P < 0.001$, ***, $P < 0.001$, ****, $P < 0.0001$.

To determine if the SypF REC domain is sufficient to confer the elevated biofilm phenotype, we introduced only that domain into the ES114 SypF protein. As shown in **Figure 3.4A**, the REC domain swap alone is sufficient for the wrinkled colony phenotype in the colony biofilm assay. Quantification of *syp'-gfp*$^+$ transcriptional reporter additionally revealed over a 3-fold increase when the REC domain from SR5 SypF is present compared to that from ES114 (**Fig. 3.4A**).

There are only three amino acid changes in the region that distinguishes ES114 and SR5 REC domains, yet no single amino acid change is sufficient to confer the SR5 phenotype in the ES114 background (**Fig. 3.S3**). Correspondingly, the SR5 SypF REC domain was able to facilitate colonization of an ES114 Δ*rscS* strain (**Fig. 3.4B**). Together, these data reveal that the SypF REC domain from SR5 is sufficient for strain-specific biofilm production and host colonization.

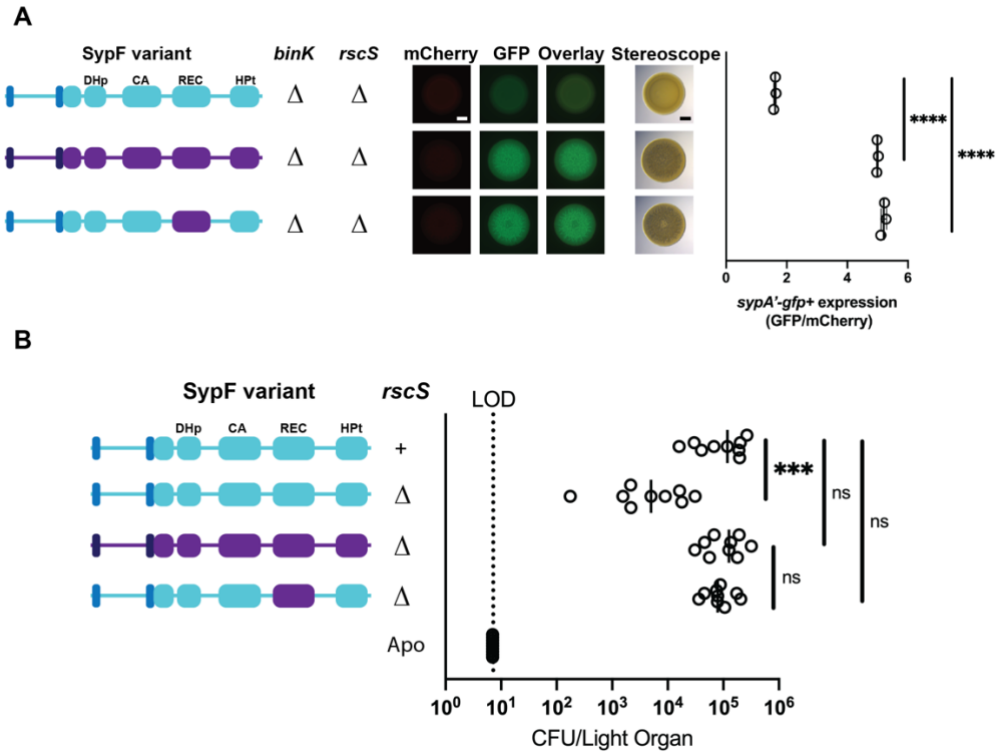


Figure 3.4. The REC domain of SR5 SypF is sufficient to confer elevated *syp* transcription and squid colonization.
A) The SR5 SypF REC domain is sufficient for increased *sypA'-gfp*$^{+}$ expression. Strains were grown at 20℃ on TBS for 72 h. Images were taken on both a Zeiss Axio Zoom fluorescence microscope and Leica brightfield stereoscope. Scale bars = 2 mm. Graph represents quantification of fluorescence intensity. GFP intensity of the entire spot was normalized to mCherry intensity. Each spot represents the average of 3 technical replicates and error bars represent standard deviation. Statistical significance determined by Student's T-test ****, P <0.0001 **B)** The SR5 SypF REC domain is sufficient for squid colonization in an ES114 Δ*rscS* background. Single strain colonization experiments with circles representing individual animals, *N* = 9. Horizontal bars represent the median for each set and the dashed line indicates the limit of detection of 7 CFU/light organ. Hatchling squid were inoculated with 4.4 x 10^2 - 5.7 x 10^3

CFU/mL bacteria. Statistical comparisons for all colonization experiments were done with the Mann-Whitney test. *P*, *** = 0.0005.

DISCUSSION

This work defines a mechanism that has evolved in *V. fischeri* to enable squid colonization, and the first that is independent of the characterized regulator RscS. It is notable that both pathways target the Syp exopolysaccharide that is a critical component of the symbiotic biofilm. We have demonstrated previously that fish symbionts that cannot efficiently colonize the squid light organ can be engineered to become outstanding colonizers by introduction of RscS from ES114 (131). Therefore, the biofilm pathway represents a key regulatory node that determines what animal host niches are available to environmental bacteria. Given the conservation of biofilm formation in the light organ symbioses across *V. fischeri* strains with distinct REC domain sequences from those described (e.g., Mediterranean isolate SA1; **Fig. 3.S4**), we speculate that additional biofilm regulatory pathways and evolutionary trajectories are present and remain to be described.

This study highlights the specificity of a two-component hybrid histidine kinase to impact a fundamental behavior in an animal symbiont. We found that three amino acid differences in the REC domain of SypF distinguishes whether a strain like SR5 can colonize the light organ of a squid host independent of the upstream phosphoryl group donor RscS, or whether a strain like ES114 is dependent on the presence of that upstream regulator. In two-component signaling, REC domains play a key role in the phosphorelay for cellular responses to environmental signals. In canonical histidine kinase-response regulator pairs, the REC domain of the response regulator interacts with the DHp domain in the histidine kinase to transfer phosphoryl groups to a conserved Asp residue in the REC domain (31, 36). This leads to the formation of either

homodimers or homomultimers for function of the effector domains (31, 36). Specific amino acids in and around the active site within the REC domain coordinate the phosphorylation of the Asp residue (31). Changes within REC domains have been shown to alter the canonical phosphotransfer between the DHp and REC domains in cognate TCS pairs (197). An amino acid substitution one residue carboxy-terminal to a conserved REC domain active site residue in the *Escherichia coli* response regulator NarL led to Nar-sensor and acetyl phosphate independent phosphorylation suggested cross-talk with non-cognate sensors (197). Compared to canonical response regulators, tethered REC domains in hybrid histidine kinases are responsible for the transfer of phosphoryl groups from the catalytic core to the HPt domain for downstream phosphorylation of the response regulator (28, 39). It seems likely that the three critical residues that distinguish the two SypF variants impact phosphoryl group equilibrium through the hybrid histidine kinase.

Put together then, we propose the following model. ES114 SypF is phosphatase-dominant, and only upon the influx of phosphoryl groups from RscS is the equilibrium shifted to sufficiently phosphorylate SypG and activate transcription from the *syp* locus. In contrast, we suggest that the mutations identified within the REC domain of SR5 SypF shift the equilibrium of this regulator to be kinase-dominant. Such a model is consistent with our data demonstrating that SR5 SypF increased colony biofilm production and *syp'-gfp*$^{+}$ transcription in ES114 strains that lack *rscS*, as well as with data in ES114 demonstrating negative regulation by SypF in a colony biofilm assay (**Fig. 3.2**) (74). This model is further supported by the data demonstrating that only the SR5 SypF can confer colonization in the absence of the upstream phosphoryl donor RscS.

What is the source of phosphoryl groups in SR5 SypF in the absence of RscS? We find it likely that under the conditions tested, phosphoryl groups flow to SR5 SypG through the SypF HPt domain, from a combination of SypF autophosphorylation at its DHp domain and/or through

other phosphoryl group transfer via SypF. Another hybrid histidine kinase, HahK, has been characterized to act in such a fashion in ES114, and it is conserved in SR5 (91.3% amino acid identity). Given that SR5 SypF functions in an ES114 Δ*rscS* background, our data are consistent with the other components of the pathway being conserved in both strains. Often, gene duplication events of two-component partners lead to the rewiring of signaling pathways (44). In the case of *V. fischeri,* although RscS and SypF are similar inner membrane hybrid histidine kinases, current evidence indicates that RscS was acquired by horizontal gene transfer (131). Therefore, we favor a model in which two distinct lineages of *V. fischeri* evolved distinct, convergent outcomes to activate *syp* biofilm genes and enable squid host colonization.

This study contributes to a body of work that suggests that major host shifts can be undergirded by tiny genomic changes. In *V. fischeri*, acquisition of positive regulator RscS resulted in altered host range for fish isolates that were previously unable to colonize the squid host (131). The *nilABC* cassette of *Xenorhabdus nematophila* is necessary and sufficient for host-specific colonization of *Steinernema carpocapsae* nematodes (141). The initiation of infection threads for nodule formation in legumes is dependent on the production of Nod factors by *Rhizobium*. While the common *nod* genes are interchangeable between species, allelic variation of the transcriptional regulator *nodD* and presence of other host specific *nod* genes alters host specificity across species (22, 139, 153, 154). Similarly, biofilm formation for colonization of the flea gut by *Yersinia pestis* relies on a nonfunctional copy of the negative regulator RcsA compared to the non-insect colonizing *Yersinia pseudotuberculosis* (198). Finally, our work in a strain that lacks the canonical sensor RscS parallels the discovery of two photosynthetic Bradyrhizobia strains that lack Nod genes, and that instead use a distinct signaling pathway to initiate symbiosis in some legume hosts (171).

Overall, this study provides evidence of a hybrid histidine kinase that serves as a key specificity factor for animal host colonization, and thus identifies a second mechanism in *V. fischeri* (after RscS) that has evolved to activate *syp* biofilm genes and enable squid host colonization. We find it likely that combining bioinformatics and functional studies in additional bobtail squid isolates will yield further insights into evolutionary transitions that underlie symbiosis specificity and host range transitions.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. *V. fischeri* and *Escherichia coli* strains used in this study are listed in **Table 3.1**. Plasmids used in this study are listed in **Table 3.2**. *V. fischeri* strains were grown at 25°C in Luria-Bertani salt (LBS) medium (25 g Difco LB broth [BD], 10 g NaCl, and 50 ml 1 M Tris buffer [pH 7.5], per liter) or at 20˚C in tryptone broth salt (TBS) medium (10 g Difco bacto-tryptone [BD], 20 g NaCl, and 50 ml 1 M Tris buffer [pH 7.5], per liter). *E. coli* strains, used for cloning and conjugation, were grown with shaking at 37°C in Luria-Bertani (LB) medium (25 g Difco LB broth [BD] per liter). Growth media were solidified with 1.5% agar (15 g Bacto agar [BD] per liter) as needed. When necessary, antibiotics were added to the medium at the following concentrations: erythromycin (Erm), 5 mg/ml for *V. fischeri*; kanamycin (Kan), 100mg/ml for *V. fischeri* and 50 mg/ml for *E. coli*; and chloramphenicol (Cam), 5 mg/ml for *V. fischeri* and 25 mg/ml for *E. coli*. The *E. coli* strain π3813 containing pKV496 is a thymidine auxotroph and was grown in LB with 50mg/ml kanamycin supplemented with 0.3 mM thymidine (177, 199, 200).

DNA synthesis and sequencing. Each of the primers listed in **Table 3.3** was synthesized by Integrated DNA Technologies. Site directed mutagenesis primers were designed using the NEBaseChanger online tool (NEB). Full plasmids from cloned constructs were verified by

Oxford Nanopore Technology (ONT) DNA sequencing at Plasmidsaurus (Eugene, OR) or the full insert sequence was verified using Sanger DNA sequencing at Functional Biosciences. Sequencing of allelic exchange-generated strains was verified by ONT whole-genome DNA sequencing at Plasmidsaurus, using linear amplicon sequencing, or whole-genome sequencing with Illumina whole-genome sequencing by MiGS (Pittsburg, PA). Sequence data were analyzed with SnapGene (GSL Biotech). For cloning PCR reactions, we used Q5 high-fidelity DNA polymerase (NEB). For sequencing PCR reactions, we used OneTaq (NEB). For diagnostic PCR reactions, we used GoTaq polymerase (Promega).

Construction of *sypEFG* swap strains. The exchange of *sypEFG* from ES114 into SR5 and vice versa was done using an allelic exchange approach modified from the laboratory's gene deletion protocol (https://doi.org/10.5281/zenodo.1470836).The previously generated pEVS79 vector was purified and used as a template for all allelic exchange vectors. For the exchange of *sypEFG* between strains, pEVS79 backbone was amplified using primer pair RYI097 F/RYI097 R. Approximately 1.6 Kb of DNA upstream and downstream of *sypEFG* was amplified using KMB_262/KMB_263 and KMB_264/KMB_265 for ES114 and KMB_270/KMB_271 and KMB_272/KMB_273 for SR5. Fragments were cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB). The constructed plasmids were transformed into chemically competent DH5α λpir cells and the resulting plasmids were sequenced by Plasmidsaurus. Plasmids with the correct sequence were named pKMB010 and pKMB011 respectively. *sypEFG* was amplified using KMB_268/KMB_269 for ES114, and KMB_276/KMB_277 for SR5. Vector backbones for pKMB010 and pKMB011 were amplified using KMB_266/KMB_267 and KMB_274/KMB_275 respectively. The ES114 *sypEFG* fragment was cloned into pKMB011 and the SR5 *sypEFG* fragment was cloned into pKMB010 using Gibson Assembly. The constructed plasmids were transformed into chemically competent DH5α λpir cells and the resulting plasmids were sequenced by Plasmidsaurus. After double

recombination of these plasmids into *V. fischeri*, candidate colonies were screened with KMB_334/KMB_335 and KMB_290/KMB_338 for ES114 *sypEFG* swap, and KMB_336/KMB_337 and KMB_290/KMB_339 for SR5 *sypEFG* swap. Positive candidates were selected and whole-genome sequenced by MiGS. Raw reads were assembled using SPAdes (v3.12.0) using the following command; spades.py – pe1-1 R1.fastq.gz –pe1-2 R2.fastq.gz -k 73,75,77 -careful -o. *syp* gene sequences from the assembled contigs were compared to our constructed reference sequence using SnapGene.

Construction of single *syp* gene swap strains. The exchange of *sypE*, *sypF*, and *sypG* from ES114 into SR5 was done using an allelic exchange approach modified from the laboratory's gene deletion protocol (https://doi.org/10.5281/zenodo.1470836). In brief, approximately 1.6 Kb of DNA upstream and downstream of each ORF was amplified from SR5 genomic DNA using primers listed in Table 3. The previously described pEVS79 vector backbone was amplified using RYI097 F/RYI097 R. Fragments were cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells. Plasmids were screened using RYI072 F/RYI072 R. Positive candidates were sent to Plasmidsaurus (Eugene, OR) for sequencing. The constructed plasmids were amplified using primers listed in Table 3. ORFs for *sypE, sypF,* and *sypG* were amplified from ES114 genomic DNA using primers listed in Table 3. Fragments were cloned into the amplified vector backbones by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells. Constructed plasmids were screened using RYI072 F/RYI072 R. Positive candidates were sequenced by Plasmidsaurus (Eugene, OR). After double recombination of these plasmids into *V. fischeri* SR5, candidate colonies were screened with KMB_279/KMB_335 for *sypE*, RYI199/KMB_325 and KMB_152/RYI198 for *sypF*, and KMB_153/DPA45 for *sypG.* Positive candidates were selected and whole-genome sequenced by MiGS. Raw reads were assembled using SPAdes

(v3.12.0) using the following command; spades.py – pe1-1 R1.fastq.gz –pe1-2 R2.fastq.gz -k 73,75,77 -careful -o. Using SnapGene, we identified the *sypC* sequence from the assembled contigs and copied approximately 7 Kb of sequence to compare to our constructed reference sequence.

To exchange *sypF* from SR5 into ES114 Δ*rscS* (MJM3010), pKMB026 was amplified using KMB_511/KMB_512. The *sypF* ORF from SR5 was amplified using KMB_513/KMB_514. The fragment was cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells. Candidate colonies were screened with RYI072 F/RYI072 R. Positive candidates were sequenced by Plasmidsaurus. After double recombination of these plasmids into *V. fischeri* ES114 Δ*rscS* (MJM3010), candidate colonies were screened with KMB_154/KMB_550. Positive candidates were selected and whole-genome sequenced by MiGS. Raw reads were assembled using SPAdes (v3.12.0) using the following command; spades.py – pe1-1 R1.fastq.gz –pe1-2 R2.fastq.gz -k 73,75,77 -careful -o. Using SnapGene, we identified the *sypC* sequence from the assembled contigs and copied approximately 7 Kb of sequence to compare to our constructed reference sequence.

Construction of barcoded deletion strains. To generate the ES114 Δ*rscS* Δ*binK*::bar Δ*sypF*::*sypF* [SR5] strain (MJM5024), the previously described pLostfoX plasmid (178, 201) was conjugated via triparental mating into MJM4888 (185) https://zenodo.org/records/1470836 (185). Candidates were selected on chloramphenicol and presence of the plasmid was confirmed using RYI072 F/RYI072 R. Genomic DNA from MJM3579 was extracted using the Qiagen DNeasy Blood and Tissue kit and transformed into the MJM4888/pLostfoX strain (200). Mutant candidates were selected on erythromycin and screened by PCR using primer pairs HB159/HB158, HB159/HB8, and HB160/HB161. Insertion of the *erm-bar* scar was confirmed via

Sanger sequencing using primers HB159, HB155, HB8, HB9, and HB158. The final barcode strain (MJM5024) was constructed via a triparental mating using donor MJM3478 (π3813/pKV496) (177) and helper strain MJM534 (CC118 λ pir/pEVS104) with MJM5023. Candidates were selected for using kanamycin and screened by PCR using the primer pairs listed above. The deletion scar was verified by Sanger sequencing using primers HB159, HB155, HB42, and HB158.

To generate the ES114 Δ*sypF::erm-bar* strain (MJM3954) deletion of *sypF* was performed following the barcode-tagged gene deletion protocol from Burgos et al. (200). In brief, the upstream homology arm was amplified using primers RYI194/RYI195 and the downstream homology arm was amplified using primers RYI196/RYI197. Homology arms were fused to either side of a third fragment containing an *erm* cassette using splicing by overhang extension PCR (SOE PCR). Mutagenic DNA was purified using the Qiagen PCR purification kit and transformed into ES114 via transformation using pLostfoX (MJM1538) (178, 201). Mutant candidates were selected using erythromycin and screened by PCR using primer pairs RYI193/RYI200, RYI193/HB8, and RYI198/RYI199. Insertion of the *erm-bar* scar was confirmed by Sanger sequencing using primers RYI193, RYI194, HB8, HB9, RYI197, RYI200, and the barcode sequence was recorded.

To generate the ES114 Δ*binK* Δ*rscS::bar* Δ*sypF*::*bar* strain (MJM4030), previously described pLostfoX-Kan plasmid (73) was conjugated via triparental mating into MJM4018. Candidates were selected on kanamycin and the resulting strain was saved as MJM4025. Genomic DNA from MJM3954 was extracted using the Qiagen DNeasy Blood and Tissue kit and transformed into MJM4025. Mutant candidates were selected on erythromycin and screened by PCR using primer pairs RYI193/RYI200, RYI193/HB8, and RYI198/199. Insertion of the *erm-bar* scar was confirmed via Sanger sequencing using primers RYI193, RYI194, HB8, HB9, RYI197 and

RYI200. The final barcode strain (MJM4030) was constructed via triparental mating with donor MJM3478 (π3813/pKV496) (177) and helper strain MJM534 (CC118 l pir/pEVS104) with MJM4026. Candidates were selected on kanamycin and screened by PCR using the primer pairs listed above. The deletion scar was verified via Sanger sequencing using primers RYI193, RYI194, HB42, RYI197 and RYI200.

To generate the ES114 Δ*rscS* Δ*sypF*::*erm-bar* strain (MJM5636), genomic DNA from MJM3954 was extracted using the Qiagen DNeasy Blood and Tissue kit and transformed into MJM3260. Mutant candidates were selected on erythromycin and screened by PCR using primer pairs RYI193/RYI200, RYI193/HB8, and RYI198/199. Insertion of the *erm-bar* scar was confirmed via linear amplicon sequencing by Plasmidsaurus.

Colony biofilm assays. Cultures were grown for approximately 18 h overnight and 8 µL was spotted onto TBS or TBS-calcium (10 mM $CaCl_2$). Spots were allowed to dry at room temperature and plates were grown at 20˚C for 72 h and imaged using a Leica M60 stereomicroscope with Leica Firecam software.

SypF HPt peptide antibody creation and purification. ProSci (Proway, CA) was given the conserved sequences of the SypF HPt domain between ES114 and SR5 and chose the peptide DASVMPELIRFYIIESKER as an epitope for polyclonal antibody production. This peptide was generated by ProSci and used to inoculate rabbits. Serum was assessed at various checkpoints to evaluate antibody production. The final bleed serum was purified by ProSci to a final antibody concentration of 30 mg/mL in PBS with 0.02% sodium azide. Purified antibody was aliquoted into small volumes and stored at -20˚C.

SDS page and Western blots. Overnight culture (1 mL) was pelleted, washed, and lysed in 1% SDS. The volume of SDS used to lyse the sample was adjusted based on the optical density at 600 nm (OD_{600}) of the overnight culture to standardize the total protein concentration in the samples. The solution was then pelleted to remove cell debris, and the supernatant was mixed 1:1 with 2X Laemmli sample buffer (Bio-Rad) and β-mercaptoethanol. Samples were heated at 80°C for 10 min and loaded onto a 10% Bio-Rad Mini-Protean TGX Stain-Free precast gel. The gel was then transferred to a polyvinylidene difluoride (PVDF) membrane and blocked overnight in 5% nonfat milk. The purified anti-SypF-HPt-peptide antibody was used as the primary antibody in a 1:1000 dilution in 0.5% nonfat milk in 1X Tris-buffered saline (TBS)-Tween 20. The secondary antibody was a 1:5,000 dilution of the Pierce goat anti-rabbit IgG (H+L)-horse-radish peroxidase (HRP) conjugate 0.4 mg/mL (lot UK293475). Washes were done in 1X TBS-Tween 20. Blots were developed using the Thermo Scientific SuperSignal West Dura extended duration substrate and analyzed using a Licor Odyssey Fc machine. Quantification of the blot was performed using densitometry with ImageJ. In brief, the mean gray value of each band was quantified and the mean gray value of the background was subtracted. The density of the bands were normalized to the total protein of their respective sample in the stain-free gel and then normalized to the density of the ES114 band for comparison.

***syp* transcriptional reporter strain construction and *in vitro* assay.** Reporter plasmid pM1422 was introduced into strains MJM1100, MJM4018, and MJM5024, MJM6088 and MJM6089 via triparental mating by mixing the pEVS104-containing helper, MJM1422 donor, and *V. fischeri* recipient. Candidates were selected on Kanamycin media, and the presence of the plasmid was confirmed by reading mCherry and GFP fluorescence of 200 µL overnight culture using a BioTek Synergy Neo2 plate reader. To assess *syp* transcription levels, strains were grown overnight for 18 h then 8µL was spotted onto TBS plates in triplicate. Spots were allowed to dry and room temperature then plates were grown at 20˚C for 72 h. Images of each

spot were taken at 24, 48, and 72 h using a Leica M60 stereomicroscope with Leica Firecam software. Fluorescent images were taken at the same time points using a Zeiss Axio Zoom.V16 large-field fluorescent stereo microscope. mCherry and GFP fluorescence readings were measured using the Zen Blue software and the polygon tool was used to select the spots and measure fluorescence intensity of the spot and a nearby background region. To calculate gene promoter activity, the background for each channel was subtracted from the region of interest. To normalize GFP to plasmid copy number, the GFP reading (reports promoter activity) was divided by the mCherry reading (constitutively expressed). This resulted in the reported mean GFP/mCherry reading for each individual spot.

Construction of *sypF* chimera strains. All plasmids were constructed using pKMB031 as the vector backbone and ES114 *sypF* fragments were amplified from gDNA. The following primer pairs were used to amplify the vector backbones and ES114 *sypF* fragments for each of the chimera constructions (vector backbone:*sypF* fragment). pKMB043 KMB_512/KMB_601:KMB_602/KMB_603; pKMB045 KMB_511/KMB_606: KMB_607/KMB_608; pKMB046 KMB_511/KMB609: KMB_607/KMB_610; pKMB047 KMB_511/KMB_611: KMB_607/KMB_612; pKMB048 KMB_511/KMB_613: KMB_607/KMB_612; pKMB049 KMB_511/KMB_613: KMB_607/KMB_612. All plasmids were constructed using Gibson assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells. Plasmid candidates were screened by PCR using primers KMB_615/KMB_616. Positive candidates were sequenced by Plasmidsaurus. Plasmids pKMB043 and pKMB045-pKMB048 underwent double recombination into *V. fischeri* ES114 Δ*rscS* (MJM3010) and ES114 Δ*binK* Δ*rscS::bar* (MJM4018). Candidate colonies for pKMB043 recombination were screened using primer pair KMB_615/KMB_155. Candidate colonies for pKMB045, pKMB046, pKMB047, and pKMB048 were screened using primer pair KMB_616/KMB_154. Plasmid pKMB049 underwent double recombination into *V.*

fischeri ES114 Δ*rscS ΔsypF::bar* (MJM5636) and ES114 *ΔbinK ΔrscS::bar ΔsypF::bar* (MJM4030). Candidate colonies were screened using primer pair KMB_616/KMB_614. Positive candidates for all strains were confirmed via whole-genome sequencing by Plasmidsaurus by comparing *syp* gene sequences from assembled contigs to our constructed reference sequence using SnapGene.

Construction of *sypF* REC domain swap strain. The previously generated vector pKMB49 was purified and used as a template for this allelic exchange vector. The vector backbone was amplified using KMB_661/KMB_662. The REC domain (AA 449-620) of SypF from SR5 was amplified from the purified pKMB031 vector using KMB_664/KMB_665. The REC domain fragment was cloned into the pKMB049 vector using Gibson assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) then transformed into DH5α λpir chemically competent cells. The resulting plasmids were sequenced by Plasmidsaurus. After double recombination of these plasmids into *V. fischeri*, candidate colonies were screened with primer pair KMB_325/KMB_550. Positive candidates were sent to Plasmidsaurus for confirmation via whole genome sequencing. *syp* gene sequences from the assembled contigs were compared to our constructed reference sequence using SnapGene.

Construction of *sypF* natural isolate swap strains. The previously generated vector pKMB031 was purified and used as a template for all allelic exchange vectors. The vector backbone was amplified using primer pair KMB_511/KMB_512. Genomic DNA was isolated from EM8.7 and mjapo6.1 using the Qiagen DNeasy Blood and Tissue Kit. The *sypF* ORF was amplified from EM8.7 using primer pair KMB_656/KMB_657 and from mjapo6.1 using KMB_658/KMB_659. *sypF* fragments were cloned into the pKMB031 vector using Gibson assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) then transformed into DH5α λpir chemically competent cells. The resulting plasmids were sequenced by

Plasmidsaurus. After double recombination of these plasmids into *V. fischeri*, candidate colonies were screened with primer pair KMB_325/KMB_550. Positive candidates were sent to Plasmidsaurus for confirmation via whole genome sequencing. *syp* gene sequences from the assembled contigs were compared to our constructed reference sequence using SnapGene.

Construction of *sypF* point mutant strains. The previously generated vector pKMB031 was purified and used as a template for all allelic exchange vectors. The point mutants were constructed using KMB_637/KMB_638 (T566S) or KMB_650/KMB_651 (E620Q) with the Q5 Site Directed Mutagenesis kit (NEB).

Vector pKMB049 was purified and used as a template for all allelic exchange vectors. The point mutants were made using KMB_639/KMB_640 (S566T), KMB_647/KMB_648 (P575L), or KMB_653/KMB_654 (Q620E) with the Q5 Site Directed Mutagenesis kit (NEB). Resulting plasmids were transformed into DH5α λpir cells and candidates were sent to Plasmidsaurus for sequencing. After double recombination of these plasmids into *V. fischeri*, candidate colonies were screened with primer pair KMB_325/KMB_550. Positive candidates were sent to Plasmidsaurus for confirmation via whole genome sequencing. *syp* gene sequences from the assembled contigs were compared to our constructed reference sequence using SnapGene.

Squid single-strain colonizations. *V. fischeri* strains were grown overnight with aeration at 25°C in LBS. Overnight cultures were diluted 1:80 in LBS and grown for 1.5 h at 25˚C. The OD600 was used to normalize the amount of each strain used to inoculate hatchling *E. scolopes* with approximately $3x10^3$ CFU/ml (ranging from 4.4×10^2 to 2.5×10^4 CFU/ml, as specified in the Figure legends) for each strain in a total volume of 40 ml of FSIO (filter-sterilized Instant Ocean) for 3 h. Experiments for *sypF* chimera strains (Fig. 3) were inoculated for 4 h due to increased age of egg laying females. Squid were then transferred to 100 ml of FSIO to stop the

inoculation and then transferred to individual vials with 4 ml of bacterium-free filter-sterilized Instant Ocean (FSIO) until approximately 48 h post inoculation (hpi) with a water change that occurred at 24 hpi. At 48 hpi, squid were transferred to 1.5-ml microcentrifuge tubes with approximately 700 µL of FSIO, and each animal's luminescence was measured using the Promega GloMax 20/20 luminometer. The squid were euthanized and surface sterilized by storage at -80°C according to standard practices (180). For determination of CFU per light organ, hatchlings were thawed and homogenized, and 50 µL of homogenate dilutions was plated onto LBS plates. Bacterial colonies from each plate were counted and recorded. Mock-treated, uncolonized hatchlings ("apo-symbiotic") were used to determine the limit of detection in the assay.

Alignment of SypF REC domain sequences. SypF amino acid sequences used to generate the phylogenetic tree of the following natural isolates were aligned using MAFFT (202) in the Align Multiple Protein Sequences tool in SnapGene; SR5, SA1, EM8.7, mjapo6.1, and, ES114 (Fig. S3). Amino acids 499-620 were displayed as the sequences for the REC domain. Presence of *rscS* was determined by performing a blastp (203) search against the genome or had previously been identified (131, 132).

Data analysis. GraphPad Prism was used to construct graphs and perform statistical analyses.

ACKNOWLEDGEMENTS

We would like to thank Denise Ludvik and Ruth Isenberg for their assistance in the generation of strains used in this paper. This work was funded by NIH grants R35GM148385 (M.J.M.). Support for trainees was provided by T32GM008349 (K.M.B.) and an NSF Graduate Research Fellowship (K.M.B).

TABLES AND SUPPLEMENTAL FIGURES

Table 3.1. Bacterial Strains

V. fischeri		
Strain	**Genotype**	**Source or Reference**
MJM1125 = SR5	natural isolate, *Sepiola robusta* squid light organ	(128, 133)
MJM1100 = ES114	natural isolate, *Euprymna scolopes* squid light organ	(101, 204)
MJM1538	MJM1100/pLostfoX	(178, 201)
MJM3010	MJM1100 Δ*rscS*	(132)
MJM2251	MJM1100 Δ*binK*	(73)
MJM3260	MJM3010/pLostfoX	This study
MJM4018	MJM2251 Δ*rscS*::bar	(123)
MJM3954	MJM1100 Δ*sypF::erm-bar*	This study
MJM4030	MJM1100 Δ*binK* Δ*rscS*::bar Δ*sypF*::bar	This study
MJM5636	MJM3010 Δ*sypF::erm-bar*	This study
MJM4025	MJM4018/pLostfoX-Kan	This study
MJM4604	MJM1100 Δ(*sypEFG*)::*sypEFG*[MJM1125]	This study
MJM4606	MJM1125 Δ(syp*EFG*)::syp*EFG*[MJM1100]	This study
MJM4608	MJM3010 Δ(*sypEFG*)::*sypEFG*[MJM1125]	This study
MJM4705	MJM1125 Δ*sypE*:: *sypE* [MJM1100]	This study
MJM4707	MJM1125 Δ*sypF*:: *sypF* [MJM1100]	This study
MJM4709	MJM1125 Δ*sypG*::*sypG* [MJM1100]	This study
MJM4888	MJM3010 Δ*sypF*::*sypF* [MJM1125]	This study
MJM5024	MJM3010 Δ*rscs* Δ*sypF*::*sypF* [MJM1125] Δ*binK*::bar	This study
MJM5965	MJM3010 Δ*sypF*:: *sypF* [AA N-225 MJM1100, 226-C MJM1125]	This study
MJM5966	MJM4018 Δ*sypF*:: *sypF* [AA N-225 MJM1100, 226-C MJM1125]	This study
MJM5968	MJM3010 Δ*sypF*:: *sypF* [AA	This study

	N-665 MJM1125, 666-C MJM1100]	
MJM5969	MJM4018 Δ*sypF*:: *sypF* [AA N-665 MJM1125, 666-C MJM1100]	This study
MJM5970	MJM3010 Δ*sypF*:: *sypF* [AA N-498 MJM1125, 499-C MJM1100]	This study
MJM5971	MJM4018 Δ*sypF*:: *sypF* [AA N-498 MJM1125, 499-C MJM1100]	This study
MJM5972	MJM3010 Δ*sypF*:: *sypF* [AA N-351 MJM1125, 352-C MJM1100]	This study
MJM5973	MJM4018 Δ*sypF*:: *sypF* [AA N-351 MJM1125, 352-C MJM1100]	This study
MJM5974	MJM5636 *sypF* [MJM1100]	This study
MJM5975	MJM4030 *sypF* [MJM1100]	This study
MJM1422	MJM1100/pM1422	(123)
MJM5997	MJM4018/pM1422	This study
MJM5998	MJM5024/pM1422	This study
MJM6088	MJM4030 *sypF* [MJM1125]	This study
MJM6089	MJM4030 *sypF* [AA N-498 MJM1100, 499-620 MJM1125, 621-C MJM1100]	This study
MJM6090	MJM4030 *sypF* [EM8.7]	This study
MJM6091	MJM4030 *sypF* [mjapo6.1]	This study
MJM6092	MJM4030 *sypF* [MJM1125 T566S]	This study
MJM6093	MJM4030 *sypF* [MJM1125 E620Q]	This study
MJM6094	MJM4030 *sypF* [MJM1100 Q620E]	This study
MJM6141	MJM4030 *sypF* [MJM1100 P575L]	This study
MJM6142	MJM4030 *sypF* [MJM1100 S566T]	This study
MJM6095	MJM6088 / pM1422	This study
MJM6096	MJM6089 / pM1422	This study
MJM6143	MJM5636 *sypF* [AA N-498 MJM1100, 499-620 MJM1125, 621-C MJM1100]	This study
E. coli		
Strain	**Genotype**	**Source or Reference**
MJM534	CC118 λ pir/pEVS104	(178)
MJM537	DH5α λ pir	Laboratory stock
MJM637	S17-1 λ pir/pUX-BF13	(205)
MJM1422	DH5α λ pir/pM1422	(73)

MJM1538	DH5α λ pir/pLostfoX	(178, 201)
MJM3478	π3813/pKV496	(177)
MJM4330	DH5α λ pir/pKMB010	This study
MJM4331	DH5α λ pir/pKMB011	This study
MJM4411	DH5α λ pir/pKMB012	This study
MJM4486	DH5α λ pir/pKMB015	This study
MJM4487	DH5α λ pir/pKMB016	This study
MJM4646	DH5α λ pir/pKMB017	This study
MJM4647	DH5α λ pir/pKMB018	This study
MJM4648	DH5α λ pir/pKMB019	This study
MJM4649	DH5α λ pir/pKMB020	This study
MJM4650	DH5α λ pir/pKMB021	This study
MJM4651	DH5α λ pir/pKMB022	This study
MJM4747	DH5α λ pir/pKMB026	This study
MJM4783	DH5α λ pir/pKMB031	This study
MJM5660	DH5α λ pir/pKMB043	This study
MJM5661	DH5α λ pir/pKMB044	This study
MJM5662	DH5α λ pir/pKMB045	This study
MJM5663	DH5α λ pir/pKMB046	This study
MJM5664	DH5α λ pir/pKMB047	This study
MJM5665	DH5α λ pir/pKMB048	This study
MJM5666	DH5α λ pir/pKMB049	This study
MJM6007	DH5α λ pir/pKMB050	This study
MJM6008	DH5α λ pir/pKMB051	This study
MJM6009	DH5α λ pir/pKMB052	This study
MJM6010	DH5α λ pir/pKMB053	This study
MJM6011	DH5α λ pir/pKMB054	This study
MJM6012	DH5α λ pir/pKMB055	This study
MJM6013	DH5α λ pir/pKMB056	This study
MJM6014	DH5α λ pir/pKMB057	This study
MJM6015	DH5α λ pir/pKMB058	This study

Table 3.2: Plasmids

Plasmid	**Description**	**Source or Reference**
pLostfoX	Arabinose-inducible TfoX for transformation (Cam^r)	(178)
pKV496	pEVS79 containing the FLP recombinase (Kan^r)	(177)
pM1422	pTM267 *sypA'-gfp*+ (Cam^r)	(73)
pKMB010	pEVS79 carrying approximately 1.6 Kb US and DS of *sypEFG* ORFs from MJM1100 (Cam^r)	This study
pKMB011	pEVS79 carrying approximately 1.6 Kb US and	This study

	DS of *sypEFG* ORFs from MJM1125 (Cam^r)	
pKMB015	pKMB011 carrying *sypEFG* ORFs from MJM1100 (Cam^r)	This study
pKMB016	pKMB010 carrying *sypEFG* ORF sfrom MJM1125 (Cam^r)	This study
pKMB017	pEVS79 carrying approximately 1.6 Kb US and DS of *sypE* ORF from MJM1125 (Cam^r)	This study
pKMB018	pEVS79 carrying approximately 1.6 Kb US and DS of ORF *sypF* from MJM1125 (Cam^r)	This study
pKMB019	pEVS79 carrying approximately 1.6 Kb US and DS of *sypG* ORF from MJM1125 (Cam^r)	This study
pKMB020	pKMB017 carrying *sypE* ORF from MJM1100 (Cam^r)	This study
pKMB021	pKMB018 carrying *sypF* ORF from MJM1100 (Cam^r)	This study
pKMB022	pKMB019 carrying *sypG* ORF from MJM1100 (Cam^r)	This study
pKMB026	pEVS79 carrying approximately 1.6 Kb US and DS of *sypF* ORF from MJM1100 (Cam^r)	This study
pKMB031	pKMB026 carrying *sypF* ORF from MJM1125 (Cam^r)	This study
pKMB043	pKMB031 carrying *sypF* [AA N-225 MJM1100, 226-C MJM1125] (Cam^r)	This study
pKMB045	pKMB031 carrying *sypF* [AA N-225 MJM1100, 226-C MJM1125] (Cam^r)	This study
pKMB046	pKMB031 carrying *sypF* [AA N-498 MJM1125, 499-C MJM1100] (Cam^r)	This study
pKMB047	pKMB031 carrying *sypF* [AA N-351 MJM1125, 352-C MJM1100] (Cam^r)	This study
pKMB049	pKMB031 carrying *sypF* ORF from MJM1100 (Cam^r)	This study
pKMB050	pKMB031 carrying *sypF* with T566S mutation (Cam^r)	This study
pKMB051	pKMB049 carrying *sypF* with Q620E mutation (Cam^r)	This study
pKMB052	pKMB031 carrying *sypF* with E620Q mutation (Cam^r)	This study

pKMB053	pKMB049 carrying *sypF* with P575 mutation (Cam^r)	This study
pKMB054	pKMB031 carrying *sypF* from EM8.7 (Cam^r	This study
pKMB055	pKMB031 carrying *sypF* from mjapo6.1 (Cam^r)	This study
pKMB056	pKMB049 carrying *sypF* [AA N-498 MJM1100, 499-620 MJM1125, 621-C MJM1100] (Cam^r)	This study
pKMB058	pKMB049 carrying *sypF* with S566T mutation (Cam^r)	This study

Table 3.3. DNA oligonucleotides for PCR amplification and sequencing.

Linker sequences are notated in lowercase letters.

Primer Name	**Sequence (5'-3')**
KMB_262	tatcgaattcctgcagcccgGAATGGTCAGATGAAATGTCATTTTTAGATTTACTGGC
KMB_263	caagcagtaaccatatttcaCGGTTTCTTTGGACTATTGCGATAAGTCTAG
KMB_264	gcaatagtccaaagaaaccgTGAAATATGGTTACTGCTTGATAGCCG
KMB_265	tctagaactagtggatccccCGAGTTTAGGACTTAAATGAATACCTAAGCTCAG
KMB_266	CGGTTTCTTTGGACTATTGCGATAAGTCTAG
KMB_267	TGAAATATGGTTACTGCTTGATAGCCG
KMB_268	gcaatcgtccaaaaaaatcgATTTTTATCAACAGTGGTGTAATCATGAATTCTACTTTAC
KMB_269	caagcagtaaccatatttcaTCATTCCGATTCTTCATAGGCTTCCC
KMB_270	tatcgaattcctgcagcccgGAGTGGTCAGATGAAATGTCATTTCTTGATTTATTAGC
KMB_271	caagcagtaaccatatttcaCGATTTTTTTGGACGATTGCTATAAGTTTAGTTG
KMB_272	gcaatcgtccaaaaaaatcgTGAAATATGGTTACTGCTTGATAGCCGTG
KMB_273	tctagaactagtggatccccGACGAGTTTAGGACTTAAATGAATACCTAAGCTC
KMB_274	TGAAATATGGTTACTGCTTGATAGCCGTG
KMB_275	CGATTTTTTTGGACGATTGCTATAAGTTTAGTTG
KMB_276	gcaatagtccaaagaaaccgATTTTTATCAACAGTGGTATGATCATGAATTCTATTTTGCTTTTTTC

KMB_277	caagcagtaaccatatttcaTCATTCCGATTCTTCATAGGCTTCCC
KMB_278	CTATTAGCAGAAGCAGGAGGGCC
KMB_279	GCAGGCTATCGAACACTATTGGTTG
KMB_280	CGCTATGAAGGCTTACCTATGGTTG
KMB_285	GCAGACCTCACACAACCATCG
KMB_286	CATTGGTATCTCTATTATTCAGTGCATCATGTTACTG
KMB_287	GAAGCTGAAGCGGCCAATAAAAGTAAAAG
KMB_288	GCCTGTGTATTCTCTGATGCTAATATTCTTGTTG
KMB_289	GATGCGAGTGTTATGCCTGAATTAATACGC
KMB_290	GCGCGAACGAGGAACTGATATC
KMB_297	taaacttatagcaatcgtccaaaaaaatcgATTTTTATCAACAGTGGTGTAATCATGAATTCTACTTTAC
KMB_298	tcacggctatcaagcagtaaccatatttcaTCATTCCGATTCTTCATAGGCTTCCC
KMB_299	tagacttatcgcaatagtccaaagaaaccgATTTTTATCAACAGTGGTATGATCATGAATTCTATTTTGCTTTTTTC
KMB_300	ttacggctatcaagcagtaaccatatttcaTCATTCCGATTCTTCATAGGCTTCCC
KMB_334	CCACCATTCGTGTACTAGGGGC
KMB_335	CCCTCAATTTCGATGATGTTAATTCCCCT
KMB_336	CCACCATTCGTGTACTAGGAGC
KMB_337	GGAAGTGTTCCTTCTGTATCAATGATATCAATTCTTTC
KMB_338	GAATAGTTTGTTGTGCATTTTTTTGAATGGTTTGGA
KMB_339	CTTGAATGGTTTGTTGTGCATTTTTCTGAATATTTTTAG
KMB_348	aaatgtcatagttagctccCGATTTTTTTGGACGATTGCTATAAGTTTAGTTG
KMB_349	gcaatcgtccaaaaaaatcgGGAGCTAACTATGACATTTCGATTAAAAACGATCATTG
KMB_350	tctagaactagtggatccccCGTCATACCATCCATTTCAGGCATGG
KMB_351	GGAGCTAACTATGACATTTCGATTAAAAACGATCATTG
KMB_352	taaacttatagcaatcgtccaaaaaaatcgATTTTTATCAACAGTGGTGTAATCATGAATTCTACTTTAC
KMB_353	gtttttaatcgaaatgtcatagttagctccTTTGCAATGTTTGCTTTTTATTGATTCTCAATTAACAGC
KMB_354	tatcgaattcctgcagcccgGAGAACCACAACGCAAATATTATTGGTCATG
KMB_355	taacatcctgttttacctagTTGCAATATTTGCTCTTTTATTGGTTCTCAATTAACAGC

KMB_356	taaaagagcaaatattgcaaCTAGGTAAAACAGGATGTTACTTATGCTACAAAAAGTATTATTAG
KMB_357	tctagaactagtggatccccGCTGTAATACGTGTGATTCGATACCTCC
KMB_358	CTAGGTAAAACAGGATGTTACTTATGCTACAAAAAGTATTATTAG
KMB_359	TTGCAATATTTGCTCTTTTATTGGTTCTCAATTAACAGC
KMB_360	accaataaaagagcaaatattgcaaGGAGCTAACTATGACATTTCGACTTAAAACG
KMB_361	gtagcataagtaacatcctgttttacctagCGAATTTTGTTTTATTCTTTTATTTTGAGAAACCTTGTTTATTTC
KMB_362	tatcgaattcctgcagcccgGGTGATGTAGCACAAGCGTTTAATTCC
KMB_363	acgtgtgattcgatacctccCGAATTTTGTTTATTCTTTTATTTTGAGAAACCTTGTTTATTTC
KMB_364	aaaagaataaacaaaattcgGGAGGTATCGAATCACACGTATTACAGC
KMB_365	GGAGGTATCGAATCACACGTATTACAGC
KMB_366	CGAATTTTGTTTATTCTTTTATTTTGAGAAACCTTGTTTATTTC
KMB_367	ttctcaaaataaaagaataaacaaaattcgCTAGGTAAAACAGGATGTTACTTATGCTACAGAAAG
KMB_368	aagctgtaatacgtgtgattcgatacctccAAAATTACGGCTATCAAGCAGTAACCATATTTC
KMB_369	tccttgcaatatttgctcttGATCATACCACTGTTGATAAAAATCGATTTTTTTGG
KMB_370	ttatcaacagtggtatgatcAAGAGCAAATATTGCAAGGAGCTAACTATG
KMB_371	AAGAGCAAATATTGCAAGGAGCTAACTATG
KMB_372	GATCATACCACTGTTGATAAAAATCGATTTTTTTGG
KMB_373	aaatcgatttttatcaacagtggtatgatcATGAATTCTACTTTACTTTTTTCAGATACAAAGCCCAC
KMB_374	catagttagctccttgcaatatttgctcttTTATTGATTCTCAATTAACAGCAATAATGCATCGTCATTG
KMB_375	agcgaattttgtttattcttAGTTAGCTCCTTGCAATATTTGCTCTTTTATTG
KMB_376	aatattgcaaggagctaactAAGAATAAACAAAATTCGCTAGGTAAAACAGGATGTTAC
KMB_377	AAGAATAAACAAAATTCGCTAGGTAAAACAGGATGTTAC
KMB_378	AGTTAGCTCCTTGCAATATTTGCTCTTTTATTG
KMB_379	taaaagagcaaatattgcaaggagctaactATGACATTTCGACTTAAAACGATCATTGGTATC
KMB_380	tgttttacctagcgaattttgtttattcttTTATTTTGAGAAA

	CCTTGTTTATTTCTTTTTTCAAGTGCG
KMB_381	caagcagtaaccatatttcaAAGTAACATCCTGTTTTACCTAGCGAATTTTGTTTATTC
KMB_382	ggtaaaacaggatgttacttTGAAATATGGTTACTGCTTGATAGCCGTG
KMB_383	AAGTAACATCCTGTTTTACCTAGCGAATTTTGTTTATTC
KMB_384	aaattcgctaggtaaaacaggatgttacttATGCTACAGAAAGTATTATTAGTCGAAGACTCC
KMB_385	agcgaattttgtttattcttGATCATACCACTGTTGATAAAAATCGATTTTTTTGG
KMB_386	ttatcaacagtggtatgatcAAGAATAAACAAAATTCGCTAGGTAAAACAGGATGTTAC
KMB_387	aatattgcaaggagctaactTGAAATATGGTTACTGCTTGATAGCCGTG
KMB_388	TTATTGATTCTCAATTAACAGCAATAATGCATCGTC
KMB_389	ATGCTACAGAAAGTATTATTAGTCGAAGACTCCAC
KMB_390	gcattattgctgttaattgagaatcaataaAAGAGCAAATATTGCAAGGAGCTAACTATG
KMB_391	gtcttcgactaataatactttctgtagcatAAGTAACATCCTGTTTTACCTAGCGAATTTTGTTTATTC
KMB_392	caagcagtaaccatatttcaAGTTAGCTCCTTGCAATATTTGCTCTTTTATTG
KMB_507	tatcgaattcctgcagcccgCCAAAAAAAATATCAACTAGACTTATCGCAATAGTCC
KMB_508	gcgaattttgttttattcttAGTTAGCTCCTTTGCAATGTTTGCTTTTTATTG
KMB_509	acattgcaaaggagctaactAAGAATAAAACAAAATTCGCTAGGTAAAACAGGATG
KMB_510	tctagaactagtggatccccGCAAGCTGTAATACGTGTGATTCGATACC
KMB_511	AAGAATAAAACAAAATTCGCTAGGTAAAACAGGATG
KMB_512	AGTTAGCTCCTTTGCAATGTTTGCTTTTTATTG
KMB_513	acattgcaaaggagctaactATGACATTTCGATTAAAAACGATCATTGGTATCTCTATTATTC
KMB_514	gcgaattttgttttattcttTTATTTTGAGAAACCTTGTTTATTTCTTTGTTCAAGTGC
KMB_601	GACCTACAACAAGCCAGAAATGAAGC
KMB_602	acattgcaaaggagctaactATGACATTTCGACTTAAAACGATCATTGGTATCTC
KMB_603	tttctggcttgttgtaggtcTTGGTAGTTTTGTTCTAGCTTTAAGGACATGTTATTAAATG
KMB_604	ACTTTCTACCTTGAAAATCATACATTTAATATACATAAAAGTTTAAATAGTG
KMB_605	tgattttcaaggtagaaagtGTTAGCTTCCATACGTGAGAAGTCGA

KMB_606	GGTGTCAGCTATTATTTGCTGAATTATCTG TTCATC
KMB_607	gcgaattttgttttattcttTTATTTTGAGAAACCTTGT TTATTTCTTTTTTCAAGTGCG
KMB_608	agcaaataatagctgacaccGATGCGAGTGTTAT GCCTGAATTAATCC
KMB_609	ATCAGAGAATACACAGGCTTTTTGTTGTT CAATATC
KMB_610	aagcctgtgtattctctgatGCTAATATTCTTGTTGT TGAGGATAATCACGC
KMB_611	CTGAACGTAGGTAGGAACATTATCAGTAA TAACTAATGAC
KMB_612	atgttcctacctacgttcagGGCGATGCTCACAGAT TACGTC
KMB_613	GGCCGCTTCAGCTTCATTTCTG
KMB_614	gaaatgaagctgaagcggccAATAAAAGTAAAAG TCGATTTTTAGCTTCCATGAGTCAC
KMB_615	GGTTATGTGCGAGGCCTAATGC
KMB_616	GGTTGCAATGATAACGGCGGTAGG
KMB_617	GTGCTCCTGGGCCTTGG
KMB_619	CAAAAGAGCGAGTAGAAACCATCTCAAAT
KMB_620	CTCTATCTGATACACACCAAAACATGCCT
KMB_621	GATGGACAGGTTCAAATATTAGTTAATGC CGAA
KMB_622	GGTTTACTCGCGATTTTAAAAGACACAAC T
KMB_637	AAAAATTCGCtCTCTATCTGATACACAC
KMB_638	TTCGTTGCCGTCATACCA
KMB_639	AAAAATCCGCaCTCTATCTGATACACAC
KMB_640	TTCGTTGCCGTCATGCCA
KMB_641	CGGCAACGAAAAAAATTCGCt
KMB_642	GATGGTTTCTACTCGCTCTTTTGATTC
KMB_643	GCAACGAAAAAAATCCGCa
KMB_644	CAAAACATGCcTATCATTGCAC
KMB_645	GTGTGTATCAGATAGAGTG
KMB_646	CACTCTATCTGATACACACCAAAACATGCc
KMB_647	CAAAACATGCtTATCATTGCAC
KMB_648	GTGTGTATCAGATAGAGAG
KMB_649	CTCTCTATCTGATACACACCAAAACATGCt
KMB_650	CCTGTTTAAAcAGGGTAAACAAG
KMB_651	TACTTAGCAAGGCATTCTAG
KMB_652	CTACTAGAATGCCTTGCTAAGTACCTGTT TAAAc
KMB_653	CCTGTTTAAAgAGGGTAAACAAG
KMB_654	TACTTAGCAAGACATTCTAG
KMB_655	GATTTACTAGAATGTCTTGCTAAGTACCT GTTTAAAg
KMB_656	acattgcaaaggagctaactATGACATTTCGATTA AAAACGATCATTGGTATCTC

KMB_657	gcgaattttgttttattcttTTATTTTGAGAAACCTTGTTTATTTCTTTGTTCAAGTG
KMB_658	acattgcaaaggagctaactATGACATTTCGATTAAAAACGATCATTGGTATCTCTATTATTCAG
KMB_659	gcgaattttgttttattcttTTATTTTGAGAAACCTTGTTTATTTCTTTGTTCAAGTGC
KMB_660	GACAGTTCTAGTTGTTTGCTTTGGGC
KMB_661	ATCAGAGAATACACAGTTTTTTTGTTGTTCAATATCC
KMB_662	GGTAAACAAGACCCTTCATTTTGCCAAG
KMB_664	aaaactgtgtattctctgatATTCTTGTTGTTGAGGATAATCACGCC
KMB_665	aatgaagggtcttgtttaccCTCTTTAAACAGGTACTTAGCAAGGCATTC
RYI072 F	GTAAAACGACGGCCAGT
RYI072 R	AGCGGATAACAATTTCACACAGG
RYI097 F	GGGATCCACTAGTTCTAGAGCGG
RYI097 R	CGGGCTGCAGGAATTCGATATCA
DPA45	ATCGAAAAACCAATCCAAGCA
HB158	gtgtatctgttgccaccgtttctcg
HB159	ctggactattcgcttgaatttggacgtg
HB8	ACAAAATTTTAAGATACTGCACTATCAACACACTCTTAAG
HB155	accatggagccaacagcaagac
HB160	gcatttggccagaacccaatacattg
HB161	gctcaaaatgacagtcagagtatcgtagg
HB42	ACGAGACGAGCTTCTTATATATGCTTCGCCAG
RYI193	cggtgagagtattgcgccaat
RYI194	aacaccgttgatttaattatctcagatattcaaatgc
RYI195	gacttgacctggatgtctctacccacaagatcgAAACAAGGTTTCTCAAAATAAAAGAATAAAACAAAATTCG
RYI196	ctggcgaagcatatataagaagctcgtctcgtCATAGTTAGCTCCTTTGCAATGTTTGC
RYI197	gtggtggcatatgaataggcacaa
RYI198	ggactcgtcaattttgcataatgcatc
RYI199	ccgatagatgatatatcaacttgtgctcct
RYI200	taagagaacggcagcattaagaacattac

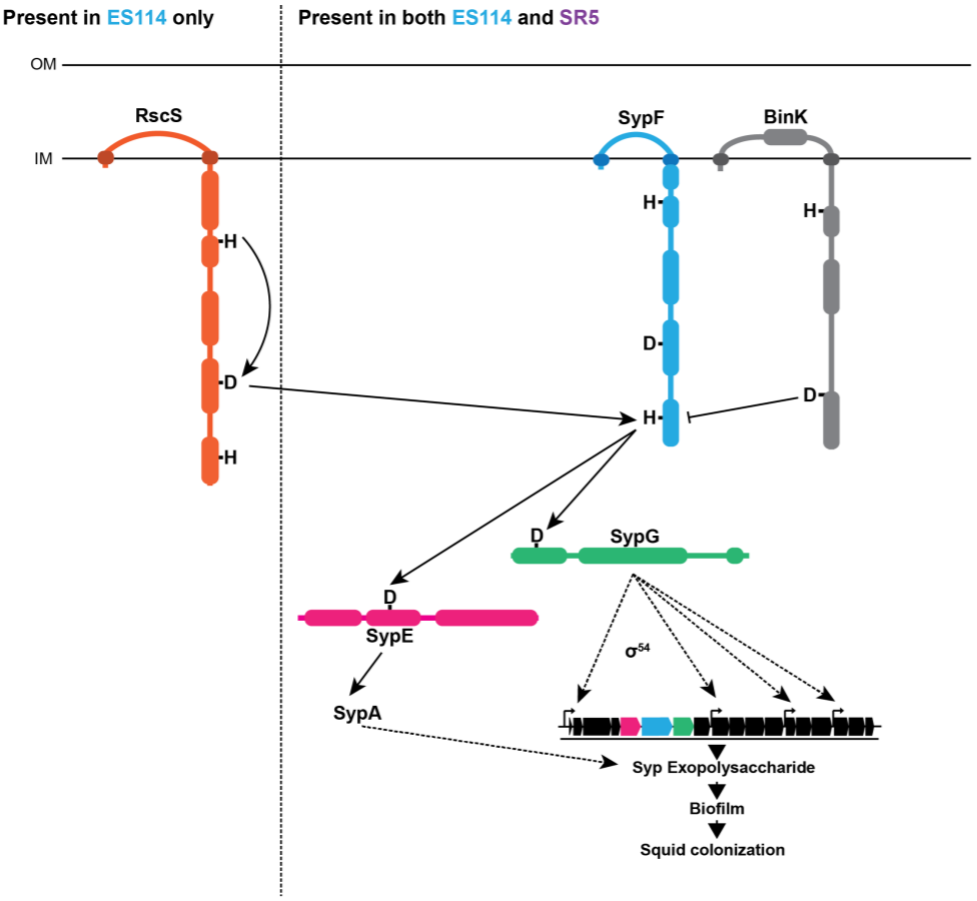


Figure 3.S1. Similarities and differences in the Syp regulation pathway between *V. fischeri* strains ES114 and SR5.
Model of Syp phosphorelay between known *syp* regulators. Solid arrows indicate phosphoryl group transfer and dashed arrows indicate activation through σ^{54} for SypG and unknown mechanisms for SypE.

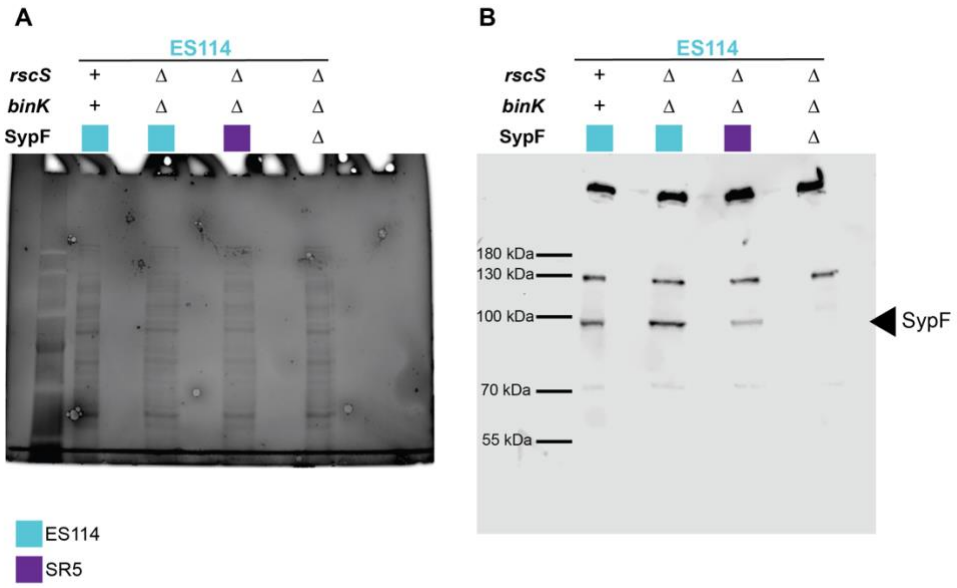


Figure 3.S2. Western blot of whole-cell lysates assessed with SypF HPt domain peptide antibody.

A) Image of stain-free polyacrylamide gel before transfer to PVDF membrane taken on a BioRad GelDoc Go gel imaging system. **B)** Image of full Western blot using SypF HPt domain peptide antibody imaged on a Licore imager. Arrow indicates full-length SypF, which is predicted to be approximately 85 kD.

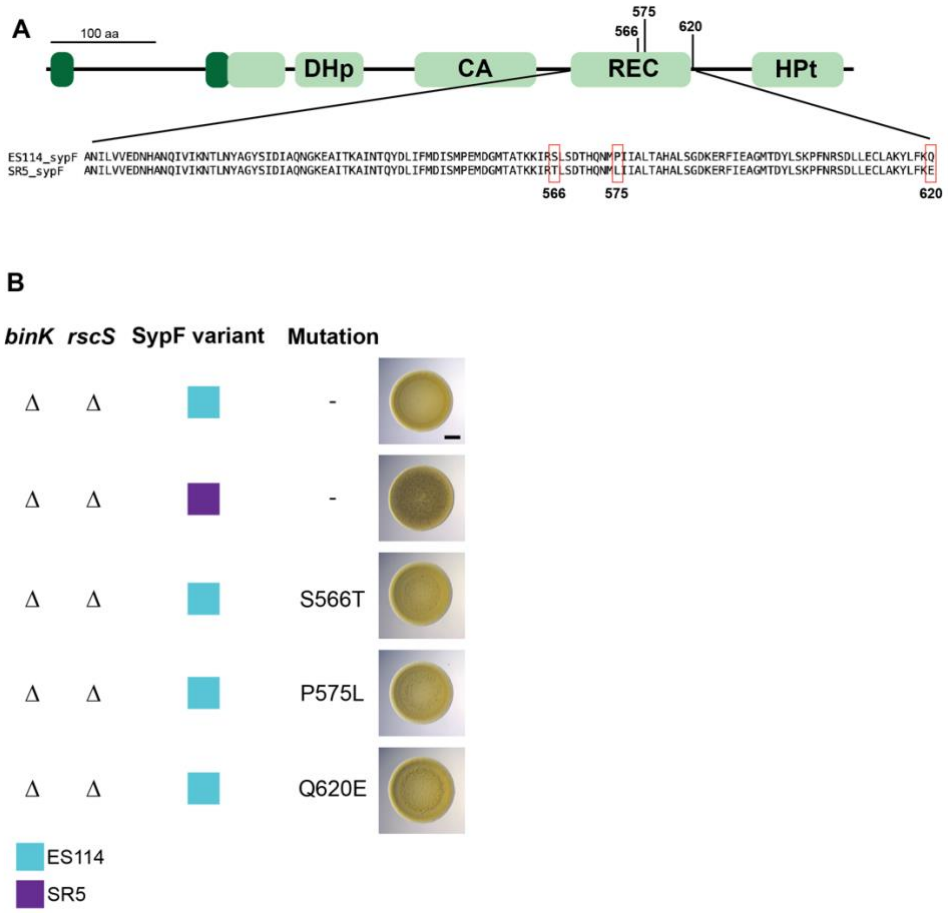
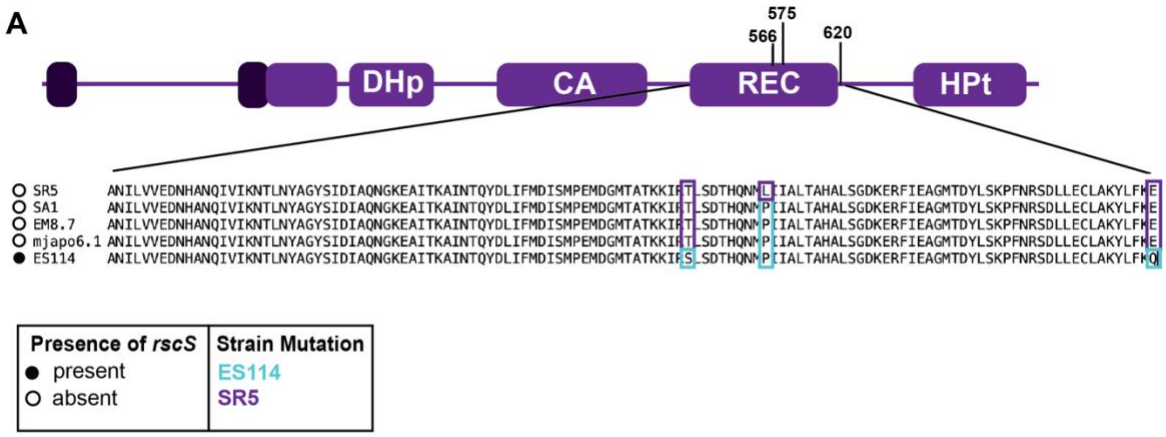


Figure 3.S3. Single amino acid mutations in the SypF REC domain are not sufficient for increased biofilm formation.

A) Amino acid alignment of the REC domain of SypF from ES114 and SR5 shows three amino acid changes. Red boxes outline amino acid changes between ES114 and SR5. **B)** Single amino acid changes of ES114 SypF are not sufficient for increased biofilm formation in ES114 Δ*binK* Δ*rscS* strains. Biofilm spot assay of strains grown at 20˚C on TBS for 72 h. Scale bar = 2 mm.



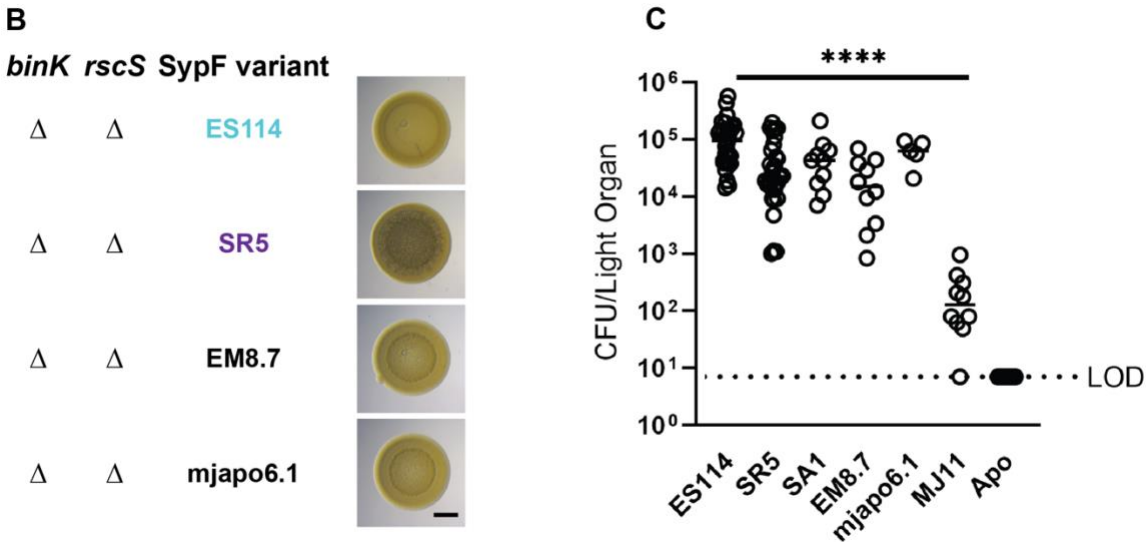


Figure 3.S4. Conserved amino acid change among natural isolates of *V. fischeri* is not sufficient for increased biofilm formation.

A) Amino acid alignment of SypF REC domain of natural isolates. The MAFFT alignment was performed using the Snapgene alignment tool. Colored boxes indicate conservation of REC domain amino acid changes with either SR5 or ES114. **B)** Natural isolate alleles of *sypF* are not sufficient for increased biofilm formation. Biofilm spot assay of ES114 Δ*rscS* Δ*binK* strain containing *sypF* ORF from natural isolates at the native locus. Strains were grown on TBS at 20°C for 72 h. Scale bar = 2 mm. **C)** Single strain colonization of natural isolates. Circles represent individual animals. Bars represent median values with the limit of detection of 7 CFU/light organ indicated by the dashed line. Hatchling squid were inoculated with $5.3x10^2$ - $1.0x10^4$ CFU/mL. Statistical significance was determined by Kruskal-Wallis test. ****, *P* = <0.0001.

Chapter 4: GacS and HahK promote symbiosis polysaccharide production in the Mediterranean squid-colonizing *Vibrio fischeri* SR5.

Katherine M. Bultman, Andrew C. Luy, Mark J. Mandel

Author contributions: KMB and MJM planned experiments, KMB and ACL conducted experiments, KMB and ACL constructed strains, KMB and MJM wrote the manuscript.

ABSTRACT

Horizontally acquired beneficial symbionts often require biofilm production to transition from a planktonic to a host associated lifestyle. Two-component signaling is one method used across bacterial species for the regulation of biofilm production. *Vibrio fischeri* uses two-component signaling to regulate production of the symbiosis polysaccharide (Syp) during the initial stages of the juvenile squid light organ colonization. In the well-studied Hawaiian isolate ES114, a phosphorelay between two hybrid histidine kinases, RscS and SypF, leads to phosphorylation of the response regulator SypG and biofilm production. The Mediterranean squid isolate SR5 lacks *rscS*, but instead has a strain specific REC domain of SypF that is sufficient for *rscS*-independent biofilm formation and squid colonization. Through the complementation of different SypF constructs, we found that similar to ES114, the histidine phosphotransfer (HPT) domain was sufficient to restore Syp production and host colonization in SR5. This led to a search for novel *syp* regulators in this isolate. We took advantage of the elevated baseline biofilm formation afforded by a Δ*binK* background to enable isolation of both biofilm-up and biofilm-down mutants. Transposon insertions in genes for the hybrid sensor kinases GacS and HahK demonstrated biofilm down phenotypes. Clean deletion of these genes resulted in decreased biofilm formation and *syp* transcription, suggesting they are potential positive biofilm regulators. Future work will determine the specific role of GacS and HahK in the phosphorelay controlling *syp,* providing insights into the diversity of biofilm regulatory networks.

INTRODUCTION

Bacteria live in a variety of environments and require the ability to adapt and respond to changing environmental conditions. Two-component signaling (TCS) occurs in both prokaryotic and eukaryotic species and functions to both recognize and respond to environmental signals

(28, 30, 44). Canonical TCS includes a histidine kinase that receives an external stimulus and catalyzes an autophosphorylation reaction leading to the downstream phosphorylation of a response regulator to elicit a cellular response (31, 36). Many response regulators act as a transcription factor with helix-turn-helix DNA binding domains resulting in a change in gene expression (31, 36). Phosphoryl group transfer between the histidine kinase and response regulator is highly specific, with co-evolved residues remaining highly conserved in the dimerization and histidine phosphotransfer (DHp) domain of the histidine kinase and the receiver (REC) domain of the response regulator (42). Approximately 25% of histidine kinases are hybrid, with tethered REC domains leading to a more elaborate phosphorelay to the downstream response regulator (36, 39). In comparison to canonical histidine kinases, tethering of the specificity DHp and REC domains enforces specificity by proximity as a lack of co-evolution between these two domain provides the potential for cross-talk between regulators and signaling pathways (45, 46). Some pathways use multikinase networks to coordinate a response to multiple environmental stimuli (47). In these networks, interactions between histidine kinases or multiple kinases phosphorylating the same response regulator results in more precise control of the downstream cellular response and bacterial behaviors (47).

In many cases, bacteria respond to specific environmental signals to transition from a planktonic lifestyle to a biofilm state during the host colonization process (57, 58, 190–194). This lifestyle switch for some bacteria is tightly controlled through two-component regulation. *Pseudomonas* species utilize a tripartite multikinase network to switch from a planktonic acute infection to a chronic biofilm state (70, 71). *Vibrio cholerae* relies on TCS to regulate biofilm formation for environmental survival and effective transmission to their host during infection (67, 152, 206). For other bacterial species, histidine kinases control the inhibition of biofilm formation (65, 73, 207). As biofilm formation is an important behavior for host colonization, precise control is necessary for the appropriate production during the colonization process.

To study the molecular mechanisms of TCS on biofilm control as part of host colonization, we focus on the marine bacterium *Vibrio fischeri* as the sole symbiont of the Hawaiian bobtail squid (*Euprymna scolopes*) light organ. During the beginning stages of colonization, *V. fischeri* are recruited from the seawater into a ciliated mucus field that surrounds the pores to the internal light organ (15, 23). Here, *V. fischeri* cells form aggregates through the production of an exopolysaccharide termed the symbiosis polysaccharide (Syp) (19). Production of Syp is regulated through the coordination of a multikinase network (112). In the canonical strain ES114, the current model suggests a phosphorelay between the hybrid sensor kinases RscS and SypF leading to phosphorylation of the downstream response regulator SypG (72, 114). In its phosphorylated state, SypG acts as an enhancer-binding protein for σ^{54}, which transcribes the *syp* locus that is required for subsequent biofilm formation (19, 116). A third hybrid sensor kinase, BinK, acts within the phosphorelay between SypF and SypG and functions to inhibit Syp production (73, 123). While the function of BinK is conserved among *V. fischeri* isolates, our previous work showed that there are two evolutionary groups that either lack *rscS* or the allele of this regulator is non-functional while still allowing squid colonization (132). The presence of *rscS* was proposed to be the key to squid specific colonization as fish isolates lacking this regulator could colonize squid if provided RscS *in trans* (131). However, identification of *rscS*-independent squid colonization demonstrated divergence within the *syp* multikinase network and the necessity of studying the diversity of these regulatory pathways within bacterial species.

We have previously identified a difference in the *syp* regulatory pathway between the Mediterranean squid isolate SR5 and the Hawaiian squid isolate ES114. While ES114 is dependent on RscS, SR5 has a divergent REC domain within SypF that is sufficient for *rscS*-independent biofilm formation and squid colonization (Ch. 3). Here we further examine biofilm regulation through SypF in SR5 and find that the HPt domain is sufficient for biofilm formation

and squid colonization. This motivated a search for novel *syp* regulators using SR5 as a representative ancestral squid-colonizing strain. Given the divergence of biofilm regulation between *V. fischeri* isolates, this work reinforces the necessity of studying strain heterogeneity in pathways that are important for host colonization.

RESULTS

The SypF HPt domain is sufficient for symbiosis polysaccharide production and squid colonization in strain SR5.

We demonstrated that the REC domain of SR5 SypF increases *syp* transcription and biofilm formation in a strain specific manner (Ch. 3). However, previous studies in ES114 have shown that only the final HPt domain of SypF is required for biofilm production and colonization (72). We therefore wanted to determine the requirement for the SR5 HPt domain. We constructed the following constructs for their ability to encode SypF function: full length SypF; full length SypF with an H250Q variant to ablate the initial phosphotransfer site; SypF with only the REC and HPt domains; and the SypF HPt domain alone (**Fig. 4.1A**). These constructs were expressed from a constitutive promoter at the neutral *Tn7* attachment site as previous attempts to complement *sypF* under its presumed native *PsypA* promoter did not result in successful complementation. To assess the biofilm formation by these constructs, we performed a colony biofilm assay in a Δ*binK* background on calcium supplemented media as this combination of strain background and media induces *syp* production in strain ES114 (62). As expected, a *sypF* deletion strain lacked wrinkled colony biofilm formation Each of the tested constructs produced rugose colony formation (**Fig. 4.1B**), demonstrating that the SR5 SypF HPt domain was sufficient to complement the Δ*sypF* mutant. In previous work, we demonstrated domain level strain specificity for SypF in which the SR5–but not ES114–SypF REC domain was sufficient to confer biofilm activation and squid colonization in the absence of upstream activator RscS.

Therefore, we next asked if the ES114 SypF HPt domain was functional to complement the SR5 Δ*sypF* mutant in the colony biofilm assay. We placed the ES114 HPt domain under the same promoter as the SR5 constructs at the *Tn7* attachment site and assessed this strain for biofilm formation. Similar to the SR5 HPt domain, the ES114 HPt domain alone resulted in rugose colony formation (**Fig. 4.1B**).

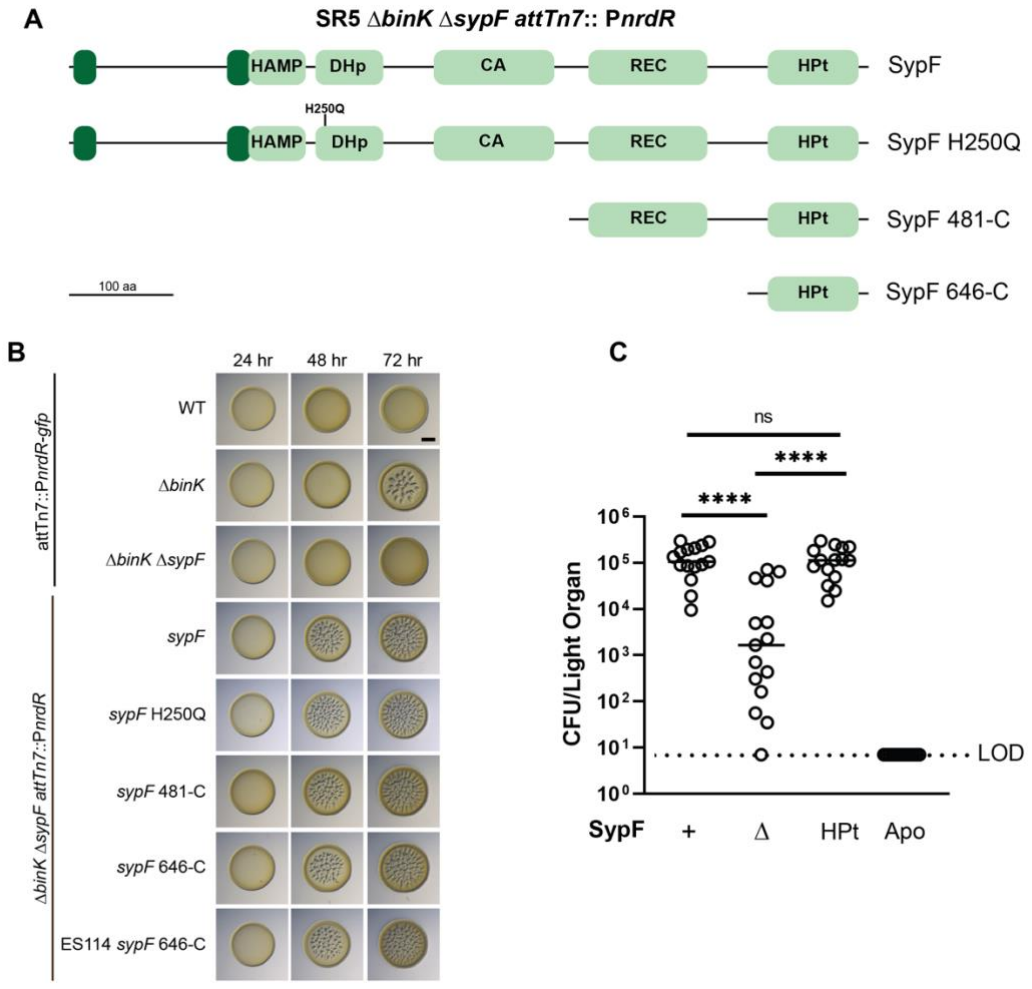


Figure 4.1. The HPt domain of SypF is sufficient for biofilm formation and squid colonization in SR5.
A) Diagram of SypF complementation constructs. Numbers correspond to the amino acid number in the protein. **B)** The HPt domain of SypF is sufficient for biofilm formation in SR5 Δ*binK* Δ*sypF* background. Biofilm spot assays of strains grown on TBS-calcium plates at 20˚C. Scale bar = 2 mm. **C)** The SypF HPt domain is sufficient for squid colonization in SR5. Single strain colonization experiment with circles representing individual animals. Horizontal bars represent the median for each set and the dashed line indicates the limit of detection of 7 CFU/light organ. Hatchling squid were inoculated with 2.4 x 10^3 - 1.1 x 10^4 CFU/mL bacteria.

Statistical comparisons for all colonization experiments were done with the Mann-Whitney test. *P*, **** < 0.0001.

The above result presented a paradox. On one hand, the two domains are highly similar between ES114 and SR5 (98/101 [97%] amino acid identity) and therefore observing a similar result in the two strains' HPt domains was not unexpected. Furthermore, the result is consistent with a previous study demonstrating that the ES114 SypF HPt domain was sufficient for colonization when expressed from a multicopy plasmid (72). On the other hand, we were surprised to observe this result given the strain specificity that we described previously for SR5 SypF (Ch. 3). We therefore investigated this effect further and asked how squid host colonization was impacted in a strain in which protein expression levels were as close to the parent as possible, with the HPt domain of *sypF* being expressed from the native *sypF* site in the chromosome. The Δ*sypF* strain showed a significant decrease in colonization compared to wild-type SR5, while expression of the HPt domain was sufficient for the bacteria to fully colonize juvenile squid (**Fig. 4.1C**).

Therefore, these results strongly support that expression of the SypF HPt domain in single copy is sufficient to restore the protein's biofilm and squid colonization phenotypes, and that this outcome is conserved in strains ES114 and SR5. Given that SR5 lacks the phosphodonor RscS, however, we reasoned that additional proteins must act to influence biofilm regulation in the Mediterranean symbiont, and we took a genetic approach to identify those factors.

Transposon screen identifies mutants with decreased or increased *syp* transcription.

Our goal was to identify both positive and negative regulators of the *syp* pathway in strain SR5. For visualization of *syp* transcription *in vitro*, we have a reporter plasmid with *PsypA'* driving the expression of *lacZ*+. However, *syp* is not expressed under normal laboratory conditions. We

took advantage of the deletion of the negative regulator BinK in ES114 increasing *syp* transcription and biofilm formation in culture and the conserved function of this regulator across isolates (73, 132). Placing the reporter plasmid into an SR5 Δ*binK* strain showed an intermediate level of blue color compared to wild-type SR5 (**Fig. 4.2C**). We therefore decided to use a Δ*binK* background for our screen as we could investigate the *syp* pathway without further genetic manipulations to our parent strain.

To create the transposon mutant library for this screen, we first introduced the mariner transposon plasmid pMarVF1 into the SR5 Δ*binK* parent strain (**Fig. 4.2A**). The resulting colonies were collected and the *PsypA'-lacZ*+ reporter plasmid was conjugated into this library and plated onto media containing X-gal to assess transcription levels. Candidates were screened for elevated ("ups") or diminished ("downs") blue color. We conducted secondary screening by re-streaking candidates onto fresh media to confirm the blue color phenotype in addition to a PCR screen for presence of the transposon (**Fig. 4.2A**). In total, we screened over 65,000 colonies and saved 244 possible candidates of which 83 had no color ("off"), 127 "downs", and 34 "ups" (**Fig. 4.2B**). For candidates that displayed an "off" phenotype, we performed an additional PCR screen to check for the presence of the reporter plasmid. These candidates were found to lack the reporter plasmid and were excluded from further analysis. Identification of transposon insertions in our candidates was determined using a combination of semi-arbitrarily primed PCR and insertion sequencing (see Materials and Methods). Candidate genes from our transposon screen are found in **Table 4.1**.

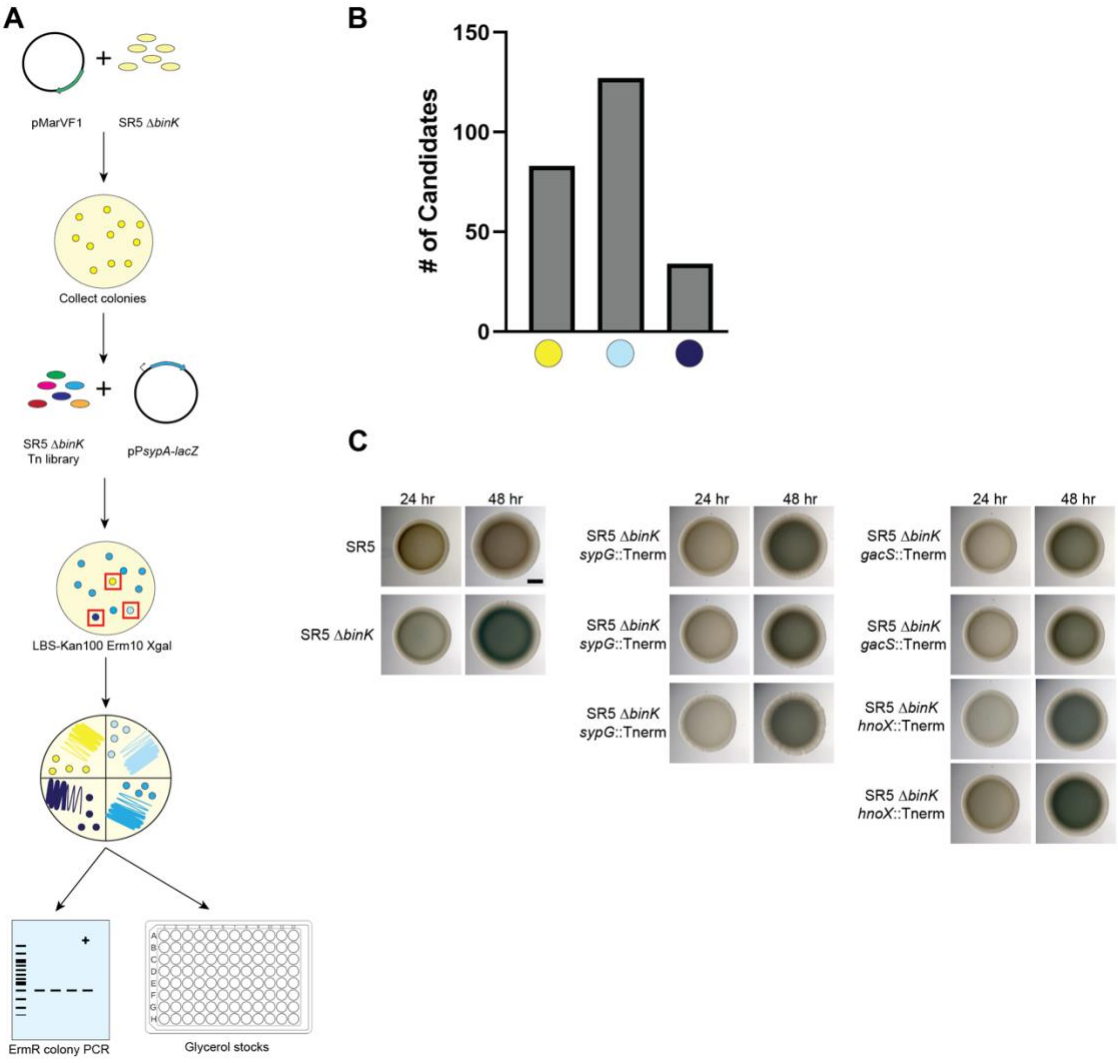


Figure 4.2. Transposon screen identifies *gacS* and *hahK* as potential *syp* regulators. **A)** Workflow of transposon screen in SR5 Δ*binK* background. **B)** Transposon screen candidate distribution. Each bar represents the number of candidates identified with that phenotype. Color corresponds to β-galactosidase activity seen in candidates compared to the parent strain. **C)** *gacS* and *hnoX* transposon candidates show decreased β-galactosidase activity. Spot assay of strains grown on LBS-Kan100 Xgal media at 25℃. Each candidate shown represents an independent insertion in the respective gene. Scale bar = 2 mm.

We wanted to find genes that function as positive regulators of *syp,* therefore we focused on candidates that showed a “down” phenotype. Four candidates were found to have independent insertions in the known response regulator *sypG*, demonstrating that this screen was correctly identifying *syp* regulators (**Fig. 4.2C**). Two other genes stood out from the identified candidates, each with multiple independent insertions, *hnoX* and *gacS*. While the *hnoX* candidates showed

only a slight “down” phenotype, this gene is directly upstream of the cytoplasmic hybrid histidine kinase HahK which has been shown to affect Syp production through the SypF HPt domain in ES114 (62). Compared to our *hnoX* candidates, the *gacS* candidates showed a “down” phenotype similar to the *sypG* mutants (**Fig. 4.2C)**. GacS is annotated as a hybrid histidine kinase and is highly conserved between ES114 and SR5 (99% amino acid identity), but its role in biofilm production in *V. fischeri* has never been investigated.

GacS and HahK impact biofilm formation and *syp* transcription in SR5.

To investigate the role of both GacS and HahK in biofilm formation in SR5, we generated clean deletions of each of these genes in an SR5 Δ*binK* background and performed colony biofilm assays. Under the conditions tested, deletion of binK produced robust wrinkled colony formation (**Fig. 4.3A**). Compared to the parent strain, deletion of *gacS* resulted in no wrinkled colony formation (**Fig. 4.3A**). Biofilm production was restored with complementation of *gacS* at the *Tn7* attachment site. HahK has been shown to act as a positive biofilm regulator in strains lacking *binK* (62). Deletion of *hahK* in our SR5 Δ*binK* strain did not produce wrinkled colony formation (**Fig. 4.3A**) For complementation, *hahK* was placed under the *V. fischeri PnrdR* promoter as this gene is encoded directly downstream of *hnoX*. Use of this promoter resulted in a partial complementation of biofilm formation (**Fig. 4.3A**).

As these genes were identified from our initial screen by a “down” phenotype for *syp* transcription, we measured the *syp* transcription levels in our deletion and complementation strains using the same reporter as our screen. The SR5 Δ*binK* strain had elevated levels of *syp* transcription as seen previously. Deletion of *gacS* resulted in an approximately 2-fold decrease in transcription that returned to the same level as the parent strain when complemented (**Fig. 4.3B**). In comparison, the *hahK* deletion decreased *syp* transcription almost 3-fold. Complementation of *hahK* under the *PnrdR* promoter increased transcription but not to the

same level as the parent strain. As HahK has been shown to be involved in biofilm formation under calcium inducing conditions (62), we also assessed *syp* transcription of these strains on media containing calcium. The overall pattern of transcription was the same between the two media conditions, however, the *gacS* deletion had a smaller reduction in *syp* transcription compared to when grown on media with no calcium (**Fig. 4.3B**).

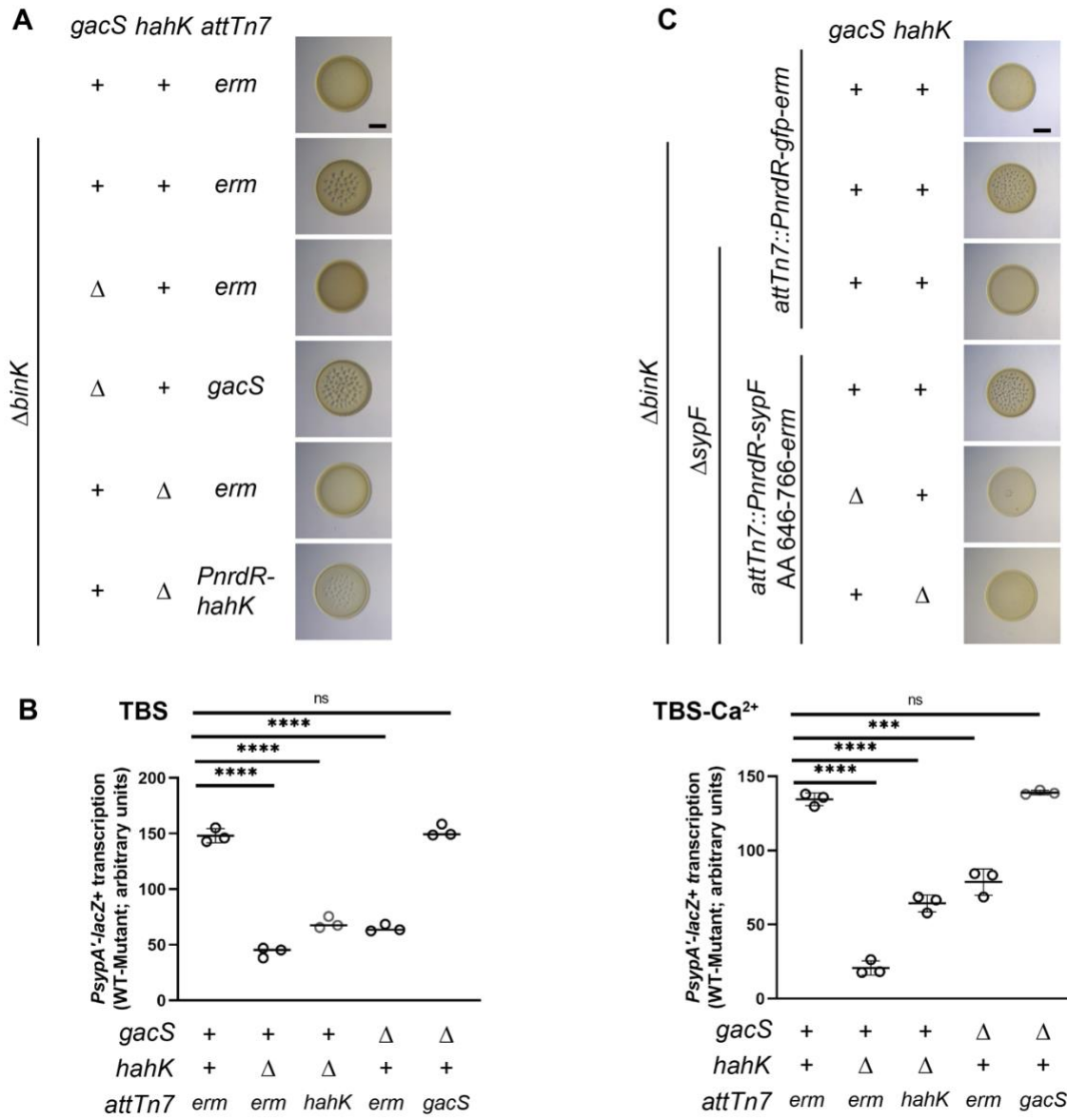


Figure 4.3. Deletion of *gacS* and *hahK* reduces biofilm formation and *syp* transcription in SR5.

A) Deletion of *gacS* or *hahK* in an SR5 Δ*binK* background reduces biofilm formation. Biofilm spot assay of strains grown at 20°C on TBS-calcium for 72 h. Scale bar = 2 mm. **B)** *syp* transcription is reduced in SR5 Δ*binK* background with *gacS* or *hahK* deletion. Strains were grown at 20°C for 48 h on TBS or TBS-calcium media. Graph represents quantification of β-galactosidase activity. Each spot represents the average of 3 technical replicates and error bars represent standard deviation. Statistical significance determined by Student's T-test, P, *** = <0.001, P, **** = <0.0001 **C)** Deletion of *gacS* or *hahK* reduces biofilm in the presence of the SypF HPt domain in an SR5 Δ*binK* background. Biofilm spot assay of strains grown at 20°C on TBS-calcium for 72 hrs. Scale bar = 2 mm.

With SR5 requiring only the HPt domain of SypF, we asked if deletion of either *gacS* or *hahK* would impact biofilm formation with only the HPt domain of SypF present. To answer this question we made clean deletions of *gacS* and *hahK* in an SR5 Δ*binK* Δ*sypF* background and then expressed the HPt domain from the *attTn7* site as in Figure 1A. Similar to when full length SypF was present, deletion of *gacS* or *hahK* in this background showed no wrinkled colony formation (**Fig. 4.3C**). These results show that in the presence of either full length SypF or the HPt domain alone, both GacS and HahK are acting as positive regulators of biofilm formation.

DISCUSSION

This study examines the molecular mechanisms of *syp* regulation and biofilm formation in the Mediterranean squid isolate SR5. It had previously been shown that regulation of this host specific exopolysaccharide is distinct from the well-studied isolate ES114 in that it lacks the hybrid sensor kinase *rscS* (132). Instead, amino acid changes within the REC domain of the hybrid sensor kinase SypF, promote *syp* transcription and biofilm formation that is sufficient for *rscS*-independent squid colonization. Here we further examined the specific regulatory strategy utilized by SR5 for *syp* transcription and subsequent biofilm formation. As this regulatory mechanism has been understudied in isolates other than ES114, our current study is one of the few to identify key players in *rscS*-independent *syp* regulation.

SypF plays a key regulatory role in Syp production acoss *V. fischeri* isolates. In ES114, it has been shown that only the HPt domain of SypF was necessary for both biofilm formation and squid colonization with RscS present (72). Here we show that the HPt domain of SypF is sufficient for biofilm formation and squid colonization in SR5. Unlike the REC domain, the HPt domain shows no strain specificity as SR5 is able to use both the ES114 and SR5 HPt domain for biofilm formation. Therefore, it is likely that the function of this domain is conserved and

phosphotransfer from an unidentified regulator could be occurring in both isolates. These data further support our model that strain specificity within the REC domain of SypF plays a key regulatory role for *rscS*-independent biofilm formation. Furthermore, these data suggest that the HPt domain is being phosphorylated by another protein as the catalytic core of SypF is not required for biofilm formation and squid colonization. In ES114, there is evidence that phosphotransfer occurs between the REC domain of RscS and the HPt domain of SypF (72). As HPt domains are known to function in a His-Asp-His-Asp multistep phosphorelay and SR5 lacks the phosphodonor RscS, it suggests that the HPt domain is receiving phosphates from another source (208, 209).

We have identified two potential positive *syp* regulators in SR5, GacS and HahK. Deletion of these genes in the presence of either full length SypF or the HPt domain alone reduced biofilm formation. This was further supported by a significant decrease in *syp* transcription when either of these genes were deleted. Both of these genes are annotated as hybrid sensor kinases and are conserved in ES114, with 99% amino acid identity for GacS and 91% for HahK. During colonization, disruption of either *gacS* or *hahK* resulted in decreased competitive fitness in ES114 (73, 122). It has been shown that HahK acts as a positive regulator of biofilm formation in ES114 (62). Under calcium inducing conditions, wrinkled colony formation is dependent on HahK and the HPt domain of SypF (62). Taken together, these data suggest a role for both of these proteins in *syp* regulation either through phosphoryl transfer to the HPt domain of SypF or through a yet unknown post transcriptional mechanism. To distinguish between these two models, it will be necessary to determine if these proteins either directly impact the phosphorylation state of the SypF HPt domain or function at or above the level of SypG within the pathway.

The role of GacS in biofilm formation and virulence factor control has been well studied in *Pseudomonas* species and is homologous to the BarA-UvrY regulatory system in *Escherichia coli (71, 210–212)*. However, only one study has looked at the function of this protein in ES114 and found that it may be responding to citrate and regulating its synthesis but a specific role in biofilm formation and colonization were not tested (213). In *Pseudomonas aeruginosa,* the GacS regulatory pathway coordinates the lifestyle switch from a planktonic acute infection to chronic biofilm state as found in cystic fibrosis patients (71, 214). The phosphorylation state of GacS is coordinated by interactions with two other hybrid histidine kinases, RetS and LadS, leading to the downstream phosphorylation of the response regulator GacA (70, 214, 215). When activated, GacA regulates the expression of two sRNA's, *rsmZ and rsmY*, resulting in the derepression of genes involved in virulence and biofilm formation (211). *V. fischeri* has homologs for both *gacA* and the sRNA's, however lacks homologs of the multikinase partners. Studies on GacA in *V. fischeri* demonstrate a role for this response regulator in luminescence and motility with a competitive disadvantage during the initial stages of colonization (216, 217). However, the role of GacA has yet to be directly studied in Syp biofilm formation. If *gacA* is required for biofilm formation, it would suggest that GacS function during *syp* regulation is likely through post transcriptional methods instead of direct interactions with either SypF or SypG similar to the regulation in other bacterial species.

Given that SR5 SypF functions in an ES114 Δ*rscS* background, our data are consistent with the model that the function of these regulators is conserved across isolates. Restoration of wrinkled colony formation in SR5 with the ES114 variant of these genes would further support this model. Since the single gene deletions show no wrinkled colony formation, it suggests that neither GacS nor HahK on their own are sufficient for biofilm formation. Here we propose two models for their function. The first, is that HahK promotes transcription of the *syp* locus through the HPt domain of SypF. Phosphorylation of GacA by GacS permits translation of the *syp* genes leading

to biofilm formation. A reduction of biofilm formation with the deletion of *gacA* would further support this model. If GacA is dispensable for biofilm formation, it would suggest our second model that GacS and HahK form a multikinase network with SypF to promote biofilm formation. If this regulation occurs through the HPt domain of SypF, the phosphotransfer between all three is expected to follow a canonical His-Asp-His-Asp relay. Based on previous data on HahK function in ES114, one possibility would be the transfer of phosphoryl groups from the catalytic domain of GacS to the REC domain of HahK before being transferred to the HPt domain of SypF (62). Pull down assays with tagged HahK and GacS would confirm interactions between these two regulators. In addition, wrinkled colony formation in a SR5 Δ*sypF*::SypF HPt background that contains a chimeric regulator of the catalytic core of GacS combined with the REC domain of HahK would suggest phosphotransfer between these regulators. *In vitro* phosphorylation assays would be necessary to confirm this model.

This work expands the current knowledge of SypF regulation and identifies previously unknown regulators that impact Syp biofilm formation. By studying the role of these proteins within this multikinase regulatory network, we gain a broader understanding of how biofilm formation is regulated across *V. fischeri* isolates. Future work will identify the direct role of GacS and HahK within the Syp regulatory pathway in SR5 in comparison with strains that are dependent on RscS leading to squid colonization.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. *V. fischeri* and *Escherichia coli* strains used in this study are listed in **Table 4.2**. Plasmids used in this study are listed in **Table 4.3**. *V. fischeri* strains were grown at 25°C in Luria-Bertani salt (LBS) medium (25 g Difco LB broth [BD], 10 g NaCl, and 50 ml 1 M Tris buffer [pH 7.5], per liter) or at 20˚C in tryptone broth salt (TBS)

medium (10 g Difco bacto-tryptone [BD], 20 g NaCl, and 50 ml 1 M Tris buffer [pH 7.5], per liter). *E. coli* strains, used for cloning and conjugation, were grown with shaking at 37°C in Luria-Bertani (LB) medium (25 g Difco LB broth [BD] per liter). Growth media were solidified with 1.5% agar (15 g Bacto agar [BD] per liter) as needed. When necessary, antibiotics were added to the medium at the following concentrations: erythromycin (Erm), 5mg/ml for *V. fischeri*; kanamycin (Kan), 100mg/ml for *V. fischeri* and 50mg/ml for *E. coli*; and chloramphenicol (Cam), 5mg/ml for *V. fischeri* and 25mg/ml for *E. coli*. The *E. coli* strain π3813 containing pKV496 is a thymidine auxotroph and was grown in LB with 50mg/ml kanamycin supplemented with 0.3 mM thymidine (177, 200).

DNA synthesis and sequencing. Each of the primers listed in **Table 4.4** was synthesized by Integrated DNA Technologies (Coralville, IA). Site directed mutagenesis primers were designed using the NEBaseChanger online tool. Full plasmids from cloned constructs were verified by Nanopore DNA sequencing at Plasmidsaurus (Eugene, OR). Insertion of constructs into the genome were verified by Nanopore DNA sequencing at Plasmidsaurus using linear amplicon sequencing or whole-genome sequencing, or was verified using Sanger DNA sequencing at Functional Biosciences via UW—Madison. Sequence data were analyzed with SnapGene. For cloning PCRs, we used Q5 high-fidelity DNA polymerase (NEB). For sequencing PCRs, we used OneTaq (NEB). For diagnostic PCR, we used GoTaq polymerase (Promega).

Construction of SR5 Δ*sypF* and Δ*binK* Δ*sypF* strains. Deletion of *sypF* was performed following the barcode-tagged gene deletion protocol from Burgos et al. (200). In brief, the upstream homology arm was amplified using primers KMB_149 and KMB_150 and the downstream homology arm was amplified using primers KMB_151 and KMB_152. Homology arms were fused to either side of a third fragment containing an *erm* cassette using splicing by overhang extension PCR (SOE PCR). Mutagenic DNA was purified using the Qiagen PCR

purification kit and transformed into SR5 via transformation using pLostfoX (MJM1539) (178, 201). Mutant candidates were selected using erythromycin and screened by PCR using primer pairs KMB_148/KMB_153, KMB_148/HB8, and KMB_154/KMB_155. Insertion of the *erm-bar* scar was confirmed by Sanger sequencing using primers KMB_148, KMB_149, HB8, HB9, KMB_152, KMB_153, and the barcode sequence was recorded. The final barcode strain (MJM4016) was constructed via a triparental mating using donor MJM3478 (π3813/pKV496) (177) and helper strain MJM534 (CC118 λpir/pEVS104) with MJM4015. Candidates were selected for using kanamycin and screened by PCR using the primer pairs listed above.The deletion scar was verified by Sanger sequencing using primers KMB_148, KMB_149, HB41, HB42, KMB_152, KMB_153.

The SR5 Δ*binK::bar* strain (MJM4037) was constructed via a triparental mating with donor MJM3478 (p3813/pKV496) (177) and helper strain MJM534 (CC118 λpir/pEVS104) with MJM3571. Candidates were selected for using kanamycin and screened by PCR using the primer pairs binK-F1/binK-R2, HB8/binK-FO, and KMB_036/KMB_037. The deletion scar was verified by Sanger sequencing using primers binK-FO, binK-F1, HB41, HB42, and binK-R2.

To generate the SR5 Δ*binK::bar* Δ*sypF::bar* strain (MJM4863), the previously described pLostfoX plasmid (178, 201) was conjugated via triparental mating into MJM4037. Candidates were selected on chloramphenicol and presence of the plasmid was confirmed using RYI072 F/RYI072 R and saved as MJM4118. Genomic DNA from MJM4015 was extracted using the Qiagen DNeasy Blood and Tissue kit and transformed into MJM4118. Mutant candidates were selected on erythromycin and screened by PCR using primer pairs KMB_148/KMB_153, KMB_148/HB8, and KMB_154/KMB_155. Insertion of the *erm-bar* scar was confirmed by Sanger sequencing using primers KMB_148, KMB_149, HB8, HB9, KMB_152, KMB_153, and the barcode sequence was recorded. The final barcode strain (MJM4863) was constructed via a

triparental mating using donor MJM3478 (π3813/pKV496) (177) and helper strain MJM534 (CC118 λpir/pEVS104) with MJM4862. Candidates were selected for using kanamycin and screened by PCR using the primer pair KMB_148/KMB_153. The deletion scar was verified by Sanger sequencing using primers KMB_148, KMB_149, HB41, HB42, KMB_152, KMB_153.

Construction of *sypF* complementation strains. A pEVS107-*PnrdR-sypF HPt* [MJM1100] plasmid was constructed by amplifying the pEVS107 backbone using primer pair DAT_385/DAT_386. The *PnrdR* promoter was amplified from pMarVF1 using the primer pair DAT_387/DAT_388. *sypF* HPt domain fragment was amplified from MJM1100 genomic DNA extracted using the Qiagen DNeasy Blood and Tissue kit using the primer pair DAT_389/DAT_390. Each fragment was cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells and selected on Kanamycin media. Candidate colonies were screened with pEVS107 F/pEVS107 R. Positive candidates were sequenced by Plasmidsaurus. The confirmed plasmid was saved as pDAT25.

All of the *sypF* complementation plasmid vector backbones were amplified from pDAT25 using the primer pair KMB_517/JFB_425. *sypF* fragments were amplified from MJM1125 genomic DNA extracted using the Qiagen DNeasy Blood and Tissue kit. Full length *sypF* was amplified using primer pair KMB_518/KMB_519, the REC-HPt fragment was amplified using primer pair KMB_521/KMB_519, and the HPt fragment was amplified using primer pair KMB_520/KMB_519. Each fragment was cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells and selected on Kanamycin media. Candidate colonies were screened with pEVS107 F/pEVS107 R. Positive candidates were sequenced by Plasmidsaurus. Confirmed plasmids were saved as pKMB027, pKMB028 and pKMB036. To generate the SypF

H250Q complement, the previously generated vector pKMB027 was purified and used as a template. The point mutant was made using KMB_515/KMB_516 with the Q5 Site Directed Mutagenesis kit (NEB). Resulting plasmids were transformed into DH5α λpir cells, selected on Kanamycin media, and screened using primer pairs KMB_526/KMB_528 (WT) and KMB_527/KMB_528 (mutant). Positive candidates were sent to Plasmidsaurus for sequencing and the confirmed plasmid was saved as pKMB032. *sypF* alleles generated in this manner were then introduced into *V. fischeri* (MJM4863) by tetraparental mating by mixing the pEVS104-containing helper, pUX-BF13-containing transposase, pEVS107 mini-*Tn7* vector-containing donor, and the *V. fischeri* recipient (185). PCR verification by amplifying around the *attTn7* site with primers Tn7 Site F and Tn7 Site R confirmed transposon insertion at the *attTn7* site. Linear amplicon sequencing using these primers was performed by Plasmidsaurus to confirm insertion and correct sequence.

Generation of *attTn7::PnrdR-gfp* strains. The control pEVS107-PnrdR-gfp plasmid (pKMB041) was constructed using the same vector backbone as the *sypF* complements. *gfp* from purified pVSV102 was amplified using the primer pair KMB_565/KMB_566. The fragment was cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells and selected for on Kanamycin media. Candidate colonies were screened with pEVS107 F/pEVS107 R. Positive candidates were sequenced by Plasmidsaurus.

pKMB041 was introduced into *V. fischeri* strains MJM1125, MJM4037, and MJM4863 by tetraparental mating by mixing the pEVS104-containing helper, pUX-BF13-containing transposase, pEVS107 mini-*Tn7* vector-containing donor, and the *V. fischeri* recipient (185). PCR verification by amplifying around the *attTn7* site with primers Tn7 Site F and Tn7 Site R

confirmed transposon insertion at the *attTn7* site. Linear amplicon sequencing using these primers was performed by Plasmidsaurus to confirm insertion and correct sequence.

Colony biofilm assays. Cultures were grown for approximately 18 hrs overnight and 8 µL was spotted onto TBS or TBS-calcium (10 mM $CaCl_2$). Spots were allowed to dry at room temperature and plates were grown at 20˚C for 72 hrs and imaged using a Leica M60 stereomicroscope with Leica Firecam software.

Construction of SypF HPt domain at native locus. The previously described pKMB021 plasmid was used as the template for the vector backbone (Ch. 3). The vector was purified and amplified using the primer pair KMB_377/KMB_378. To amplify the HPt domain of SypF from SR5, pKMB036 was purified and the fragment was amplified using the primer pair KMB_567/KMB_568. The fragment was cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells and selected for on Chloramphenicol media. Candidate colonies were screened with RYI072 F/RYI072 R. Positive candidates were sequenced by Plasmidsaurus. The resulting plasmids was saved as pKMB042 and conjugated into the *V. fischeri* recipient (MJM1125) by triparental mating with helper plasmid pEVS104, selecting for the chloramphenicol resistance of the plasmid backbone. Single recombinants in *V. fischeri* were screened for maintaining chloramphenicol resistance. To obtain double recombinants, single recombinants were then grown without antibiotics and patched onto LBS and LBS-Cam to find isolates that lost the antibiotic resistance cassette. These candidates were then verified with PCR using the primer pair KMB_325/KMB_550 and this fragment was sent for linear amplicon sequencing at Plasmidsarus.

Construction of mariner transposon library and *syp* transcription reporter screen. The mariner transposon plasmid for *V. fischeri* pMarVF1 was conjugated into the DAP auxotroph β3914 and selected for on LB-Carb100 DAP(0.3mM) and saved as MJM1431. The mariner transposon library was generated in the MJM4037 background using the laboratory's mariner library in *V. fischeri* protocol (https://zenodo.org/records/1470836). In brief, MJM1431 and MJM4037 overnight cultures were combined and spotted onto LBS plates and grown overnight. Mating spots were resuspended in LBS and transposon mutants were selected for on Erythromycin media.

An SR5 *sypA'-lacZ*+ transcriptional reporter plasmid was constructed by amplifying *PsypA'* from MJM1125 using primer pair MJM-475F/MJM-476R. This fragment was cloned into the SphI-SpeI sites in pADK701 to generate the *lacZ* transcriptional fusion.

Transposon mutants were collected and used as the recipient strain in a triparental mating with MJM1421 to insert the reporter plasmid. The resulting pool of mutants were selected and screened on media containing Erythromycin, Kanamycin, and X-gal. Plates were allowed to grow for 48 hrs at 25˚C and visually screened for colonies with either none, elevated or diminished blue color compared to the parent strain. Potential candidates were re-streaked onto media containing Erythromycin, Kanamycin, and X-gal to confirm the β-galactosidase phenotype. Transposon insertion for all candidates were confirmed via PCR screening with KMB_128 and KMB_129. For candidates with no color, presence of the reporter plasmid was confirmed via PCR screening with M13F (-41) and MJM-431 R.

Identification of transposon mutant candidates. Candidates were identified either using semi-arbitrarily primed PCR or insertion sequencing. For the semi-arbitrarily primed PCR, the laboratory arbitrarily primed PCR protocol was followed (https://zenodo.org/records/1470836). In

brief, genomic DNA was isolated from each candidate using the Qiagen DNeasy Blood and tissue kit. First round PCR was performed using primers ARB1 and MJM-440. Reactions were purified using the Qiagen PCR purification kit. Second round PCR was performed using ARB2 and MJM-477. Reactions were purified using the Qiagen PCR purification kit and sent for Sanger sequencing using MJM-477. Candidate genes were identified by a blastn (203) search against the SR5 genome.

For insertional sequencing, candidates were pooled by column and row of the 96-well plate and genomic DNA was isolated using the Qiagen DNeasy blood and tissue kit. For each sample 1 µg library DNA was prepared for insertion sequencing (INSeq) as published (218). Briefly, linear amplification from both ends of the transposon was performed with a biotinylated primer to allow for the transposon ends and neighboring chromosomal sequence to be affixed to magnetic streptavidin beads. A second round of amplification completed the double-stranded molecule. Digestion with the type II endonuclease MmeI recognized sequences within the transposon inverted repeat, leading to cleavage in the flanking chromosomal DNA: 20 bp from the recognition site and 16 bp from the end of the transposon. Adapters were ligated and limited amplification released the short (~125 bp) transposon– chromosome junction sequences from the beads. These enriched junctions were then sequenced by Illumina single-end 50-bp sequencing on a HiSEq. 2000 at the Tufts University Core Facility for Genomics. Candidates were identified using pyinseq (219) and comparison of genes found in each row and column.

Transposon mutant spot assay. The reporter plasmid pM1421 was conjugated into MJM1125 and MJM4037 via triparental mating. Candidates were selected for on Kanamycin media and the presence of the plasmid was confirmed via PCR screening with M13F (-41) and MJM-431 R and saved as MJM1440 and MJM4056 respectively. Transposon mutant candidates, MJM1440, and MJM4056 were grown in liquid culture for approximately 18 h at 25˚C in LBS-Kan media

and then 8 µL of culture was spotted onto LBS media containing Kanamycin and X-gal. Spots were allowed to dry at room temperature and plates were grown at 25˚C for 48 h and imaged using a Leica M60 stereomicroscope with Leica Firecam software.

Construction of SR5 Δ*binK::bar* Δ*gacS:bar* strain. Deletion of *gacS* was performed following the barcode-tagged gene deletion protocol from Burgos et al. (200). In brief, the upstream homology arm was amplified using primers KMB_418/KMB_419 and the downstream homology arm was amplified using primers KMB_422/KMB_423. Homology arms were fused to either side of a third fragment containing an erm cassette using splicing by overhang extension PCR (SOE PCR). Mutagenic DNA was purified using the Qiagen PCR purification kit and transformed into SR5 Δ*binK::bar* via transformation using pLostfoX (MJM4118) (178, 201). Mutant candidates were selected using erythromycin and screened by PCR using primer pairs KMB_417/KMB_424, KMB_417/HB8, and KMB_420/KMB_421. Insertion of the *erm-bar* scar was confirmed by linear amplicon sequencing using primers KMB_417, KMB_424 and the barcode sequence was recorded. The final barcode scar strain (MJM4966) was constructed via a triparental mating with donor MJM3478 (p3813/pKV496) (177) and helper strain MJM534 (CC118 λpir/pEVS104) with MJM4965. Candidates were selected for using kanamycin and screened by PCR using the primer pair KMB_417/KMB_424. The deletion scar was verified by linear amplicon sequencing.

Construction of SR5 Δ*hahK:bar* and *ΔbinK::bar ΔhahK::bar* strains. Deletion of *hahK* was performed following the barcode-tagged gene deletion protocol from Burgos et al. (200). In brief, the upstream homology arm was amplified using primers KMB_102/KMB_103 and the downstream homology arm was amplified using primers KMB_106/KMB_107. Homology arms were fused to either side of a third fragment containing an erm cassette using splicing by

overhang extension PCR (SOE PCR). Mutagenic DNA was purified using the Qiagen PCR purification kit and transformed into SR5 via transformation using pLostfoX (MJM1539) (178, 201). Mutant candidates were selected using erythromycin and screened by PCR using primer pairs KMB_101/KMB_108, KMB_101/HB8, and KMB_104/KMB_105. Insertion of the *erm-bar* scar was confirmed by Sanger sequencing using primers KMB_101, KMB_102, HB8, HB9, KMB_107, and KMB_108 and the barcode sequence was recorded. The final barcode scar strain (MJM3953) was constructed via a triparental mating with donor MJM3478 (p3813/pKV496) (177) and helper strain MJM534 (CC118 λpir/pEVS104) with MJM3952. Candidates were selected for using kanamycin and screened by PCR using the primers listed above. The deletion scar was verified by Sanger sequencing using primers KMB_101, KMB_102, HB41, HB42, KMB_107, and KMB_108.

To delete *hahK* in the SR5 *ΔbinK::bar* strain (MJM4037), the same SOE product from above was transformed into MJM4037 via transformation using pLostfoX (MJM4118). Mutant candidates were selected using erythromycin and screened by PCR using primer pairs KMB_101/KMB_108, KMB_101/HB8, and KMB_104/KMB_105. Insertion of the *erm-bar* scar was confirmed by Sanger sequencing using primers KMB_101, KMB_102, HB8, HB9, KMB_107, and KMB_108. The final barcode scar strain (MJM4156) was constructed via a triparental mating with donor MJM3478 (p3813/pKV496) (177) and helper strain MJM534 (CC118 l pir/pEVS104) with MJM4155. Candidates were selected for using kanamycin and screened by PCR using the primers listed above. The deletion scar was verified by Sanger sequencing using primers KMB_101, KMB_102, HB41, HB42, KMB_107, and KMB_108.

Construction of *attTn7::Erm* strains. The previously described pEVS107 plasmid was then introduced into *V. fischeri* (MJM1125, MJM4037, MJM4966, MJM4156) by tetraparental mating by mixing the pEVS104-containing helper, pUX-BF13-containing transposase, pEVS107 mini-

Tn7 vector-containing donor (MJM658), and the *V. fischeri* recipient (185). PCR verification by amplifying around the *attTn7* site with primers Tn7 Site F and Tn7 Site R confirmed transposon insertion at the *attTn7* site. Linear amplicon sequencing using these primers was performed by Plasmidsaurus to confirm insertion and correct sequence.

Construction of gene complementation strains. The previously described plasmid pEVS107 was purified and used as a template for the vector backbone. The plasmid was amplified using the primer pair JFB_424/JFB_425. The open reading frame of *gacS* and approximately 300 bp US and DS of the gene were amplified using the primer pair KMB_579/KMB_580. The fragment was cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into DH5α λpir chemically competent cells and selected for on Kanamycin media. Candidate colonies were screened with pEVS107 F/pEVS107 R. Positive candidates were sequenced by Plasmidsaurus and saved as pACL04. The plasmid generated was then introduced into *V. fischeri* (MJM4966) by tetraparental mating by mixing the pEVS104-containing helper, pUX-BF13-containing transposase, pEVS107 mini-*Tn7* vector-containing donor, and the *V. fischeri* recipient (185). PCR verification by amplifying around the *attTn7* site with primers Tn7 Site F and Tn7 Site R confirmed transposon insertion at the *attTn7* site. Linear amplicon sequencing using these primers was performed by Plasmidsaurus to confirm insertion and correct sequence.

To complement *hahK*, the previously described plasmid pKMB027 was purified and used as a template for the vector backbone. The plasmid and *PnrdR* promoter was amplified using the primer pair KMB_517/JFB_425. The open reading frame of *hahK* was amplified from SR5 genomic DNA that was purified with the Qiagen DNeasy Blood and Tissue kit using the primer pair KMB_673/KMB_674. The fragment was cloned into the vector backbone by Gibson Assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB) and transformed into

DH5α λpir chemically competent cells and selected for on Kanamycin media. Candidate colonies were screened with pEVS107 F/pEVS107 R. Positive candidates were sequenced by Plasmidsaurus and saved as pACL03. The plasmid generated was then introduced into *V. fischeri* (MJM4156) by tetraparental mating by mixing the pEVS104-containing helper, pUX-BF13-containing transposase, pEVS107 mini-*Tn7* vector-containing donor, and the *V. fischeri* recipient (185). PCR verification by amplifying around the *attTn7* site with primers Tn7 Site F and Tn7 Site R confirmed transposon insertion at the *attTn7* site. Linear amplicon sequencing using these primers was performed by Plasmidsaurus to confirm insertion and correct sequence.

Construction of SR5 Δ*binK::bar* Δ*sypF::bar attTn7 sypF* HPt with *gacS* or *hahK* deletions. To make deletions in the MJM4863 background, the previously described pLostfoX plasmid (73, 178) was moved into the *V. fischeri* recipient using triparental mating with helper plasmid pEVS104 and selecting for the chloramphenicol resistance of the plasmid backbone. The presence of the plasmid was confirmed using RYI072 F/RYI072 R and saved as MJM5814.

To generate the *gacS* deletion in this strain, gDNA from MJM4965 was extracted using the Qiagen DNeasy Blood and Tissue kit and transformed into MJM5814 via natural transformation. Mutant candidates were selected using erythromycin and screened by PCR using primer pairs KMB_417/KMB_424, KMB_417/HB8, and KMB_420/KMB_421. Insertion of the *erm-bar* scar was confirmed by linear amplicon sequencing using the primer pair KMB_417/KMB_424. The final barcode scar strain (MJM6000) was constructed via a triparental mating with donor MJM3478 (p3813/pKV496) (177) and helper strain MJM534 (CC118 λpir/pEVS104) with MJM5999. Candidates were selected for using kanamycin and screened by PCR using the primer pair KMB_417/KMB_424. The deletion scar was verified by linear amplicon sequencing.

To generate the *hahK* deletion in this strain, gDNA from MJM3952 was extracted using the Qiagen DNeasy Blood and Tissue kit and transformed into MJM5814 via natural transformation. Mutant candidates were selected using erythromycin and screened by PCR using primer pairs KMB_101/KMB_108, KMB_101/HB8, and KMB_104/KMB_105. Insertion of the *erm-bar* scar was confirmed by linear amplicon sequencing using the primer pair KMB_101/KMB_108. The final barcode scar strain (MJM6004) was constructed via a triparental mating with donor MJM3478 (p3813/pKV496) (177) and helper strain MJM534 (CC118 λpir/pEVS104) with MJM6003. Candidates were selected for using kanamycin and screened by PCR using the primer pair KMB_101/KMB_108. The deletion scar was verified by linear amplicon sequencing. Maintenance of the Δ*binK::bar* and Δ*sypF::bar* cassettes were confirmed via PCR using the primer pairs KMB_036/KMB_037 and KMB_154/KMB_155 respectively.

The sypF HPt domain was introduced into MJM6000 and MJM6004 by tetraparental mating by mixing the pEVS104-containing helper, pUX-BF13-containing transposase, pEVS107 mini-*Tn7* vector-containing donor (MJM4890), and the *V. fischeri* recipient (185). PCR verification by amplifying around the *attTn7* site with primers Tn7 Site F and Tn7 Site R confirmed transposon insertion at the *attTn7* site. Linear amplicon sequencing using these primers was performed by Plasmidsaurus to confirm insertion and correct sequence.

***syp* transcriptional reporter strain construction and *in vitro* assay.** The reporter plasmid pM1421 was conjugated into the following *V. fischeri* recipients (MJM3646, MJM4389, MJM5813, MJM5799, MJM6080, and MJM6102) by triparental mating with helper plasmid pEVS104, selecting for the Kanamycin resistance of the plasmid backbone. Potential candidates were screened for the presence of pM1421 using the primer pair M13 F (-41)/MJM-431 R. The resulting strains were grown for approximately 18 hrs at 25˚C in LBS-Kan media and then 8 µL of culture was spotted onto both TBS and TBS-calcium plates containing X-gal (100

µg/mL) and Kanamycin. Spots were allowed to dry at room temperature, then grown for 48 hrs at 20˚C.

Spots were transferred onto printer paper and allowed to dry at room temperature. Once dry, spots were scanned and saved as a tiff file. Quantification of the spots was performed using ImageJ. In brief, the mean gray value of each spot was quantified and the mean gray value of the background was subtracted. The mean gray values of the technical triplicates for each biological replicate were averaged and values reported as arbitrary units.

Data analysis. SnapGene was used to align sequencing data to reference to confirm plasmid and strain construction. GraphPad Prism was used to construct graphs and perform statistical analyses.

TABLES AND SUPPLEMENTAL FIGURES

Table 4.1. Candidate genes from *syp* transcription transposon screen

Candidates with diminished ("down") color	
Disrupted Gene	**Predicted function**
VFSR5_0100	GTP-binding protein
VFSR5_0125	bifunctional phosphopantothenoylcysteine decarboxylase/
VFSR5_0184	acetyltransferase
VFSR5_0257	arginine repressor
Ig(VFSR5_0269 - VFSR5_0270)	hypothetical protein - hypothetical protein
VFSR5_0365	Magnesium transporter
VFSR5_0370	RNA-polymerase factor sigma-54
VFSR5_0302	divalent anion:sodium symporter family protein
VFSR5_0465 rrmJ	23S rRNA methyltransferase
VFSR5_0523 mutS	DNA mismatch repair MutS
VFSR5_0714 glyA	serine hydroxymethyltransferase

VFSR5_0735	tRNA s(4)U8 sulfurtransferase
VFSR5_0759 proA	gamma-glutamyl phosphate reductase
VFSR5_0797	uroporphrin III-C-methyltransferase
VFSR5_0821 clpP	ATP-dependent Clp protease proteolytic subunit
VFSR5_0829 asnB	asparagine synthetase B
Ig(VFSR5_0939 - lrp)	alanine dehydrogenase - leucine responsive transcriptional regulator
VFSR5_1240	tyrosine-specific transport protein
VFSR5_1269	coproporphyrinogen III oxidase
VFSR5_1535	hypothetical protein
VFSR5_1734 ansA	cytoplasmic asparaginase 1
VFSR5_1775	coniferyl aldehyde dehydrogenase
VFSR5_1803	long-chain-fatty-acid -- CoA ligase
VFSR5_1983 flgM	Negative regulator of flagellin synthesis flgM
VFSR5_1997	cysteine synthase A
VFSR5_2099	DnaJ
VFSR5_2125	amino-acid carrier protein AlsT
VFSR5_2148 gacS	hybrid sensory histidine kinase
IG VFSR5_2253 – VFSR5_2254	hypoxanthine phosphoribosyltransferase - LitR
VFSR5_2313	phosphate transport regulator
VFSR5_2314	low-affinity inorganic phosphate transporter
VFSR5_2518	RNA-binding protein
VFSR5_2563	phosphoenolpyruvate carboxykinase
VFSR5_2598	cyclic nucleotide binding protein
VFSR5_A0091 hnoX	Heme NO binding protein
VFSR5_A0975 sypG	Sigma-54 interacting response regulator
VFSR5_A1059	transcriptional regulator MarR family
VFSR5_A0169	hypothetical protein
VFSR5_A0625	hypothetical protein
VFSR5_A0876	LuxD acyl transferase
VFSR5_A0896	bifunctional PTS system fructose specific transporter subunit IIA/HPr protein
VFSR5_A0903	VgrG protein
VFSR5_A1058	hydrogenase cytochrome B-type subunit
VFSR5_A1101	putative transporter
VFSR5_1976 flgE	flagellar hook protein FlgE
uphT	sugar phosphate antiporter
Candidates with enhanced (“up”) color	
Disrupted Gene	**Predicted function**
VFSR5_0171	putative carbamoyl-phosphate synthase large chain
VFSR5_0178	glycosyltransferase
VFSR5_0256	immunogenic protein
VFSR5_0456	cis-trans isomerase

VFSR5_0523 mutS	DNA mismatch repair MutS
VFSR5_0611	HTH-type transcriptional regulator AscG
VFSR5_0821 clpP	ATP-dependent Clp protease proteolytic subunit ClpP
VFSR5_0822 clpX	ATP-dependent Clp protease proteolytic subunit ClpX
VFSR5_0831	protein mlc
VFSR5_0924	DNA-binding transcriptional regulator HexR
IG (VFSR5_0924 - VFSR5_0925)	DNA-binding transcriptional regulator HexR - glutamate decarboxylase
VFSR5_1345	sensor protein LuxQ-like protein
VFSR5_1938 flhF	flagellar biosynthesis regulator
VFSR5_A0147	hypothetical protein
VFSR5_A0625	hypothetical protein
VFSR5_A0637	lipoprotein, putative
VFSR5_A0937	trimethylamine-N-oxide reductase
VFSR5_A1101	putative transporter
VFSR5_0727 luxQ	autoinducer 2 sensor kinase/phosphatase LuxQ

Table 4.2. Bacterial Strains

V. fischeri		
Strain	**Genotype**	**Source or Reference**
MJM1125 = SR5	natural isolate, squid light organ	(128, 133)
MJM1100 = ES114	natural isolate, squid light organ	(101, 204)
MJM1440	MJM1125/pM1421	This study
MJM1539	MJM1125/pLostfoX	This study
MJM3571	MJM1125 Δ*binK::erm-bar*	(132)
MJM4015	MJM1125 Δ*sypF::erm-bar*	This study
MJM4016	MJM1125 Δ*sypF::bar*	This study
MJM4037	MJM1125 Δ*binK::bar*	This study
MJM4118	MJM4037/pLostfoX	This study
MJM4056	MJM4037/pM1421	This study
MJM4862	MJM4037 Δ*sypF::erm-bar*	This study
MJM4863	MJM4037 Δ*sypF::bar*	This study
MJM4998	MJM4863 *attTn7::PnrdR-gfp-erm*	This study
MJM4999	MJM4863 *attTn7::PnrdR-SR5 SypF-erm*	This study
MJM5000	MJM4863 *attTn7::PnrdR-SR5 SypF H250Q-erm*	This study
MJM5002	MJM4863 *attTn7::PnrdR-SR5 sypF 646-C-erm*	This study

MJM5001	MJM4863 *attTn7::PnrdR-SR5 SypF 481-C-erm*	This study
MJM5015	MJM4037 *attTn7::PnrdR-gfp-erm*	This study
MJM5033	MJM1125 *attTn7::PnrdR-gfp-erm*	This study
MJM4965	MJM4037 Δ*gacS::erm-bar*	This study
MJM4966	MJM4037 Δ*gacS::bar*	This study
MJM3952	MJM1125 Δ*hahK::erm-bar*	This study
MJM3953	MJM1125 Δ*hahK::bar*	This study
MJM4155	MJM4037 Δ*hahK::erm-bar*	This study
MJM4156	MJM4037 Δ*hahK::bar*	This study
MJM3646	MJM1125 *attTn7::erm*	This study
MJM4389	MJM4037 *attTn7::erm*	This study
MJM5813	MJM4966 *attTn7::erm*	This study
MJM6080	MJM4156 *attTn7::erm*	This study
MJM5799	MJM4966 *attTn7::gacS-erm*	This study
MJM6102	MJM4156 *attTn7::PnrdR-hahK-erm*	This study
MJM5520	MJM1125 Δ*sypF::sypF 646-C*	This study
MJM5814	MJM4863/pLostfoX	This study
MJM5848	MJM3646/pM1421	This study
MJM4416	MJM4389/pM1421	This study
MJM5849	MJM5813/pM1421	This study
MJM5811	MJM5799/pM1421	This study
MJM6136	MJM6080/pM1421	This study
MJM6139	MJM6102/pM1421	This study
MJM5999	MJM4863 Δ*gacS::erm-bar*	This study
MJM6000	MJM4863 Δ*gacS::bar*	This study
MJM6003	MJM4863 Δ*hahK::erm-bar*	This study
MJM6004	MJM4863 Δ*hahK::bar*	This study
MJM6018	MJM6000 *attTn7::PnrdR-sypF 646-C-erm*	This study
MJM6019	MJM6004 a*ttTn7::PnrdR-sypF 646-C-erm*	This study
E. coli		
Strain	**Genotype**	**Source or Reference**
MJM534	CC118 λ pir/pEVS104	(178)
MJM537	DH5α λ pir	Laboratory stock
MJM542	DH5α λpir/pVSV102	(186)
MJM637	S17-1 λ pir/pUX-BF13	(205)
MJM684	DH5α λ pir/pMarVF1	(122)
MJM658	DH5α λ pir/pEVS107	(185)
MJM1421	DH5α λpir/pM1421	This study
MJM1431	β3914/pMarVF1	This study
MJM1538	DH5α λ pir/pLostfoX	(178, 201)
MJM3478	π3813/pKV496	(177)

MJM4650	DH5α λpir/pKMB021	Ch. 3
MJM4748	DH5α λ pir/pKMB027	This study
MJM4749	DH5α λ pir/pKMB028	This study
MJM4784	DH5α λ pir/pKMB032	This study
MJM4890	DH5α λ pir/pKMB036	This study
MJM4977	DH5α λ pir/pKMB041	This study
MJM5103	DH5α λ pir/pKMB042	This study
MJM5240	DH5α λ pir/pDAT25	This study
MJM6081	DH5α λpir/pACL03	This study
MJM6103	DH5α λpir/pACL04	This study

Table 4.3. Plasmids

Plasmid	**Description**	**Source or Reference**
pVSV102	Constitutive GFP (Kan^r)	(186)
pEVS107	Mini-Tn7 mobilizable vector (Kan^r Erm^r)	(185)
pLostfoX	Arabinose-inducible TfoX for transformation (Cam^r)	(178)
pMarVF1	*V. fischeri* mariner transposon vector (Amp^R, Erm^R)	(122)
pM1421	pAKD701 *sypA'SR5-lacZ*+ (Kan^r)	This study
pKMB021	pKMB018 carrying *sypF* ORF from MJM1100 (Cam^r)	Ch. 3
pKMB027	pEVS107 encoding *PnrdR-SR5 sypF-erm* (Kan^r)	This study
pKMB028	pEVS107 encoding *PnrdR-SR5 sypF AA 481-C-erm* (Kan^r,Erm^r)	This study
pKMB032	pEVS107 encoding PnrdR-SR5 sypF H250Q-erm (Kan^r,Erm^r)	This study
pKMB036	pEVS107 encoding PnrdR-SR5 sypF AA 646-C-erm (Kan^r,Erm^r)	This study
pKMB041	pEVS107 encoding *PnrdR-gfp-erm* (Kan^r,Erm^r)	This study
pKMB042	pKMB021 carrying SR5 *sypF AA 646-C* (Cam^r)	This study
pDAT25	pEVS107 encoding *PnrdR-sypF HPt [MJM1100]-erm* (Kan^r,Erm^r)	This study
pACL04	pEVS107 encoding *PnrdR-hahK-erm* (Kan^r,Erm^r)	This study

pACL03	pEVS107 encoding SR5 *gacS* (Kan^r,Erm^r)	This study

Table 4.4. DNA oligonucleotides for amplification and sequencing

Linker sequences are notated by lowercase letters.

Primer Name	**Sequence (5'-3')**
KMB_036	CCACAATAGCAGAATACAAATTCGCTG
KMB_037	CTCAAAATGACAGTCAGAGTATCGTAGGC
KMB_101	CGTGACTCCAGGAAGCAATACACC
KMB_102	CCACCTATTTGTGAAAACTCAGCCATTGC
KMB_103	ctggcgaagcatatataagaagctcgtctcgtCATTTATTTTGCCTTAATCTTTTATGATTTGGTTAAGGTGAAACG
KMB_104	GCCTCTCCAAATACAATGTGGCGC
KMB_105	GCCAGCCAACCTGTAAACTGATGTATC
KMB_106	gacttgacctggatgtctctacccacaagatcgTGGGGAGCAAAGAACATATAAAAAAGCCTC
KMB_107	GTCAATAATGTTGGTGCCCTACTGGG
KMB_108	CGAAACAAACGACCACGAAACACC
KMB_128	CCCTAACAAACAGAGGTATAAAATTGTTGGG
KMB_129	GTTTACTTTGGCGTGTTTCATTGCTTGATG
KMB_148	CTCTGAGTGGAGCACCAACTTAGTTC
KMB_149	CAGCAAAACCAAATGGCAGAGAAGC
KMB_150	ctggcgaagcatatataagaagctcgtctcgtCATAGTTAGCTCCTTGCAATATTTGCTCTTTTATTG
KMB_151	gacttgacctggatgtctctacccacaagatcgAAACAAGGTTTCTCAAAATAAAAGAATAAACAAAATTCGCTAG
KMB_152	GGCGGAAGGTGCTCTATTGCTAC
KMB_153	GGGTGAGTTGTAGGGTAGCGG
KMB_154	GCGGTCAATACTGTTGCCAATAAAGG
KMB_155	CGTAATCGGTGAGCATCGCCC
KMB_325	CGAGGCATTGATATCATTCAAGATCAAACTGATG
KMB_377	AAGAATAAACAAAATTCGCTAGGTAAAACAGGATGTTAC
KMB_378	AGTTAGCTCCTTGCAATATTTGCTCTTTTATTG
KMB_417	GACGCCACTAACTCCTGACAATGC
KMB_418	CAACACCTTCAATTCCCACCACTG
KMB_419	ctggcgaagcatatataagaagctcgtctcgtCATATAAATTCGATTCGTTTAACTTCCTTGAGGGC
KMB_420	CTGCATGGTGAATTAGATACCCTAAAAAATGG

KMB_421	GGCCTAAACCTGTACCACCGTAAC
KMB_422	gacttgacctggatgtctctacccacaagatcgCTGCTATTGAGCAATACATAACTGAGTATCTAATTAAATAATAAAAC
KMB_423	GCGCGAGGTACTAAATACCCAGATC
KMB_424	GCCTTTACCCCATTGCTAATGAAATGG
KMB_515	CGATGAGTCAaGAAATACGAAC
KMB_516	AAGCTAAAAATCGGCTTTTAC
KMB_517	TTGCCTCCTAGTTTTTTCAATTAGTGTACAAGAAAAG
KMB_518	ttgaaaaaactaggaggcaaATGACATTTCGATTAAAAACGATCATTGGTATCTCTATTATTC
KMB_519	ctagtggccaggtacctcgaAAGTAACATCCTGTTTTACCTAGCGAATTTTG
KMB_521	ttgaaaaaactaggaggcaaAAAATAGAACAACAAAGGGATATTGAACAACAAAAAGC
KMB_522	ttgaaaaaactaggaggcaaATGCTACAAAAAGTATTATTAGTCGAAGACTCCACT
KMB_523	ctagtggccaggtacctcgaTCAAGCAGTAACCATATTTCATCATTCCGATTC
KMB_526	CCTAAAACTGAATTCATAGGTGTTCGTATTTCG
KMB_527	CCTAAAACTGAATTCATAGGTGTTCGTATTTCA
KMB_528	CTTGGGAATGCCTTAACTCGCAATC
KMB_550	GGCCATGAGTAATTGCATAGTCTTAATTCAACATTAC
KMB_565	ttgaaaaaactaggaggcaaATGGCTAGCAAAGGAGAAGAACTCTTC
KMB_566	ctagtggccaggtacctcgaTCAGTTGTACAGTTCATCCATGCCATG
KMB_567	aatattgcaaggagctaactATGGATAATGAGCTATTATTAGTAGATGAACAGATAATTCAGC
KMB_568	agcgaattttgtttattcttTTATTTTGAGAAACCTTGTTTATTTCTTTGTTCAAGTGC
KMB_579	tagagggccctaggcgcgccCGCTTTTTCACCAGGCAATACTCC
KMB_580	ctagtggccaggtacctcgaGAGTTAAAACAGAAAGGGCGTTTTTTAGC
KMB_673	ttgaaaaaactaggaggcaaATGGATGTTGATAAAACAATTGAATTATTGAATCGG
KMB_647	ctagtggccaggtacctcgaCTCTGTTGCAATGGCATTAAGTGGC
binK-F1	GAAATTACCATGGAGCCAACAGCAAGAC
binK-R2	GGCATCATTATGGCAACCATTAAAGACG
binK-FO	CCGTTAATACTGGATTATTCGCTTGAATTTGAACG
HB8	ACAAAATTTTAAGATACTGCACTATCAACACACTCTTAAG
HB9	GGGAGGAAATAAtCTAGAATGCGAGAGTA

	GG
HB41	CGATCTTGTGGGTAGAGACATCCAGGTCAAGTCcagccccgctctagtttgGGAATCAAGTGCATGAGCGCTgaag
HB42	ACGAGACGAGCTTCTTATATATGCTTCGCCAG
RYI072 F	GTAAAACGACGGCCAGT
RYI072 R	AGCGGATAACAATTTCACACAGG
JFB_424	GGCGCGCCTAGGGCCCTC
JFB_425	TCGAGGTACCTGGCCACTAGTAGATCTCTG
pEVS107 F	ACCTATCAAGGTGTACTGCCTTCC
pEVS107 R	GTCGTTAAATGCCCTTTACCTGT
Tn7 site F	TGTTGATGATACCATTGAAGCTAAA
Tn7 site R	TTGCTGTATGTATTTGCTGATGA
M13 F (-41)	CGCCAGGGTTTTCCCAGTCACGAC
MJM-431 R	CTTGGTTTCATCAGCCATCC
MJM-475F	gcGCATGCCCGGGgccttacttggacacgaatca
MJM-476R	gcACTAGTCTAGAttagtccatatcaccttgaactgatagc
DAT_385	TCGAGGTACCTGGCCACTAGTAGATCTCTG
DAT_386	GGCGCGCCTAGGGCCCTC
DAT_387	TAGAGGGCCCTAGGCGCGCCGTAATTGGCATCATAAACATTTTC
DAT_388	CATTATCCATTTGCCTCCTAGTTTTTTCAATTAG
DAT_389	TAGGAGGCAAATGGATAATGAGCTATTATTAGTAG
DAT_390	CTAGTGGCCAGGTACCTCGATTATTTTGAGAAACCTTGTTTATTTC
ARB1	GGCCACGCGTCGACTAGTACNNNNNNNNNNGATAT
ARB2	GGCCACGCGTCGACTAGTAC
MJM-440	TCAACACACTCTTAAGTTTGCTTC
MJM-477	TTCCATAACTTCTTTTACGTTTCC

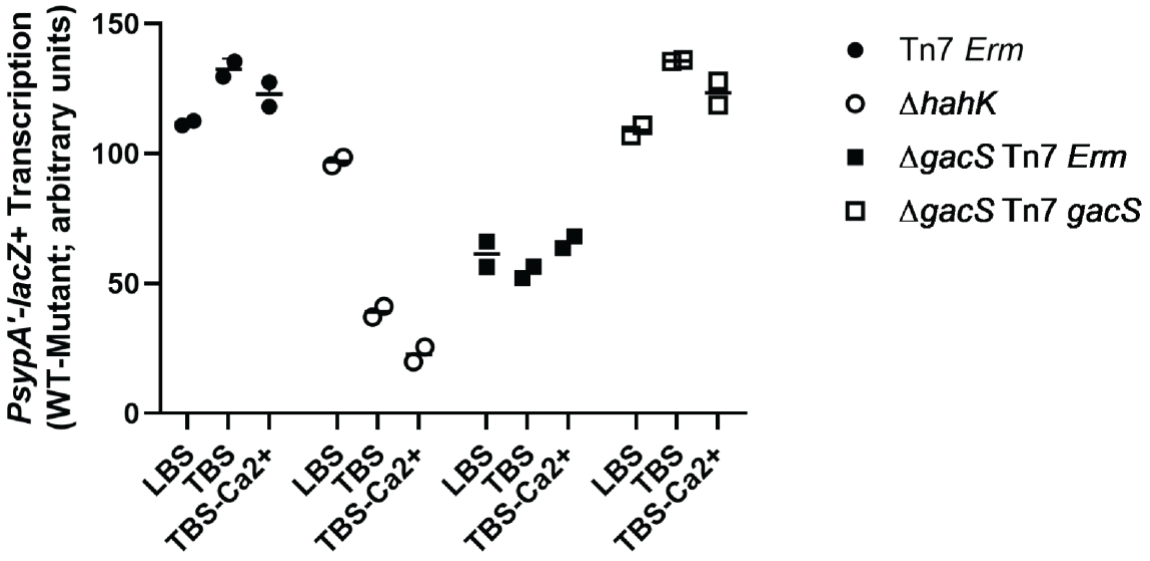


Figure 4.S1. Decrease in *syp* transcription for *hahK* deletion is dependent on media type in SR5 Δ*binK* background.
Strains were grown for 48 h at 25°C on LBS and 20°C on TBS or TBS-calcium media. Graph represents quantification of β-galactosidase activity. Each spot represents the average of 3 technical replicates and error bars represent standard deviation.

ACKNOWLEDGEMENTS

We would like to thank Denise Ludvik for her assistance in the construction of a plasmid used in this study.

Chapter 5: Summative Discussion, Conclusions, and Future Directions

Katherine M. Bultman

Variation in biofilm regulation impacts host colonization

For horizontally acquired bacteria, the production of biofilm is an important part of the lifestyle transition from planktonic living to associating with a host. Biofilm formation is a tightly regulated process that allows for bacteria to group together and stick to surfaces including host epithelial tissue (51, 52). During the initial stages of colonization, *V. fischeri* use biofilm formation to form aggregates within the host mucus before continuing into the pores of the host light organ (15, 19). Regulation of this biofilm formation is dependent on the coordination of a two-component signaling network, however studies on the function of the regulators involved have been limited to a single Hawaiian squid isolate ES114. Presence of a specific regulator, RscS, in ES114 had previously been thought to drive squid-specific biofilm regulation but further analysis of diverse isolates showed that not all contain this gene (131, 133). The overarching goal of my work was to investigate the diversity of biofilm regulation across *V. fischeri* isolates and how it impacts squid colonization.

In Chapter 2, I first looked at the species as a whole and identified three distinct evolutionary groups across *V. fischeri* that all require biofilm production for colonization. Through these studies, I found that two groups were not reliant on the function of RscS for biofilm production and subsequent squid colonization. While the ancestral-like isolates do not contain the gene, analysis of a subset of Hawaiian isolates revealed a point mutation leading to a nonfunctional RscS and an additional point mutation in the downstream response regulator SypE.

Biofilm regulation and squid colonization had previously only been studied in the context of the dependence on RscS. Therefore, in Chapter 3 I asked how strains that lack this regulator initiate biofilm formation to colonize the squid host focusing on the ancestral-like Mediterranean squid isolate SR5. I found that the diversity within hybrid sensor kinase SypF lead to *rscS*-

independent biofilm formation and squid colonization. Further investigation into the functional differences of SypF between SR5 and ES114 showed an increase in biofilm production and *syp* transcription that was dependent on mutations within the REC domain of this regulator. These data led to my current model that in contrast to negative regulation through phosphatase activity by ES114 SypF, SR5 SypF is kinase dominant resulting in increased phosphoryl group flow to SypG and biofilm production (**Fig. 5.1**). Furthermore, this difference in function is driven by the natural diversity within the REC domain.

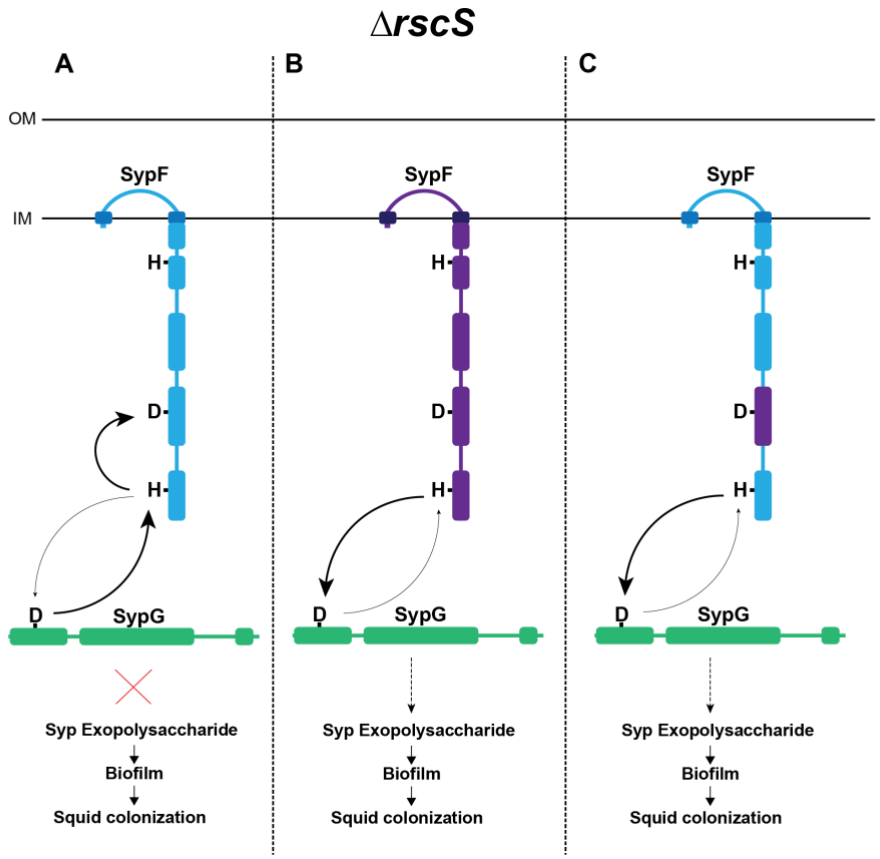


Figure 5.1. Model of SR5 SypF activity as positive regulator of biofilm formation. Arrows indicate phosphoryl group flow with thickness of arrows representing predicted amount. **A)** ES114 SypF is phosphatase dominant. The REC domain removes more phosphoryl groups from the HPt domain leading to less phosphorylation of the downstream response regulator SypG and Syp production. **B)** SR5 SypF is kinase dominant therefore the REC domain is predicted to remove fewer phosphoryl groups from the HPt domain for increased phosphoryl flow to SypG. **C)** The SR5 REC domain confers kinase dominant activity for increased phosphoryl flow to SypG.

My work in Chapter 4 built upon my discoveries in Chapter 3 and found that similar to ES114, only the HPt domain of SypF in SR5 is necessary and sufficient for both biofilm formation and squid colonization. This led to the question of what protein phosphorylates the HPt domain if it is not from SypF itself. A transposon screen for additional regulators identified two hybrid histidine kinases, GacS and HahK, as potential positive regulators within this pathway. Each of these genes when deleted resulted in a decrease in *syp* transcription and biofilm production suggesting a novel role for these proteins in SR5 biofilm regulation.

The Syp regulatory pathway in SR5 involves a multikinase network that must function in concert for proper Syp production during colonization. The conserved function of the negative regulator BinK and its known function in ES114 suggests that this HK is stimulated by the host to facilitate biofilm production though the derepression of *syp* expression in all *V. fischeri* isolates. Full length SypF function is predicted to be strictly for regulation of the phosphorylation state of the HPt domain compared to autokinase activity as only the HPt domain is necessary for biofilm production and colonization. Whether SypF functions as a positive or negative regulator is dependent on the specific variant of the REC domain. Expression of *syp* genes is also dependent on GacS and HahK, however the exact roles of these proteins is still unknown.
I propose two models for their function (**Fig. 5.2**). In the first model I hypothesize that HahK promotes the transcription of the *syp* locus through the HPt domain of SypF. The GacS/GacA system permits the translation of the *syp* genes through a post transcriptional mechanism that is dependent on the phosphorylation of GacA by GacS leading to biofilm formation. Alternatively, I propose in my second model that GacS and HahK work together to phosphorylate the HPt domain of SypF. In this model, I predict that the phosphoryl groups are transferred from the catalytic core of GacS to the REC domain of HahK which phosphorylates the SypF HPt domain. Distinguishing between these two models will be dependent on the role of GacA in biofilm

formation. If GacA is necessary for biofilm formation it suggests the first model over the second, however further experiments would be needed to confirm if the GacS/GacA system in *V. fischeri* functions in a similar manner to other bacteria.

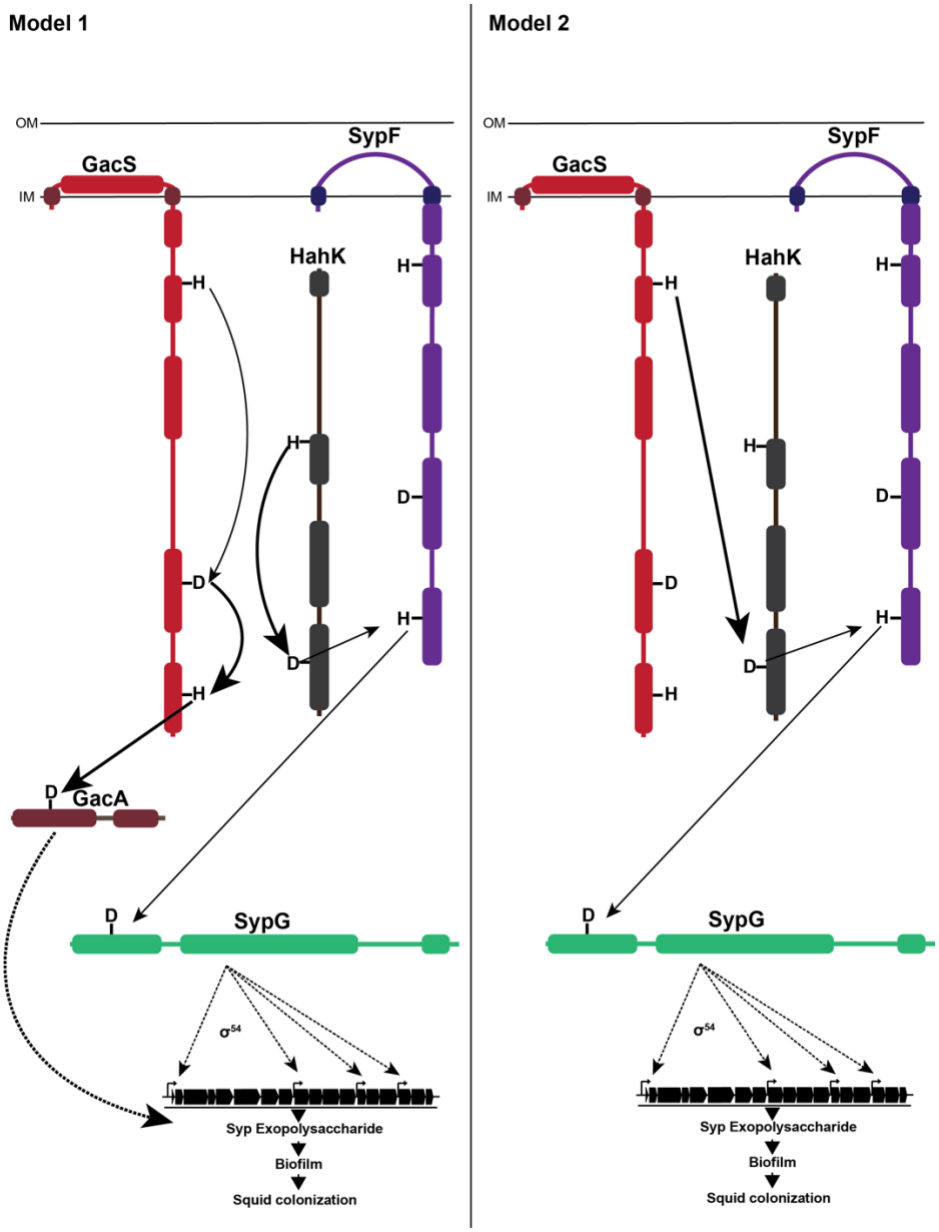


Figure 5.2. Two models of *syp* regulation in SR5 via GacS and HahK.
Solid arrows indicate predicted phosphoryl group flow. Dashed arrows indicate activation through σ^{54} for SypG and unknown mechanisms for GacA. Bolded arrows highlight differences in phosphate flow between the two models. Each model shows one representative phosphotransfer, however others are possible. **Model 1**: The HPt domain of SypF is

phosphorylated by HahK. GacS phosphorylates its cognate response regulator GacA leading to post transcriptional regulation of *syp* expression. **Model 2**: Phosphoryl groups are transferred from GacS to HahK leading to the phosphorylation of the HPt domain of SypF for *syp* expression.

Overall, this work has advanced our understanding of the diversity in host-specific biofilm regulation across *V. fischeri*. The necessity of biofilm production for squid colonization across multiple host species suggests a conserved evolutionary pressure for the maintenance of a functional level of Syp production. Here I identified both conservation of regulator function (BinK, HahK) but also regions of the pathway that may be under more evolutionary pressure (RscS, SypF, SypE) to facilitate squid colonization. Furthermore, these studies provide additional support for small genomic changes resulting in regulatory changes which impact host colonization and specificity.

The *Vibrio*-squid system as a model for studying the diversity of biofilm regulation

The use of biofilm formation is one tactic for associating with host tissue and can play a large role in the colonization process (51, 52, 54, 220). Phenotypic studies have identified variation in biofilm formation across natural isolates that correspond to niche specificity (136, 151, 221, 222). While some bacteria use the secondary messenger c-di-GMP, others use two-component signaling to regulate the switch to a biofilm state (56, 67, 74, 207, 223–225). The natural diversity in biofilm regulation both within and across bacterial species has been shown to alter host specificity (131, 198). However, most of the studies on the specific mechanism for biofilm regulation as part of host colonization have come from pathogenic species (18, 65, 67, 225, 226). It is therefore necessary to study the diverse mechanisms for biofilm formation and its role in host specificity in beneficial symbionts.

Some of the many challenges of identifying the molecular mechanism of biofilm formation as part of host colonization by beneficial symbionts is the complexity of the communities in which they associate and/or a lack of genetic tools. Utilizing the naturally occurring *Vibrio*-squid system, we can take advantage of binary relationship to study the colonization process by *V. fischeri.* Other valuable features of this system are the ability to raise aposymbiotic juveniles for targeted colonization experiments and easy visualization of the bacterium during various stages of colonization (75). In addition, *V. fischeri* is genetically tractable and we have a variety of genetic tools such as transposon libraries, allelic exchange, and methods for rapid gene deletion that allow for both large scale and targeted analysis of gene function (122, 177, 200). Through our ability to make genetic manipulations across isolates coupled with raising aposymbiotic juveniles, we can determine the role of specific genes and variants during host colonization. Each of these genetic techniques allowed me to assess the function of the genomic diversity within the *syp* regulators and identify potential novel regulators within the pathway. In Chapter 3, I relied heavily on allelic exchange to insert the different SypF variants into the native chromosomal locus in both ES114 and SR5 background strains. Although this process takes longer compared to the rapid methods, I was able to maintain the native regulation of this gene in single copy and make mutants without maintaining antibiotic resistance markers or the insertion of in-frame scars. This meant that I assessed only the genomic changes in the ORF between the SypF variants to determine the functional differences in SypF between ES114 and SR5.

The use of multiple different genetic tools has allowed for the *syp* biofilm pathway to be relatively well studied in the Hawaiian squid isolate ES114. Work done primarily by the Visick lab has identified the function of the regulators discussed in Chapter 1. Our knowledge from this previous work provides a point of reference and comparison when assessing the function of these regulators across diverse isolates. Increased biofilm production from the point mutation in

the C-terminal kinase domain of SypE in the Group A strain MB11B1 (**Fig. 2.8**) led us to hypothesize that the resulting protein is phosphatase-dominant compared to ES114 (161). Similarly, based on the mutations in the REC domain of SypF in SR5 being sufficient for increased biofilm formation and squid colonization (**Fig. 3.4**), we proposed that SypF in SR5 is kinase-dominant compared to the phosphatase-dominant function of the ES114 SypF (74). To analyze the effect of the SR5 REC domain on downstream phosphorylation of SypG, I had aimed to create a tagged version of the protein that could be used for Phostag analysis. Unfortunately, the C-terminal tagged SypG that was expressed from the *Tn7* attachment site was unable to complement a *sypG* deletion in colony biofilm assays. However, I could use the successful allelic exchange genetic approach similar to how my SypF variants were constructed to add a tag to SypG in the chromosome.

As phosphorylation of SypG by the hybrid histidine kinase SypF leads to increased *syp* transcription (114), my work took advantage of two *syp* transcriptional reporter plasmids to assess the function of two-component regulators within the pathway. Strains using these reporters could be constructed easily and quickly with triparental mating making this a valuable tool. A fluorescent reporter with *PsypA'* driving the expression of *gfp*+ and a constitutive *mCherry* has been used predominantly for assessing transcription levels in culture, but recent work in *V. fischeri* has expanded the utilization to plate-based colony biofilms and monitoring of *syp* expression during the colonization process (73, 123, 226, 227). I was able to quantify and compare the *syp* transcription levels in strains with SypF variants to demonstrate that the REC domain from SR5 SypF increased both transcription and biofilm formation (**Fig. 3.4**).

The fluorescent reporter could be used in large scale screens, but requires additional steps in the screen process to identify candidates with a change in fluorescence. For example, to assess candidates by plate reader, colonies would first need to be arrayed in 96-well plates and a

specific threshold for what determines elevated or diminished fluorescence would have to be set. Colony picking robots that can read fluorescence levels could be used instead, but requires access to that equipment. For faster identification of transposon mutants with changes in *syp* transcription, I therefore turned to a reporter plasmid with *PsypA'* driving *lacZ*+ expression as this allowed for visual screening of my candidates. Utilizing this method, I manually screened over 65,000 colonies and was able to select candidates with variation in transcription levels compared to the parent strain which was later confirmed with clean deletions (**Fig 4.2, Fig 4.3**). To make my analysis more quantitative, I adapted a method for assessing changes in transcription levels using ImageJ based on methods previously used for Congo Red binding (228). Another benefit to the *lacZ*+ reporter is that it is amenable to undergraduate studies where access to a fluorescent microscope or plate reader is limited or unavailable. This tool could therefore be used in a laboratory course or undergraduate research program for the assessment of biofilm regulation in bacteria.

Access to native squid populations from various geographic locations, including Hawaii, since the initial studies on the *Vibrio*-squid system have provided an opportunity to build a large collection of natural isolates. While most of these isolates are from *E. scolopes* collected off the coast of Oahu, our collection encompasses many of the geographic locations and various squid and fish hosts that *V. fischeri* has been found to form relationships with (128, 131, 135, 229). Not only can we identify the phenotypic diversity of these isolates including squid colonization, but also the genomic diversity. With the decreasing costs of sequencing, more studies in the field have begun to compare the genomic diversity across isolates and how it relates to the host range and evolutionary relationships between isolates (131–133, 135, 137, 230). The number of large scale genomic analyses comparing bacterial species and isolates has been steadily increasing with the lowered cost barrier to whole genome and metagenomic sequencing in addition to the accessibility of bioinformatic software. By combining these genomic studies and

genetic tools, we can test the functional implications of the identified isolate heterogeneity for bacterial behaviors that are important during the colonization process. In Chapter 2, our phylogenetic analysis of *V. fischeri* natural isolates led to the classification of three different evolutionary groups that all demonstrate different biofilm regulatory strategies for squid colonization (**Fig. 2.2**). Furthermore, the comparison of the *syp* loci across representative isolates from each evolutionary group (**Fig. 2.7**) served as a launching point for investigating the diversity in the biofilm regulators that permits *rscS*-independent colonization by the ancestral-like Mediterranean squid isolate SR5.

Strain heterogeneity: One size does not fit all

Across bacterial species, the study of specific molecular mechanisms of colonization behaviors typically utilize the same isolate. For *V. fischeri*, the Hawaiian squid isolate ES114 has been the predominant isolate used for investigating the specific pathways involved in colonization of the squid host. While the study of a single isolate is useful for building foundational knowledge, it provides a narrow view of colonization behaviors and the pathways that regulate them. In recent years, there has been an increase in the studies that have investigated the diversity of bacterial behaviors that impact colonization within bacterial species (130, 135, 151, 231–234). Within *V. fischeri*, strain heterogeneity has been identified in processes including type VI secretion (T6SS), aggregate formation, and dominance during the colonization process (135–137, 230, 235). A select group of *V. fischeri* are highly competitive and don't co-colonize with other isolates, but the exact mechanism for this dominance is unknown (130). While some of these isolates contain a T6SS, others that co-colonize well also contain this machinery (137, 235). The variation across colonization phenotypes highlights the need to look into the mechanisms of colonization and competition among diverse isolates. My work in Chapter 2 identified that out of the three defined evolutionary groups, only Group B isolates are dependent on RscS for squid

colonization (**Fig. 2.5**), although production of Syp is required ubiquitously (**Fig. 2.3, Fig. 2.6**). This means that while the target genes for biofilm production are the same across the species, there is variation in how they are regulated. Comparison of the 18-gene *syp* locus across representative isolates from each evolutionary group identified the two-component regulators SypE and SypF as the least similar (**Fig. 2.7**). Selective pressure on these genes for *syp* regulation amongst diverse isolates may be the basis for proper Syp production during the colonization process. This led me to focus on the functional variation of these regulators within *rscS*-independent colonization.

Genomic changes for host specificity

Genomic variation within bacterial species can lead to the ability to colonize specific hosts (138, 143, 144). Variation of genomic islands or presence and absence of specific genes is a common mechanism for host switching (131, 138, 139, 189, 236)**.** In *V. fischeri*, the acquisition of *rscS* was thought to be the determining factor for squid colonization (131). However, my work demonstrates that while this gene is important for some squid isolates, it is only one of many strategies for biofilm formation and squid colonization. In addition to variation in genomic content, a single domain or residue change within genes that are vital for colonization can alter host specificity (138, 143, 144). In *Salmonella typhimurium*, bovine colonizing isolates contain a FimH variant that has a single amino acid change leading to preferential binding to bovine cells compared to human isolates (143). My work in Chapter 3 shows that the specificity of the REC domain of SypF from SR5 results in a novel regulatory mechanism for biofilm formation and *rscS*-independent squid colonization (**Fig. 3.4**). In addition, a single point mutation within the response regulator SypE in MB11B1 also led to increased biofilm formation (**Fig 2.8**). While this mutation on its own did not rescue colonization of strains lacking *rscS* (**Fig. 2.9**), I hypothesize that variation in the other two-component regulators, such as SypF, in conjunction with this

mutation may hold the key for Group A specific colonization. Sequencing and comparative analyses of genes involved in the colonization process will continue to identify mutations that result in a change in host specificity not only for *V. fischeri*, but symbionts more broadly.

Similarities in two-component signaling for biofilm formation

In addition to regulatory differences identified across evolutionary groups, there are also similarities between isolates. The negative regulator BinK has preserved function across all isolates tested (**Fig. 2.10**). As this regulator functions to turn off *syp* expression and biofilm production (73, 123), it suggests a conserved evolutionary pressure for the repression of biofilm formation during the squid colonization process, either for dispersal from the aggregates or for expression of *syp* only within the host mucus. Similarly, the histidine kinase HahK in ES114 has been shown to positively regulate biofilm production (62). I show a similar function for this regulator in SR5 as deletion of *hahK* results in a decrease in biofilm formation and *syp* transcription (**Fig. 4.3**). The conserved function of these proteins in the regulatory pathway suggest that diversity within this pathway lies predominantly in the two-component regulators in the *syp* locus and their functions.

Although there is strain specificity within the REC domain of SypF, this is not the case for all domains. In both ES114 and SR5, the HPt domain alone of SypF is sufficient for biofilm formation and squid colonization (**Fig. 4.1**) (72). Similar to the REC domain, there are also three amino acid changes in the HPt domain between the isolates. However, this domain shows no strain specificity as SR5 is able to utilize both its own and the HPt domain from ES114 for biofilm formation (**Fig 4.1**). This suggests that the role of the SypF HPt domain between isolates remains the same.

Altogether my results demonstrate the necessity to study the heterogeneity of colonization behaviors within bacterial species as variation within the regulation of these vital behaviors can lead to alteration of host colonization. Future studies focused on the function of biofilm regulators across *V. fischeri* isolates will aid in our understanding of how both gene acquisition and loss of function lead to the diversity in biofilm regulation across the species.

Future Directions

The findings presented in this dissertation provide new insights into the diversity of biofilm regulation by two-component regulators within a beneficial symbiont. My work includes the first studies to investigate the molecular mechanisms of biofilm formation in isolates other than the Hawaiian squid colonizer ES114. These data demonstrate the impact of this natural variation on host colonization and specificity across the *V. fischeri* species. However, questions still remain about how conserved these regulatory strategies are within evolutionary groups. For example, the increased biofilm production by SR5 SypF does not appear to be conserved among other Group C isolates (**Fig. 3.S4**), however it remains to be tested if SypF variants from other Group C squid isolates would be sufficient for RscS-independent colonization. It is possible that multiple different regulatory strategies evolved within this ancestral group to facilitate colonization of different squid hosts. For Group A isolates, current studies have focused on the Hawaiian isolate MB11B1. Although the mutations in the *syp* regulators have been found across multiple isolates in this group, it remains to be determined if other Group A isolates also have these mutations and if they confer the competitive advantage of this group or if there are other mechanisms at play. As biofilm formation is a key determinant of host specificity, understanding the conservation of these mutations across Group A and potential altered function will provide additional insights into the different mechanisms utilized by *V. fischeri* for biofilm formation and subsequent squid colonization. The different regulatory strategies for biofilm formation all lead to

the ability of these *V. fischeri* isolates to colonize the squid host, further emphasizing the importance of biofilm formation for host specificity.

As a large portion of my work focused on biofilm formation for squid colonization by SR5, the next steps for further study of the SR5 Syp regulatory strategy are outlined below.

Experiment 1: Identify the amino acids in the SypF REC domain that are sufficient for SR5-specific *rscS*-independent colonization. The REC domain of SR5 SypF is both necessary and sufficient for increased *syp* transcription and biofilm formation in an *rscS*-independent manner, but the specific mechanism of action remains unknown. Three amino acid changes were identified in the REC domain between ES114 and SR5, however no single point mutation was sufficient for these phenotypes. Comparison with other natural isolates suggest that the two conserved amino acid changes across isolates compared to ES114 (S566T and Q620E) are also not sufficient. I hypothesize that there must be at least two amino acid changes, one of which being the P575L change, is driving the change in function. While these amino acid changes are not in the active site of the REC domain, they may impact the interaction with the other domains. Therefore pinpointing the exact changes generating the functional differences could direct further study in the specific mechanism for promoting biofilm formation. To identify which amino acid changes are necessary, every combination of double mutations within the REC domain would be constructed and placed at the native site in an ES114 Δ*rscS* Δ*binK* strain. Colony biofilm assays would be performed as described in Chapter 3. Combinations of mutations that result in increased biofilm formation would be investigated further to determine their role in *syp* transcription and squid colonization.

Experiment 2: Determine if GacS functions within the *syp* regulatory pathway. In Chapter 4, I demonstrated that both GacS and HahK impact biofilm formation in SR5, even in the

presence of the SypF HPt domain alone. As HPt domains function as part of a phosphorelay, phosphoryl groups are transferred from a REC domain to the HPt and then transferred to the downstream response regulator (33). Therefore, if the HPt domain of SypF is not acting in a phosphorelay with the domains in SypF, then another protein is likely donating phosphates to this domain. HahK has been shown to function through the HPt domain of SypF in ES114 (62), however it is yet to be determined if GacS functions directly within the *syp* regulatory network or in an indirect manner. Previous studies have utilized either *sypG*-overexpression or a phosphomimetic allele of *sypG* to determine a direct role for the negative regulator BinK within the *syp* pathway (73, 123). A wrinkled phenotype in a colony biofilm assays of an SR5 Δ*binK* Δ*sypE* Δ*gacS* strain with a phosphomimetic SypG would demonstrate that GacS functions at or above the level SypG within the biofilm regulatory pathway (120).

Experiment 3: Identify the role of GacA in biofilm formation and squid colonization. The GacS/GacA two-component system has been well-studied in *Pseudomonas* species and other Gram-negative bacteria (70, 210, 214, 215, 237). In this system, GacS phosphorylates GacA leading to an upregulation of two sRNAs that inhibit a translational repressor. While *V. fischeri gacA* mutants have reduced luminescence, motility and the ability to initiate colonization, the role of this response regulator in biofilm formation is not understood (216). It is also unclear if GacA is the cognate response regulator for GacS in *V. fischeri*. If a gacA deletion strain shows a defect in biofilm formation and squid colonization, it suggests that GacA functions to promote biofilm formation for colonization. To test if GacS and GacA form a regulatory pair, a phosphomimetic *gacA* allele would be placed at the native site in a *gacS* deletion strain and both biofilm formation and colonization would again be assessed. Restoration of biofilm formation and colonization in this background would suggest a cognate GacS/GacA pair and that these proteins are likely to participate in post transcriptional biofilm regulation similar to the role of the SypE/SypA pair (161).

Conclusions

My work has built foundational knowledge on the diversity of biofilm formation across the *V. fischeri* species and how this diversity impacts host colonization. The *Vibrio*-squid system provides a unique opportunity to investigate the diversity in the molecular mechanisms of biofilm formation due to the large collection of natural isolates and the ability to not only manipulate the genome but also easily assess host colonization. Here I identified three evolutionary groups within *V. fischeri* that all demonstrate different biofilm regulatory strategies. I further analyzed the specific strategy of the ancestral-like isolate SR5 and identified a key protein domain that leads to host specificity and additional positive regulators whose function within the biofilm regulatory pathway are largely under-studied. This work suggests that the previously characterized *syp* regulatory pathway in ES114 represents only one of many different mechanisms that targets the same 18-gene locus for biofilm production. As this biofilm is necessary during the squid colonization process, the specific regulatory mechanisms found across isolates determine the ability to colonize the squid host and may reflect differences in host adaptations. Collectively, this work provides a foundation for the expansion of our understanding of biofilm regulation in *V. fischeri*, and lays the groundwork for research into the other specific mechanisms across isolates. Future work should focus on the expansion of these studies into the other evolutionary groups across *V. fischeri* and investigate the impact of these differing strategies on host specificity.

Appendix A: Draft genome sequences of type VI secretion system-encoding *Vibrio fischeri* strains FQ-A001 and ES401

Katherine M. Bultman, Andrew G. Cecere, Tim Miyashiro, Alecia N. Septer, Mark J. Mandel

A version of this appendix is published as:

Bultman KM, Cecere AG, Miyashiro T, Septer AN, Mandel MJ. Draft Genome Sequences of Type VI Secretion System-Encoding *Vibrio fischeri* Strains FQ-A001 and ES401. Microbiol Resour Announc. 2019 May 16;8(20):e00385-19. doi: 10.1128/MRA.00385-19.

Author contributions: KMB performed the genome assembly, TM and MJM provided sequencing data, KMB and MJM wrote the manuscript.

ABSTRACT

The type VI secretion system (T6SS) facilitates lethal bacterial-bacterial competition through direct contact. Comparative genomics has facilitated study of these systems in *Euprymna scolopes* squid host-colonizing *Vibrio fischeri*. Here, we report the draft genome sequences of FQ-A001 and ES401, two lethal *V. fischeri* strains that encode the T6SS.

MAIN TEXT

The bacterial-encoded Type VI Secretion System (T6SS) is a membrane-embedded syringe-like structure that delivers effectors to bacterial, and in some cases, eukaryotic cells (238). A role for T6SS has been demonstrated in mediating strain separation of incompatible *Vibrio fischeri* strains during colonization of *Euprymna scolopes* juvenile squid (235). *V. fischeri* FQ-A001 (226) and ES401 (76) are strains that were isolated decades apart, yet both exhibit the lethal phenotype, defined as the ability to kill *V. fischeri* ES114 in direct competition (235). Experimental analysis of FQ-A001 has demonstrated that a specific T6SS locus (T6SS2) is responsible for the lethal phenotype observed (235). Draft genomes for FQ-A001 and ES401 were obtained as follows, both of which have an intact T6SS2 locus.

Strain FQ-A001 was isolated from an *E. scolopes* adult female (21 mm mantle length) from Kaneohe Bay in 2015, and was cultivated in the laboratory on LBS medium (226). For Illumina MiSeq sequencing (2 x 250 bp; Penn State Genomics Core Facility), genomic DNA was isolated using the MasterPure DNA Purification Kit (Epicentre, Madison, WI) and the library was constructed using the TruSeq DNA PCR-Free kit (Illumina, San Diego, CA). For Pacbio RSII SMRT sequencing (UNC Chapel Hill High Throughput Sequencing Facility), genomic DNA was isolated by phenol-chloroform extraction, and the sequencing library was prepared using the SMRTbell library prep kit (10 kb size selection). Approximately 9.4 x 10^8 bp (MiSeq) and 1.1 x 10^9 bp (RSII) of sequence data was obtained, yielding approximately 470-fold coverage of the

FQ-A001 genome. *De novo* assembly of the FQ-A001 genome was conducted using SPAdes version 3.13.0 (with --careful -k 127) (239), yielding four contigs > 1,000 bp. Contigs were reordered using Mauve Contig Mover with ES114 as a reference (204, 240, 241). For exploratory analysis, the genome was annotated with Prokka 1.13.3 (242), and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP; March 4, 2019) was used to annotate the contigs for deposition to GenBank (243).

Strain ES401 was isolated from an *E. scolopes* juvenile (3.5 mm mantle length) from Maunalua Bay in 1990, and can be cultivated in the laboratory on LBS medium (76). For Illumina MiSeq sequencing (2 x 300 bp; University of Wisconsin Biotechnology Center), genomic DNA was isolated using the DNeasy Blood & Tissue Kit, Gram-negative bacterial protocol (Qiagen USA, Germantown, MD) and the library was constructed using the TruSeq Nano kit (Illumina, San Diego, CA). Pacbio sequencing was conducted as for FQ-A001 above. Approximately 2.7 x 10^8 bp (MiSeq) and 1.8 x 10^9 bp (RSII) of sequence data was obtained, yielding approximately 450-fold coverage of the ES401 genome. Assembly, annotation, and GenBank deposition of the resulting 6 contigs > 1,000 bp were conducted as for FQ-A001 above with the following changes in software parameters: SPAdes (--careful -k 125); and PGAP (April 2, 2019).

For quality control, the original Illumina and Pacbio basecalling resulted in 0 sequences flagged as poor quality by FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Upon submission to PGAP, an additional fifth contig in the FQ-A001 sequence contained only PhiX phage sequence and was excluded from the assembly and further analysis. PGAP additionally identified Illumina adapter sequence at one location in the ES401 genome; we conducted Sanger sequencing of this region, which clarified the correct sequence. The correct sequence was resubmitted to PGAP and the surrounding genomic context was not affected.

Analysis using progressiveMauve (snapshot 2015_02_25, macOS) revealed genomes that are collinear with MJ11 and the presence of the lethal strain-specific T6SS on chromosome II in both FQ-A001 and ES401 (Table 1) (131, 235, 244). Default parameters for software were used except where additional parameters are noted above.

DATA AVAILABILITY

The data for this paper are available at NCBI as: FQ-A001 genome SJSX00000000 (version SJSX01000000 for this report); FQ-A001 reads SRR8647323-SRR8647324; ES401 genome SRJG00000000 (version SRJG01000000); ES401 reads SRR8708068-SRR8708069.

ACKNOWLEDGMENTS

We gratefully acknowledge Ned Ruby for providing strain ES401 and Ella Rotman for assistance with sample preparation. This work was funded by NIH grants R35GM119627 (to M.J.M.), R01GM129133 (T.M.), and R21AI117262 (M.J.M.); and NSF grant IOS-1757297 (M.J.M.). Trainee support was provided on NIH grant T32GM008349 (K.M.B.). This work used the Jetstream service via the Extreme Science and Engineering Discovery Environment (XSEDE) through allocation MCB180165 (245, 246).

TABLES

Table A.1. *Vibrio fischeri* genomes described in this report.

Strain	Year isolated	Genome size (bp)	No. of contigs	N_{50}	Total no. of genes	G+C content (%)	GenBank accession no.	Genes corresponding to MJ11 T6SS2

FQ-A001	2015	4,286,900	4	2,417,280	4,015	38.3	SJSX00000000	VFFQA001_15465 - VFFQA001_15635
ES401	1990	4,282,653	6	1,351,222	4,011	38.3	SRJG00000000	VFES401_15680 - VFES401_15865

REFERENCES

1. Oulhen N, Schulz BJ, Carrier TJ. 2016. English translation of Heinrich Anton de Bary's 1878 speech, "Die Erscheinung der Symbiose" ("De la symbiose"). Symbiosis 69:131–139.

2. Dimijian GG. 2000. Evolving together: the biology of symbiosis, part 1. Proc 13:217–226.

3. McFall-Ngai MJ. 2002. Unseen forces: the influence of bacteria on animal development. Dev Biol 242:1–14.

4. Douglas AE. 2014. Symbiosis as a general principle in eukaryotic evolution. Cold Spring Harb Perspect Biol 6.

5. Moran NA. 2006. Symbiosis. Current Biology https://doi.org/10.1016/j.cub.2006.09.019.

6. Ruby E, Henderson B, McFall-Ngai M. 2004. We Get By with a Little Help from Our (Little) Friends. Science 303:1305–1307.

7. Xu J, Gordon JI. 2003. Honor thy symbionts. Proc Natl Acad Sci U S A 100:10452–10459.

8. Brooks JF 2nd, Hooper LV. 2020. Interactions among microbes, the immune system, and the circadian clock. Semin Immunopathol 42:697–708.

9. Hooper LV, Littman DR, Macpherson AJ. 2012. Interactions between the microbiota and the immune system. Science 336:1268–1273.

10. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. 2007. Development of the human infant intestinal microbiota. PLoS Biol 5:e177.

11. Cho I, Blaser MJ. 2012. The human microbiome: at the interface of health and disease. Nat Rev Genet 13:260–270.

12. Bright M, Bulgheresi S. 2010. A complex journey: transmission of microbial symbionts. Nat Rev Microbiol 8:218–230.

13. Russell SL. 2019. Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: a systematic review and meta-analysis. FEMS Microbiol Lett 366.

14. Gage DJ. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing *Rhizobia* during nodulation of temperate legumes. Microbiol Mol Biol Rev 68:280–300.

15. Nyholm SV, Stabb EV, Ruby EG, McFall-Ngai MJ. 2000. Establishment of an animal-bacterial association: recruiting symbiotic vibrios from the environment. Proc Natl Acad Sci U S A 97:10231–10235.

16. Obeng N, Bansept F, Sieber M, Traulsen A, Schulenburg H. 2021. Evolution of Microbiota-host associations: The microbe's perspective. Trends Microbiol 29:779–787.

17. Rinaudi LV, Giordano W. 2010. Bacterial Biofilms: Role in Rhizobium–Legume Symbiosis, p. 325–335. *In* Khan, MS, Musarrat, J, Zaidi, A (eds.), Microbes for Legume Improvement. Springer Vienna, Vienna.

18. Vestby LK, Grønseth T, Simm R, Nesse LL. 2020. Bacterial Biofilm and its Role in the Pathogenesis of Disease. Antibiotics (Basel) 9.

19. Yip ES, Grublesky BT, Hussa EA, Visick KL. 2005. A novel, conserved cluster of genes promotes symbiotic colonization and sigma-dependent biofilm formation by *Vibrio fischeri*.

Mol Microbiol 57:1485–1498.

20. Nyholm SV, McFall-Ngai MJ. 2004. The winnowing: establishing the squid-vibrio symbiosis. Nat Rev Microbiol 2:632–642.

21. Douglas AE. 2008. Conflict, cheats and the persistence of symbioses. New Phytol 177:849–858.

22. van Rhijn P, Vanderleyden J. 1995. The Rhizobium-plant symbiosis. Microbiol Rev 59:124–142.

23. Mandel MJ, Schaefer AL, Brennan CA, Heath-Heckman EAC, Deloney-Marino CR, McFall-Ngai MJ, Ruby EG. 2012. Squid-derived chitin oligosaccharides are a chemotactic signal during colonization by *Vibrio fischeri*. Appl Environ Microbiol 78:4620–4626.

24. Masson-Boivin C, Sachs JL. 2018. Symbiotic nitrogen fixation by *Rhizobia*—the roots of a success story. Curr Opin Plant Biol 44:7–15.

25. Essock-Burns T, Bongrand C, Goldman WE, Ruby EG, McFall-Ngai MJ. 2020. Interactions of Symbiotic Partners Drive the Development of a Complex Biogeography in the Squid-Vibrio Symbiosis. MBio 11.

26. Fronk DC, Sachs JL. 2022. Symbiotic organs: the nexus of host–microbe evolution. Trends Ecol Evol 37:599–610.

27. Afzal I, Shinwari ZK, Sikandar S, Shahzad S. 2019. Plant beneficial endophytic bacteria: Mechanisms, diversity, host range and genetic determinants. Microbiol Res 221:36–49.

28. Ishii E, Eguchi Y. 2021. Diversity in Sensing and Signaling of Bacterial Sensor Histidine Kinases. Biomolecules 11.

29. Stewart RC. 2010. Protein histidine kinases: assembly of active sites and their regulation in signaling pathways. Curr Opin Microbiol 13:133–141.

30. Singh D, Gupta P, Singla-Pareek SL, Siddique KHM, Pareek A. 2021. The Journey from Two-Step to Multi-Step Phosphorelay Signaling Systems. Curr Genomics 22:59–74.

31. Zschiedrich CP, Keidel V, Szurmant H. 2016. Molecular Mechanisms of Two-Component Signal Transduction. J Mol Biol 428:3752–3775.

32. Hoch JA. 2000. Two-component and phosphorelay signal transduction. Curr Opin Microbiol 3:165–170.

33. Zhang W, Shi L. 2005. Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. Microbiology 151:2159–2173.

34. Wuichet K, Cantwell BJ, Zhulin IB. 2010. Evolution and phyletic distribution of two-component signal transduction systems. Curr Opin Microbiol 13:219–225.

35. Salazar ME, Laub MT. 2015. Temporal and evolutionary dynamics of two-component signaling pathways. Curr Opin Microbiol 24:7–14.

36. Capra EJ, Laub MT. 2012. Evolution of Two-Component Signal Transduction Systems. Annual Review of Microbiology https://doi.org/10.1146/annurev-micro-092611-150039.

37. Bourret RB, Silversmith RE. 2010. Two-component signal transduction. Curr Opin Microbiol 13:113–115.

38. Perraud AL, Weiss V, Gross R. 1999. Signaling pathways in two-component phosphorelay systems. Trends Microbiol 7:115–120.

39. Gao R, Stock AM. 2009. Biological insights from structures of two-component proteins. Annu Rev Microbiol 63:133–154.

40. Laub MT, Goulian M. 2007. Specificity in two-component signal transduction pathways. Annu Rev Genet 41:121–145.

41. Podgornaia AI, Laub MT. 2013. Determinants of specificity in two-component signal transduction. Curr Opin Microbiol 16:156–162.

42. Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. 2008. Rewiring the specificity of two-component signal transduction systems. Cell 133:1043–1054.

43. Alm E, Huang K, Arkin A. 2006. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. PLoS Comput Biol 2:e143.

44. Nocedal I, Laub MT. 2022. Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity. Elife 11.

45. Capra EJ, Perchuk BS, Ashenberg O, Seid CA, Snow HR, Skerker JM, Laub MT. 2012. Spatial tethering of kinases to their substrates relaxes evolutionary constraints on specificity. Mol Microbiol 86:1393–1403.

46. Ortet P, Fochesato S, Bitbol A-F, Whitworth DE, Lalaouna D, Santaella C, Heulin T, Achouak W, Barakat M. 2021. Evolutionary history expands the range of signaling interactions in hybrid multikinase networks. Sci Rep 11:11763.

47. Francis VI, Porter SL. 2019. Multikinase Networks: Two-Component Signaling Networks Integrating Multiple Stimuli. Annu Rev Microbiol 73:199–223.

48. Jiang M, Shao W, Perego M, Hoch JA. 2000. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. Mol Microbiol 38:535–542.

49. Henke JM, Bassler BL. 2004. Three parallel quorum-sensing systems regulate gene expression in *Vibrio harveyi*. J Bacteriol 186:6902–6914.

50. Bhuwan M, Lee H-J, Peng H-L, Chang H-Y. 2012. Histidine-containing phosphotransfer protein-B (HptB) regulates swarming motility through partner-switching system in *Pseudomonas aeruginosa* PAO1 strain. J Biol Chem 287:1903–1914.

51. Donlan RM. 2002. Biofilms: microbial life on surfaces. Emerg Infect Dis 8:881–890.

52. Branda SS, Vik S, Friedman L, Kolter R. 2005. Biofilms: the matrix revisited. Trends Microbiol 13:20–26.

53. Flemming H-C, Wingender J, Szewzyk U, Steinberg P, Rice SA, Kjelleberg S. 2016. Biofilms: an emergent form of bacterial life. Nat Rev Microbiol 14:563–575.

54. Shibata S, Yip ES, Quirke KP, Ondrey JM, Visick KL. 2012. Roles of the structural symbiosis polysaccharide (*syp*) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in *Vibrio fischeri*. J Bacteriol 194:6736–6747.

55. Hengge R, Gründling A, Jenal U, Ryan R, Yildiz F. 2016. Bacterial Signal Transduction by Cyclic Di-GMP and Other Nucleotide Second Messengers. J Bacteriol 198:15–26.

56. Valentini M, Filloux A. 2016. Biofilms and Cyclic di-GMP (c-di-GMP) Signaling: Lessons from *Pseudomonas aeruginosa* and Other Bacteria. J Biol Chem 291:12547–12555.

57. Yildiz FH, Visick KL. 2009. Vibrio biofilms: so much the same yet so different. Trends Microbiol 17:109–118.

58. Chatterjee S, Samal B, Singh P, Pradhan BB, Verma RK. 2020. Transition of a solitary to a biofilm community life style in bacteria: a survival strategy with division of labour. Int J Dev Biol 64:259–265.

59. Liu X, Zhang K, Liu Y, Zou D, Wang D, Xie Z. 2020. Effects of Calcium and Signal Sensing Systems on *Azorhizobium caulinodans* Biofilm Formation and Host Colonization. Front Microbiol 11:563367.

60. Chodur DM, Coulter P, Isaacs J, Pu M, Fernandez N, Waters CM, Rowe-Magnus DA. 2018. Environmental Calcium Initiates a Feed-Forward Signaling Circuit That Regulates Biofilm Formation and Rugosity in *Vibrio vulnificus*. MBio 9.

61. Broder UN, Jaeger T, Jenal U. 2016. LadS is a calcium-responsive kinase that induces acute-to-chronic virulence switch in *Pseudomonas aeruginosa*. Nat Microbiol 2:16184.

62. Tischler AH, Lie L, Thompson CM, Visick KL. 2018. Discovery of calcium as a biofilm-promoting signal for *Vibrio fischeri* reveals new phenotypes and underlying regulatory complexity. J Bacteriol https://doi.org/10.1128/JB.00016-18.

63. Whitehead NA, Barnard AM, Slater H, Simpson NJ, Salmond GP. 2001. Quorum-sensing in Gram-negative bacteria. FEMS Microbiol Rev 25:365–404.

64. Bjelland AM, Sørum H, Tegegne DA, Winther-Larsen HC, Willassen NP, Hansen H. 2012. LitR of *Vibrio salmonicida* is a salinity-sensitive quorum-sensing regulator of phenotypes involved in host interactions and virulence. Infect Immun 80:1681–1689.

65. Hansen H, Bjelland AM, Ronessen M, Robertsen E, Willassen NP. 2014. LitR is a repressor of *syp* genes and has a temperature-sensitive regulatory effect on biofilm formation and colony morphology in *Vibrio (Aliivibrio) salmonicida*. Appl Environ Microbiol 80:5530–5541.

66. Liu C, Sun D, Zhu J, Liu W. 2018. Two-Component Signal Transduction Systems: A Major Strategy for Connecting Input Stimuli to Biofilm Formation. Front Microbiol 9:3279.

67. Kitts G, Rogers A, Teschler JK, Park JH, Trebino MA, Chaudry I, Erill I, Yildiz FH. 2023. The Rvv two-component regulatory system regulates biofilm formation and colonization in *Vibrio cholerae*. PLoS Pathog 19:e1011415.

68. Schaefers MM, Wang BX, Boisvert NM, Martini SJ, Bonney SL, Marshall CW, Laub MT, Cooper VS, Priebe GP. 2021. Evolution towards Virulence in a *Burkholderia* Two-Component System. MBio 12:e0182321.

69. Wang BX, Cady KC, Oyarce GC, Ribbeck K, Laub MT. 2021. Two-Component Signaling Systems Regulate Diverse Virulence-Associated Traits in *Pseudomonas aeruginosa*. Appl Environ Microbiol 87.

70. Chambonnier G, Roux L, Redelberger D, Fadel F, Filloux A, Sivaneson M, de Bentzmann S, Bordi C. 2016. The Hybrid Histidine Kinase LadS Forms a Multicomponent Signal Transduction System with the GacS/GacA Two-Component System in *Pseudomonas aeruginosa*. PLoS Genet 12:e1006032.

71. Valentini M, Gonzalez D, Mavridou DA, Filloux A. 2018. Lifestyle transitions and adaptive pathogenesis of *Pseudomonas aeruginosa*. Curr Opin Microbiol 41:15–20.

72. Norsworthy AN, Visick KL. 2015. Signaling between two interacting sensor kinases promotes biofilms and colonization by a bacterial symbiont. Mol Microbiol 96:233–248.

73. Brooks JF 2nd, Mandel MJ. 2016. The Histidine Kinase BinK Is a Negative Regulator of Biofilm Formation and Squid Colonization. J Bacteriol 198:2596–2607.

74. Thompson CM, Marsden AE, Tischler AH, Koo J, Visick KL. 2018. *Vibrio fischeri* Biofilm Formation Prevented by a Trio of Regulators. Appl Environ Microbiol 84.

75. Mandel MJ, Dunn AK. 2016. Impact and Influence of the Natural Vibrio-Squid Symbiosis in Understanding Bacterial-Animal Interactions. Front Microbiol 7:1982.

76. Ruby EG, Lee KH. 1998. The *Vibrio fischeri-Euprymna scolopes* Light Organ Association: Current Ecological Paradigms. Appl Environ Microbiol 64:805–812.

77. Lee KH, Ruby EG. 1992. Detection of the Light Organ Symbiont, *Vibrio fischeri* in Hawaiian Seawater by using *lux* Gene Probes. Appl Environ Microbiol 58:942–947.

78. Lee KH, Ruby EG. 1994. Effect of the Squid Host on the Abundance and Distribution of Symbiotic *Vibrio fischeri* in Nature. Appl Environ Microbiol 60:1565–1571.

79. McFall-Ngai M. 2008. Hawaiian bobtail squid. Curr Biol 18:R1043–4.

80. Nawroth JC, Guo H, Koch E, Heath-Heckman EAC, Hermanson JC, Ruby EG, Dabiri JO, Kanso E, McFall-Ngai M. 2017. Motile cilia create fluid-mechanical microhabitats for the active recruitment of the host microbiome. Proc Natl Acad Sci U S A 114:9510–9516.

81. McFall-Ngai MJ, Ruby EG. 1991. Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. Science 254:1491–1494.

82. Nyholm SV, Deplancke B, Gaskins HR, Apicella MA, McFall-Ngai MJ. 2002. Roles of *Vibrio fischeri* and nonsymbiotic bacteria in the dynamics of mucus secretion during symbiont colonization of the *Euprymna scolopes* light organ. Appl Environ Microbiol 68:5113–5122.

83. Nyholm SV, McFall-Ngai MJ. 2003. Dominance of *Vibrio fischeri* in secreted mucus outside the light organ of *Euprymna scolopes*: the first site of symbiont specificity. Appl Environ Microbiol 69:3932–3937.

84. McAnulty SJ, Nyholm SV. 2016. The Role of Hemocytes in the Hawaiian Bobtail Squid, *Euprymna scolopes*: A Model Organism for Studying Beneficial Host-Microbe Interactions. Front Microbiol 7:2013.

85. Yip ES, Geszvain K, DeLoney-Marino CR, Visick KL. 2006. The symbiosis regulator *rscS* controls the *syp* gene locus, biofilm formation and symbiotic aggregation by *Vibrio fischeri*. Mol Microbiol 62:1586–1600.

86. Wolfe AJ, Millikan DS, Campbell JM, Visick KL. 2004. *Vibrio fischeri* σ^{54} Controls Motility, Biofilm Formation, Luminescence, and Colonization. Appl Environ Microbiol 70:2520–2524.

87. Davidson SK, Koropatnick TA, Kossmehl R, Sycuro L, McFall-Ngai MJ. 2004. NO means “yes” in the squid-vibrio symbiosis: nitric oxide (NO) during the initial stages of a beneficial association. Cell Microbiol 6:1139–1151.

88. Wang Y, Dunn AK, Wilneff J, McFall-Ngai MJ, Spiro S, Ruby EG. 2010. *Vibrio fischeri* flavohaemoglobin protects against nitric oxide during initiation of the squid-Vibrio symbiosis. Mol Microbiol 78:903–915.

89. Dunn AK, Karr EA, Wang Y, Batton AR, Ruby EG, Stabb EV. 2010. The alternative oxidase (AOX) gene in *Vibrio fischeri* is controlled by NsrR and upregulated in response to nitric oxide. Mol Microbiol 77:44–55.

90. Thompson CM, Tischler AH, Tarnowski DA, Mandel MJ, Visick KL. 2019. Nitric oxide inhibits biofilm formation by *Vibrio fischeri* via the nitric oxide sensor HnoX. Mol Microbiol 111:187–203.

91. Graf J, Dunlap PV, Ruby EG. 1994. Effect of transposon-induced motility mutations on colonization of the host light organ by *Vibrio fischeri*. J Bacteriol 176:6986–6991.

92. Millikan DS, Ruby EG. 2004. *Vibrio fischeri* flagellin A is essential for normal motility and for symbiotic competence during initial squid light organ colonization. J Bacteriol 186:4315–4325.

93. Millikan DS, Ruby EG. 2002. Alterations in *Vibrio fischeri* motility correlate with a delay in symbiosis initiation and are associated with additional symbiotic colonization defects. Appl Environ Microbiol 68:2519–2528.

94. Deloney-Marino CR, Visick KL. 2012. Role for *cheR* of *Vibrio fischeri* in the Vibrio-squid symbiosis. Can J Microbiol 58:29–38.

95. McCann J, Stabb EV, Millikan DS, Ruby EG. 2003. Population dynamics of *Vibrio fischeri* during infection of *Euprymna scolopes*. Appl Environ Microbiol 69:5928–5934.

96. Heath-Heckman EAC, McFall-Ngai MJ. 2011. The occurrence of chitin in the hemocytes of invertebrates. Zoology 114:191–198.

97. Ruby EG, Asato LM. 1993. Growth and flagellation of *Vibrio fischeri* during initiation of the sepiolid squid light organ symbiosis. Arch Microbiol 159:160–167.

98. Schwartzman JA, Koch E, Heath-Heckman EAC, Zhou L, Kremer N, McFall-Ngai MJ, Ruby EG. 2015. The chemistry of negotiation: rhythmic, glycan-driven acidification in a symbiotic conversation. Proc Natl Acad Sci U S A 112:566–571.

99. Montgomery MK, McFall-Ngai M. 1994. Bacterial symbionts induce host organ morphogenesis during early postembryonic development of the squid *Euprymna scolopes*. Development 120:1719–1729.

100. Koropatnick TA, Engle JT, Apicella MA, Stabb EV, Goldman WE, McFall-Ngai MJ. 2004. Microbial factor-mediated development in a host-bacterial mutualism. Science 306:1186–1188.

101. Boettcher KJ, Ruby EG. 1990. Depressed light emission by symbiotic *Vibrio fischeri* of the sepiolid squid *Euprymna scolopes*. J Bacteriol 172:3701–3706.

102. Nyholm SV, McFall-Ngai MJ. 1998. Sampling the light-organ microenvironment of *Euprymna scolopes:* description of a population of host cells in association with the bacterial symbiont *Vibrio fischeri*. Biol Bull 195:89–97.

103. Wier AM, Nyholm SV, Mandel MJ, Massengo-Tiassé RP, Schaefer AL, Koroleva I, Splinter-Bondurant S, Brown B, Manzella L, Snir E, Almabrazi H, Scheetz TE, Bonaldo M de F, Casavant TL, Soares MB, Cronan JE, Reed JL, Ruby EG, McFall-Ngai MJ. 2010. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proc Natl Acad Sci U S A 107:2259–2264.

104. Thompson LR, Nikolakakis K, Pan S, Reed J, Knight R, Ruby EG. 2017. Transcriptional characterization of *Vibrio fischeri* during colonization of juvenile *Euprymna scolopes*. Environ Microbiol 19:1845–1856.

105. Pan M, Schwartzman JA, Dunn AK, Lu Z, Ruby EG. 2015. A Single Host-Derived Glycan Impacts Key Regulatory Nodes of Symbiont Metabolism in a Coevolved Mutualism. MBio 6:e00811.

106. Bourgois JJ, Sluse FE, Baguet F, Mallefet J. 2001. Kinetics of light emission and oxygen consumption by bioluminescent bacteria. J Bioenerg Biomembr 33:353–363.

107. Ruby EG, McFall-Ngai MJ. 1999. Oxygen-utilizing reactions and symbiotic colonization of the squid light organ by *Vibrio fischeri*. Trends Microbiol 7:414–420.

108. Crookes WJ, Ding L-L, Huang QL, Kimbell JR, Horwitz J, McFall-Ngai MJ. 2004. Reflectins: the unusual proteins of squid reflective tissues. Science 303:235–238.

109. Ruby EG. 1996. Lessons from a cooperative, bacterial-animal association: The *Vibrio fischeri–Euprymna scolopes* Light Organ Symbiosis. Annu Rev Microbiol 50:591–624.

110. Ruby EG, McFall-Ngai MJ. 1992. A squid that glows in the night: development of an animal-bacterial mutualism. J Bacteriol 174:4865–4870.

111. Visick KL, Skoufos LM. 2001. Two-component sensor required for normal symbiotic colonization of *Euprymna scolopes* by *Vibrio fischeri*. J Bacteriol 183:835–842.

112. Visick KL. 2009. An intricate network of regulators controls biofilm formation and colonization by *Vibrio fischeri*. Mol Microbiol 74:782–789.

113. Geszvain K, Visick KL. 2008. The hybrid sensor kinase RscS integrates positive and negative signals to modulate biofilm formation in *Vibrio fischeri*. J Bacteriol 190:4437–4446.

114. Darnell CL, Hussa EA, Visick KL. 2008. The putative hybrid sensor kinase SypF coordinates biofilm formation in *Vibrio fischeri* by acting upstream of two response regulators, SypG and VpsR. J Bacteriol 190:4941–4950.

115. Hussa EA, O'Shea TM, Darnell CL, Ruby EG, Visick KL. 2007. Two-component response regulators of *Vibrio fischeri*: identification, mutagenesis, and characterization. J Bacteriol 189:5825–5838.

116. Ray VA, Eddy JL, Hussa EA, Misale M, Visick KL. 2013. The syp enhancer sequence plays a key role in transcriptional activation by the σ^{54}-dependent response regulator SypG and in biofilm formation and host colonization by *Vibrio fischeri*. J Bacteriol 195:5402–5412.

117. Hussa EA, Darnell CL, Visick KL. 2008. RscS functions upstream of SypG to control the *syp* locus and biofilm formation in *Vibrio fischeri*. J Bacteriol 190:4576–4583.

118. Morris AR, Visick KL. 2010. Control of biofilm formation and colonization in *Vibrio fischeri*: a role for partner switching? Environ Microbiol 12:2051–2059.

119. Morris AR, Darnell CL, Visick KL. 2011. Inactivation of a novel response regulator is necessary for biofilm formation and host colonization by *Vibrio fischeri*. Mol Microbiol 82:114–130.

120. Morris AR, Visick KL. 2013. Inhibition of SypG-induced biofilms and host colonization by the negative regulator SypE in *Vibrio fischeri*. PLoS One 8:e60076.

121. Dufour A, Haldenwang WG. 1994. Interactions between a *Bacillus subtilis* anti-sigma factor (RsbW) and its antagonist (RsbV). J Bacteriol 176:1813–1820.

122. Brooks JF 2nd, Gyllborg MC, Cronin DC, Quillin SJ, Mallama CA, Foxall R, Whistler C, Goodman AL, Mandel MJ. 2014. Global discovery of colonization determinants in the squid symbiont *Vibrio fischeri*. Proc Natl Acad Sci U S A 111:17284–17289.

123. Ludvik DA, Bultman KM, Mandel MJ. 2021. Hybrid Histidine Kinase BinK Represses *Vibrio fischeri* Biofilm Signaling at Multiple Developmental Stages. J Bacteriol 203:e0015521.

124. Iyer LM, Anantharaman V, Aravind L. 2003. Ancient conserved domains shared by animal soluble guanylyl cyclases and bacterial signaling proteins. BMC Genomics 4:5.

125. Wang Y, Dufour YS, Carlson HK, Donohue TJ, Marletta MA, Ruby EG. 2010. H-NOX-mediated nitric oxide sensing modulates symbiotic colonization by *Vibrio fischeri*. Proc Natl Acad Sci U S A 107:8375–8380.

126. Lee KH, Ruby EG. 1994. Competition between *Vibrio fischeri* strains during initiation and maintenance of a light organ symbiosis. J Bacteriol 176:1985–1991.

127. Nishiguchi MK, Ruby EG, McFall-Ngai MJ. 1998. Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in Sepiolid squid-vibrio symbioses. Appl Environ Microbiol 64:3209–3213.

128. Fidopiastis PM, von Boletzky S, Ruby EG. 1998. A new niche for *Vibrio logei*, the predominant light organ symbiont of squids in the genus *Sepiola*. J Bacteriol 180:59–64.

129. Wollenberg MS, Ruby EG. 2009. Population structure of *Vibrio fischeri* within the light

organs of *Euprymna scolopes* squid from Two Oahu (Hawaii) populations. Appl Environ Microbiol 75:193–202.

130. Bongrand C, Ruby EG. 2019. Achieving a multi-strain symbiosis: strain behavior and infection dynamics. ISME J 13:698–706.

131. Mandel MJ, Wollenberg MS, Stabb EV, Visick KL, Ruby EG. 2009. A single regulatory gene is sufficient to alter bacterial host range. Nature 458:215–218.

132. Rotman ER, Bultman KM, Brooks JF 2nd, Gyllborg MC, Burgos HL, Wollenberg MS, Mandel MJ. 2019. Natural Strain Variation Reveals Diverse Biofilm Regulation in Squid-Colonizing *Vibrio fischeri*. J Bacteriol 201.

133. Gyllborg MC, Sahl JW, Cronin DC 3rd, Rasko DA, Mandel MJ. 2012. Draft genome sequence of *Vibrio fischeri* SR5, a strain isolated from the light organ of the Mediterranean squid Sepiola robusta. J Bacteriol 194:1639.

134. Wollenberg MS, Ruby EG. 2012. Phylogeny and fitness of *Vibrio fischeri* from the light organs of *Euprymna scolopes* in two Oahu, Hawaii populations. ISME J 6:352–362.

135. Bongrand C, Koch EJ, Moriano-Gutierrez S, Cordero OX, McFall-Ngai M, Polz MF, Ruby EG. 2016. A genomic comparison of 13 symbiotic *Vibrio fischeri* isolates from the perspective of their host source and colonization behavior. ISME J 10:2907–2917.

136. Koehler S, Gaedeke R, Thompson C, Bongrand C, Visick KL, Ruby E, McFall-Ngai M. 2018. The model squid-vibrio symbiosis provides a window into the impact of strain- and species-level differences during the initial stages of symbiont engagement. Environ Microbiol doi:10.1111/1462–2920.14392.

137. Bongrand C, Moriano-Gutierrez S, Arevalo P, McFall-Ngai M, Visick KL, Polz M, Ruby EG. 2020. Using Colonization Assays and Comparative Genomics To Discover Symbiosis Behaviors and Factors in *Vibrio fischeri*. MBio 11.

138. Kirzinger MWB, Stavrinides J. 2012. Host specificity determinants as a genetic continuum. Trends Microbiol 20:88–93.

139. Spaink HP, Wijffelman CA, Pees E, Okker RJH, Lugtenberg BJJ. 1987. *Rhizobium* nodulation gene *nodD* as a determinant of host specificity. Nature 328:337–340.

140. Gerlach RG, Hensel M. 2007. *Salmonella* pathogenicity islands in host specificity, host pathogen-interactions and antibiotics resistance of *Salmonella enterica*. Berl Munch Tierarztl Wochenschr 120:317–327.

141. Cowles CE, Goodrich-Blair H. 2008. The *Xenorhabdus nematophila nilABC* genes confer the ability of Xenorhabdus spp. to colonize *Steinernema carpocapsae* nematodes. J Bacteriol 190:4121–4128.

142. Killiny N, Almeida RPP. 2011. Gene regulation mediates host specificity of a bacterial pathogen. Environ Microbiol Rep 3:791–797.

143. Yue M, Han X, De Masi L, Zhu C, Ma X, Zhang J, Wu R, Schmieder R, Kaushik RS, Fraser GP, Zhao S, McDermott PF, Weill F-X, Mainil JG, Arze C, Fricke WF, Edwards RA, Brisson D, Zhang NR, Rankin SC, Schifferli DM. 2015. Allelic variation contributes to bacterial host

specificity. Nat Commun 6:8754.

144. Eswarappa SM, Janice J, Nagarajan AG, Balasundaram SV, Karnam G, Dixit NM, Chakravortty D. 2008. Differentially evolved genes of *Salmonella* pathogenicity islands: insights into the mechanism of host specificity in *Salmonella*. PLoS One 3:e3829.

145. Kwong WK, Moran NA. 2015. Evolution of host specialization in gut microbes: the bee gut as a model. Gut Microbes 6:214–220.

146. Vayssier-Taussat M, Le Rhun D, Deng HK, Biville F, Cescau S, Danchin A, Marignac G, Lenaour E, Boulouis HJ, Mavris M, Arnaud L, Yang H, Wang J, Quebatte M, Engel P, Saenz H, Dehio C. 2010. The Trw type IV secretion system of *Bartonella* mediates host-specific adhesion to erythrocytes. PLoS Pathog 6:e1000946.

147. Jalan N, Aritua V, Kumar D, Yu F, Jones JB, Graham JH, Setubal JC, Wang N. 2011. Comparative genomic analysis of *Xanthomonas axonopodis pv. citrumelo* F1, which causes citrus bacterial spot disease, and related strains provides insights into virulence and host specificity. J Bacteriol 193:6342–6357.

148. Rakov AV, Mastriani E, Liu S-L, Schifferli DM. 2019. Association of *Salmonella* virulence factor alleles with intestinal and invasive serovars. BMC Genomics 20:429.

149. Frese SA, Mackenzie DA, Peterson DA, Schmaltz R, Fangman T, Zhou Y, Zhang C, Benson AK, Cody LA, Mulholland F, Juge N, Walter J. 2013. Molecular characterization of host-specific biofilm formation in a vertebrate gut symbiont. PLoS Genet 9:e1004057.

150. Schilcher K, Horswill AR. 2020. Staphylococcal Biofilm Development: Structure, Regulation, and Treatment Strategies. Microbiol Mol Biol Rev 84.

151. Tovi N, Frenk S, Hadar Y, Minz D. 2018. Host Specificity and Spatial Distribution Preference of Three *Pseudomonas* Isolates. Front Microbiol 9:3263.

152. Guo Y, Rowe-Magnus DA. 2011. Overlapping and unique contributions of two conserved polysaccharide loci in governing distinct survival phenotypes in *Vibrio vulnificus*. Environ Microbiol 13:2888–2990.

153. Long SR. 1996. *Rhizobium* symbiosis: *nod* factors in perspective. Plant Cell 8:1885–1898.

154. Roche P, Maillet F, Plazanet C, Debellé F, Ferro M, Truchet G, Promé JC, Dénarié J. 1996. The common *nodABC* genes of *Rhizobium meliloti* are host-range determinants. Proc Natl Acad Sci U S A 93:15305–15310.

155. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community variation in human body habitats across space and time. Science 326:1694–1697.

156. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA. 2009. Topographical and temporal diversity of the human skin microbiome. Science 324:1190–1192.

157. Mark Welch JL, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG. 2016. Biogeography of a human oral microbiome at the micron scale. Proc Natl Acad Sci U S A 113:E791–E800.

158. Ruby EG. 2008. Symbiotic conversations are revealed under genetic interrogation. Nat Rev Microbiol 6:752–762.

159. Visick KL, Ruby EG. 2006. *Vibrio fischeri* and its host: it takes two to tango. Curr Opin Microbiol 9:632–638.

160. Jones BW, Nishiguchi MK. 2004. Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca: Cephalopoda). Mar Biol 144:1151–1155.

161. Morris AR, Visick KL. 2013. The response regulator SypE controls biofilm formation and colonization through phosphorylation of the *syp*-encoded regulator SypA in *Vibrio fischeri*. Mol Microbiol 87:509–525.

162. Pankey MS, Foxall RL, Ster IM, Perry LA, Schuster BM, Donner RA, Coyle M, Cooper VS, Whistler CA. 2017. Host-selected mutations converging on a global regulator drive an adaptive leap towards symbiosis in bacteria. Elife 6:e24414.

163. Nyholm SV, Nishiguchi MK. 2008. The evolutionary ecology of a sepiolid squid-vibrio association: from cell to environment. Vie Milieu Paris 58:175–184.

164. Mandel MJ. 2010. Models and approaches to dissect host-symbiont specificity. Trends Microbiol 18:504–511.

165. Geszvain K, Visick KL. 2008. Multiple factors contribute to keeping levels of the symbiosis regulator RscS low. FEMS Microbiol Lett 285:33–39.

166. Elliott KT, DiRita VJ. 2008. Characterization of CetA and CetB, a bipartite energy taxis system in *Campylobacter jejuni*. Mol Microbiol 69:1091–1103.

167. Antonov I, Coakley A, Atkins JF, Baranov PV, Borodovsky M. 2013. Identification of the nature of reading frame transitions observed in prokaryotic genomes. Nucleic Acids Res 41:6514–6530.

168. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

169. Rusch DB, Rowe-Magnus DA. 2017. Complete Genome Sequence of the Pathogenic *Vibrio vulnificus* Type Strain ATCC 27562. Genome Announc 5:e00907–17.

170. Ray VA, Visick KL. 2012. LuxU connects quorum sensing to biofilm formation in *Vibrio fischeri*. Mol Microbiol 86:954–970.

171. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre J-C, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang W-S, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Verméglio A, Médigue C, Sadowsky M. 2007. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. Science 316:1307–1312.

172. Bonaldi K, Gargani D, Prin Y, Fardoux J, Gully D, Nouwen N, Goormachtig S, Giraud E. 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic *Bradyrhizobium* Sp. strain ORS285: the Nod-dependent versus the Nod-independent

symbiotic interaction. Mol Plant Microbe Interact 24:1359–1371.

173. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772.

174. Swofford DL. 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods) 4th edn. Sinauer.

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

176. Ronquist F, van der Mark P, Huelsenbeck JP. 2009. Bayesian phylogenetic analysis using MrBayes, 2nd edn. Cambridge University Press.

177. Visick KL, Hodge-Hanson KM, Tischler AH, Bennett AK, Mastrodomenico V. 2018. Tools for Rapid Genetic Engineering of *Vibrio fischeri*. Appl Environ Microbiol 84.

178. Pollack-Berti A, Wollenberg MS, Ruby EG. 2010. Natural transformation of *Vibrio fischeri* requires *tfoX* and *tfoY*. Environ Microbiol 12:2302–2311.

179. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008.

180. Naughton LM, Mandel MJ. 2012. Colonization of *Euprymna scolopes* squid by *Vibrio fischeri*. J Vis Exp e3758.

181. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–85.

182. Lee K-H. 1994. Ecology of *Vibrio fisheri*: the light organ symbiont of the Hawaiian sepiolid squid *Euprymna scolopes*. University of Southern California.

183. Boettcher KJ, Ruby EG. 1994. Occurrence of plasmid DNA in the sepiolid squid symbiont *Vibrio fischeri*. Curr Microbiol 29:279–286.

184. Nishiguchi MK, Nair VS. 2003. Evolution of symbiosis in the *Vibrionaceae*: a combined approach using molecules and physiology. Int J Syst Evol Microbiol 53:2019–2026.

185. Stabb EV, Ruby EG. 2002. RP4-based plasmids for conjugation between *Escherichia coli* and members of the *Vibrionaceae*. Methods Enzymol 358:413–426.

186. Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb EV. 2006. New rfp- and pES213-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and *lux* expression in situ. Appl Environ Microbiol 72:802–810.

187. McFall-Ngai M, Nyholm SV, Castillo MG. 2010. The role of the immune system in the initiation and persistence of the *Euprymna scolopes–Vibrio fischeri* symbiosis. Semin Immunol 22:48–53.

188. Hooper LV. 2009. Do symbiotic bacteria subvert host immunity? Nat Rev Microbiol 7:367–

374.

189. Matuszewska M, Murray GGR, Harrison EM, Holmes MA, Weinert LA. 2020. The Evolutionary Genomics of Host Specificity in *Staphylococcus aureus*. Trends Microbiol 28:465–477.

190. Ganesan Ramya, Wierz Jürgen C., Kaltenpoth Martin, Flórez Laura V. 2022. How It All Begins: Bacterial Factors Mediating the Colonization of Invertebrate Hosts by Beneficial Symbionts. Microbiol Mol Biol Rev 86:e00126–21.

191. Powell JE, Leonard SP, Kwong WK, Engel P, Moran NA. 2016. Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. Proc Natl Acad Sci U S A 113:13887–13892.

192. Ciche TA, Kim K-S, Kaufmann-Daszczuk B, Nguyen KCQ, Hall DH. 2008. Cell Invasion and Matricide during *Photorhabdus luminescen*s Transmission by *Heterorhabditis bacteriophora* Nematodes. Appl Environ Microbiol 74:2275–2287.

193. Kim JK, Kwon JY, Kim SK, Han SH, Won YJ, Lee JH, Kim C-H, Fukatsu T, Lee BL. 2014. Purine biosynthesis, biofilm formation, and persistence of an insect-microbe gut symbiosis. Appl Environ Microbiol 80:4374–4382.

194. Murphy MP, Caraher E. 2015. Residence in biofilms allows *Burkholderia cepacia* complex (Bcc) bacteria to evade the antimicrobial activities of neutrophil-like dHL60 cells. Pathog Dis 73:ftv069.

195. Koch EJ, Miyashiro T, McFall-Ngai MJ, Ruby EG. 2014. Features governing symbiont persistence in the squid-vibrio association. Mol Ecol 23:1624–1634.

196. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. 2023. InterPro in 2022. Nucleic Acids Res 51:D418–D427.

197. Huynh TN, Lin H-Y, Noriega CE, Lin AV, Stewart V. 2015. Cross Talk Inhibition Nullified by a Receiver Domain Missense Substitution. J Bacteriol 197:3294–3306.

198. Sun Y-C, Hinnebusch BJ, Darby C. 2008. Experimental evidence for negative selection in the evolution of a *Yersinia pestis* pseudogene. Proc Natl Acad Sci U S A 105:8097–8101.

199. Le Roux F, Binesse J, Saulnier D, Mazel D. 2007. Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene *vsm* by use of a novel counterselectable suicide vector. Appl Environ Microbiol 73:777–784.

200. Burgos HL, Burgos EF, Steinberger AJ, Suen G, Mandel MJ. 2020. Multiplexed Competition in a Synthetic Squid Light Organ Microbiome Using Barcode-Tagged Gene Deletions. mSystems 5.

201. Brooks JF 2nd, Gyllborg MC, Kocher AA, Markey LEH, Mandel MJ. 2015. TfoX-based genetic mapping identifies *Vibrio fischeri* strain-level differences and reveals a common lineage of laboratory strains. J Bacteriol 197:1065–1074.

202. Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066.

203. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

204. Mandel MJ, Stabb EV, Ruby EG. 2008. Comparative genomics-based investigation of resequencing targets in *Vibrio fischeri*: focus on point miscalls and artefactual expansions. BMC Genomics 9:138.

205. Bao Y, Lies DP, Fu H, Roberts GP. 1991. An improved Tn7-based system for the single-copy insertion of cloned genes into chromosomes of Gram-negative bacteria. Gene 109:167–168.

206. Guo Y, Rowe-Magnus DA. 2010. Identification of a c-di-GMP-regulated polysaccharide locus governing stress resistance and biofilm and rugose colony formation in *Vibrio vulnificus*. Infect Immun 78:1390–1402.

207. Mangalea MR, Borlee BR. 2022. The NarX-NarL two-component system regulates biofilm formation, natural product biosynthesis, and host-associated survival in *Burkholderia pseudomallei*. Sci Rep 12:203.

208. Appleby JL, Parkinson JS, Bourret RB. 1996. Signal transduction via the multi-step phosphorelay: not necessarily a road less traveled. Cell 86:845–848.

209. Janiak-Spens F, Sparling DP, West AH. 2000. Novel role for an HPt domain in stabilizing the phosphorylated state of a response regulator domain. J Bacteriol 182:6673–6678.

210. Tomenius H, Pernestig A-K, Jonas K, Georgellis D, Möllby R, Normark S, Melefors O. 2006. The *Escherichia coli* BarA-UvrY two-component system is a virulence determinant in the urinary tract. BMC Microbiol 6:27.

211. Lapouge K, Schubert M, Allain FH-T, Haas D. 2008. Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. Mol Microbiol 67:241–253.

212. Workentine ML, Chang L, Ceri H, Turner RJ. 2009. The GacS-GacA two-component regulatory system of *Pseudomonas fluorescens*: a bacterial two-hybrid analysis. FEMS Microbiol Lett 292:50–56.

213. Septer AN, Bose JL, Lipzen A, Martin J, Whistler C, Stabb EV. 2015. Bright luminescence of *Vibrio fischeri* aconitase mutants reveals a connection between citrate and the Gac/Csr regulatory system. Mol Microbiol 95:283–296.

214. Goodman AL, Kulasekara B, Rietsch A, Boyd D, Smith RS, Lory S. 2004. A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in *Pseudomonas aeruginosa*. Dev Cell 7:745–754.

215. Ryan Kaler KM, Nix JC, Schubot FD. 2021. RetS inhibits *Pseudomonas aeruginosa* biofilm formation by disrupting the canonical histidine kinase dimerization interface of GacS. J Biol Chem 297:101193.

216. Whistler CA, Ruby EG. 2003. GacA regulates symbiotic colonization traits of *Vibrio fischeri*

and facilitates a beneficial association with an animal host. J Bacteriol 185:7202–7212.

217. Whistler CA, Koropatnick TA, Pollack A, McFall-Ngai MJ, Ruby EG. 2007. The GacA global regulator of *Vibrio fischeri* is required for normal host tissue responses that limit subsequent bacterial colonization. Cell Microbiol 9:766–778.

218. Goodman AL, Wu M, Gordon JI. 2011. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. Nat Protoc 6:1969–1980.

219. Mark Mandel EB. 2015. pyinseq. https://github.com/mjmlab/pyinseq. Retrieved 6 December 2019.

220. Powell E, Ratnayeke N, Moran NA. 2016. Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. Mol Ecol 25:4461–4471.

221. Sanchez CJ Jr, Mende K, Beckius ML, Akers KS, Romano DR, Wenke JC, Murray CK. 2013. Biofilm formation by clinical isolates and the implications in chronic infections. BMC Infect Dis 13:47.

222. Fernández-Gómez P, Figueredo A, López M, González-Raurich M, Prieto M, Alvarez-Ordóñez A. 2021. Heterogeneity in biofilm formation and identification of biomarkers of strong biofilm formation among field isolates of *Pseudomonas* spp. Food Res Int 148:110618.

223. Poulin MB, Kuperman LL. 2021. Regulation of Biofilm Exopolysaccharide Production by Cyclic Di-Guanosine Monophosphate. Front Microbiol 12:730980.

224. Tagua VG, Molina-Henares MA, Travieso ML, Nisa-Martínez R, Quesada JM, Espinosa-Urgel M, Ramos-González MI. 2022. C-di-GMP and biofilm are regulated in *Pseudomonas putida* by the CfcA/CfcR two-component system in response to salts. Environ Microbiol 24:158–178.

225. Mikkelsen H, Sivaneson M, Filloux A. 2011. Key two-component regulatory systems that control biofilm formation in *Pseudomonas aeruginosa*. Environ Microbiol 13:1666–1681.

226. Sun Y, LaSota ED, Cecere AG, LaPenna KB, Larios-Valencia J, Wollenberg MS, Miyashiro T. 2016. Intraspecific Competition Impacts *Vibrio fischeri* Strain Diversity during Initial Colonization of the Squid Light Organ. Appl Environ Microbiol 82:3082–3091.

227. Miyashiro T, Wollenberg MS, Cao X, Oehlert D, Ruby EG. 2010. A single *qrr* gene is necessary and sufficient for LuxO-mediated regulation in *Vibrio fischeri*. Mol Microbiol 77:1556–1567.

228. Isenberg RY, Christensen DG, Visick KL, Mandel MJ. 2022. High Levels of Cyclic Diguanylate Interfere with Beneficial Bacterial Colonization. MBio 13:e0167122.

229. Ast JC, Urbanczyk H, Dunlap PV. 2009. Multi-gene analysis reveals previously unrecognized phylogenetic diversity in *Aliivibrio*. Syst Appl Microbiol 32:379–386.

230. Guckes KR, Cecere AG, Wasilko NP, Williams AL, Bultman KM, Mandel MJ, Miyashiro T. 2019. Incompatibility of *Vibrio fischeri* Strains during Symbiosis Establishment Depends on Two Functionally Redundant *hcp* Genes. J Bacteriol 201.

231. Kuehnast T, Cakar F, Weinhäupl T, Pilz A, Selak S, Schmidt MA, Rüter C, Schild S. 2018. Comparative analyses of biofilm formation among different *Cutibacterium acnes* isolates. Int J Med Microbiol 308:1027–1035.

232. Wilske B, Barbour AG, Bergström S, Burman N, Restrepo BI, Rosa PA, Schwan T, Soutschek E, Wallich R. 1992. Antigenic variation and strain heterogeneity in *Borrelia* spp. Res Microbiol 143:583–596.

233. Proutière A, Drebes Dörr NC, Bader L, Stutzmann S, Metzger LC, Isaac S, Chiaruttini N, Blokesch M. 2023. Sporadic type VI secretion in seventh pandemic *Vibrio cholerae*. Microbiology 169.

234. Spencer BL, Job AM, Robertson CM, Hameed ZA, Serchejian C, Wiafe-Kwakye CS, Mendonça JC, Apolonio MA, Nagao PE, Neely MN, Korotkova N, Korotkov KV, Patras KA, Doran KS. 2023. Heterogeneity of the group B streptococcal type VII secretion system and influence on colonization of the female genital tract. Mol Microbiol 120:258–275.

235. Speare L, Cecere AG, Guckes KR, Smith S, Wollenberg MS, Mandel MJ, Miyashiro T, Septer AN. 2018. Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. Proc Natl Acad Sci U S A 115:E8528–E8537.

236. Lavín JL, Binnewies TT, Pisabarro AG, Ussery DW, García-Lobo JM, Oguiza JA. 2010. Differences in two-component signal transduction proteins among the genus *Brucella*: implications for host preference and pathogenesis. Vet Microbiol 144:478–483.

237. Heeb S, Haas D. 2001. Regulatory roles of the GacS/GacA two-component system in plant-associated and other Gram-negative bacteria. Mol Plant Microbe Interact 14:1351–1363.

238. Coulthurst S. 2019. The Type VI secretion system: a versatile bacterial weapon. Microbiology 165:503–515.

239. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.

240. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics 25:2071–2073.

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

242. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

243. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624.

244. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

245. Stewart CA, Turner G, Vaughn M, Gaffney NI, Cockerill TM, Foster I, Hancock D, Merchant N, Skidmore E, Stanzione D, Taylor J, Tuecke S. 2015. Jetstream: a self-provisioned, scalable science and engineering cloud environment, p. 1–8. *In* Proceedings of the 2015 XSEDE Conference on Scientific Advancements Enabled by Enhanced Cyberinfrastructure - XSEDE '15. ACM Press, New York, New York, USA.

246. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, Roskies R, Scott JR, Wilkins-Diehr N. 2014. XSEDE: Accelerating Scientific Discovery. Computing in Science Engineering 16:62–74.