Self-recognition signaling and ecological variation in the bacterial pathogen Proteus mirabilis

Citation

Chittor, Achala. 2021. Self-recognition signaling and ecological variation in the bacterial pathogen Proteus mirabilis. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37368228

Terms of Use

Share Your Story

Accessibility

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Molecular and Cellular Biology

have examined a dissertation entitled

Self-recognition signaling and ecological variation in the bacterial pathogen Proteus mirabilis

presented by Achala Chittor

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Colleen M. Cavanaugh (Apr 21, 2021 17:00 EDT)

Typed name: Prof. Colleen Cavanaugh

Signature _____

Richard Losick (Apr 30, 2021 16:15 EDT)

Typed name: Prof. Richard Losick

Signature _____

Elena Kramer

Typed name: Prof. Elena Kramer

Signature _____

Typed name: Prof.

Signature _____

Typed name: Prof.

Date: April 21, 2021

Self-recognition signaling and ecological variation in the bacterial pathogen

Proteus mirabilis

A dissertation presented

by

Achala Chittor

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biology

Harvard University

Cambridge, Massachusetts

April 2021

Dissertation Advisor: Professor Karine A. Gibbs   By: Achala Chittor

Self-recognition signaling and ecological variation in the bacterial pathogen *Proteus mirabilis*

Abstract

Considering molecular interactions within the cell and with the environment can advance our understanding of opportunistic pathogens within a framework of ecology and evolution. For the opportunistic pathogen *Proteus mirabilis*, self recognition impacts swarming, a collective motility associated with virulence. During swarming, neighboring cells inject an identity protein that induces a stress response and iteratively excludes non-kin cells. The interactions of Ids signaling within the cell and the broader ecology of collective behavior were previously unknown. During my dissertation research, I showed that a conserved serine transporter, SdaC, is the dominant serine transporter during swarming and essential for self recognition. Analysis of SdaC single-residue variants and homologs revealed that the open conformation of SdaC is necessary and sufficient for self recognition and is conserved in SdaC from *Escherichia coli* but not the similar serine transporter YhaO in *P. mirabilis*. We hypothesize that a specific molecular interface is likely exposed when SdaC is in an open conformation that facilitates inner membrane integration of the injected identity protein. Notably, the majority of sequenced *P. mirabilis* isolates share SdaC and self-recognition factors, whether found in infections or from animal or environmental reservoirs, raising questions about the phylogenetic history of *P. mirabilis*. To address this gap, we performed whole-genome sequencing and behavioral characterizations of asymptomatic animal-associated *P. mirabilis* strains.

We found that genes for virulence and swarming are in the core genome, while identity proteins are in the non-core. Further, I showed that poor swarmers were attenuated in virulence and, in one case, this phenotype could be traced to a nonsense mutation in the flagellar gene *fliF*. Our approach, using diverse isolates and natural variation, uncovered unexpected ecotype similarity, opening a path to identify new molecular mechanisms of virulence.

Table of Contents

List of Figures

"Don't let anyone rob you of your imagination, your creativity, or your curiosity.

It's your place in the world; it's your life.

Go on and do all you can with it, and make it the life you want to live"

Mae C. Jemison, scientist and first African American woman in space

This work is dedicated to all those who are driven by creativity and curiosity

to ask tough questions and continue learning in the face of adversity.

Page intentionally left blank

Chapter 1

Molecular mechanisms of self recognition: identity determinants and beyond

<u>Self recognition is a fundamental behavior for collectivity</u>

Organisms across the tree of life discern between self and non-self. However, the benefits and challenges broadly differ between organisms that are multicellular versus unicellular. In multicellular individuals like humans, self recognition allows immune cells to protect against foreign entities and informs the spatial distribution of neurons. By contrast, for many colonial organisms, cooperation between individuals can be beneficial and carries a risk of compromising autonomy. Clonal individuals (i.e., siblings) can join together to form a larger colony that covers more space, has increased access to resources, and is more environmentally robust. An ongoing threat is that distantly related organisms can become incorporated and potentially exploit the colony, hijacking the cooperative benefits for their own survival. To possibly protect against this potential conflict, self-recognition systems emerged, likely independently in bacteria and eukaryotes. This essay explores how these various self-recognition systems appear to have converged on similar operational rules.

A model proposes that self recognition comes down to three steps: detection, recognition, and discrimination or assortment [Figure 1.1, (Grice & Degnan, 2015)]. First, an individual detects another nearby individual. Second, communication of identity allows for recognition of the individual as self or non-self. Third, the individual takes action to merge with self and separate from non-self leading to positive assortment and enrichment of kin (Grice & Degnan, 2015). For many systems, it is still unclear through what molecular mechanism(s) the recognition in step two could cause assortment in step three. By studying self recognition in a bacterial model in comparison to
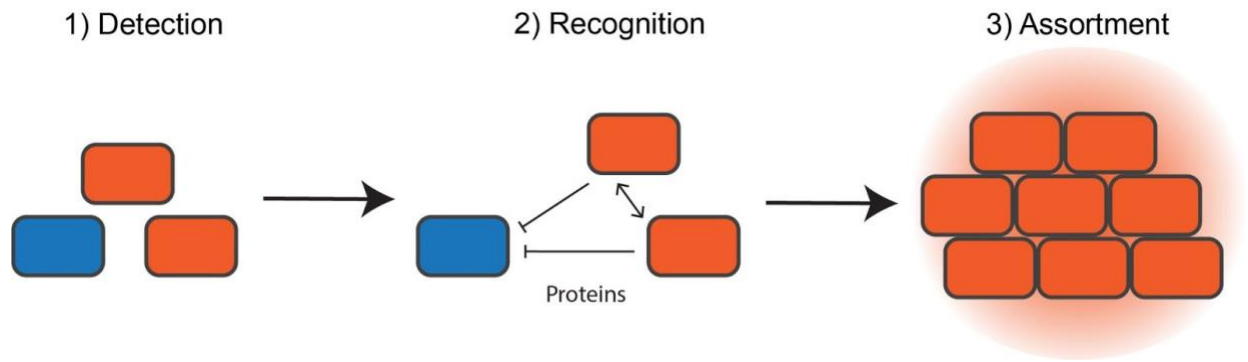
Figure 1.1: Three universal features of self recognition
First, individuals detect others in close proximity. Second, individuals use secreted or surface-exposed protein signals to indicate their identity and recognize one another as self or non-self. Third, related individuals positively assort through fusion or collective motility to form a larger colony of related cells, while non-self individuals are excluded and antagonized.

eukaryotic models, underlying operational rules (that are shared across the tree of life) could emerge.

This review of self recognition in clonal, single-celled populations reveals several shared concepts between bacteria and eukaryotes (Figure 1.2). Also examined are areas lacking critical information, particularly as concerning molecular mechanisms, with suggestions for closing these knowledge gaps. Colonial eukaryotes, such as marine invertebrates and filamentous fungi, employ self recognition for fusion with histocompatible individuals and rejection of foreign individuals, resulting in a growing colony enriched with kin (siblings) over time (Gonçalves et al., 2020; Nicotra, 2019). On the other hand, the amoeba *Dictyostelium discoideum* utilizes self recognition to regulate multicellular cooperation including collective motility and fruiting body formation (Kundert & Shaulsky, 2019). Likewise, the bacterium *Proteus mirabilis* uses self recognition to selectively engage with kin during collective motility across surfaces; non-self cells are excluded from the leading edges of the migrating colony and unrelated swarm colonies repel one another (Gibbs et al., 2008; Tipping & Gibbs, 2019; Wenren et al., 2013).

Shared concepts begin with how the identifying information is found and communicated. In these diverse systems, polymorphic gene loci encode identity information and, through protein-protein interactions, transmit this information between cells to induce cooperative or antagonistic behavioral outcomes. Recognition of self corresponds to different behavioral outcomes that depend on the organism or cell type. However, the proteins encoding identity information themselves appear to follow certain rules. These (self-)identity proteins contain variable regions, are often localized to the
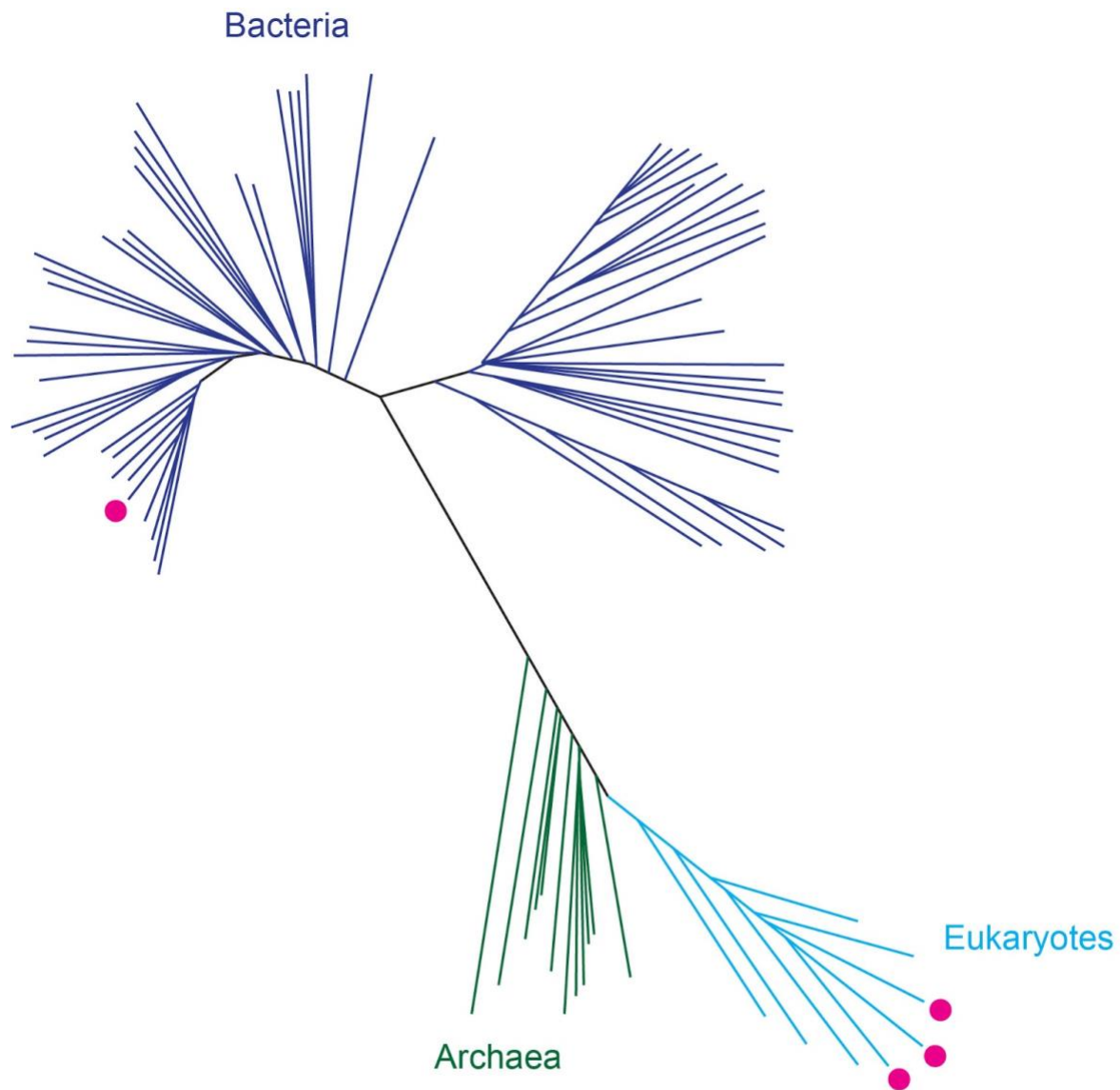
Figure 1.2: Evolutionarily distant self-recognition systems follow shared rules.
An artistic representation of the evolutionary tree of life (Hug et al., 2016). This figure is
not an accurate phylogenetic tree. The pink dots represent the eukaryotic organisms
(including animals, fungi, and protists) and the bacterium *P. mirabilis*, whose
independent, convergent self-recognition systems are described in this chapter.

cell surface or cell envelope, and interact homotypically or with a partner protein.

In the following sections, I will review a subset of self-recognition systems from three diverse eukaryotic systems: marine invertebrates, filamentous fungi, and amoeba. Within this context, I will end by considering the bacterial opportunistic pathogen, *P. mirabilis*, and how our research, including data in this dissertation, has uncovered some potentially fundamental properties of self-recognition signaling.

Marine invertebrates

Colonial invertebrates such as sponges and corals build up larger colonies from organismal units. Larger colonies are more likely to survive various environmental pressures such as predation and physical disruption (Nicotra, 2019). However, fusion means sharing of space and resources plus a potential loss of germline control. As smaller colonies interact with each other, fusion occurs if they are histocompatible, but rejection occurs if they are not (Nicotra, 2019).

In the ascidian *Botryllus schlosseri*, individual zooids are connected to one another through a shared circulatory system. The colony is encapsulated in a cellulose-based substance called the tunic. At the edge of the colony are projections called the ampullae, which contain a layer of epithelial cells at their tip involved in allorecognition (Nicotra, 2019; Taketa & De Tomaso, 2015). Colonies with compatible haplotypes fuse, otherwise cytotoxic products are released that damage the local tissue and lead to colony separation. These fusion and rejection outcomes occur relatively quickly: within 24-48 hours (Taketa & De Tomaso, 2015). Behavioral outcomes are determined by identity genes that vary between strains.

Polymorphisms in two tightly linked genes, *fuhc-sec* and *fuhc-tm*, as well as a nearby gene, *Botryllus histocompatibility factor* (BHF), are correlated with allorecognition outcomes. The gene *fuhc-sec* encodes a secreted protein that contains epidermal growth factor (EGF)-like domains, while the gene *fuhc-tm* encodes a single-pass transmembrane protein with extracellular immunoglobulin (Ig)-like domains (Nydam et al., 2013). The *BHF* gene encodes an intracellular protein without detectable domains (Voskoboynik et al., 2013). There are also two other nearby genes that seem to play a role in allorecognition. The gene *fester* encodes a transmembrane protein with a short consensus/complement repeat (SCR) domain also found in complement receptors; different isoforms of the gene can be expressed through alternative splicing (Nyholm et al., 2006). However, polymorphisms in *fester* do not correlate with allorecognition outcomes (Nydam & De Tomaso, 2012). The gene *uncle fester* is similar to *fester*, but is not polymorphic (McKitrick et al., 2011). The proteins fester and/or uncle fester possibly act as receptors for the *fuhc* ligand (Nyholm et al., 2006; Taketa & De Tomaso, 2015). Alternatively, *fester* and *uncle fester* may be required for other aspects of allorecognition signaling as non-identity partners. Defined molecular mechanisms remain elusive.

Self recognition is also found in the cnidarian *Hydractinia symbiolongicarpus*, which grows on gastropod shells. Like *B. schlosseri,* colonies of individual polyps are connected by shared vasculature. Two interacting colonies will fuse if they are compatible, otherwise rejection occurs by the release of harmful nematocysts (Nicotra, 2019). Partial compatibility leads to transient fusion before the colonies eventually separate due to autophagy and necrosis (Buss et al., 2012). The genetic determinants

for this recognition system that have been identified so far are *Allorecognition 1 (Alr1)* and *Allorecognition 2 (Alr2),* both of which are highly polymorphic. Each gene encodes a single-pass transmembrane protein with extracellular regions containing immunoglobulin-like (Ig-like) domains (Nicotra et al., 2009; Rosa et al., 2010). The two proteins, Alr1 and Alr2, are predicted to bind through homophilic interactions (Karadge et al., 2015). Both Alr1 and Alr2 also contain ITAM-like or ITIM-like motifs in the cytoplasmic tail that may be involved in downstream signaling (Nicotra et al., 2009). However, colonies collected from the environment that have matching *Alr1* and *Alr2* alleles can still reject each other (Nicotra, 2019). Other proteins may act as non-identity partners during allorecognition in *H. symbiolongicarpus*, revealing a knowledge gap in the collection of proteins needed for self recognition.

As described, approaching colonies detect one another by interacting through peripheral cells for both *B. schlosseri* and *H. symbiolongicarpus*. Recognition is determined by identity genes *fuhc-sec, fuhc-tm*, and *BHF* in *B. schlosseri*, and *Alr1* and *Alr2* in *H. symbiolongicarpus*; regions with variable amino acid sequences are found in all. Despite the evolutionary separation between these species, for both the interactions of the encoded proteins determines assortment through fusion of compatible colonies or cell death and colony separation. Non-identity partners such as *fester* may be required for signal transduction. Thus, clonal cells could selectively fuse with adjacent related colonies to grow larger colonies. The resultant communities are more robust to environmental stressors and have increased access to space and resources. Further, each step of the proposed recognition model is retained: detection, recognition, and discrimination or assortment.

<u>Filamentous fungi</u>

These steps of self recognition are also visible in filamentous fungi and more details of the molecular mechanism are known. In the filamentous fungus *Neurospora crassa*, cells fuse with other genetically related cells to create a multinucleated syncytium (Glass & Dementhon, 2006; Gonçalves et al., 2020). Several distinct checkpoints regulate the fusion process. If cells are compatible at the *doc* locus (determinant of communication), they will chemotropically grow towards each other to promote proximity and fusion events (Heller et al., 2016). Once contact has occurred, but before fusion, incompatibility between the linked genes *cwr-1* and *cwr-2* can abort the fusion process. The gene *cwr-1* encodes a polysaccharide monooxygenase and the gene *cwr-2* encodes a predicted transmembrane protein (Gonçalves et al., 2019). After fusion occurs, there is another checkpoint called germling-regulated death (GRD), determined by compatibility at the *sec-9/plp-1* and *rcd-1* loci (Daskalov et al., 2019; Heller et al., 2018). Of note, PLP-1 is similar to NOD-like receptor (NLR) proteins, which are components of innate immunity in plants and animals. SEC-9 contains a SNARE domain, which usually mediates vesicle fusion to a target membrane. PLP-1 localizes to the cell periphery, and becomes activated if it interacts with incompatible SEC-9, triggering cell death (Heller et al., 2018). The *rcd-1-1* and *rcd-1-2* genes encode proteins of unknown function (Daskalov et al., 2019). In sum, multiple self-recognition loci, and checkpoints, regulate migration and compatibility prior to hyphal fusion (i.e., joining efforts with siblings).

The converse is a discerning penalty for non-clonal cells. After hyphal fusion, rejection can induce programmed cell death based on heterokaryon incompatibility that

often leads to macroscopic separation of strains (Glass & Dementhon, 2006). The genetic determinants of this vegetative compatibility occur in multiple, unlinked *het* loci. These loci often contain a polymorphic gene linked to a gene encoding a HET domain, which is a domain of unknown function. Outside of the HET domain, *het* genes do not appear to share any sequence identity or domain structure. Predicted HET domain genes are specific to, and conserved in, filamentous ascomycetes. In *N. crassa*, there are at least 11 loci that contribute to allorecognition (Zhao et al., 2015). Three of these loci are molecularly characterized as predicted HET domain proteins and their partners: *mat/tol*, *un-24/het-6*, and *het-c/pin-c*. Proteins TOL, HET-6, and PIN-C contain HET domains. Their partners vary in predicted function: MAT is a predicted transcription factor, HET-C is a predicted plasma membrane protein, and UN-24 is a predicted ribonucleotide reductase (Glass & Dementhon, 2006). The partner proteins are predicted to interact with each other to determine recognition outcomes (Kaneko et al., 2006). In sum, HET domain proteins and their partners determine recognition and compatibility after hyphal fusion.

Again, each step of the proposed recognition model is retained: detection, recognition, and discrimination or assortment. In the fungus *N. crassa*, cells can detect one another through secreted signals and cell-cell contact. Recognition before fusion is determined by identity genes *cwr-1, cwr-2, rcd-1-1* and *rcd-1-2.* After fusion, recognition is determined by interaction between HET domain proteins and their partners. Distinguishing self from non-self leads to hyphal fusion and assortment of kin into a single syncytium, where nutrients and other resources are shared across a single network-like cell. The multiple molecular steps may be due to the complexity of fungal

biology, or more likely, access to a model organism with more tractable genetics and biochemistry.

<u>Social amoeba</u>

Tractability in genetics, biochemistry, evolution (in terms of time), and imaging lowers the barriers for uncovering shared principles in self recognition. One such model is the slime mold *D. discoideum.* In response to starvation, asocial, individual cells of the slime mold *D. discoideum* stop dividing and enter a social phase. Cells aggregate into collectively moving colonies, termed "slugs," that ultimately form fruiting bodies and release spores (Kundert & Shaulsky, 2019). Only a subset of cells end up in the spore because many other cells sacrifice themselves to construct the spore-supporting stalk. To prevent some cells from cheating their way into the spore, kin are selected during the developmental process. Aggregation and coordination of development is correlated with the linked, polymorphic genes *tgrB1* and *tgrC1* and the binding interaction of the encoded proteins (Benabentos et al., 2009; Gruenheit et al., 2017). TgrB1 and TgrC1 are polymorphic identity determinants during fruiting body formation.

TgrB1 and TgrC1 contain extracellular Ig-like domains with single transmembrane domains. Both proteins function in cell-cell adhesion. TgrB1 may be the receptor that can signal downstream based on its binding interaction with the ligand TgrC1 (Hirose et al., 2017). During *D. discoideum* development, cells coordinate movement using chemotaxis driven by extracellular cAMP gradients; however, the TgrB1-TgrC1 complex formed between cells appears to orient the direction of cellular migration within multicellular slugs (Fujimori et al., 2019). Inactivation of either the *tgrB1* or *tgrC1* gene leads to defects in developmental processes including aggregation and

fruiting body formation (Benabentos et al., 2009). Therefore, TgrB1 and TgrC1 play a multifaceted role in *D. discoideum* development beyond cell-cell adhesion that may rely on non-identity partners. For self recognition, again, each step of the proposed recognition model is retained: detection, recognition, and discrimination or assortment.

When *D. discoideum* populations transition to a social phase, cells aggregate through contact-dependent interactions. Cells recognize one another through interaction of the membrane proteins TgrB1 and TgrC1, which promotes cell-cell adhesion and coordinates migration. Aggregation and motility drives the assortment of related cells into clonal fruiting bodies that release spores into the environment to find new nutrient-rich areas for growth.

Open questions among single-cell eukaryotes

Many identity genes are described under the "greenbeard" theory. This idea is that a specific so-called greenbeard gene can act as a relatedness factor for cooperative behavior (Madgwick et al., 2019). However, few genes nicely fit into this theory, especially since many evolutionary questions remain about how such a gene would be maintained. Part of the greenbeard theory is the idea that a single gene or gene locus is responsible for all aspects of identity signaling: the display of a phenotypic marker, perception of the marker, and modulation of the social response.

However, this single-gene framework may be limiting our understanding of self recognition by ignoring important interactions and signaling pathways within the organism. For example, in many recognition systems, the identity determinants themselves are known and characterized. However, it is not completely understood how the recognition event, usually a protein-protein interaction, is actually translated into a

behavioral change, especially in the eukaryotic model systems. Further, interactions

with non-identity partners may be crucial for generating allelic diversity and maintaining

self-recognition systems. While there are bound to be many differences between

organisms, the downstream responses to allorecognition observed in both marine

invertebrates and filamentous fungi are similar to the immune response in plants and

animals (Heller et al., 2018; Oren et al., 2013). Perhaps there are underlying properties

that hold not only for identity determinants, but also for identity signal transduction

across the tree of life.

These other components of self-recognition signaling could, at least in part,

regulate the evolution of identity determinants (Figure 1.3). Polymorphic proteins are

thought to be maintained through balancing selection or negative frequency-dependent

selection (Hedrick, 2007). This is in contrast to directional selection where alleles

increase in frequency to fixation, reducing variability in the gene. Essentially, alleles are

hypothesized to be selected when rare because they are more likely to accurately

encode self in a population. This property can promote allelic diversity in a population as

long as new alleles are continuously generated (Nydam, 2011). Balancing selection can

be detected through specific genetic signatures such as positive selection ($d_N/d_S>1$) and

high within-population variation (low F_{ST} compared to neutral loci), which were shown

for identity factors is likely required for the signaling pathway linking recognition to

behavioral outcomes. Identifying other proteins that are not polymorphic identity

determinants, some of which may not even be encoded in the self-recognition genetic

locus, is important for understanding the molecular mechanisms. These critical proteins

may exert additional constraints on the evolution of self-recognition proteins, particularly
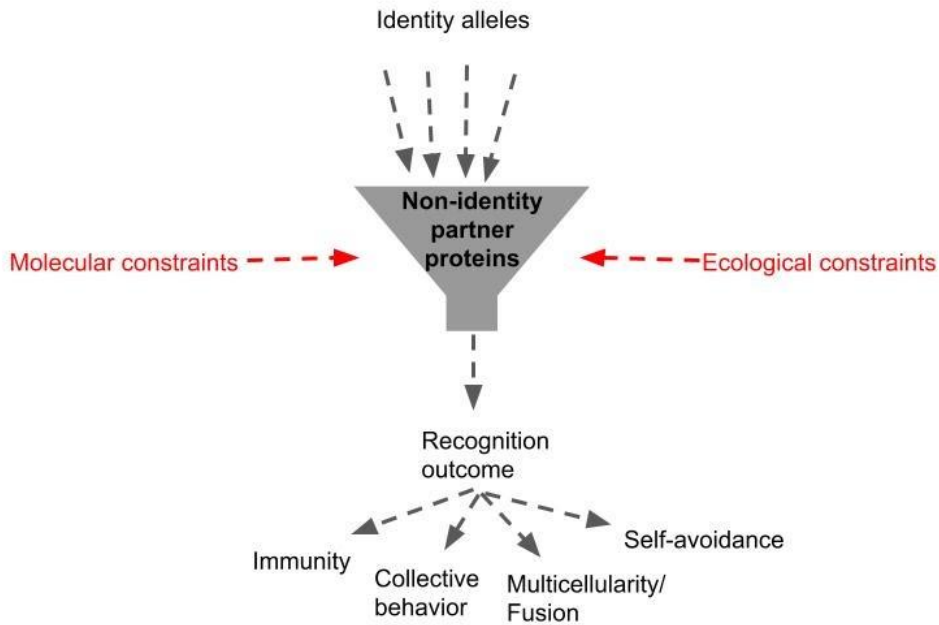
Figure 1.3: Protein-protein interactions outside of identity determinants may constrain the evolution of identity alleles due to broader signal fidelity. Non-identity partner proteins may act as a selective filter for the generation of novel identity alleles that can reliably and accurately determine recognition outcomes and are therefore subject to positive selection. In red, molecular and ecological constraints on the non-identity partners may constrain the non-identity partners, indirectly affecting the function of identity alleles as well. Recognition outcomes are diverse across different organisms and cell types, but there may be underlying properties of identity signal transduction and corresponding evolutionary constraints that hold across different kinds of organisms.

if they experience greater structural constraints or contribute to ecological fitness (Figure 1.3). Further, there may be a common set of as yet unknown critical proteins or pathways that are frequently co-opted for self-recognition signaling.

Proteus mirabilis as a model system for self-recognition signaling

The link between recognition and assortment is still a black box for many eukaryotic systems. It is currently difficult to elucidate interactions between identity determinants and non-identity partners required for signal transduction due to the biological complexity of many eukaryotic organisms. By contrast, the model experimental system, *P. mirabilis,* is a tractable bacterial system for elucidating signal transduction mechanisms required for self recognition (Figure 1.4). As stated above, tractability in genetics, biochemistry, evolution (in terms of time), and imaging reduces barriers for uncovering molecular mechanisms of self recognition.

P. mirabilis is an opportunistic pathogen. This bacterium causes catheter-associated urinary tract infections in humans, but can also be found in healthy humans and animals at low abundance as a gut commensal (Armbruster et al., 2018; Drzewiecka, 2016). Populations exhibit collective swarming motility across solid surfaces that is associated with virulence (Armbruster & Mobley, 2012; Morgenstein et al., 2010; Rather, 2005). During collective swarm motility, cells interact with their neighbors and exchange identity information; one such pathway is the Ids (identification of self) non-lethal self-recognition system (Gibbs et al., 2008; Wenren et al., 2013). The identity signal IdsD is injected into a neighboring cell, where it can interact with its partner protein IdsE for recognition [Figure 1.4, (Saak & Gibbs, 2016)]. Binding of IdsD and IdsE distinguishes self from non-self, leading to normal swarm colony expansion. In

a strain where a different IdsE isoform is expressed, IdsD and IdsE do not bind; every cell perceives their neighbors as non-self, resulting in reduced swarm colony expansion [Figure 1.4, (Cardarelli et al., 2015; Saak & Gibbs, 2016; Tipping & Gibbs, 2019)]. IdsD and IdsE are non-lethal identity determinants that mediate self recognition during swarming.

Parallels are apparents between *P. mirabilis* Ids-mediated recognition and the eukaryotic recognition systems described above. For two swarm colonies approaching one another, Ids-mediated self recognition between cells at the leading edges results in merging of the colonies while non-self recognition leads to separation of the colonies by a clear, visible boundary (Gibbs et al., 2008). IdsD is secreted directly into neighboring cells, which requires interaction with non-identity partners in the donor cell: a chaperone protein and the type VI secretion system (Saak, 2016; Zepeda-Rivera, 2018). IdsD induces cell stress in recipient non-self cells; the intracellular mechanism likely requires non-identity partners (Tipping & Gibbs, 2019). By contrast, self recognition in the bacterium *Myxococcus xanthus* employs two identity proteins TraA and TraB to mediate outer membrane fusion, and subsequent exchange of its components, between neighboring cells (Cao et al., 2019; Cossey et al., 2019). A multi-component secretion system is likely not involved, but there is no definitive answer as yet. Further, each step of the proposed recognition model occurs: detection, recognition, and discrimination or assortment.
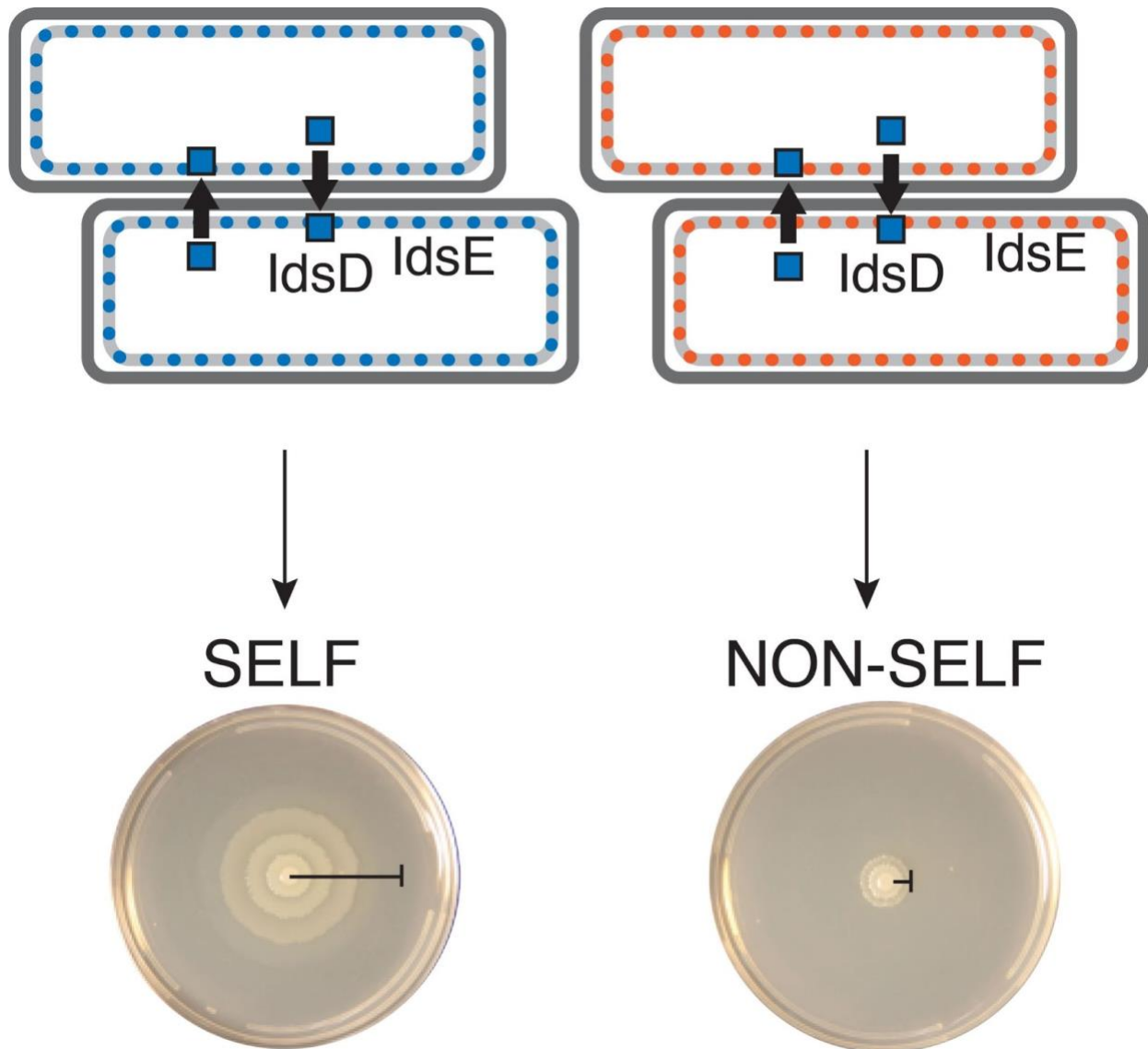
Figure 1.4: Identity influences the collectivity needed for colony migration.
During swarming, cells are in physical contact and exchange protein signals through the type VI secretion system (Wenren et al., 2013). The identity signal IdsD (square icon) is transferred from one cell to another where it likely interacts with its partner IdsE (circle icon), which is localized along the inner membrane (Saak & Gibbs, 2016). On the left, IdsD and IdsE are compatible, interact, and lead to recognition of the cell as self; the result is proficient expansion of the swarm colony (Cardarelli et al., 2015; Saak & Gibbs, 2016). On the right, if IdsD and IdsE are not compatible (in this case IdsE has been swapped with a different isoform), the identity proteins do not interact and the cell is recognized as non-self (Cardarelli et al., 2015; Saak & Gibbs, 2016). IdsD transiently induces a developmental switch to a low-metabolism state in individual non-self cells, resulting in a >7% reduction in overall swarm expansion (Saak & Gibbs, 2016; Tipping & Gibbs, 2019).

The serine transporter SdaC is a non-identity partner in *P. mirabilis*

In this thesis, I show that a conserved serine transporter, SdaC, moonlights in self recognition. I propose that SdaC is likely as an inner membrane receptor for the identity protein IdsD (Chapter 2). My research revealed the following about this potential receptor. SdaC is the dominant serine transporter during swarm motility. This protein, which contains a conserved LeuT fold, samples an open and closed conformation during serine transport. Only the open conformation, and not serine transport, is required for self recognition. While the SdaC ortholog from *Escherichia coli* is sufficiently conserved to substitute for the *P. mirabilis* SdaC in self recognition, a homologous serine transporter YhaO in *P. mirabilis* is unable to replace SdaC for self recognition, but can for serine transport. There may be a specific molecular interface that is exposed when SdaC is in its open conformation that is required for interaction with the identity signal IdsD. Ecological and structural constraints on SdaC may also constrain the evolution of IdsD.

In Chapter 3, I investigate the phylogenetics of *P. mirabilis,* which was previously focused on human-associated isolates. I was the experimental lead on a collaborative project to analyze a pangenome dataset with associated phenotypic characterization to investigate strain-level diversity and population structure across *P. mirabilis* ecotypes (Chapter 3). This was accomplished using newly sequenced *P. mirabilis* isolates from asymptomatic animals in a Charles River Laboratories facility and executed in collaboration with colleagues in the Gibbs and Cavanaugh laboratories. Our findings were as follows. The genotypes and phenotypes were largely similar. Minor phenotypic variation does exist. In one example, a single nucleotide polymorphism (SNP) in a core

flagellar gene drastically affected swarming and virulence. These results highlight the impact subtle sequence variation can have on gene expression and phenotype. We found that identity determinants required for self recognition, which are known to vary between strains, were predictably in the variable regions of the pangenome, but the machinery required for secretion and non-identity partners are in the core genome. This dataset can be used to further investigate sequence variation at self-recognition loci, protein coevolution indicative of potential protein-protein interactions, and functional identity protein domains.

This dissertation concludes with additional thoughts on proposed models and experiments. Throughout this research, shared features of self recognition from bacteria to colonial eukaryotes emerge. More broadly, each of these systems with collective states critical for development share each step of self recognition: detection, recognition, and discrimination or assortment. But, of course, the devil is in the details.

References:

Armbruster, C. E., & Mobley, H. L. (2012). Merging mythology and morphology: The multifaceted lifestyle of *Proteus mirabilis*. *Nat Rev Microbiol*, *10*(11), 743–754. https://doi.org/10.1038/nrmicro2890

Armbruster, C. E., Mobley, H. L. T., & Pearson, M. M. (2018). Pathogenesis of *Proteus mirabilis* Infection. *EcoSal Plus*, *8*(1), 10.1128/ecosalplus.ESP-0009–2017. https://doi.org/10.1128/ecosalplus.ESP-0009-2017

Benabentos, R., Hirose, S., Sucgang, R., Curk, T., Katoh, M., Ostrowski, E. A., Strassmann, J. E., Queller, D. C., Zupan, B., Shaulsky, G., & Kuspa, A. (2009). Polymorphic Members of the lag Gene Family Mediate Kin Discrimination in *Dictyostelium*. *Current Biology*, *19*(7), 567–572. https://doi.org/10.1016/j.cub.2009.02.037

Buss, L. W., Anderson, C., Westerman, E., Kritzberger, C., Poudyal, M., Moreno, M. A., & Lakkis, F. G. (2012). Allorecognition Triggers Autophagy and Subsequent Necrosis in the Cnidarian *Hydractinia symbiolongicarpus*. *PLoS ONE*, *7*(11), e48914. https://doi.org/10.1371/journal.pone.0048914

Cao, P., Wei, X., Awal, R. P., Müller, R., & Wall, D. (2019). A Highly Polymorphic Receptor Governs Many Distinct Self-Recognition Types within the *Myxococcales* Order. *MBio*, *10*(1), e02751-18.

Cardarelli, L., Saak, C., & Gibbs, K. A. (2015). Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding. *MBio*, *6*(3), e00251-15.

Cossey, S. M., Yu, Y.-T. N., Cossu, L., & Velicer, G. J. (2019). Kin discrimination and outer membrane exchange in *Myxococcus xanthus*: Experimental analysis of a natural population. *Plos One*, *14*(11), e0224817.

Daskalov, A., Gladieux, P., Heller, J., & Glass, N. L. (2019). Programmed cell death in *Neurospora crassa* is controlled by the allorecognition determinant rcd-1. *Genetics*, *213*(4), 1387–1400.

Drzewiecka, D. (2016). Significance and Roles of *Proteus spp.* Bacteria in Natural Environments. *Microb Ecol*, *72*(4), 741–758. https://doi.org/10.1007/s00248-015-0720-6

Fujimori, T., Nakajima, A., Shimada, N., & Sawai, S. (2019). Tissue self-organization based on collective cell migration by contact activation of locomotion and chemotaxis. *Proceedings of the National Academy of Sciences*, *116*(10), 4291–4296.

Gibbs, K. A., Urbanowski, M. L., & Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. *Science*, *321*(5886), 256–259. https://doi.org/10.1126/science.1160033

Glass, N. L., & Dementhon, K. (2006). Non-self recognition and programmed cell death in filamentous fungi. *Current Opinion in Microbiology*, *9*(6), 553–558. https://doi.org/10.1016/j.mib.2006.09.001

Gonçalves, A. P., Heller, J., Rico-Ramírez, A. M., Daskalov, A., Rosenfield, G., & Glass, N. L. (2020). Conflict, Competition, and Cooperation Regulate Social Interactions in Filamentous Fungi. *Annual Review of Microbiology*, *74*, 693–712.

Gonçalves, A. P., Heller, J., Span, E. A., Rosenfield, G., Do, H. P., Palma-Guerrero, J., Requena, N., Marletta, M. A., & Glass, N. L. (2019). Allorecognition upon fungal cell-cell contact determines social cooperation and impacts the acquisition of multicellularity. *Current Biology*, *29*(18), 3006-3017. e3.

Grice, L. F., & Degnan, B. M. (2015). How to Build an Allorecognition System: A Guide for Prospective Multicellular Organisms. In I. Ruiz-Trillo & A. M. Nedelcu (Eds.), *Evolutionary Transitions to Multicellular Life* (Vol. 2, pp. 395–424). Springer Netherlands. https://doi.org/10.1007/978-94-017-9642-2_19

Gruenheit, N., Parkinson, K., Stewart, B., Howie, J. A., Wolf, J. B., & Thompson, C. R. (2017). A polychromatic 'greenbeard' locus determines patterns of cooperation in a social amoeba. *Nature Communications*, *8*(1), 1–9.

Hedrick, P. W. (2007). Balancing selection. *Current Biology*, *17*(7), R230–R231. https://doi.org/10.1016/j.cub.2007.01.012

Heller, J., Clavé, C., Gladieux, P., Saupe, S. J., & Glass, N. L. (2018). NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. *Proceedings of the National Academy of Sciences*, *115*(10), E2292–E2301.

Heller, J., Zhao, J., Rosenfield, G., Kowbel, D. J., Gladieux, P., & Glass, N. L. (2016). Characterization of Greenbeard Genes Involved in Long-Distance Kind Discrimination in a Microbial Eukaryote. *PLOS Biology*, *14*(4), e1002431. https://doi.org/10.1371/journal.pbio.1002431

Hirose, S., Chen, G., Kuspa, A., & Shaulsky, G. (2017). The polymorphic proteins TgrB1 and TgrC1 function as a ligand–receptor pair in *Dictyostelium* allorecognition. *J Cell Sci*, *130*(23), 4002–4012.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., & Ise, K. (2016). A new view of the tree of life. *Nature Microbiology*, *1*(5), 1–6.

Kaneko, I., Dementhon, K., Xiang, Q., & Glass, N. L. (2006). Nonallelic Interactions Between *het-c* and a Polymorphic Locus, *pin-c*, Are Essential for Nonself Recognition and Programmed Cell Death in *Neurospora crassa. Genetics*, *172*(3), 1545–1555. https://doi.org/10.1534/genetics.105.051490

Karadge, U. B., Gosto, M., & Nicotra, M. L. (2015). Allorecognition Proteins in an Invertebrate Exhibit Homophilic Interactions. *Current Biology*, *25*(21), 2845–2850. https://doi.org/10.1016/j.cub.2015.09.030

Kundert, P., & Shaulsky, G. (2019). Cellular allorecognition and its roles in *Dictyostelium* development and social evolution. *The International Journal of Developmental Biology*, *63*(8-9–10), 383.

Madgwick, P. G., Belcher, L. J., & Wolf, J. B. (2019). Greenbeard genes: Theory and reality. *Trends in Ecology & Evolution*, *34*(12), 1092–1103.

McKitrick, T. R., Muscat, C. C., Pierce, J. D., Bhattacharya, D., & De Tomaso, A. W. (2011). Allorecognition in a Basal Chordate Consists of Independent Activating and Inhibitory Pathways. *Immunity*, *34*(4), 616–626. https://doi.org/10.1016/j.immuni.2011.01.019

Morgenstein, R. M., Szostek, B., & Rather, P. N. (2010). Regulation of gene expression during swarmer cell differentiation in Proteus mirabilis. *FEMS Microbiol Rev*, *34*(5), 753–763. https://doi.org/10.1111/j.1574-6976.2010.00229.x

Nicotra, M. L. (2019). Invertebrate allorecognition. *Current Biology*, *29*(11), R463–R467.

Nicotra, M. L., Powell, A. E., Rosengarten, R. D., Moreno, M., Grimwood, J., Lakkis, F. G., Dellaporta, S. L., & Buss, L. W. (2009). A Hypervariable Invertebrate Allodeterminant. *Current Biology*, *19*(7), 583–589. https://doi.org/10.1016/j.cub.2009.02.040

Nydam. (2011). Creation and maintenance of variation in allorecognition loci. *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2011.00079

Nydam, M. L., & De Tomaso, A. W. (2012). The fester locus in *Botryllus schlosseri* experiences selection. *BMC Evolutionary Biology*, *12*(1), 249. https://doi.org/10.1186/1471-2148-12-249

Nydam, M. L., Netuschil, N., Sanders, E., Langenbacher, A., Lewis, D. D., Taketa, D. A., Marimuthu, A., Gracey, A. Y., & De Tomaso, A. W. (2013). The Candidate Histocompatibility Locus of a Basal Chordate Encodes Two Highly Polymorphic Proteins. *PLoS ONE*, *8*(6), e65980. https://doi.org/10.1371/journal.pone.0065980

Nydam, M. L., Stephenson, E. E., Waldman, C. E., & De Tomaso, A. W. (2017). Balancing selection on allorecognition genes in the colonial ascidian *Botryllus schlosseri*. *Developmental & Comparative Immunology*, *69*, 60–74.

Nyholm, S. V., Passegue, E., Ludington, W. B., Voskoboynik, A., Mitchel, K., Weissman, I. L., & De Tomaso, A. W. (2006). Fester, a Candidate Allorecognition Receptor from a Primitive Chordate. *Immunity*, *25*(1), 163–173. https://doi.org/10.1016/j.immuni.2006.04.011

Oren, M., Paz, G., Douek, J., Rosner, A., Amar, K. O., & Rinkevich, B. (2013). Marine invertebrates cross phyla comparisons reveal highly conserved immune

machinery. *Immunobiology*, *218*(4), 484–495. https://doi.org/10.1016/j.imbio.2012.06.004

Rather, P. N. (2005). Swarmer cell differentiation in *Proteus mirabilis. Environ Microbiol*, *7*(8), 1065–1073. https://doi.org/10.1111/j.1462-2920.2005.00806.x

Rosa, S. F. P., Powell, A. E., Rosengarten, R. D., Nicotra, M. L., Moreno, M. A., Grimwood, J., Lakkis, F. G., Dellaporta, S. L., & Buss, L. W. (2010). *Hydractinia* Allodeterminant *alr1* Resides in an Immunoglobulin Superfamily-like Gene Complex. *Current Biology*, *20*(12), 1122–1127. https://doi.org/10.1016/j.cub.2010.04.050

Saak, C. C., & Gibbs, K. A. (2016). The Self-Identity Protein IdsD Is Communicated between Cells in Swarming *Proteus mirabilis* Colonies. *Journal of Bacteriology*, *198*(24), 3278–3286.

Taketa, D. A., & De Tomaso, A. W. (2015). Botryllus schlosseri allorecognition: Tackling the enigma. *Developmental & Comparative Immunology*, *48*(1), 254–265.

Tipping, M. J., & Gibbs, K. A. (2019). Peer pressure from a *Proteus mirabilis* self-recognition system controls participation in cooperative swarm motility. *PLoS Pathogens*, *15*(7), e1007885.

Voskoboynik, A., Newman, A. M., Corey, D. M., Sahoo, D., Pushkarev, D., Neff, N. F., Passarelli, B., Koh, W., Ishizuka, K. J., Palmeri, K. J., Dimov, I. K., Keasar, C., Fan, H. C., Mantalas, G. L., Sinha, R., Penland, L., Quake, S. R., & Weissman, I. L. (2013). Identification of a Colonial Chordate Histocompatibility Gene. *Science*, *341*(6144), 384–387. https://doi.org/10.1126/science.1238036

Wenren, L. M., Sullivan, N. L., Cardarelli, L., Septer, A. N., & Gibbs, K. A. (2013). Two independent pathways for self-recognition in Proteus mirabilis are linked by type VI-dependent export. *MBio*, *4*(4), e00374-13. https://doi.org/10.1128/mBio.00374-13

Zhao, J., Gladieux, P., Hutchison, E., Bueche, J., Hall, C., Perraudeau, F., & Glass, N. L. (2015). Identification of allorecognition loci in *Neurospora crassa* by genomics and evolutionary approaches. *Molecular Biology and Evolution*, *32*(9), 2417–2432.

Chapter 2

The conserved serine transporter SdaC moonlights to enable self recognition

Achala Chittor and Karine A. Gibbs

Abstract

       Cells can use self recognition to achieve cooperative behaviors. Self-recognition genes principally evolve in tandem with partner self-recognition alleles. However, other constraints on protein evolution could exist. Here, we have identified an interaction outside of self-recognition loci that could constrain the sequence variation of a self-recognition protein. We show that during collective swarm expansion in *Proteus mirabilis*, self-recognition signaling co-opts SdaC, a serine transporter. Serine uptake is crucial for bacterial survival and colonization. Single-residue variants of SdaC reveal that self recognition requires an open conformation of the protein; serine transport is dispensable. A distant ortholog from *Escherichia coli* is sufficient for self recognition; however, a homologous serine transporter, YhaO, is not. Thus, SdaC couples self recognition and serine transport, likely through a shared molecular interface. Understanding molecular and ecological constraints on self-recognition proteins can provide insights into the evolution of self recognition and emergent collective behaviors.

Introduction

 Self recognition regulates diverse processes in organisms across the tree of life. These vital roles include cell-cell communication, morphogenesis, and cooperation. For example, individual neurons express their identity through a unique set of clustered protocadherin proteins that interact between cells to inform neuronal self-avoidance and proper circuit formation (Kostadinov & Sanes, 2015; Lefebvre et al., 2012; Molumby et al., 2016). Kin cells of filamentous fungi can fuse to share resources in a syncytial lifestyle (Fischer & Glass, 2019; Gonçalves et al., 2020). And social microbes can identify, and coordinate with, kin during group migrations and fruiting body formation (Asfahl & Schuster, 2017; Gibbs et al., 2008; Gruenheit et al., 2017; Hirose et al., 2017; Pathak et al., 2013; Wenren et al., 2013). Self-recognition proteins often contain regions—stretches of amino acids—that vary between lineages to barcode a range of identities. The molecular mechanisms that regulate variation in self-recognition proteins are not fully elucidated.

 From studies on eukaryotes, we know that protein-protein interactions affect how self-recognition proteins change over time. For clustered protocadherins, residues co-evolve across the homophilic interaction interface (Nicoludis et al., 2015, 2016, 2019). Allelic diversity of self-recognition genes results from balancing selection (Gruenheit et al., 2017; Noonan et al., 2003; Wu, 2005; Zhao et al., 2015). Still, no comprehensive model unifies genetics and molecular mechanisms. A similar gap between gene evolution and protein biochemistry exists for microbial self recognition. One can directly address this schism by studying the broader interaction networks that shape self-recognition evolution, both in individual microbial cells and populations.

Many microbes use self recognition to engage in collective behaviors selectively with kin. Microbial populations can build large, structured biofilms or migrate collectively across surfaces (Strassmann et al., 2011; West et al., 2007). For example, *P. mirabilis* cells engage in a collective migration known as swarming, allowing populations to cover surfaces efficiently. Individual cells communicate using self-recognition proteins to discern clonal siblings from others. Siblings gain preferred access to collective motility; non-self cells enter a transient altered state, resulting in exclusion from the swarming population (Cardarelli et al., 2015; Saak & Gibbs, 2016; Tipping & Gibbs, 2019). Unlike many cooperative self-recognition proteins found on the cell's surface (Hirose et al., 2017; Pathak et al., 2013)*, P. mirabilis* cells inject the protein signal IdsD into adjacent neighbors. Mechanisms for localization and downstream signaling in the recipient cell are unknown.

For delivery into the recipient cell, IdsD may interact with a receptor protein residing in the inner membrane. Models propose that IdsD localizes to the inner membrane of recipient *P. mirabilis* cells (Cardarelli et al., 2015; Zepeda-Rivera et al., 2018), but there is less clarity about the delivery mechanism. A comparable model is the contact-dependent inhibition toxin, CdiA. In *Escherichia coli*, CdiA can promote collective behavior through cell-cell adhesion. Transferred from one cell into its neighbor, CdiA requires outer and inner membrane receptors for delivery to a recipient cell's cytoplasm (Willett et al., 2015). Uptake depends on the interaction between CdiA's modular translocation domain and a species-specific host receptor (Ruhe et al., 2013, 2017). None of IdsD's species-specific binding partners—the chaperone IdsC in the donor cell and its self-recognition partner IdsE in the recipient cell (Cardarelli et al.,

2015; Zepeda-Rivera et al., 2018)—could reasonably act as a receptor for delivery to the inner membrane. Although receptors on the inner or outer membrane have remained elusive, identifying these proteins is necessary to describe self-recognition pathways and elucidate further constraints on IdsD's evolution.

With this in mind, consider bacteriophage-host interactions as a framework for studying the molecular evolution of self-recognition proteins. Many phage receptors are nutrient transporters on the outer and inner membranes. Phages exploit these host transporters, leading to rapid coevolution of competing proteins—an "arms race," as reported for outer membrane proteins (de Jonge et al., 2019; Hampton et al., 2020). While mutations in the receptor can disrupt interactions and prevent killing, modifying these transporters can also carry fitness costs due to the importance of nutrient uptake (Mangalea & Duerkop, 2020; Meaden et al., 2015). Receptors for self-recognition proteins, particularly nutrient transporters, could likewise carry both costs and benefits. Discerning these interactions could reveal molecular constraints on the evolution of self-recognition systems.

Ids-mediated self recognition in *P. mirabilis* provides an opportunity to examine this concept. Cells can bypass recognition-triggered limitations by removing the self-recognition genes or the cell-to-cell transport system (Saak & Gibbs, 2016; Zepeda-Rivera et al., 2018). Here we show that another mechanism is the disruption of the serine transporter SdaC. Self-recognition signaling specifically uses SdaC during collective motility. By analyzing single-residue variants to alter protein state, we show self-recognition signaling requires an open conformation but not serine transport. Supporting the importance of a protein interface, we saw that an ortholog from *E. coli* is

sufficient for self recognition. A related serine transporter in *P. mirabilis*, YhaO, is not.

Therefore, SdaC moonlights to couple self recognition with serine transport, revealing a

critical interaction and a potential regulator for self-recognition protein evolution.

Results

Organisms can escape self-recognition control using several strategies. Spontaneous mutations that restore collective motility have removed self-recognition communication, either by disrupting self-recognition genes themselves or disrupting the necessary transport system (Saak & Gibbs, 2016). To identify additional mechanisms, we modified an unbiased assay for isolating self-recognition escape mutants emerging from an initial swarm colony [(Figure 2.1A), (Saak & Gibbs, 2016)]. We selected twelve novel independent mutant strains for further study. Whole-genome sequencing showed that these strains contained mutations in the gene *sdaC*; all, except two, were deleterious to a full-length protein (Figure 2.1B). We reasoned that *sdaC* disruption might constitute another mode of escape from recognition-based swarm exclusion.

Direct deletion and complementation of *sdaC* in a clean genetic background could confirm this proposed role. We generated a non-polar deletion of *sdaC* in the non-self Δ*idsE* background (Zepeda-Rivera et al., 2018), resulting in strain ACH01. In a colony of the Δ*idsE* strain, each cell perceives others as non-self, resulting in restricted collective swarm motility (Figure 2.1C). Strain ACH01 showed a relative three-fold increase in collective swarm motility. By contrast, plasmid-based expression of *sdaC* in ACH01 restricted motility back to the non-self strain's level (Figure 2.1C). The rescue of collective motility in ACH01 is not due to a change in the self-recognition protein's transport. The *sdaC* gene was dispensable for IdsD secretion as measured using a self-recognition assay based on boundary formation between colliding swarming populations (Figure A8). Thus, *sdaC* is necessary for self-recognition signal transduction in the receiving cell.

Figure 2.1: SdaC is required for Ids-mediated self recognition and is the dominant serine transporter during swarming.

A) Strains containing *de novo* mutations emerge (marked with arrows) from a restricted swarm colony of strain BB2000 Δ*ids* pIdsBBΔ*idsE* (Saak & Gibbs, 2016). Images of swarm plate at the end of standard two day incubation (left) and after incubating an additional one week (right). B) Isolated mutations occur throughout SdaC protein sequence; relative location denoted by tick marks. Magenta lines mark missense mutations. Black lines denote nonsense mutations. See Table A8 for specific sites. C) Swarm radius measured from swarm assay of BB2000 empty vector, Δ*idsE* empty vector, ACH01 [BB2000 Δ(*idsE*, *sdaC*)] empty vector, and ACH01 pSdaC. Representative plate images shown below. D) Brief prediction of serine uptake and utilization pathway in *P. mirabilis*, which was constructed based on the homologies of SdaA, SdaB, and SdaC to *E. coli* proteins. E) Growth curve of BB2000 empty vector, ACH04 [BB2000 Δ(*sdaA*, *sdaB*)] empty vector, ACH05 [BB2000 Δ(*sdaA*, *sdaB-sdaC*)] empty vector, and ACH05 pSdaC in minimal medium (left) and minimal medium plus 10mM L-serine (right) for three biological replicates. F) Serine concentration calculated from LC-MS results of total swarmer cell lysates of BB2000, ACH04, and ACH05 and normalized to total protein content for two technical replicates. G) Swarm radius measured (from edge of inoculum) from swarm assay of ACH04 empty vector, ACH05 empty vector, and ACH05 pSdaC. CFUs per swarm colony shown in Figure A9. Representative plate images in Figure A11. H) Swarm radius measured from swarm assay of strains BB2000, ACH04, ACH04 Δ*sstT*, ACH04 Δ*yhaO*.
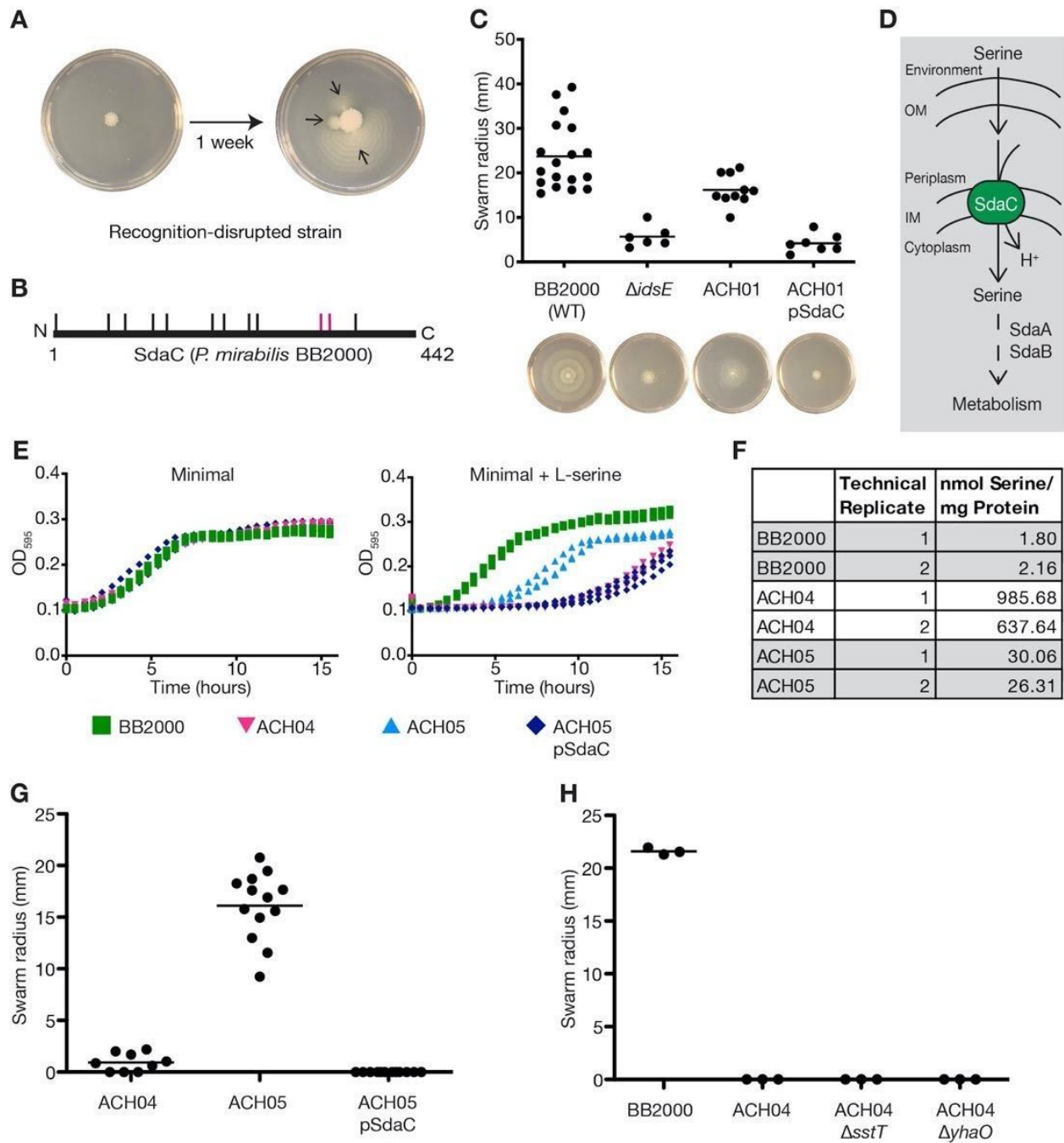
Figure 2.1: SdaC is required for Ids-mediated self recognition and is the dominant serine transporter during swarming. (Continued)

The SdaC protein is a predicted membrane-bound serine transporter but without confirmed function in *Proteus* spp. In *E. coli* and other enteric Gram-negative bacteria, SdaC is an integral inner membrane protein that brings serine into the cell coupled to a proton (Hama et al., 1988; Shao et al., 1994; Velayudhan et al., 2004). Serine deaminases (SdaA and SdaB) primarily metabolize imported serine to pyruvate (Shao & Newman, 1993; Su et al., 1989; Velayudhan et al., 2004). *Proteus* spp. contain genes with sequence similarity to these serine import and utilization genes (Figure 2.1D). We confirmed membrane localization of *P. mirabilis* SdaC by expressing a mCherry fusion from its native promoter (Figure A10). As with *E. coli* (Baba et al., 2006), deletion of *sdaC* does not produce any significant growth or motility defects in *P. mirabilis* (Figure A9). Serine uptake and utilization in *P. mirabilis* via SdaA, SdaB, and SdaC resemble *E. coli* (Figure 2.1D). Therefore, we can use what is known in *E. coli* to develop tools to investigate the function of SdaC further.

Serine transport should contribute to the internal serine pool and could shift growth dynamics. Deleting the *E. coli* serine deaminases increases internal serine concentration, causing physiological effects, including growth defects in minimal medium and cell wall instability (Hama et al., 1990, 1991; X. Zhang et al., 2010; X. Zhang & Newman, 2008). Informed by these results, we made a *P. mirabilis* strain with a serine-dependent reporter phenotype. Removing the comparable serine deaminases (the *sdaA* and *sdaB* genes) resulted in strain ACH04 (Figure A9). Consistent with serine toxicity observed in *E. coli*, this strain had the expected delayed growth in a minimal medium containing exogenous serine but not in the absence of serine (Figure 2.1E). Swarm colony cells were also lysed and subjected to LC-MS analysis to measure serine

concentrations. ACH04 cells contained about 400-fold higher intracellular serine than the wild-type parent (Figure 2.1F). They also displayed less swarm expansion than wildtype (Figure 2.1G). Based on subsequent experiments (Figure A9) and earlier reports (Little et al., 2019; X. Zhang et al., 2010), serine-induced cell wall instability in strain ACH04 probably causes the reduced growth rate and swarm colony expansion in nutrient-rich conditions. Nonetheless, swarm expansion in the ACH04 strain background can serve as a readout for internal serine levels.

We reasoned that if SdaC were a serine transporter in *P. mirabilis*, then deletion of *sdaC* would alleviate defects in strain ACH04. ACH05, the engineered strain containing a *sdaC* deletion in the ACH04 background, showed a partial rescue of growth in a minimal medium with serine (Figure 2.1E). Internal serine concentrations decreased by ~ 30-fold compared to ACH04 (Figure 2.1F), while swarm expansion increased by roughly 8-fold (Figure 2.1G). Adding back SdaC to ACH05, through plasmid-based expression, reproduced the growth and colony expansion defects (Figure 2.1E, G). However, two other predicted serine transporters, *sstT* and *yhaO* (Connolly et al., 2016; Ogawa et al., 1998), could play a role. To examine any potential contributions to internal serine during swarming, we individually removed *sstT* and *yhaO* from the ACH04 background. Neither strain showed swarm colony expansion and instead looked like the ACH04 parental strain (Figure 2.1H, A12). Thus, SdaC is indeed a serine transporter in *P. mirabilis* and is dominant during collective motility.

Of the possible ways in which SdaC could function in self-recognition, two seemed most probable. SdaC serine transport could regulate downstream self-recognition signaling by modulating internal serine levels. Alternatively, a specific

conformation of SdaC could be a required binding interface. Indeed, point mutations in

SdaC already hinted at a molecular mechanism for its function in signal transduction.

Two independent full-length SdaC disruptions emerged from the suppressor screen:

G328V and G332R (Figure 2.1B). These residues sit in a predicted interface that

stabilizes the open conformation in LeuT, a similar transport protein [Figure 2.2A,

(Krishnamurthy & Gouaux, 2012)]. In combination with the assay toolkit, additional

reduced-function point mutations could reveal SdaC's role in self recognition.

A structural model for SdaC resembles LeuT-fold proteins, providing a template

for generating variants with biased conformations. We used the structure prediction

program I-TASSER (Roy et al., 2010; Yang et al., 2015; Y. Zhang, 2008) to make this

model based on the primary amino acid sequence and solved structures of similar

proteins (Figure 2.2A). LeuT-fold transporters sample at least three conformations: open

(outward-facing), closed (inward-facing), and an intermediate state (Bozzi et al., 2019;

Krishnamurthy & Gouaux, 2012). A V222W mutation would reasonably bias the SdaC

protein to the open conformation (Figure 2.2A-B); an equivalent conversion in NRAMP

favors the open conformation (Bozzi et al., 2019)]. The mutations in the suppressor

strains, G328V and G332R, are predicted to bias SdaC to the closed conformation

[Figure 2.2A-B, (Krishnamurthy & Gouaux, 2012)]. Altering either of two residues (I115A

and H325A) in the substrate-binding site would allow non-specific transport of additional

amino acids (Figure 2.2A-B) while continuing to sample both open and closed

conformations. Each mutation was introduced independently into *sdaC* expressed from

its native promoter on a plasmid. The engineered variants were visualized using an N-

terminal mCherry fusion comparable to the wildtype and localized to the *P. mirabilis*

Figure 2.2: SdaC's conformation determines serine transport and IdsD signaling.
A) PyMol figures of a structural model for *P. mirabilis* SdaC from I-TASSER (C-score = -0.82, TM-score = 0.61±0.14, RMSD = 8.9±4.6 Å). Targeted residues are labeled. B) Predicted conformational bias and serine transport changes are shown as a protein cartoon for each mutation: V222W is open-biased, I115A and H325A are non-specific in substrate affinity, G328R and G332V are closed-biased. Summary of (B-E) results are shown on the right with a check (success) or cross (failure). C) Growth curves of each strain, expressed from the native promoter in the ACH05 [BB2000 Δ(*sdaA*, *sdaB-sdaC*)] strain background, in minimal medium (left) or minimal medium plus 10 mM L-serine (right) for three biological replicates. ACH05 empty vector and ACH05 pSdaC data were copied from Figure 2.1E and colored in red for comparison. D) Swarm radius measured from assay of same strains from (C). ACH05 empty vector and ACH05 pSdaC are copied from Figure 2.1G and colored in red for comparison. Blue cell icon indicates that intracellular serine levels are high in the ACH05 background when SdaC is active, leading to restricted swarm expansion. E) Swarm radius measured from swarm assay of each mutant expressed from the native promoter in the ACH01 [BB2000 Δ(*idsE*, *sdaC*)] strain background. ACH01 empty vector and ACH01 pSdaC are copied from Figure 2.1C and colored in red for comparison. Red cell icon indicates that IdsD and downstream recognition signaling are active in the ACH01 background when SdaC is active, leading to restricted swarm expansion.
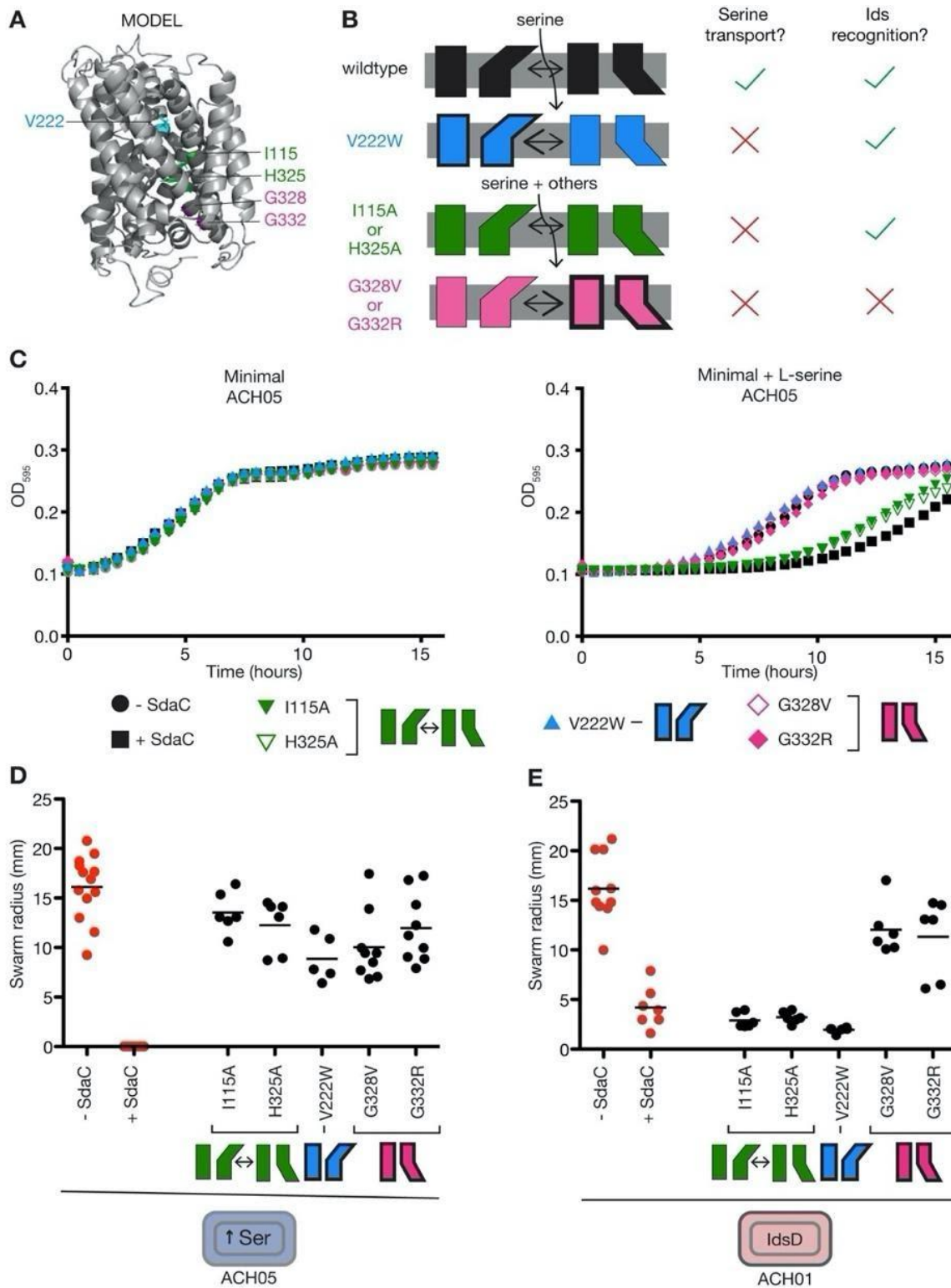
Figure 2.2: SdaC's conformation determines serine transport and IdsD signaling. (Continued)

cell envelope (Figure A10). The mutant strains provide molecular levers to distinguish between contributions to serine transport versus self recognition. Bias toward an open or closed conformation should restrict serine transport compared to that of the wildtype and the mutant that transports non-specifically. Therefore, we tested transport function in the ACH05 background, which lacks the serine deaminases. All strains grew equivalently in minimal medium (Figure 2.2C). As shown earlier, ACH05 has lowered internal serine and grows like wild-type in a minimal medium with exogenous serine (Figure 2.1E). The addition of transgenic SdaC resulted in attenuated growth (Figure 2.1E). The open-biased (V222W) and closed-biased (G328V, G332R) variants grew like ACH05 in minimal medium plus excess serine (Figure 2.2C). However, the non-specific variants (I115A, H325A) showed attenuated growth in minimal medium plus excess serine, much like ACH05 pSdaC (Figure 2.2C). This reduced liquid growth did not translate to altered swarm expansion. All mutant strains expanded beyond a radius of 10 mm, similar to ACH05, instead of being restricted as observed in ACH05 pSdaC (Figure 2.2D). These results support that serine transport is not itself required for swarm colony expansion under these conditions.

Suppose self-recognition relies on SdaC-mediated serine transport. In that case, the conformations biased to open or closed should prevent Ids-mediated recognition signaling, allowing swarm colonies of BB2000 Δ*idsE* (all non-self cells) to expand. To test this hypothesis, we introduced each SdaC variant into the ACH01 background (Figure 2.1C). The closed-biased variants (G328V and G332R) showed increased swarm colony expansion (Figure 2.2E), consistent with these mutations emerging from the original suppressor screen (Figure 2.1A-B). By contrast, the open-biased (V222W)

38

and non-specific variants (I115A and H325A) exhibited restricted swarm expansion

comparable to that of ACH01 pSdaC, which is the wild-type protein (Figure 2.2E).

Altogether, these results indicate that self-recognition signaling requires for SdaC to

sample the open conformation, but not transport serine. These SdaC functions are

distinct and overlapping.

The natural sequence variation among similar proteins provides an avenue for

understanding what molecular aspects of SdaC might be critical for self recognition.

SdaC from *E. coli* (SdaC-Ecol) is a diverged ortholog whose function is understood

(Shao et al., 1994). As stated earlier, SdaC is in a broader conserved pathway with

SdaA and SdaB, apparently shared between *E. coli* and *P. mirabilis.* On the other hand,

YhaO, another predicted serine/H+ symporter in *P. mirabilis*, is a distant homolog. In *E.*

coli, YhaO transports primarily D-serine, and to a lesser degree, L-serine (Connolly et

al., 2016). YhaO (*P. mirabilis*) shares much less sequence identity with SdaC-Pmir than

SdaC-Ecol; yet, conserved residues are visible throughout the protein, especially in the

predicted transmembrane domains (Figure 2.3). I-TASSER-based predictions for the

structures of SdaC-Ecol and YhaO also have a LeuT fold and share the same I-

TASSER templates as SdaC-Pmir (Figure A14). SdaC-Pmir, SdaC-Ecol, and YhaO

share sequence similarities and predicted tertiary structures.

Suppose sampling an open conformation is the critical molecular mechanism for

SdaC's function in self recognition. In that case, SdaC-Ecol and YhaO should be able to

replace SdaC in *P. mirabilis*. To interrogate this hypothesis, we expressed SdaC-Ecol

and YhaO from an inducible promoter in the ACH05 background. The resultant strains

showed growth defects in minimal medium with added serine but not in minimal
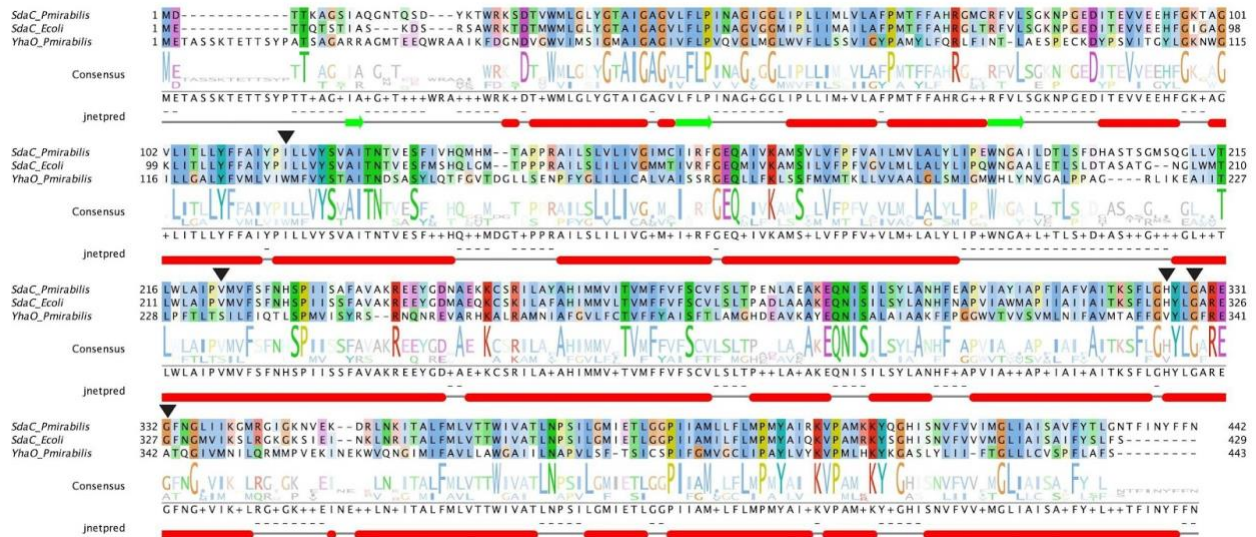
Figure 2.3: Sequence alignment of SdaC orthologs from *P. mirabilis* and *E. coli* and the homolog YhaO from *P. mirabilis.*
Aligned to each other are SdaC from *P. mirabilis* BB2000 (BB2000_0742), SdaC from *E. coli* K-12 MG1655, and YhaO from BB2000 (BB2000_2747) using Clustal Omega (Madeira et al., 2019; Sievers et al., 2011). The alignment was modified in Jalview (Waterhouse et al., 2009) to show the consensus sequence and logo plot as well as secondary structure prediction using JPred (red ovals for alpha-helices and green arrows for beta-sheets). Targeted residues are labeled with black arrowheads.

medium (Figure 2.4A). Therefore, both SdaC-Ecol and YhaO are sufficient to substitute for SdaC's serine transport function in *P. mirabilis*. Next, SdaC-Ecol and YhaO were expressed from an inducible promoter in the non-self ACH01 background and subjected to the swarm expansion assay. Strains producing SdaC-Pmir or SdaC-Ecol showed similarly restricted swarm expansion (Figure 2.4B). The strain producing YhaO expanded more fully (Figure 2.4B). Despite the sequence divergence, only SdaC-Ecol could substitute for SdaC and reproduce self recognition.

Theoretically, YhaO might not sufficiently sample an open conformation, so we introduced a point mutation to bias the protein. We constructed an open-biased S234W variant of YhaO analogous to the V222W variant of SdaC. Unlike wild-type YhaO, there was no growth defect for the open-biased variant expressed in ACH05 when grown in a minimal medium with serine (Figure 2.4C). This disrupted serine transport is consistent with the results of the open-biased SdaC variant (Figure 2.2C). We then moved the open-biased variant into the ACH01 background. The resultant strain showed swarm expansion similar to wild-type YhaO (Figure 2.4D). YhaO in the open conformation is sufficiently different from SdaC that it does not work in self recognition (Figure 2.4D). While SdaC's transport function is conserved among similar proteins, there is specificity to its conformation and sequence needed for self-recognition signal transduction.
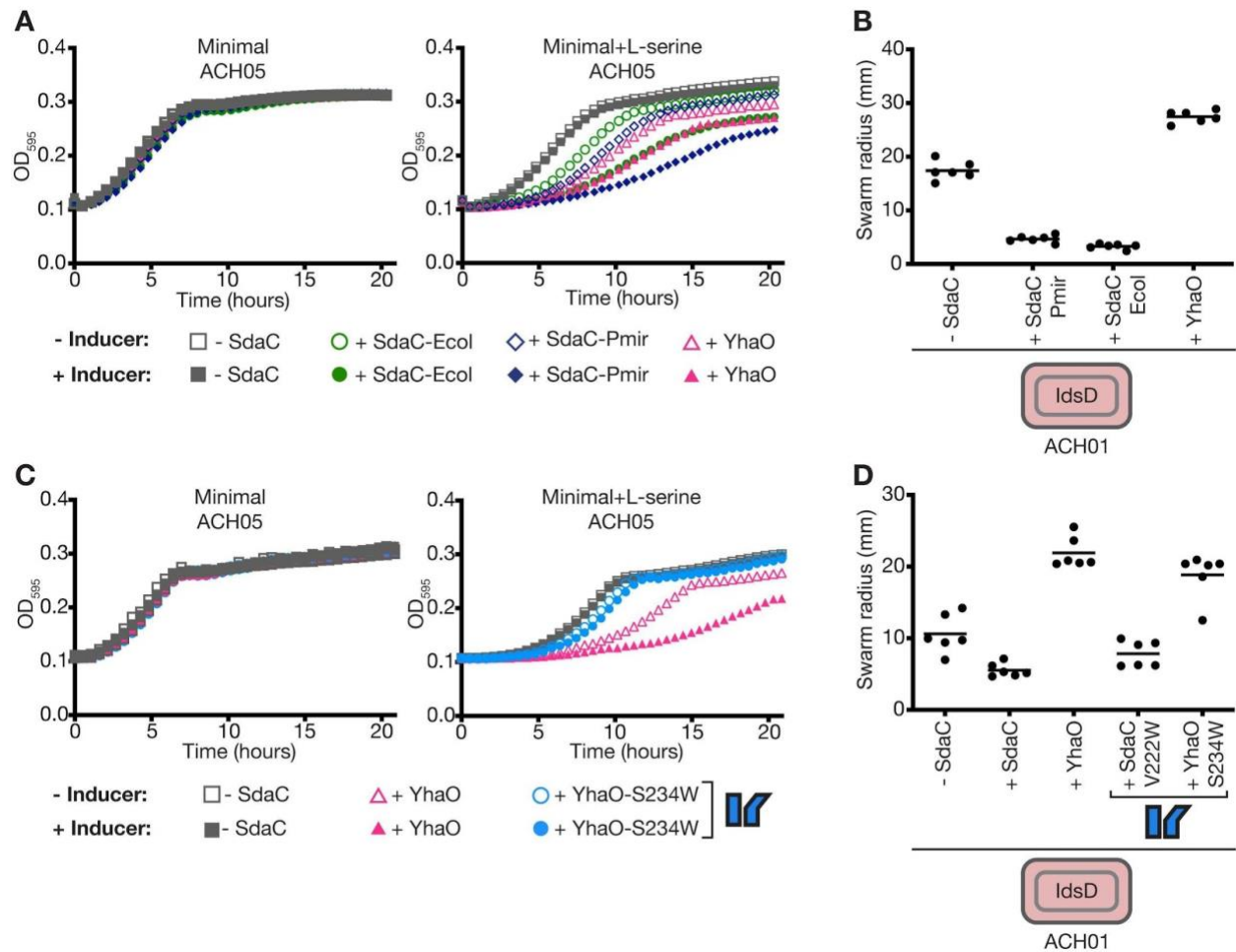
Figure 2.4: SdaC homologs share a similar predicted structure and function but only the SdaC ortholog from *E. coli* can complement in self recognition.
A) Growth curve of ACH05 [BB2000 Δ(*sdaA, sdaB-sdaC*)] expressing an empty vector, pTet-SdaC-Pmir, pTet-SdaC-Ecol, or pTet-YhaO without inducer (empty icons) or with inducer (filled icons) in minimal medium (left) and minimal medium plus 10mM L-serine (right). Mean of six biological replicates shown. B) Swarm radius measured from swarm assay of ACH01 [BB2000 Δ(*idsE, sdaC*)] carrying empty vector, pTet-SdaC-Pmir, pTet-SdaC-Ecol, or pTet-YhaO. Red cell icon indicates that IdsD and downstream recognition signaling are active in the ACH01 strain background when SdaC is active, leading to restricted swarm expansion. C) Growth curve of ACH05 expressing an empty vector, pTet-YhaO, or pTet-YhaO-S234W (open-biased) with inducer (filled icons) or without inducer (empty icons) in minimal medium (left) and minimal medium plus 10mM L-serine (right). Mean of three biological replicates shown. D) Swarm radius measured from swarm assay of ACH01 carrying empty vector, pTet-SdaC-Pmir, pTet-SdaC-Pmir-V222W (open-biased), pTet-YhaO, or pTet-YhaO-S234W (open-biased). Red cell icon indicates that IdsD and downstream recognition signaling is active in the ACH01 strain background when SdaC is active, leading to restricted swarm expansion.

Discussion

SdaC, a serine transporter, moonlights in the self-recognition signaling pathway. The transduced self-identity signal, IdsD, is a type VI secretion substrate predicted to localize to a recipient cell's inner membrane (Cardarelli et al., 2015; Zepeda-Rivera et al., 2018). Though inner membrane transporters are not known as receptors for type VI secretion substrates, we propose that SdaC functions as a receptor to promote IdsD insertion. Consistent with this hypothesis, disrupting SdaC also provides resistance to microcin V and phage C1 (Gérard et al., 2005; Likhacheva et al., 1996). Multiple inner membrane proteins are nutrient transporters and receptors for incoming proteins. For example, the contact-dependent inhibition protein CdiA requires specific inner membrane transporters for proposed insertion (Ruhe et al., 2017; Willett et al., 2015). Based on work in CdiA, a predominant hypothesis is that the membrane protein's functions work independently: nutrient transport versus translocation of protein. Our data expands this prior model. For SdaC, self recognition and nutrient transport are interdependent, likely affecting the protein's evolutionary trajectory.

Cells need SdaC to sample an open conformation for either self recognition or serine transport to occur. Like CdiA receptors (Willett et al., 2015), self-recognition signaling and collective motility do not specifically require serine transport (Figure 2.2D-E). Removing SdaC allows cells to bypass Ids-mediated self recognition (Figure 2.1C). Non-self populations regained collective motility when SdaC was not functional (Figure 2.1A-C, Figure 2.2E). Mutations that biased SdaC to an open conformation were sufficient to permit Ids-mediated self recognition regardless of serine transport (Figure 2.2E). However, the open conformation is only accessible when SdaC undergoes the

conformational dynamics needed for active transport. These two functions of SdaC are distinct but not independent.

High internal serine concentrations are toxic, inducing cell envelope stress and fitness defects of slower growth and no swarming (Figures 2.1E-G, 2.2C-E, A9). Stopping serine transport by deleting SdaC rescues serine toxicity in mutant strains lacking serine deaminases (Figure 2.1E-G). However, deleting the two other serine transporters, SstT and YhaO, does not relieve serine toxicity during swarming (Figure 2.1H). Therefore, internal serine levels, partially controlled by SdaC activity, are important during collective motility. And it is during this collective motility that self recognition occurs (Tipping & Gibbs, 2019). Expanding upon proposals for phage receptors, the coupling of SdaC functions may limit the emergence of mutations in both pathways.

Ecological context could constrain SdaC evolution. In bacteriocin and phage receptors, the local ecology and associated fitness trade-offs impact the emergence of intersectional mutations (Feldgarden & Riley, 1999; Inglis et al., 2016; Mangalea & Duerkop, 2020; Meaden et al., 2015). Serine is a crucial metabolite for urinary tract and gut pathogens (Barroso-Batista et al., 2020; Brauer et al., 2019; Connolly et al., 2016; Kitamoto et al., 2019; Velayudhan et al., 2004). Moreover, serine homeostasis is vital. Elevated internal serine poisons cells, leading to growth defects and susceptibility to cell envelope stress [Figure A9, (X. Zhang et al., 2010; X. Zhang & Newman, 2008)]. Too little serine starves cells of a significant amino acid. SdaC is the critical serine transporter during swarming (Figure 2.1E-H), a behavior correlated with disease (Armbruster et al., 2018; Kearns, 2010; Schaffer & Pearson, 2015). Further, the SdaA,

SdaB, and SdaC serine uptake pathway holds for distantly related species, suggesting evolutionary conservation and importance (Figure 2.3, 2.4A). Our results support that SdaC is a bifunctional conserved molecular interface, constrained by *Proteus*' ecology.

SdaC structure may limit plasticity and exploration of sequence space in buried regions. Often, exposed loops of receptor proteins are the main interaction interface based on outer membrane proteins co-opted by phages and toxins (Chatterjee & Rothenberg, 2012; Kleanthous, 2010; Ruhe et al., 2013). However, this model does not appear to hold for SdaC from *E. coli* and *P. mirabilis*. Sequences for the exposed periplasmic loops are ~ 65% identical (Figure 2.3). At the same time, the transmembrane regions are ~ 86% identical (Figure 2.3). Yet, *E. coli* SdaC can function in self recognition (Figure 2.4B) and transport (Figure 2.4A). By contrast, the *P. mirabilis* YhaO protein also aligns with SdaC (Figure 2.3) and transports serine (Figure 2.4A). However, YhaO cannot substitute for self recognition, even when biased into an open conformation (Figure 2.4B-D). Our data suggest that the membrane-localized pocket, predicted to be exposed conditionally in the open conformation, is the interaction interface (Figure 2.2). SdaC may follow conventional evolution ideas for integral membrane helical proteins, which are predicted to evolve slowly in buried regions due to molecular constraints (Oberai et al., 2009). Multi-conformation proteins are further constrained (Sharir-Ivry & Xia, 2017). The synergy between ecological fitness and structural constraints could slow the rate of SdaC sequence changes.

SdaC conservation can potentially regulate IdsD sequence drift to preserve signal fidelity. We have discussed the constraints on SdaC, but what constraints exist for the identity signal, IdsD? Self-recognition genes contain polymorphic regions (self-

identity barcodes) that can serve as a proxy for relatedness (de Oliveira et al., 2019). A

dominant model is that evolution of self-recognition proteins is constrained by

interactions with themselves or other self-recognition proteins, driving variation in their

interaction interface (Cardarelli et al., 2015; Hirose et al., 2017; Pathak et al., 2013).

IdsD does bind other self-recognition proteins such as its chaperone IdsC (Zepeda-

Rivera et al., 2018) and its partner recognition protein IdsE (Cardarelli et al., 2015).

Extending this model, we hypothesize that SdaC, which is not a recognition protein, acts

as an additional source of selective pressure. IdsD requires SdaC in a sequence-

specific manner. Although predicted to have a similar structure and function, YhaO

cannot replace SdaC for recognition (Figure 2.4). For uptake, IdsD must retain

compatibility with SdaC. Therefore, SdaC potentially acts as a bottleneck for signal

transduction in the receiving cell.

Molecular crosstalk mirrors observed interactions between nutrient availability,

collective behavior, and self recognition in many organisms. Collective behaviors are

associated with nutrient limitation in other microbes (Kundert & Shaulsky, 2019; Wall,

2014), fungi (Gonçalves et al., 2020), and plants (Palmer et al., 2016). Collective

behaviors can allow for sharing of nutrients and promote developmental processes such

as fruiting body formation. Self-recognition signaling allows preferential collective action

with kin, an advantage during nutrient limitation. Functional coupling between self

recognition and organism-relevant pathways such as nutrient transport, as shown for

SdaC, could constrain identity signal evolution. For collective behaviors, nutrient uptake

is a crucial regulator. Still, for other contexts such as self-avoidance and syncytial

fusion, there may be other core proteins that are evolutionarily constrained. Conserved

non-recognition proteins might also anchor other self-recognition proteins. By exploring

external interactions in the multi-level context of the organism, population, and

environment, we gain a better understanding of the different constraints on the evolution

of self-recognition genes.

Materials and Methods

Bacterial strains and media

The strains and plasmids used in this study are described in Table 2.1. *P. mirabilis*

strains were maintained on low swarm (LSW) agar (Belas et al., 1991). CM55 blood

agar base agar (Oxoid, Basingstoke, England) was used for swarm-permissive nutrient

plates. Overnight cultures of all strains were grown at 37°C in LB broth under aerobic

conditions. For growth curve assays, cells were grown in minimal medium [M9 salts (3

g/L KH_2PO_4, 6.8 g/L Na_2HPO_4, 0.5 g/L NaCl, 1.0 g/L NH_4Cl), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$,

0.2% glucose] supplemented with 10mM L-serine (VWR, Beantown chemical,

BT128350) when stated. Kanamycin (Corning, Corning, NY) was used at a

concentration of 35 µg/ml for plasmid maintenance and was added to swarm and

growth media when appropriate. Other antibiotics were used at the following

concentrations for transforming plasmids into *P. mirabilis*: 15 µg/ml tetracycline

(Amresco Biochemicals, Solon, OH), and 25 µg/ml streptomycin (Sigma-Aldrich, St.

Louis, MO). Anhydrotetracycline (Sigma-Aldrich, St. Louis, MO) was used to induce

gene expression from the Tet promoter at a concentration of 10 nM in the medium when

stated.

Random mutagenesis of IdsE and spontaneous suppressor collection

Plasmids pIdsBB-IdsE-mut1 and pIdsBB-IdsE-mut2 were constructed by amplifying the

idsE gene using oAC006 and oAC007 (Table A9) from the pIdsBB expression system

containing a C-terminal GFPmut2 fusion (Gibbs et al., 2008) using error-prone PCR with

the GeneMorph II Random Mutagenesis Kit (Agilent, Santa Clara, CA) and ligated back

into the same pIdsBB expression vector using the restriction enzymes SacI and BamHI.

Swarm-capable spontaneous mutants of the BB2000 Δ*ids* strains carrying pIdsBB-IdsE-mut1 or pIdsBB-IdsE-mut2 were isolated. Starting from frozen stocks, stable recovery of swarm expansion was verified. Plasmids were miniprepped and retransformed into the BB2000 Δ*ids* strain background to screen for plasmid-based suppressors, which were removed from further analysis. Boundary assays were performed to screen for defects in production or secretion of IdsD, and these suppressors were removed from further analysis. Seven of the remaining suppressor mutants were whole-genome sequenced (suppressors 1-7, Table A8).

The second set of swarm-capable spontaneous mutants were isolated from BB2000 Δ*idsE* (Zepeda-Rivera et al., 2018) carrying pTet-IdsE-mut3 and from ACH06 [BB2000 Δ(*sdaA*, *sdaB*, *idsE*)]. The *idsE*-mut3 sequence was amplified from pIdsBB-IdsE-T246A-S247A-T248A using oAC006 and oAC041 (Table A9) and ligated into the pTet vector (Zepeda-Rivera et al., 2018) using enzymes SacI and Bsu36I. Starting from frozen stocks, stable recovery of swarm expansion was verified. For suppressors derived from BB2000 Δ*idsE* pTet-IdsE-mut3, the promoter and gene were sequenced using Sanger sequencing (Genewiz, Inc., South Plainfield, NJ) to confirm causative mutations were chromosomal. Boundary assays were performed to screen for defects in production or secretion of IdsD, and these suppressor mutant strains were excluded from further analysis. Five of the remaining suppressor strains were whole-genome sequenced (suppressors 8-12, Table A8).

Whole genome sequencing and variant calling

For the first round of sequencing (suppressors 1-7, Table A8), isolates were subjected to phenol-chloroform extractions to isolate genomic DNA (gDNA). gDNA was sheared using a Covaris S220 system (Covaris, Woburn, MA), and a library for whole-genome sequencing was prepared using the PrepX ILM DNA library kit (Takara Biosciences, Mountain View, CA) for the Apollo 324 next-generation sequencing (NGS) library prep system (Takara Biosciences). The library was sequenced as 75-bp paired-end reads using an Illumina NextSeq 500 system (Illumina, San Diego, CA). The Bauer Core Facility at Harvard University performed all genome sequencing.

For subsequent sequencing of the second set of mutants (suppressors 8-12, Table A8), gDNA was isolated as above, but library prep using KAPA HyperPrep kit (Roche, Wilmington, MA) was performed by the Bauer Core Facility. The library was sequenced as 150-bp paired-end reads using an Illumina NextSeq 500 system (Illumina, San Diego, CA) by the Bauer Core Facility. Breseq (Deatherage & Barrick, 2014) was used to perform variant calling of Illumina NextSeq reads against the BB2000 reference genome (GenBank accession no. CP004022).

Plasmid construction

Restriction digestion was performed using restriction enzymes described (New England BioLabs, Ipswich, MA). Ligations were resolved in OneShot Omnimax2 T1R competent cells (Thermo Fisher Scientific, Waltham, MA) or SM10λpir (Simon et al., 1983). The resultant plasmids were confirmed by Sanger sequencing (Genewiz, Inc., South Plainfield, NJ), and correct resultant plasmids were then transformed into *P. mirabilis* as described previously (Gibbs et al., 2008) using *E. coli* conjugative strain MFDpir

(Ferrières et al., 2010). Table A9 contains the nucleotide sequences for listed primers, all of which contain the prefix "oAC."

For pSdaC, the *sdaC* gene (BB2000_0742) including ~1kb upstream putative promoter region was amplified using oAC072 and oAC071 and ligated into the pTet vector (Zepeda-Rivera et al., 2018) using restriction enzymes NheI and PshAI to construct pSdaC. The empty vector is derived from pSdaC, containing no promoter or gene of interest but the rest of the backbone including kanamycin resistance and origin of replication is intact.

For pTet-SdaC-Pmir, the *sdaC* gene was amplified using oAC208 and oAC071 from pSdaC. Next, the sequence encoding the Tet promoter was amplified from pTet-FLAG-IdsE-mut3 with the addition of a 3XFLAG tag using oAC068, oAC196, oAC197, oAC198, and oAC207. The Tet promoter and *sdaC* gene were joined using overlap extension PCR with oAC068 and oAC071. The insert was ligated into the pTet vector using NheI and PshAI. For pTet-SdaC-Ecol, the *sdaC* gene was amplified from *E. coli* K-12 MG1655 using oAC177 and oAC178. Next, the Tet promoter sequence was amplified from the pTet vector using oAC68 and oAC185. The *sdaC* and Tet promoter sequences were joined using overlap extension PCR with oAC68 and oAC178. The insert was ligated into the pTet vector using NheI and Bsu36I. For pTet-YhaO, the *yhaO* gene (BB2000_2747) was amplified from *P. mirabilis* BB2000 using oAC187 and oAC188. Next, the Tet promoter sequence was amplified from the pTet vector using oAC68 and oAC186. The *yhaO* gene and Tet promoter sequence were joined using

overlap extension PCR with oAC68 and oAC188. The insert was ligated into the pTet vector using NheI and Bsu36I.

SdaC variants were constructed by amplifying the *sdaC* gene from pSdaC using complementary primers containing the mutation along with oAC072 (SdaC native promoter) or oAC068 (Tet promoter) and oAC071 in overlap extension PCR. The *sdaC* gene containing the mutation was ligated back into the pSdaC vector using enzymes NheI and PshAI. The primers for each mutation were oAC264 and oAC265 for I115A, oAC268 and oAC269 for H325A, oAC293 and oAC294 for V222W, oAC118 and oAC119 for G332R, and oAC227 and oAC228 for G328V. The YhaO variant was constructed using complementary primers oAC314 and oAC315 containing the S324W mutation along with flanking primers oAC68 and oAC188 in overlap extension PCR. The insert was ligated into pTet-YhaO with enzymes NheI and Bsu36I.

Strain construction

All chromosomal deletions were performed as described earlier using pKNG101-derived suicide vectors (Saak & Gibbs, 2016).

For strain ACH01 [BB20000 Δ(*idsE*, *sdaC*)], 500bp regions adjacent on either side to *sdaC* (BB2000_0742) with restriction sites were amplified using oAC046-049, and ligated into pKNG101 using restriction enzymes ApaI and SpeI. The resulting vector was mated into BB2000 Δ*idsE*. Matings were subjected to antibiotic selection on LSW agar (with 15 g/ml Tet and 25 g/ml Strep). Candidate strains were subjected to sucrose counterselection as described (Sturgill et al., 2002). Double recombinants were

confirmed by PCR of the surrounding regions using oAC113 and oA114. ACH01 was confirmed by whole genome sequencing.

For strain ACH04 [BB2000 Δ(*sdaA*, *sdaB*)], the regions flanking the *sdaB* (BB2000_0741) gene were amplified using overlap extension PCR with oAC050-053 and ligated into pKNG101 using enzymes ApaI and SpeI. The resulting vector was mated into BB2000 and subjected to sucrose selection. Double recombinants were confirmed by colony PCR of the surrounding region using oAC113 and oAC114. The regions flanking the *sdaA* gene (BB2000_1697) were amplified using overlap extension PCR with oAC141-144 and ligated into pKNG101 using enzymes ApaI and SpeI. The resulting vector was mated into BB2000 Δ*sdaB* and subjected to sucrose selection. Double recombinants were confirmed by PCR of the surrounding region using oAC153 and oAC154.

For strain ACH05 [BB2000 Δ(*sdaA*, *sdaB*-*sdaC*)], the *sdaA* deletion vector used to construct ACH04 was mated into BB2000 and confirmed by PCR of the *sdaA* region using oAC153 and oAC154. The regions flanking the *sdaB*-*sdaC* genes were amplified using overlap extension PCR with oAC46, oAC53, oAC54, and oAC55 and ligated into pKNG101 using enzymes ApaI and SpeI. The resulting vector was mated into Δ*sdaA* and subjected to sucrose selection. Double recombinants were confirmed by PCR of the surrounding region using oAC113 and oAC114. Strain ACH05 was confirmed by whole genome sequencing.

For strain ACH04 Δ*sstT*, the regions flanking the *sstT* gene (BB2000_0146) were amplified using overlap extension PCR with oAC145-oAC148 and ligated into pKNG101

using enzymes ApaI and SpeI. The resulting vector was mated into ACH04 and subjected to sucrose selection. ACH04 Δ*sstT* was confirmed using PCR amplification of the surrounding region using oAC156 and oAC157. For strain ACH04 Δ*yhaO*, the regions flanking the *yhaO* gene (BB2000_2747) were amplified using overlap extension PCR with oAC161-oAC164 and ligated into pKNG101 using enzymes ApaI and SpeI. The resulting vector was mated into ACH04 and subjected to sucrose selection. ACH04 Δ*yhaO* was confirmed using PCR amplification of the surrounding region using oAC179 and oAC180.

Growth curve

Overnight cultures were normalized to an optical density at 600 nm (OD600) of 0.1 in minimal medium [M9 salts (3 g/L KH_2PO_4, 6.8 g/L Na_2HPO_4, 0.5 g/L NaCl, 1.0 g/L NH_4Cl), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% glucose] supplemented with 10mM L-serine (VWR, Beantown chemical, BT128350) when stated. Both were supplemented with kanamycin for plasmid maintenance when appropriate. Normalized cultures were grown overnight at 37°C, with periodic shaking, in a Tecan Infinite 200 PRO microplate reader (Tecan, Männedorf, Switzerland).

Swarm expansion

Overnight cultures were normalized to an OD600 of 0.1, and swarm-permissive nutrient plates supplemented with kanamycin were inoculated with 2 μl of normalized culture in the center. Plates were incubated at room temperature for two days, and the radii of actively migrating swarms starting from the edge of the inoculum were measured using Fiji (ImageJ) (Schindelin et al., 2012).

LC-MS of L-serine in *P. mirabilis* cells

P. mirabilis cells were harvested by centrifugation from three swarm-permissive plates after incubation at 37°C for 16 to 20 hours. Cells were sequentially resuspended and centrifuged in 1ml of 100% LB, 80% LB, 60% LB, 40% LB, 20% LB, water (LC-MS grade water from Thermo Fisher Scientific, Waltham, MA). Cell pellets were flash frozen in liquid nitrogen and stored at -80°C. Pellets were resuspended in 1 ml cold LC-MS grade methanol (Sigma-Aldrich, St. Louis, MO) and lysed by vortexing for 10 minutes with cell disruptor beads (0.1-mm diameter; Electron Microscopy Sciences, Hatfield, PA). Lysate was transferred to an 8ml glass vial with an additional 1ml cold methanol rinse of the lysis tube. 4 mL of cold LC-MS grade chloroform (Sigma-Aldrich, St. Louis, MO) was added and samples were vortexed for 1 minute. 2 mL of LC-MS grade water containing 0.1 µL of MSK-A2-1.2 stable isotope-labeled amino acid standards (Cambridge Isotope Laboratories, Tewksbury, MA) was added before vortexing for 1 minute. After centrifuging for 10min at 3000rpm, 3.5 ml of the upper aqueous phase was transferred to a new glass vial and stored at -80°C. After removing the organic chloroform phase, the remaining interphase was dried completely before resuspending in Tris-buffered saline (50mM Tris-Cl, 150mM NaCl, pH 7.6) for protein quantification using a Bradford Assay (BioRad). Aqueous phase was evaporated under nitrogen flow and used for LC-MS, which was performed by the Small Molecule Mass Spectrometry Core Facility at Harvard University.

Bioinformatics

We used I-TASSER (Roy et al., 2010; Yang et al., 2015; Y. Zhang, 2008) to predict the

protein structure of SdaC-Pmir. Figures of structural models were made using PyMOL

(The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Amino acid

sequences in Figure 3 were aligned using Clustal Omega (Madeira et al., 2019; Sievers

et al., 2011). Sequence alignments were visualized in Jalview (Waterhouse et al.,

2009).

Table 2.1: Strains used in this study

Strain	Name in this study	Description	Reference	KAG#	AC#
Proteus mirabilis					
BB2000 wild-type strain	BB2000	wild-type *P. mirabilis* BB2000 strain	Belas et al., 1991	KAG 0001	AC52
BB2000 Δ*ids* carrying pIdsBBΔ*idsE*	Δ*ids* pIdsBBΔ*idsE*	BB2000 Δ*ids* carrying a vector expressing the *ids* locus from BB2000 but with *idsE* deleted (CCS06)	Saak & Gibbs, 2016	KAG 1923	AC21
BB2000 carrying empty vector	BB2000 empty vector	BB2000 carrying a plasmid without promoter-gene insert to confer antibiotic resistance	This study	KAG 4183	AC232
BB2000 Δ*idsE*	Δ*idsE*	BB2000 with a chromosomal *idsE* deletion	Zepeda-Rivera et al., 2018	KAG 3126	AC336
BB2000 Δ*idsE* carrying empty vector	Δ*idsE* empty vector	BB20000 Δ*idsE* carrying empty vector	This study	KAG 3739	AC104
BB2000 Δ(*idsE*, *sdaC*)	ACH01	BB20000 Δ*idsE* with a chromosomal *sdaC* (BB2000_0742) deletion	This study	KAG 3974	AC120
BB2000 Δ(*idsE*, *sdaC*) carrying empty vector	ACH01 empty vector	ACH01 carrying empty vector	This study	KAG 4187	AC236
BB2000 Δ(*idsE*, *sdaC*) carrying pSdaC	ACH01 pSdaC	ACH01 carrying a plasmid expressing *sdaC* from its native promoter	This study	KAG 4203	AC252
BB2000 Δ(*sdaA*, *sdaB*)	ACH04	BB2000 with a chromosomal deletion of *sdaA* (BB2000_1697) and *sdaB* (BB2000_0741)	This study	KAG 4303	AC338
BB2000 Δ(*sdaA*, *sdaB*) carrying empty vector	ACH04 empty vector	ACH04 carrying an empty vector	This study	KAG 4386	AC410

Table 2.1: Strains used in this study (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
BB2000 Δ(*sdaA*, *sdaB-sdaC*)	ACH05	ACH04 with a chromosomal deletion of *sdaC*	This study	KAG 4328	AC393
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying empty vector	ACH05 empty vector	ACH05 carrying an empty vector	This study	KAG 4388	AC466
BB2000 Δ(*sdaA*, *sdaB, sstT*)	ACH04 *ΔsstT*	ACH04 with a chromosomal deletion of *sstT* (BB2000_0146)	This study	KAG 4359	AC431
BB2000 Δ(*sdaA*, *sdaB, yhaO*)	ACH04 *ΔyhaO*	ACH04 with a chromosomal deletion of *yhaO* (BB2000_2747)	This study	KAG 4361	AC433
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pSdaC	ACH05 pSdaC	ACH05 carrying pSdaC	This study	KAG 4391	AC469
BB2000 Δ(*idsE*, *sdaC*) carrying pSdaC-I115A	ACH01 pSdaC-I115A	ACH01 carrying a modified pSdaC plasmid with I115A amino acid change	This study	KAG 4446	AC528
BB2000 Δ(*idsE*, *sdaC*) carrying pSdaC-H325A	ACH01 pSdaC-H325A	ACH01 carrying a modified pSdaC plasmid with H325A amino acid change	This study	KAG 4447	AC529
BB2000 Δ(*idsE*, *sdaC*) carrying pSdaC-V222W	ACH01 pSdaC-V222W	ACH01 carrying a modified pSdaC plasmid with V222W amino acid change	This study	KAG 4469	AC553
BB2000 Δ(*idsE*, *sdaC*) carrying pSdaC-G328V	ACH01 pSdaC-G328V	ACH01 carrying a modified pSdaC plasmid with G328V amino acid change	This study	KAG 4836	AC521
BB2000 Δ(*idsE*, *sdaC*) carrying pSdaC-G332R	ACH01 pSdaC-G332R	ACH01 carrying a modified pSdaC plasmid with G332R amino acid change	This study	KAG 4834	AC522

Table 2.1: Strains used in this study (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pSdaC-I115A	ACH05 pSdaC-I115A	ACH05 carrying a modified pSdaC plasmid with I115A amino acid change	This study	KAG 4448	AC530
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pSdaC-H325A	ACH05 pSdaC-H325A	ACH05 carrying a modified pSdaC plasmid with H325A amino acid change	This study	KAG 4449	AC531
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pSdaC-V222W	ACH05 pSdaC-V222W	ACH05 carrying a modified pSdaC plasmid with V222W amino acid change	This study	KAG 4468	AC552
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pSdaC-G328V	ACH05 pSdaC-G328V	ACH05 carrying a modified pSdaC plasmid with G328V amino acid change	This study	KAG 4394	AC472
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pSdaC-G332R	ACH05 pSdaC-G332R	ACH05 carrying a modified pSdaC plasmid with G332R amino acid change	This study	KAG 4392	AC470
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pTet-3xFLAG-SdaC	ACH05 pTet-SdaC-Pmir	ACH01 carrying a modified pSdaC plasmid with anhydrotetracycline inducible promoter and N-terminal 3xFLAG tag	This study	KAG 4840	AC684
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pTet-SdaC-Ecoli	ACH05 pTet-SdaC-Ecoli	ACH01 carrying a plasmid expressing *E. coli* K12 MG1655 *sdaC* from an inducible anhydrotetracycline promoter	This study	KAG 4838	AC683

Table 2.1: Strains used in this study (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pTet-YhaO	ACH05 pTet-YhaO	ACH01 carrying a plasmid expressing *yhaO* from an inducible anhydrotetracycline promoter	This study	KAG 4725	AC669
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pTet-YhaO-S234W	ACH05 pTet-YhaO-S234W	ACH01 carrying a plasmid expressing *yhaO* with S234W amino acid change from an inducible anhydrotetracycline promoter	This study	KAG 4740	AC685
BB2000 Δ(*idsE, sdaC*) carrying pTet-3xFLAG-SdaC	ACH01 pTet-SdaC-Pmir	ACH01 carrying a modified pSdaC plasmid with anhydrotetracycline inducible promoter and N-terminal 3xFLAG tag	This study	KAG 4474	AC558
BB2000 Δ(*idsE, sdaC*) carrying pTet-SdaC-Ecoli	ACH01 pTet-SdaC-Ecoli	ACH01 carrying a plasmid expressing *E. coli* K-12 MG1655 *sdaC* from an inducible anhydrotetracycline promoter	This study	KAG 4752	AC691
BB2000 Δ(*idsE, sdaC*) carrying pTet-YhaO	ACH01 pTet-YhaO	ACH01 carrying a plasmid expressing *yhaO* from an inducible anhydrotetracycline promoter	This study	KAG 4842	AC402
BB2000 Δ(*idsE, sdaC*) carrying pTet-3xFLAG-SdaC	ACH01 pTet-SdaC-Pmir-V222W	ACH01 carrying a modified pSdaC plasmid with anhydrotetracycline inducible promoter and N-terminal 3xFLAG tag and the V222W amino acid change	This study	KAG 4780	AC695

Table 2.1: Strains used in this study (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
BB2000 Δ(*idsE, sdaC*) carrying pTet-YhaO-S234W	ACH01 pTet-YhaO-S234W	ACH01 carrying a plasmid expressing *yhaO* with S234W amino acid change from an inducible anhydrotetracycline promoter	This study	KAG 4683	AC630
Escherichia coli					
One Shot Omnimax 2 T1R competent cells		Cloning strain for vectors	Thermo Fisher Scientific, Waltham, MA		
MFDpir		Mu-free mating strain for *P. mirabilis*	Ferrières et al., 2010		
SM10λpir		Mating strain for moving pKNG101 into *P. mirabilis*	Simon et al., 1983		

References

Armbruster, C. E., Mobley, H. L. T., & Pearson, M. M. (2018). Pathogenesis of *Proteus mirabilis* Infection. *EcoSal Plus*, *8*(1), 10.1128/ecosalplus.ESP-0009–2017. https://doi.org/10.1128/ecosalplus.ESP-0009-2017

Asfahl, K. L., & Schuster, M. (2017). Social interactions in bacterial cell–cell signaling. *FEMS Microbiology Reviews*, *41*(1), 92–107. https://doi.org/10.1093/femsre/fuw038

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology*, *2*(1), 2006.0008.

Barroso-Batista, J., Pedro, M. F., Sales-Dias, J., Pinto, C. J. G., Thompson, J. A., Pereira, H., Demengeot, J., Gordo, I., & Xavier, K. B. (2020). Specific Eco-evolutionary Contexts in the Mouse Gut Reveal *Escherichia coli* Metabolic Versatility. *Current Biology*, *30*(6), 1049-1062.e7. https://doi.org/10.1016/j.cub.2020.01.050

Belas, R., Erskine, D., & Flaherty, D. (1991). Transposon mutagenesis in *Proteus mirabilis*. *J Bacteriol*, *173*(19), 6289–6293.

Bozzi, A. T., Zimanyi, C. M., Nicoludis, J. M., Lee, B. K., Zhang, C. H., & Gaudet, R. (2019). Structures in multiple conformations reveal distinct transition metal and proton pathways in an Nramp transporter. *Elife*, *8*, e41124.

Brauer, A. L., White, A. N., Learman, B. S., Johnson, A. O., & Armbruster, C. E. (2019). D-Serine degradation by *Proteus mirabilis* contributes to fitness during single-species and polymicrobial catheter-associated urinary tract infection. *MSphere*, *4*(1).

Cardarelli, L., Saak, C., & Gibbs, K. A. (2015). Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding. *MBio*, *6*(3), e00251-15.

Chatterjee, S., & Rothenberg, E. (2012). Interaction of Bacteriophage l with Its *E. coli* Receptor, LamB. *Viruses*, *4*(11), 3162–3178. https://doi.org/10.3390/v4113162

Connolly, J. P., Gabrielsen, M., Goldstone, R. J., Grinter, R., Wang, D., Cogdell, R. J., Walker, D., Smith, D. G., & Roe, A. J. (2016). A Highly Conserved Bacterial D-Serine Uptake System Links Host Metabolism and Virulence. *PLoS Pathog*, *12*(1), e1005359. https://doi.org/10.1371/journal.ppat.1005359

de Jonge, P. A., Nobrega, F. L., Brouns, S. J., & Dutilh, B. E. (2019). Molecular and evolutionary determinants of bacteriophage host range. *Trends in Microbiology*, *27*(1), 51–63.

de Oliveira, J. L., Morales, A. C., Stewart, B., Gruenheit, N., Engelmoer, J., Brown, S. B., de Brito, R. A., Hurst, L. D., Urrutia, A. O., & Thompson, C. R. (2019). Conditional expression explains molecular evolution of social genes in a microbe. *Nature Communications*, *10*(1), 1–12.

Deatherage, D. E., & Barrick, J. E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. In *Engineering and analyzing multicellular systems* (pp. 165–188). Springer.

Feldgarden, M., & Riley, M. A. (1999). The phenotypic and fitness effects of colicin resistance in *Escherichia coli* K-12. *Evolution*, *53*(4), 1019–1027.

Ferrières, L., Hémery, G., Nham, T., Guérout, A.-M., Mazel, D., Beloin, C., & Ghigo, J.-M. (2010). Silent mischief: Bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *Journal of Bacteriology*, *192*(24), 6418–6427.

Fischer, M. S., & Glass, N. L. (2019). Communicate and fuse: How filamentous fungi establish and maintain an interconnected mycelial network. *Frontiers in Microbiology*, *10*, 619.

Gérard, F., Pradel, N., & Wu, L.-F. (2005). Bactericidal activity of colicin V is mediated by an inner membrane protein, SdaC, of *Escherichia coli*. *Journal of Bacteriology*, *187*(6), 1945–1950.

Gibbs, K. A., Urbanowski, M. L., & Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. *Science*, *321*(5886), 256–259. https://doi.org/10.1126/science.1160033

Gonçalves, A. P., Heller, J., Rico-Ramírez, A. M., Daskalov, A., Rosenfield, G., & Glass, N. L. (2020). Conflict, Competition, and Cooperation Regulate Social Interactions in Filamentous Fungi. *Annual Review of Microbiology*, *74*, 693–712.

Gruenheit, N., Parkinson, K., Stewart, B., Howie, J. A., Wolf, J. B., & Thompson, C. R. (2017). A polychromatic 'greenbeard' locus determines patterns of cooperation in a social amoeba. *Nature Communications*, *8*(1), 1–9.

Hama, H., Kayahara, T., Tsuda, M., & Tsuchiya, T. (1991). Inhibition of homoserine dehydrogenase I by L-serine in *Escherichia coli*. *The Journal of Biochemistry*, *109*(4), 604–608.

Hama, H., Shimamoto, T., Tsuda, M., & Tsuchiya, T. (1988). Characterization of a novel L-serine transport system in *Escherichia coli*. *Journal of Bacteriology*, *170*(5), 2236–2239.

Hama, H., Sumita, Y., Kakutani, Y., Tsuda, M., & Tsuchiya, T. (1990). Target of serine inhibition in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, *168*(3), 1211–1216.

Hampton, H. G., Watson, B. N., & Fineran, P. C. (2020). The arms race between bacteria and their phage foes. *Nature*, *577*(7790), 327–336.

Hirose, S., Chen, G., Kuspa, A., & Shaulsky, G. (2017). The polymorphic proteins TgrB1 and TgrC1 function as a ligand–receptor pair in *Dictyostelium* allorecognition. *Journal of Cell Science*, *130*(23), 4002–4012.

Inglis, R. F., Scanlan, P., & Buckling, A. (2016). Iron availability shapes the evolution of bacteriocin resistance in *Pseudomonas aeruginosa*. *The ISME Journal*, *10*(8), 2060.

Kearns, D. B. (2010). A field guide to bacterial swarming motility. *Nature Reviews Microbiology*, *8*(9), 634–644.

Kitamoto, S., Alteri, C. J., Rodrigues, M., Nagao-Kitamoto, H., Sugihara, K., Himpsl, S. D., Bazzi, M., Miyoshi, M., Nishioka, T., & Hayashi, A. (2019). Dietary l-serine confers a competitive fitness advantage to *Enterobacteriaceae* in the inflamed gut. *Nature Microbiology*, 1–10.

Kleanthous, C. (2010). Swimming against the tide: Progress and challenges in our understanding of colicin translocation. *Nature Reviews Microbiology*, *8*(12), 843–848.

Kostadinov, D., & Sanes, J. R. (2015). Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. *Elife*, *4*, e08964.

Krishnamurthy, H., & Gouaux, E. (2012). X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature*, *481*(7382), 469.

Kundert, P., & Shaulsky, G. (2019). Cellular allorecognition and its roles in *Dictyostelium* development and social evolution. *The International Journal of Developmental Biology*, *63*(8-9–10), 383.

Lefebvre, J. L., Kostadinov, D., Chen, W. V., Maniatis, T., & Sanes, J. R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature*, *488*(7412), 517–521.

Likhacheva, N. A., Samsonov, V. V., Samsonov, V. V., & Sineoky, S. P. (1996). Genetic control of the resistance to phage C1 of *Escherichia coli* K-12. *J Bacteriol*, *178*(17), 5309–5315.

Little, K., Austerman, J., Zheng, J., & Gibbs, K. A. (2019). Cell shape and population migration are distinct steps of Proteus mirabilis swarming that are decoupled on high-percentage agar. *Journal of Bacteriology*, *201*(11), e00726-18.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R., Potter, S. C., & Finn, R. D. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641.

Mangalea, M. R., & Duerkop, B. A. (2020). Fitness trade-offs resulting from bacteriophage resistance potentiate synergistic antibacterial strategies. *Infection and Immunity*.

Meaden, S., Paszkiewicz, K., & Koskella, B. (2015). The cost of phage resistance in a plant pathogenic bacterium is context-dependent. *Evolution*, *69*(5), 1321–1328.

Molumby, M. J., Keeler, A. B., & Weiner, J. A. (2016). Homophilic protocadherin cell-cell interactions promote dendrite complexity. *Cell Reports*, *15*(5), 1037–1050.

Nicoludis, J. M., Green, A. G., Walujkar, S., May, E. J., Sotomayor, M., Marks, D. S., & Gaudet, R. (2019). Interaction specificity of clustered protocadherins inferred from sequence covariation and structural analysis. *Proceedings of the National Academy of Sciences*, *116*(36), 17825–17830.

Nicoludis, J. M., Lau, S.-Y., Schärfe, C. P., Marks, D. S., Weihofen, W. A., & Gaudet, R. (2015). Structure and sequence analyses of clustered protocadherins reveal antiparallel interactions that mediate homophilic specificity. *Structure*, *23*(11), 2087–2098.

Nicoludis, J. M., Vogt, B. E., Green, A. G., Schärfe, C. P., Marks, D. S., & Gaudet, R. (2016). Antiparallel protocadherin homodimers use distinct affinity-and specificity-mediating regions in cadherin repeats 1-4. *Elife*, *5*, e18449.

Noonan, J. P., Li, J., Nguyen, L., Caoile, C., Dickson, M., Grimwood, J., Schmutz, J., Feldman, M. W., & Myers, R. M. (2003). Extensive linkage disequilibrium, a common 16.7-kilobase deletion, and evidence of balancing selection in the human protocadherin α cluster. *The American Journal of Human Genetics*, *72*(3), 621–635.

Oberai, A., Joh, N. H., Pettit, F. K., & Bowie, J. U. (2009). Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proceedings of the National Academy of Sciences*, *106*(42), 17747–17750.

Ogawa, W., Kim, Y. M., Mizushima, T., & Tsuchiya, T. (1998). Cloning and expression of the gene for the Na+-coupled serine transporter from *Escherichia coli* and characteristics of the transporter. *J Bacteriol*, *180*(24), 6749–6752.

Palmer, A. G., Ali, M., Yang, S., Parchami, N., Bento, T., Mazzella, A., Oni, M., Riley, M. C., Schneider, K., & Massa, N. (2016). Kin recognition is a nutrient-dependent inducible phenomenon. *Plant Signaling & Behavior*, *11*(9), e1224045.

Pathak, D. T., Wei, X., Dey, A., & Wall, D. (2013). Molecular recognition by a polymorphic cell surface receptor governs cooperative behaviors in bacteria. *PLoS Genet*, *9*(11), e1003891.

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725.

Ruhe, Z. C., Nguyen, J. Y., Xiong, J., Koskiniemi, S., Beck, C. M., Perkins, B. R., Low, D. A., & Hayes, C. S. (2017). CdiA effectors use modular receptor-binding domains to recognize target bacteria. *MBio*, *8*(2).

Ruhe, Z. C., Wallace, A. B., Low, D. A., & Hayes, C. S. (2013). Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *MBio*, *4*(4), e00480-13.

Saak, C. C., & Gibbs, K. A. (2016). The Self-Identity Protein IdsD Is Communicated between Cells in Swarming *Proteus mirabilis* Colonies. *Journal of Bacteriology*, *198*(24), 3278–3286.

Schaffer, J. N., & Pearson, M. M. (2015). *Proteus mirabilis* and Urinary Tract Infections. *Microbiol Spectr*, *3*(5). https://doi.org/10.1128/microbiolspec.UTI-0017-2013

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., & Schmid, B. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.

Shao, Z., Lin, R. T., & Newman, E. B. (1994). Sequencing and characterization of the *sdaC* gene and identification of the *sdaCB* operon in *Escherichia coli* K12. *Eur J Biochem*, *222*(3), 901–907.

Shao, Z., & Newman, E. B. (1993). Sequencing and characterization of the *sdaB* gene from *Escherichia coli* K-12. *European Journal of Biochemistry*, *212*(3), 777–784.

Sharir-Ivry, A., & Xia, Y. (2017). The impact of native state switching on protein sequence evolution. *Molecular Biology and Evolution*, *34*(6), 1378–1390.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., & Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539.

Simon, R., Priefer, U., & Pühler, A. (1983). A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria. *Bio/Technology*, *1*(9), 784–791.

Strassmann, J. E., Gilbert, O. M., & Queller, D. C. (2011). Kin discrimination and cooperation in microbes. *Annu Rev Microbiol*, *65*, 349–367. https://doi.org/10.1146/annurev.micro.112408.134109

Sturgill, G. M., Siddiqui, S., Ding, X., Pecora, N. D., & Rather, P. N. (2002). Isolation of *lacZ* fusions to *Proteus mirabilis* genes regulated by intercellular signaling: Potential role for the sugar phosphotransferase (Pts) system in regulation. *FEMS Microbiology Letters*, *217*(1), 43–50.

Su, H. S., Lang, B. F., & Newman, E. B. (1989). L-serine degradation in *Escherichia coli* K-12: Cloning and sequencing of the *sdaA* gene. *Journal of Bacteriology*, *171*(9), 5095–5102.

Tipping, M. J., & Gibbs, K. A. (2019). Peer pressure from a *Proteus mirabilis* self-recognition system controls participation in cooperative swarm motility. *PLoS Pathogens*, *15*(7), e1007885.

Velayudhan, J., Jones, M. A., Barrow, P. A., & Kelly, D. J. (2004). L-serine catabolism via an oxygen-labile L-serine dehydratase is essential for colonization of the avian gut by *Campylobacter jejuni*. *Infection and Immunity*, *72*(1), 260–268.

Wall, D. (2014). Molecular recognition in myxobacterial outer membrane exchange: Functional, social and evolutionary implications. *Molecular Microbiology*, *91*(2), 209–220.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.

Wenren, L. M., Sullivan, N. L., Cardarelli, L., Septer, A. N., & Gibbs, K. A. (2013). Two independent pathways for self-recognition in *Proteus mirabilis* are linked by type VI-dependent export. *MBio*, *4*(4), e00374-13. https://doi.org/10.1128/mBio.00374-13

West, S. A., Diggle, S. P., Buckling, A., Gardner, A., & Griffin, A. S. (2007). The social lives of microbes. *Annu. Rev. Ecol. Evol. Syst.*, *38*, 53–77.

Willett, J. L., Gucinski, G. C., Fatherree, J. P., Low, D. A., & Hayes, C. S. (2015). Contact-dependent growth inhibition toxins exploit multiple independent cell-entry pathways. *Proceedings of the National Academy of Sciences*, *112*(36), 11341–11346.

Wu, Q. (2005). Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. *Genetics*, *169*(4), 2179–2188.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7.

Zepeda-Rivera, M. A., Saak, C. C., & Gibbs, K. A. (2018). A proposed chaperone of the bacterial type VI secretion system functions to constrain a self-identity protein. *J Bacteriol*. https://doi.org/10.1128/JB.00688-17

Zhang, X., El-Hajj, Z. W., & Newman, E. (2010). Deficiency in L-serine deaminase interferes with one-carbon metabolism and cell wall synthesis in *Escherichia coli* K-12. *J Bacteriol*, *192*(20), 5515–5525. https://doi.org/10.1128/JB.00748-10

Zhang, X., & Newman, E. (2008). Deficiency in l-serine deaminase results in abnormal growth and cell division of *Escherichia coli* K-12. *Molecular Microbiology*, *69*(4), 870–881.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 40.

Zhao, J., Gladieux, P., Hutchison, E., Bueche, J., Hall, C., Perraudeau, F., & Glass, N. L. (2015). Identification of allorecognition loci in *Neurospora crassa* by genomics and evolutionary approaches. *Molecular Biology and Evolution*, *32*(9), 2417–2432.

Chapter 3

A parallel genotype-phenotype study of ecological variation in populations of the opportunistic pathogen *Proteus mirabilis*

Achala Chittor, Zehan Zhou, Daniel Utter, Colleen Cavanaugh, and Karine A. Gibbs

Abstract

The bacterium *Proteus mirabilis* resides in guts and can cause infections if it journeys to the bladder. Disease, and likely colonization, involve its collective migration ("swarming"). Human clinical isolates comprise most research on swarming and genome diversity. However, *P. mirabilis* resides in a range of animals and environments. An open question is whether genomic or functional differences delineate isolates in animals versus humans; if not, animals could serve as reservoirs for human infection. Therefore, we examined strains isolated from asymptomatic research animals to explore ecological variation. Overall, *P. mirabilis* genomes shared critical metabolism and virulence genes in the core genome and did not cluster by the host. However, swarm migration and virulence, measured by waxworm killing, varied—despite motility and virulence genes residing in the core genome. By combining the paired phenotype and genotype data, we were able to pinpoint a nonsense mutation in a gene of one isolate as the cause of reduced motility and virulence. Thus, a similar approach in other opportunistic pathogens would provide a deeper dataset to understand the functional output of rare sequence variation and the eco-evolutionary pressures that shape collective behavior.

Introduction

        Genetic determinants of virulence have long been sought-after in bacterial

pathogens. Before next-generation sequencing and genomics were readily available,

the primary focus was on genomic regions called pathogenicity islands. These regions

encoding virulence factors can be mobilized between strains via plasmids or phage to

enable novel phenotypes that improve fitness and virulence (Gal-Mor & Finlay, 2006;

Hacker & Kaper, 2000). Whole-genome sequencing of multiple strains of a species led

to the model of the pangenome. Some genes are the essence of the species and

shared between strains, while others are confined to specific strains or even unique to a

single strain (Medini et al., 2005; Vernikos et al., 2015). In many organisms, virulence

and host-specific genes outside of a species-shared set of genes (or "core") can drive

functional specialization in different niches (Anani et al., 2020; Cross et al., 2021; Murfin

et al., 2015). Pangenome analysis can therefore be used to identify sets of genes that

are context-specific.

        However, the power of pangenome analysis often comes from a broad genome

pool consisting of different ecotypes and sufficient sequence variation between groups.

For many pathogens, such as *Proteus mirabilis*, genomes outside of clinical isolates are

still underrepresented. While *P. mirabilis* is a harmful urinary tract pathogen in

individuals with long-term catheters, it also exists as a gut commensal at very low

abundance in healthy people (Armbruster et al., 2018). It's hypothesized that the same

strain of *P. mirabilis* present in the gut can become a pathogen in the urinary tract

(Mathur et al., 2005). Likely needed to ascend into the urinary tract, *P. mirabilis*

migrates across surfaces using collective swarm motility behavior; virulence genes are

co-expressed during swarming (Allison et al., 1994; Howery et al., 2016; Pearson et al., 2010; Rather, 2005). Further, *P. mirabilis* strains use identity genes to recognize kin and maintain a clonal population separate from other strains during swarming (Gibbs et al., 2008; Wenren et al., 2013). Are virulence genes, swarm motility, and self recognition conserved across strains? Or are they only present in a subset of emergent pathogenic strains? To answer this, we aimed to better understand the phylogenetic history of *P. mirabilis* outside of human infection.

In conjunction with comparative genomics analysis of diverse ecotypes, there is a need for phenotypic characterization. The same molecular factors that are relevant for human disease can contribute to environmental fitness. For example, in *Vibrio cholerae,* virulence factors that are important for human infection also serve a clear purpose in marine environments; the same proteins that mediate attachment to intestinal epithelial cells are also implicated in colonization of copepods and crustaceans (Sakib et al., 2018)*.* Studying an organism outside of pathogenesis can recontextualize the genes and behaviors associated with virulence within the organism's broader ecology.

P. mirabilis is part of the ecologically diverse *Morganellaceae* family. Other genera *Providencia* and *Morganella* are opportunistic pathogens in humans, whereas genera *Xenorhabdus and Photorhabdus* are nematode-associated insect pathogens (Armbruster et al., 2018; Drzewiecka, 2016; Sajnaga & Kazimierczak, 2020). While swarm behavior and homologous identity genes appear to exist across these different genera, their ecological impact remains unclear. Studying *P. mirabilis* ecotypes from both genomics and microbiology perspectives can provide a more complete

understanding of the different ecological pressures that shape collective behavior in *P. mirabilis* and related taxa.

Studying the phenotypic outputs of the genome may shed light on organisms that are both long-term commensals and pathogens. In *Escherichia coli*, both a commensal and pathogen, a disconnect between isolates and their host species phylogeny suggests that these organisms may act as generalists that can readily adapt to different environment (Murphy et al., 2021; Tenaillon et al., 2010). There may be environmental factors or gene regulatory mechanisms that determine these transitions. In this case, the gene content of the genome is a first step, but may not be sufficient to fully understand the impact of gene expression. Therefore, we deepened our pangenome analysis by pairing it with phenotypic analysis to pinpoint rare mutations with a large impact on behavior and virulence.

In this work, we performed whole-genome sequencing on a set of *P. mirabilis* isolates from asymptomatic animals and assessed swarm motility. Comparing these isolates with a reference human clinical isolate, we found that they all contained critical virulence genes and most were fully capable of swarm expansion. However, we observed defects in swarm motility in a subset of isolates; these also showed decreased virulence in a waxworm model. For one such isolate, we were able to identify the putative cause as a nonsense mutation in the flagellar *fliF* gene by mining the genomic data. Our approach provides an example of how genotype-phenotype analysis of different ecotypes can connect natural genetic variation to major effects on behavior, leading to potential advances in understanding the molecular mechanisms that regulate virulence in opportunistic pathogens.

Genomes share key metabolism and virulence genes

To study the opportunistic pathogen *P. mirabilis* beyond clinical isolates, we received a set of *P. mirabilis* strains as pure cultures from Charles River Laboratories (Wilmington, MA) that were isolated from asymptomatic animals. Charles River Laboratories is a global company that supports biotech, pharmaceutical, agrochemical, academic, and government institutions with research animals and other resources. Animals are housed based on two different customer types: biotech and production. For biotech, animals are used for basic research and preclinical applications such as drug discovery and drug safety. For production, research animals generate antibodies and provide other blood products such as plasma. Research animals are routinely screened by mass spectrometry for pathogenic organisms such as *P. mirabilis* which can impact research applications. When detected, bacteria were cultured as single isolates and shipped to us, which we labeled as STAR isolates. We received thirty *P. mirabilis* strains from mice, nine from rats, one from a chicken, and one from a rabbit isolated from either the gastrointestinal tract, the respiratory tract, or the animal's environment (Table A5). This set of *P. mirabilis* STAR strains were isolated from various asymptomatic research animals from two housing areas within Charles River Laboratories.

We first aimed to compare the gene content of these isolates. The genomes of these isolates were sequenced using Illumina NextSeq, and assembled using SPAdes (Bankevich et al., 2012). We found close relatedness between the genomes in terms of

the overall sequence identity, phylogeny, and synteny (Figure A15). We next compared the whole genome assemblies with each other and with the clinical isolate BB2000, a reference strain with a closed genome (Sullivan et al., 2013) using the pangenome tool Anvi'o [Figure 3.1, (Eren et al., 2015)]. Each row corresponds to a different strain labeled on the left; the BB2000 strain is highlighted in red. The host from which the strain was isolated is marked on the left by different colors. All of the predicted genes in all genomes are laid out horizontally and are not in linear genome order. If a gene is present in a given strain, it is marked as a vertical dark line. Genes that are present in all genomes make up the core genome, which was approximately 86% of the genes in the BB2000 genome. The core genome and the variable regions on either side are marked below. The variable regions contain genes that are present in some strains and not others. Genes that are strain-specific can be used to cluster the genomes with others that are more similar, resulting in the dendrogram and three major clusters labeled on the right. The core genome contained genes we would predict are essential to *P. mirabilis* physiology.

The core genome contained genes involved in core metabolic pathways (Table 3.1). These genes are involved in glycolysis, pentose phosphate, Enter-Doudoroff, gluconeogenesis, and the TCA cycle. *P. mirabilis* swarm motility is an energetically costly process that relies on efficient central metabolism pathways and availability of specific nutrients in the surrounding environment (Alteri et al., 2012; Armbruster et al., 2013). *P. mirabilis* utilization of central metabolism is unique from *E. coli* within the same host environment during urinary tract infection (Alteri et al., 2015), so we selected genes in metabolic pathways that are critical for *P. mirabilis* swarming and

75
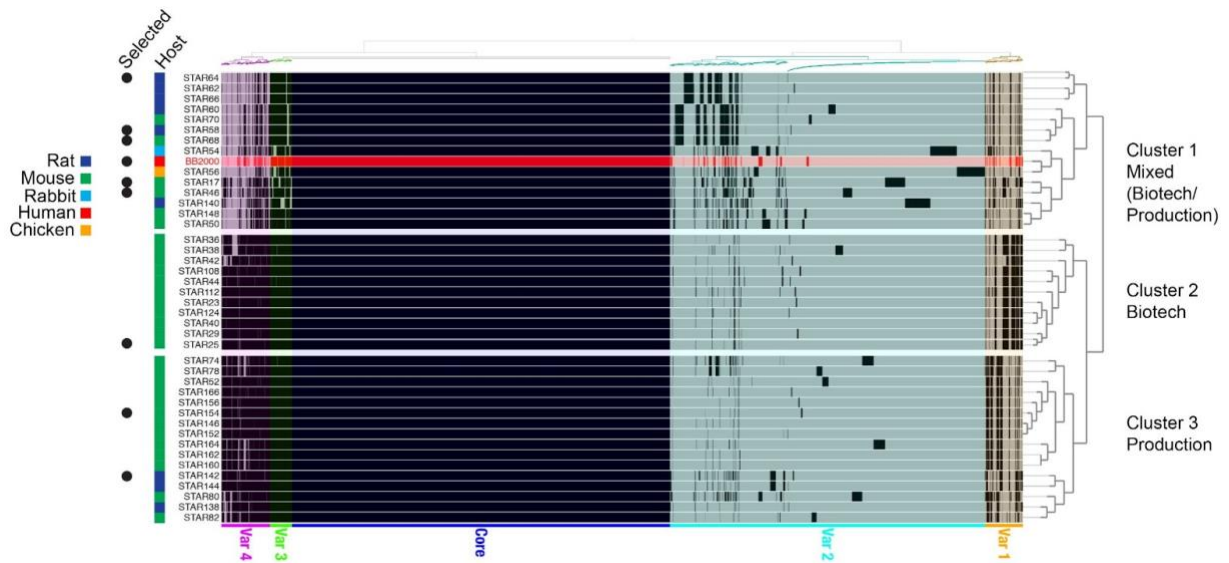
Figure 3.1: Isolate genomes share critical metabolism and virulence genes and do not cluster based on host-specific genes.
Pangenome of *P. mirabilis* STAR isolates and BB2000 human clinical isolate (highlighted in red). Genomes are not in linear gene order. Gene presence is denoted by dark shading and gene absence by lighter shading. Core and variable regions are labeled below and highlighted in different colors. Representative strains shown in later figures are labeled on the left by the strain name and cover different hosts. Strains are clustered together based on similarity of presence of variable genes. The resulting clusters are labeled on the right alongside the dendrogram with the correlated housing location in Charles River Laboratories.

Table 3.1: Metabolism genes in the core genome

Metabolism Gene	Pathway	PMID
pfkA	Glycolysis	PMI3203
pgi	Glycolysis	PMI2754
tpiA	Glycolysis	PMI3205
gnd	Pentose Phosphate	PMI0655
talB	Pentose Phosphate	PMI0006
edd	Entner-Doudoroff	PMI2760
sdhB	TCA cycle	PMI0568
frdA	TCA cycle	PMI3588
fumC	TCA cycle	PMI1296
pckA	Gluconeogenesis	PMI3015

pathogenesis. These central metabolism genes are present in the core genome and shared across STAR isolates.

The core genome also contained key *P. mirabilis* virulence genes (Table 3.2). Flagella and fimbriae are involved in motility, surface sensing and attachment to the host epithelium (Debnath et al., 2018). Hemolysin further promotes epithelial invasion and nitrogen metabolism, in part through urease activity, is a significant nutrient source in urine (Armbruster et al., 2018). Although the STAR strains were isolated from asymptomatic animals, they share these virulence genes with clinical isolate BB2000.

Genomes differ in identity, LPS, mobile element, and some virulence genes

The variable regions contained self-recognition genes and LPS biosynthesis genes which were expected to vary between strains (Table 3.3.) Self-recognition genes are polymorphic, allowing kin to recognize one another during swarming (Gibbs et al., 2008; Wenren et al., 2013). A bottleneck of strain-level competition during swarm migration could impact the clonality of infection. Lipopolysaccharides (LPS) are large molecules that decorate the outer membrane of gram-negative bacteria that impact membrane permeability and are recognized by the host immune system. Modification of LPS allows pathogens to more readily evade host defenses, and variation in LPS biosynthesis genes may indicate co-evolution with the host (Simpson & Trent, 2019). Additionally, LPS alters the integrity of the outer membrane, which can impact swarm motility and sensitivity to environmental stressors in *P. mirabilis* (Little et al., 2018). In contrast to the phylosymbiosis model (Brooks et al., 2016), large sets of host-specific genes in variable regions did not cause genomes to cluster by host. Variable region genes may modulate interaction between bacteria and the environment.

Table 3.2: Virulence genes in the core genome

Virulence gene	Function	PMID
tesB	Lipid metabolism: acyl-CoA thioesterase	PMI0128
mrpJ	Transcriptional regulator: fimbria-mediated attachment and flagellar motility	PMI0271
PMI0720	Cold shock protein	PMI0720
ompA	Outer Membrane Protein	PMI0785
cspA	Cold shock protein	PMI0913
PMI0999	Lipid metabolism	PMI0999
fim8J	Fimbrial gene	PMI1470
fliE	Flagellar gene	PMI1629
fliF	Flagellar gene	PMI1630
flhD	Flagellar gene	PMI1672
rcsB	Capsular synthesis regulator component B	PMI1730
pmfA	Fimbrial gene (major type 1 subunit fimbrin [pilin])	PMI1877
speA	Biosynthetic arginine decarboxylase	PMI2094
pmpA	MrpJ-regulated Fimbrial gene	PMI2223
glnD	Part of nitrogen sensing/metabolism	PMI2287
atfA	Major subunit of type 1 fimbria; pilin	PMI2728
PMI3637	Part of nitrogen sensing/metabolism	PMI3637
hpmA	Hemolysin	PMI2057

Table 3.3: Identity genes involved in strain-level variation

Gene Name	Function	BB2000 ID	Region
IdsD - 1034	Self recognition	BB2000_3005	Var 4
IdsD - 1072	Self recognition	N/A	Var 2
IdsE	Self recognition	BB2000_3006	Var 3
IdrD	Self recognition	BB2000_0825	Core
IdrE	Self recognition	BB2000_0826	Var 1
BB2000_3203	LPS biosynthesis	BB2000_3203	Var 2
BB2000_3208	LPS biosynthesis	BB2000_3208	Var 2

We observed many phage-related genes and other mobile elements in the variable regions. Phage defense systems including CRISPR/Cas genes (Var 4) and type I restriction endonucleases were also in the variable regions, consistent with what we would predict if there is ongoing co-evolution with phage. We also found genes with predicted virulence functions outside of the core genome including those involved in siderophore biosynthesis, heme utilization, and pilus assembly. These may be exchanged in conjunction with other transferred genes on plasmids or as part of integrative and conjugative elements. Exchange of genetic material outside of the core genome can provide ways for new functions and phenotypes to evolve.

These variable region gene functions are consistent with recently identified variation between *P. mirabilis* isolates in humans and *Nasonia* wasps (Cross et al., 2021), suggesting that these are flexible regions of *P. mirabilis* genome that are potentially experiencing selection. However, instead of by host, we observed that strains clustered by housing within the Charles River facility. Based on metadata for the isolates, clusters two and three seem to correspond to Biotech and Production consumer types, respectively (Figure 3.1). Charles River may contain a dominant *P. mirabilis* lineage that has recently evolved by localized phage interactions and horizontal gene transfer in the two housing areas. Cluster 1, by contrast, contains a mix of biotech and production animals and contains a broader host range. These strains are potentially environmentally associated since they cluster along with diverse strains from NCBI, while the other two clusters remain well-defined (data not shown). The host genotype may also play a selective role as the set of rat isolates in cluster 1 is all from

the rat strain SHRS and may represent a host-associated lineage that resists

colonization by other strains.

From genomic information alone, we are somewhat limited in the conclusions we

can draw about how these strains behave or perform. Regulation of gene expression at

the transcriptional and translational level as well as gene interactions add complexity to

how the genome actually manifests functionally. We therefore studied all of the strains

experimentally and observed subtle phenotypic differences in traits such as smell,

appearance of the swarm colony, and boundary formation with other strains (Table A6).

We therefore wanted to further investigate any differences in swarm expansion between

strains that share metabolism and virulence genes but may differ in gene expression or

other regulatory mechanisms.

Swarm colony expansion is reduced in some isolates, leading to reduced *in vivo*

virulence

Growth and swarm motility are shown here for a subset of representative isolates

that includes different hosts and genome clusters. We first compared liquid growth in

nutrient-rich media, which was similar across the STAR isolates and BB2000,

suggesting that core metabolism genes are sufficient for liquid growth (Figure 3.2A,

Figure A16). There appear to be two growth phases that could correspond to diauxic

growth on different substrates and growth rates vary slightly in both phases (Figure

3.2A). For swarm motility, most isolates are fully capable of swarming comparable to

BB2000, but we see a few isolates that are poor swarmers or even non-swarming under

our assay conditions (Figure 3.2B, Figure A16). The swarm defect is found in multiple
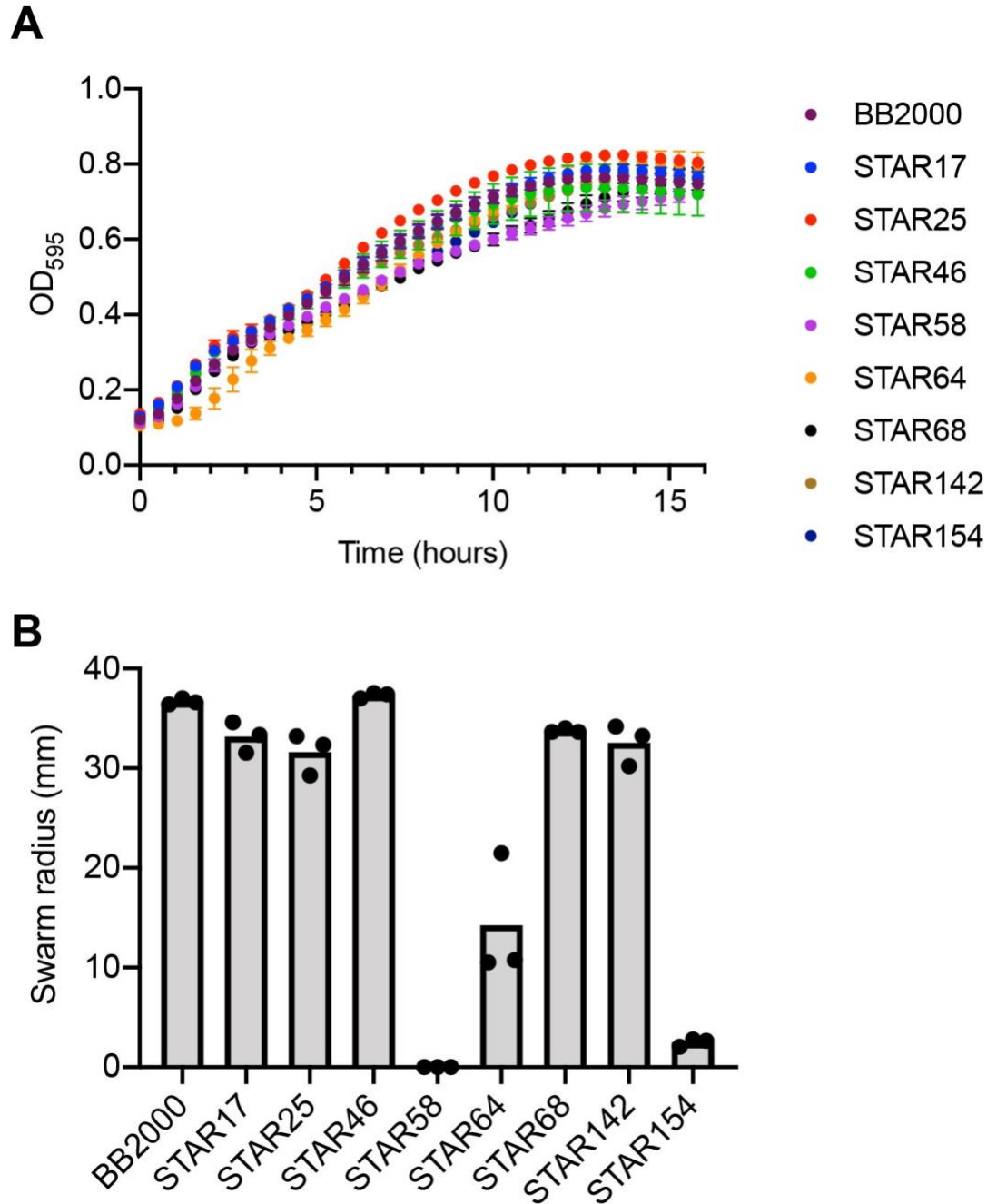
Figure 3.2: Variation in growth and swarm motility across STAR isolates.
A) Growth curve of representative STAR isolates and BB2000. Mean of three biological replicates with standard deviation error bars is shown. B) Swarm expansion of representative STAR isolates and BB2000. Three biological replicates are shown as points and the mean is shown as a grey filled bar.

hosts (mouse and rat) and genome clusters (clusters 1 and 3). A broader range of phenotypic variation may stably exist in commensal isolates of *P. mirabilis.*

To test virulence capability of the representative strains, we used the waxworm *Galleria mellonella* as a model system; as the worms become infected and die, they also darken in color (Figure 3.3A). As a non-pathogenic control, we used BB2000 *Δugd* since deletion of the LPS biosynthesis gene leads to reduced cell invasion and inhibition of swarming (Jiang et al., 2010; Little, 2017). As positive controls, we used the clinical isolates BB2000 and HI4320, as well as BB2000 *ΔrcsB* because deletion of *rcsB* in HI4320 leads to hyperswarming and increased virulence in *Galleria* (Howery et al., 2016). We found that all of the strains we tested killed at least 50% of the worms by two days post injection (Figure 3.3B). This is somewhat surprising given that these isolates were taken from asymptomatic animals. However, given that *P. mirabilis* is predicted to exist as a commensal and as a pathogen, virulence may be encoded in the core genome and contextually activated or emergent. The three poor swarmer strains also showed attenuated killing: STAR58, STAR64, and STAR154 (Figure 3.3B). STAR58 also killed more slowly than STAR64 or STAR154. Virulence in waxworms is partially determined by proficiency of swarm motility.

The STAR58 strain consists of non-motile swarmer cells, most likely due to a nonsense mutation in the flagellar gene *fliF*

The swarm cycle is conserved, but swarm motility is variable. The swarm cycle, surface-based development, and the link between cell length and swarm motility have
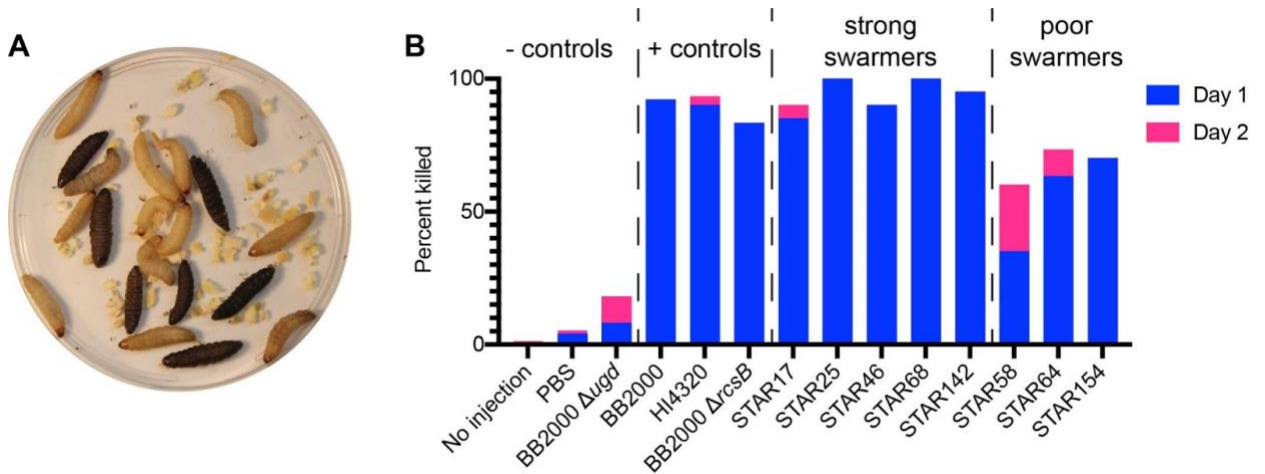
Figure 3.3: Poor swarmers are attenuated in virulence compared to strong swarmers.
A) Representative plates of *Galleria mellonella* waxworms during virulence assay. Cream-colored worms are healthy and alive while dark-colored worms are dead. B) The average percent of waxworms killed is shown over a two day period (killed on the first day in blue, killed between the first and second day in pink). Sample size of at least 20 worms per strain. Negative controls are no injection, PBS, and BB2000 *Δugd*. Positive controls are BB2000, HI4320, and BB2000 *ΔrcsB*. BB2000 and representative STAR isolates are separated into strong and poor swarmers (STAR58, STAR64, and STAR154).

been studied in detail in *P. mirabilis* (Armbruster & Mobley, 2012; Little et al., 2019; Morgenstein et al., 2010) as well as in other robust swarmers such as *Vibrio parahaemolyticus* (Broberg et al., 2011; Gode-Potratz et al., 2011). When *P. mirabilis* cells are grown on a surface, liquid-grown cells upregulate the flagellar master regulator *flhDC* and elongate into swarmer cells, which is accompanied by the expression of virulence-associated genes [Figure 3.4A, (Pearson et al., 2010)]. Using phase microscopy of swarming populations, we observed that cells from all the strains exhibit cellular elongation as an indicator of entry into the swarm cycle and surface adaptation (Figure 3.4B). STAR64 and STAR154 show multicellular rafting and migration out of the inoculum, indicating that there may be a defect in macroscale migration (Figure 3.4B). The swarm cycle and swarmer cell development may be core phenotypes to *P. mirabilis* as a species.

STAR58 cells, unlike STAR64 and STAR154, were completely non-motile on surfaces (Figure 3.4B). Flagella are used to propel bacteria both through liquid and across surfaces. We therefore interrogated the motility phenotype further by performing a swim motility assay where cells swim outwards through low percentage agar. A defect in flagellar production would be predicted to cause general motility defects. Indeed, STAR58 was the only strain with no swim motility, likely due to a flagellar defect (Figure 3.4C). Motility is known to affect virulence through its multifaceted impact on surface attachment, invasion, and colony formation (Duan et al., 2013; Josenhans & Suerbaum, 2002). The lack of motility may explain why STAR58 was slower to kill *Galleria* than the other poor swarmers (Figure 3.3B). STAR58 is defective in swim and swarm motility, suggesting that a defect in flagellar production is at the root of the phenotype.
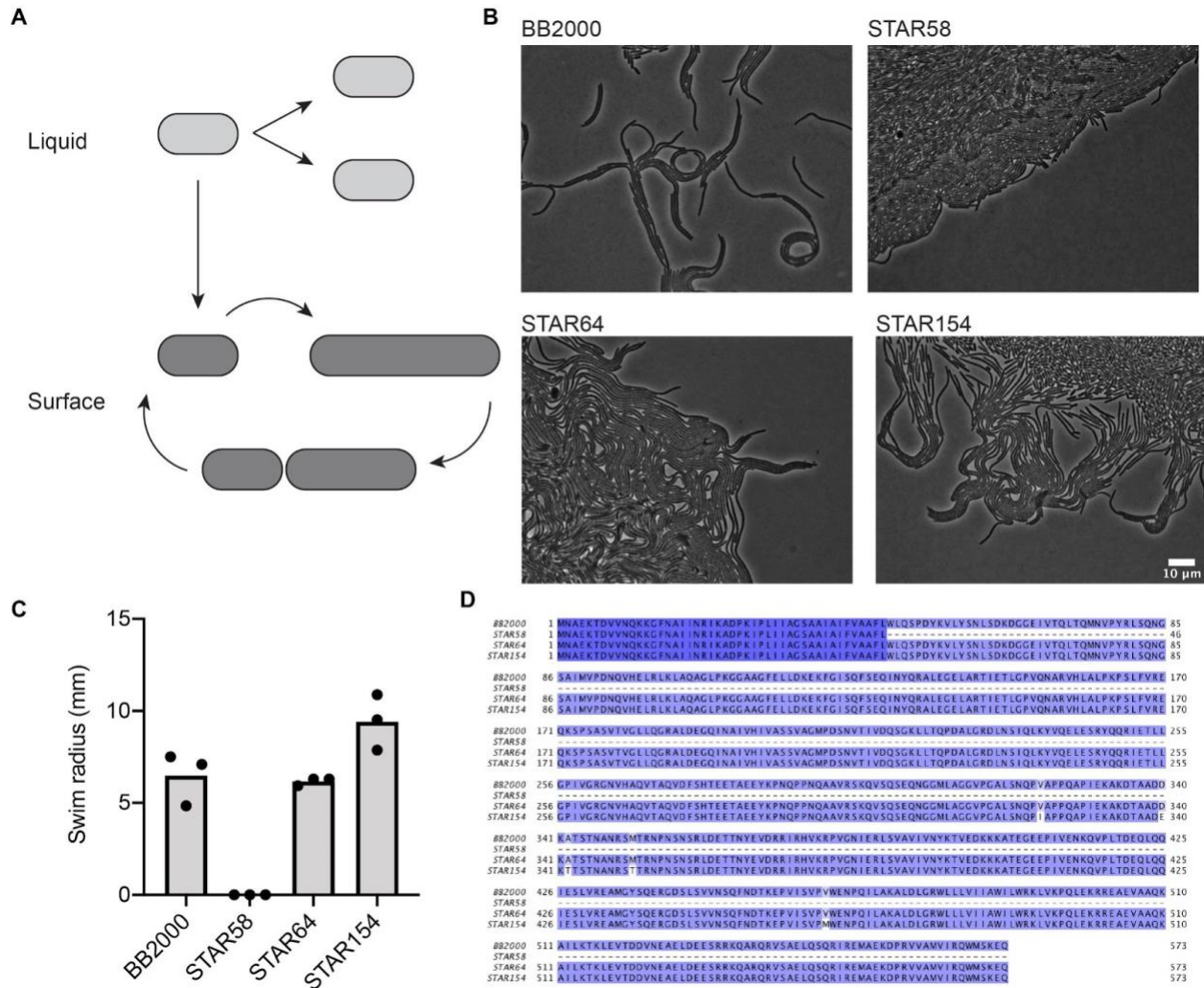
Figure 3.4: Swarmer cell development and swim motility point to FliF truncation in STAR58.

A) *P. mirabilis* cells grown in liquid divide as short cells. When liquid-grown cells encounter a surface, they differentiate into elongated, hyperflagellated swarmer cells that periodically divide back into short consolidated cells and then elongate into swarmer cells. B) Phase-contrast microscopy of swarmer cells from strains BB2000, STAR58, STAR64, and STAR154. C) Distance migrated in swim motility assay for same strains as in (B). D) Clustal Omega (Sievers et al., 2011) alignment of FliF amino acid sequences from the same strains as in (B) highlighted according to sequence conservation using Jalview (Waterhouse et al., 2009). STAR58 FliF is truncated to 46 amino acids due to a nonsense mutation.

The motility defect in STAR58 was traced back to a putative causative nonsense mutation in *fliF*. When we looked at the genome sequence of STAR58 and compared it to the other genomes, we found a nonsense mutation early in the flagellar MS ring gene *fliF*, resulting in a truncated protein of only 46 amino acids (Figure 3.4D). This mutation and the non-motile swarmer cell phenotype are consistent with the phenotype that was observed previously for a transposon mutation in *fliF* (Belas & Suvanasuthi, 2005). It is interesting to find a mutation in a key flagellar gene and virulence factor that can occur naturally and persist within a host without being rapidly outcompeted. Because the truncation is due to a single nucleotide polymorphism (SNP), it was not immediately apparent in the pangenome analysis and was still included as part of the core genome. Rare SNPs that disproportionately impact behavior can be identified using genotypes and phenotypes together and can potentially lead to an ecology-focused understanding of molecular mechanisms of virulence.

Discussion

This exploratory study has revealed potential areas for advancement in genomics and microbiology. The majority of published genomes and reference strains for *P. mirabilis* come from a disease context in humans, so including isolates from asymptomatic animals expands the available sequence data. Upon performing the pangenome analysis, virulences genes were shared between strains (Figure 3.1), suggesting that many of the well-studied virulence genes are not unique to human infection. We also did not find large sets of host-specific genes, as observed previously for *P. mirabilis* (Cross et al., 2021). Molecular interactions and gene expression could be equally (or more) important for pathogenicity or persistence, especially in long-term residents or generalists such as *P. mirabilis*. The level of phenotypic variation between isolates may widen depending on the context. Phenotypic assays, when coupled to genomic datasets, could pinpoint rare mutations or differences in gene expression that are tied to the organism's ecology.

When we investigated the swarming behavior of these isolates, we found that most were proficient swarmers, but some had reduced swarm expansion. We then validated that strains with reduced swarm motility were also less virulent (Figure 3.3). These variants were readily isolated from asymptomatic animals, suggesting that motility may be important not for commensals, but potentially for the transition from commensal to pathogen. By contrast, all isolates showed swarmer cell differentiation, suggesting the swarmer cell state, separate from collective motility, may provide an advantage for commensalism. Further, we found a nonsense *fliF* mutation as a naturally occurring variant, which was unexpected given the importance of *fliF* for motility and

89

virulence (Figure 3.4). These results highlight how capturing a spectrum of phenotypes that are stable in a subset of ecological contexts can help to interrogate specific molecular mechanisms of virulence.

In studying pathogens like *P. mirabilis*, clinical microbiology has often overlooked ecology and how the genome and behavior are shaped by different pressures in different environments. For example, our understanding of *V. cholerae* and other pathogens in the *Vibrionaceae* family was hindered by an intense focus on its behavior in a human host, ignoring how it survives in its marine environment (Sakib et al., 2018). For pathogens that are also long-term commensals, fitness in both phases will shape its genome and physiology. To fully understand *P. mirabilis* as an opportunistic pathogen, we need to sample and study diverse isolates.

P. mirabilis cells use self recognition to selectively engage with kin during swarming, but the ecology of self recognition is unknown. We observed that the genes that determine strain identity vary between STAR isolates (Table 3.3), but components of the secretion machinery and other self-recognition factors are found in the core genome. Our hypothesis is that self recognition is important for collective motility outside of pathogenesis and may be regulated by environmental factors. This genome dataset can be used to further interrogate the molecular evolution of self-recognition proteins by studying sequence variation. Protein alignments and phylogenies can be applied experimentally to understanding the sequence space and coevolution of self-recognition proteins and their partners within the biochemical and structural constraints of protein-protein interactions.

There is a broader need for genome data outside clinical isolates, ideally paired with phenotypic data to advance our understanding of opportunistic pathogens. By using closely related environmental or commensal isolates that exhibit a different range of phenotypic variation, we can link genotypic and phenotypic data with minimal genetic diversity between isolates. This kind of microbial GWAS in a more high throughput manner can reveal naturally occurring high impact SNPs that correlate with changes in virulence-associated behaviors such as swarm motility that can be screened experimentally (Read & Massey, 2014; San et al., 2020). This approach has recently proven useful for understanding other pathogens such as *Campylobacter jejuni* (Bandoy & Weimer, 2020) and *Listeria monocytogenes* (Hsu et al., 2020). Thus, we can begin to bridge the gap that exists between molecular microbiology and computational genomics through an interdisciplinary approach studying ecological variation.

<u>Materials and Methods</u>

Bacterial strains and media

The strains used in this study are described in Table 3.4. Isolates from Charles River Laboratories are denoted as "STAR." Metadata for STAR isolates can be found in Table A5. *P. mirabilis* strains were maintained on low swarm (LSW) agar (Belas et al., 1991). CM55 blood agar base agar (Oxoid, Basingstoke, England) was used for swarm-permissive nutrient plates. Tryptone agar (1% Tryptone, 0.5% NaCl, 0.3% Difco Noble Agar) was used for swim assays. Overnight cultures of all strains were grown at 37°C in LB broth under aerobic conditions. For growth curve assays, cells were grown in LB broth.

Growth curve

Overnight cultures were normalized to an optical density at 600 nm (OD600) of 0.1 in LB medium. Normalized cultures were grown overnight at 37°C, with periodic shaking, in a Tecan Infinite 200 PRO microplate reader (Tecan, Männedorf, Switzerland).

Swarm expansion assay

Overnight cultures were normalized to an OD600 of 0.1, and swarm-permissive nutrient plates were inoculated with 2 µl of normalized culture in the center. Plates were incubated at 30°C for 16 hours, and the radii of actively migrating swarms starting from the edge of the inoculum were measured using Fiji (ImageJ) (Schindelin et al., 2012).

Swim motility assay

Overnight cultures were normalized to an OD600 of 0.1, and tryptone agar plates were inoculated with normalized culture using an inoculation needle stabbed into the center.

Plates were incubated at room temperature for 24 hours, and the radii of motility starting from the inoculation point were measured using Fiji (ImageJ) (Schindelin et al., 2012).

Microscopy

One-millimeter-thick swarm-permissive CM55 agar pads were inoculated with overnight cultures and incubated overnight at room temperature. The agar pads were then incubated at 30°C in a modified humidity chamber. After three to four hours, the pads were imaged by phase-contrast microscopy (10 ms exposure) using a Leica DM5500B microscope system (Leica Microsystems, Buffalo Grove, IL) and a CoolSnap HQ2 cooled charge-coupled device camera (Photometrics, Tucson, AZ). MetaMorph (version 7.8.0.0; Molecular Devices, Sunnyvale, CA) was used for image acquisition.

Boundary Assay

Cultures of each isolate were grown overnight in LB at 37°C. Each CM55 swarm-permissive plate was first labeled with three distinct sections of the plate, one per strain. Each section was inoculated in the center with 2 µL of overnight culture. Plates were incubated at room temperature for 24h. As a control, BB2000 was included as one of the isolates on each plate to confirm expected boundary formation. Each isolate was tested against three other STAR isolates in addition to BB2000.

***Galleria mellonella* virulence assay**

The protocol for *P. mirabilis* preparation was adapted from previous work (Howery et al., 2016). Overnight cultures of *P. mirabilis* strains were grown overnight at 37°C with shaking. 10 µL of the culture was inoculated into fresh LB broth and grown for 2 hours

at 37ºC until the OD reached approximately 0.6. Dilutions of 10^{-6} were prepared in fresh

LB to result in approximately 10 CFUs in a 10µL inoculation volume.

Waxworms were ordered from DubiaRoaches (Wichita, KS). Upon receipt, any larvae

that were dead or discolored were discarded. Waxworms were then incubated at 16ºC

for at least 2 days for recovery before performing injections. Waxworm injection protocol

was adapted from previous work (Hernandez et al., 2019). Waxworms were placed in

petri dishes on ice for 30 minutes prior to injection. For injection, waxworms were placed

in between two petri dishes, a 60 mm plate placed in a 100mm one, that were joined

together by attaching small binder clips to either side and tying rubber bands around

binder clips from both plates. A 100µL Hamilton syringe (Sigma Aldrich, St. Louis, MO)

with sterile 27G 1/2 inch needle (BD, Franklin Lakes, NJ) was used to inject larvae with

10 CFUs of *P. mirabilis* diluted culture into the last proleg. The syringe and needle were

cleaned between waxworm injections with three washes in 70% ethanol and three

washes in filter-sterilized PBS. A new sterile needle was used for each strain. Two

negative controls were used for each experiment: a no-injection control and a sterile

PBS control. After injection, waxworms were moved into a new 100mm petri dish (20

waxworms per petri dish) and incubated at 37ºC for 48 hours, counting and discarding

any dead waxworms at 24 hour and 48 hour timepoints post-injection.

Whole-genome sequencing and genome assembly

Bacterial isolates were subjected to phenol-chloroform extractions (Sambrook &

Russell, 2006) to isolate genomic DNA (gDNA). Library prep using KAPA HyperPrep kit

(Roche, Wilmington, MA) was performed by the Bauer Core Facility. The library was

sequenced as 150-bp paired-end reads using an Illumina NextSeq 500 system

(Illumina, San Diego, CA) by the Bauer Core Facility.

STAR25 and STAR64 were chosen as representative strains to assemble higher quality

genomes using long reads. STAR25 and STAR64 gDNA were submitted for nanopore

sequencing by the Microbial Genome Sequencing Center (Pittsburgh, PA).

Trimmomatic (Bolger et al., 2014) and SPAdes (Bankevich et al., 2012) were used to

assemble genome assemblies. Genomes were assembled using both short Illumina and

long Nanopore reads using the HYBRID-SPAdes assembly option (Antipov et al., 2016).

Anvi'o (Eren et al., 2015) was used for pangenome analysis.

Table 3.4: *Proteus mirabilis* strains used in this study

Strain	Description	Source
BB2000	Human clinical isolate	(Belas et al., 1991)
HI4320	Human clinical isolate	(Pearson et al., 2008)
BB2000 *ΔrcsB*	BB2000 with a chromosomal deletion of the *rcsB gene*	(Little et al., 2018)
BB2000 *Δugd*	BB2000 with a chromosomal deletion of the *ugd gene*	(Little, 2017)
STAR 17	CR metadata in supplementary table A5	This study
STAR 23	CR metadata in supplementary table A5	This study
STAR 25	CR metadata in supplementary table A5	This study
STAR 29	CR metadata in supplementary table A5	This study
STAR 36	CR metadata in supplementary table A5	This study
STAR 38	CR metadata in supplementary table A5	This study
STAR 40	CR metadata in supplementary table A5	This study
STAR 42	CR metadata in supplementary table A5	This study
STAR 44	CR metadata in supplementary table A5	This study
STAR 46	CR metadata in supplementary table A5	This study
STAR 50	CR metadata in supplementary table A5	This study
STAR 52	CR metadata in supplementary table A5	This study
STAR 54	CR metadata in supplementary table A5	This study
STAR 56	CR metadata in supplementary table A5	This study
STAR 58	CR metadata in supplementary table A5	This study
STAR 60	CR metadata in supplementary table A5	This study
STAR 62	CR metadata in supplementary table A5	This study
STAR 64	CR metadata in supplementary table A5	This study

Table 3.4: *Proteus mirabilis* strains used in this study (Continued)

Strain	Description	Source
STAR 66	CR metadata in supplementary table A5	This study
STAR 68	CR metadata in supplementary table A5	This study
STAR 70	CR metadata in supplementary table A5	This study
STAR 74	CR metadata in supplementary table A5	This study
STAR 78	CR metadata in supplementary table A5	This study
STAR 80	CR metadata in supplementary table A5	This study
STAR 82	CR metadata in supplementary table A5	This study
STAR 108	CR metadata in supplementary table A5	This study
STAR 112	CR metadata in supplementary table A5	This study
STAR 124	CR metadata in supplementary table A5	This study
STAR 138	CR metadata in supplementary table A5	This study
STAR 140	CR metadata in supplementary table A5	This study
STAR 142	CR metadata in supplementary table A5	This study
STAR 144	CR metadata in supplementary table A5	This study
STAR 146	CR metadata in supplementary table A5	This study
STAR 148	CR metadata in supplementary table A5	This study
STAR 152	CR metadata in supplementary table A5	This study
STAR 154	CR metadata in supplementary table A5	This study
STAR 156	CR metadata in supplementary table A5	This study
STAR 160	CR metadata in supplementary table A5	This study
STAR 162	CR metadata in supplementary table A5	This study
STAR 164	CR metadata in supplementary table A5	This study
STAR 166	CR metadata in supplementary table A5	This study

References

Allison, C., Emody, L., Coleman, N., & Hughes, C. (1994). The role of swarm cell differentiation and multicellular migration in the uropathogenicity of Proteus mirabilis. *J Infect Dis*, *169*(5), 1155–1158.

Alteri, C. J., Himpsl, S. D., Engstrom, M. D., & Mobley, H. L. (2012). Anaerobic respiration using a complete oxidative TCA cycle drives multicellular swarming in Proteus mirabilis. *MBio*, *3*(6), e00365-12.

Alteri, C. J., Himpsl, S. D., & Mobley, H. L. T. (2015). Preferential Use of Central Metabolism In Vivo Reveals a Nutritional Basis for Polymicrobial Infection. *PLoS Pathogens*, *11*(1), e1004601. https://doi.org/10.1371/journal.ppat.1004601

Anani, H., Zgheib, R., Hasni, I., Raoult, D., & Fournier, P.-E. (2020). Interest of bacterial pangenome analyses in clinical microbiology. *Microbial Pathogenesis*, *149*, 104275.

Antipov, D., Korobeynikov, A., McLean, J. S., & Pevzner, P. A. (2016). hybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics*, *32*(7), 1009–1015. https://doi.org/10.1093/bioinformatics/btv688

Armbruster, C. E., Hodges, S. A., & Mobley, H. L. (2013). Initiation of swarming motility by Proteus mirabilis occurs in response to specific cues present in urine and requires excess L-glutamine. *J Bacteriol*, *195*(6), 1305–1319. https://doi.org/10.1128/JB.02136-12

Armbruster, C. E., & Mobley, H. L. (2012). Merging mythology and morphology: The multifaceted lifestyle of Proteus mirabilis. *Nat Rev Microbiol*, *10*(11), 743–754. https://doi.org/10.1038/nrmicro2890

Armbruster, C. E., Mobley, H. L. T., & Pearson, M. M. (2018). Pathogenesis of Proteus mirabilis Infection. *EcoSal Plus*, *8*(1), 10.1128/ecosalplus.ESP-0009–2017. https://doi.org/10.1128/ecosalplus.ESP-0009-2017

Bandoy, D. D. R., & Weimer, B. C. (2020). Biological Machine Learning Combined with Campylobacter Population Genomics Reveals Virulence Gene Allelic Variants Cause Disease. *Microorganisms*, *8*(4), 549. https://doi.org/10.3390/microorganisms8040549

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Belas, R., Erskine, D., & Flaherty, D. (1991). Transposon mutagenesis in Proteus mirabilis. *J Bacteriol*, *173*(19), 6289–6293.

Belas, R., & Suvanasuthi, R. (2005). The ability of *Proteus mirabilis* to sense surfaces and regulate virulence gene expression involves FliL, a flagellar basal body protein. *J Bacteriol*, *187*(19), 6789–6803. https://doi.org/10.1128/JB.187.19.6789-6803.2005

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Broberg, C. A., Calder, T. J., & Orth, K. (2011). Vibrio parahaemolyticus cell biology and pathogenicity determinants. *Microbes and Infection*, *13*(12–13), 992–1001. https://doi.org/10.1016/j.micinf.2011.06.013

Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J., & Bordenstein, S. R. (2016). Phylosymbiosis: Relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biology*, *14*(11), e2000225.

Cross, K. L., Leigh, B. A., Hatmaker, E. A., Mikaelyan, A., Miller, A. K., & Bordenstein, S. R. (2021). *Genomes of gut bacteria from* Nasonia *wasps shed light on phylosymbiosis and microbe-assisted hybrid breakdown* [Preprint]. Microbiology. https://doi.org/10.1101/2021.02.13.431100

Debnath, I., Stringer, A. M., Smith, S. N., Bae, E., Mobley, H. L., Wade, J. T., & Pearson, M. M. (2018). MrpJ Directly Regulates Proteus mirabilis Virulence Factors, Including Fimbriae and Type VI Secretion, during Urinary Tract Infection. *Infection and Immunity*, *86*(10), e00388-18.

Drzewiecka, D. (2016). Significance and Roles of Proteus spp. Bacteria in Natural Environments. *Microb Ecol*, *72*(4), 741–758. https://doi.org/10.1007/s00248-015-0720-6

Duan, Q., Zhou, M., Zhu, L., & Zhu, G. (2013). Flagella and bacterial pathogenicity: Flagella and bacterial pathogenicity. *Journal of Basic Microbiology*, *53*(1), 1–8. https://doi.org/10.1002/jobm.201100335

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319.

Gal-Mor, O., & Finlay, B. B. (2006). Pathogenicity islands: A molecular toolbox for bacterial virulence. *Cellular Microbiology*, *8*(11), 1707–1719.

Gibbs, K. A., Urbanowski, M. L., & Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. *Science*, *321*(5886), 256–259. https://doi.org/10.1126/science.1160033

Gode-Potratz, C. J., Kustusch, R. J., Breheny, P. J., Weiss, D. S., & McCarter, L. L. (2011). Surface sensing in Vibrio parahaemolyticus triggers a programme of gene expression that promotes colonization and virulence: Surface-responsive

gene expression. *Molecular Microbiology*, *79*(1), 240–263. https://doi.org/10.1111/j.1365-2958.2010.07445.x

Hacker, J., & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annual Reviews in Microbiology*, *54*(1), 641–679.

Hernandez, R. J., Hesse, E., Dowling, A. J., Coyle, N. M., Feil, E. J., Gaze, W. H., & Vos, M. (2019). Using the Wax moth larva Galleria mellonella infection model to detect emerging bacterial pathogens. *PeerJ*, *6*, e6150.

Howery, K. E., Clemmer, K. M., & Rather, P. N. (2016). The Rcs regulon in *Proteus mirabilis*: Implications for motility, biofilm formation, and virulence. *Curr Genet*, *62*(4), 775–789. https://doi.org/10.1007/s00294-016-0579-1

Hsu, C.-Y., Cairns, L., Hobley, L., Abbott, J., O'Byrne, C., & Stanley-Wall, N. R. (2020). Genomic Differences between *Listeria monocytogenes* EGDe Isolates Reveal Crucial Roles for SigB and Wall Rhamnosylation in Biofilm Formation. *Journal of Bacteriology*, *202*(7), e00692-19, /jb/202/7/JB.00692-19.atom. https://doi.org/10.1128/JB.00692-19

Jiang, S. S., Lin, T. Y., Wang, W. B., Liu, M. C., Hsueh, P. R., & Liaw, S. J. (2010). Characterization of UDP-Glucose Dehydrogenase and UDP-Glucose Pyrophosphorylase Mutants of Proteus mirabilis: Defectiveness in Polymyxin B Resistance, Swarming, and Virulence. *Antimicrobial Agents and Chemotherapy*, *54*(5), 2000–2009. https://doi.org/10.1128/AAC.01384-09

Josenhans, C., & Suerbaum, S. (2002). The role of motility as a virulence factor in bacteria. *International Journal of Medical Microbiology*, *291*(8), 605–614.

Little, K. (2017). *Outer Membrane Structures Regulate Proteus Mirabilis Swarm Motility* [Doctor of Philosophy]. Harvard University.

Little, K., Austerman, J., Zheng, J., & Gibbs, K. A. (2019). Cell shape and population migration are distinct steps of Proteus mirabilis swarming that are decoupled on high-percentage agar. *Journal of Bacteriology*, *201*(11), e00726-18.

Little, K., Tipping, M. J., & Gibbs, K. A. (2018). Swarmer cell development of the bacterium Proteus mirabilis requires the conserved enterobacterial common antigen biosynthesis gene rffG. *Journal of Bacteriology*, *200*(18), e00230-18.

Mathur, S., Sabbuba, N. A., Suller, M. T. E., Stickler, D. J., & Feneley, R. C. L. (2005). Genotyping of urinary and fecal *Proteus mirabilis* isolates from individuals with long-term urinary catheters. *European Journal of Clinical Microbiology & Infectious Diseases*, *24*(9), 643–644. https://doi.org/10.1007/s10096-005-0003-0

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, *15*(6), 589–594. https://doi.org/10.1016/j.gde.2005.09.006

Morgenstein, R. M., Szostek, B., & Rather, P. N. (2010). Regulation of gene expression during swarmer cell differentiation in Proteus mirabilis. *FEMS Microbiol Rev*, *34*(5), 753–763. https://doi.org/10.1111/j.1574-6976.2010.00229.x

Murfin, K. E., Lee, M.-M., Klassen, J. L., McDonald, B. R., Larget, B., Forst, S., Stock, S. P., Currie, C. R., & Goodrich-Blair, H. (2015). Xenorhabdus bovienii Strain Diversity Impacts Coevolution and Symbiotic Maintenance with Steinernema spp. Nematode Hosts. *MBio*, *6*(3), e00076-15. https://doi.org/10.1128/mBio.00076-15

Murphy, R., Palm, M., Mustonen, V., Warringer, J., Farewell, A., Parts, L., & Moradigaravand, D. (2021). Genomic Epidemiology and Evolution of *Escherichia coli* in Wild Animals in Mexico. *MSphere*, *6*(1), e00738-20, /msphere/6/1/mSph.00738-20.atom. https://doi.org/10.1128/mSphere.00738-20

Pearson, M. M., Rasko, D. A., Smith, S. N., & Mobley, H. L. (2010). Transcriptome of swarming *Proteus mirabilis*. *Infect Immun*, *78*(6), 2834–2845. https://doi.org/10.1128/IAI.01222-09

Pearson, M. M., Sebaihia, M., Churcher, C., Quail, M. A., Seshasayee, A. S., Luscombe, N. M., Abdellah, Z., Arrosmith, C., Atkin, B., Chillingworth, T., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Walker, D., Whithead, S., Thomson, N. R., … Mobley, H. L. T. (2008). Complete Genome Sequence of Uropathogenic Proteus mirabilis, a Master of both Adherence and Motility. *Journal of Bacteriology*, *190*(11), 4027–4037. https://doi.org/10.1128/JB.01981-07

Rather, P. N. (2005). Swarmer cell differentiation in *Proteus mirabilis*. *Environ Microbiol*, *7*(8), 1065–1073. https://doi.org/10.1111/j.1462-2920.2005.00806.x

Read, T. D., & Massey, R. C. (2014). Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: A new direction for bacteriology. *Genome Medicine*, *6*(11), 109. https://doi.org/10.1186/s13073-014-0109-z

Sajnaga, E., & Kazimierczak, W. (2020). Evolution and taxonomy of nematode-associated entomopathogenic bacteria of the genera Xenorhabdus and Photorhabdus: An overview. *Symbiosis*, *80*(1), 1–13.

Sakib, S. N., Reddi, G., & Almagro-Moreno, S. (2018). Environmental Role of Pathogenic Traits in *Vibrio cholerae*. *Journal of Bacteriology*, *200*(15), e00795-17, /jb/200/15/e00795-17.atom. https://doi.org/10.1128/JB.00795-17

Sambrook, J., & Russell, D. W. (2006). Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harbor Protocols*, *2006*(1), pdb.prot4455. https://doi.org/10.1101/pdb.prot4455

San, J. E., Baichoo, S., Kanzi, A., Moosa, Y., Lessells, R., Fonseca, V., Mogaka, J., Power, R., & de Oliveira, T. (2020). Current Affairs of Microbial Genome-Wide

Association Studies: Approaches, Bottlenecks and Analytical Pitfalls. *Frontiers in Microbiology*, *10*, 3119. https://doi.org/10.3389/fmicb.2019.03119

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., & Schmid, B. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., & Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539.

Simpson, B. W., & Trent, M. S. (2019). Pushing the envelope: LPS modifications and their consequences. *Nature Reviews Microbiology*, *17*(7), 403–416. https://doi.org/10.1038/s41579-019-0201-x

Sullivan, N. L., Septer, A. N., Fields, A. T., Wenren, L. M., & Gibbs, K. A. (2013). The Complete Genome Sequence of Proteus mirabilis Strain BB2000 Reveals Differences from the P. mirabilis Reference Strain. *Genome Announc*, *1*(5), e00024-13. https://doi.org/10.1128/genomeA.00024-13

Tenaillon, O., Skurnik, D., Picard, B., & Denamur, E. (2010). The population genetics of commensal *Escherichia coli. Nature Reviews Microbiology*, *8*(3), 207–217.

Vernikos, G., Medini, D., Riley, D. R., & Tettelin, H. (2015). Ten years of pan-genome analyses. *Current Opinion in Microbiology*, *23*, 148–154. https://doi.org/10.1016/j.mib.2014.11.016

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.

Wenren, L. M., Sullivan, N. L., Cardarelli, L., Septer, A. N., & Gibbs, K. A. (2013). Two independent pathways for self-recognition in Proteus mirabilis are linked by type VI-dependent export. *MBio*, *4*(4), e00374-13. https://doi.org/10.1128/mBio.00374-13

Chapter 4

Discussion & Future Directions

<u>Structural modeling, mutational analysis, and phylogeny as means to study protein</u>

<u>function</u>

IdsD and IdsE are identity determinants in *Proteus mirabilis* that lead to self or

non-self recognition during swarming (Gibbs et al., 2008). IdsD is transferred from one

cell to another through the type VI secretion system (Saak & Gibbs, 2016). When IdsE

is not present or does not bind to IdsD in the recipient cell, IdsD triggers a transient

stress response that reduces swarm expansion and excludes cells from the swarm front

(Cardarelli et al., 2015; Saak & Gibbs, 2016; Tipping & Gibbs, 2019). Factors in addition

to the Ids proteins and the secretion system were likely involved, though were

previously unknown.

In Chapter 2, I sequenced suppressor mutations in an unbiased screen that were

able to bypass swarm motility constraints due to non-self recognition (Figure 2.1). I

discovered that Ids signaling requires the predicted serine transporter gene *sdaC*

(Figure 2.1). SdaC, which is a LeuT-fold protein, likely samples an open and closed

conformation, which the results in Chapter 2 support. Sampling of the open

conformation is required for Ids signaling (Figure 2.2). As IdsD is predicted to localize to

the inner membrane along with its partner IdsE (Cardarelli et al., 2015; Zepeda-Rivera

et al., 2018), I hypothesize that IdsD interacts with the open conformation of SdaC to

insert into the inner membrane of the recipient cell.

Next steps require detecting a protein-protein interaction between IdsD and

SdaC. I worked on developing both *in vitro* and *in vivo* methods (Appendix B) to achieve

this goal, though these directions were ultimately not fruitful. SdaC is a membrane

protein that was difficult to solubilize and overexpress. *In vitro* immunoprecipitation

assays resulted in only a marginal amount of IdsD specifically eluting with SdaC (Figure A3). For *in vivo* colocalization, I observed IdsD degradation when overexpressed in *E. coli*, again resulting in marginal, if any, protein (Figure A5). Co-expression of the chaperone protein IdsC did not produce more full-length IdsD and is likely unable to reduce IdsD degradation (Figure A6). Thus, I decided to focus on the molecular mechanism of SdaC during self recognition.

Nutrient transporters such as SdaC are often highly conserved. Similar structures were studied in different conformations (Bozzi et al., 2019; Krishnamurthy & Gouaux, 2012). I was able to take advantage of this information to model the structure of SdaC using I-TASSER (Zhang, 2008). Rather than using random alanine scanning throughout the protein to systematically create SdaC variants, I aimed to use a more precise approach. Using the model as a starting point, I could begin to form a hypothesis about the function of point mutations that arose in the suppressor screen; both resulting amino acid mutations were predicted to bias the protein to a closed conformation. I then designed point mutants to bias the conformation or change substrate specificity (Figure 2.2), using the point mutants to probe the impact of protein function or conformation on self recognition.

Using phylogenetics, I considered homologs of SdaC to assess the range of sequence specificity: an ortholog from *E. coli* and a paralogous serine transporter YhaO in *P. mirabilis*. I showed that YhaO is a similar protein in terms of serine transport function, but does not enable self recognition in the same way as SdaC does (Figure 2.4). I then swapped SdaC and YhaO regions corresponding to a periplasmic loop to find the interaction interface while minimally disrupting or destabilizing the structure

(Figure A7). While I was unable to rewire the specificity by swapping the EL4 loop, a similar approach could be used to determine if other periplasmic regions of SdaC are required or if IdsD interacts with membrane-localized residues in the periplasm-facing vestibule of SdaC. Since I have not yet tested if all IdsD homologs use SdaC as a receptor, especially comparing those from the two IdsD subfamilies (1034 and 1072), it would be interesting to test if SdaC is the only IdsD receptor across strains. If not, similar swaps or mutations could be constructed for IdsD based on different strain variants to identify the interaction region of IdsD.

Through the combined use of strain construction, phylogenetics, structural models, and behavioral assays, I investigated the molecular basis of SdaC function in self recognition. This multifaceted approach has helped develop a preliminary model for the role of SdaC structure and conformational dynamics in self recognition and serine transport.

Interactions and crosstalk between metabolism, self recognition, and swarm motility

In Chapter 2, to investigate the function of SdaC in serine transport, I developed rapid and scalable assays for serine transport *in vivo*. I first constructed a high-serine strain by removing two primary serine metabolism genes, *sdaA* and *sdaB*. Mass spectrometry was used to confirm that removal of these enzymes resulted in accumulation of internal serine when SdaC was present (Figure 2.1). Next, I used what was known in *E. coli* about the impact of serine levels on cell physiology to develop two different assays as readouts for serine levels: swarm motility and growth in minimal medium when serine is added. I showed that, when SdaC is present, high intracellular serine levels suppress growth in minimal medium with serine and inhibit swarm

expansion (Figure 2.1). The high-serine strain in both liquid growth and swarm assays was then used to test the serine transport activity of SdaC variants (Figure 2.2).

While the high-serine strain was useful for detecting serine transport, I do not think this brute force method is an accurate measure of how more subtle fluctuations of internal serine levels may affect swarm motility. For example, differential regulation of serine metabolism during the swarm cycle could also affect the level of SdaC activity. Similarly, serine availability outside the cell, especially in a context outside laboratory conditions, could affect how SdaC samples different conformations. Serine is known to be an important nutrient both in the gut and in the urinary tract (Barroso-Batista et al., 2020; Brauer et al., 2019; Kitamoto et al., 2019). If sensory and regulatory pathways related to serine availability and serine metabolism are important for swarming, competition, and virulence, it's possible that SdaC is also a part of that regulatory network through differential regulation of its structure.

I showed that SdaC in the open conformation, but not the closed conformation, can enable self recognition (Figure 2.2). Conformational dynamics can represent, in part, a readout of serine availability inside or outside the cell. Thus the structure of SdaC would potentially link Ids signaling to metabolic and environmental changes regarding levels of serine, a key nutrient and molecular building block. High levels of serine in the environment could cause SdaC to sample the open conformation more frequently for serine uptake. I would then predict that more frequent exposure of the open conformation would also cause IdsD to more rapidly interact with SdaC and insert into the inner membrane for self recognition. In this way, SdaC structure may, in part, mediate serine metabolism's regulation of self recognition.

This connection between self recognition and serine metabolism is intriguing because IdsD induces an altered cell state in non-self cells. This state of cell stress includes high levels of the alarmone (p)ppGpp and increased antibiotic tolerance (Tipping & Gibbs, 2019). All of the molecular mechanisms through which IdsD induces an altered cell state are unknown, but it is possible that the effect is related to serine uptake.

One hypothesis is that IdsD binding to SdaC transiently reduces serine uptake, triggering the synthesis of (p)ppGpp to signal amino acid limitation. Indeed, the stringent response, which involves (p)ppGpp synthesis, is often induced by adding a serine analog called serine hydroxamate (Patacq et al., 2020). However, given our current model that IdsD requires SdaC to insert into the inner membrane before interacting with IdsE, even if the receiving cell encodes the cognate IdsE, IdsD-SdaC interaction would still occur. Based on the binding interaction with IdsD, IdsE would need to separately impact serine levels to counteract the effects of IdsD in kin cells. Thus, IdsD interaction with SdaC could be reducing serine uptake and triggering the production of (p)ppGpp to alter cell state, but this would require IdsE signaling in both self and non-self cells.

Alternatively, IdsD may interact with a downstream target, separate from SdaC, that is involved in serine metabolism. If IdsD is exploiting serine metabolism to impact cell physiology to induce stress in non-self cells, then serine uptake is required for IdsD activity. In this case, IdsD delivery into the recipient cell's inner membrane is linked to its downstream activation of stress. If IdsD evolved to target serine metabolism first, then co-opting the serine transporter may have evolved second because any recipient cell

that has the hypothetical serine target would also be likely to have the serine transporter SdaC.

Interaction with non-identity partners may constrain identity genes

SdaC is a membrane protein composed of a bundle of transmembrane helices that are packed together. Hydrophobic residues and helical interactions within the membrane are often constrained (Oberai et al., 2009). In the suppressor screen, I identified two mutant strains with G328V and G332R mutations in full-length SdaC (Figure 2.1). Based on my structural model of SdaC and conformational changes in LeuT-fold proteins, the G328V and G332R mutations were predicted to bias SdaC to a closed conformation (Figure 2.2). I did not isolate any mutant strains that could bypass non-self signaling without disrupting serine transport. Based on these results, mutations that arise to evade non-self recognition may also tend towards disrupting serine transport. In Chapter 2, I showed that SdaC is the dominant transporter during swarming (Figure 2.1). There may be a lower tolerance for mutations due to fitness tradeoffs, leading to purifying selection on SdaC. Consistent with this hypothesis, SdaC appears to be highly conserved across gammaproteobacteria. I showed that SdaC from *E. coli* is sufficiently conserved to replace SdaC from *P. mirabilis* and enable self recognition (Figure 2.4). IdsD may induce stress in other species during polymicrobial infection since *P. mirabilis* is frequently found along with *E. coli* and other species in the urinary tract. Fitness tradeoffs and structural constraints on SdaC may prevent evasion of non-self signaling in *P. mirabilis* and in other species.

Proteins that interact with IdsD can be important regulators of IdsD evolution. In other self recognition systems, there is often a focus on the variable identity proteins,

but the mechanism of non-identity partner proteins remains unknown (Chapter 1). IdsD depends on non-identity partners for signaling. IdsD interacts with a conserved chaperone protein IdsC in the donor cell prior to secretion through the type VI secretion system (Zepeda-Rivera et al., 2018). In Chapter 2, I showed that IdsD also depends on SdaC in the receiving cell (Figure 2.1). SdaC and IdsC are highly conserved across *P. mirabilis* strains. SdaC and IdsC represent non-identity partners that are required for self-recognition signaling through predicted interaction with IdsD.

One of the cruxes of self-recognition systems is that new alleles are constantly emerging and subjected to selection (de Oliveira et al., 2019). However, for these newly generated alleles to effectively signal self or non-self, the proteins must navigate the myriad of protein-protein interactions required—IdsD itself interacts with IdsC, SdaC, IdsE, and potentially other factors over the course of self recognition. If these interactions are not present for new alleles of IdsD, self recognition cannot occur. If IdsC and SdaC are under purifying selection, IdsD does not need to co-evolve and keep up with its non-identity partners in addition to its partner IdsE. It is unknown if the variable region of IdsD that determines interaction with IdsE is also the same region of IdsD that interacts with non-identity proteins IdsC and SdaC. If the variable region is involved in interactions with non-identity partners, the sequence space of IdsD and IdsE is partially constrained by the specificity and interface of those interactions.

Ecological variation as an untapped resource

In the same way that SdaC homologs allowed me to explore a range of phylogenetic variation, whole-genome variation across diverse ecological isolates provided insight into the ecology of *P. mirabilis*. This was possible through collaboration

with Dr. Daniel Utter and Dr. Colleen Cavanaugh, who brought a complementary expertise in genomics analysis to the project. Zehan Zhou, who worked with me on this project, and I were able to learn from them and develop these genomics tools as resources for the Gibbs lab going forward. For core research areas in the lab, such as the study of strain identity and self recognition, this genome dataset can be used to interrogate evolution of identity genes that are in the non-core in relation to the rest of the genome; components of the secretion machinery and other factors required for self recognition are in the core genome. Charles River Laboratories provides an interesting setting for this work because strain-level competition is occurring within animal host species as well as within a research facility with barriers to gene flow.

In Chapter 3, we successfully expanded into an underrepresented ecological niche by studying *P. mirabilis* isolates from asymptomatic animals, enriching our data through the inclusion of natural variation. By doing so, we identified patterns of behavioral variation in swarm motility within an ecological framework. Swarm motility has long been associated with virulence through the study of gene knockouts and co-regulation of genetic programs in clinical isolates (Howery et al., 2016; Pearson et al., 2010; Rather, 2005). However, the role of swarm motility outside of virulence, for example during a commensal lifestyle, is not well understood.

Our null hypothesis was that if *P. mirabilis* is not isolated from an infection, the incidence of swarm motility, as well as virulence capability, should be reduced if it is not being selected for, especially since it is thought to be an energetically costly behavior. We instead showed that the majority of isolates were able to swarm, and even those that were poor swarmers were clearly entering the swarm cycle as evidenced by our

observation of cell elongation on surfaces. Key virulence genes were present in the core genome and shared across all STAR isolates. Thus, there may be an unappreciated ecological role for swarming and virulence-associated genes beyond the context of pathogenesis. Also possible is that there are regulatory programs in place to control gene expression when *P. mirabilis* adapts to different situations. Transcriptional or translational data through RNAseq or proteomics methods would begin to answer if virulence genes or swarm motility genes are expressed outside of human infection. Changes in gene function or gene expression may occur more frequently within a population if those genes are not active.

Tapping into phenotypic variation may offer a more fluid range of phenotypes in certain ecotypes, which can lead to the uncoupling of different aspects of behavior. For example, we were able to disconnect the swarm cycle, surface motility and swim motility to look at each component's contribution to virulence. Genomic data led to the identification of specific genetic changes such as a SNP that results in truncation of the flagellar protein FliF in STAR58, leading to a non-motile swarmer cell population (Chapter 3). This phenotype resulting from FliF disruption was previously described using transposon mutagenesis in *P. mirabilis* (Belas & Suvanasuthi, 2005). However, there are many caveats to systematic mutagenesis screens including polar effects within an operon or important gene interactions that are missed in a single-gene approach. Studying the range of genetic and phenotypic variation present in different ecological contexts provides a relatively unbiased and ecologically-minded approach that is complementary to using techniques based in single reference strains.

Galleria mellonella: a model for virulence

112

As part of the work in Chapter 3, we adapted a protocol for using *Galleria mellonella* waxworms as a model for studying virulence. We found it important to connect behavioral and phenotypic differences back to virulence for impact and relevance. Waxworms are an emerging virulence model for probing the innate immune system (Hernandez et al., 2019; Ramarao et al., 2012). Adding virulence assays to standing procedures in the Gibbs Lab could provide fruitful results.

Our initial results reinforced models in the field that motility and swarmer cell development are important factors, but they are not necessarily the only ones. For example, we used BB2000 *Δugd* as an avirulent strain, which does not swarm and is deficient in LPS biosynthesis (Little, 2017); it was not clear if the reduced killing by some strains was due to cell envelope stress, sensitivity to the host innate immune system, inhibition of swarm development, or a combination of the three. This could be further explored by using LPS and O-antigen mutants previously constructed in the lab by Kristin Little (Little et al., 2018; Little et al., 2019). We could also screen LPS composition by extracting LPS from cells and visualizing the components using a silver-stained polyacrylamide gel. Significant changes such as differences in the banding pattern or loss of certain modifications can be traced back to sequence variation in the genome and tested for virulence in the waxworm model.

We could also further explore the role of motility outside of swarming by using chemotaxis mutants that I previously constructed. Having a convenient model system to test virulence potential may also inform how our research on self recognition and strain-level competition translates to human diseases. Several self-recognition strains in the lab could also be used to continue previous work on *P. mirabilis* Ids signaling in *C.*

elegans by Martha Zepeda-Rivera (Zepeda Rivera, 2018). Down the road, we could

potentially develop *in situ* techniques to study spatial aspects of strain-level interaction,

competition, and territoriality within host tissues by using labeled strains.

References

Barroso-Batista, J., Pedro, M. F., Sales-Dias, J., Pinto, C. J. G., Thompson, J. A., Pereira, H., Demengeot, J., Gordo, I., & Xavier, K. B. (2020). Specific Eco-evolutionary Contexts in the Mouse Gut Reveal *Escherichia coli* Metabolic Versatility. *Current Biology, 30*(6), 1049-1062.e7. https://doi.org/10.1016/j.cub.2020.01.050

Belas, R., & Suvanasuthi, R. (2005). The ability of *Proteus mirabilis* to sense surfaces and regulate virulence gene expression involves FliL, a flagellar basal body protein. *J Bacteriol, 187*(19), 6789–6803. https://doi.org/10.1128/JB.187.19.6789-6803.2005

Bozzi, A. T., Zimanyi, C. M., Nicoludis, J. M., Lee, B. K., Zhang, C. H., & Gaudet, R. (2019). Structures in multiple conformations reveal distinct transition metal and proton pathways in an Nramp transporter. *Elife, 8*, e41124.

Brauer, A. L., White, A. N., Learman, B. S., Johnson, A. O., & Armbruster, C. E. (2019). D-Serine degradation by *Proteus mirabilis* contributes to fitness during single-species and polymicrobial catheter-associated urinary tract infection. *MSphere, 4*(1).

Cardarelli, L., Saak, C., & Gibbs, K. A. (2015). Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding. *MBio, 6*(3), e00251-15.

de Oliveira, J. L., Morales, A. C., Stewart, B., Gruenheit, N., Engelmoer, J., Brown, S. B., de Brito, R. A., Hurst, L. D., Urrutia, A. O., & Thompson, C. R. (2019). Conditional expression explains molecular evolution of social genes in a microbe. *Nature Communications, 10*(1), 1–12.

Gibbs, K. A., Urbanowski, M. L., & Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. Science, 321(5886), 256–259. https://doi.org/10.1126/science.1160033

Hernandez, Rafael J., Elze Hesse, Andrea J. Dowling, Nicola M. Coyle, Edward J. Feil, Will H. Gaze, and Michiel Vos. "Using the Wax Moth Larva Galleria Mellonella Infection Model to Detect Emerging Bacterial Pathogens." PeerJ 6 (2019): e6150.

Howery, K. E., Clemmer, K. M., & Rather, P. N. (2016). The Rcs regulon in *Proteus mirabilis*: Implications for motility, biofilm formation, and virulence. *Curr Genet, 62*(4), 775–789. https://doi.org/10.1007/s00294-016-0579-1

Kitamoto, S., Alteri, C. J., Rodrigues, M., Nagao-Kitamoto, H., Sugihara, K., Himpsl, S. D., Bazzi, M., Miyoshi, M., Nishioka, T., & Hayashi, A. (2019). Dietary l-serine confers a competitive fitness advantage to *Enterobacteriaceae* in the inflamed gut. *Nature Microbiology*, 1–10.

Krishnamurthy, H., & Gouaux, E. (2012). X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature*, *481*(7382), 469.

Little, K. (2017). *Outer Membrane Structures Regulate Proteus Mirabilis Swarm Motility* [Doctor of Philosophy]. Harvard University, Graduate School of Arts & Sciences.

Little, K., Austerman, J., Zheng, J., & Gibbs, K. A. (2019). Cell shape and population migration are distinct steps of *Proteus mirabilis* swarming that are decoupled on high-percentage agar. *Journal of Bacteriology*, *201*(11), e00726-18.

Little, K., Tipping, M. J., & Gibbs, K. A. (2018). Swarmer cell development of the bacterium *Proteus mirabilis* requires the conserved enterobacterial common antigen biosynthesis gene rffG. *Journal of Bacteriology*, *200*(18), e00230-18.

Oberai, A., Joh, N. H., Pettit, F. K., & Bowie, J. U. (2009). Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proceedings of the National Academy of Sciences*, *106*(42), 17747–17750.

Patacq, C., Chaudet, N., & Létisse, F. (2020). Crucial Role of ppGpp in the Resilience of *Escherichia coli* to Growth Disruption. *Msphere*, *5*(6).

Pearson, M. M., Rasko, D. A., Smith, S. N., & Mobley, H. L. (2010). Transcriptome of swarming *Proteus mirabilis. Infect Immun*, *78*(6), 2834–2845. https://doi.org/10.1128/IAI.01222-09

Ramarao, Nalini, Christina Nielsen-Leroux, and Didier Lereclus. "The Insect Galleria Mellonella as a Powerful Infection Model to Investigate Bacterial Pathogenesis." JoVE (Journal of Visualized Experiments), no. 70 (2012): e4392.

Rather, P. N. (2005). Swarmer cell differentiation in *Proteus mirabilis. Environ Microbiol*, *7*(8), 1065–1073. https://doi.org/10.1111/j.1462-2920.2005.00806.x

Saak, C. C., & Gibbs, K. A. (2016). The Self-Identity Protein IdsD Is Communicated between Cells in Swarming *Proteus mirabilis* Colonies. *Journal of Bacteriology*, *198*(24), 3278–3286.

Tipping, M. J., & Gibbs, K. A. (2019). Peer pressure from a *Proteus mirabilis* self-recognition system controls participation in cooperative swarm motility. *PLoS Pathogens*, *15*(7), e1007885.

Zepeda Rivera, Martha Adriana. (2018). *Molecular Mechanism for Targeting a Self-Identity Protein to the Type VI System in Proteus Mirabilis*. [Doctoral of Philosophy]. Harvard University, Graduate School of Arts & Sciences.

Zepeda-Rivera, M. A., Saak, C. C., & Gibbs, K. A. (2018). A proposed chaperone of the bacterial type VI secretion system functions to constrain a self-identity protein. *J Bacteriol*. https://doi.org/10.1128/JB.00688-17

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 40.

Appendix A

Identification of functional domains of self-recognition protein IdsE

Achala Chittor, Amy Hao, Akeem Pinnock, and Karine A. Gibbs

Abstract

The opportunistic pathogen *Proteus mirabilis* uses swarm motility to migrate across surfaces. One regulator of swarm expansion is self recognition between cells. Neighboring cells in a swarming population exchange the identity signal protein IdsD, which interacts with its partner protein IdsE in recipient cells. If IdsD and IdsE are compatible, then swarming is proficient, otherwise swarming is restricted. IdsE does not fit the model of a standard immunity protein, since it contains a C-terminal region that is separate from the variable region implicated in IdsD interaction. Using random mutagenesis, we found that mutations in IdsE across the length of the protein, including in the variable region, can disrupt swarm expansion completely or partially. However, mutations in between the variable region and C-terminal domain do not affect swarm expansion. We propose that the C-terminal domain may mediate downstream signaling by IdsE based on binding affinity with IdsD.

Introduction

      Proteus mirabilis is a gram-negative bacterium that is an opportunistic human pathogen. On surfaces, *P. mirabilis* cells enter a developmental cycle that oscillates between short, non-motile consolidated cells and long, motile swarmer cells that are elongated and hyperflagellated (Gibbs & Greenberg, 2011). Swarm motility is, in part, regulated by the Ids (Identification of self) proteins IdsD and IdsE; when IdsD and IdsE are from the same strain, they interact and swarm motility is more efficient (Cardarelli et al., 2015; Saak & Gibbs, 2016). IdsD and IdsE both contain variable regions that vary extensively between strains, and exchanging these variable regions can rewire IdsD-IdsE binding and population behavior outcomes (Cardarelli et al., 2015). IdsD is predicted to be transferred between cells through the type VI secretion system and interact with its partner IdsE in the inner membrane of the receiving cell (Saak & Gibbs, 2016).

      IdsE does not fit neatly into the predominant model of immunity proteins because it also contains a sizeable C-terminal region that is conserved. In this study, we aim to understand the functional domains of IdsE: which regions are important for IdsD-IdsE interaction and does the C-terminal region serve an additional regulatory function in downstream signaling? Here, we performed random mutagenesis on IdsE and screened the swarm expansion phenotype to identify functional domains for IdsD interaction and downstream signaling. We found that the C-terminal domain may play a role in downstream signaling and that there may be a flexible linker between the variable region and the C-terminal domain.

Results

To identify functional domains of IdsE, we decided to use random mutagenesis as an unbiased method for creating a library of IdsE variants. IdsE was expressed under its native promoter with the rest of the Ids locus with a C-terminal GFP fluorescent tag. The IdsE variant-containing plasmids were cloned into *E. coli* and mated in *P. mirabilis* for functional screening (Figure A1). To rapidly screen IdsE variants, the clones were grown in 96-well plates and inoculated on swarm-permissive plates in batch. The swarm phenotype was recorded and the GFP fluorescence in the swarm colony was measured using a Tecan plate reader.

The swarm expansion phenotype and IdsE-GFP fluorescence were used to categorize the clones into four classes. Class A clones were GFP-negative and showed restricted swarming. Class B clones were GFP-positive and restricted in swarming. Class C clones were GFP-negative and proficient in swarming. Class D clones were GFP-positive and proficient in swarming. The breakdown of these classes for the initial set of mutants is shown in Table A1. While we were hoping to collect the Class B clones expressing full-length IdsE (GFP-positive) that resulted in defects in swarm expansion, we found a very low proportion of these (roughly 2%). Instead, we observed a greater incidence in Class C (roughly 77%). When we did a first pass of the plasmid sequencing, we found that class C clones tended to have large regions of the p*ids*BB plasmid missing. The mutations in *idsE* may have reduced overall fitness, leading to a prevalence of plasmid deletions that would increase fitness by turning off Ids activity.
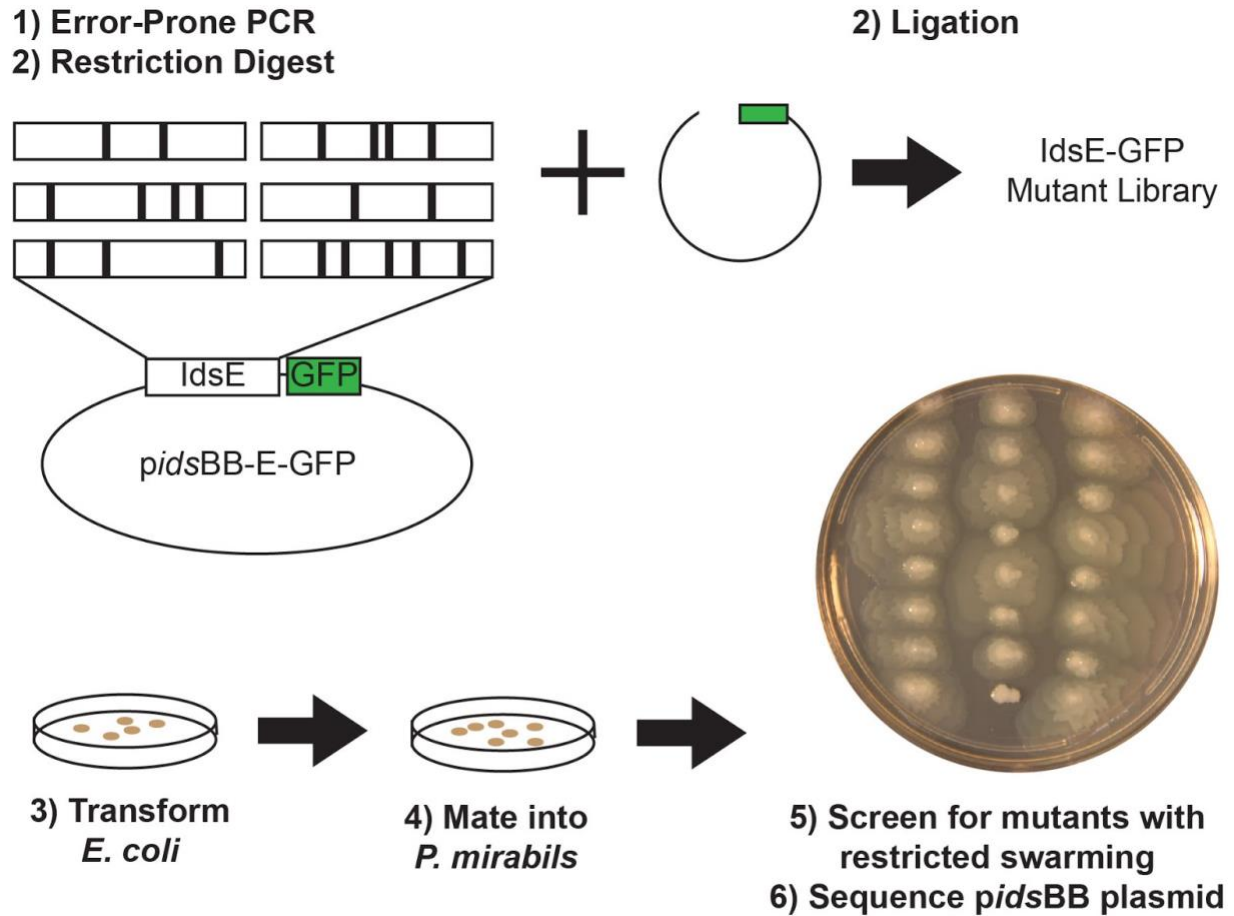
1) Error-Prone PCR
2) Restriction Digest

2) Ligation

IdsE-GFP
Mutant Library

IdsE GFP

p*ids*BB-E-GFP

3) Transform
E. coli

4) Mate into
P. mirabils

5) Screen for mutants with restricted swarming
6) Sequence p*ids*BB plasmid

Figure A1: Experimental setup for IdsE random mutagenesis screen.
The *idsE* sequence was amplified using an error-prone polymerase and ligated back into a p*ids*BB expression vector with a C-terminal GFP fusion. These ligation products were transformed into an *E. coli* cloning strain and then into an *E. coli* mating strain to be mated into *P. mirabilis* BB2000 Δ*ids*. *P. mirabilis* clones were screened for swarm expansion as well as IdsE-GFP fluorescence. A subset of plasmids were miniprepped and sequenced.

Table A1: IdsE mutagenesis screen class totals

	GFP fluorescence	Swarm expansion	Number	Percent
Class A	Negative	Restricted	18	8
Class B	Positive	Restricted	5	2
Class C	Negative	Proficient	183	77
Class D	Positive	Proficient	33	14
Total			239	100

We then decided to take a subset of sequences and submit them for sequencing of the entire plasmid. We included a set of clones from each class to get a broader sense of where the mutations in IdsE were occurring and the functional impact on swarm expansion. The sequencing results are summarized in Table A2.

We mapped these full-length IdsE sequence mutations onto the amino acid sequence based on whether they resulted in proficient or restricted swarm expansion (Figure A2). Overall, the mutations were scattered throughout the sequence, validating our random, unbiased experimental approach. We found that for proficient swarm expansion, the mutations in IdsE were clustered in the first predicted transmembrane region as well as between the variable region and the predicted cytoplasmic C-terminal domain. For restricted swarm expansion, IdsE mutations occurred throughout the protein, including the C-terminal domain. Also, the restricted swarm expansion phenotype varied from intermediate to severe (Figure A2). All full-length IdsE variants that resulted in restricted swarm expansion recovered swarm expansion when expressed in a strain background defective in IdsD secretion, indicating that any IdsE downstream signaling requires IdsD secretion.

Table A2: IdsE sequence mutations

Plate ID	Sample ID	Amino Acid mutations	Class
prelim	2	truncation at 43	A
prelim	5	F37L L48M F77V N90T L158P V182D P272Q	A
092716S17A	D3	T66I, L159F, S225*	A
092716S17A	H3	T66I, L159F, S225*	A
092716S17B	B3	T66I, L159F, S225*	A
092716S17B	G11	K91N, truncation at 184	A
092716S17B	H11	Y96H, Y104*	A
110716S17A1	C1	F37C, P114A, L133I, V167L, F175S, truncation at 203	A
110716S17A1	D1	E81*	A
110716S17A1	E1	V230F, K281*	A
110716S17A1	E5	V230F, K281*	A
110716S17A1	E9	S131F, I160F, truncation at 222	A
110716S17A3	E5	V230F, K281*	A
110716S17B1	E1	truncation at 134	A
110716S17B1	E3	F37C, P114A, L133I, V167L, F175S, truncation at 203	A
110716S17B2	G1	E81K, Y262*	A
110716S17B2	G7	deletion plasmid	A
110716S17B2	G9	I132N, truncation at 162	A
020917plate3	B1	Y121*	A
092716S17A	E11	K85N, L113V, N126K, I166N, N188Y, T248N, N260D	B
092716S17A	G9	K85N, L113V, N126K, I166N, N188Y, T248N, N260D	B
110716S17A2	E1	D151Y, T192S, T214I, L215F, P288L, K303Q	B
110716S17A3	E9	D151Y, T192S, T214I, L215F, P288L, K303Q	B
110716S17B2	G5	L48Q, F64C, I208F, V230L	B
020917plate5	G11	Q45L, D95V	B
prelim	1	K197T, Y241F, truncation at 321 (in GFP)	C
092716S17A	C11	E81* (deletion plasmid)	C

Table A2: IdsE sequence mutations (continued)

| Plate ID	Sample ID	Amino Acid mutations	Class
092716S17A	F5	deletion plasmid	C
092716S17A	G3	deletion plasmid	C
092716S17B	E9	deletion plasmid	C
110716S17A2	G3	D151Y, T192S, T214I, L215F, P288L, K303Q (and disrupted promoter)	C
110716S17A3	G9	deletion plasmid	C
020917plate2	D1	K51E, E81*	C
020917plate3	A1	K51E, E81*	C
020917plate3	F3	K51E, E81*	C
prelim	3	R55L G63A S84L L93F R194S R202H T216I	D
prelim	4	V74M F217Y F284S	D
092716S17A	C9	K107N, T147A, V198I	D
092716S17A	F3	K177R, G207D, I208T	D
092716S17A	F7	K107N, T147A, V198I	D
092716S17B	B7	L162I	D
092716S17B	F5	T97I	D
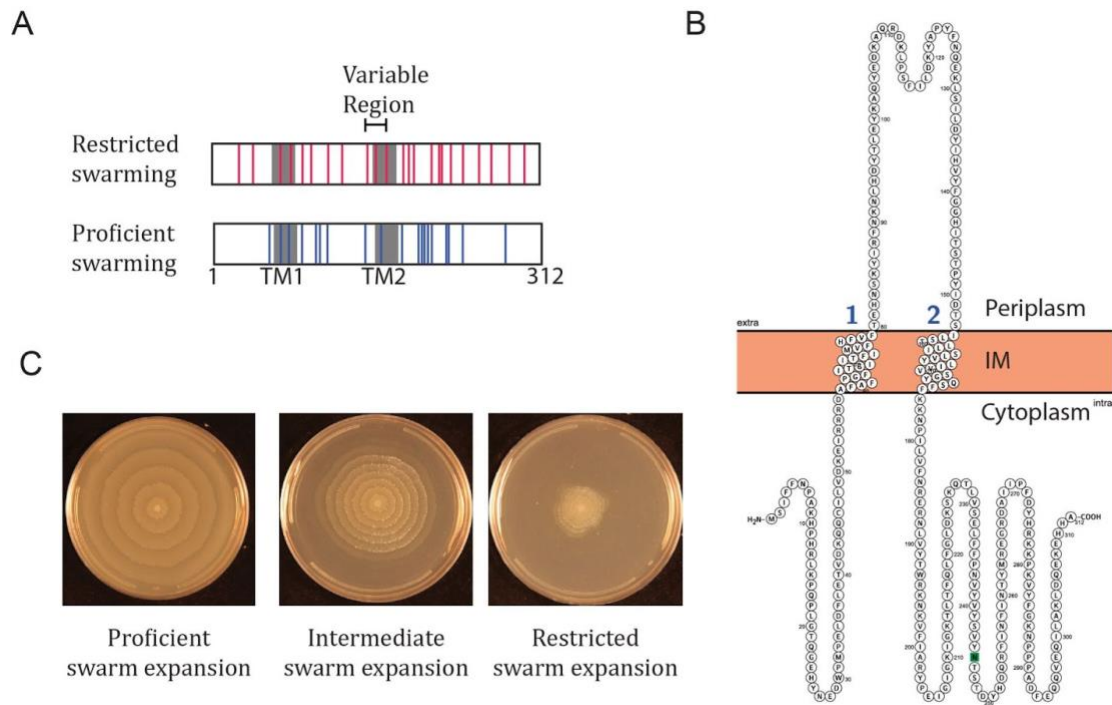
Figure A2: Distribution of sequence variation in IdsE and range of swarm phenotypes.
A) Mutations in IdsE from screen (only predicted full-length IdsE) mapped along amino acid sequence from restricted and proficient swarmers. B) Predicted IdsE membrane topology generated using Protter. C) Swarm expansion phenotypes produced by IdsE mutants range from proficient to intermediate to fully restricted.

Discussion

In our random mutagenesis of IdsE, we generated many deletion plasmids,

indicating that there may be a fitness defect from mutating IdsE that leads to large

deletion events in the p*ids*BB plasmid. We observed mutations throughout the IdsE

sequence, including in the C-terminal domain, that resulted in restricted swarm

expansion, indicating the importance of the C-terminal domain for Ids signaling. The C-

terminal domain is, in fact, homologous to other bacterial sequences that do not appear

to contain the N-terminal region of IdsE. We also observed a higher incidence of

mutations in the region between the variable region and the C-terminal domain that did

not disrupt swarm expansion, so this region may be a flexible linker region between

specialized domains. IdsE constructs lacking parts of the C-terminal domain exhibit

weakened binding to IdsD and intermediate swarming (Pinnock, 2018). Therefore, it

may not be just the variable region that is required for IdsD interaction even though the

variable region sequence determines strain-specific binding (Cardarelli et al., 2015).

The 246-248 TST to AAA mutation does not appear to disrupt IdsD binding, but does

change the appearance of swarming and can even restrict swarm expansion when

overexpressed (data not included here), possibly through IdsE signaling activity.

However, all signaling is dependent on IdsD secretion, which allows for IdsD-IdsE

interaction in the receiving cell. We propose that IdsE can potentially titrate the

downstream signaling response and recognition outcome according to the strength of its

binding with IdsD. It would be interesting to test this hypothesis by further investigating

specific residues in the C-terminal domain that emerged in this study and identifying

downstream signaling factors that interact with IdsE.

<u>Materials and Methods</u>

Bacterial strains and media

The strains and plasmids used in this study are described in Table A3. *P. mirabilis* strains were maintained on low swarm (LSW) agar (Belas et al., 1991). CM55 blood agar base agar (Oxoid, Basingstoke, England) was used for swarm-permissive nutrient plates. Overnight cultures of all strains were grown at 37°C in LB broth under aerobic conditions. Kanamycin (Corning, Corning, NY) was used at a concentration of 35 µg/ml for plasmid maintenance and was added to swarm and growth media when appropriate.

Random mutagenesis of IdsE

Plasmids were constructed by amplifying the *idsE* gene using oAC006 and oAC007 (Table S2) from the pIdsBB expression system containing a C-terminal GFPmut2 fusion (Gibbs et al., 2008) using error-prone PCR with the GeneMorph II Random Mutagenesis Kit (Agilent, Santa Clara, CA) and ligated back into the same pIdsBB expression vector using the restriction enzymes SacI and BamHI. The resultant plasmids were transformed into *E. coli* Omnimax before transforming into *E. coli* S17λ*pir* and mating into *P. mirabilis* BB2000 Δ*ids.* Mutant plasmids were generated in four sets, labeled by date: prelim, 092716, 110716, and 020917. Plates within the set are labeled with a letter and number (e.g. A1). The well number is indicated by its position in the 96-well plate (e.g. G5).

Phenotypic classification

Plasmids were categorized into 4 classes based on GFP fluorescence (from IdsE-GFP) and swarm expansion. Only select strains from set 020917 were categorized, so this set was excluded completely from the totals in Table A1. Overnight cultures grown in 96-well plates were replica plated onto both CM55 swarm-permissive medium and LSW non-swarming medium and grown for two days at room temperature. Strains BB2000 Δ*ids* empty vector, BB2000 Δ*ids* pIdsBB-*idsE-GFP,* BB2000 Δ*ids* pIdsBB-Δ*idsE,* BB2000 Δ*ids* pIdsBB-Δ*idsCD, and* BB2000 Δ*ids* pIdsBB-IdsE-ex were used as controls. Glycerol was added to a final concentration of 30% to the 96-well plates containing liquid culture and they were labeled with the designated "plate ID" and stored at -80C. From the swarm plate, GFP fluorescence was measured using the Tecan Infinite 200 PRO microplate reader (Tecan, Männedorf, Switzerland). Strains with restricted swarm expansion were identified. To validate the restricted swarm phenotype, overnight cultures were normalized to an OD600 of 0.1, and CM55 swarm-permissive nutrient plates were inoculated with 2 µl of normalized culture in the center. Plates were incubated for two days at room temperature. The media was supplemented with kanamycin to maintain the plasmid. Full length IdsE mutants that resulted in restricted swarming were re-transformed into a BB2000 Δ*ids* *vipA*-T95G background to test for function without IdsD secretion.

Plasmid sequencing

For a subset of 66 strains, the plasmids were miniprepped and quantified using the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA). The pIdsBB-IdsE variants were sequenced by Seqwell (Beverly, MA). Some samples were removed from analysis that did not produce high quality sequencing data or assemblies. All

deletion plasmids (roughly half the length and missing most of the *ids* locus) from class

D were also removed from further analysis since class D should have positive GFP

fluorescence from IdsE expression. Mutations in IdsE from the remaining 42 plasmids

are listed in Table A2. Mutations from full-length IdsE (highlighted in red in Table A2)

are mapped in Figure A2.

Table A3: Strains used in Appendix A

Strain	Description	Source	KAG#	AC#
P. mirabilis				
BB2000 *Δids* c. pBBR2-GFP	Wild-type BB2000 strain with the *ids* locus deleted carrying an empty vector (promoterless GFP)	Karine Gibbs	KAG67	AC18
BB2000 *Δids* c. pidsBB-*ΔidsCD*	BB2000 *Δids* carrying a plasmid expressing the *ids* locus from its native promoter with *idsCD* removed.	Karine Gibbs	KAG23	AC20
BB2000 *Δids* c. pidsBB-*ΔidsE*	BB2000 *Δids* carrying a plasmid expressing the *ids* locus from its native promoter with *idsE* removed.	Christina Saak	KAG1923	AC21
BB2000 *Δids* c. pidsBB-IdsE-exchange	BB2000 *Δids* carrying a plasmid expressing the *ids* locus from its native promoter with *idsE* containing a variable region swap with HI4320	Christina Saak	KAG1483	AC22
BB2000 *Δids* c. pidsBB-IdsE-GFP	BB2000 *Δids* carrying a plasmid expressing the *ids* locus from its native promoter with a C-terminal GFP fusion on IdsE	Christina Saak	KAG1455	AC23
BB2000 *Δids* *vipA*-T95G	BB2000 *Δids* containing a point mutation T95G in VipA that disrupts type VI secretion	Christina Saak	KAG2115	AC67
E. coli				
One Shot Omnimax 2 T1R	Cloning strain	Thermo Fisher Scientific, Waltham, MA	KAG2183	
S17λ*pir*	Mating strain	(Simon et al., 1983)	KAG68	

References

Belas, R., Erskine, D., & Flaherty, D. (1991). Transposon mutagenesis in Proteus mirabilis. *J Bacteriol*, *173*(19), 6289–6293.

Cardarelli, L., Saak, C., & Gibbs, K. A. (2015). Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding. *MBio*, *6*(3), e00251-15.

Gibbs, K. A., & Greenberg, E. P. (2011). Territoriality in Proteus: Advertisement and aggression. *Chem Rev*, *111*(1), 188–194. https://doi.org/10.1021/cr100051v

Gibbs, K. A., Urbanowski, M. L., & Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. *Science*, *321*(5886), 256–259. https://doi.org/10.1126/science.1160033

Pinnock, A. (2018). *A structure-function analysis of IdsE, a self-identity protein involved in the self-and non-self-recognition system of* Proteus mirabilis [Unpublished bachelor's thesis]. Harvard University.

Saak, C. C., & Gibbs, K. A. (2016). The Self-Identity Protein IdsD Is Communicated between Cells in Swarming Proteus mirabilis Colonies. *Journal of Bacteriology*, *198*(24), 3278–3286.

Simon, R. U., Priefer, U., & Pühler, A. (1983). A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. *Bio/technology*, *1*(9), 784-791.

Appendix B

Investigation of interaction between self-recognition protein IdsD and

serine transporter SdaC

Achala Chittor and Karine A. Gibbs

Abstract:

We hypothesize that the serine transporter SdaC acts as an inner membrane receptor for the self-recognition protein IdsD, which is transferred between *Proteus mirabilis* cells during collective swarm expansion. Here we show that there is some evidence of IdsD-SdaC interaction *in vitro*, and describe the initial setup and troubleshooting of an *in vivo* interaction assay in *E. coli.* We also show that the EL4 periplasmic loop can be swapped between SdaC and YhaO without affecting self recognition signaling. Therefore, the membrane-localized pocket may in fact be the interaction interface rather than the periplasmic loops. Further exchange of regions of SdaC and YhaO will allow us to identify regions of SdaC that are important for interaction with IdsD.

Introduction:

      The serine transporter SdaC is required for self recognition and regulation of swarm expansion in *Proteus mirabilis*. Based on single residue mutations in SdaC that bias the conformation to open or closed, we previously concluded that the open conformation, without serine transport, is sufficient for self recognition (Chapter 2). Based on reports that SdaC acts as an inner membrane receptor for microcin V and phage C1 (Gérard et al., 2005; Likhacheva et al., 1996) and that the secreted contact-dependent inhibition protein CdiA can utilize different inner membrane proteins for delivery into the host cell (Willett et al., 2015), we proposed that SdaC acts as receptor to tether IdsD to the inner membrane so that it can insert into the membrane. IdsD is a predicted membrane protein and so is its recognition partner IdsE.

      In this work, we used *in vitro* and *in vivo* experimental approaches to detect IdsD interaction with SdaC. We first used co-immunoprecipitation to pull down SdaC and test for the presence of IdsD. Using a previously-described assay (Lim & Bernhardt, 2019), We next expressed SdaC as a polar localized bait and tested for polar localization of periplasmic IdsD using microscopy of live *E. coli* cells. We also continued to develop genetic constructs based on our previous results that YhaO has a similar predicted structure (LeuT fold) and can transport serine (Chapter 2). Based on the predicted structure and our hypothesis that IdsD is interacting with SdaC in the periplasm, we hypothesized that exposed loops may be required for binding; extracellular loops of outer membrane proteins are often implicated in binding (Chatterjee & Rothenberg, 2012; Ruhe et al., 2013; Sharma et al., 2007). We therefore swapped a periplasmic loop between SdaC and YhaO to test if IdsD activity and self recognition is also swapped.

<u>Results:</u>

To test if SdaC physically interacts with IdsD, we used a batch co-immunoprecipitation assay to pull down 3xFLAG-SdaC or FLAG-BAP in swarmer cell lysates and test for the presence of IdsD in the elution fraction. Sigma70, a housekeeping protein, is visible in the load and flow-through for both samples, but not in the elution, as expected (Figure A3). The anti-FLAG blot shows clear elution of both 3xFLAG-SdaC and FLAG-BAP. IdsD was detected in the elution fractions from 3xFLAG-SdaC samples, but not for FLAG-BAP samples (Figure A3). The presence of IdsD in the 3xFLAG-SdaC elution indicates that IdsD and SdaC interact *in vitro*. However, there is still a large amount of IdsD in the flow-through fraction as well, which is IdsD that did not bind to SdaC. It's possible that the interaction is weak or transient. It's also possible that the *in vitro* conditions are not completely effective for stabilizing the interaction. IdsD is predicted to be secreted into the periplasm of the recipient cell, so the interaction with SdaC would also be predicted to take place in the periplasm, and the *in vitro* setup does not recapitulate that reducing environment.

To study the IdsD-SdaC interaction *in vivo*, we turned to an assay from the Bernhardt lab that utilizes the polar-localized protein PopZ from *Caulobacter crescentus* (Lim & Bernhardt, 2019). The bait vector (pHCL149) co-expresses PopZ with a tagged GFP N-terminally fused to the protein of interest or can be used as a control on its own with the single transmembrane domain. The prey vector (pHCL151) is a periplasmically localized (DsbA signal sequence) N-terminal mScarlet fusion to the protein of interest. The bait protein localizes to the poles due to interaction with polar-localized PopZ in the cytoplasm. If the prey protein interacts with the bait protein, in this case in the
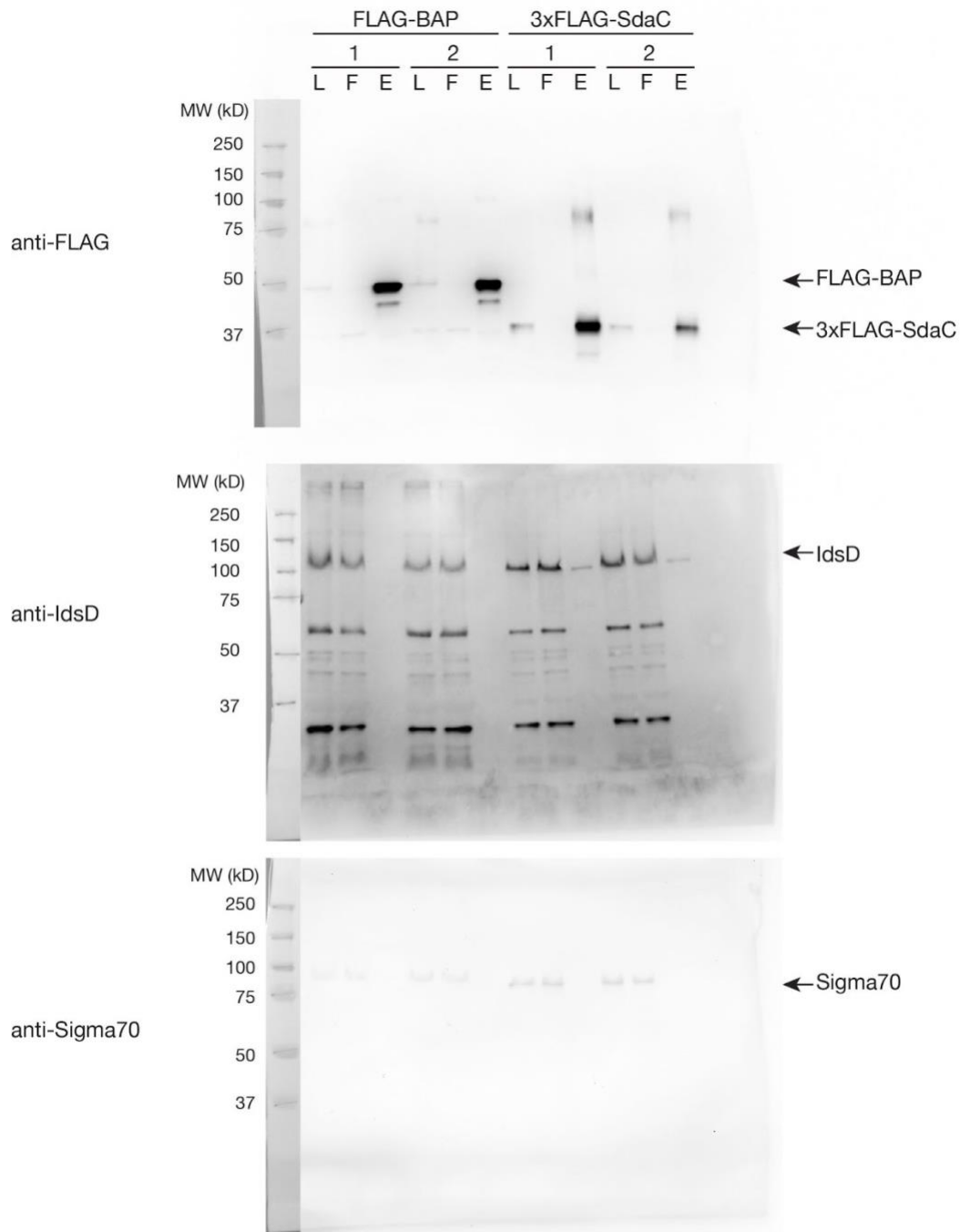
Figure A3: SdaC interacts with IdsD *in vitro*.
Western blots on swarmer cells from ACH01 (BB2000 *ΔidsE ΔsdaC*) carrying an empty vector (with FLAG-BAP added as a control) or ptet-3xFLAG-SdaC, where FLAG-tagged proteins were pulled down using anti-FLAG resin and tested for IdsD presence using an anti-IdsD antibody. Anti-FLAG antibody was used to confirm presence of 3xFLAG-SdaC and FLAG-BAP. Anti-Sigma70 was used as a housekeeping control. Fractions shown are load (L), flow-through (FT), and elution (E) for two biological replicates.

periplasm, the prey protein should also localize to the poles with the bait. We inserted SdaC into the bait vector in place of the transmembrane domain and inserted IdsD into the prey vector. If SdaC and IdsD interact *in vivo* in the periplasm, we would expect mScarlet-IdsD localization to be polar when expressed with GFP-SdaC but not the GFP control. We did initial microscopy of both the bait (GFP) and prey (RFP) along with a membrane dye (DAPI) to better visualize the polar-localized bait (Figure A4). While the two control proteins and GFP-SdaC showed the expected localization (GFP at the poles, RFP along the periphery), mScarlet-IdsD was forming puncta that were not consistently colocalized with SdaC or the GFP control and sometimes occurred in the midcell rather than at the poles.

To further investigate why IdsD was exhibiting such an abnormal subcellular localization pattern, we decided to perform periplasmic extraction followed by western blot analysis to detect IdsD in the periplasm (Figure A5). We used both an anti-RFP antibody and the anti-IdsD antibody, as well as anti-GFP for the bait proteins and anti-G6PD antibody as a housekeeping control along with a general Coomassie stain. We found that while the control mScarlet protein was clearly visible in the periplasmic fraction using the anti-RFP antibody, mScarlet-IdsD was much more faint. Interestingly, because the IdsD antibody recognizes the N-terminal sequence and because mScarlet is also an N-terminal fusion, we detected several degradation bands in the periplasmic fraction with a darker band around the size of mScarlet. This band indicated that a very minimal section of the N-terminus from IdsD was staying fused to mScarlet after mScarlet-IdsD was degraded. Based on this result, we concluded that the mScarlet we observed in the microscopy experiment is coming from a degraded form of
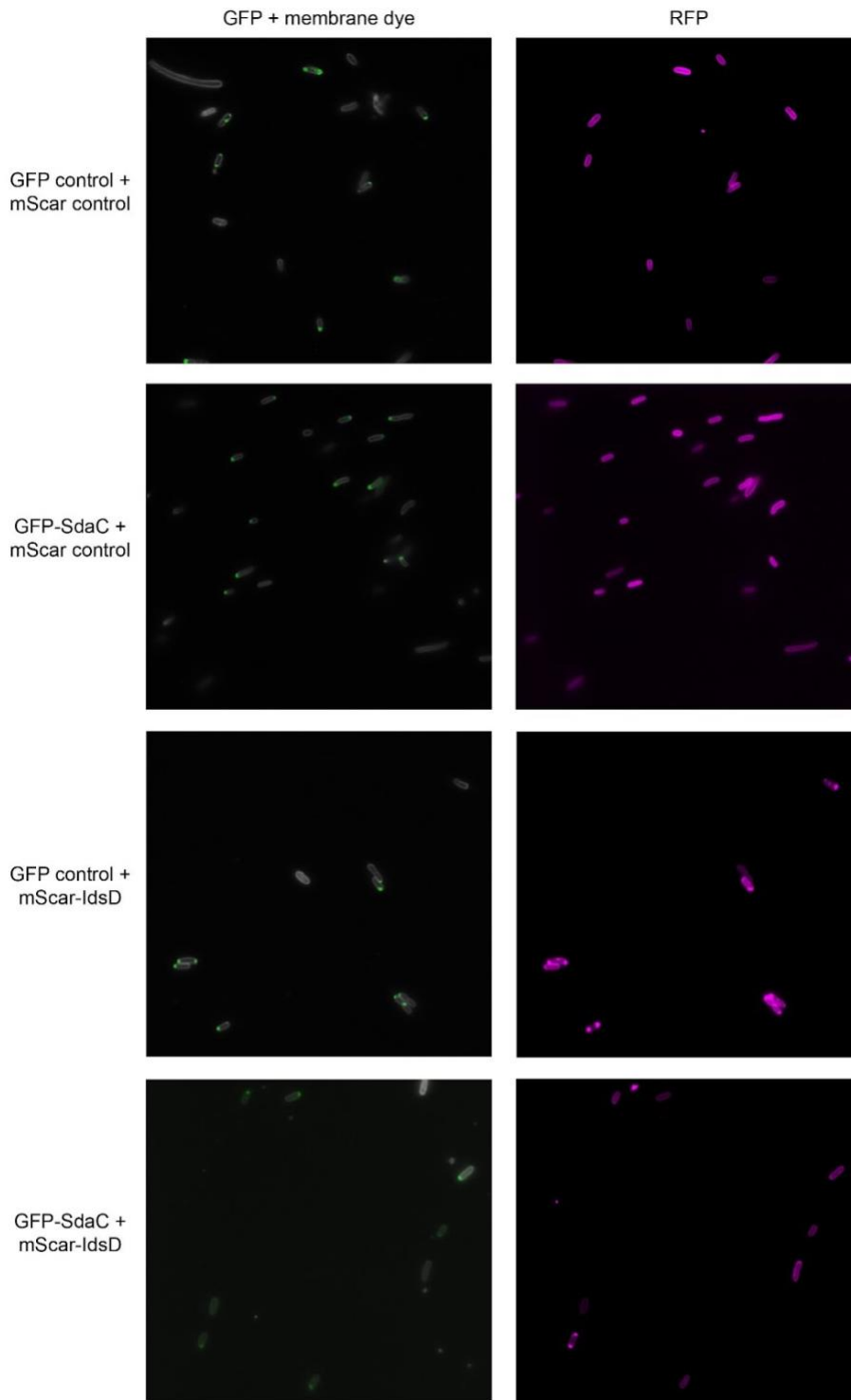
Figure A4: Initial IdsD-SdaC interaction assay using PopZ polar localization.
GFP plus membrane dye (TMA-DPH in the DAPI channel) on left and RFP on right for
strains co-expressing either GFP control or GFP-SdaC with DsbA$_{ss}$-mScar control or
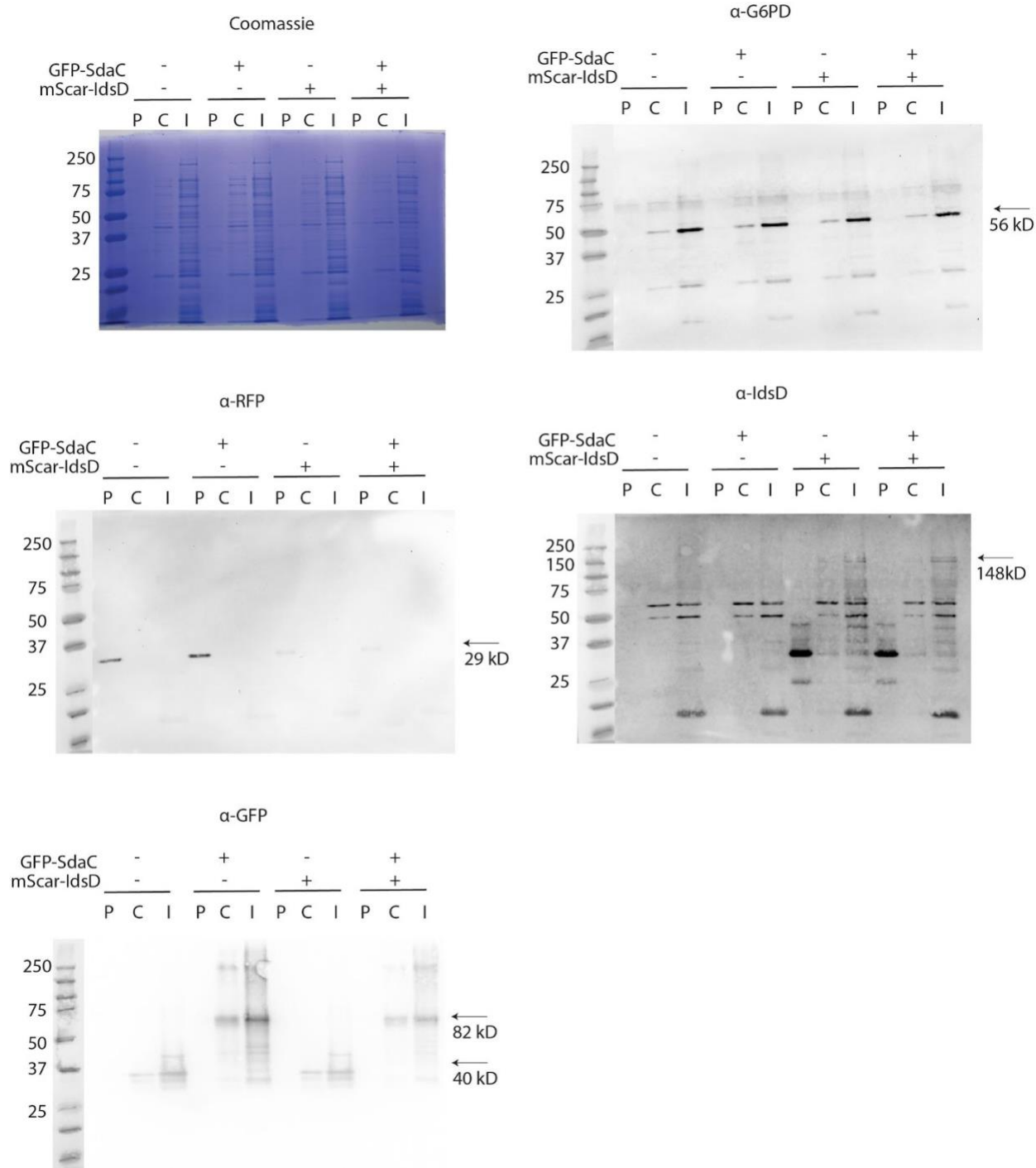DsbA$_{ss}$-mScar-IdsD

Figure A5: IdsD is not stable as a full-length construct in PopZ assay conditions. Coomassie stain and western blots for anti-G6PD, anti-RFP, anti-IdsD, and anti-GFP for cells co-expressing a combination of either GFP control or GFP-SdaC and DsbA$_{ss}$-mScar control or DsbA$_s$mScar-IdsD. Fractions are periplasmic (P), soluble (C), and insoluble (I). G6PD is a housekeeping control.

mScarlet-IdsD rather than the full length protein. Thus, the interaction assay is not going to effectively detect any SdaC-IdsD protein-protein interaction because IdsD is not stable.

To try to stabilize IdsD, we thought co-expression of its chaperone, IdsC, might help (Zepeda-Rivera et al., 2018). We inserted the *idsC* sequence ahead of DsbA$_s$mScar-IdsD to mimic the operon and transcriptional organization in the native *ids* locus. We then repeated the periplasmic extraction and western blot analysis on strains expressing the mScarlet vector alone: the control, IdsD, or IdsC with IdsD (Figure A6). We found that when IdsD was co-expressed with IdsC, we still did not detect full-length mScar-IdsD in the periplasmic fraction. In fact, even the degradation bands, including the mScarlet-sized band, became much fainter, indicating increased degradation than without IdsC. It was at this point that we decided not to move forward with the PopZ assay for now, given that we do not understand the best way to stably overexpress IdsD in the periplasm of *E. coli* for this kind of *in vivo* interaction assay.

When we previously looked at SdaC homologs, we found that *E. coli* SdaC, but not YhaO, was not able to support self recognition even though they were both similar proteins in terms of serine transport and predicted structure (Chapter 2). Since we hypothesize that IdsD interacts with SdaC in the periplasm to allow for IdsD to insert into the inner membrane, it is possible that an exposed periplasmic loop is required for IdsD binding. Many of the loops are very short, and another long loop is variable between *P. mirabilis* and *E. coli*. This leaves what is known as the EL4 loop or extracellular loop 4 (Krishnamurthy & Gouaux, 2012). The EL4 loop is actually important for transport function, as the tip of the loop fits into the exposed substrate-binding
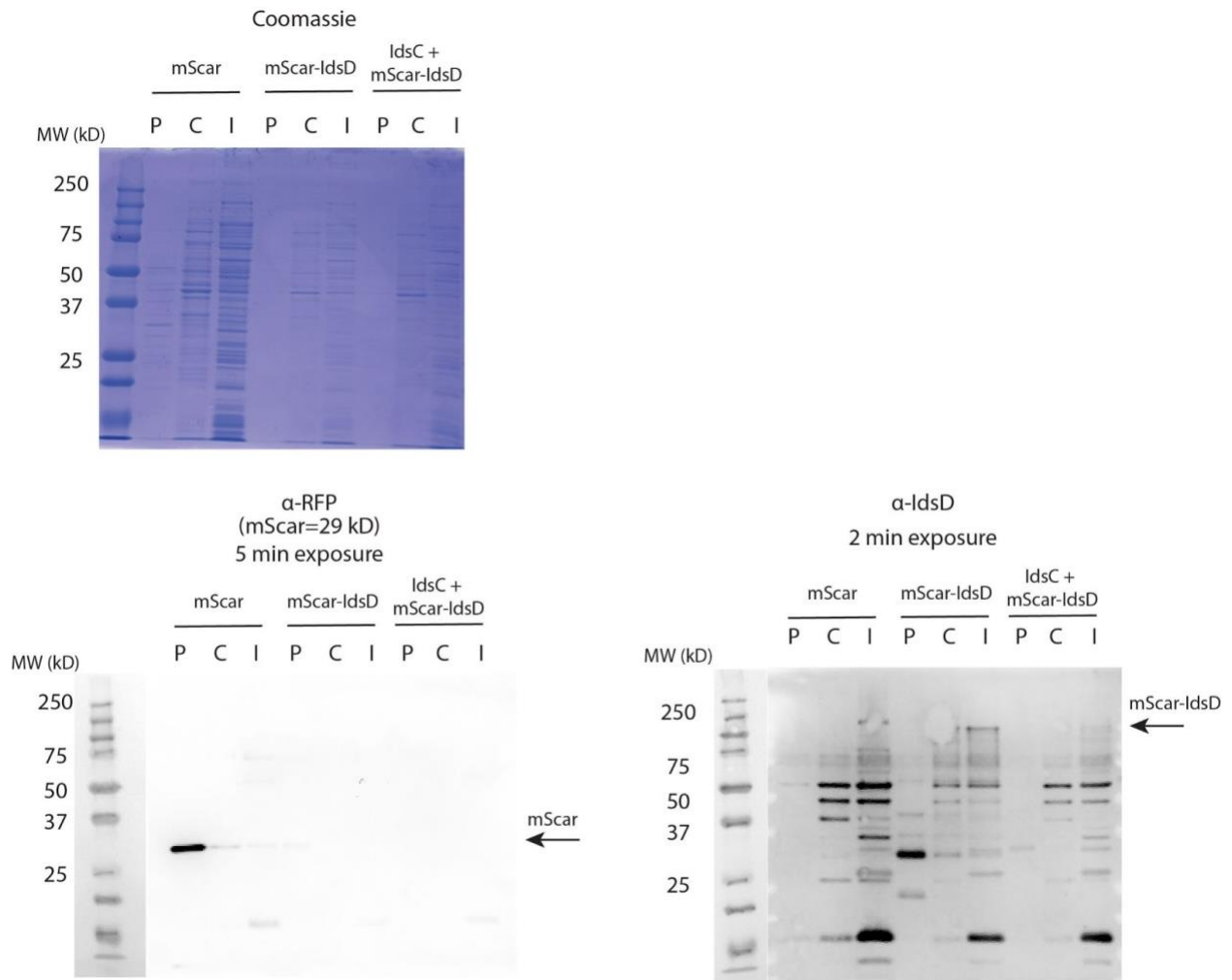
Figure A6: Co-expression of chaperone IdsC does not improve IdsD stability in PopZ assay.
Coomassie stain and western blots for anti-RFP and anti-IdsD for cells expressing either DsbA$_s$mScar-IdsD or IdsC co-expressed with DsbA$_s$mScar-IdsD. Fractions are periplasmic (P), soluble (C), and insoluble (I).
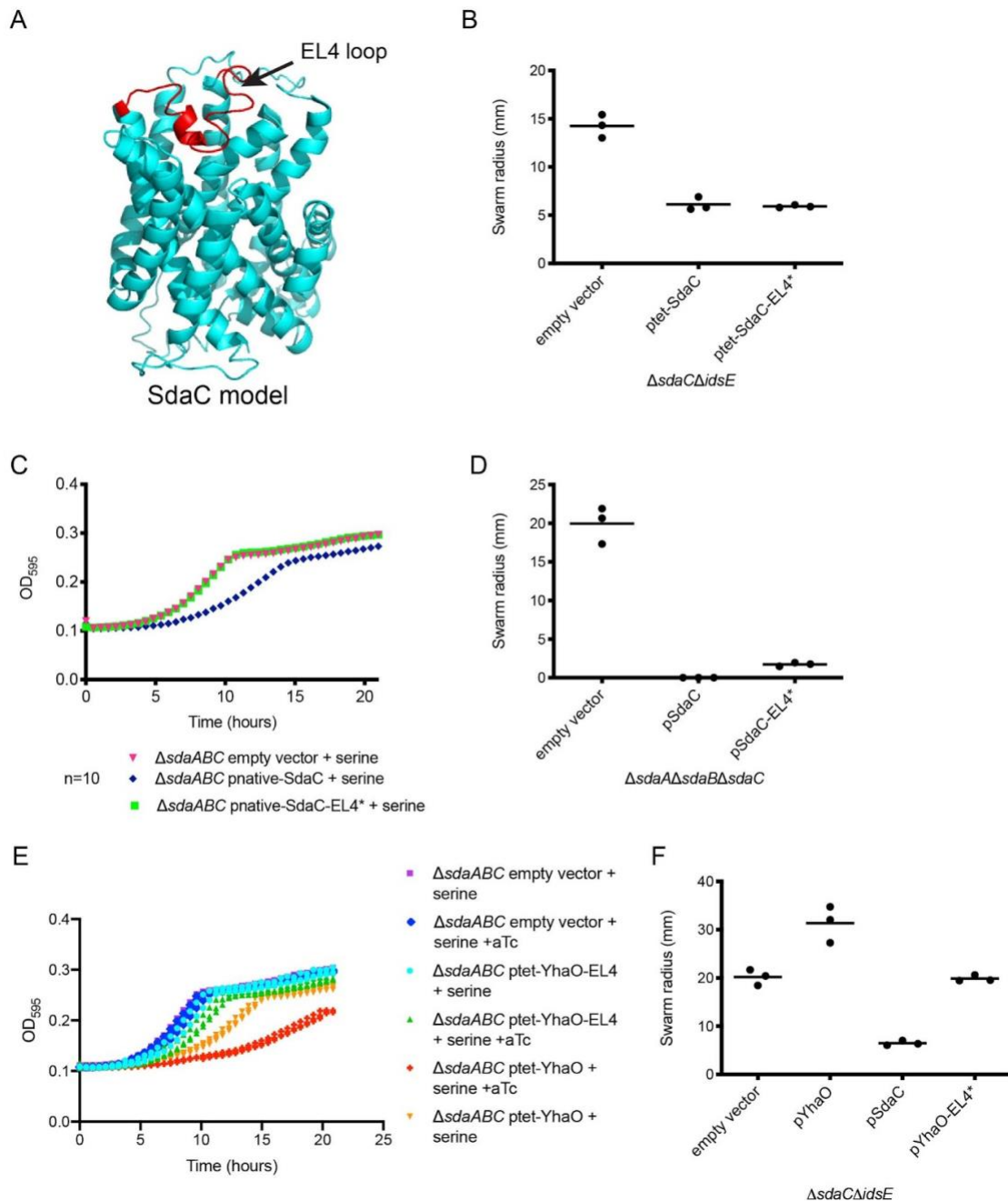
Figure A7: Periplasmic loop "EL4" in SdaC is required for serine transport, but does not appear to be required for self recognition.
A) I-TASSER model of SdaC in cyan with predicted periplasmic EL4 loop highlighted in red using Pymol. B) Swarm expansion for BB2000 *ΔsdaCΔidsE* (ACH01) carrying empty vector, ptet-SdaC, or ptet-SdaC-EL4*. C) Growth curve for BB2000 *ΔsdaAΔsdaBC* (ACH05) carrying empty vector, pSdaC, or pSdaC-EL4* for an average of 10 biological replicates. D) Swarm expansion average for ACH05 carrying empty vector, pSdaC, or pSdaC-EL4*. E) Growth curve for ACH05 carrying empty vector, ptet-YhaO, or ptet-YhaO-EL4* with or without aTc inducer. F) Swarm expansion for ACH01 carrying empty vector, ptet-YhaO, ptet-SdaC, or ptet-YhaO-EL4*.

143

periplasm-facing pocket (Figure A7). While the functionally relevant tip of the loop is fairly conserved even between SdaC and YhaO, the sides of the loop around where the loop hinges off from the transmembrane helices is more variable. Since the loop is known to shift position between the open and closed conformations (Bozzi et al., 2019) and the open conformation is specifically required for self recognition, we hypothesized that IdsD may interact with the EL4 loop only when SdaC is in the open conformation.

To test this hypothesis, we decided to swap the EL4 loop sequence between SdaC and YhaO. If IdsD binds the EL4 loop and this binding is specific to the SdaC sequence rather than the YhaO sequence, swapping the loops between the two proteins should also swap self recognition function. When we measured swarm expansion of the non-self strain expressing SdaC-EL4* (EL4 swapped in from YhaO), swarm expansion was still restricted, comparable to with the wildtype SdaC. We also tested SdaC-EL4* expressed from its native promoter in the high serine strain background, and observed no growth defect in minimal medium plus serine, but an intermediate swarm defect, indicating some loss of transport activity. The YhaO-EL4* (EL4 swapped from SdaC) also showed no growth defect when expressed in the high serine background in minimal medium plus serine, indicating a loss of transport function. When expressed in the non-self background, there was no restriction of swarm expansion, indicating YhaO-EL4* is still unable to function in self recognition. These results suggest that the EL4 loop is not required for hypothesized IdsD interaction. However, further characterization and additional constructs would provide a more complete picture since both EL4 swaps appear to have defects in serine transport, which may affect conformational dynamics or folding in unforeseen ways.

Discussion

IdsD interacts with SdaC *in vitro*, although weakly. Because SdaC is so hydrophobic and has a tendency to aggregate in solution, it is hard to optimize the co-immunoprecipitation experiment further to improve binding conditions. At this point, it is unclear if the interaction during self recognition *in vivo* is itself weak or transient or if the experimental conditions are not fully optimized. We hoped that the PopZ interaction assay would provide a solution to this issue by allowing us to study the IdsD-SdaC interaction *in vivo* in *E. coli* cells. However, we found that IdsD is not stable under these conditions of overexpression in the periplasm, even when co-expressed with its chaperone, IdsC. An alternative for future work may be to truncate IdsD to mimic potential processing that IdsD undergoes during secretion and reduce the size of the protein to the minimal domain required for SdaC interaction, likely near the variable region and C-terminus.

The periplasmic loop EL4 does not seem to be required for interaction with IdsD. We swapped the predicted EL4 sequence between SdaC and YhaO, which disrupted serine transport, but did not change the effect on IdsD activity and self recognition. Given the requirement for the open conformation, it's possible that IdsD interacts with the membrane-localized transmembrane helices when they are exposed to the periplasm. This would be really interesting since transmembrane domains of membrane proteins tend to be more structurally constrained than the aqueous, exposed portions outside of the membrane (Oberai et al., 2009). If the interaction interface is more constrained than for outer membrane receptors (Ruhe et al., 2013), there may be unforeseen additional constraints on IdsD evolution.

Materials and Methods

Bacterial strains and media

P. mirabilis strains were maintained on low swarm (LSW) agar (Belas et al., 1991).

CM55 blood agar base agar (Oxoid, Basingstoke, England) was used for swarm-

permissive nutrient plates. Overnight cultures of all strains were grown at 37°C in LB

broth under aerobic conditions. For growth curve assays, cells were grown in minimal

medium [M9 salts (3 g/L KH_2PO_4, 6.8 g/L Na_2HPO_4, 0.5 g/L NaCl, 1.0 g/L NH_4Cl), 2 mM

$MgSO_4$, 0.1 mM $CaCl_2$, 0.2% glucose] supplemented with 10mM L-serine (VWR,

Beantown chemical, BT128350) when stated. Kanamycin (Corning, Corning, NY),

tetracycline (VWR, Radnor, PA), and chloramphenicol (Sigma Aldrich, St. Louis, MO)

were used at a concentration of 35 µg/ml, 15 µg/ml, or 35 µg/ml, respectively, for

plasmid maintenance and was added to swarm and growth media when appropriate.

Anhydrotetracycline (Sigma-Aldrich, St. Louis, MO) was used to induce gene

expression from the Tet promoter at a concentration of 10 nM in the medium when

stated.

Swarm expansion

Overnight cultures were normalized to an OD600 of 0.1, and swarm-permissive nutrient

plates supplemented with kanamycin were inoculated with 2 µl of normalized culture in

the center. Plates were incubated at room temperature for two days, and the radii of

actively migrating swarms starting from the edge of the inoculum were measured using

Fiji (ImageJ) (Schindelin et al., 2012).

Growth curve

Overnight cultures were normalized to an optical density at 600 nm (OD600) of 0.1 in

minimal medium [M9 salts (3 g/L KH_2PO_4, 6.8 g/L Na_2HPO_4, 0.5 g/L NaCl, 1.0 g/L NH_4Cl),

2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% glucose] supplemented with 10 mM L-serine (VWR,

Beantown chemical, BT128350) and 10 nM anhydrotetracycline as an inducer when

stated. Both were supplemented with kanamycin for plasmid maintenance when

appropriate. Normalized cultures were grown overnight at 37°C, with periodic shaking,

in a Tecan Infinite 200 PRO microplate reader (Tecan, Männedorf, Switzerland).

Anti-FLAG co-immunoprecipitation assay

Anti-FLAG immunoprecipitation assays were performed and analyzed as described

previously for *Proteus mirabilis* cells (Cardarelli et al., 2015). Modifications are as

follows. *P. mirabilis* cells were harvested by centrifugation from 5 swarm-permissive

plates after incubation at 37°C for 16 to 20 h. Sample buffer was modified by adding 8M

Urea (Sigma Aldrich, St. Louis, MO) to improve solubility. Protein samples were

incubated at 37C for 1h to limit aggregation that can occur from boiling.

SDS-PAGE and Western Blot analysis

Assays were performed and analyzed as described previously (Cardarelli et al., 2015).

Modifications as follows. Samples for the immunoprecipitation experiment were

separated using 10% Tris-tricine polyacrylamide gel electrophoresis, whereas all other

samples for the PopZ experiments were separated using 12% Tris-tricine

polyacrylamide gel electrophoresis. Polyclonal primary antibodies (and dilutions) were

as follows: rabbit anti-IdsD (1:2,000), rabbit anti-RFP (GenScript, Piscataway, NJ),

rabbit anti-G6PD (Assaypro, St. Charles, MO). Monoclonal primary antibodies (and

dilutions) were rabbit anti-FLAG (1:4,000; Sigma-Aldrich, St. Louis, MO) and mouse anti-σ^{70} (1:1,000; Thermo Fisher Scientific, Waltham, MA). Secondary antibodies were goat anti-rabbit or goat anti-mouse antibodies conjugated to horseradish peroxidase (HRP) (polyclonal; dilution, 1:5,000; SeraCare Life Sciences, Milford, MA). The Western blots are not quantitative. For coomassie staining, gels were left in the stain (45% ethanol, 5% acetic acid, 0.5% coomassie) overnight with shaking and then incubated in destain solution (45% ethanol, 5% acetic acid) for 1 hour with shaking before imaging on a light box.

Microscopy

TMA-DPH membrane dye (Thermo Fisher Scientific, Waltham, MA) was added to the cultures 10 minutes prior to imaging. Pads (1xM9 salts, 1.5% agarose) of approximately 0.5mm thickness were inoculated with 5μl droplets of induced culture and dried for 5 minutes before adding a coverslip. The pads were imaged by phase-contrast microscopy (10ms exposure) and fluorescence microscopy (rfp, gfp, and dapi channels, 500ms exposure) using a Leica DM5500B microscope system (Leica Microsystems, Buffalo Grove, IL) and a CoolSnap HQ2 cooled charge-coupled device (CCD) camera (Photometrics, Tucson, AZ). MetaMorph (version 7.8.0.0; Molecular Devices, Sunnyvale, CA) was used for image acquisition.

PopZ protein production and periplasmic extraction

PopZ assay protocol was adapted from the Bernhardt group (Lim & Bernhardt, 2019). Overnight cultures were grown in LB with 1% glucose (to prevent protein expression that may affect fitness) containing chloramphenicol (Cm) and tetracycline (Tet). In a

baffled 125ml flask, 25ml LB broth with antibiotics was inoculated with 500μl from the overnight culture and grown for 2 hours shaking at 37°C. Cells were pelleted by centrifuging cultures for 10min at 4000rpm and resuspend in 25ml induction medium (M9, 0.2% casamino acids, 0.2% glycerol, 100μM IPTG, 0.2% Arabinose, Cm+Tet) and grown for 2 hours shaking at 37°C. Cells were pelleted by centrifuging for 10min at 4000rpm and decanting the supernatant. The following protein preparation steps were performed in a 4°C cold room. The periplasmic extraction protocol was adapted from method 2 (Cold osmotic shock with MgCl2) of a recent study (Malherbe, 2019). Roche cOmplete protease inhibitor cocktail (Sigma Aldrich, St. Louis, MO) was added to all buffers at a 25x dilution. The cell pellet was washed with 850μl PBS. Cell suspension was transferred to 1.5ml tubes and spun down for 10min at 4000rpm. Cell pellet was next resuspended in 900μl spheroplast buffer (0.1 M Tris pH 8.0, 500 mM sucrose, 0.5 mM EDTA pH 8.0) and incubated for 5min. Cells were spun down for 10min at 4000rpm and resuspended in 400μl 1mM MgCl2. After 15s, 20μl 20mM MgSO4 was added. Cells were spun down for 10min at 4000rpm. A 250μl periplasmic fraction was collected into a separate tube and mixed with 250μl sample buffer and incubated at 37°C for 1 hour before storing at -80°C. After decanting the remaining supernatant, cells were resuspended in 1ml lysis buffer (50mM Tris pH 7.4, 150mM NaCl, 1mM EDTA, 1% Triton X-100). The cell suspension was then transferred to a chilled 2ml tube with ~100mg of 0.1mm cell disruptor beads (Scientific Industries, Bohemia, NY). The tubes were wrapped in parafilm and placed in a 50ml falcon tube which was vortexed at full speed for 10 minutes. Cells were then spun down at 15,000rpm for 5 min. A 250μl soluble fraction was collected into a fresh tube and mixed with 250μl sample buffer and

incubated at 37°C for 1 hour before storing at -80°C. The remaining pellet was resuspended in 250μl sample buffer for the insoluble fraction and incubated at 37°C for 1 hour before storing at -80°C.

Plasmid construction

Restriction digestion was performed using restriction enzymes described (New England BioLabs, Ipswich, MA). Oligos and gBlocks were ordered from IDT (Coralville, IA). Ligations were resolved in OneShot Omnimax2 T1R competent cells (Thermo Fisher Scientific, Waltham, MA) or BW25141pir cells (Datsenko & Wanner, 2000). The resultant plasmids were confirmed by Sanger sequencing (Genewiz, Inc., South Plainfield, NJ), and correct resultant plasmids were then transformed into *P. mirabilis* as described previously (Gibbs et al., 2008) using *E. coli* conjugative strain MFDpir (Ferrières et al., 2010).

For ptet-SdaC-EL4* and pSdaC-EL4*, the EL4 sequence from *yhaO* with surrounding regions was ordered as a gBlock (5'-
TGATTATGTTAGTACTTGCTTTTCCTATGACCTTTTTTGCACACCGTGGTATGTGTC
GCTTTGTGCTATCAGGTAAAAATCCAGGTGAAGATATTACTGAAGTGGTTGAAGAA
CACTTTGGTAAAACAGCAGGTGTATTAATCACTCTGCTCTATTTCTTTGCTATTTATC
CTATCTTGTTGGTTTACAGTGTCGCTATCACTAATACGGTAGAAAGCTTTATTGTTC
ACCAAATGCATATGACAGCACCACCTCGTGCGATTTTATCACTGGTACTGATTGTC
GGTATTATGTGTATCATTCGCTTTGGTGAACAAGCTATTGTTAAAGCAATGAGTGTC
TTAGTATTCCCATTCGTTGCCATCTTGATGGTACTTGCTTTATACCTCATTCCTGAAT
GGAATGGAGCCATTCTTGATACGCTCTCTTTTGATCACGCATCAACTTCTGGTATGA
GCCAAGGCTTGTTAGTAACCTTATGGTTAGCTATCCCTGTGATGGTGTTCTCTTTTA

ACCACTCACCTATCATCTCTGCTTTTGCTGTGGCAAAACGCGAAGAGTACGGTGAC

AATGCAGAGAAAAATGTTCACGCATCTTAGCTTATGCTCATATTATGATGGTGATC

ACTGTTATGTTCTTCGTGTTTAGCTGCGTATTTAGCTTAGGACATGATGAAGCAGTA

AAAGCTTATGAGCAAAATATCTCTGCATTAGCGATTGCGGCAAAATTCTTCCCTGG

CGGTGTTATCGCTTACATTGCACCTTTCATTGCCTTTGTTGCGATTACTAAATCTTT

CTTAGGTCACTATTTAGGGGCTCGTGAAGGTTTCAATGGCTTAATTATCAAAGGGA

TGCGCGGAATTGGTAAGAATGTCGAAAAGATCGTTTAAACAAAATCACAG -3') and

ligated into the ptet-SdaC (pAC49) or pSdaC (pAC12) vector using restriction enzymes

ScaI and PmeI. For ptet-YhaO-EL4*, the EL4 sequence from *sdaC* with surrounding

regions was ordered was a gBlock (5'-

TGCCGGACGTTTAATTAAAGAAGCCATTATCACACTACCTTTTACATTAACTTCTATT

TTGTTTATCCAAACATTAAGCCCAATGGTTATCTCTTATCGTAGTCGTAACCAAAAT

CGTGAAGTAGCGCGTCATAAAGCATTACGTGCAATGAATATCGCTTTCGGTGTGTT

GTTCTGTACGGTGTTTTTCTATGCAATCTCCTTCACATTAGCAATGACACCAGAAAA

TTTAGCTGAAGCAAAAGAGCAAAATATCAGTATTCTGTCTTATCTTGCTAACCACTT

TGAAGCACCCTGGGTAACCGTAGTGAGTGTGATGTTGAATATTTTCGCAGTAATGA

CCGCATTCTTTGGTGTCTATCTCGGTTTCCGTGAAGCAACACAAGGAATTGTGATG

AATATTCTGCAACGCATGATGCCCGTTGAAAAAATCAATGAAAAATGGGTACAAAA

CGGCATTATGATCTTTGCTGTGCTGTTAGCTTGGGGCGCCATTATTCTCAATGCAC

CAGTACTGAGCTTCACCTCAATTTGTAGTCCTATCTTTGGTATGGTTGGTTGTCTTA

TCCCTGCATACCTAGTGTACAAAGTACCGATGTTACATAAATACAAAGGCGCATCA

CTTTATCTGATTATCTTTACGGGTCTGCTACTTTGTGTTTCTCCGTTCTTGGCATTCT

CATAGTATCCCTGAGGCCAGTTTGCT -3') and ligated into ptet-YhaO (pAC43) using

restriction enzymes PacI and Bsu36I. For pAC79, the *sdaC* sequence was amplified

from pAC12 using oAC310 (5'-

TTCTGGCGGTGGATCCGATACAACCAAAGCTGGTTCTATCGCACA -3') and oAC311

(5'- CTTAGTCGACCTCGAGTTAGTTAAAGAAATAATTAATGAAAGTATTACCTAAGGT

-3') and ligated into pHCL149 using the restriction enzymes BamHI and XhoI. For

pAC84, the *idsD* sequence was amplified from pIdsBB using oAC308 (5'-

TCTGGGCGGTGGATCCACTGGAGAAGTGAATGAGAAATATTTAAC -3') and oAC309

(5'- GATAAGCTTACTCGAGGATATTCTCACTGTTAATAAAGCCTAAAAGCTGA -3')

and ligated into the pHCL151 vector using restriction enzymes BamHI and XhoI. For

pAC90, the *idsC* sequence with mScarlet as part of it was ordered as a gBlock (5'-

AGGAGATATACATATGCTCTTGAGTCCAAATCCCCTCTATAAAGCGTATTGGGTTG

CTCAATGCCGTTATACTCGCAACGGTGAACAATTCAAGGGGGTGATGACCGTAGC

AGGTACAAGTCAATCACAAGCTATTAAGCAGATGCGCCAGTACTTTACGGCTCACC

CAGGTGAATATACCTTTGCGGACTATGACACATTAATCCCTTTAATCACCCATATTG

AACAAAGTTCAACCTTAGAATTACCGTTAATACGGCAAGTACGTGAGCAACATAAT

GCAAAGGTTTCAGCCGTATTAGTGGATAAATGCAACCTCACACACCCAAGACCGTC

AGAAAAAGGCGACATTCATTACCGTGAGGGGCAACCTACGTTTATTGAATATTCGC

ATCTCTATGTCGTCATTGACAGTGGGGAATACCACCGCCAAACCGGGCAACATCTT

GTACCGAAACTGCATGGCTCACAACTGCCATGGAAATCACTCTATCAAGGAGAAAC

CCAAGACAGCCTTGAAGATAAAGCCCCTTATTTGGTACACATTGCCGCCAATCAAG

CCGGTCAGCGGTTTCTGGCTCATTACTTGAATTTACCACATAAAGCGAGTCTCGGA

TTATTTATCAATAGCCTCAAACCCTTTACCGATATTCACCGGCAAATGCGAAAACTC

ACCTATTTATATAATCAAAAACTGGAGAGTTGGAATTTCTTTCGTTTTTATGATGTTA

AGCACTTTATCCCATTTATTGAGTCTTTGACTCACGGACAGTTAATTAATGTGGCCA

ATGGGGTAAATGCGTTCTACGGCTATAGTGCACAATACCCCGATGGGGTTGAAATC

ACCTTTCACCCAGATTATCTGTATGACGGCAGTAAGCGAGAGCCGTTATTTATTAAT

ACCTATTTATACAATCACTACGCGAATATCACACAGATGCAAACTGTGGCTAAAGCT

AAGGCACTGATTGAACAATTTTCTCAGGTAGAAGGGGATGAGTTAGAGGGTGACG

CATTAATGGGCTACTGTATACACGCAGCAAATTGCAGTTTTTTAGACGATATTCATC

AATCAAAAGCGTTATTGTACGATTTGCAAGCTCGCTATTTGTGCCGTCATCAACCG

AGAACATGGCAGATCGCCAATGAAAAGCTGCACCTTATAAATACAACCAAGTTTT

ATTGAGTTACCACCGTTATATCGCCTGCTTAAATACCCAAGGAGAAATGAAATGAG

AAGGAGATGAGCTCATGAAAAGATTTGGCTGGCGCTGGCTGGTTTAGTTTTAGCG

TTTAGCGCATCGGCGGCGCAGTATGAAGGATCTATGGTTTCTAAAGGTGAAGCAG

TTATCAAGGAGTTCATGCGCTTTAAGGTTCACATGGAAGGGAGCGTTAACGGACAC

GAGTTCGAGATCGAAGGGGAGGGGGAAGGCCGCCCTTATGAGGGTACTCAAACT

GCCAAGCTGAAGGTCACAAAGGGGGGTCCCTTGCCGTTTTCTTGGGACATTCTGT

CACCTCAGTTCATGTACGGGTCCCGCGCGTTTATTAAGCATCCGGCTGACATCCCA

GATTACTATAAGCAGTCATTCCCGGAAGGGTTCAAATGGGAACGCGTGATGAACTT

CGAAGATGGGGGAGCCGTCACTGTTACCCAGGACACTAGCCTGGAGGATGGCAC

GTTGATTTATAAGGTCAAGCTGCGTGGAACAAATTTTCCCCCAGATGGGCCAGTTA

TGCAGAAGAAAACAATGGGGTGGGAAGCGAGTACGGAGCGCTTATACCCCGAAGA

TGGCGTCCTGAAAGGGGACATCAAAATGGCCCTGCGCCTGAAGGACGGGGGACG

TTATTTAGCGGACTTCAAGACAACCTATAAAGCTAAGAAACCCGTGCAAATGCCCG

GTGCTTACAACGTGGACCGCAAGCTGGATATCACAAGCCATAATGAAGATTACACC

GTGGTAGAGCAGTACGAGCGCAGTGAGGGCCGCCACTCAACCGGAGGTATGGAT

GAACTGTATAAGTCTGGCGGTGGTTCTCCGGCTGGTCTGGGCGGTGGATCCGGTA

CCACTG -3') and subcloned into a TOPO vector by incubating the gel-extracted gBlock

with Taq polymerase and dATP to add the required overhangs before ligating into the

vector following the TOPO TA cloning kit directions (Thermo Fisher Scientific, Waltham,

MA). The insert was then amplified using oAC327 (5'-

AAGAAGGAGATATACATATGCTCTTGAGTCCAAATCCCCTCTATA -3') and oAC332

(5'- TTCTCCAGTGGTACCGGATCCACCGCCCAG -3') and ligated into pAC84 using

restriction enzymes NdeI and BamHI.

Table A4: Strain used in Appendix B

Strain	Description	Source	KAG#	AC#
P. mirabilis				
ACH01 empty vector	BB2000 Δ(*idsE, sdaC*) carrying an empty vector	Achala Chittor (Chapter 2)	KAG4187	AC236
ACH01 ptet-SdaC	BB2000 Δ(*idsE, sdaC*) carrying a vector that expresses SdaC with a 3xFLAG N-terminal tag induced from an anhydrotetracycline promoter	Achala Chittor	KAG4474	AC558
ACH01 ptet-SdaC-EL4*	BB2000 Δ(*idsE, sdaC*) carrying ptet-SdaC vector where EL4 loop is swapped with YhaO	Achala Chittor	KAG4687	AC634
ACH01 ptet-YhaO	BB2000 Δ(*idsE, sdaC*) carrying ptet-YhaO vector	Achala Chittor (Chapter 2)	KAG4474	AC558
ACH01 ptet-YhaO-EL4*	BB2000 Δ(*idsE, sdaC*) carrying ptet-YhaO vector where EL4 loop is swapped with SdaC	Achala Chittor	KAG4685	AC632
ACH05 empty vector	BB2000 Δ(*sdaA, sdaB-C*) carrying an empty vector	Achala Chittor (Chapter 2)	KAG4388	AC466
ACH05 pSdaC	BB2000 Δ(*sdaA, sdaB-C*) carrying pSdaC vector where SdaC is expressed from its native promoter	Achala Chittor (Chapter 2)	KAG4391	AC469
ACH05 pSdaC-EL4*	BB2000 Δ(*sdaA, sdaB-C*) carrying pSdaC vector where SdaC is expressed from its native promoter and the EL4 loop sequence is swapped with YhaO	Achala Chittor		AC679

Table A4: Strain used in Appendix B (continued)

Strain	Description	Source	KAG#	AC#
E. coli				
DH5alpha pHCL149	DH5alpha strain carrying pHCL149 which co-expresses PopZ and a tagged, membrane-anchored GFP that will localize to the poles due to interaction with PopZ under an arabinose promoter	(Lim & Bernhardt, 2019) Addgene #134457	KAG4532	AC603
DH5alpha pHCL151	DH5alpha strain carrying pHCL151 which expresses DsbA$_{ss}$-mScarlet under a lacZ promoter	(Lim & Bernhardt, 2019) Addgene #134455	KAG4534	AC605
BW25141pir pHCL149 pHCL151	BW25141pir strain carrying both pHCL149 and pHCL151	Achala Chittor		AC663
BW25141pir pAC79 pHCL151	BW25141pir strain carrying pAC79, a pHCL149 derivative with SdaC inserted as the bait protein (GFP) with the pHCL151 control vector	Achala Chittor		AC664
BW25141pir pHCL149 pAC84	BW25141pir strain carrying pAC84, a pHCL151 derivative with IdsD inserted as the prey protein (mScarlet) with the pHCL149 control vector	Achala Chittor		AC654
BW25141pir pHCL79 pAC84	BW25141pir strain carrying pAC79 and pAC84	Achala Chittor		AC655
BW25141pir pAC84	BW25141 carrying pAC84	Achala Chittor		AC702

Table A4: Strain used in Appendix B (continued)

Strain	Description	Source	KAG#	AC#
BW25141pir pAC90	BW25141pir strain carrying pAC90, a pHCL151 derivative with IdsD inserted as the prey protein (mScarlet) and co-expressed with IdsC	Achala Chittor		AC700
BW25141pir pHCL151	BW25141pir strain carrying pHCL151	Achala Chittor		AC701
Omnimax	Cloning strain	Thermo Fisher Scientific (Waltham, MA)	KAG2183	
MFDpir	Mating strain	(Ferrières et al., 2010)	KAG3731	
BW25141pir	Pir+ cloning and expression strain	(Datsenko & Wanner, 2000) CGSC #7635	KAG4719	AC652

References

Belas, R., Erskine, D., & Flaherty, D. (1991). Transposon mutagenesis in *Proteus mirabilis*. *J Bacteriol*, *173*(19), 6289–6293.

Bozzi, A. T., Zimanyi, C. M., Nicoludis, J. M., Lee, B. K., Zhang, C. H., & Gaudet, R. (2019). Structures in multiple conformations reveal distinct transition metal and proton pathways in an Nramp transporter. *Elife*, *8*, e41124.

Cardarelli, L., Saak, C., & Gibbs, K. A. (2015). Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding. *MBio*, *6*(3), e00251-15.

Chatterjee, S., & Rothenberg, E. (2012). Interaction of Bacteriophage l with Its *E. coli* Receptor, LamB. *Viruses*, *4*(11), 3162–3178. https://doi.org/10.3390/v4113162

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*, *97*(12), 6640–6645. https://doi.org/10.1073/pnas.120163297

Ferrières, L., Hémery, G., Nham, T., Guérout, A.-M., Mazel, D., Beloin, C., & Ghigo, J.-M. (2010). Silent mischief: Bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *Journal of Bacteriology*, *192*(24), 6418–6427.

Gérard, F., Pradel, N., & Wu, L.-F. (2005). Bactericidal activity of colicin V is mediated by an inner membrane protein, SdaC, of *Escherichia coli*. *Journal of Bacteriology*, *187*(6), 1945–1950.

Gibbs, K. A., Urbanowski, M. L., & Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. *Science*, *321*(5886), 256–259. https://doi.org/10.1126/science.1160033

Krishnamurthy, H., & Gouaux, E. (2012). X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature*, *481*(7382), 469.

Likhacheva, N. A., Samsonov, V. V., Samsonov, V. V., & Sineoky, S. P. (1996). Genetic control of the resistance to phage C1 of *Escherichia coli* K-12. *J Bacteriol*, *178*(17), 5309–5315.

Lim, H. C., & Bernhardt, T. G. (2019). A PopZ-linked apical recruitment assay for studying protein–protein interactions in the bacterial cell envelope. *Molecular Microbiology*, *112*(6), 1757–1768. https://doi.org/10.1111/mmi.14391

Oberai, A., Joh, N. H., Pettit, F. K., & Bowie, J. U. (2009). Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proceedings of the National Academy of Sciences*, *106*(42), 17747–17750.

Ruhe, Z. C., Wallace, A. B., Low, D. A., & Hayes, C. S. (2013). Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *MBio*, *4*(4), e00480-13.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., & Schmid, B. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.

Sharma, O., Yamashita, E., Zhalnina, M. V., Zakharov, S. D., Datsenko, K. A., Wanner, B. L., & Cramer, W. A. (2007). Structure of the Complex of the Colicin E2 R-domain and Its BtuB Receptor THE OUTER MEMBRANE COLICIN TRANSLOCON. *Journal of Biological Chemistry*, *282*(32), 23163–23170.

Willett, J. L., Gucinski, G. C., Fatherree, J. P., Low, D. A., & Hayes, C. S. (2015). Contact-dependent growth inhibition toxins exploit multiple independent cell-entry pathways. *Proceedings of the National Academy of Sciences*, *112*(36), 11341–11346.

Zepeda-Rivera, M. A., Saak, C. C., & Gibbs, K. A. (2018). A proposed chaperone of the bacterial type VI secretion system functions to constrain a self-identity protein. *J Bacteriol*. https://doi.org/10.1128/JB.00688-17
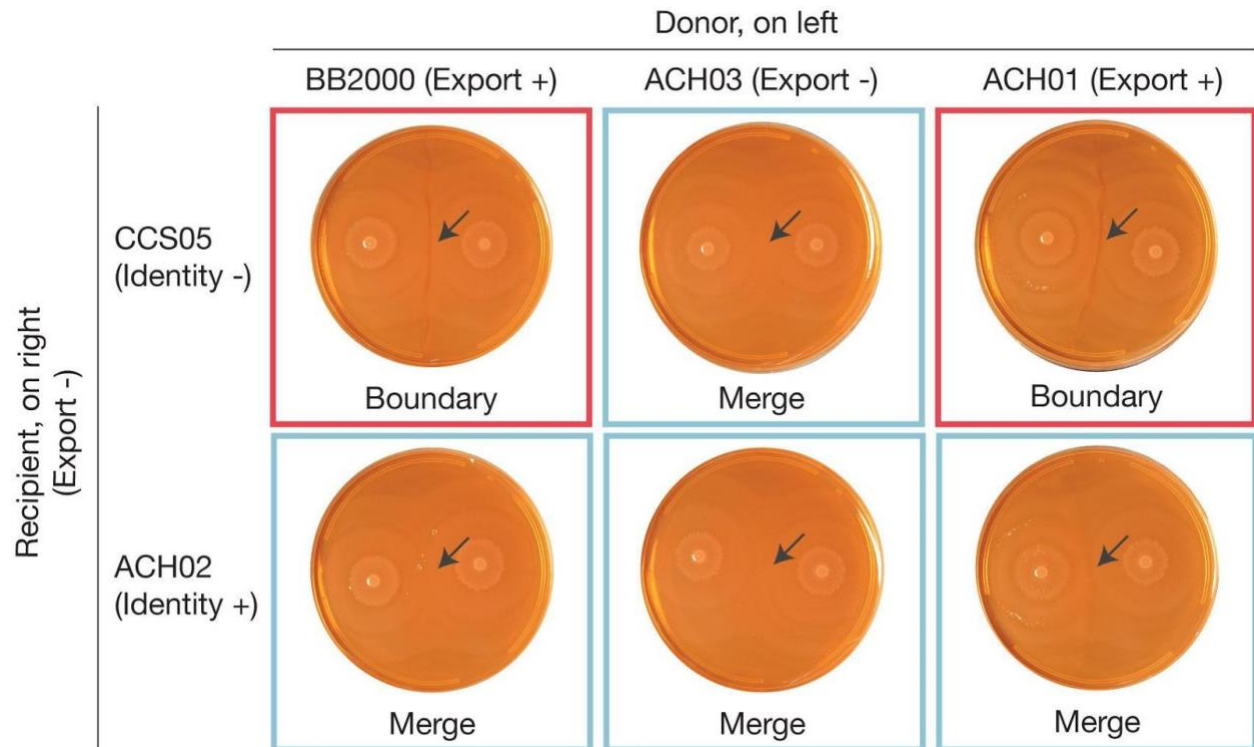
Appendix C

Supplementary Data

Figure A8: The *sdaC* gene is not required for IdsD secretion between cells during swarming.
Boundary assay results shown for donor strains BB2000 (Export +), ACH03 (Export -), and ACH01 on the left side of the plate against recipient strains CCS05 (Identity -, Export -) and ACH02 (Identity +, Export -) on the right side of the plate. Intersection of strains is marked with a black arrow. Boundaries are outlined in a red box and merges are outlined in a blue box.
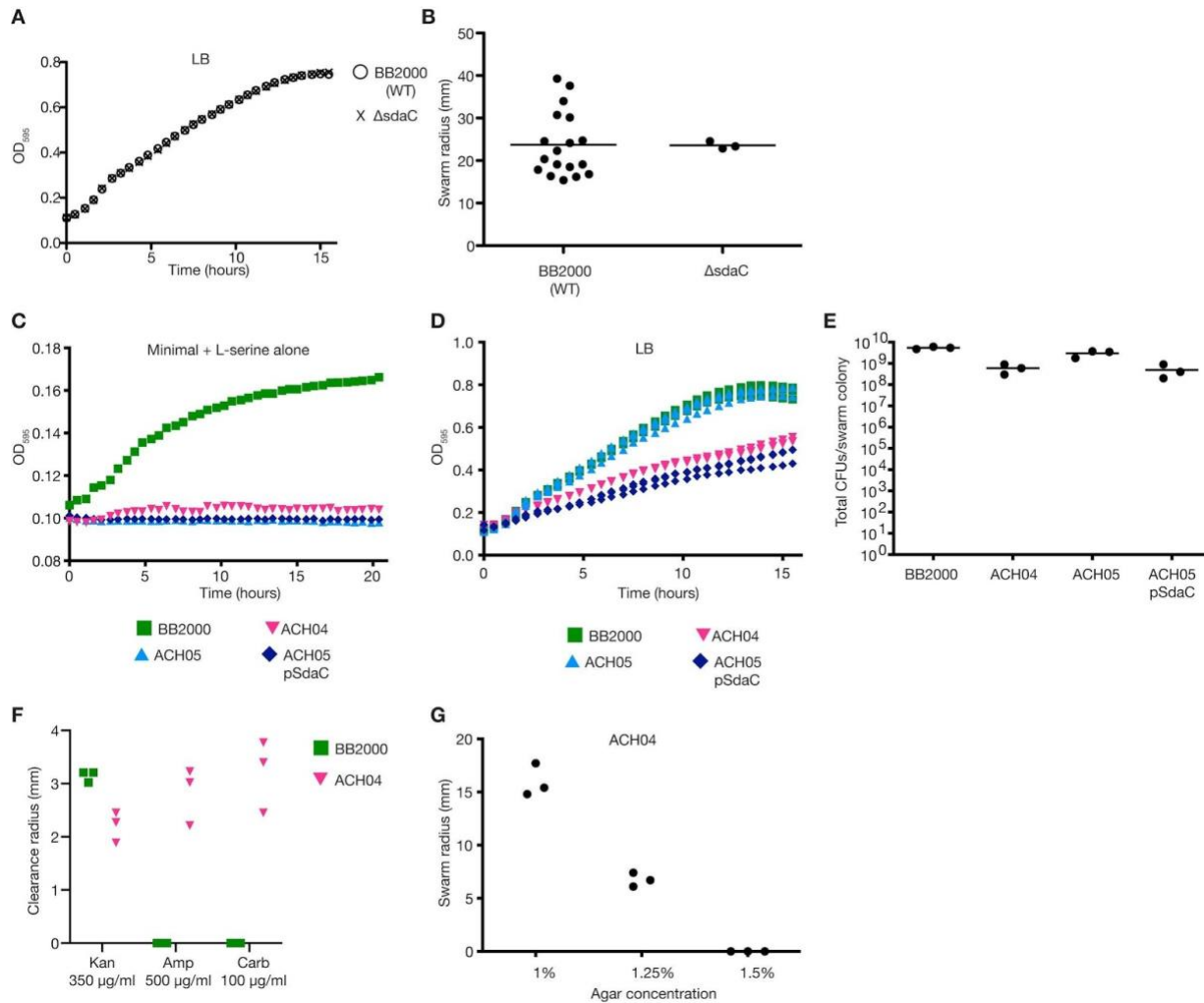
Figure A9: SdaC activity in a high serine strain background reduces swarm expansion, slows growth, and destabilizes the cell wall.
A) Growth curve of BB2000 and BB2000 Δ*sdaC* in LB broth. **B)** Swarm radius measured from swarm assay of BB2000 Δ*sdaC* empty vector with BB2000 empty vector data copied from Figure 1C for comparison. **C)** Growth of BB2000 empty vector, ACH04 [BB2000 Δ(*sdaA*, *sdaB*)] empty vector, ACH05 [BB2000 Δ(*sdaA*, *sdaB-sdaC*)] empty vector, and ACH05 pSdaC in minimal medium without glucose and 10mM serine as the only carbon source. Averages shown for three biological replicates. **D)** Growth of BB2000 empty vector, ACH04 empty vector, ACH05 empty vector, and ACH05 pSdaC in LB for three biological replicates. **E)** CFUs per swarm colony for BB2000 empty vector, ACH04 empty vector, ACH05 empty vector, and ACH05 pSdaC for three biological replicates. **F)** Antibiotic clearance radius by discs soaked in Kan (350μg/ml), Amp (500μg/ml), or Carb (100μg/ml) for swarm colonies of strains BB2000 and ACH04. **G)** Swarm radius measured from swarm assay of ACH04 on LB medium containing 1%, 1.25%, or 1.5% agar.
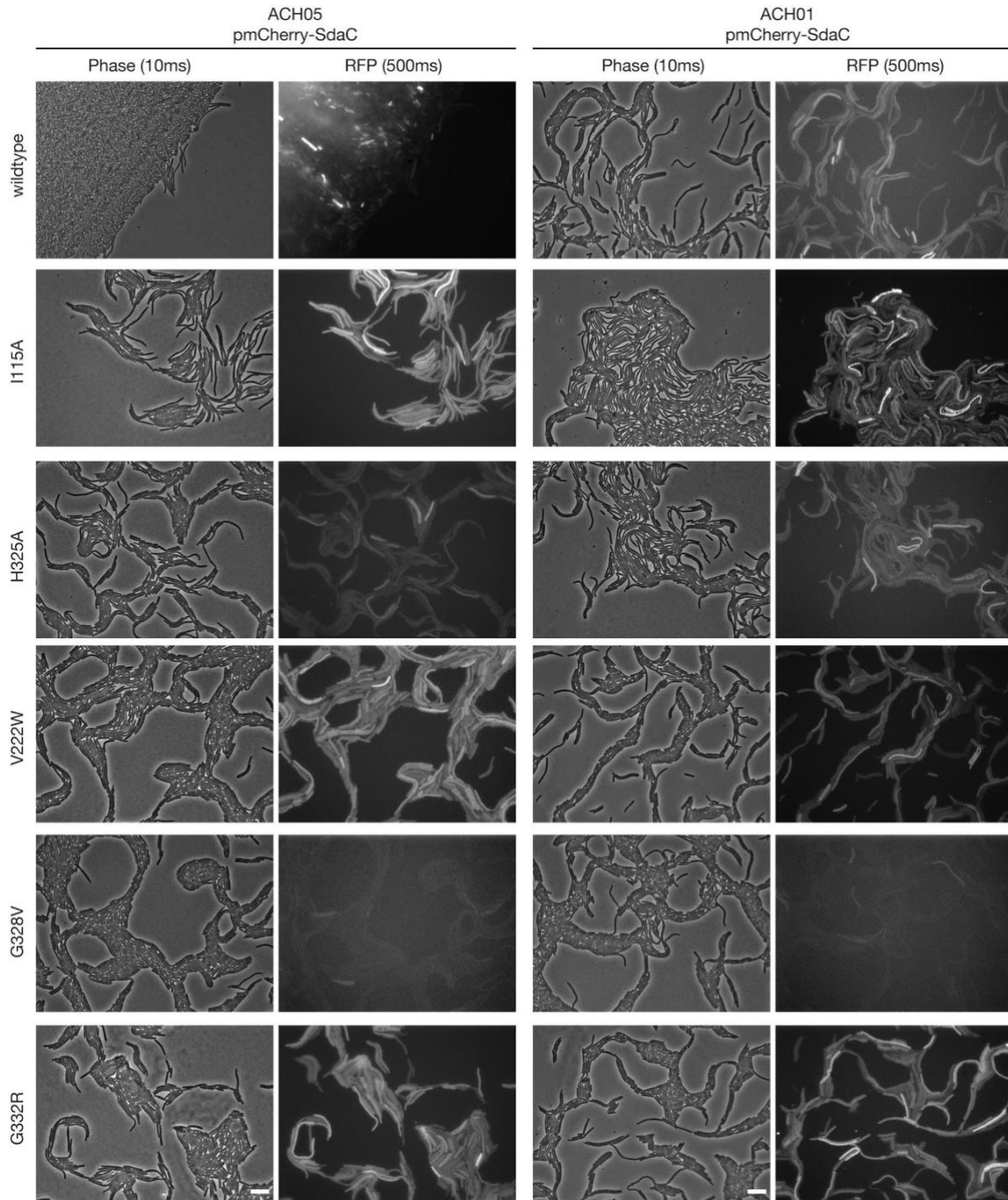
Figure A10: Fluorescent fusion protein for wildtype SdaC and SdaC variants.
Representative epifluorescence micrographs from phase and RFP channels of
pmCherry-SdaC (wildtype or individual residue variants I115A, H325A, V222W, G328V,
and G332R) expressed in ACH05 [BB2000 Δ(*sdaA, sdaB-sdaC*)] and ACH01 [BB2000
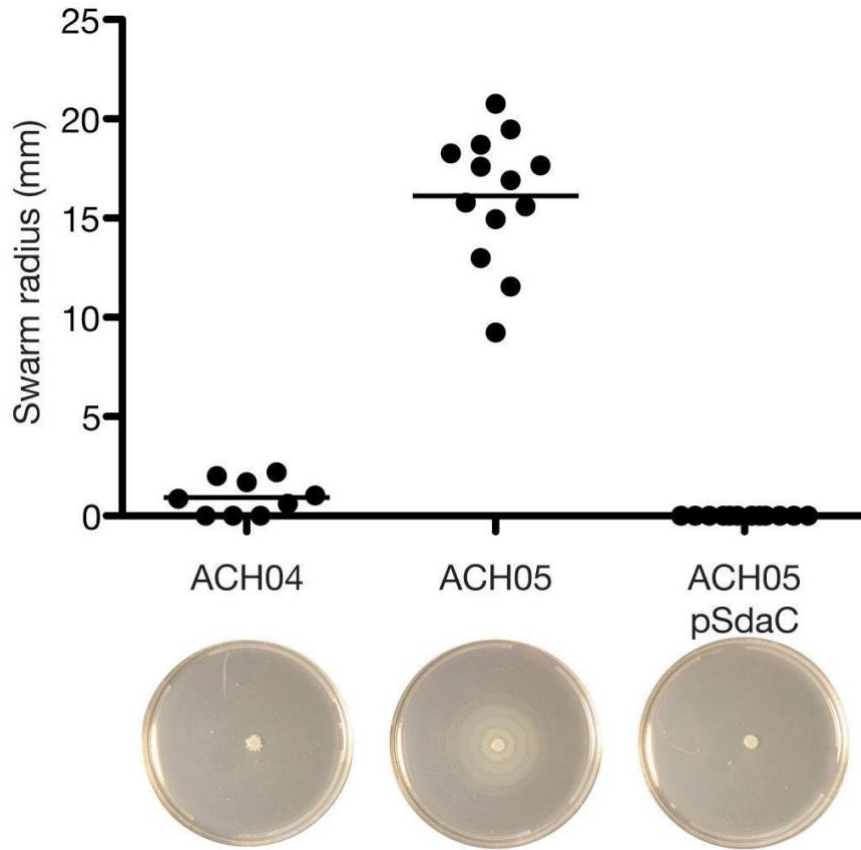Δ(*idsE, sdaC*)]. Scale bar on the lower right of the last row of phase contrast images is
10μm.

Figure A11: SdaC activity restricts swarm expansion when *sdaA* and *sdaB* are deleted.
Representative swarm plate images for Figure 1G swarm assay of ACH04 [BB2000 Δ(*sdaA*, *sdaB*)] empty vector, ACH05 [BB2000 Δ(*sdaA*, *sdaB-sdaC*)] empty vector, and ACH05 pSdaC.
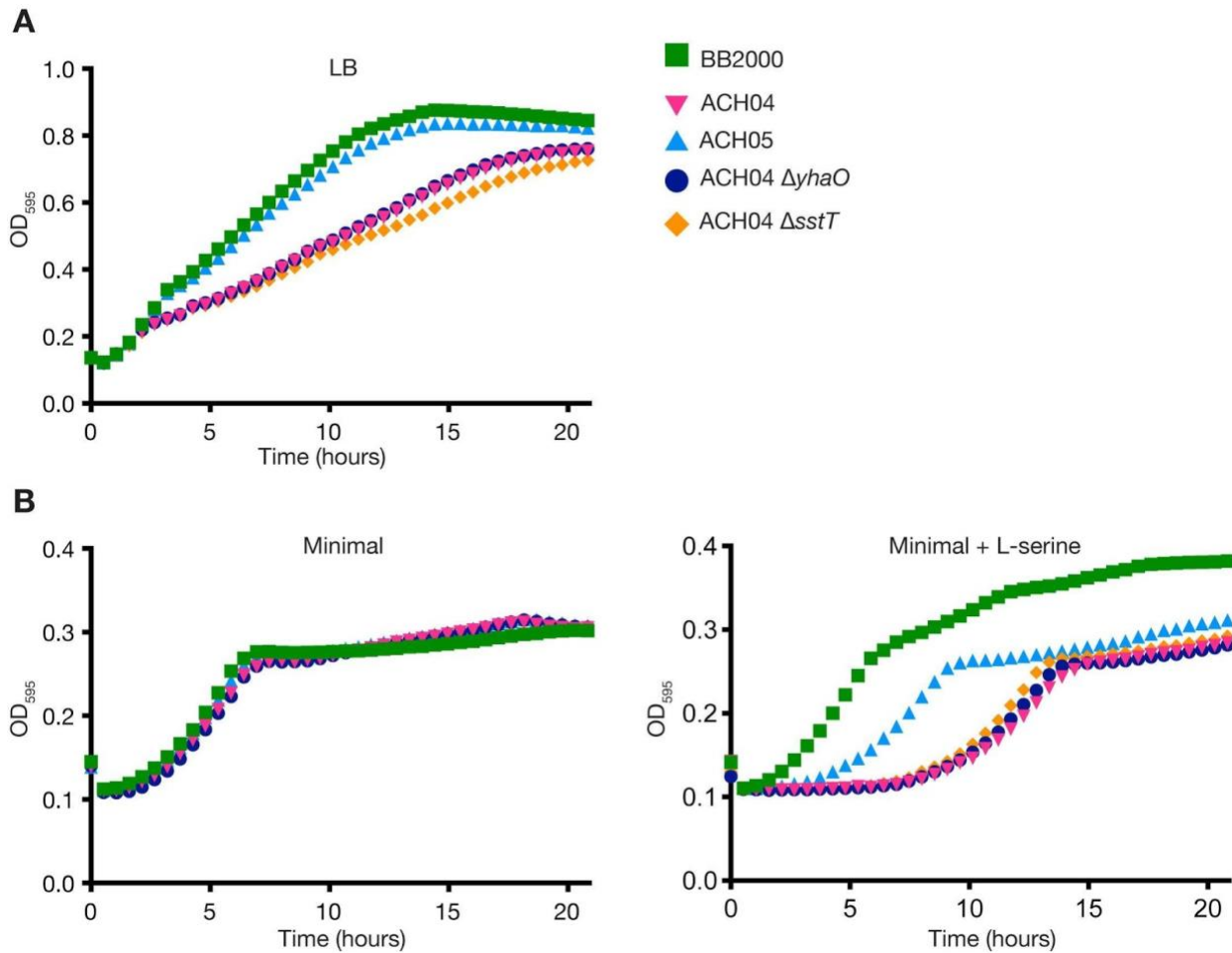
Figure A12: Deletion of neither *sstT* nor *yhaO* in the ACH04 background rescues growth in LB or minimal medium compared to deletion of *sdaC*.
A) Growth curve in LB of BB2000, ACH04 [BB2000 Δ(*sdaA*, *sdaB*)], ACH05 [BB2000 Δ(*sdaA*, *sdaB-sdaC*)], ACH04 Δ*sstT*, and ACH04 Δ*yhaO.* **B)** Growth curve in minimal medium (left) and minimal medium plus 10mM L-serine (right) of BB2000, ACH04, ACH05, ACH04 Δ*sstT*, and ACH04 Δ*yhaO.*

A

IdsD amino acids sequence within 1034 subfamily across 27 P. mirabilis isolates

Variable Region

B

SdaC amino acids sequence across 65 P. mirabilis isolates
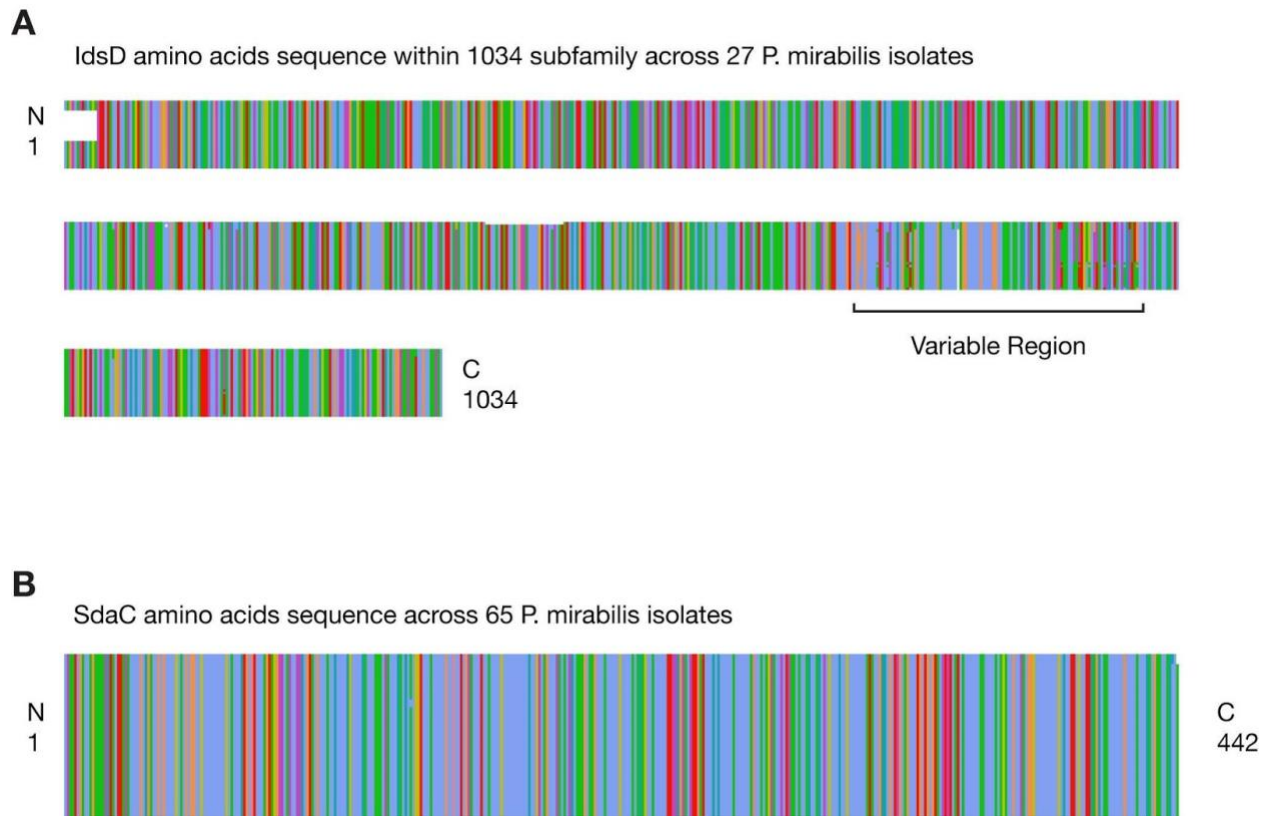

Figure A13: IdsD is polymorphic across *P. mirabilis* isolates while SdaC is conserved throughout the protein sequence.
A) Alignment of 1034 subfamily IdsD amino acid sequences from 27 *P. mirabilis* isolate genomes available through NCBI. Alignment performed using MUSCLE and visualized with AlignmentViewer. B) Alignment of SdaC amino acid sequences from 65 *P. mirabilis* isolate genomes available through NCBI. Alignment performed using MUSCLE and visualized with AlignmentViewer.
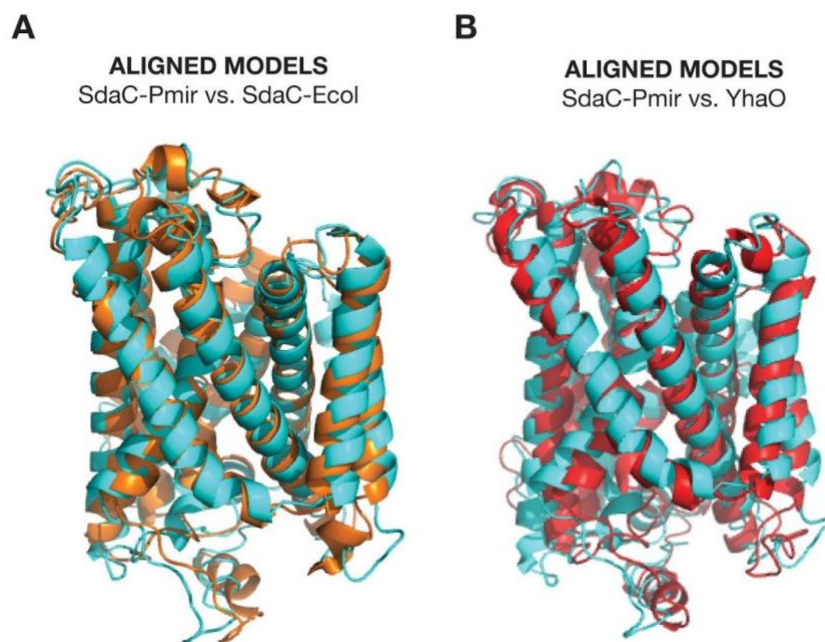
Figure A14: Alignment of SdaC and YhaO I-TASSER models.
A) PyMol alignment of structural model of SdaC-Pmir from I-TASSER (blue) aligned to structural model of SdaC-Ecol from I-TASSER (orange) (C-score = -0.64, TM-score = 0.63±0.13, RMSD = 8.4±4.5Å). B) PyMol alignment of structural model of SdaC from I-TASSER (blue) aligned to structural model of YhaO from I-TASSER (red) (C-score = -0.55, TM-score = 0.64±0.13, RMSD = 8.3±4.5Å).
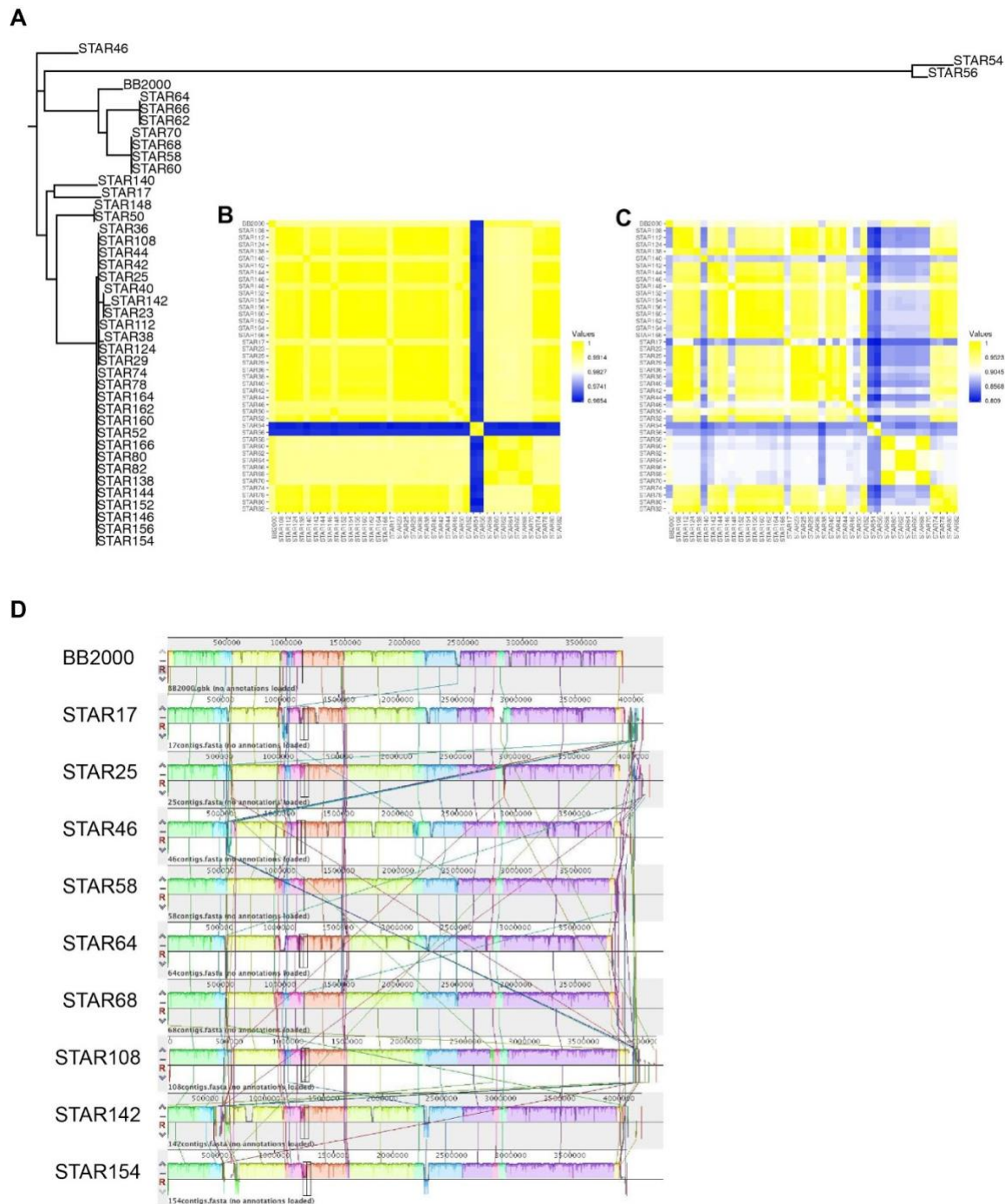
Figure A15: Genome sequence similarity across STAR isolates.
A) Phylogenetic tree of the STAR strains and BB2000. B) Heatmap (blue=low, yellow=high) of average nucleotide identity (ANI) in coding regions of the STAR strains and BB2000. C) Heatmap (blue=low, yellow=high) of ANI across the whole genome of the STAR strains and BB2000. D) Alignment of genomes (with contigs reordered against BB2000) of the representative STAR strains and BB2000 using Mauve (Darling et al., 2004).
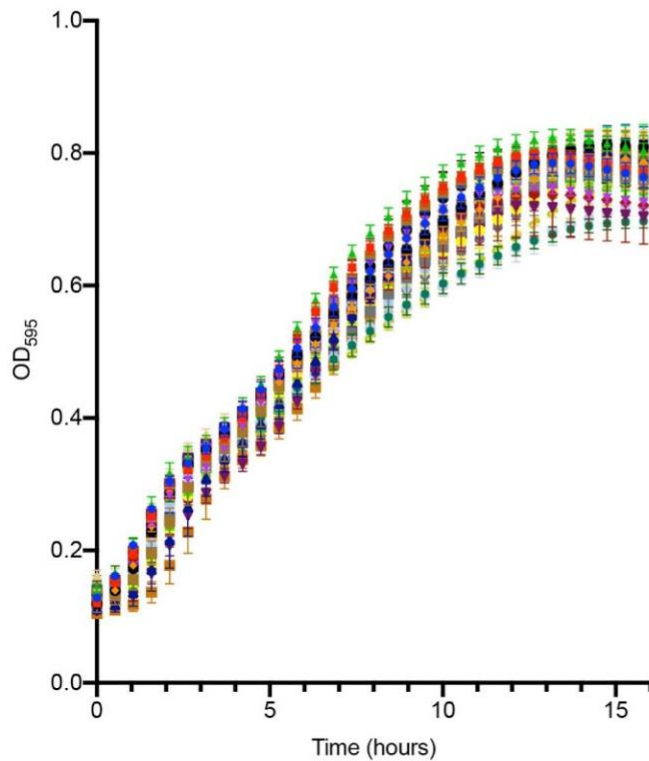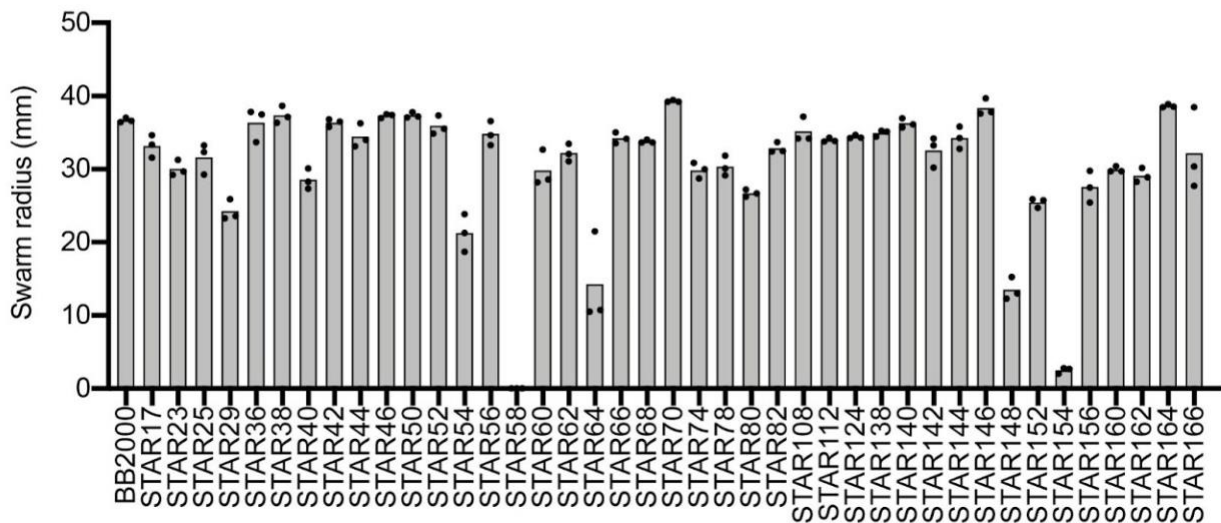
Figure A16: Growth and swarm behavior across all STAR isolates.
A) Growth curve of all STAR strains and BB2000. Average of three biological replicates is shown with standard deviation error bars. B) Swarm expansion radius for all STAR strains and BB2000. Three biological replicates are shown as well as the average as a grey filled bar.

Table A5: Metadata from Charles River Laboratories

Strain	Test Date	Customer Type	LTM Order Number	Sample ID	Culture Site	Host species	Host strain	Age	Sex	# Sub-Culture
STAR 17	6/18/2018	Biotech	2018030662	2	Environmental	Mouse				1
STAR 23	7/26/2018	Biotech	2018032984	2	GI Tract	Mouse	Sentinel			2
STAR 25	7/26/2018	Biotech	2018031979	4	GI Tract	Mouse				2
STAR 29	7/30/2018	Biotech	2018034860	6	GI Tract	Mouse				2
STAR 36	8/10/2018	Biotech	2018035281	5	Respiratory	Mouse	SW	16 Weeks	F	2
STAR 38	8/10/2018	Biotech	2018035281	9	Respiratory	Mouse	SW	16 Weeks	F	2
STAR 40	8/10/2018	Biotech	2018034860	6	Environmental	Mouse				2
STAR 42	8/10/2018	Biotech	2018034860	7	Environmental	Mouse				2
STAR 44	8/10/2018	Biotech	2018034860	9	Environmental	Mouse				2
STAR 46	8/10/2018	Biotech	2018035278	4	GI Tract	Mouse	SW	17 Weeks	F	2
STAR 50	8/10/2018	Biotech	2018035287	2	Respiratory	Mouse	SW		F	2
STAR 52	8/10/2018	Biotech	2018035287	6	Respiratory	Mouse	SW		F	2
STAR 54	8/10/2018	Production	2018003546	1	Respiratory	Rabbit	NZW	RB	F	1
STAR 56	8/16/2018	Production	2018000942	3	GI tract	Chicken				1
STAR 58	9/6/2018	Production	201807154	2	Respiratory	rat	SHRSP	RB	M	2
STAR 60	9/6/2018	Production	201807154	5	GI tract	rat	SHRSP	56-70 do	M	2
STAR 62	9/6/2018	Production	201807154	6	Respiratory	rat	SHRSP	56-70 do	M	2
STAR 64	9/6/2018	Production	201807154	8		rat	SHRSP	56-70 do	F	2
STAR 66	9/6/2018	Production	201807154	5		rat	SHRSP	56-70 do	M	2
STAR 68	9/6/2018	Production	2018025883	5	GI tract	Mouse	CF-i	56-70 do	M	2
STAR 70	9/6/2018	Production	2018025883	7	GI tract	Mouse	CF-i	56-70 do	F	2
STAR 74	9/6/2018	Production	2018007148	2	GI tract	Mouse	CD-1	RB	M	2
STAR 78	9/6/2018	Production	2018007148	7	GI tract	Mouse	CD-1	56-70 do	F	2
STAR 80	9/6/2018	Production	2018037713	3	Environmental	Mouse	HO, Nude		F	1

Table A5: Metadata from Charles River Laboratories (Continued)

Strain	Test Date	Customer Type	LTM Order Number	Sample ID	Culture Site	Host species	Host strain	Age	Sex	# Sub-Culture
STAR 82	9/6/2018	Production	2018037713	4	Environmental	Mouse	HO, Nude		F	1
STAR 108	9/6/2018	Biotech	2018038738	7	Environmental	Mouse				1
STAR 112	9/6/2018	Biotech	2018038683	2	Respiratory	Mouse				1
STAR 124	9/6/2018	Biotech	2018038683	1	GI Tract	Mouse				1
STAR 138	9/6/2018	Production	2018007241	2	GI Tract	Rat	ZDF FA/+	RB	M	2
STAR 140	9/6/2018	Production	2018007241	3	GI Tract	Rat	ZDF FA/+	RB	F	2
STAR 142	9/6/2018	Production	2018007241	5	GI Tract	Rat	ZDF FA/+	56-70 do	M	2
STAR 144	9/6/2018	Production	2018007241	6	GI Tract	Rat	ZDF FA/+	56-70 do	M	2
STAR 146	9/6/2018	Production	2018007223	5	Respiratory	Mouse	C57BL/6	56-70 do	M	2
STAR 148	9/6/2018	Production	2018007223	1	GI Tract	Mouse	C57BL/6	RB	M	2
STAR 152	9/6/2018	Production	2018007223	5	GI Tract	Mouse	C57BL/6	56-70 do	M	2
STAR 154	9/6/2018	Production	2018007223	6	GI Tract	Mouse	C57BL/6	56-70 do	M	2
STAR 156	9/6/2018	Production	2018007223	7	GI Tract	Mouse	C57BL/6	56-70 do	F	2
STAR 160	9/6/2018	Production	2018016101	2	Respiratory	Mouse	CD-I	RB	M	2
STAR 162	9/6/2018	Production	2018016101	7	Respiratory	Mouse	CD-I	56-70 do	M	2
STAR 164	9/6/2018	Production	2018016101	1	GI Tract	Mouse	CD-I	RB	M	2
STAR 166	9/6/2018	Production	2018016101	7	GI Tract	Mouse	CD-I	56-70 do	F	2

Table A6: Phenotypic and observational data for STAR isolates

Strain	Replicate	Swarm Radius (mm)	Growth rate 1 (OD/hr)	Growth rate 2 (OD/hr)	Max OD	Swarm on LSW	Round pock marks	Striated pock marks	Oval pock marks	Strength of smell (1=strong)	# Swarm rings	# Boundaries
STAR 17	1	33.33	0.045	0.024	0.787	1	1	1	0	1	3	4
STAR 17	2	34.63	0.048	0.024	0.798	1	1	1	0	1	3	4
STAR 17	3	31.56	0.046	0.023	0.769	1	1	1	0	1	3	4
STAR 23	1	29.23	0.051	0.028	0.791	0	0	0	1	2	1	2
STAR 23	2	29.67	0.049	0.028	0.802	0	0	0	1	2	1	2
STAR 23	3	31.27	0.052	0.030	0.805	0	0	0	1	2	1	2
STAR 25	1	33.22	0.051	0.031	0.828	0	1	0	1	3	1	2
STAR 25	2	32.34	0.046	0.028	0.810	0	1	0	1	3	1	2
STAR 25	3	29.27	0.053	0.026	0.834	0	1	0	1	3	1	2
STAR 29	1	25.89	0.042	0.031	0.753	0	0	1	0	2	2	2
STAR 29	2	23.27	0.039	0.032	0.742	0	0	1	0	2	2	2
STAR 29	3	23.58	0.040	0.029	0.745	0	0	1	0	2	2	2
STAR 36	1	37.49	0.048	0.021	0.797	0	1	0	0	1	1	1
STAR 36	2	33.69	0.047	0.021	0.800	0	1	0	0	1	1	1
STAR 36	3	37.84	0.048	0.020	0.780	0	1	0	0	1	1	1
STAR 38	1	36.35	0.044	0.021	0.820	0	1	0	0	2	2	1
STAR 38	2	37.15	0.044	0.020	0.811	0	1	0	0	2	2	1
STAR 38	3	38.64	0.045	0.020	0.807	0	1	0	0	2	2	1
STAR 40	1	30.09	0.042	0.030	0.785	0	0	1	0	3	1	1
STAR 40	2	28.26	0.041	0.028	0.801	0	0	1	0	3	1	1
STAR 40	3	27.34	0.041	0.028	0.770	0	0	1	0	3	1	1
STAR 42	1	35.79	0.038	0.026	0.826	0	1	0	0	3	3	1
STAR 42	2	36.80	0.047	0.025	0.807	0	1	0	0	3	3	1
STAR 42	3	36.50	0.038	0.026	0.787	0	1	0	0	3	3	1

Table A6: Phenotypic and observational data for STAR isolates (Continued)

Strain	Rep-licate	Swarm Radius (mm)	Growth rate 1 (OD/hr)	Growth rate 2 (OD/hr)	Max OD	Swarm on LSW	Round pock marks	Striated pock marks	Oval pock marks	Strength of smell (1=strong)	# Swarm rings	# Boundaries
STAR 44	1	34.01	0.033	0.029	0.721	0	0	1	0	1	3	2
STAR 44	2	36.28	0.039	0.028	0.715	0	0	1	0	1	3	2
STAR 44	3	33.11	0.041	0.029	0.723	0	0	1	0	1	3	2
STAR 46	1	37.53	0.046	0.025	0.775	1	1	0	0	2	3	2
STAR 46	2	37.01	0.044	0.023	0.774	1	1	0	0	2	3	2
STAR 46	3	37.41	0.042	0.020	0.665	1	1	0	0	2	3	2
STAR 50	1	37.08	0.039	0.016	0.697	0	1	0	0	1	4	4
STAR 50	2	37.79	0.044	0.016	0.708	0	1	0	0	1	4	4
STAR 50	3	37.20	0.041	0.016	0.687	0	1	0	0	1	4	4
STAR 52	1	34.88	0.045	0.023	0.763	0	1	0	1	3	4	2
STAR 52	2	37.35	0.046	0.022	0.777	0	1	0	1	3	4	2
STAR 52	3	35.53	0.045	0.023	0.794	0	1	0	1	3	4	2
STAR 54	1	23.86	0.043	0.028	0.769	0	0	0	0	2	3	4
STAR 54	2	21.28	0.042	0.027	0.766	0	0	0	0	2	3	4
STAR 54	3	18.70	0.045	0.026	0.752	0	0	0	0	2	3	4
STAR 56	1	36.57	0.048	0.022	0.815	0	1	0	0	3	2	4
STAR 56	2	34.65	0.047	0.022	0.803	0	1	0	0	3	2	4
STAR 56	3	33.26	0.050	0.025	0.809	0	1	0	0	3	2	4
STAR 58	1	0.00	0.045	0.019	0.735	0	0	0	0	2	0	4
STAR 58	2	0.00	0.049	0.016	0.718	0	0	0	0	2	0	4
STAR 58	3	0.00	0.045	0.017	0.705	0	0	0	0	2	0	4
STAR 60	1	32.68	0.040	0.022	0.784	0	0	1	0	1	2	4
STAR 60	2	28.59	0.044	0.025	0.777	0	0	1	0	1	2	4
STAR 60	3	28.21	0.041	0.021	0.742	0	0	1	0	1	2	4

Table A6: Phenotypic and observational data for STAR isolates (Continued)

Strain	Repl-icate	Swarm Radius (mm)	Growth rate 1 (OD/hr)	Growth rate 2 (OD/hr)	Max OD	Swarm on LSW	Round pock marks	Striated pock marks	Oval pock marks	Strength of smell (1= strong)	# Swarm rings	# Boundaries
STAR 62	1	33.46	0.023	0.023	0.763	0	1	0	0	2	3	2
STAR 62	2	31.09	0.047	0.022	0.821	0	1	0	0	2	3	2
STAR 62	3	32.08	0.042	0.018	0.752	0	1	0	0	2	3	2
STAR 64	1	10.52	0.040	0.025	0.777	0	0	0	0	2	3	2
STAR 64	2	21.49	0.047	0.027	0.839	0	0	0	0	2	3	2
STAR 64	3	10.74	0.043	0.026	0.791	0	0	0	0	2	3	2
STAR 66	1	34.13	0.044	0.025	0.805	0	1	1	0	1	3	2
STAR 66	2	35.00	0.042	0.023	0.814	0	1	1	0	1	3	2
STAR 66	3	33.54	0.045	0.024	0.796	0	1	1	0	1	3	2
STAR 68	1	34.03	0.040	0.021	0.790	1	1	0	0	1	2	4
STAR 68	2	33.66	0.043	0.019	0.763	1	1	0	0	1	2	4
STAR 68	3	33.64	0.044	0.018	0.731	1	1	0	0	1	2	4
STAR 70	1	39.22	0.040	0.024	0.748	1	0	0	1	1	3	4
STAR 70	2	39.44	0.038	0.022	0.759	1	0	0	1	1	3	4
STAR 70	3	39.24	0.040	0.022	0.765	1	1	0	1	1	3	4
STAR 74	1	29.97	0.041	0.022	0.787	0	0	1	0	2	1	2
STAR 74	2	28.73	0.042	0.024	0.794	0	0	1	0	2	1	2
STAR 74	3	30.84	0.044	0.021	0.801	0	0	1	0	2	1	2
STAR 78	1	31.84	0.046	0.025	0.809	0	1	1	0	2	3	2
STAR 78	2	30.10	0.045	0.022	0.814	0	1	1	0	2	3	2
STAR 78	3	29.16	0.045	0.022	0.803	0	1	1	0	2	3	2
STAR 80	1	26.67	0.046	0.023	0.832	0	1	0	1	3	3	2
STAR 80	2	27.20	0.049	0.023	0.808	0	1	0	1	3	3	2
STAR 80	3	26.23	0.050	0.022	0.773	0	1	0	1	3	3	2

Table A6: Phenotypic and observational data for STAR isolates (Continued)

Strain	Replicate	Swarm Radius (mm)	Growth rate 1 (OD/hr)	Growth rate 2 (OD/hr)	Max OD	Swarm on LSW	Round pock marks	Striated pock marks	Oval pock marks	Strength of smell (1=strong)	# Swarm rings	# Boundaries
STAR82	1	32.47	0.048	0.021	0.787	0	1	0	1	3	2	1
STAR82	2	33.70	0.048	0.022	0.787	0	1	0	1	3	2	1
STAR82	3	32.42	0.049	0.022	0.784	0	1	0	1	3	2	1
STAR108	1	34.18	0.046	0.023	0.817	0	0	1	0	1	3	1
STAR108	2	34.19	0.046	0.022	0.800	0	0	1	0	1	3	1
STAR108	3	37.16	0.046	0.022	0.779	0	0	1	0	1	3	1
STAR112	1	34.30	0.048	0.030	0.804	0	1	0	1	1	2	2
STAR112	2	33.76	0.044	0.026	0.805	0	1	0	1	1	2	2
STAR112	3	33.86	0.048	0.025	0.763	0	1	0	1	1	2	2
STAR124	1	34.67	0.048	0.029	0.815	0	0	0	1	2	2	2
STAR124	2	34.26	0.046	0.027	0.784	0	0	0	1	2	2	2
STAR124	3	34.29	0.044	0.027	0.804	0	0	0	1	2	2	2
STAR138	1	35.26	0.048	0.023	0.774	0	1	1	1	1	2	4
STAR138	2	35.13	0.049	0.021	0.815	0	1	1	1	1	2	4
STAR138	3	34.44	0.050	0.023	0.804	0	1	1	1	1	2	4
STAR140	1	36.11	0.050	0.022	0.812	0	1	1	0	1	3	2
STAR140	2	36.96	0.049	0.019	0.784	0	1	1	0	1	3	2
STAR140	3	35.71	0.051	0.018	0.767	0	1	1	0	1	3	2
STAR142	1	33.24	0.043	0.020	0.784	0	0	1	1	2	3	1
STAR142	2	34.19	0.043	0.020	0.782	0	0	1	1	2	3	1
STAR142	3	30.20	0.044	0.019	0.769	0	0	1	1	2	3	1
STAR144	1	32.77	0.047	0.020	0.802	0	0	1	0	1	3	1
STAR144	2	34.25	0.049	0.020	0.807	0	0	1	0	1	3	1
STAR144	3	35.80	0.047	0.019	0.796	0	0	1	0	1	3	1

Table A6: Phenotypic and observational data for STAR isolates (Continued)

Strain	Re pl- icat e	Swar m Radi us (mm)	Grow th rate 1 (OD/ hr)	Grow th rate 2 (OD/ hr)	Max OD	Swar m on LSW	Rou nd pock mark s	Striat ed pock mark s	Oval poc k mar ks	Stren gth of smell (1= strong)	# Swar m rings	# Boundar ies
STAR1 46	1	37.8 1	0.04 4	0.02 2	0.8 24	0	1	1	0	3	3	1
STAR1 46	2	39.6 7	0.04 5	0.02 1	0.8 37	0	1	1	0	3	3	1
STAR1 46	3	37.5 9	0.04 3	0.02 3	0.8 37	0	1	1	0	3	3	1
STAR1 48	1	13.0 2	0.04 3	0.02 3	0.8 45	0	0	1	0	3	2	1
STAR1 48	2	12.3 1	0.04 3	0.02 1	0.8 14	0	0	1	0	3	2	1
STAR1 48	3	15.2 3	0.04 1	0.02 1	0.8 22	0	0	1	0	3	2	1
STAR1 52	1	25.7 5	0.04 3	0.02 1	0.7 71	0	0	1	1	3	3	1
STAR1 52	2	25.8 9	0.04 5	0.02 1	0.7 78	0	0	1	1	3	3	1
STAR1 52	3	24.7 0	0.04 9	0.02 0	0.7 73	0	0	1	1	3	3	1
STAR1 54	1	2.03	0.04 1	0.02 3	0.7 89	0	0	0	0	2	3	1
STAR1 54	2	2.80	0.04 1	0.02 2	0.7 97	0	0	0	0	2	3	1
STAR1 54	3	2.66	0.03 8	0.02 3	0.7 81	0	0	0	0	2	3	1
STAR1 56	1	27.4 9	0.04 3	0.02 0	0.7 98	0	1	1	0	2	4	1
STAR1 56	2	25.4 3	0.04 3	0.01 9	0.7 74	0	1	1	0	2	4	1
STAR1 56	3	29.7 5	0.04 9	0.02 0	0.7 84	0	1	1	0	2	4	1
STAR1 60	1	29.7 0	0.04 4	0.02 4	0.8 26	0	1	1	0	2	3	1
STAR1 60	2	30.4 0	0.04 4	0.02 3	0.8 25	0	1	1	0	2	3	1
STAR1 60	3	29.7 3	0.04 5	0.02 2	0.7 99	0	1	1	0	2	3	1
STAR1 62	1	28.9 0	0.04 2	0.02 3	0.8 00	0	1	0	0	2	3	1
STAR1 62	2	28.2 9	0.04 1	0.02 1	0.7 84	0	1	1	0	2	4	1
STAR1 62	3	30.1 5	0.04 3	0.02 2	0.8 06	0	1	1	0	2	4	1
STAR1 64	1	38.4 6	0.04 2	0.02 4	0.8 18	0	1	1	0	1	3	1
STAR1 64	2	38.8 9	0.04 2	0.02 5	0.8 16	0	1	1	0	1	3	1
STAR1 64	3	38.5 4	0.04 2	0.02 5	0.8 11	0	1	1	0	1	3	1

Table A6: Phenotypic and observational data for STAR isolates (Continued)

Strain	Re pl- icat e	Swar m Radi us (mm)	Grow th rate 1 (OD/ hr)	Grow th rate 2 (OD/ hr)	Max OD	Swar m on LSW	Rou nd pock mark s	Striat ed pock mark s	Oval poc k mar ks	Stren gth of smell (1= strong)	# Swar m rings	# Boundar ies
STAR1 66	1	38.4 6	0.04 2	0.02 4	0.8 33	0	1	1	0	3	3	1
STAR1 66	2	30.3 7	0.04 2	0.02 3	0.8 13	0	1	1	0	3	5	1
STAR1 66	3	27.7 0	0.04 1	0.02 1	0.7 47	0	0	1	1	3	6	1
BB200 0	1	37.0 1	0.03 7	0.02 5	0.7 75	0	1	1	0	2	2	4
BB200 0	2	36.6 1	0.04 4	0.02 5	0.7 60	0	1	1	0	2	2	4
BB200 0	3	36.4 0	0.04 3	0.02 5	0.7 64	0	1	1	0	2	2	4

Table A7: Strains used in supplementary data

Strain	Name in this study	Description	Reference	KAG#	AC#
Proteus mirabilis					
BB2000	BB2000	wild-type *P. mirabilis* BB2000 strain	Belas et al., 1991	KAG 0001	AC52
BB2000 Δ*ids* carrying pIdsBB-IdsE[K85N, L113V, N126K, I166N, N188Y, T248N, N260D]-GFP	BB2000 Δ*ids* pIdsBB-IdsE-mut1	BB2000 Δ*ids* complemented with *ids* operon expressed from its native promoter where IdsE has a C-terminal GFPmut2 and contains mutations K85N, L113V, N126K, I166N, N188Y, T248N, N260D	This study	KAG 4830	AC56
BB2000 Δ*ids* carrying pIdsBB-IdsE[L48Q, F64C, I208F, V230L]-GFP	BB2000 Δ*ids* pIdsBB-IdsE-mut2	BB2000 Δ*ids* complemented with *ids* operon expressed from its native promoter where IdsE has a C-terminal GFPmut2 and contains mutations L48Q, F64C, I208F, V230L	This study	KAG 4832	AC58
BB2000 Δ*idsE* carrying pTet-FLAG-IdsE[T246A, S247A, T248A]	BB2000 Δ*idsE* pTet-IdsE-mut3	BB2000 Δ*idsE* carrying vector where IdsE contains an N-terminal FLAG tag and T246A, S247A, T248A mutations	This study	KAG 4086	AC 156
BB2000 carrying empty vector	BB2000 empty vector	BB2000 carrying a plasmid without promoter-gene insert to confer antibiotic resistance	This study	KAG 4183	AC 232
BB2000 Δ*ids* $tssB_{T95G}$	CCS05	BB2000 Δ*ids* $tssB_{T95G}$; deficient in T6SS-mediated transport	Saak & Gibbs, 2016	KAG 2115	AC67
BB2000 Δ*idsE*	Δ*idsE*	BB2000 with a chromosomal *idsE* deletion	Zepeda-Rivera et al., 2018	KAG 3126	AC 336

Table A7: Strains used in supplementary data (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
BB2000 Δ(*idsE*, *sdaC*)	ACH01	BB2000 Δ*idsE* with a chromosomal deletion of *sdaC* (BB2000_0742)	This study	KAG 3974	AC 120
BB2000 *tssB*$_{T95G}$	ACH02	BB2000 with a mutation in *tssB* (BB2000_0821) to inactivate type VI secretion	This study	KAG 4370	AC 447
BB2000 Δ*idsE* *tssB*$_{T95G}$	ACH03	BB2000 Δ*idsE* with a mutation in *tssB* (BB2000_0821) to inactivate type VI secretion	This study	KAG 4263	AC 290
BB2000 Δ(*sdaA*, *sdaB*)	ACH04	BB2000 with chromosomal deletions of *sdaA* (BB2000_1697) and *sdaB* (BB2000_0741)	This study	KAG 4303	AC 338
BB2000 Δ(*sdaA*, *sdaB*) carrying empty vector	ACH04 empty vector	ACH04 carrying an empty vector	This study	KAG 4303	AC 338
BB2000 Δ(*sdaA*, *sdaB-sdaC*)	ACH05	ACH04 with a chromosomal deletion of *sdaC*	This study	KAG 4328	AC 393
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying empty vector	ACH05 empty vector	ACH05 carrying an empty vector	This study	KAG 4388	AC 466
BB2000 Δ(*sdaA*, *sdaB*, *sstT*)	ACH04 Δ*sstT*	ACH04 with a chromosomal deletion of *sstT* (BB2000_0146)	This study	KAG 4359	AC 431
BB2000 Δ(*sdaA*, *sdaB*, *yhaO*)	ACH04 Δ*yhaO*	ACH04 with a chromosomal deletion of *yhaO* (BB2000_2747)	This study	KAG 4361	AC 433
BB2000 Δ(*sdaA*, *sdaB-sdaC*) carrying pSdaC	ACH05 pSdaC	ACH05 carrying pSdaC	This study	KAG 4391	AC 469

Table A7: Strains used in supplementary data (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
BB2000 Δ(*sdaA, sdaB, idsE*)	AC06	ACH04 with a chromosomal deletion of *idsE*	This study	KAG 4305	AC 340
BB2000 Δ*sdaC*	Δ*sdaC*	BB2000 with a chromosomal deletion of *sdaC*	This study	KAG 3972	AC118
BB2000 Δ(*idsE, sdaC*) carrying pmCherry-SdaC	ACH01 pmCherry-SdaC	ACH01 carrying a modified pSdaC plasmid with an N-terminal mCherry fluorescent protein after the start codon	This study	KAG 4473	AC 557
BB2000 Δ(*idsE, sdaC*) carrying pmCherrySdaC-I115A	ACH01 pmCherrySdaC-I115A	ACH01 carrying pmCherry-SdaC with I115A amino acid change	This study	KAG 4506	AC 580
BB2000 Δ(*idsE, sdaC*) carrying pmCherrySdaC-H325A	ACH01 pmCherrySdaC-H325A	ACH01 carrying pmCherry-SdaC with H325A amino acid change	This study	KAG 4498	AC 573
BB2000 Δ(*idsE, sdaC*) carrying pmCherrySdaC-V222W	ACH01 pmCherrySdaC-V222W	ACH01 carrying pmCherry-SdaC with V222W amino acid change	This study	KAG 4497	AC 572
BB2000 Δ(*idsE, sdaC*) carrying pmCherrySdaC-G328V	ACH01 pmCherrySdaC-G328V	ACH01 carrying pmCherry-SdaC with G328V amino acid change	This study	KAG 4499	AC 574
BB2000 Δ(*idsE, sdaC*) carrying pmCherrySdaC-G332R	ACH01 pmCherrySdaC-G332R	ACH01 carrying pmCherry-SdaC with G332R amino acid change	This study	KAG 4500	AC 575
BB2000 Δ(*sdaA, sdaB-sdaC*) carrying pmCherry-SdaC	ACH05 pmCherry-SdaC	ACH05 carrying pmCherry-SdaC	This study	KAG 4472	AC 556
BB2000 Δ(*sdaA, sdaB-sdaC*) carrying pmCherrySdaC-I115A	ACH05 pmCherrySdaC-I115A	ACH05 carrying pmCherry-SdaC with I115A amino acid change	This study	KAG 4507	AC 581

Table A7: Strains used in supplementary data (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
BB2000 Δ(*sdaA, sdaB-sdaC*) carrying pmCherrySdaC-H325A	ACH05 pmCherrySdaC-H325A	ACH05 carrying pmCherry-SdaC with H325A amino acid change	This study	KAG 4502	AC 577
BB2000 Δ(*sdaA, sdaB-sdaC*) carrying pmCherrySdaC-V222W	ACH05 pmCherrySdaC-V222W	ACH05 carrying pmCherry-SdaC with V222W amino acid change	This study	KAG 4501	AC 576
BB2000 Δ(*sdaA, sdaB-sdaC*) carrying pmCherrySdaC-G328V	ACH05 pmCherrySdaC-G328V	ACH05 carrying pmCherry-SdaC with G328V amino acid change	This study	KAG 4503	AC 578
BB2000 Δ(*sdaA, sdaB-sdaC*) carrying pmCherrySdaC-G332R	ACH05 pmCherrySdaC-G332R	ACH05 carrying pmCherry-SdaC with G332R amino acid change	This study	KAG4504	AC 579
STAR 17	STAR 17	Info in Table A5	This study		
STAR 23	STAR 23	Info in Table A5	This study		
STAR 25	STAR 25	Info in Table A5	This study		
STAR 29	STAR 29	Info in Table A5	This study		
STAR 36	STAR 36	Info in Table A5	This study		
STAR 38	STAR 38	Info in Table A5	This study		
STAR 40	STAR 40	Info in Table A5	This study		
STAR 42	STAR 42	Info in Table A5	This study		
STAR 44	STAR 44	Info in Table A5	This study		
STAR 46	STAR 46	Info in Table A5	This study		
STAR 50	STAR 50	Info in Table A5	This study		
STAR 52	STAR 52	Info in Table A5	This study		
STAR 54	STAR 54	Info in Table A5	This study		
STAR 56	STAR 56	Info in Table A5	This study		
STAR 58	STAR 58	Info in Table A5	This study		
STAR 60	STAR 60	Info in Table A5	This study		
STAR 62	STAR 62	Info in Table A5	This study		
STAR 64	STAR 64	Info in Table A5	This study		

Table A7: Strains used in supplementary data (Continued)

Strain	Name in this study	Description	Reference	KAG#	AC#
STAR 66	STAR 66	Info in Table A5	This study		
STAR 68	STAR 68	Info in Table A5	This study		
STAR 70	STAR 70	Info in Table A5	This study		
STAR 74	STAR 74	Info in Table A5	This study		
STAR 78	STAR 78	Info in Table A5	This study		
STAR 80	STAR 80	Info in Table A5	This study		
STAR 82	STAR 82	Info in Table A5	This study		
STAR 108	STAR 108	Info in Table A5	This study		
STAR 112	STAR 112	Info in Table A5	This study		
STAR 124	STAR 124	Info in Table A5	This study		
STAR 138	STAR 138	Info in Table A5	This study		
STAR 140	STAR 140	Info in Table A5	This study		
STAR 142	STAR 142	Info in Table A5	This study		
STAR 144	STAR 144	Info in Table A5	This study		
STAR 146	STAR 146	Info in Table A5	This study		
STAR 148	STAR 148	Info in Table A5	This study		
STAR 152	STAR 152	Info in Table A5	This study		
STAR 154	STAR 154	Info in Table A5	This study		
STAR 156	STAR 156	Info in Table A5	This study		
STAR 160	STAR 160	Info in Table A5	This study		
STAR 162	STAR 162	Info in Table A5	This study		
STAR 164	STAR 164	Info in Table A5	This study		
STAR 166	STAR 166	Info in Table A5	This study		

Table A8: Suppressor strains used in supplementary data

Suppressor strain background	SdaC mutation	Description	Reference
BB2000 Δ*ids* pIdsBB-IdsE[K85N, L113V, N126K, I166N, N188Y, T248N, N260D]-GFP	L371*	Suppressor 1, results in non-functional SdaC	This study
BB2000 Δ*ids* pIdsBB-IdsE[K85N, L113V, N126K, I166N, N188Y, T248N, N260D]-GFP	Δ68bp (366-433)	Suppressor 2, results in non-functional SdaC	This study
BB2000 Δ*ids* pIdsBB-IdsE[K85N, L113V, N126K, I166N, N188Y, T248N, N260D]-GFP	E243*	Suppressor 3, results in non-functional SdaC	This study
BB2000 Δ*ids* pIdsBB-IdsE[L48Q, F64C, I208F, V230L]-GFP	Δ1bp (244)	Suppressor 4, results in non-functional SdaC	This study
BB2000 Δ*ids* pIdsBB-IdsE[L48Q, F64C, I208F, V230L]-GFP	G332R	Suppressor 5, results in full-length SdaC with single residue change	This study
BB2000 Δ*ids* pIdsBB-IdsE[L48Q, F64C, I208F, V230L]-GFP	Δ1bp (200)	Suppressor 6, results in non-functional SdaC	This study
BB2000 Δ*ids* pIdsBB-IdsE[L48Q, F64C, I208F, V230L]-GFP	Δ49bp (575-623)	Suppressor 7, results in non-functional SdaC	This study
BB2000 Δ(*sdaA, sdaB, idsE*)	+2bp (363)	Suppressor 8, results in non-functional SdaC	This study
BB2000 Δ(*sdaA, sdaB, idsE*)	G192*	Suppressor 9, results in non-functional SdaC	This study
BB2000 Δ(*sdaA, sdaB, idsE*)	Δ2194bp (sdaBC and 751bp upstream and 123bp downstream)	Suppressor 10, results in non-functional SdaC	This study
BB2000 Δ*idsE* pTet-FLAG-IdsE[T246A, S247A, T248A]	G328V	Suppressor 11, results in full-length SdaC with single residue change	This study
BB2000 Δ*idsE* pTet-FLAG-IdsE[T246A, S247A, T248A]	Δ1bp (725)	Suppressor 12, results in non-functional SdaC	This study

Table A9: Primers used in this study

Primer name	Primer sequence
oAC006	GGCCCATGCCTGAGCTCGATT
oAC007	CAGCTGATCCGGATCCCGCA
oAC041	ACTACCTCAGGGATACTACGCATGG
oAC046	ATCGGGGCCCGCGAAAGTTAAAATAATGTTTTA
oAC047	TAAACTTAAATACTCTCCGAAATAACGCGGT
oAC048	AGAGTATTTAAGTTTATCGATGGCTACTTTC
oAC049	ATTAACTAGTATAGTTCGCCATTATAGGCGCTG
oAC050	ATTAGGGCCCATACATTGCACCTTTCATTGCCT
oAC051	GTGTTAAAAATATTCGCCTCCCTAATTTAAAGGC
oAC052	CGAATATTTTTAACACTTTGACATGATTGTTACCCA
oAC053	ATTAACTAGTATAAGCACCAATTTACCCTGTTT
oAC054	GTGTTAAAAATACTCTCCGAAATAACGCGGTAA
oAC055	AGAGTATTTTTAACACTTTGACATGATTGTTAC
oAC068	AAATGGCTAGCTTAAGACCCACTTTCACATTTAAG
oAC071	CGCCGACACGGGTCACGCTG
oAC072	ATGGGCAAATGGCTAGCTCCTATTAGGATAAGAAA
oAC113	GTGATCACGGTAGTTTATCC
oAC114	TGTAGAGCTACATGAGTGAT
oAC118	TTTAGGGGCTCGTGAACGTTTCAATGGCTTAAT
oAC119	ATTAAGCCATTGAAACGTTCACGAGCCCCTAAA
oAC139	GCATACTCATACAAGGAGCTTA
oAC140	GCTCTTGAGCTTCTGCATTT
oAC141	AACCCGGGCCCATATCCCAGTATGGTTTCTGTGAT
oAC142	ATAAAATCAGGCTAAGTCGAAAACGCTAATCACGT
oAC143	ATTAGCGTTTTCGACTTAGCCTGATTTTATTTCTA
oAC144	ACTATACTAGTATATAATGAGATTTATTTTAAGACAGGCTACTCC
oAC145	AACCCGGGCCCATATTTCCAGTTTAGGATACGTTG
oAC146	TACTACTTATACTCTAGTTATTCCATTACTTAAAATAATATTTAAAAAAGCAATT
oAC147	AGTAATGGAATAACTAGAGTATAAGTAGTAAATAGATATAAAGAGTAGTAAGACA
oAC148	ACTATACTAGTATATTTGTGTGAGCTTGATCAAACACCAA
oAC153	TCTCATGCCGGACAAGTTGC
oAC154	ACAATTGACCCCCAACCAAA
oAC156	GGCTTCCTAAAACTGAGTCA
oAC157	GCAGGGAACAGAATTAGCAC
oAC161	AACCCGGGCCCATATTATATAATTAACACTCTTTT
oAC162	AACCGCAGTTGACGAATTAGTACTCACTTTTTATATTGTAATTGT

Table A9: Primers used in this study (Continued)

Primer name	Primer sequence
oAC163	AAAGTGAGTACTAATTCGTCAACTGCGGTTCGAGTTAGCTGTTAT
oAC164	ACTATACTAGTATATTAGCAATACTGACAATGGCATAACCTGACT
oAC177	ATTTCGGAGAGTATTATGGAAACGACTCAAACCAG
oAC178	CTGGCCTCAGGGATATTAGCTGAACAGAGAGTAGA
oAC179	CGGTTTACGCACTTCTTGTC
oAC180	TCTTCAGCACAAACCGTCGC
oAC185	TTGAGTCGTTTCCATAATACTCTCCGAAATTTTCTCTATC
oAC186	ATTAGTACTCACTTTTTTCTCTATCACTGATAGGGAGTGGTAAAA
oAC187	TCAGTGATAGAGAAAAAAGTGAGTACTAATATGGAAACGGCTTCC
oAC188	CTGGCCTCAGGGATACTATGAGAATGCCAAGAACGGAGAAACACA
oAC196	ATCGTCTTTGTAGTCTTTATCGTCATCGTCTTTGTAGTCC
oAC197	GTAGTCTTTATCGTCATCGTCTTTGTAGTCTTTATCGTCA
oAC198	TTGTATCTTTATCGTCATCGTCTTTGTAGTCTTTATCGTC
oAC207	CCCTGTGCGATAGAACCAGCTTTGGTTGTATCTTTATCGT
oAC208	TACAACCAAAGCTGGTTCTATCGCACAGGGTAATA
oAC226	CTGTGCGATAGAACCAGCTTTGGTTGTATCCTTGTACAGCTCGTCCATGC
oAC227	TTAGGTCACTATTTAGTGGCTCGTGAAGGTTTC
oAC228	GAAACCTTCACGAGCCACTAAATAGTGACCTAA
oAC264	TTTGCTATTTATCCTGCCTTGTTGGTTTACAGT
oAC265	ACTGTAAACCAACAAGGCAGGATAAATAGCAAA
oAC268	AAATCTTTCTTAGGTGCCTATTTAGGGGCTCGT
oAC269	ACGAGCCCCTAAATAGGCACCTAAGAAAGATTT
oAC293	TGGTTAGCTATCCCTTGGATGGTGTTCTCTTTT
oAC294	AAAAGAGAACACCATCCAAGGGATAGCTAACCA
oAC295	CTCGCCCTTGCTCACCATAATACTCTCCGAAATAACGCGG
oAC296	TCGGAGAGTATTATGGTGAGCAAGGGCGAGGAGGATAACA
oAC314	CACTACCTTTTACATTAACTTGGATTTTGTTTATCCAAACATT
oAC315	AATGTTTGGATAAACAAAATCCAAGTTAATGTAAAAGGTAGTG

Materials and Methods:

Bacterial strains and media

The strains and plasmids used in the supplementary information are described in Tables A7 and A8. *P. mirabilis* strains were maintained on low swarm (LSW) agar (Belas et al., 1991). CM55 blood agar base agar (Oxoid, Basingstoke, England) was used for swarm-permissive nutrient plates, except when LB agar was used as a substitute in Figure S2G to vary agar concentration. Overnight cultures of all strains were grown at 37°C in LB broth under aerobic conditions. For growth curve assays, cells were grown in LB or minimal medium [M9 salts (3 g/L KH_2PO_4, 6.8 g/L Na_2HPO_4, 0.5 g/L NaCl, 1.0 g/L NH_4Cl), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% glucose] supplemented with 10mM L-serine (VWR, Beantown chemical, BT128350) when stated. Kanamycin (Corning, Corning, NY) was used at a concentration of 35 μg/ml for plasmid maintenance and was added to swarm and growth media when appropriate. Other antibiotics were used as follows for transforming plasmids into *P. mirabilis*: 15 μg/ml tetracycline (Amresco Biochemicals, Solon, OH), and 25 μg/ml streptomycin (Sigma-Aldrich, St. Louis, MO).

Plasmid construction

Restriction digestion using restriction enzymes described (New England BioLabs, Ipswich, MA). Ligations were resolved in OneShot Omnimax2 T1R competent cells (Thermo Fisher Scientific, Waltham, MA) or SM10λpir (Simon et al., 1983). The resultant plasmids were confirmed by Sanger sequencing (Genewiz, Inc., South Plainfield, NJ), and correct resultant plasmids were then transformed into *P. mirabilis* as described previously (Gibbs et al., 2008) using *E. coli* conjugative strain MFDpir

(Ferrières et al., 2010). Table S2 contains the nucleotide sequences for listed primers, all of which contain the prefix "oAC."

For pmCherry-SdaC, the native promoter for the *sdaC* gene was amplified from pSdaC using oAC072 and oAC295. Next, the gene encoding mCherry was amplified using oAC296 and oAC226. SdaC was amplified from pSdaC using oAC208 and oAC071. PCR products were joined together using overlap extension PCR with oAC072 and oAC071. The full insert was ligated into the pSdaC backbone using NheI and PshAI. Point mutations were made in the pmCherry-SdaC as described for point mutations in pSdaC using the same primers.

Strain construction

All chromosomal deletions were performed as described earlier using pKNG101-derived suicide vectors (Saak & Gibbs, 2016).

For strains ACH02 (BB2000 *tssB*$_{T95G}$) and ACH03 (BB2000 *tssB*$_{T95G}$ Δ*idsE*), pCS34 (Saak & Gibbs, 2016) was mated into BB2000 to construct ACH02 and BB2000 Δ*idsE* to construct ACH03. Matings were subjected to antibiotic selection on LSW agar (with 15 g/ml Tet and 25 g/ml Strep). Candidate strains were subjected to sucrose counterselection as described (Sturgill et al., 2002). Double recombinants were confirmed by PCR of the SNP-containing region of *tssB* (BB2000_0821) using oAC139 and oAC140.

For strain BB2000 Δ*sdaC*, 500bp regions adjacent on either side to *sdaC* with restriction sites were amplified using oAC046-049, and ligated into pKNG101 using restriction enzymes ApaI and SpeI. The resulting vector was mated into BB2000. Matings were

subjected to antibiotic selection on LSW agar (with 15 g/ml Tet and 25 g/ml Strep).

Candidate strains were subjected to sucrose counterselection as described (Sturgill et

al., 2002). For BB2000 Δ*sdaC*, double recombinants were confirmed by PCR using

oAC113 and oA114.

For strain ACH06 [BB2000 Δ(*sdaA*, *sdaB*, *idsE)*], the regions flanking the *sdaB* gene

were amplified using overlap extension PCR with oAC050-053 and ligated into

pKNG101 using enzymes ApaI and SpeI. The resulting vector was mated into BB2000

Δ*idsE* and subjected to sucrose selection. Double recombinants were confirmed by

colony PCR of the region using oAC113 and oAC114. The regions flanking the *sdaA*

gene were amplified using overlap extension PCR with oAC141-144 and ligated into

pKNG101 using enzymes ApaI and SpeI. The resulting vector was mated into BB2000

Δ(*idsE, sdaB*) and subjected to sucrose selection. Double recombinants were confirmed

by PCR of the surrounding region using oAC153 and oAC154. ACH06 was confirmed

by whole genome sequencing.

Growth curve

Overnight cultures were normalized to an an optical density at 600 nm (OD600) of 0.1 in

LB medium or minimal medium [M9 salts (3 g/L KH_2PO_4, 6.8 g/L Na_2HPO_4, 0.5 g/L NaCl,

1.0 g/L NH_4Cl), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% glucose] supplemented with 10mM L-

serine (VWR, Beantown chemical, BT128350) when stated. Glucose was omitted from

the minimal medium to test utilization of serine as a carbon source. Medium was

supplemented with kanamycin for plasmid maintenance when appropriate. Normalized

cultures were grown overnight at 37°C, with periodic shaking, in a Tecan Infinite 200 PRO microplate reader (Tecan, Männedorf, Switzerland).

Swarm expansion and boundary assay

Overnight cultures were normalized to an OD600 of 0.1, and swarm-permissive nutrient plates, supplemented with kanamycin when appropriate, were inoculated with 2 µl of normalized culture in the center. To vary agar concentration, LB medium with various percentages of agar was substituted for CM55 (Oxoid, Basingstoke, England) as the swarm-permissive nutrient medium. Plates were incubated at room temperature for two days, and the radii of actively migrating swarms starting from the edge of the inoculum were measured using Fiji (ImageJ) (Schindelin et al., 2012). For STAR strains, plates were incubated for 16 hours at 30°C. Colony-forming units (CFUs) per swarm colony were calculated by resuspending all of the cells from a single plate in LB at the end of the swarm assay experiment and plating dilutions onto LSW agar plates that were incubated overnight at 37°C for growth of single colonies. Boundary assays were performed as reported (Wenren et al., 2013).

Microscopy

One-millimeter-thick swarm-permissive agar pads supplemented with kanamycin were inoculated with overnight cultures and incubated overnight at room temperature. The agar pads were then incubated at 37°C in a modified humidity chamber. After five to six hours, the pads were imaged by phase-contrast microscopy (10 ms exposure) and fluorescence microscopy (RFP channel, 500 ms exposure) using a Leica DM5500B microscope system (Leica Microsystems, Buffalo Grove, IL) and a CoolSnap HQ2

cooled charge-coupled device camera (Photometrics, Tucson, AZ). MetaMorph (version 7.8.0.0; Molecular Devices, Sunnyvale, CA) was used for image acquisition.

Antibiotic clearance assay

50 µl of overnight culture was spread evenly across a swarm-permissive nutrient plate. 6-mm blank paper disks were soaked into antibiotics [Kan (350μg/ml), Amp (500μg/ml), or Carb (100μg/ml)] or water as a control before placing on plates. After incubating plates at 37°C overnight, the radius of clearance surrounding the disk was measured.

Bioinformatics

I-TASSER (Roy et al., 2010; Yang et al., 2015; Zhang, 2008) was used to predict the protein structure of SdaC-Pmir, SdaC-Ecol, and YhaO. Figures aligning structural models were made using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Additional IdsD and SdaC sequences from *P. mirabilis* genomes were identified by using tblastn (NCBI) with the *P. mirabilis* amino acid sequence as the query. Sequences with less than 90% query coverage and IdsD sequences from the 1072 subfamily were excluded. Sequence alignments were generated using MUSCLE (Edgar, 2004; Madeira et al., 2019). AlignmentViewer (Reguant et al., 2020) was used to visualize MUSCLE alignments of IdsD and SdaC for Figure S6. Mauve was used to reorder contigs and align genomes of representative STAR isolates and BB2000 (Darling et al., 2004). ANI values for STAR strains and BB2000 were calculated through Anvi'o (Eren et al., 2015) using PyANI (Pritchard et al., 2016). PhyloPhlAn 3.0 was used to construct the phylogenetic tree of STAR strains and BB2000 (Asnicar et al., 2020).

References

Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., ... &
    Segata, N. (2020). Precise phylogenetic analysis of microbial isolates and
    genomes from metagenomes using PhyloPhlAn 3.0. *Nature communications*,
    11(1), 1-10.

Belas, R., Erskine, D., & Flaherty, D. (1991). Transposon mutagenesis in *Proteus
    mirabilis*. *J Bacteriol*, *173*(19), 6289–6293.

Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment
    of conserved genomic sequence with rearrangements. *Genome research*, *14*(7),
    1394-1403.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and
    high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., &
    Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform
    for 'omics data. *PeerJ*, *3*, e1319.

Ferrières, L., Hémery, G., Nham, T., Guérout, A.-M., Mazel, D., Beloin, C., & Ghigo, J.-
    M. (2010). Silent mischief: Bacteriophage Mu insertions contaminate products of
    Escherichia coli random mutagenesis performed using suicidal transposon
    delivery plasmids mobilized by broad-host-range RP4 conjugative machinery.
    Journal of Bacteriology, *192*(24), 6418–6427.

Gibbs, K. A., Urbanowski, M. L., & Greenberg, E. P. (2008). Genetic determinants of
    self identity and social recognition in bacteria. *Science*, *321*(5886), 256–259.
    https://doi.org/10.1126/science.1160033

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P.,
    Tivey, A. R., Potter, S. C., & Finn, R. D. (2019). The EMBL-EBI search and
    sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–
    W641.

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2016).
    Genomics and taxonomy in diagnostics for food security: soft-rotting
    enterobacterial plant pathogens. *Analytical Methods*, *8*(1), 12-24.

Reguant, R., Antipin, Y., Sheridan, R., Dallago, C., Diamantoukos, D., Luna, A., Sander,
    C., & Gauthier, N. P. (2020). AlignmentViewer: Sequence Analysis of Large
    Protein Families. *F1000Research*, *9*(213), 213.

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated
    protein structure and function prediction. *Nature Protocols*, *5*(4), 725.

Saak, C. C., & Gibbs, K. A. (2016). The Self-Identity Protein IdsD Is Communicated between Cells in Swarming *Proteus mirabilis* Colonies. *Journal of Bacteriology*, *198*(24), 3278–3286.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., & Schmid, B. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.

Simon, R., Priefer, U., & Pühler, A. (1983). A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria. *Bio/Technology*, *1*(9), 784–791.

Sturgill, G. M., Siddiqui, S., Ding, X., Pecora, N. D., & Rather, P. N. (2002). Isolation of *lacZ* fusions to *Proteus mirabilis* genes regulated by intercellular signaling: Potential role for the sugar phosphotransferase (Pts) system in regulation. *FEMS Microbiology Letters*, *217*(1), 43–50.

Wenren, L. M., Sullivan, N. L., Cardarelli, L., Septer, A. N., & Gibbs, K. A. (2013). Two independent pathways for self-recognition in *Proteus mirabilis* are linked by type VI-dependent export. *MBio*, *4*(4), e00374-13. https://doi.org/10.1128/mBio.00374-13

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7.

Zepeda-Rivera, M. A., Saak, C. C., & Gibbs, K. A. (2018). A proposed chaperone of the bacterial type VI secretion system functions to constrain a self-identity protein. *J Bacteriol*. https://doi.org/10.1128/JB.00688-17

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 40.