

Abstract

The genotype-to-phenotype (G2P) problem is a fundamental challenge in understanding how genetic information controls the collective phenotypic outputs of multicellular organisms. To address this problem, this thesis focuses on *Myxococcus xanthus*, a model organism widely used for studying social behavior and morphological differentiation in bacteria. The social behaviors of *M. xanthus* are mediated by complex gene regulatory networks, making it an excellent model organism for studying the G2P problem. This thesis presents two studies that provide valuable insights into the G2P problem in *M. xanthus*.

The first study investigates the dynamics of biofilm morphogenesis, an essential process for bacteria survival, using image capture and analysis techniques from custom-designed microscopes. Biofilms are complex structures composed of different cell types and gene expression patterns. The study produced a topological map of the process of biofilm formation in wild-type *M. xanthus*, which allowed the identification and characterization of even subtle mutations associated with different mutations with statistical significance. Stochastic variation was mapped, and this methodology allowed us to distinguish previously non-distinguishable genotypes of *M. xanthus* using phenotype data. The approach used in this study has the potential to enhance our understanding of the genetic and environmental factors that contribute to the development of *M. xanthus* and other organisms.

The second study employs a "common garden" approach to investigate the impact of transcriptional regulators on development in *M. xanthus*. By recording subtle differences in traditional phenotype assays across a library of mutant transcriptional regulators, the study shows there are uncharacterized regulatory genes that play significant roles in regulating biofilm dynamics and gene expression during development. The study identified sigma factors that have an impact on sporulation fitness and characterized the effects of these regulators on the development pathways. The findings of this study supply a deeper understanding of the G2P problem in *M. xanthus* and could have broader implications for understanding the development of other organisms.

By investigating the complex interplay between genotype and phenotype, this thesis aims to shed light on fundamental mechanisms underlying multicellular development, and the potential for these findings to be applied in the fields of biotechnology and medicine.

Myxococcus xanthus: an approach to phenotyping complex biofilms

by

Eduardo Caro

B.S., University of Central Florida, 2017

Dissertation

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biology

Syracuse University
June 2023

Copyright © Eduardo Caro 2023

All Rights Reserved

Acknowledgments

I would like to thank my collaborators, mentors, friends, and family for their continuous support through graduate school. This thesis represents a summary of my work and accomplishments of scientific research from myself, my lab, and all the many science discussions I have had with those during my time at Syracuse University.

To my collaborators, for helping me improve my interdisciplinary knowledge of physics and computer science and be better at communicating non-biology.

To my mentors, for helping me develop my critical thinking and leadership skills as a scientist and encouraging me to help and mentor others.

To my family and friends, who pushed me to be myself throughout the process and supported my endeavors as a scientist.

Most importantly, this thesis is dedicated to my parents; my dad, who gave me the confidence to pursue higher education, and my mom who teaches me every day that working hard is always the right thing to do. I am proud to present the following work as a collection of all these contributions.

Thank you,

Eddie

Contents

CHAPTER 1: INTRODUCTION	1
1.1 GENOTYPE TO PHENOTYPE (G2P)	1
A. UNDERSTANDING DNA AS A BLUEPRINT FOR LIFE	1
B. TECHNIQUES FOR IDENTIFYING GENOTYPE-TO-PHENOTYPE CONNECTIONS	3
C. GENOTYPE-TO-PHENOTYPE IN MODEL ORGANISMS	5
D. PITFALLS OF GENOTYPE-TO-PHENOTYPE APPROACHES	9
1.2 MYXOCOCCUS XANTHUS	12
A. MULTICELLULARITY, SELF-ORGANIZATION, AND EMERGENT BEHAVIORS	12
B. THE <i>M. XANTHUS</i> GENOME	15
1.3 <i>M. XANTHUS</i> EMERGENT BEHAVIORS AND THEIR USE IN GENOTYPE-TO-PHENOTYPE STUDIES	16
A. TRANSPOSON LIBRARY MAPPING	16
B. UNDERSTANDING THE ROLE OF REGULATORY GENES IN MORPHOGENESIS	17
C. SOCIAL MICROBES AS MODELS OF MULTICELLULARITY	18
D. REFERENCES	21
CHAPTER 2: STOCHASTIC BOUNDS IN THE STUDY OF AGGREGATION DYNAMICS PROVIDE A MEANS TO DIFFERENTIATE BETWEEN SOCIAL BACTERIA.	30
ACKNOWLEDGEMENTS	30
A. INTRODUCTION	30
B. RESULTS	35
C. DISCUSSION	47

D. CONCLUSIONS	53
E. METHODS	54
F. FUNDING ACKNOWLEDGMENTS	58
G. REFERENCES:	60
CHAPTER 3: A 'COMMON GARDEN' APPROACH TO INVESTIGATING THE ROLE OF KEY SIGMA (σ) FACTORS IN <i>MYXOCOCCUS XANTHUS</i> DEVELOPMENT	
	65
A. PROJECT SUMMARY	65
B. INTRODUCTION	66
G. RESULTS	84
H. DISCUSSION	86
I. CONCLUSION	88
J. MATERIALS AND METHODS	89
SUPPLEMENTARY DATA	95
D. REFERENCES	99
CHAPTER 4: CONCLUSIONS	
	107
A. SUMMARY	107

CHAPTER 1: INTRODUCTION

1.1 GENOTYPE TO PHENOTYPE (G2P)

A. UNDERSTANDING DNA AS A BLUEPRINT FOR LIFE

The central dogma of molecular biology is a critical concept that highlights the process by which genetic information flows within the cell.^{1,2} This flow is vital for the maintenance of cellular function and the emergence of higher-order structures that underpin the function and development of organisms. For this to occur, the genetic blueprint, DNA (deoxyribonucleic acid), must function as a primary repository for storing cellular instructions over generations, thus maintaining genome stability. This mechanism was first proposed by Francis Crick in 1958, subsequently refined in 1970, and is now established as the central dogma of molecular biology.³ DNA is transcribed into messenger RNA (mRNA), which is subsequently translated into polypeptides that fold into functional proteins. These proteins and signals are controlled by intermediate regulatory proteins that are essential for maintaining cellular homeostasis in an often-changing environment, which allows for cellular growth, response to stimuli, or the emergence of multicellularity.^{4,5}

Genetic architecture refers to the genetic basis underlying the phenotypic traits observed in an organism. It encompasses the number, distribution, and effects of genetic variants that contribute to a particular trait or disease. The genetic architecture of a trait can range from simple, involving a single gene with a large effect, to complex, involving multiple genes with small effects and gene-gene and gene-environment interactions.⁵⁹ Understanding the genetic architecture of a trait is essential for predicting its heritability, identifying genetic

risk factors for disease, and developing targeted interventions.⁶⁰ Advances in high-throughput genotyping and sequencing technologies have allowed for the identification of genetic variants across the genome, enabling researchers to uncover the genetic basis of complex traits.⁶¹

The genotype of an organism refers to the whole collection of nucleotide sequences that make up its DNA, which can contain both major and minor variations resulting in distinct versions of genes known as alleles.⁶ To identify and quantify the effects of genetic differences between two alleles, a phenotype is observed and characterized, which refers to an organism's observable traits resulting from its genetic makeup. These traits can be physical, such as hair and eye color; or inherent, such as blood type and susceptibility to diseases; and can provide insight into the function of specific genes. This approach, known as genotype-to-phenotype, has been used to identify the function of specific DNA sequences, although more drastic mutations can sometimes arise that impair a gene's function, even sometimes completely.^{7,8}

Mutations can occur anywhere throughout a genome. Changes in the DNA sequences found between genes might have no effect or can affect regulatory binding sites near genes, altering gene expression indirectly, but not through a sequence change. Other mutations can occur within the DNA of protein-coding sequences of genes and can be categorized further based on their impact on the corresponding amino acid. Silent mutations do not change the original folding of the amino acids and do not disrupt protein function.⁹ These mutations are typically assumed to have no impact on an organism's fitness by measures typically employed to study gene expression and function. However, more drastic changes in nucleotide sequences can result in missense mutations, which introduce new amino acids, or nonsense mutations, which create stop signals that prematurely terminate the protein chain.¹⁰ These types of

mutations can disrupt protein function, as the appropriate protein structures may fail to form. It is important to note that alleles in protein-coding sequences could make variants that fold slightly differently but are still functional - these are "mutated" but just not in the most relevant portion of the protein to affect function. However, studying divergences in functioning proteins and genes can better understand how a system phenotype changes and evolves concerning its genetic structure. This G2P approach can reveal correlations in living systems, which can help answer a variety of questions related to health, evolution, and natural systems at the mechano-physiological level.⁷

B. TECHNIQUES FOR IDENTIFYING GENOTYPE-TO-PHENOTYPE CONNECTIONS

One method to investigate these complex systems is to induce changes within the system and examine their effects. In the context of understanding the role of genes in organisms, a conventional method is to study the repercussions of gene malfunctions on the system's behavior.¹¹ For instance, identifying the gene responsible for bacterial motility and creating a mutation that impairs its function is akin to removing a building block's base and analyzing the resultant collapse. Proteins are complex molecules made up of smaller building blocks called amino acids, which are linked together in a specific sequence. Just like a block tower, where each block must be placed in a specific order to maintain its stability, proteins require the correct sequence of amino acids to fold into their proper shape and perform their specific function. If even one amino acid is out of place, it can disrupt the entire structure of the protein, much like removing a block from a tower can cause it to collapse. This is why DNA mismatching can have significant consequences for the functioning of cells and organisms.

To better understand the genotype-to-phenotype relationship, it is also important to consider how the environment plays a crucial role in shaping emergent phenotypes. Phenotypic plasticity, whereby organisms with identical genotypes exhibit different phenotypes under different environmental conditions, provides another means to investigate the functions of genes that are not visibly disturbed by mutations.^{5,12} Subjecting a system to a range of environmental conditions maximizes the likelihood of uncovering conditions leading to molecular events that can lead to unique phenotypes. For some microbial systems, the substrate stiffness that a cell travels on can modulate and alter gene expression, leading to the production of different proteins, such as pili for stiffer surfaces, or an increase in exopolysaccharide production for softer surfaces. Moreover, the temperature can alter the genetic and proteomic expression of a system, whereby a warmer system would be kinetically more favorable and drive some metabolic reactions, while a cooler system would not have as many random molecular collisions. Therefore, considering the role of the environment in the genotype-to-phenotype relationship can enable a more holistic understanding of how complex systems function by providing more emergent phenotypes to occur.

Additionally, in the field of genetics, it is important to consider penetrance and its role in the expression of phenotypes when understanding the relationship between genotype and phenotype. Penetrance refers to the percentage of individuals carrying a particular genotype who exhibit the associated phenotype. It is influenced by genetic and environmental factors, including age, sex, and gene-gene interactions. The concept of penetrance is particularly relevant in the context of genetic diseases, where the presence of a disease-causing mutation may not always result in disease manifestation. For example, a mutation in the BRCA1 gene is

known to increase the risk of breast and ovarian cancer. However, not all individuals with this mutation will develop cancer, and the penetrance of this mutation varies depending on the population studied and other genetic and environmental factors.⁵⁸ Understanding penetrance is important for genetic counseling and clinical management of individuals with genetic mutations.

In particular, the investigations of social microbes have shed light on the evolution and maintenance of multicellularity by demonstrating how cooperative behaviors can emerge and persist in simple organisms without the need for more complicated developmental pathways.⁶⁹ For example, canalization is a genetic concept that describes the ability of a genotype to produce a consistent and stable phenotype, despite environmental variation. In other words, canalization allows the robustness of genetic networks and regulatory mechanisms to buffer against environmental inputs that may otherwise disrupt or alter the phenotype.⁷⁰ This phenomenon is particularly important in the study of complex traits, such as development and disease, as it helps explain why some phenotypes remain unaffected by environmental variation while others are more sensitive.

C. GENOTYPE-TO-PHENOTYPE IN MODEL ORGANISMS

Model organisms have been utilized to investigate cellular and developmental mechanisms that may have relevance broadly to other organisms and that may not be readily available or ethical to study in humans. These models are non-human, yet exhibit comparable genetic information and, for some systems, developmental programming, enabling comparative analysis across multiple studies.¹³ In addressing a biological question, scientists must

contemplate several factors, including time, cost, feasibility, ethics, and others, when selecting an appropriate model. However, due to the conservation of biology across phylogenetically related species, genetic commonalities across distinct species and systems can be identified and related to each other. Important examples across some bacteria and fungi include:

- *S. cerevisiae*, commonly known as baker's yeast, is a popular model eukaryotic model organism for molecular biology research due to its simple genetics, fast growth, and easy manipulation. Its genome has been sequenced and found to have many homologs to human genes, making it useful for studying basic cellular processes and human diseases. The yeast's unique characteristics have led to its use in the study of aging and age-related diseases, and the development of genetic tools and resources has made it accessible for genetic manipulation and high-throughput screens. *S. cerevisiae's* generation time of a few hours along with the lifespan of two to three weeks has also made it an attractive organism for studying the fundamental mechanisms of aging.^{74,75}
- *Dictyostelium discoideum*, fungi, is a soil slime mold (social amoeba) that behaves similarly to prokaryotes but has the genetic composition of a eukaryote¹⁴. *D. discoideum* has both unicellular and multicellular stages of life that can be studied for the emergence of specialization via its capacity of undergoing development. Under starvation conditions, the amoeba proceeds to make an "aggregation competent" community that can cluster together and differentiate at the cellular level. The "slug" of differentiated cells can then migrate to the soil surface and progress to its last phase of

morphogenesis—the formation of a fruiting body consisting of a basal disc, a stalk, and a droplet of spores at the top of the stalk.¹⁵

- *Cryptococcus neoformans*, is a fungal pathogen that causes life-threatening meningitis in immunocompromised individuals. Its genome has been fully sequenced, and it exhibits a range of unique features, including the ability to grow at elevated temperatures and the presence of a polysaccharide capsule that protects against host defenses.⁶² These features make *C. neoformans* an excellent model organism for studying the molecular mechanisms of fungal pathogenesis and the evolution of virulence traits. Additionally, *C. neoformans* exhibits a haploid-dominant life cycle, making it an ideal system for genetic manipulation and the study of gene function.⁶³
- *V. cholerae* is a bacterium that causes cholera, a severe diarrheal disease affecting millions worldwide. Its ability to colonize and cause disease in the human gut is closely linked to the expression of virulence genes, making it an ideal model organism for studying the molecular basis of bacterial pathogenesis.⁶⁴ Furthermore, *V. cholerae* exhibits a high degree of genetic diversity, with multiple strains and serotypes that differ in their virulence and environmental adaptations. This diversity provides a valuable opportunity to study the evolution of bacterial populations and the molecular mechanisms driving adaptation to changing environments.⁶⁵

- *Escherichia coli*, a prokaryote, is considered the most influential regarding the progress of molecular biology. Some of the most notable milestones were the interpretation of the nucleotide code, semi-conservative DNA replication, the molecular mechanism of DNA transcription, the elucidation of the life cycle of lytic and lysogenic bacterial viruses, the concept of gene regulatory mechanisms, and the discovery of restriction enzymes. Restriction enzymes can identify specific sequences of DNA that can then break the bonds at these sites.^{16–21}

These unicellular organisms can exist in multicellular developmental states in many cases or undergo developmental progression and are thus powerful for looking at G2P in simpler, yet often biomedically relevant contexts, including:

- A. Developing new treatments for bacterial or fungal infections, researchers can identify potential targets for new antibiotics and other treatments. This can help to combat the growing problem of antibiotic resistance, which occurs when bacteria or fungi evolve to become resistant to existing drugs.
- B. Improving agricultural and industrial processes: Bacteria and fungi are used in a variety of industrial and agricultural settings, including the production of food, fuel, and other products. By understanding the genetics of these bacteria, researchers can improve their performance and efficiency.
- C. Understanding the role of organisms capable of multicellularity in environmental processes: Bacteria and fungi play important roles in many natural processes, including the cycling of nutrients in ecosystems and the decomposition of organic matter.

Researchers can better understand their role in these processes and how they may be affected by environmental changes.

D. PITFALLS OF GENOTYPE-TO-PHENOTYPE APPROACHES

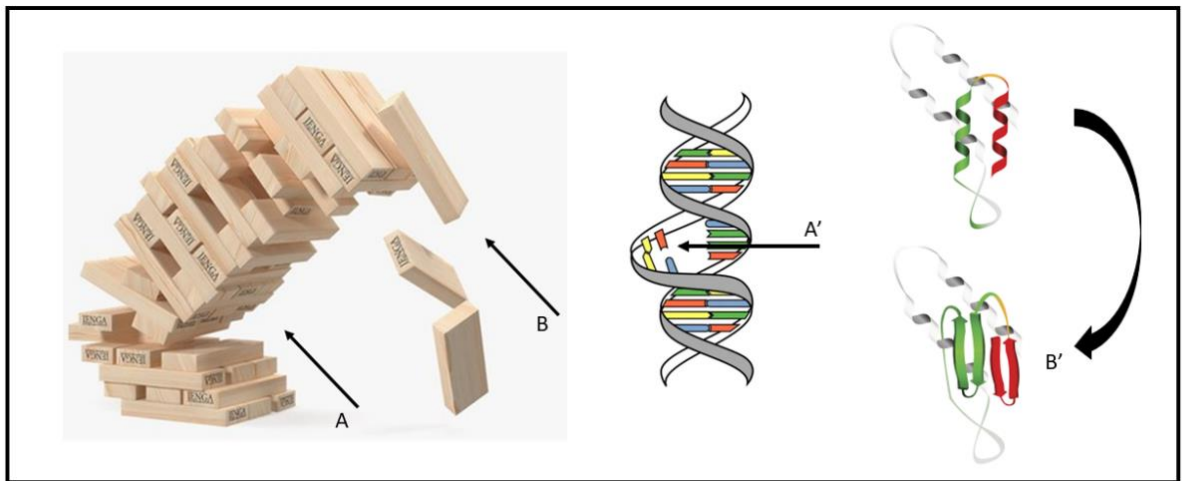


Figure 1: Inferring Genetic Information from Mutations

The investigation of how genetic information translates into observable traits, or the genotype-to-phenotype relationship is often complex. To illustrate, consider removing a block from a tower. Figure 1; A) If there is enough damage to support blocks, the tower will B) collapse. However, the tower's structural redundancy usually requires multiple blocks to be removed. Similarly, some genes can have redundant functions that compensate for the loss of functions in a biological system, which permits the occasional mutation without compromising its overall function.^{22,23} A') As these mutations build in the genome, genetic drift, or the increase or decrease of different alleles in populations based on random inheritance when

there is little or no selection on them can occur, which can then B') allow for the emergence of new emergent behaviors or phenotypes.

Gene duplications can also have effects on whether mutations in genes will be acted on by natural selection. When a gene is duplicated or two or more genes unrelated by evolutionary descent perform similar functions and the inactivation of one has little or no impact on the biological phenotype, this phenomenon is known as genetic redundancy. Such redundancy is common in genes that arise from duplication events, and while it may seem unnecessary, it serves several advantages.²⁴ Firstly, it allows organisms to avoid accumulating lethal mutations in a single gene by creating duplicates of the genetic information. Secondly, it leads to the diversification of protein functions, which gives rise to specialized evolutionary functions.²⁵ However, this genetic compensation mechanism makes it difficult to distinguish the individual evolutionary functions of these genes. While one solution is to knock out both genes of interest, this still fails to provide information on the impact of a single gene. Therefore, it is necessary to develop new metrics that can distinguish between otherwise indiscernible phenotypes to provide a more sensitive genetic annotation.¹¹

Another problem when inferring phenotypes is variation in gene expression; numerous stochastic factors could influence emerging phenotypes by altering gene expression, and the precise number of these factors is unknown.²⁶ Muller's morphs are a set of definitions used to describe the diverse types of morphological variation that can arise within a population. These definitions were proposed by the evolutionary biologist J.B.S. Haldane and were later elaborated on by the geneticist Hermann Muller.⁶⁶ Muller's morphs include monomorphic, dimorphic, trimorphic, and polymorphic traits. Monomorphic traits exhibit little or no variation

within a population, while dimorphic traits exhibit variation between two distinct forms.

Trimorphic traits exhibit variation between three distinct forms, and polymorphic traits exhibit continuous variation across a range of forms. The study of Muller's morphs has been essential in understanding the genetic basis of phenotypic variation and the evolution of complex traits.⁶⁷ It has also been used to study the role of gene-environment interactions in shaping phenotypic variation and the maintenance of genetic diversity within populations.⁶⁸

Muller's work on gene mutations further provided insight into the effects of genetic variations on phenotypes, which has been crucial for understanding the molecular basis of development and disease. Muller identified five categories of gene mutations: amorphs (null mutations), hypomorphs (partial loss-of-function mutations), hypermorphs (enhanced activity mutations), antimorphs (dominant negative mutations), and neomorphs (new function mutations). These diverse types of mutations can be caused by various genetic changes, including point mutations, deletions, duplications, and insertions. To determine the type of mutation, different genetic tests can be performed, such as complementation tests, gene dosage tests, and biochemical assays. Complementation tests can determine if two mutations affecting the same gene complement each other or not, while gene dosage tests can assess the effect of various levels of gene expression. Biochemical assays can reveal the activity and function of the protein product. By understanding the type of mutation and its effects on gene function, researchers can better predict how genetic variations may affect a population's phenotype. Muller's work has been instrumental in establishing the foundation of modern genetics and has paved the way for further understanding of the genetic basis of disease.^{76,77}

A genetically homogeneous population of cells may have different degrees of nutrient exposure, oxygen parameters, and thermal fluctuations as a consequence of the environment, which could result in variations in gene expression within the cells.^{27,28} Stochastic effects are now known to influence both individual cells and populations when determining new genotype-to-phenotype relationships. The methodology of introducing mutations to an organism can also be biased or have indirect effects. One example of gene regulation is co-expressing genes that are in operons from the promoter sequences. Operons comprise a group of genes that are transcribed and translated together to form multiple proteins from a single mRNA.²⁹ However, in some cases, a single mutated gene can affect the entire operon, complicating direct genotype-to-phenotype interpretations. Disrupting gene expression timing can result in a diverse range of phenotypes and have detrimental effects on cell integrity and lead to lethality. It is important not only for the correct genetic information to be expressed but also for the information to be delivered on time, as mistimed information can lead to early or late execution of molecular mechanisms, metabolic or structural costs, and cell death.³⁰

1.2 MYXOCOCCUS XANTHUS

A. MULTICELLULARITY, SELF-ORGANIZATION, AND EMERGENT BEHAVIORS

Analogous to tissues present in multicellular models in higher organisms, cells in a biofilm coordinate using intracellular and extracellular signals to direct a group response that typically benefits the swarm as a group.³⁵ When the biofilm population detects a sufficient lack of nutrients, internal gene expression changes cause cellular changes in the transcriptome.⁴³ Initially, a cell will have starvation-induced intracellular signals that direct changes in gene

expression. Alarmones known as (p)ppGpp affect the transcriptome by binding to RNA polymerase, reducing intracellular GTP levels, and priming the production of extracellular signals. A stringent response occurs, which typically inhibits transcription of rRNA operons and upregulates genes necessary for multicellularity and downregulates those necessary to the vegetative lifecycle.⁴⁷

By studying responses to the phenotype via mutation in the genotype, we can investigate genetic multi-step processes in structural organization and biofilm stress responses.³⁸ Using genomic methodologies, we now understand that bacteria within a biofilm regulate and metabolize a matrix of extracellular polymeric substances (EPS), which can further help to protect them from environmental stressors, such as temperature fluctuations and predation.^{40,44} Common medicinal examples include *Lactobacillus*, which ferments glucose and forms plaque on teeth, and *Staphylococcus*, which persists on medical instruments that lead to nosocomial infection.^{32,33,36} In addition, some microorganisms within biofilms can regulate their transcriptome (inherent gene expression pattern) to enter a dormant or dormant-like state, which helps them to survive during times of stress or when nutrients are scarce; under stress from exposure to antibiotics³⁷, or enable hedge-betting, diversification of resources to different cell fates that are determined through stochasticity.

M. xanthus is a gram-negative, soil-dwelling bacterium found in various environments, including soil, water, and the surface of plants. It is known for its complex life cycle and is used as a model organism to study multicellularity.^{39,46} The golden delta proteobacteria, including *M. xanthus*, are found in temperate soil environments and, like *D. discoideum*, are capable of multicellularity through a developmental mechanism triggered by starvation. In its vegetative

state, millions of rod-shaped cells exist as a motile swarm capable of secreting lytic enzymes. When in contact with prey, *M. xanthus* cells penetrate the prey colony and secrete lytic enzymes, surviving as a dynamic biofilm that can form emergent behaviors such as rippling. When starved, the population of cells can initiate a developmental program using a cascade of genetic signals to exchange extracellular and intracellular signals to form aggregates when nutrients are no longer present. As the local population density of *M. xanthus* increases, the bacteria form dense, multicellular aggregates called fruiting bodies filled with differentiated and environmentally resistant myxospores. An increase in cellular motility during this developmental response leads to the cells encountering one another and emergent population behaviors can appear that are phenotypically distinguishable when viewed using microscopy techniques.⁴⁸

Myxospores ensure survival during starvation or dehydration and germinate when sufficient nutrients are detected, allowing a return to the vegetative state. Within these mounds, the cells mature and differentiate into myxospores, metabolically dormant spores that can germinate when nutrients return. After maturing over the course of twenty-four to seventy-two hours, these cells are physically and thermally more resistant than their vegetative state and are a survival mechanism for high-stress-low nutrient environments. Interestingly, it is also possible to chemically induce sporulation in *M. xanthus* by submerging the cells in a glycerol solution. However, the resulting spores have a different physical structure and express different genomic pathways to progress through sporulation.⁴⁶

B. THE *M. XANTHUS* GENOME

The growth and development of an *M. xanthus* population are regulated by the integration of signals from outside the cells and physiological signals from within.⁴¹ The species has a large genome, 9,139,763 bp (GenBank accession no. CP000113), that has evolved with its many (or several) complex phenotypes. A sizable number of gene duplication events have allowed for the emergence of sensory systems for complex environmental perception relative to other myxobacteria. Divergence of function includes gene regulatory families like enhancer-binding proteins (EBPs), two-component systems (TCS), and sigma factors (σ^{70}) which can promote evolutionary divergence.⁴⁹

Subsequently, various efforts have been made to annotate the genome and understand the molecular basis of the organism's behavior. One study used transcriptomics and proteomics to identify and categorize genes involved in fruiting body development in *M. xanthus*. The authors identified numerous genes involved in diverse processes, including signal transduction, cytoskeleton organization, and metabolism, suggesting that complex regulation and coordination are necessary for fruiting body development.⁵⁰ Another study used transcriptomics to examine the regulatory network underlying social motility in *M. xanthus*. The authors found that motility is regulated by multiple transcriptional regulators, including members of the luxR family and the response regulator protein A.⁵¹ These studies highlight the importance of genome annotation in understanding the complex behavior of *M. xanthus* and demonstrate the potential for transcriptomics and proteomics to provide insights into the underlying mechanisms of complex behaviors in bacteria.

The molecular techniques for studying facets of growth, motility, sporulation, and manipulating the genome of *M. xanthus* have been well documented.^{38,45,46,48} *M. xanthus* is capable of dense growth in liquid cultures with high nutrients, which can be quantified by spectrophotometry for use in molecular studies of vegetative growth. When observing motility during sporulation, which is a conserved survival mechanism that can be observed across other phylogenies, the intricate emergent changes in phenotype offer a more sensitive approach to genetic manipulation. For example, *M. xanthus* has two distinct motility systems that allow it to travel across surfaces: A-motility (Gliding or Adventurous motility) is based on the propulsion of a single cell using exopolysaccharides, or S-motility (Social motility), that is thought to be like a form of twitching motility using pili and fibrils from multiple cells to move as a swarm.³⁹ These motility mechanisms offer diverse survival advantages based on whether the surface is hard and stiff or soft and malleable. Notably, only S-motility is required for wild-type fruiting body morphogenesis, which was observed in an A-motility mutant via glycerol differentiation. These emergent behaviors from understanding the genome of *M. xanthus* are reasons for it to be an attractive model for understanding drivers of genome evolution and modeling.^{42,49}

1.3 *M. XANTHUS* EMERGENT BEHAVIORS AND THEIR USE IN GENOTYPE-TO-PHENOTYPE STUDIES

A. TRANSPOSON LIBRARY MAPPING

To address the genotype-to-phenotype problem in biofilms, researchers have developed comparative models for assessing the genetic mechanisms leading to different phenotypes for the same mutation in different genetic backgrounds. These models can help in understanding

how genes and genetic variation contribute to biofilm phenotype and identifying genes that regulate biofilm development. To understand the regulatory mechanisms of fruiting body morphogenesis, a genomic approach using mutagenesis transposons was implemented in 1986 by Dale Kaiser's lab to map out key genes within the genome of *M. xanthus*.⁵⁴ The Kaiser lab used transposon mutagenesis to randomly insert transposons into the genome of *M. xanthus*. They then screened the mutants for defects in developmental processes such as fruiting body formation and spore production. By analyzing the location of the transposons in the genome, the researchers were able to identify the genes disrupted by the transposons and implicated in development. This approach, known as transposon mapping, allowed the Kaiser lab to systematically identify key developmental genes in *M. xanthus*. Separately, the Beta-gal/lacZ reporter methods showed when the expression of a gene was turned on. Using this method, the group identified thirty-six key genes necessary for the regulation of development. However, when these identified genes were mutated, only seven of the thirty-six strains displayed abnormal phenotypic defects in sporulation or aggregation characteristics. These results suggested that the other twenty-nine genes were developmentally regulated but not essential for development in *M. xanthus*, thereby suggesting developmental redundancy.⁵⁴

B. UNDERSTANDING THE ROLE OF REGULATORY GENES IN MORPHOGENESIS

To elucidate the function of the twenty-nine other genes, it was important to produce another methodology that was capable of either determining the gene regulatory networks of each gene from a molecular approach or ascertaining phenotypic differences that were previously indistinguishable from a systems biology approach. EBPs, enhancer-binding proteins,

were discovered as a novel mechanism of gene regulation for multicellularity in *M. xanthus* using the molecular approach.⁵⁵ Enhancer binding proteins (EBPs) play a critical role in the development of *M. xanthus*. EBPs are transcriptional regulators that bind to specific DNA sequences, called enhancers, to control gene expression. In *M. xanthus*, EBPs help to coordinate gene expression and regulation in response to environmental signals, such as changes in nutrient availability or population density. Studies have shown that EBPs play a key role in regulating several aspects of *M. xanthus* development, including self-organization and the formation of fruiting bodies.^{45,53} For example, members of the LuxR family of EBPs have been shown to control gene expression in response to changes in population density, leading to the coordination of multicellular behaviors such as social motility and fruiting body development. EBPs also regulate gene expression during developmental transitions, such as the switch from vegetative growth to fruiting body development. This is accomplished through the binding of EBPs to enhancers located upstream of target genes, leading to changes in gene expression that are necessary for developmental progression and sporulation.^{56,57} Overall, the role of EBPs in *M. xanthus* development is to coordinate gene expression and regulation in response to environmental signals, leading to complex and coordinated multicellular behavior and highlighting a novel signaling pathway that had not been recognized previously.

C. SOCIAL MICROBES AS MODELS OF MULTICELLULARITY

The ease of manipulating social microbes in the laboratory makes them ideal subjects for genetic, biochemical, and physiological studies. *M. xanthus* has several key features that make it a valuable subject of study. First, it undergoes a clear and well-defined developmental

process from individual cells to complex fruiting bodies, making it an ideal system for studying the transition from unicellular to multicellularity. Second, the genome of *M. xanthus* has been sequenced and well-characterized, making it easy to manipulate its genes and study the effects on development and behavior. Third, *M. xanthus* is a social bacterium that can coordinate its behavior and form complex structures through cell-cell communication and cooperation, making it a valuable model for studying the evolution of multicellularity and the development of social behavior in organisms.⁷³

Myxobacteria can also be model organisms for studying canalization because they exhibit complex social behaviors, including the ability to form multicellular fruiting bodies and engage in cooperative predation.⁷¹ The genetic and regulatory basis of canalization in Myxobacteria appears to be tightly controlled, enabling the production of reproducible patterns of development, even in the face of environmental variation and stochasticity. Thus, understanding the genetic and regulatory basis is crucial for studying emergent behaviors in complex systems. Emergent behaviors arise from the interactions of multiple components in a system, rather than being directly encoded in the individual components themselves. Myxobacteria provide an excellent model system for studying emergent behaviors because their social behaviors require the coordination of many individual cells, each with its genetic program. By studying the genetic basis of canalization in Myxobacteria, we can gain insight into the regulatory mechanisms that enable emergent behaviors and the evolution of complex systems. This knowledge has important implications for the study of complex biological systems, as it can help predict how changes to individual components may affect the emergent behavior of the system.⁷¹⁻⁷²

In conclusion, *M. xanthus* stands out as a valuable model organism for investigating variation in multicellularity within biofilms and the diverse phenotypes that can emerge, owing to its remarkable features and ease of genetic manipulation and behavioral observation. Nonetheless, the task of annotating genes that do not display severe difference phenotypes when mutated may pose challenges when studying the organism's genome through a single-gene mutation approach, due to genetic redundancy. Moreover, the phenotypic variation observed in wild-type *M. xanthus* during macroscopic development is inherently noisy and can mask uncharacterized phenotypes. Expanding on methods to profile genomes, phenomes, and transcriptomes is a vital process regardless of the system of study; a systems biology approach when the reductionist biology approach may not work. As a result, a more refined and systematic phenotyping approach that exploits population dynamics to visualize differences in genotypes was adopted, and this approach will be elaborated upon in the forthcoming two chapters of this thesis.

D. REFERENCES:

1. Li, G.-W. & Xie, X. S. Central dogma at the single-molecule level in living cells. *Nature* **475**, 308–315 (2011).
2. Franklin, S. & Vondriska, T. M. Genomes, Proteomes, and the Central Dogma. *Circ Cardiovasc Genet* **4**, 576–576 (2011).
3. CRICK, F. H. C., BARNETT, L., BRENNER, S. & WATTS-TOBIN, R. J. General Nature of the Genetic Code for Proteins. *Nature* **192**, 1227–1232 (1961).
4. Stearns, S. C., & H. R. F. *Evolution: An Introduction (2nd ed.)*. Oxford University Press. (2005).
5. Fusco, G. & Minelli, A. Phenotypic plasticity in development and evolution: facts and concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 547–556 (2010).
6. Pigliucci, M. Genotype–phenotype mapping and the end of the ‘genes as blueprint’ metaphor. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 557–566 (2010).
7. Dean, A. M. & Thornton, J. W. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* **8**, 675–688 (2007).
8. Shakhnovich, B. E., Deeds, E., Delisi, C. & Shakhnovich, E. Protein structure and evolutionary history determine sequence space topology. *Genome Res* **15**, 385–392 (2005).
9. Zheng, S., Kim, H. & Verhaak, R. G. W. Silent Mutations Make Some Noise. *Cell* **156**, 1129–1131 (2014).

10. Zhang, Z., Miteva, M. A., Wang, L. & Alexov, E. Analyzing Effects of Naturally Occurring Missense Mutations. *Comput Math Methods Med* **2012**, 1–15 (2012).
11. Simon, M. M. *et al.* Current strategies for mutation detection in phenotype-driven screens utilising next generation sequencing. *Mammalian Genome* **26**, 486–500 (2015).
12. Sommer, R. J. Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. *Genetics* **215**, 1–13 (2020).
13. Hunter, P. The paradox of model organisms. *EMBO Rep* **9**, 717–720 (2008).
14. Bozzaro, S. The Model Organism *Dictyostelium discoideum*. in 17–37 (2013). doi:10.1007/978-1-62703-302-2_2.
15. Martín-González, J., Montero-Bullón, J. & Lacal, J. *Dictyostelium discoideum* as a non-mammalian biomedical model. *Microb Biotechnol* **14**, 111–125 (2021).
16. Linn, S. & Arber, W. Host specificity of DNA produced by *Escherichia coli*, X. In vitro restriction of phage fd replicative form. *Proceedings of the National Academy of Sciences* **59**, 1300–1306 (1968).
17. Jacob, F., Perrin, D., Sánchez, C. & Monod, J. L'opéron: groupe de gènes à expression coordonnée par un opérateur [C. R. Acad. Sci. Paris 250 (1960) 1727–1729]. *C R Biol* **328**, 514–520 (2005).
18. Lwoff, A. LYSOGENY. *Bacteriol Rev* **17**, 269–337 (1953).
19. Ellis, E. L. & Delbrück, M. THE GROWTH OF BACTERIOPHAGE. *Journal of General Physiology* **22**, 365–384 (1939).

20. Stevens, A. Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* B. *Biochem Biophys Res Commun* **3**, 92–96 (1960).
21. LEHMAN, I. R., BESSMAN, M. J., SIMMS, E. S. & KORNBERG, A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J Biol Chem* **233**, 163–70 (1958).
22. Rodrigo, G. & Poyatos, J. F. Genetic Redundancies Enhance Information Transfer in Noisy Regulatory Circuits. *PLoS Comput Biol* **12**, e1005156 (2016).
23. Krakauer, D. C. & Plotkin, J. B. Redundancy, antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences* **99**, 1405–1409 (2002).
24. Thomas, J. H. Thinking about genetic redundancy. *Trends in Genetics* **9**, 395–399 (1993).
25. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy. *Nature* **388**, 167–171 (1997).
26. Spudich, J. L. & Koshland, D. E. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
27. Thattai, M. & van Oudenaarden, A. Stochastic Gene Expression in Fluctuating Environments. *Genetics* **167**, 523–530 (2004).
28. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451–464 (2005).
29. Osbourn, A. E. & Field, B. Operons. *Cellular and Molecular Life Sciences* **66**, 3755–3775 (2009).

30. Aymoz, D. *et al.* Timing of gene expression in a cell-fate decision system. *Mol Syst Biol* **14**, (2018).
31. Berg, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* **71**, 587–603 (1978).
32. Boles BR, Horswill AR. Staphylococcal biofilm disassembly. *Trends in Microbiology*. 2011;19(9):449-455.
33. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: a common cause of persistent infections. *Science*. 1999;284(5418):1318-1322.
34. Flemming H-C, Wingender J. The biofilm matrix. *Nature Reviews Microbiology*. 2010;8(9):623-633.
35. Götz F. Staphylococcus and biofilms. *Molecular Microbiology*. 2002;43(6):1367-1378.
36. Høiby N, Bjarnsholt T, Givskov M, Molin S, Ciofu O. Antibiotic resistance of bacterial biofilms. *International Journal of Antimicrobial Agents*. 2010;35(4):322-332.
37. Jakobsen TH, Bjarnsholt T, Jensen PO, Givskov M, Høiby N. Targeting quorum sensing in *Pseudomonas aeruginosa* biofilms: current and emerging inhibitors. *Future Microbiology*. 2013;8(7):901-921.
38. Jelsbak, L. & Sørensen, L. Pattern formation: Fruiting body morphogenesis in *Myxococcus xanthus*. *Curr Opin Microbiol* **3**, 637–642 (2000).
39. Kaiser, D. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proceedings of the National Academy of Sciences* **76**, 5952–5956 (1979).

40. Kaplan JB. Biofilm dispersal: mechanisms, clinical implications, and potential therapeutic uses. *Journal of Dental Research*. 2010;89(3):205-218.
41. Kaplan, H. B., Kuspa, A. & Kaiser, D. Suppressors that permit A-signal-independent developmental gene expression in *Myxococcus xanthus*. *J Bacteriol* **173**, 1460–1470 (1991).
42. Kazazian, H. H. Mobile Elements: Drivers of Genome Evolution. *Science* (1979) **303**, 1626–1632 (2004).
43. Kjelleberg S, Molin S. Is there a role for quorum sensing signals in bacterial biofilms? *Current Opinion in Microbiology*. 2002;5(3):254-258.
44. Ko, M. S. H. A stochastic model for gene induction. *J Theor Biol* **153**, 181–194 (1991).
45. Kroos, L., Kuspa, A. & Kaiser, D. A global analysis of developmentally regulated genes in *Myxococcus xanthus*. *Dev Biol* **117**, 252–266 (1986).
46. Muñoz-Dorado, J., Marcos-Torres, F. J., García-Bravo, E., Moraleda-Muñoz, A. & Pérez, J. Myxobacteria: Moving, killing, feeding, and surviving together. *Front Microbiol* **7**, 1–18 (2016).
47. Irving, S.E., Choudhury, N.R. & Corrigan, R.M. The stringent response and physiological roles of (pp)pGpp in bacteria. *Nat Rev Microbiol* **19**, 256–271 (2021).
<https://doi.org/10.1038/s41579-020-00470-y>
48. Taylor, Rion G, and Roy D Welch. “Recording multicellular behavior in *Myxococcus xanthus* biofilms using time-lapse microcinematography.” *Journal of visualized experiments: JoVE* ,42 2038. 6 Aug. 2010, doi:10.3791/2038

49. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB. Evolution of sensory complexity recorded in a myxobacterial genome. *Proc Natl Acad Sci U S A*. 2006 Oct 10;103(41):15200-5. doi: 10.1073/pnas.0607335103. Epub 2006 Oct 2. Erratum in: *Proc Natl Acad Sci U S A*. 2006 Dec 19;103(51):19605. Eisen, J [corrected to Eisen, J A]. PMID: 17015832; PMCID: PMC1622800.
50. McLoon, Anna L et al. "Transcriptomic analysis of the *Myxococcus xanthus* FruA regulon, and comparative developmental transcriptomic analysis of two fruiting body forming species, *Myxococcus xanthus* and *Myxococcus stipitatus*." *BMC genomics* vol. 22,1 784. 1 Nov. 2021, doi:10.1186/s12864-021-08051-w
51. Lloyd, Daniel G, and David E Whitworth. "The Myxobacterium *Myxococcus xanthus* Can Sense and Respond to the Quorum Signals Secreted by Potential Prey Organisms." *Frontiers in microbiology* vol. 8 439. 14 Mar. 2017, doi:10.3389/fmicb.2017.00439
52. Kaiser, Dale et al. "Myxobacteria, polarity, and multicellular morphogenesis." *Cold Spring Harbor perspectives in Biology* vol. 2,8 (2010): a000380. doi:10.1101/cshperspect.a000380
53. Ma M, Welch RD, Garza AG. The σ_{54} system directly regulates bacterial natural product genes. *Sci Rep*. 2021 Feb 26;11(1):4771. doi: 10.1038/s41598-021-84057-4. PMID: 33637792; PMCID: PMC7910581.
54. Kroos, L, and D Kaiser. "Construction of Tn5 lac, a transposon that fuses lacZ expression to exogenous promoters, and its introduction into *Myxococcus xanthus*." *Proceedings of the*

National Academy of Sciences of the United States of America vol. 81,18 (1984): 5816-20.

doi:10.1073/pnas.81.18.5816

55. Sarwar Z, Garza AG. Two-Component Signal Transduction Systems That Regulate the Temporal and Spatial Expression of *Myxococcus xanthus* Sporulation Genes. *J Bacteriol.* 2015;198(3):377-385. Published 2015 Sep 14. doi:10.1128/JB.00474-15
56. Giglio KM, Zhu C, Klunder C, Kummer S, Garza AG. The enhancer binding protein Nla6 regulates developmental genes that are important for *Myxococcus xanthus* sporulation. *J Bacteriol.* 2015;197(7):1276-1287. doi:10.1128/JB.02408-14
57. Ma M, Garza AG, Lemon DJ, et al. Identifying the Gene Regulatory Network of the Starvation-Induced Transcriptional Activator Nla28. *J Bacteriol.* 2022;204(12):e0026522. doi:10.1128/jb.00265-22
58. Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro AN, Iversen ES, Couch FJ, Goldgar DE. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am J Hum Genet.* 2007 Nov;81(5):873-83. doi: 10.1086/521032. Epub 2007 Sep 6. PMID: 17924331; PMCID: PMC2265654.
59. Wray, N.R., Ripke, S., Mattheisen, M. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 50, 668–681 (2018). <https://doi.org/10.1038/s41588-018-0090-3>

60. Manolio, T., Collins, F., Cox, N. et al. Finding the missing heritability of complex diseases. *Nature* 461, 747–753 (2009). <https://doi.org/10.1038/nature08494>
61. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet.* 2012 Jan 13;90(1):7-24. doi: 10.1016/j.ajhg.2011.11.029. PMID: 22243964; PMCID: PMC3257326.
62. Perfect JR, Casadevall A. Cryptococcosis. *Infect Dis Clin North Am.* 2002 Dec;16(4):837-74, v-vi. doi: 10.1016/s0891-5520(02)00036-3. PMID: 12512184.
63. Idnurm, Alexander, et al. "Deciphering the model pathogenic fungus *Cryptococcus neoformans*." *Nature Reviews Microbiology* 3.10 (2005): 753-764.
64. Watnick, Paula, and Roberto Kolter. "Biofilm, city of microbes." *Journal of Bacteriology* 182.10 (2000): 2675-2679.
65. Reidl, Joachim, and Karl E. Klose. "Vibrio cholerae and cholera: out of the water and into the host." *FEMS microbiology reviews* 26.2 (2002): 125-139.
66. Muller, H. J. 1932. Further studies on the nature and causes of gene mutations. *Proceedings of the 6th International Congress of Genetics*, pp
67. (2003). *Developmental Plasticity and Evolution*. *Nature*. xx.
68. Mitchell-Olds T, Willis JH, Goldstein DB. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat Rev Genet.* 2007 Nov;8(11):845-56. doi: 10.1038/nrg2207. PMID: 17943192.

69. Higgs, P. I., & Cho, K. H. (2012). Myxobacteria: models for social development. *Annual review of microbiology*, 66, 1-23.
70. Hayes, E. S., & Velicer, G. J. (2019). Nontransitive competitive interactions mediated through allelopathic effects in Myxobacteria. *Journal of bacteriology*, 201(20), e00381-19.
71. Kaiser, D. (2004). Signaling in myxobacteria. *Annual review of microbiology*, 58, 75-98.
72. Nariya, H., & Inouye, S. (2008). A signaling protein of the aspartate chemotaxis system is responsible for the induction of natural competence in *Myxococcus xanthus*. *Journal of bacteriology*, 190(7), 2874-2878.
73. Müller, S., Strack, S. N., Hoefler, B. C., Straight, P. D., & Kearns, D. B. (2014). *Myxococcus xanthus* exhibits cell death with similarities to the developmental apoptosis of vertebrates. *Proceedings of the National Academy of Sciences*, 111(1), 445-450.
74. Botstein, D., & Fink, G. R. (2011). Yeast: an experimental organism for 21st Century biology. *Genetics*, 189(3), 695-704. doi 10.1534/genetics.111.130765;
75. Kaerberlein, M. (2010). Lessons on longevity from budding yeast. *Nature*, 464(7288), 513-519. doi: 10.1038/nature08981)
76. Muller, H. J. (1932). Further Studies on the Nature and Causes of Gene Mutations. *Proceedings of the Sixth International Congress of Genetics*, 213-255.
77. Spradling, A. C., & Rubin, G. M. (1982). Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science*, 218(4570), 341-347. doi 10.1126/science.6289435)

CHAPTER 2: STOCHASTIC BOUNDS IN THE STUDY OF AGGREGATION DYNAMICS PROVIDE A MEANS TO DIFFERENTIATE BETWEEN SOCIAL BACTERIA.

ACKNOWLEDGEMENTS

I would like to acknowledge the collaborative efforts of the Welch and Patteson Lab, presented in this chapter. This research represents a team effort, and I am grateful for the contributions of each, and every member involved.

A. INTRODUCTION

Development is an energetically expensive and complicated part of the life cycle of many organisms. For a genetically specified multicellular phenotype to incarnate (or manifest), the process requires the stepwise self-organization of increasingly ordered states and an active and robust dampening response to disruptive forces, such as environmental stress and mutation^{1,2,3}. The control systems for development involve non-linear redundant branching and intersecting intracellular and intercellular signal transduction pathways that provide spatiotemporal coordination of the transcriptional, translational, and post-translational events required for the organism to manifest^{4,5}. This complicated genotype-to-phenotype problem, abbreviated as G2P, is the broad task of understanding the iterative process of genetic cause and phenotypic effect that eventually results in biological emergence.

We use the Gram-negative delta-proteobacterium *Myxococcus xanthus* as a model organism. In the laboratory on an agar substrate, *M. xanthus* exists as a single-species motile biofilm called a swarm. Under nutrient-rich conditions, a swarm will expand across an agar

surface as its component cells grow and move (swarming). In contrast, under non-nutritive (starvation) conditions, a swarm will not expand, although its component cells are moving around faster within it. Instead, a starving swarm will transform over a period of approximately one day, during which swarm cells organize into a discrete number of mound-shaped aggregates distributed non-randomly across the swarm area, with each aggregate harboring a bolus of thousands of cells. Over the next few days, the cells at the center of each aggregate differentiate into quiescent myxospores, at which point the aggregates are considered to have matured into a fruiting body. The entire process represents a rudimentary but robust form of multicellular development and can therefore be studied as an example of G2P³¹.

Development of a multicellular prokaryote, such as the formation of an *M. xanthus* fruiting body, may be less complicated than the development of a eukaryote, such as the formation of a mouse, fly, or flatworm, but it is still complicated enough to involve hundreds of genes arranged in branching networks of intersecting pathways. Many of the genes known to be involved in these networks and pathways were first identified through mutation and phenotypic characterization³². It is a foundational protocol in developmental biology: first, a wild-type (WT) strain is selected and its development phenotype is characterized, then mutations are introduced into the WT genome to create new (mutant) strains, and their development phenotypes are characterized, then the phenotypes of the mutant strains are compared to WT and, if a strain displays a significant deviation, the gene(s) and other genetic elements (s) affected by the mutation are deemed more likely to be involved in development. This information can then be used for genome annotation and to guide future research.

Permutations of this protocol have been used many times on *M. xanthus* to provide preliminary biological process annotations for a significant percentage of the genes in its genome. These annotations are more impactful when they can be used for comparison and, to be objectively comparable, their underlying phenotypic data must be quantitative. A mutant strain's descriptive (qualitative) phenotype characterization can be used for comparison to WT only if its deviation is catastrophic, in which case mimics being quantitative because the possible outcomes are binary (success/failure). Phenotypic deviations from WT that are less than catastrophic are still potentially quantitative if measurable features (dimensions) can be identified that match the descriptive characterization or at least aspects of it. If the quantification of features is reproducible and statistically significant, they can be used to differentiate mutant strains from WT, even when their descriptive characterizations are identical.

A requisite condition to establish significance when characterizing and quantifying a developmental phenotype is a defined boundary that distinguishes the phenotype of WT from a near-WT mutant. There are at least three confounding factors that make this difficult. First, developing biological systems exhibit inherent phenotypic stochasticity, and efforts at holding genome sequence and experimental conditions constant can only reduce the variation to a non-trivial baseline. Second, biological systems are also phenotypically robust to the impact of mutation because evolution guides the genes, networks, and pathways that control development to incorporate redundancies due, at least in part, to mechanisms like duplication and divergence.⁵⁰ Third, development processes are often difficult to observe, record, and analyze in replicates sufficient to establish significance.

Statistical techniques that are useful in the face of highly stochastic events such as gene expression are a topic of active interdisciplinary research³³, and the inherent ability of living systems to submit to statistical study is an open epistemological question³⁴. From a strictly statistical perspective, however, if a baseline variation can be determined for WT, any differences distinct from that baseline are capable of distinguishing non-WT phenotypes. It is an interesting idea that must be verified through experiments.

In this study, we describe the design, construction, and operation of an experimental setup for observing many instances of *M. xanthus* development and an analysis pipeline that quantifies atypical features of the development phenotype and displays them as a map. We then employ this setup to distinguish between WT *M. xanthus* development and four near-WT mutants.

Figure 1: Stochasticity Is Inherent to Multicellular Behaviors in Social Bacteria.

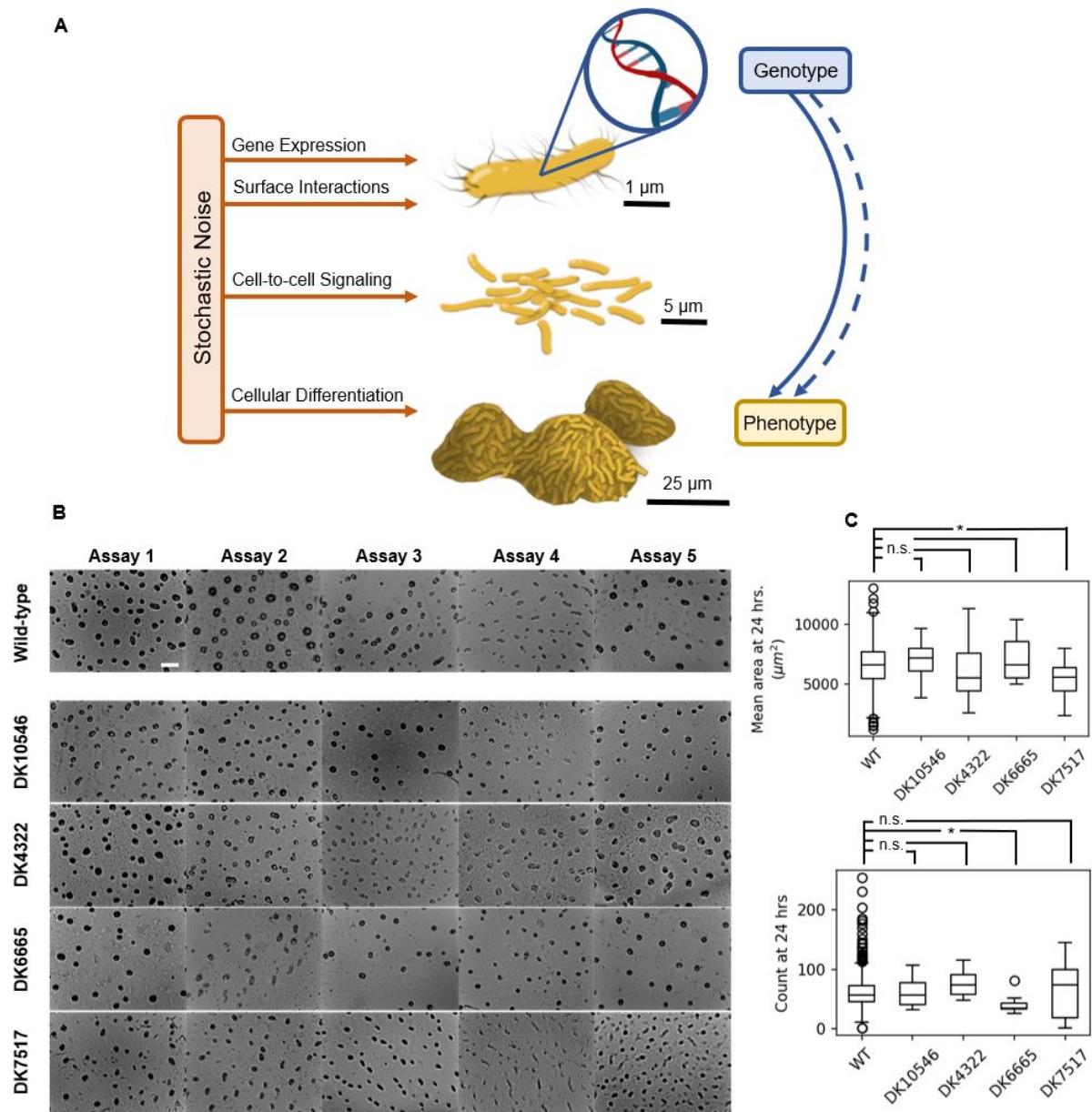


Figure 1: A bacterial colony **(A)** undergoing fruiting body development is exposed to stochastic noise on multiple scales. At the cellular level, gene expression depends on thermally driven

chemical events, and environmental factors such as variations in temperature and humidity introduce further uncertainty. Thus, both direct and indirect effects of genotype arrive at a final phenotype, via multiple developmental paths. **(B)** Images of final developmental phenotype for separate aggregation assays at 24 hours post-inoculation. Pictured is a range of outcomes from the wild-type *M. xanthus* strain as well as the four mutant strains used in this study. Aggregates are visible as dark spots, seen from above. Scale bar, 250 μm . **(C)** The average final area and final count of wild-type aggregates and those for four mutant strains are reported with boxplots. Although there are some differences in these typical metrics of comparison, there is considerable overlap between the wild-type and each of the mutant strains. $N > 500$ measurements for wild type, taken over 25 different days; $N = 15$ measurements for each mutant strain, taken over 2 different days.

B. RESULTS

A time-lapse brightfield micro cinematography assay of *M. xanthus* development is initiated upon settling cells from liquid suspension onto an agar surface. The entire development assay is recorded at 40X magnification and 1 image per minute, spanning at least 24 hours and comprising 1440 frames. Initially, local cell density varies across the field of view, with denser areas appearing darker. Black spots are occasionally visible in the first hour, which are either small tight cell clumps transferred from suspension or motes of dust. Movement becomes evident in subsequent hours as the grey areas move and change shape, and black spots disperse. Over the next several hours, the different grey areas fade, and the entire field of

view becomes a more homogeneous population. The movement then appears to speed up and become more synchronous. Transient darker grey regions begin to emerge across the field of view, and the population appears effervescent. Subsequently, darker, and less transient ridges appear, moving in an amoeboid or sluglike fashion and eventually consolidating into the initial population of aggregates. These initial aggregates are unstable, merging if they collide, while others divide in binary fission. Aggregates undergo coarsening when a significant subset of the aggregates distributed across the field of view shrink and disappear. The remaining aggregates become more stationary and stable. The space between aggregates is teeming with cells and cell movement throughout these aggregation dynamics. Once the stable field of aggregates is established, the cell activity seems to diminish.

Figure 2: High-throughput time series acquisition setup.

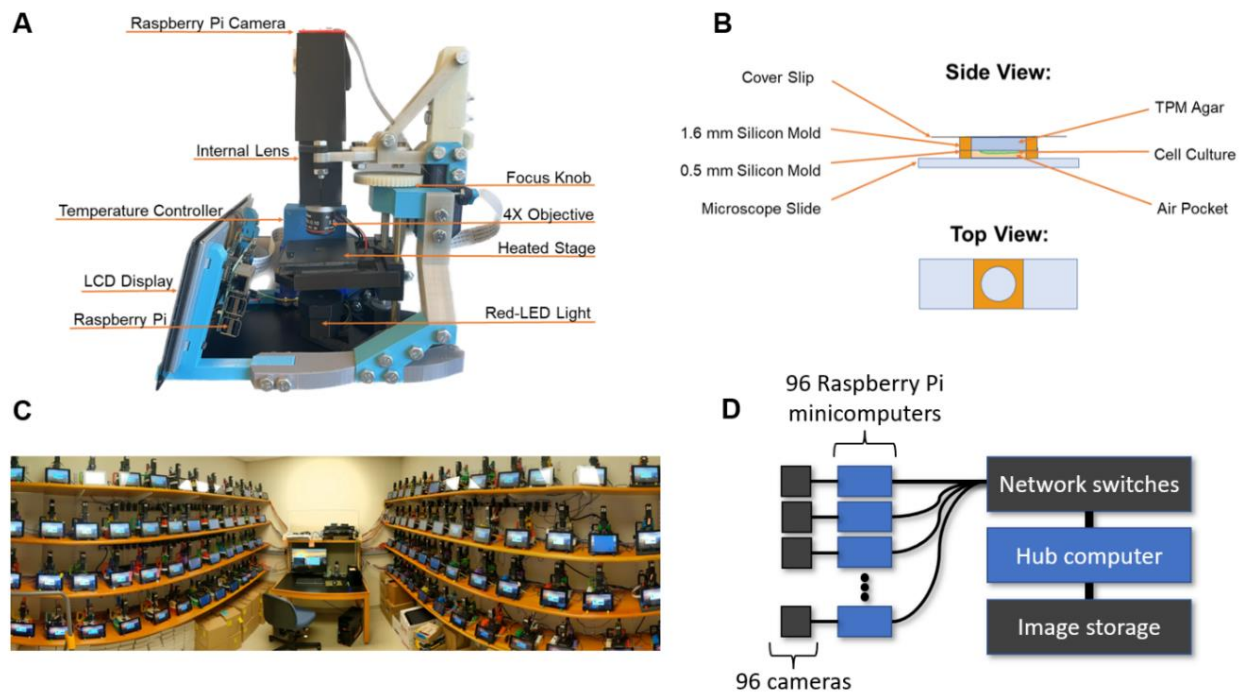


Figure 2: A single microscope **(A)**, at 1.2 kg and 24x19x25 cm³, with a 3D-printed armature, 4X objective lens, light source, heated stage, camera, and Raspberry Pi microcomputer. **(B)** Slide assembly for each developmental experiment. Sandwiched between a glass coverslip and a glass slide, two silicone gaskets create a sealed enclosure containing a disk of non-nutritive agarose on which a colony of *M. xanthus* has been inoculated. Aggregate development is imaged over 24 hours with one image taken each minute. **(C)** Panoramic photograph of full image acquisition setup including 96 microscopes and central hub computer. **(D)** Basic network architecture for centralized image storage and control of all 96 microscopes.

In our study, we observed 500 wild-type movies, with each fruiting body assay requiring a sealed chamber with sufficient temperature, oxygenation, and humidity for development to occur. Cells are inoculated from liquid culture and sealed in each slide assembly, where the aggregates begin to form. We report the spectrum of development behaviors, which manifested to varying degrees. Coarsening varied, as did the length of the slug phase and the occurrence of aggregation. Rippling, previously described, appeared intermittently and could dominate the entire field of view. We also observed pulsing, a distinct wavelike pattern where an aggregate propagates a signal throughout the field of view. Additionally, fruiting body merging, dividing, and other features varied in intensity and duration, even in replicates designed to be biologically and technically perfect (or identical). Although the developmental events followed a linear order, their timing was not synchronized among the movies. These findings suggest that the wild-type developmental process exhibits significant phenotypic variability.

Using previously established and novel metrics to characterize differences in fruiting body morphogenesis metrics, we map and differentiate distributions of fruiting body behavior in DK1622 *M. xanthus* to identify and separate previously indistinguishable phenotypes. Using a dataset of wild-type aggregation time series acquired over 25 separate days, we quantify the range of developmental phenotypes by measuring ten quantitative metrics for each video. We choose three metrics related to timing: start time, when aggregation begins; peak time, when the area occupied by aggregates reaches its maximum; and stability time when the number of aggregates becomes stable. We also measure the mean and standard deviation in the average aggregate area at two-time points: peak time, and 24 hours. We measure the number of identifiable aggregates at both peak time and 24 hours. Finally, we measure the fraction of aggregates that appear and then disperse 24 hours before the inoculation. By establishing metrics for wild-type behavior, our method distinguishes parameters that may be under a genetic influence and identifies new characteristics in phenotype that can be compared.

We next map the wild-type dataset in a visualizable way. Because we use ten phenotypic metrics, each time series may be represented by a point in a 10-dimensional phenotype space, where points closer together are more phenotypically similar than points far apart. To reduce the number of dimensions but retain the structure of our dataset, we use principal component analysis (PCA), to reduce the number of dimensions from ten to two. The resulting metrics from each time series are mapped to a point in a 2D phenotype space. The two dimensions of this space are called PC1 and PC2, the first and second principal components, respectively. PC1 and PC2 are each a single numerical measure that is a mathematical composite of multiple quantitative features, each weighted differently. Principal

component analysis guarantees that PC1 and PC2 are the metrics that display the most variation across the wild-type dataset as compared to any other linearly independent combination of the input metrics. Between just PC1 and PC2, most of the variance across the full dataset (56%) is accounted for. The distribution of points in this map constitutes the wild-type phenotype profile.

Figure 3: Quantitative breadth of wild-type phenotype.

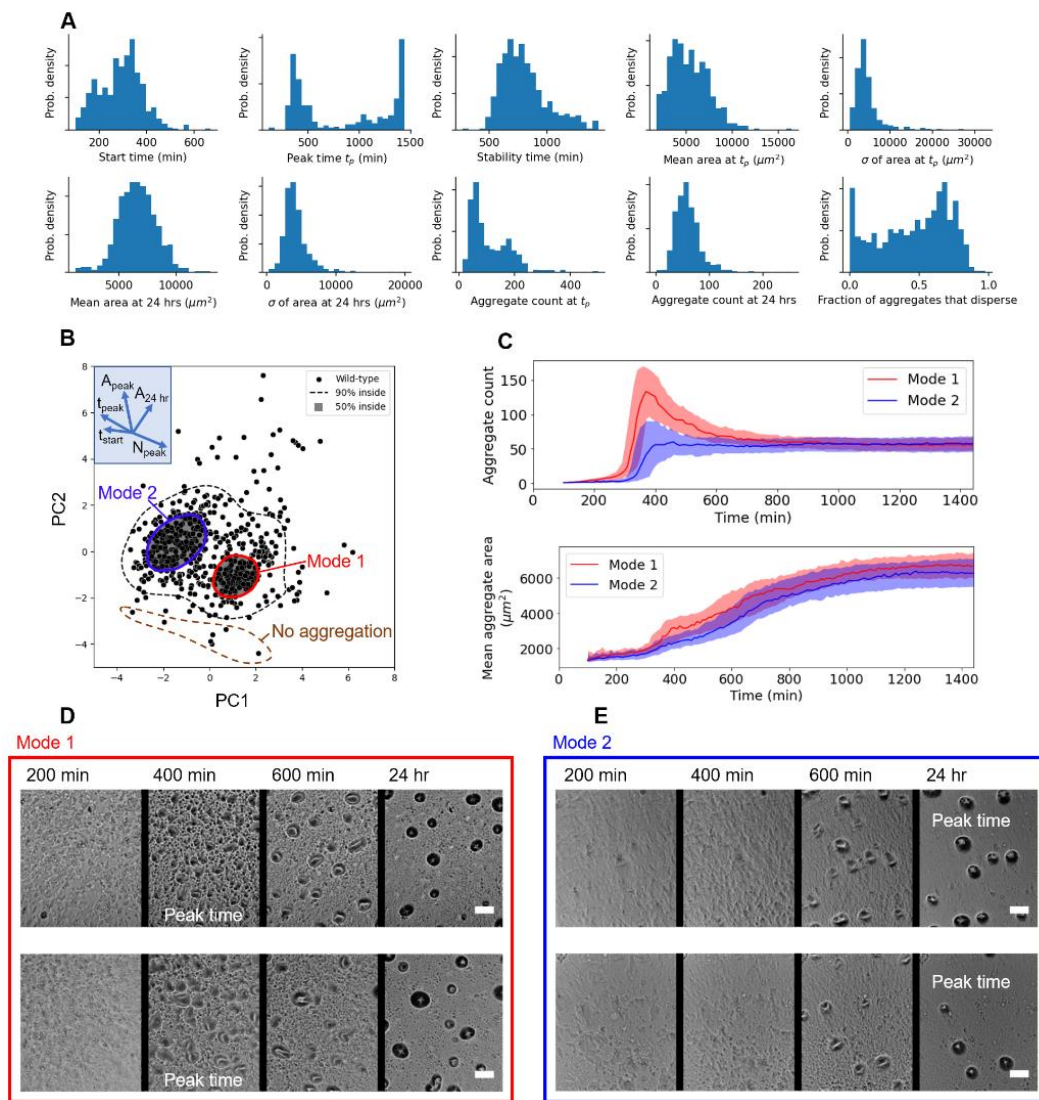


Figure 3: Histograms **(A)** display the range of phenotypic metrics across over 500 wild-type aggregate development time series. Bimodal shapes in peak time (when the total aggregate area is maximum), aggregate count at peak time, and a fraction of aggregates that disperse reflect the two most common groupings of metrics. Long-tailed distributions, such as standard dev. (σ) of the area and aggregate count (both at peak time and after 24 hours) indicate the presence of abnormal phenotypes. All y-axes display probability density. **(B)** By using PCA to combine information from all ten metrics, each wild-type time series is plotted as a single data point in a phenotypic feature space. PC1 primarily measures aggregate area, and PC2 correlates with the number and timing of aggregates. For example, while moving in the direction of the arrow labeled " N_{peak} ," data points will have higher numbers of aggregates at peak time. Units of PC1 and PC2 are arbitrary, although the origin at (0,0) represents average behavior across the full wild-type dataset. A contour is drawn enclosing 90% of the data points, separating typical phenotypes from rare phenotypes. Within typical behavior, two separate clusters, Mode 1, and Mode 2, contain 50% of the wild-type data points. **(C)** Curves displaying the total number of aggregates over time (top) and the mean area of aggregates over time (bottom) illustrate the developmental differences and similarities between the two wild-type modes. The central line represents the median at each time point, and the colored bands span the 25th to 75th percentiles at each time point, i.e., half the data about the median. In Mode 1, a larger number of aggregates develop at an earlier time, most of which disperse. The final number of aggregates is comparable for both modes. The rates of increase in the mean area are also similar across the two modes. **(D&E)** Two representative time series each for Mode 1 and Mode 2 phenotypes at three relevant time points. Mode 1 displays many, dense aggregates that form

early and then disperse. This causes an early peak time. Mode 2 displays aggregates that form later, most of which persist through the 24 hours of development, slowly growing in the area and darkening. This causes a late peak time. Scale bar 100 μm .

The unique weighted combination of metrics that make up PC1 and PC2 indicate key metrics that can distinguish behavior. The weights are bounded between -1 and 1, with larger absolute values indicating more strongly weighted metrics. Both PC1 and PC2 contain a mix of all ten metrics, with no one metric standing out in significance over the others, but rather groups of metrics being more significant. In this study, the top-weighted metrics of PC1 (with weights given in parentheses) are the number of aggregates at peak time (0.47), a fraction of aggregates that disperse (0.44), peak time (-0.43), and start time (-0.39). For PC2, the top metrics are mean area at peak time (0.57), the standard deviation in area at peak time (0.46), the standard deviation in area at 24 hours (0.42), and the mean area at 24 hours (0.40). In summary, PC1 is primarily shared between timing and the total number of fruiting bodies that form, in diametrical opposition. That is when aggregation starts and peaks at an earlier time, the number of fruiting bodies tends to be larger, and vice-versa. PC2 is a variable independent from PC1 that mostly characterizes the area. Thus, large aggregates can present in large or small numbers and do so early or late relative to average wild-type behavior.

We observe two primary modes of aggregate formation, as shown by the two shaded regions in Figure 3B. What we term “Mode 1” features aggregates that start forming and peak in the total aggregate area sooner than other wild-type assays. Mode 1 aggregates are generally numerous, small, and dark at peak time, but a large fraction of them disappear before 24 hours of development. These aggregates tend to be dynamic and lack a well-defined shape

until after peak time (Figure 3D). In contrast, Mode 2 aggregation is less mature early on, with either no visible aggregates or aggregates with fewer layers of cells able to block light (Figure 3E). These aggregates are more static and form with more well-defined shapes, and more of them tend to persist through the 24 hours of development. Because these aggregates tend to persist once they form, the time of peak total area is late for Mode 2, when stable aggregates are still growing slowly. Although there are fewer Mode 2 aggregates at peak time than most wild-type assays, the mean number and size of these aggregates at 24 hours are equal to that of Mode 1, as well as wild-type assays in general. Both modes demonstrate more consistently sized aggregates than other wild-type assays, both at peak time and at 24 hours. Exceptional phenotypes observed in our wild-type dataset include those that produce unusually large fruiting bodies. These occur by a variety of mechanisms, such as large aggregates forming either extremely early with defined shapes from initial formation or extremely late with shapes that only appear visible towards the end of 24 hours. These abnormal behaviors are present at the margins of PCA phenotype space because they represent a confluence of multiple abnormal metrics, revealing more information than standard statistical tests on one metric at a time. Some rare behaviors observed include failure to aggregate, which occurred in about 2% of wild-type assays, and failure for aggregates to stabilize after 24 hours, which occurred in about 17% of wild-type assays.

These metrics are extracted with a custom Python image processing algorithm that identifies and measures each aggregate, as described in Methods. Values for these metrics across the wild-type dataset are shown in Figure 3A, with the distributions illustrating averages and variation for each metric. Peak time, aggregate count at peak time, and the fraction of

aggregates that disperse exhibit bimodal distributions. Long-tailed distributions, such as the standard deviation (σ) of area and aggregate count (both at peak time and after 24 hours) indicate the presence of abnormal phenotypes with extreme values in these metrics.

We chose four mutant strains to compare with our nominal wild-type strain DK1622, each with 60 to 80 replicates collected over two to six separate days. These strains were chosen to be developmentally like wild type to test the sensitivity of our methods. In preliminary experiments, all four strains produced a set of three replicates that were manually identified as “near-wild-type” and displayed a final aggregate size and number that could not be distinguished from wild-type with a student’s t-test. Three mutant strains contain insertions of simple reporter genes, such as DK10546 producing GFP (See Methods for more details on each strain). Analyzing these strains tests the assumption that introducing reporter genes into a prokaryotic genome will not significantly impact cellular behavior or emergent phenotypes, an important preliminary consideration before their use in other experiments. When more replicates had been analyzed, standard statistical tests distinguished one strain, DK7517, as distinct from wild type because it produces smaller than average aggregates (Fig. 1). Because the distribution of wild-type final mean areas is non-Gaussian as measured by a Shapiro-Wilk normality test, the Kolmogorov-Smirnov test for distinguishing two distributions was chosen as the standard test in rather than a student’s t-test, which assumes normality of the underlying distributions.

The developmental data for the additional replicates of the mutant strains are projected onto the same PC1, and PC2 axes that were defined for the wild-type data. This allows direct comparison and visualization of multiple metrics simultaneously. The typical behavior and

variability of each mutant strain's development are captured by two regions: a contour is drawn that captures 50% of the data points, creating an effective median region in PCA space. We then choose a wider contour that captures 90% of data points to serve as a boundary for abnormal phenotypes. By comparing the distribution of the mutant strain points to that of the wild-type, a p-value can be calculated for the null hypothesis that the mutant data points are drawn from the wild-type distribution. This p-value depends on the number of points found inside the 50% contour and the number inside the 90% contour and was calculated with bootstrapping, as described in Methods. This is a nonparametric, data-driven statistical method that makes no assumptions about the dataset *a priori*, allowing for multi-modal distributions which are likely to arise in living systems. Validation can be confirmed on multiple subsamples to quantify the impact of day-to-day variation on the p-value.

Figure 4: Deviation of Near-Wild-Type Mutant Strains from Wild-Type Behavior.

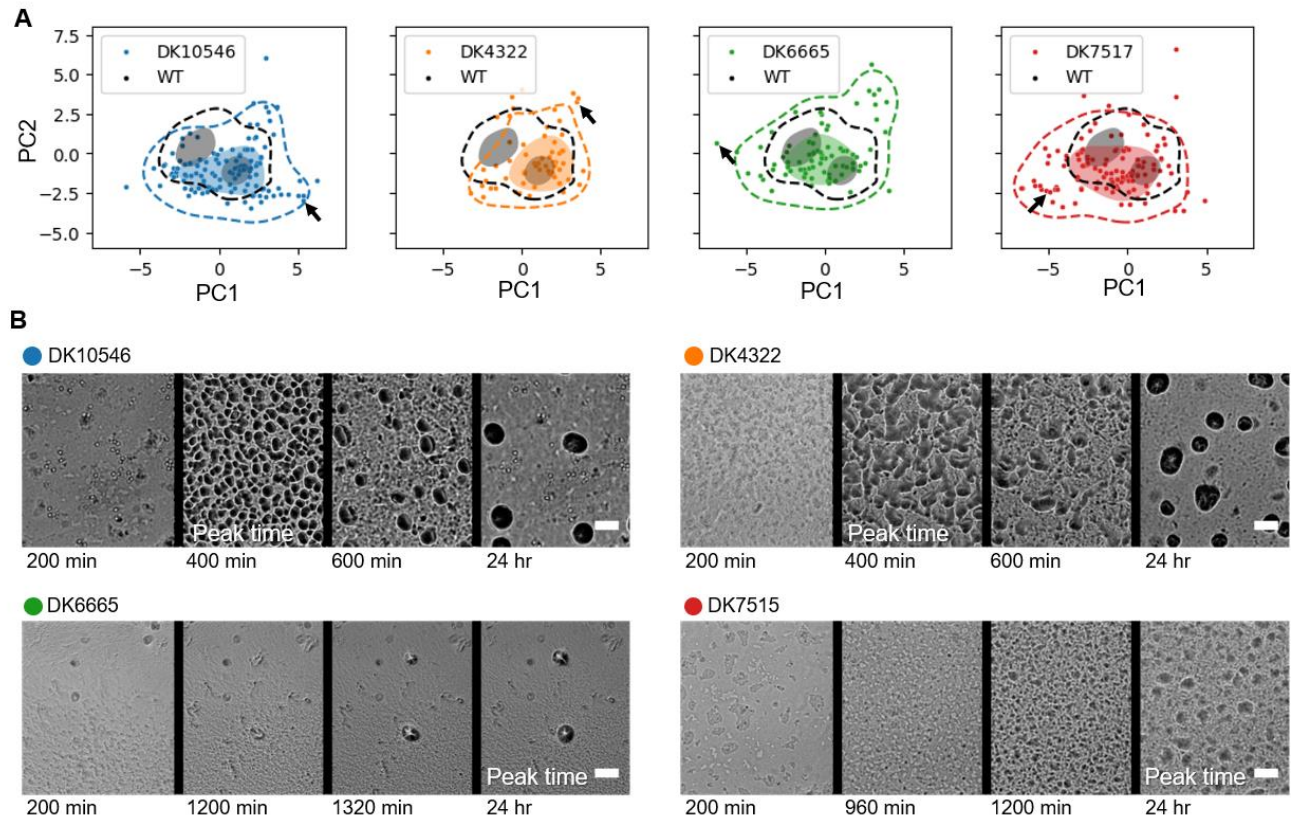


Figure 4: Each mutant development time series (**A**) is plotted as a single data point in phenotype space, as measured by the collective metrics PC1 and PC2. For each respective strain, dashed contours enclose 90% of data points, and the shaded region(s) enclose 50% of data points. The 90% and 50% contours for wild type are shown for reference. The deviation of the mutant phenotype from the wild-type is determined by the departure of the mutant distribution from the wild-type distribution. Statistically significant departures from the wild-type distribution are measured for all four mutant strains, with p-values calculated for subsamples of only 15 replicates each. These p-values are calculated from many random samplings drawn from the wild-type dataset (Methods). Arrows point to the time series shown

in **(B)** Time series of phenotypes expressed rarely in wild type are shown at three relevant time points for each mutant strain. Scale bar 100 μm .

All four strains demonstrated a subtle yet statistically significant departure from wild-type behavior. Strains DK10546 and DK4322 showed a preference for Mode 1 behavior, with Mode 2 being rarely expressed, unlike in wild type. Some mutant replicates that exhibited rare behaviors are highlighted in Fig. 4B. Replicates of DK10546 displayed more extreme versions of Mode 1 behavior, in which many small aggregates form early on, all of which disperse by 24 hours. DK4322 replicates also displayed a more extreme version of Mode 1 behavior in which aggregates at peak time, although distinct, had very irregular shapes. The final aggregates were slightly larger and more varied in area than typical wild-type assays. The aggregates of some DK6665 replicates formed from sparse, small points that formed late and grew steadily over the course of 24 hours. This nucleation was seldom expressed in wild type. This strain also displayed difficulty in the dispersal of the random initial cell clumps that are present at inoculation. In wild type, these initial clumps always disperse, and final aggregate positions do not correlate with these initial clumps. Finally, the abnormal behavior of DK7517 replicates, which involved late aggregates that never significantly darkened, was also a noticeable deviation from even exceptional wild-type behavior. Among the mutant strains, about 2% failed to aggregate (the same fraction as wild-type), and 30% to 45% of mutant assays failed to stabilize after 24 hours, a significant increase from the 17% observed in wild-type.

C. DISCUSSION

Each instance of fruiting body formation is the result of the combined effects of the genetic background, environmental factors both controllable – such as temperature or substrate stiffness – and uncontrollable – such as local pockets of varying initial cell density, or the changes in gene expression that uniquely unfold for that specific population of cells. Genetic changes may be better described by how they affect the odds of a multiplicity of outcomes. We observe that most mutants behave like wild type most of the time. This is a significant reinterpretation of the meaning of the “phenotype” that results from a given genotype, not as a guaranteed outcome, but as a reshuffling of outcomes. This description is appropriate to the physics of living systems, which are tuned through evolution to be poised at the center of a variety of behaviors, ready to adapt to rapid changes either in the organism or its environment. Techniques in the biostatistics community are consistent with this perspective, such as probabilistic latent variable models (Sankaran 2019), which complement the analysis presented in this study.

The statistical method reported here characterizes phenotype in terms of abnormal behavior, either locally – i.e., groupings of behavior that fall within the broad scope of wild-type but are still outside the norm – or globally – i.e., behaviors that are never expressed in the entire wild-type profile. Because measures of mean behavior can fail to capture variations, such as a shifting distribution that happens not to be skewed, the study of abnormalities is a fruitful ground for distinguishing the effects of single-gene mutations, especially when enough replicates can be performed to reliably observe abnormal behavior³⁵.

From our results, we also suggest with confidence a failure rate of the DK1622 genome to not

create mature fruiting bodies using the methods provided here. It is important to distinguish between strains with fruiting bodies that fail to aggregate and germinate spores, relative to those that cannot aggregate yet still manage to sporulate. It had been previously reported that up to 10% of the WT cell lines survive germination past development (Bradley et al., 2016), however, it is unknown to what degree there is a lack of fruiting morphogenesis during these types of sporulation assays. The cause for this abnormal phenotype could be internal mutations in genes or disruptions in signaling pathways, or a physical issue preventing them from assembling (i.e., the density of cells or lack of motility). Using the time-lapse imaging, we show that 2% of the movies captured resulted in dynamics that did not yield aggregation of fruiting bodies, which would account for only a fraction of these instances. During these movies, the *M. xanthus* population was treated under the same conditions yet no fruiting bodies were observed. This does not address what could be happening internally to cells in the remaining instances where fruiting bodies mature under normal dynamics but remain unable to germinate WT colonies post-starvation, however, there is now a quantification of that failure rate in this system which is crucial to mechano-physio modeling. This perspective also emphasizes the need to vet exceptional data to ensure that they represent genuine but rare behavior and are not simply abnormal due to a failure of data processing, such as inconsistent imaging conditions. As far as possible, the metrics chosen in this study were selected to minimize dependence on imaging setup, and manual vetting was performed on points that fell on the margins of PCA phenotype space.

The methods we report in this work are not unique to bacterial development. A similar analysis could be performed for any stochastic system that can: 1) Have many comparable

replicates prepared, 2) Have multiple relevant metrics measured for each replicate. This method can thus be compared to a similar general analysis framework, such as machine learning. Machine learning is powerful in that metrics do not need to be chosen in advance. However, this comes at the cost of the transparency in how categorization is accomplished, and the need for training a model on input data categorized by some other method. It is often true that a system of interest has several obvious aspects that are amenable to measurement with image processing. In the case where the dynamics of the system are relevant, our method is also attractive to use with time series image data, which in raw form can be multiple gigabytes in size and is difficult to work with on large scales. The initial choice of phenotypic metrics represents a large simplification of the unprocessed image data that ensure phenotypic relevance is preserved over image acquisition noise. The further reduction of the phenotypic dataset from ten to two dimensions not only produces a phenotypic map that is sufficiently navigable to reveal the overall structure and guide the investigation of individual data points but also avoids a well-known problem in data science associated with the so-called “curse of dimensionality.” This issue occurs in high-dimensional datasets, where geometry tends to make the distance between neighboring points like the distance across the dataset, making “similarity” in terms of distance essentially meaningless (Beyer 1999).

Although 60 replicates or more were analyzed for each mutant strain in this study, a strain that appears like wild-type to the eye can be distinguished from wild-type with fewer replicates. By analyzing the distribution in PCA space of many subsamples of wild-type aggregation, we find that only 15 replicates spread over two days are needed to establish departure from wild-type behavior for each of the mutant strains above. Notably, for this same

sample size, standard statistical tests based on individual metrics can only distinguish one strain, DK7517, as distinct from wild-type, and with less statistical power than the method used in this study. Because the distribution of final mean areas is non-Gaussian as measured by a Shapiro-Wilk normality test, the Kolmogorov-Smirnov test for distinguishing two distributions was chosen as the standard test rather than a student's t-test, which assumes normality of the underlying distributions.

Our results indicate that there can be indirect effects on fruiting body formation dynamics in *M. xanthus* due to the use of common reporters. Reporter genes are used as effective markers for successful transformation and are used for quantitative assays, which require them to not obstruct or alter the mechanism of study. GFP, a 28kDa green fluorescent protein, allows for the precise visualization of proteins using UV light. Although small enough to diffuse from the cytosol into a nucleus in eukaryotes, there are inherent indirect costs to attaching these tags to a molecule of interest. By inserting this extra DNA, another introduction of molecular noise via transcription and translation steps is added during these biochemical reactions.^{36–38} These non-target effects can cause cellular differences in expression changes which contribute to the stochastic variation we see in the overall cell population³⁹.

Another reporter gene, Tn5 lac, is a promoter-less trp-lac fusion that was designed to identify strains that specifically increase beta-galactosidase expression at some point during *M. xanthus* development as developmental markers. Transposons are a diverse class of mobile genetic elements that can promote genetic rearrangements without a requirement for sequence homology⁴⁰. The Tn5 transposon was inserted so lac Z transcription occurred with exogenous promoters and their promoter strength was quantified⁴¹ to identify genes that were

expressed during *M. xanthus* fruiting body morphogenesis. By attaching to the promoter, it was assumed that lacZ expression would occur in parallel with gene-specific myxospore development without disrupting gene function. However, Tn5 transposon insertions can promote adjacent deletions⁴² which can then disrupt regulatory regions and lead to changes in phenotype from differences in gene expression⁴³. The ability to differentiate these transposon insertion strains from wild type emphasizes the need to assess the impact of reporter genes, especially in biophysical studies that focus on developmental dynamics, where differences may be easier to observe.

These results point to a method of gene annotation that is sufficiently sensitive to identify the impact of single-gene mutations that would otherwise be imperceptible. Each mutant strain becomes associated with a signature distribution in a PCA space that can be defined and used by any laboratory. Strains with sufficiently similar distributions can then be said to share a function because they impact development in a demonstrably comparable way. If the signature distribution of well-understood genes is reported, newly characterized genes of unknown function can be compared to those benchmark distributions. Notably, these signatures are agnostic of any specific biological model and are based only on visually observable characteristics. Future work can process a library of single-gene knockout strains with unknown functions. Over time, this also develops an overall phenome concerning fruiting body development that is quantitative and creates a common language of comparison for the function of many different genes. As more such experiments are done in this framework, either through high-throughput imaging methods like those described here or by the collective efforts of many researchers, developmental phenotypes that are uncommon will be revealed. These

unusual events serve to define a boundary on multicellular behavior and can expose the regulatory mechanisms of fruiting body formation when stretched to their limits by stochastic factors alone. These “exceptions” can teach us much about the “rule.”

Among these sources of behavioral change, we expect that small aspects of the experimental protocol will have subtle but measurable effects. Over the course of the experiments carried out for this work, a new protocol variable was confirmed that is not normally controlled for in *M. xanthus* culture, namely the age of the agar plate containing colonies to be harvested for liquid bacterial culture. It is expected that reintroducing bacteria to liquid culture will “reset” their metabolic state regardless of what state they were in before, but our analysis revealed preliminary evidence that a colony would “remember” the age of the plate it was harvested from and produce fewer fruiting bodies if the colony grew on the plate for at least three days. Although the mechanism of this memory is unknown, the effect has been consistent, and we expect that other such protocol variables exist that have a measurable phenotypic impact.

The sensitivity of the methods presented here may also be used to measure phenotypic response to changes in a variety of environmental variables. Like running replicates of single-gene mutants, running replicates with differing substrates will reveal subtle or overt changes in phenotype. This future work could address the missing environmental information of the genotype-phenotype problem and expand the bounds of “wild-type behavior” as a function of environmental conditions. The contour bounding abnormal wild-type behavior encompasses exactly what has not yet been characterized to have a specific cause, providing both a measure of ignorance of relevant physical and biological mechanisms, and a way to characterize how

much knowledge is gained when subregions can be assigned a root cause, and thus separated from wild-type.

D. CONCLUSIONS

The biological significance of this study highlights how detailed characterization of stochastic variation in phenotype can be used to differentiate otherwise identical appearing phenotypes due to different gene mutations in a social bacterium that relies on collective behavior for survival. The study also suggests that even subtle reporter genes that are widely assumed to be benign and not impacting cellular behavior, can still produce measurable changes in aggregation development. The findings highlight the importance of understanding the dynamics of collective behavior in social organisms like *M. xanthus*, as slight changes in gene expression, environment, and other seemingly minor stochastic factors can affect the behavior of the entire group. This can have implications for understanding the evolution of social behavior and how it is shaped by genetic and environmental factors. Moreover, the study demonstrates the potential of using these methods to characterize development by mapping out phenotype in a visualizable way that incorporates quantitative aspects that can be measured in collective, living systems. This provides a language of data presentation that can be used for gene annotation and investigating the impact of environmental variables on the genotype-phenotype problem.

Overall, this study highlights the importance of understanding the subtle effects of gene expression on social behavior and the potential of using visualizable data presentation methods to investigate the dynamics of collective behavior in living systems.

E. METHODS

Imaging setup: The setup can simultaneously collect time-series images for 96 experiments using an array of compact and identical microscopes controlled by a central computer. Each of the 96 microscopes is equipped with a single 4X objective lens, a Peltier device that maintains stage and sample temperature, a red-light source, and a camera controlled by a Raspberry Pi, a single-board minicomputer. The 3D-printed armature and assembly hardware serve to keep all components firmly in place and provide a focus knob for higher image quality.

To ensure uniform control of all microscopes and central storage of their time series output, each Raspberry Pi unit is networked via ethernet and two 64-port network switches to a central hub computer. This computer runs PIServer software, which boots each Raspberry Pi from a single operating system image, allowing software to be changed and updated for all Raspberry Pi units simultaneously. Custom software written in Python provides a convenient GUI to control image acquisition from each camera via SSH and organize output in a centralized image storage location.

Cell culture: Long-term stock cultures were recovered on nutrient-rich CTTYE media agar (1% Casein Peptone (Remel, San Diego, CA, USA), 0.5% Bacto Yeast Extract (BD Biosciences, Franklin Lakes, NJ, USA), 10 mM Tris (pH 8.0), 1 mM $\text{KH}(\text{H}_2)\text{PO}_4$ (pH 7.6), 8 mM MgSO_4). Cells

were harvested from the plates and used to inoculate broth cultures in CTTYE with vigorous shaking at 32°C and grown to an approximate density of 4×10^8 cells/mL (100 Klett or 0.7 A_{550}). Cells were centrifuged to remove the nutrient broth, washed in TPM buffer (10 mM Tris (pH 7.6), 1 mM $\text{KH}(\text{H}_2)\text{PO}_4$, 8 mM MgSO_4), and resuspended to a final concentration of 4×10^9 cells/mL. For the development assay, approximately 4×10^7 cells (10 μL aliquots) were spotted onto a TPM agar slide, a nutrient-limited medium then incubated on the microscope stage at 32°C for 24 hours. TPM slides were prepared as previously described⁴⁴.

The strains used in this study are as follows:

Strain	Genotype	Phenotype
DK1622	WT	The nominal strain was genetically modified from a naturally occurring <i>M. xanthus</i> isolate ⁴⁵ . This was done to establish a stable baseline for fruiting body development assays, as strains isolated directly from soil have a high rate of developmental failure in a laboratory setting.
DK10546	<i>pilAp</i> -GFP	The constitutive GFP labeled strain is used to track motility and cell dynamics during development. Used as an experimental control for fluorescence microscopy, the construct was generated by fusing a copy of the <i>pilA</i> promoter to the coding sequence of GFP, which was then re-inserted into the <i>M. xanthus</i> chromosome ⁴⁶ . This study

		showed an increased likelihood of early aggregation with many dispersing aggregates over wild type for this strain.
DK6665	Tn5 Ω 6658 <i>sasB</i>	This mutant was generated from a reporter attached to the developmental gene <i>sasB</i> ⁴⁷ . Previous experimental work observed no visible phenotypic impact on mutations to this development gene; the mutant can still proceed through development via other regulatory channels. This study observed initial cell clumps having an unusually high impact on final aggregates due to a lack of dispersal of initial cell clumps relative to wild type.
DK4322	<i>spiA::Tn5/lac</i>	The strain is a reporter fusion for the developmental gene <i>spi</i> with <i>lacZ</i> for β -galactosidase assays. The <i>spi</i> gene is induced at 2 hrs into development and is developmentally regulated by C signal pathways. In previous work, the transposon was characterized as not interfering with development or affecting spore production ⁴⁸ . This study showed a higher likelihood of irregular aggregate shapes at early times for the mutant relative to wild type.
DK7517	TA::Tn5/ <i>lac</i>	Generated via a Tn5-LacZ insertion into a TA synthesis gene, as a reporter gene involved in toxin and antitoxin production also for β -galactosidase assays ⁴⁹ to isolate

		regulatory mutants. This reporter fusion was shown to be expressed during vegetative growth while peaking during the lag phase. In this study, late and immature aggregates were more likely to develop in this strain than wild type.
--	--	--

Image processing pipeline: Phenotype was automatically quantified for each fruiting body aggregation assay in this study by running 144 individuals. TIFF images (ten minutes between each frame over 24 hours of total development) from each time series through a custom Python image processing and analysis pipeline to identify in each frame which pixels could belong to a fruiting body, based on their gray value. The information for the position and geometry of each aggregate was filtered to remove noise and spurious aggregates. This detailed data summary for each time series then had a list of ten specific numbers extracted from it, each of which captures one overall feature, such as the time at which aggregation began or the average size of final fruiting bodies. The values of these ten metrics together were then used in further analysis.

Statistical methods: To calculate p-values that test the null hypothesis of mutant development datapoints in PCA space being drawn from the same distribution as the wild-type development datapoints, we first generate the contours for the wild-type PCA data by starting with Gaussian kernel density estimation (KDE) and using standard root-finding techniques to draw contours from the density estimate that capture 50% and 90% of the PCA datapoints. An appropriate kernel size for the KDE is validated by using 75% of the wild-type dataset, and ensuring that, across many subsamples of the remaining 25% (verification data), the

distribution of enclosed points is centered on the appropriate percentage. When this distribution is skewed, it indicates overfitting of the original contour. With these contours drawn, we then use a data-driven statistical technique like bootstrapping. Given sample size N , 10,000 samples of that size are drawn from the wild-type dataset. Each subsample has a characteristic pair of numbers, (n_{50}, n_{90}) , which corresponds to the number of points in the sample that fall inside the 50% and 90% contours, respectively. Once the distribution of these pairs for wild-type data is known, n_{50} and n_{90} are calculated for a sample of mutant PCA data points of size N . The fraction of wild-type videos that have both n_{50} and n_{90} greater than the mutant sample's values of n_{50} and n_{90} gives the p-value or the probability that a sample of wild-type data of size N would exhibit the same distribution. Contours for mutant strains are shown for visualization only and do not figure into the calculation of the p-values.

F. FUNDING ACKNOWLEDGMENTS

The authors acknowledge funding from NSF MCB 2026747 and NSF DEB 2033942 awarded to A.P. and NSF MCB 1856665 and NSF DMS-NIGMS 1903160 awarded to R.W.

G. REFERENCES:

1. Li, G.-W. & Xie, X. S. Central dogma at the single-molecule level in living cells. *Nature* **475**, 308–315 (2011).
2. Franklin, S. & Vondriska, T. M. Genomes, Proteomes, and the Central Dogma. *Circ Cardiovasc Genet* **4**, 576–576 (2011).
3. CRICK, F. H. C., BARNETT, L., BRENNER, S. & WATTS-TOBIN, R. J. General Nature of the Genetic Code for Proteins. *Nature* **192**, 1227–1232 (1961).
4. Stearns, S. C., & H. R. F. *Evolution: An Introduction (2nd ed.)*. Oxford University Press. (2005).
5. Fusco, G. & Minelli, A. Phenotypic plasticity in development and evolution: facts and concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 547–556 (2010).
6. Pigliucci, M. Genotype–phenotype mapping and the end of the ‘genes as blueprint’ metaphor. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 557–566 (2010).
7. Dean, A. M. & Thornton, J. W. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* **8**, 675–688 (2007).
8. Shakhnovich, B. E., Deeds, E., Delisi, C. & Shakhnovich, E. Protein structure and evolutionary history determine sequence space topology. *Genome Res* **15**, 385–392 (2005).
9. Zheng, S., Kim, H. & Verhaak, R. G. W. Silent Mutations Make Some Noise. *Cell* **156**, 1129–1131 (2014).
10. Zhang, Z., Miteva, M. A., Wang, L. & Alexov, E. Analyzing Effects of Naturally Occurring Missense Mutations. *Comput Math Methods Med* **2012**, 1–15 (2012).

11. Simon, M. M. *et al.* Current strategies for mutation detection in phenotype-driven screens utilising next generation sequencing. *Mammalian Genome* **26**, 486–500 (2015).
12. Sommer, R. J. Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. *Genetics* **215**, 1–13 (2020).
13. Hunter, P. The paradox of model organisms. *EMBO Rep* **9**, 717–720 (2008).
14. Bozzaro, S. The Model Organism *Dictyostelium discoideum*. in 17–37 (2013). doi:10.1007/978-1-62703-302-2_2.
15. Martín-González, J., Montero-Bullón, J. & Lacal, J. *Dictyostelium discoideum* as a non-mammalian biomedical model. *Microb Biotechnol* **14**, 111–125 (2021).
16. Linn, S. & Arber, W. Host specificity of DNA produced by *Escherichia coli*, X. In vitro restriction of phage fd replicative form. *Proceedings of the National Academy of Sciences* **59**, 1300–1306 (1968).
17. Jacob, F., Perrin, D., Sánchez, C. & Monod, J. L'opéron: groupe de gènes à expression coordonnée par un opérateur [C. R. Acad. Sci. Paris 250 (1960) 1727–1729]. *C R Biol* **328**, 514–520 (2005).
18. Lwoff, A. LYSOGENY. *Bacteriol Rev* **17**, 269–337 (1953).
19. Ellis, E. L. & Delbrück, M. THE GROWTH OF BACTERIOPHAGE. *Journal of General Physiology* **22**, 365–384 (1939).
20. Stevens, A. Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* B. *Biochem Biophys Res Commun* **3**, 92–96 (1960).

21. LEHMAN, I. R., BESSMAN, M. J., SIMMS, E. S. & KORNBERG, A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J Biol Chem* **233**, 163–70 (1958).
22. Rodrigo, G. & Poyatos, J. F. Genetic Redundancies Enhance Information Transfer in Noisy Regulatory Circuits. *PLoS Comput Biol* **12**, e1005156 (2016).
23. Krakauer, D. C. & Plotkin, J. B. Redundancy, antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences* **99**, 1405–1409 (2002).
24. Thomas, J. H. Thinking about genetic redundancy. *Trends in Genetics* **9**, 395–399 (1993).
25. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy. *Nature* **388**, 167–171 (1997).
26. Spudich, J. L. & Koshland, D. E. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
27. Thattai, M. & van Oudenaarden, A. Stochastic Gene Expression in Fluctuating Environments. *Genetics* **167**, 523–530 (2004).
28. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451–464 (2005).
29. Osbourn, A. E. & Field, B. Operons. *Cellular and Molecular Life Sciences* **66**, 3755–3775 (2009).
30. Aymoz, D. *et al.* Timing of gene expression in a cell-fate decision system. *Mol Syst Biol* **14**, (2018).
31. Muñoz-Dorado, J., Marcos-Torres, F. J., García-Bravo, E., Moraleda-Muñoz, A. & Pérez, J. Myxobacteria: Moving, killing, feeding, and surviving together. *Front Microbiol* **7**, 1–18 (2016).

32. Jelsbak, L. & S gaard-Andersen, L. Pattern formation: Fruiting body morphogenesis in *Myxococcus xanthus*. *Curr Opin Microbiol* **3**, 637–642 (2000).
33. Vuong, H., Shedden, K., Liu, Y. & Lubman, D. M. Outlier-based differential expression analysis in proteomics studies. *J Proteomics Bioinform* **4**, 116–122 (2011).
34. Dougherty, E. R. & Shmulevich, I. *Send Orders of Reprints at reprints@benthamscience.org On the Limitations of Biological Knowledge. Current Genomics* vol. 13 (2012).
35. Smiti, A. A critical overview of outlier detection methods. *Computer Science Review* vol. 38 Preprint at <https://doi.org/10.1016/j.cosrev.2020.100306> (2020).
36. Singh, U. N. Polyribosomes and unstable messenger RNA: A stochastic model of protein synthesis. *J Theor Biol* **25**, 444–460 (1969).
37. Berg, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* **71**, 587–603 (1978).
38. Ko, M. S. H. A stochastic model for gene induction. *J Theor Biol* **153**, 181–194 (1991).
39. Nunes, F. *et al.* Non-Target Effects of Green Fluorescent Protein (GFP)-Derived Double-Stranded RNA (dsRNA-GFP) Used in Honeybee RNA Interference (RNAi) Assays. *Insects* **4**, 90–103 (2013).
40. Lodge, J. K., Weston-Hafer, K. & Berg, D. E. Transposon Tn5 target specificity: preference for insertion at G/C pairs. *Genetics* **120**, 645–650 (1988).
41. Kroos, L., Kuspa, A. & Kaiser, D. A global analysis of developmentally regulated genes in *Myxococcus xanthus*. *Dev Biol* **117**, 252–266 (1986).
42. Jilk, R. A., Makris, J. C., Borchardt, L. & Reznikoff, W. S. Implications of Tn5-associated adjacent deletions. *J Bacteriol* **175**, 1264–1271 (1993).

43. Kazazian, H. H. Mobile Elements: Drivers of Genome Evolution. *Science* (1979) **303**, 1626–1632 (2004).
44. Taylor, R. G. & Welch, R. D. Recording Multicellular Behavior in *Myxococcus xanthus* Biofilms using Time-lapse Microcinematography. *Journal of Visualized Experiments* (2010) doi:10.3791/2038.
45. Kaiser, D. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proceedings of the National Academy of Sciences* **76**, 5952–5956 (1979).
46. Wall, D. & Kaiser, D. Alignment enhances the cell-to-cell transfer of pilus phenotype. *Proceedings of the National Academy of Sciences* **95**, 3054–3058 (1998).
47. Kaplan, H. B., Kuspa, A. & Kaiser, D. Suppressors that permit A-signal-independent developmental gene expression in *Myxococcus xanthus*. *J Bacteriol* **173**, 1460–1470 (1991).
48. Kroos, L., Kuspa, A. & Kaiser, D. A global analysis of developmentally regulated genes in *Myxococcus xanthus*. *Dev Biol* **117**, 252–266 (1986).
49. Varon, M. & Rosenberg, E. Transcriptional regulation of genes required for antibiotic TA synthesis in *Myxococcus xanthus*. *FEMS Microbiol Lett* **136**, 203–208 (1996).
50. Guillaume Diss et al., Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* 355, 630-634 (2017). DOI:10.1126/science.aai7685

CHAPTER 3: A 'COMMON GARDEN' APPROACH TO INVESTIGATING THE ROLE OF KEY SIGMA (σ) FACTORS IN *MYXOCOCCUS XANTHUS* DEVELOPMENT

A. PROJECT SUMMARY

Understanding the molecular mechanisms that underlie bacterial social behavior and developmental plasticity can provide insights into microbial ecology, evolution, and biotechnology applications. This study employs a common garden approach to investigate the impacts of sigma (σ) factors on development in *Myxococcus xanthus*, a model organism for studying social behavior and morphological differentiation in bacteria. Common-garden experiments involve the comparison of genetically distinct strains, families or populations under identical environmental conditions. We grew multiple *M. xanthus* strains under controlled environmental conditions and analyzed their growth and morphological characteristics. Our results demonstrate that there are additional uncharacterized σ factor genes in the family that play a significant role in regulating biofilm dynamics and gene expression during development. We show that different σ factors have diverse impacts on biofilm formation and dynamics; additionally, they are important for regulating the expression of key genes associated with developmental processes. Our study also provides insights into the molecular mechanisms underlying the impacts of environmental factors on bacterial development, and how changes in gene expression contribute to developmental plasticity. Overall, our findings advance our understanding of bacterial social behavior and developmental plasticity and provide new targets for further research in this field.

B. INTRODUCTION

Biofilms are attractive models for understanding gene regulation in prokaryotes.

Biofilms are complex communities of bacteria that play a critical role in promoting the survival of prokaryotes in harsh conditions.¹ Their structural organization and genetic regulation make them an attractive model for studying how multicellular behaviors, such as communication, metabolism, and gene expression, are regulated in prokaryotes. This is particularly important given the significant medical and environmental impact of biofilms, which can be difficult to eliminate due to their resistance to antibiotics and other stressors. *Myxococcus xanthus* is a well-known model organism for studying genetic regulatory mechanisms in biofilms.⁹ This proteobacterium displays distinct motility behaviors when cultivated on nutrient-rich or starvation media, forming predatory biofilms that expand and ripple on nutrient-rich media, and aggregating into fruiting bodies on starvation media. Sporulation requires dynamic changes in gene expression making it an attractive model for investigating the relationship between multicellular mechanisms and genetic regulatory pathways.

Single-gene inactivation of regulatory genes in *M. xanthus* can result in defects in traditional phenotypes such as motility, growth rate, and germination success.¹⁰ By analyzing sporulation efficiency, researchers can identify genetic regulatory modules that govern the development, evolution, and disease genetics of organisms.⁸ Investigating the relationship between biofilm development and genetic regulatory pathways is critical in understanding genetic regulatory modules, and analyzing sporulation efficiency can provide insights into these

mechanisms. Overall, biofilms provide a unique and diverse platform for studying gene regulation in prokaryotes. Their complex structure and genetic regulation make them an ideal model for investigating how multicellular behaviors are regulated, particularly in the context of bacterial survival in harsh conditions. *M. xanthus* provides a useful model organism for studying genetic regulatory mechanisms in biofilms, highlighting the importance of investigating the relationship between multicellular mechanisms and genetic regulatory pathways in prokaryotes.

Figure 1: Wild-type *Myxococcus xanthus* lifecycle.

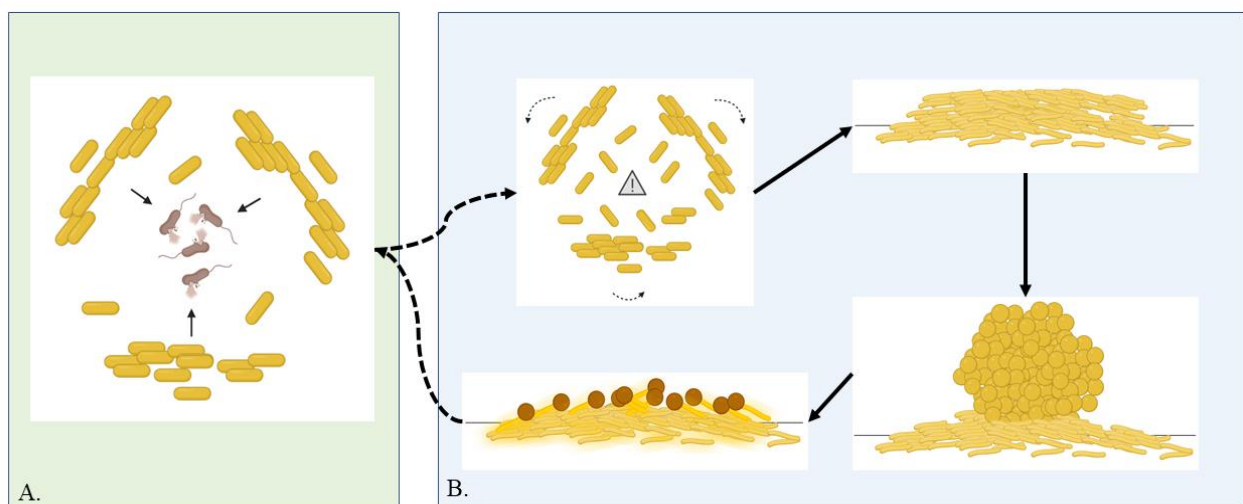


Figure 1: *Myxococcus xanthus* is a gram-negative, soil-dwelling bacterium found in various environments, including soil, water, and the surface of plants. It is known for its complex life cycle and is used as a model organism to study multicellularity. (A) When in contact with prey, *M. xanthus* cells penetrate the prey colony and secrete lytic enzymes, surviving as a dynamic biofilm that can form emergent behaviors such as rippling. (B) When starved, the population of cells can initiate a developmental program using a cascade of genetic signals to exchange

extracellular and intracellular signals to form aggregates when nutrients are no longer present. As the local population density of *M. xanthus* increases, the bacteria form dense, multicellular aggregates called fruiting bodies filled with differentiated and environmentally resistant myxospores. Myxospores ensure survival during starvation or dehydration and germinate when sufficient nutrients are detected, allowing a return to the vegetative state.

***Myxococcus xanthus*'s Gene Regulatory Network is Complex**

A gene regulatory network (GRN) is a complex system of interactions between genes and other molecular components that control the expression of genes in a cell. This system plays a crucial role in determining cellular identity and function. Understanding the molecular basis of the stringent response, or starvation response, in *M. xanthus* can provide valuable insights into the regulatory mechanisms underlying cellular differentiation during nutrient starvation.¹¹ Currently, we know that the stringent response in *M. xanthus* is regulated by a network of four core modules, Nla24, Mrp, FruA, and EBPs, which control transcription during spore development. This bacterial stringent response connects nutrient starvation with gene transcription via the stringent factor RelA, which detects the presence of deacylated tRNA in the ribosome as a signal of amino acid starvation and synthesizes the alarmone (p)ppGpp.¹¹ When nutrients are scarce, the prokaryotic protein RelA is believed to induce RNA polymerase stalling, allowing other transcriptional regulatory elements to bind to DNA.

Nla24 module: The Nla24 module is essential for the formation of fruiting bodies and motility in *M. xanthus*. NtrC-like activators (Nla) bind DNA sequences upstream of σ^{54} binding sites and play a vital role in this process. Of the 28 nla genes, 8 have been found to cause developmental

defects. Nla24 hydrolyzes ATP to increase exopolysaccharide (EPS) production, which is important for biofilm formation and swarming in *B. subtilis* and social motility regulation in *M. xanthus*. This results in higher expression of c-di-GMP, which binds to Nla24 to enhance myxospore sporulation and aggregation via the Dif chemosensory system. This pathway not only responds to starvation but also plays a role in quorum sensing by detecting adjacent cells' type IV pili and lipids.¹²⁻¹⁶

Mrp module: The Mrp module is a crucial regulator of gene expression in *M. xanthus*, with the autoregulator MrpC binding to nearly 300 genes involved in a range of cellular processes, including protein kinase activity, transcriptional regulation, extracellular signaling, and motility. Two signal transduction pathways, Pkn and Esp, and three proteins, MrpA, MrpB, and MrpC, are responsible for MrpC production. Increased MrpC expression is observed in aggregating cells, suggesting a correlation with developmental progress. Through nutrient proteolysis, MrpC can inhibit gene expression related to sporulation commitment during *M. xanthus* development.¹⁷⁻¹⁹

FruA module: The expression of FruA plays a crucial role in positive feedback loops through the Frz chemosensory system, which is involved in extracellular signaling that temporally regulates gliding motility during aggregation. Additionally, it works parallel to MrpC binding, and the combinatorial regulation of both pathways ensures that starving cells are present nearby and at a sufficiently high density to devote resources to sporulation. This has significant biological implications for the regulation of cell behavior in *M. xanthus*, particularly in the context of social motility and developmental progression.²⁰

Enhancer Binding Proteins (EBPs) module: EBPs are auxiliary proteins involved in transcription factor cascades, which bind to promoter sites 100 bps or more upstream to initiate and recruit transcription of σ^{54} polymerases. The EBP cascade comprises ActB and MXAN 4899, which positively autoregulate extracellular signaling responses necessary for development. These signals serve as cues for density and proximity and are regulated in combination by two more EBPs.^{21,22}

The four interconnected modules, Nla24, Mrp, FruA, and EBPs; regulate transcription in *M. xanthus* and ensure proper myxospore development. However, these modules require yet even further expression regulation, which has yet to be extensively characterized.

Additional σ Factors May Play Roles In *M. Xanthus* Development

M. xanthus development is finely controlled by gene expression, but the extent of other transcriptional regulators is still unknown. Prokaryotic σ factors comprise another widely distributed and important family of transcriptional regulators. These proteins bind reversibly to RNA polymerase (RNAP) and mediate the site-directed transcription of bacterial genes. During RNA synthesis initiation, σ factors perform three key functions: (1) recruiting RNAP holoenzyme to sequence-specific promoters, (2) melting double-stranded promoter DNA and stabilizing it as an open (single-stranded) complex, and (3) interacting with other transcription factors to regulate gene expression.²³ This mechanism is essential for vegetative growth and general cellular maintenance, performed as housekeeping expression (σ^A in *Escherichia coli*), or as a stress-specific response to the cell's environment, such as heat, toxicity, or starvation.²⁴ Previous studies have identified at least five key σ factors involved in gene expression during

development shown in Table 1. Additionally, RpoN (σ^{54}) is upregulated throughout development and is responsible for turning on a large subset of stress response genes.

Table 1: Primary σ Factors of *M. xanthus*

σ factor	MXAN#	Function	References
SigA	MXAN_5204	Housekeeping σ factor controls the transcription of genes involved in growth and metabolism.	[36]
SigB	MXAN_3357	Essential for the expression of late-stage development genes, specifically the maturation of spores before dormancy	[37,38]
SigC	MXAN_6209	Acts as a nutrient sensor for <i>M. xanthus</i> that negatively regulates the initiation of development	[38]
SigD	MXAN_2957	Involved in early and late developmental gene expression	[39]
SigE	MXAN_6759	Controls transcription of genes involved in cell envelope biosynthesis involved in fruiting body maturation	[38,40]
SigF	MXAN_2437	Required for fruiting body formation during development as well as social motility during vegetative growth	[57]

Figure 2: Classification of σ factors

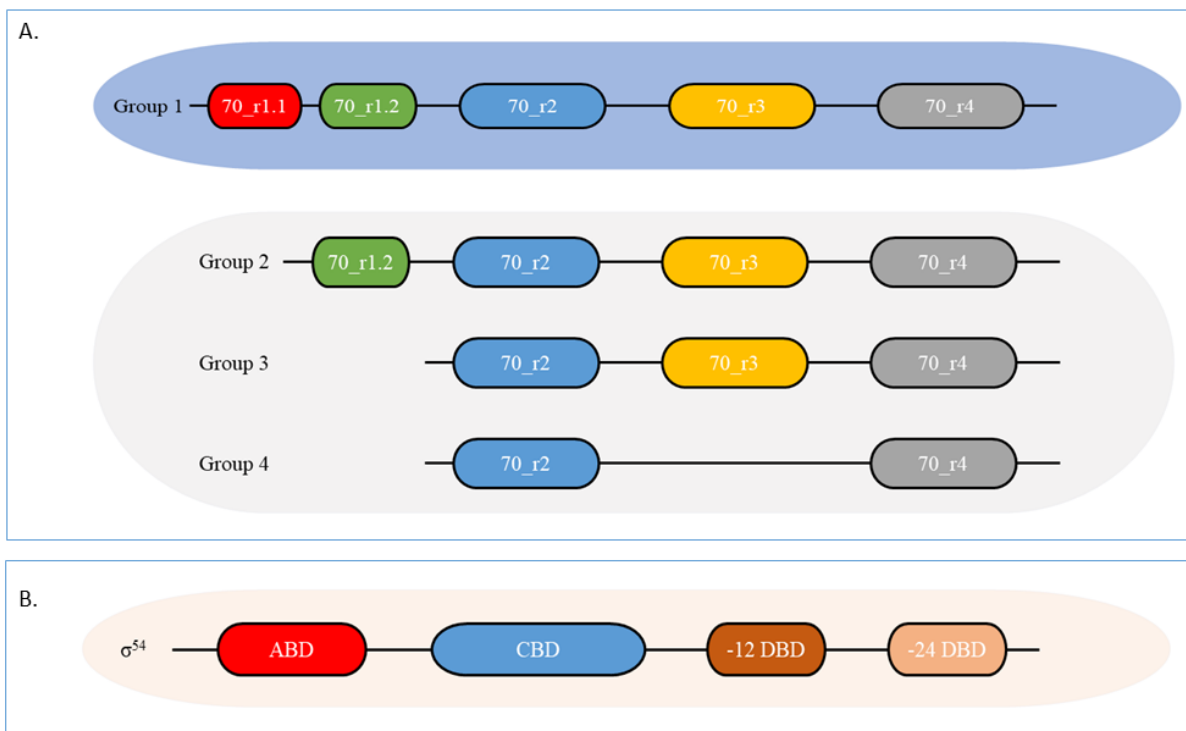


Figure 2: Classification of σ factors. A) Groups 1 to 4 in the σ^{70} family differ by the presence and absence of four conserved regions. σ^{54} factors differ in structure and mechanism of action but are critical for genomic stress responses. Group 1 factors have all four conserved domains within their structure. Group 2 is closely related to Group 1 but lacks the 70_r1.1, which acts as a gatekeeper in the selective binding of promoters and is considered non-essential by function. Group 3 is structurally and functionally diverse but usually contains 70_r2, 70_r3, and 70_r4 domains. Group 4 members possess the 70_r2 and 70_r4 domains at a minimum but are more phylogenetically diverse.^{25,26} At least two domains are required for a protein to be considered a σ factor: 70_r2, where binding to RNAP occurs, and 70_r4, where transcription initiation occurs

at sequence-specific promoters. Domains 70_r1 and 70_r3 serve as additional regulatory points during transcription. B) Whereas the σ^{70} holoenzyme is sufficient for transcription, σ^{54} is incapable of transitioning from the closed complex to the open complex on the DNA without the assistance of an EBP to hydrolyze ATP due to its activator binding domain (ABD). The remaining polymerase stabilization occurs at the core binding domain (CBD) with promoter specificity and transcription initiation at the -12DBD and -24DBD.^{27,28}

In addition to their fundamental role in cellular processes, σ factors are also known to be essential for multicellular development in prokaryotes. For instance, in *Bacillus subtilis*, σ factors (σ^F , σ^E , σ^G , and σ^K) direct the temporal and spatial expression of developmental genes, critical for spore morphogenesis.²⁹⁻³² Similarly, in *Streptomyces coelicolor*, the σ factor BldN regulates the expression of developmental genes during aerial hyphae formation.³³ Moreover, σ factors are also known to play crucial roles in pathogenic bacteria such as *Streptococcus pneumoniae* and *Pseudomonas aeruginosa*, where they regulate virulence gene expression.^{34,35} Ultimately, σ factors are essential transcriptional regulators in prokaryotes that mediate the transcription of all bacterial genes. They play critical roles in various cellular processes, including development and stress responses, as well as virulence gene expression in pathogenic bacteria.

Our study aimed to identify any uncharacterized σ factors that play a role in myxospore development in *M. xanthus* and to determine their relationship with previously characterized development σ factors. To achieve this, we first identified all σ factors within the DK1622 genome and generated a mutant library with each σ factor inactivated via homologous recombination. We then characterized aspects from the phenotype of each mutant in terms of

growth, motility, and sporulation. Utilizing phenotype assays as indicators of fitness is useful in studying transcriptional regulators in *M. xanthus* because these assays provide a reliable means of measuring the effects of gene regulation on cellular behavior and function. By manipulating the expression of transcriptional regulators and observing the resulting changes in phenotype, researchers can gain insight into how these regulators control key cellular processes such as growth, motility, and sporulation. Phenotype assays are also highly quantitative, allowing for precise measurements of fitness and other important parameters, and can help shed light on the underlying mechanisms of bacterial adaptation and survival.

We then identified changes in the expression of known developmental genes using RT-qPCR. Studying relative gene expression changes in key developmental pathways in *M. xanthus* mutants can provide valuable insights into the regulatory mechanisms governing these processes. For example, a study by Shimkets et al. (2000) analyzed the transcriptome of *M. xanthus* cells lacking the σ^F gene, which encodes a σ factor involved in sporulation. The study found that the expression of genes involved in sporulation and stress responses was significantly altered in the mutant cells compared to wild-type cells. Similarly, a study by Kroos et al. (2010) investigated the role of σ factor σ^E in *M. xanthus* development. The study found that σ^E plays a critical role in the initiation of developmental gene expression and is required for the transcription of several key genes involved in cell differentiation. By analyzing relative gene expression changes in *M. xanthus* σ factor mutants, we can gain a deeper understanding of the regulatory networks controlling development in this bacterium. To investigate global perturbations to the gene regulatory network, 7 developmental genes were screened in mutants with inactivated σ factors (via RT-qPCR) for abnormal patterns of expression. The

expression level of RNA Polymerase D, RpoD (MXAN5204), was used as a reference gene to normalize global transcription patterns. The reference gene is expressed at a relatively constant level in all cells and is used as an internal control to correct for variations in the amount of RNA or cDNA template added to the reaction mix, as well as variations in the efficiency of the reverse transcription and amplification steps. To ensure the validity of the reporters, we first screened the wild-type DK1622 to ensure the development pattern in our strain matched prior reports in the literature.

The results collected from this study suggest that *M. xanthus* has additional σ factors that are important for myxospore development, including 2 uncharacterized σ factor-like genes. These findings add to our understanding of the transcriptional pathways that regulate myxospore development for spatial and temporal mapping of the *M. xanthus* gene regulatory network.

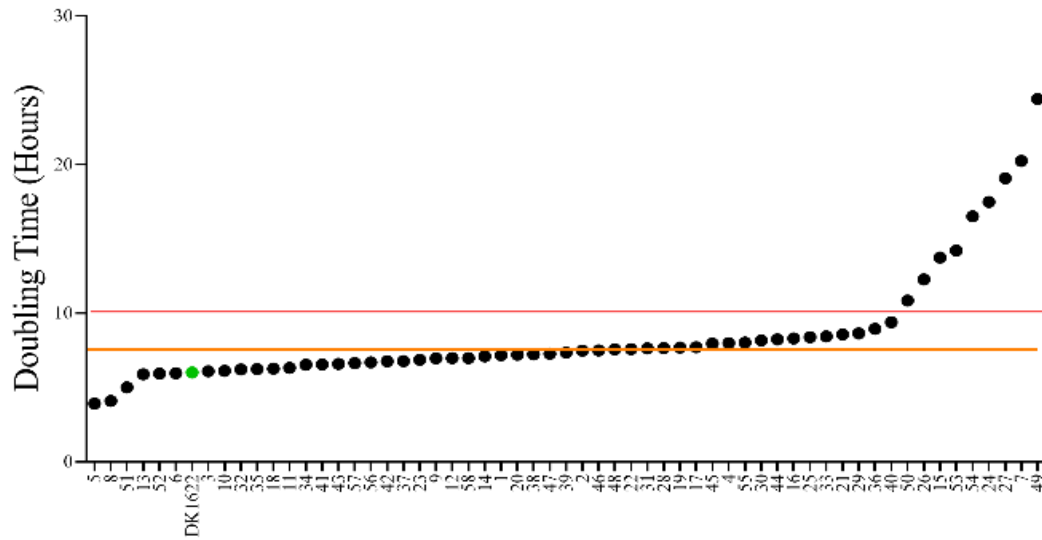
Results

Annotation of σ factors in the *M. xanthus* genome.

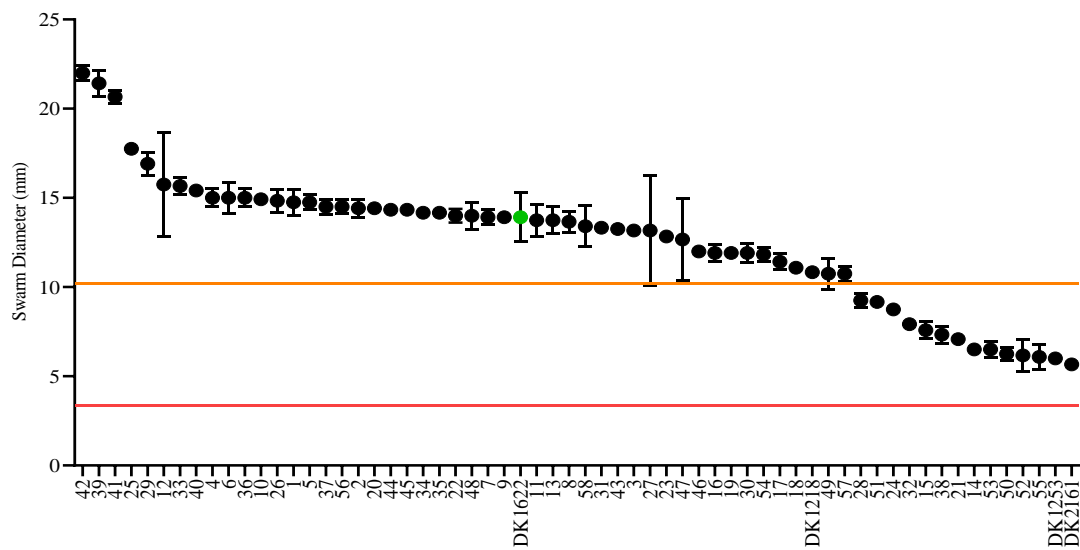
The sequence of *M. xanthus* was obtained from GenBank (<http://www.ncbi.nlm.nih.gov/GenBank/>) with the accession number NC_008095.1. Each predicted ORF in the genome was annotated using multiple databases. σ factor genes in the *M. xanthus* genome were reviewed and identified primarily using the databases Pfam and Interpro. Additional tools such as BLAST GenBank, and secondary structure predicting software such as SWISS-Model was used to assist in the selection. Manual curation was adopted to complete the annotation. To ensure that all putative σ factor ORFs were identified in the genome, a list of Pfam accession IDs associated with these ORFs was compiled and used to search the rest of the genome.⁴⁶⁻⁵² We went through the *M. xanthus* genome to identify annotated genes to encode at least two of the four characteristic domains of σ^{70} factors Figure 2. ORFs annotated as hypothetical or putative proteins that contained the required characteristic domains were manually checked using Pfam and psi-blast. BLAST searches indicate that 58 of these *M. xanthus* genes are likely to encode σ factors, and the highly conserved central domains (70_r2 and 70_r4) characteristic of the σ^{70} family of proteins was found in each case. Of the 58 genes identified in the *M. xanthus* genome sequence, 46 were previously annotated as σ factors. The remaining 12 genes were selected on similar sequence homology to the respective domains searched. These 58 genes and their annotations can be found in **S1**.

Figure 3: Fitness Assessment of Quantitative Phenotypes

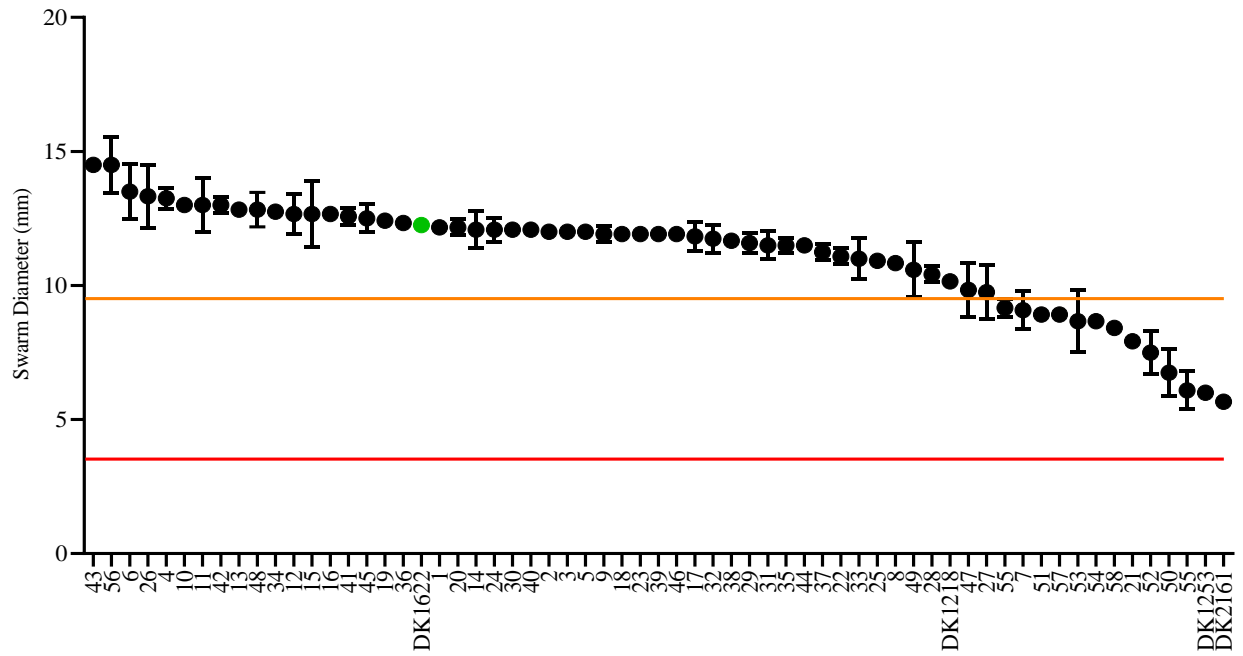
A. Doubling Time



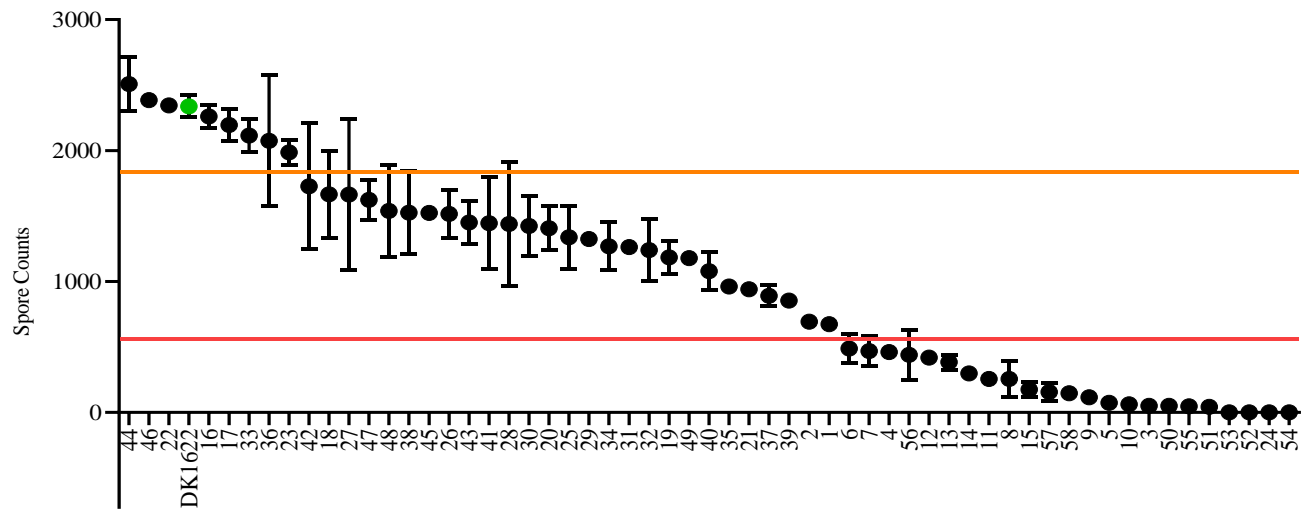
B. Soft Agar Motility



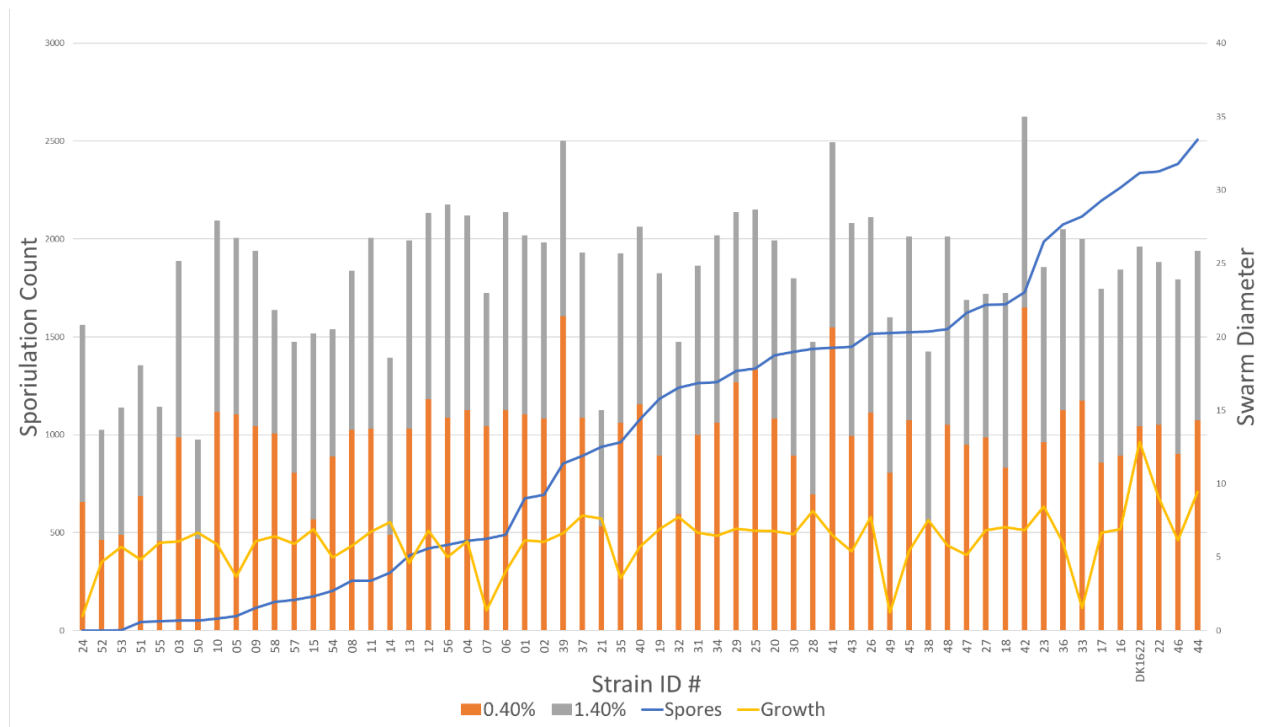
C. Hard Agar Motility



D. Germination Count



E. Phenotype Assays Overview



F. Mutants Significantly Defective for Sporulation

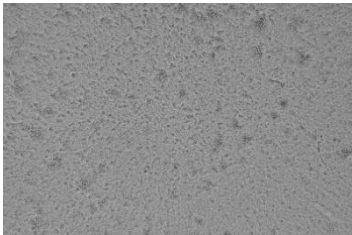
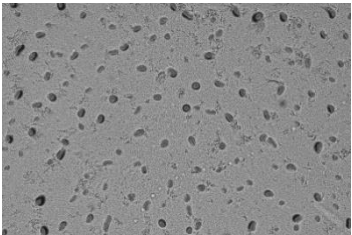
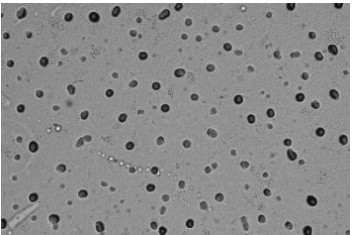
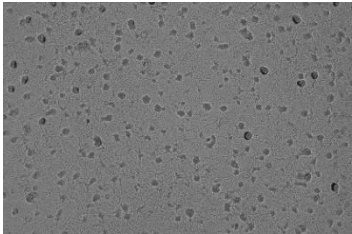
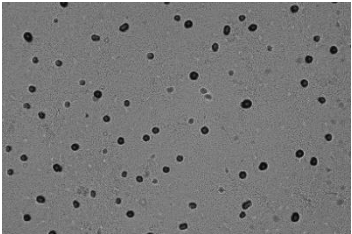
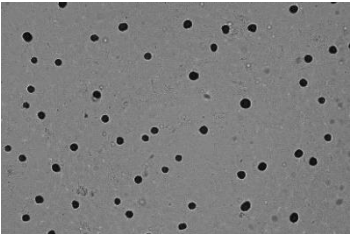
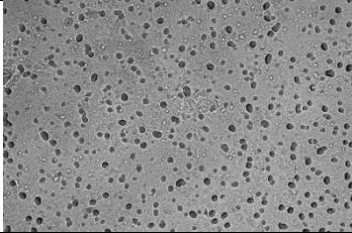
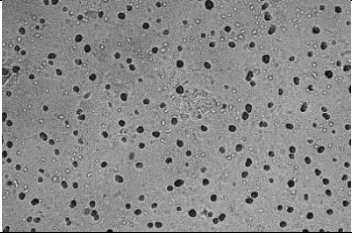
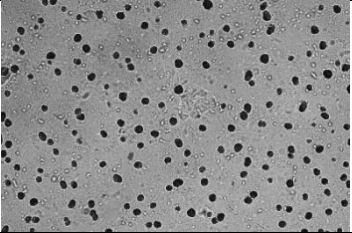
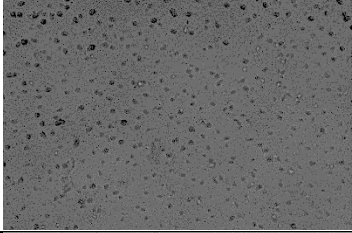
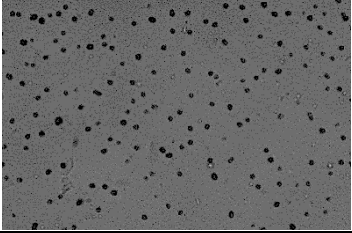
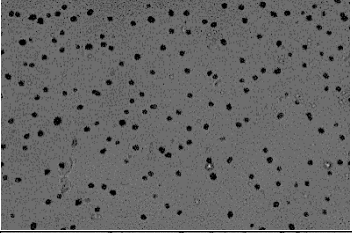
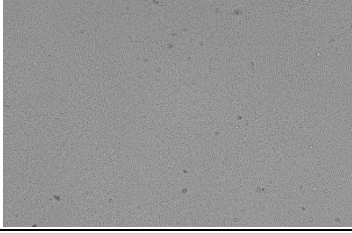
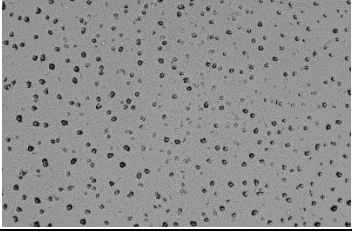
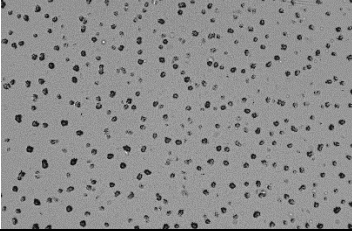
MX #	24	54	53	51	55	03	50	10	05	09	58	57	15	54	08	11	14	13	12	56	4	7	6
0.4 %			X		X		X										X						
1.4 %					X		X																
Growth	X	X	X										X	X								X	

Figure 3: Common Garden Phenotyping and Identification of Global Trends in Behavior. The DK1622 strain and 58 mutant strains were evaluated using different phenotypic assays, with the data organized in descending order from healthy to sickly for each assay. The results were

plotted on graphs, with the strains on the x-axes and the mean of three replicates for each assay on the y-axes. The order of the mutant strains differed for each graph. In each graph, DK1622 was represented by a green dot, while >75% and >25% DK1622 were represented by orange and red lines, respectively. A) The max linear growth rate was calculated during the log growth, and the doubling time was determined by fitting a line of best fit from three growth curves. B) The swarm diameter of 61 strains, including 58 mutants, DK1622, and motility control strains, was arranged in order from largest to smallest. The swarming motility was evaluated on soft agar (0.4%) and C) hard agar (1.4%). D) The viable spore count was also determined. The uncharacterized σ factor family was then related to previously characterized primary σ factors based on germination count. E) Graph E highlight trends that could be identified from looking at all phenotype parameters collectively. Mutants that were low in sporulation success (the blue line) typically had lower levels of motility (orange + grey bars) as well as lower rates of growth (orange line). However, there were notable outliers that could still were successful at sporulation but were significantly impacted in motility and growth, and vice versa. F.) This study emphasized development and therefore we wanted to characterize the mutants with the most significant impacts on phenotype, therefore we chose to investigate the sickliest mutants for sporulation for further analysis. The table above further supports our hypothesis as we identified all the previously recognized σ factors essential for sporulation in *M. xanthus*, in addition to 16 other putative developmental genes.

Figure 4.

A. Time-Lapse Microscopy Images

MX/MXAN#	6 Hours	12 Hours	24 Hours
WT DK1622			
7 MXAN_6681			
33 MXAN_4316			
49 MXAN_0233			
51 MXAN_0785			

B. Time Lapse Microscopy Map PCA

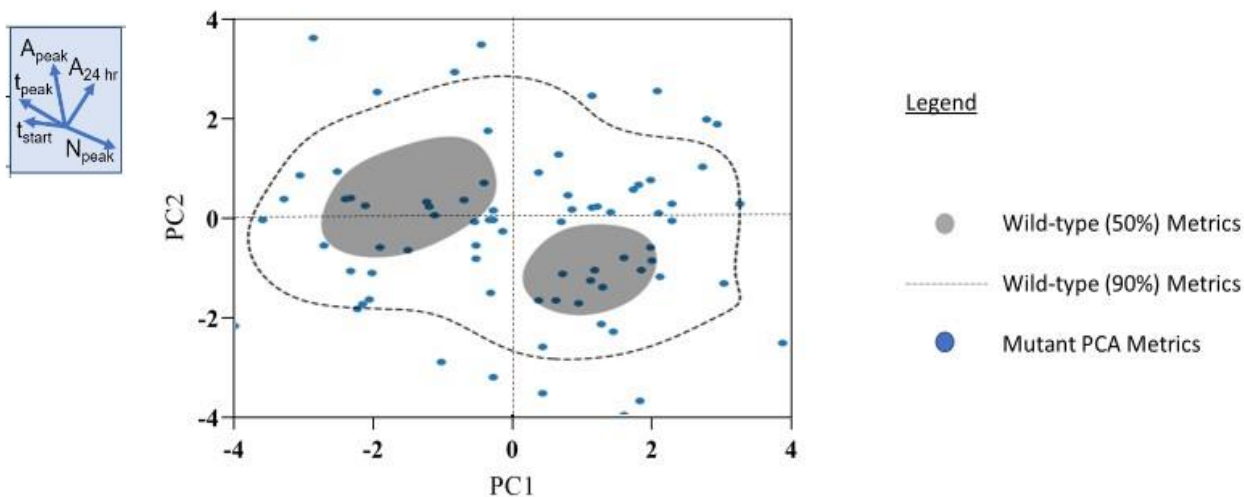


Figure 4. **Multivariate Metrics Analysis Separates Development Outliers using Timelapse**

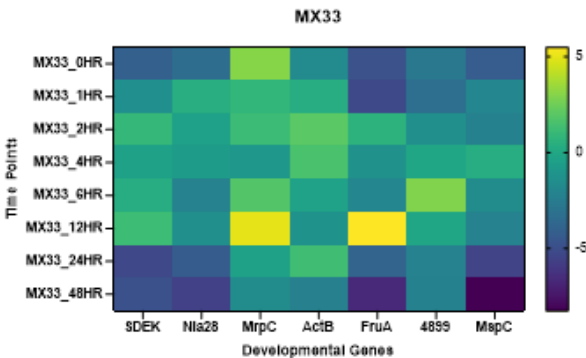
Microscopy (A) Image analysis is useful for identifying trends in myxobacterial development.

The 6HR, 12HR, and 24HR timepoints are shown recording the fruiting body morphogenesis of a subset of *M. xanthus* σ factor mutant strains via time-lapse imaging for 24 hours of development. Metrics relating to timing (start time, peak time, and stability time) selected, as well as the mean and standard deviation in the average aggregate area at peak time and 24 hours. The number of identifiable aggregates at peak time and 24 hours was also measured, along with the fraction of aggregates that appeared and dispersed. B.) 100 movies were mapped to a 2D phenotype space using principal component analysis. These results were screened against wild-type to determine whether the aggregation dynamics of the mutants were statistically significantly different from wild-type *M. xanthus*.

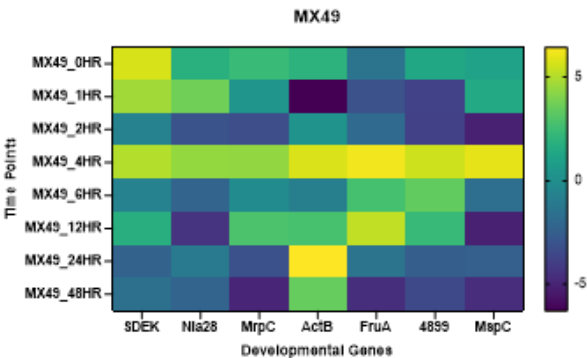
Figure 5. Identification of Deviations from Wild-Type Expression using RT-qPCR

A.

Strain	33
Spore (%)	100
0.4%	+
1.4%	+
Growth	+

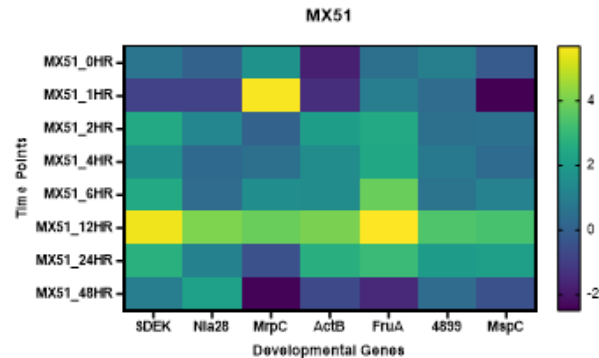


Strain	49
Spore (%)	50%
0.4%	+
1.4%	+
Growth	+



B.

Strain	51
Spore (%)	0
0.4%	+
1.4%	+
Growth	+



Strain	7
Spore (%)	0
0.4%	+
1.4%	+
Growth	X

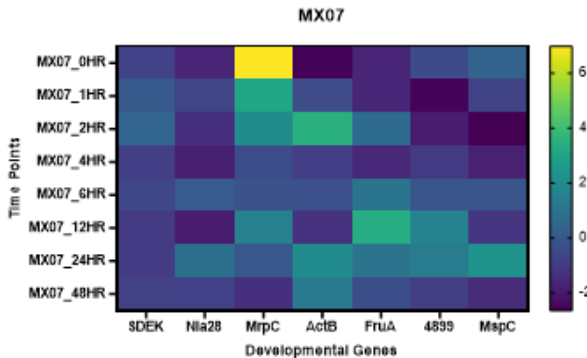


Figure 5: Identification of Deviations from Wild-Type Expression using RT-qPCR. To identify outliers from wild-type gene expression in *M. xanthus*, we used RT-qPCR to measure the expression of seven essential genes critical for development regulation. These genes were chosen as proxy indicators of wild-type development. We measured the expression (Blue showing low expression fold, yellow showing high expression fold) of developmental genes at time points 0, 1, 2, 6, 12, 24, and 48 hours in DK1622 and normalized it to housekeeping expression using RpoD. The mutants were also quantified for expression in each developmental gene, normalized to levels of wild-type expression, and mapped as a heatmap to identify differences in expression timing and quantity. In *M. xanthus*, changes in global gene expression of the core regulatory modules lead to phenotypic changes. These changes include cell aggregation into fruiting bodies or the transition of vegetative cells into spores. **Supplemental Table S2** details the literature on these genes and the rationale for their choosing.

G. RESULTS

Global trends in phenotype show the diversity of phenotypes caused by the inactivation of different σ factors. To visualize global trends in the σ factor family, all phenotypes were observed from replicate assays for quantitative comparison. When observing growth rates, strains that exhibited severe growth defects grew at rates >25% DK1622, could not produce germinating spores, and moved ~50% of the distance DK1622. These cut-offs were established for consistency across assays for ease of visualization. In addition, many of these strains were sensitive to freezing and required 100-fold concentration to increase the chances of preserving them long-term at -80 C due to their frailty. When observing motility on either agar, stiffness seemed to have no strong apparent effect on the family, with no mutation severely affecting

the swarming rate. When observing sporulation, however, there was a significant impact on the ability to germinate successfully from collected spores. About 1 in 3 mutants had a severe phenotype < 25% sporulation success only about 1 in 6 mutants had near wild-type levels of sporulation success.

Timelapse imaging for multivariable analysis identifies σ factors outliers. We analyzed a dataset of 100 movies from Wildtype and 4 mutants (recorded over 3 days as technical replicates), measuring quantitative metrics for each video.⁴¹ These quantitative metrics were used to establish a PCA of morphogenesis phenotype data to then plot against phenotype variation vectors in the population. This method of analysis allows us to see the global distribution of aggregation dynamics from the movies collected and identify outliers with rare emergent behaviors and plot them against the distribution of behaviors observed from the prior phenotype assays. This allowed us to identify outliers in morphogenesis that still retained wild-type levels of assay behavior for narrowing down potential strains as regulatory targets. We postulated that those that had wild-type levels of motility during vegetative studies would behave normally during development, which was supported by our 85% inclusive rate. However, our results highlight that although we screened for σ factor mutants that behaved similarly to wild-type dynamics, based on motility and growth, there were still parameters of development that differed that could be contributing to the sporulation defects we observed in the remaining 15% of movies. Using this method resulted in putative strains for further characterization by RT-qPCR. For this study, we chose 3 to characterize further as well as one non-defective sporulating mutant.

σ factor mutations cause changes in key gene developmental expression. The expression patterns of the developmental genes selected were measured using RT-qPCR throughout development and normalized to the mutant's overall gene expression (RpoD normalization) and then normalized again to DK1622 expression. Prioritizing mutations that significantly impacted sporulation, yet retained wild-type motility parameters, we then checked for inhibited and outlier expression of essential proxy genes throughout development to identify whether any of the established critical pathways of *M. xanthus* transcription regulation were affected by the σ factor mutations. Using this approach, we identified σ factors that may be potential regulatory genes for developmental gene expression pathways in *M. xanthus*.

H. DISCUSSION

Growth rate and motility are strong indicators of sporulation fitness in *M. xanthus*.

This study investigates the role of σ factors in growth and sporulation in *M. xanthus*. Previous literature annotations and references to σ factors in the DK1622 *M. xanthus* genome have presented varied numbers of genes in the family from 44 to 56. We conducted a comprehensive study of all 58 predicted σ factors and σ factors-like genes and characterized their phenotypes and expression of key developmental genes. Most mutations to this family impaired viable myxospore formation in *M. xanthus*, but no single mutation proved lethal under vegetative conditions. We further provide evidence that there is not an “essential” housekeeping σ factor during development, but rather a possibility of other σ factors maintaining gene regulation. The heatmaps identify deviations from wildtype, however further characterization of the entire sigma factor family would be necessary for understanding

whether this impact was biologically significant. This agrees with prior literature discussing transcriptome changes during the stringent response. At 6 hours, there is a noticeable change in *M. xanthus* expression, which promotes the binding of other σ factors to RNA polymerase, allowing for other σ factors to assume gene regulation and dictate the cellular fate.

Vegetative fitness is important for aggregation dynamics.

The use of both vegetative and developmental assays in *M. xanthus* is a powerful approach for identifying outliers, which are strains that deviate significantly from the average behavior of the population. This combined approach allows for the detection of strains that exhibit anomalous vegetative growth or developmental behavior and can reveal genetic or environmental factors that contribute to the variability and plasticity of bacterial social behavior and development. Identifying the “worst” candidates using phenotype assays allowed us to verify out PCA methods for biological significance and allowed for the identification of subtle outliers in *M. xanthus*; important for understanding the molecular mechanisms underlying bacterial social behavior and developmental plasticity. These 16 outliers were attractive targets for understanding deviations in profile from wild-type behaviors.

Key Development Genes Can Identify Changes In Transcriptome Patterns By Proxy In Mutants.

Expression profiles can provide valuable insights into a gene's function and its impact on phenotype. For example, FruA expression at various time points has been shown to have pleiotropic effects on fruiting body count and size, depending on the expression levels of its trans-regulatory gene.⁴² Similarly, early MrpC expression has been found to result in early

aggregation of swarms and sporulation occurring outside of fruiting bodies.⁴³ To better understand the relationship between gene expression and phenotype in *M. xanthus* development, a systematic study of a large gene family that regulates essential stress-response genes would be useful. We did not find much discernably different or consistently different for any of the mutants selected, however, a larger study could shed light on the significance of differences in the expression of essential regulatory genes and their impact on phenotype.

I. CONCLUSION

The study of bacterial development and social behavior is an area of great interest for both basic and applied research. Understanding the molecular mechanisms that underlie bacterial social behavior and developmental plasticity can provide insights into microbial ecology, evolution, and biotechnology applications. *Myxococcus xanthus*, a model organism for studying bacterial social behavior, exhibits remarkable developmental plasticity in response to environmental cues. The σ factor family has been implicated in regulating many aspects of *M. xanthus* development, including multicellular behavior and biofilm formation. By characterizing the impacts of σ factors on *M. xanthus* development using a common garden approach, our study sheds light on the molecular mechanisms underlying bacterial social behavior and developmental plasticity and highlights their pleiotropic effects in fruiting body dynamics and regulating gene expression.

The findings of our study have several important implications for future research. We highlight critical sigma factors important for development based on standard laboratory conditions and their impact to phenotype based on parameters that are established in literature. Additionally,

our study suggests that certain σ factors may have distinct roles in regulating gene expression in *M. xanthus*. Our RT-qPCR results suggest certain sigma factors may have time-sensitive roles, and others may be pathway specific. Future research could aim to identify specific target genes and pathways regulated by σ factors, and how these pathways contribute to bacterial social behavior and developmental plasticity. It is also possible that the phenotype assays here failed to capture the scope of the effects of the σ factors' contributions to development and future work may expand on other assays of quantitative phenotype. One potential direction for further explore is how environmental factors interact with σ factor signaling to regulate bacterial development. For example, it would be interesting to examine the impacts of nutrient availability, temperature, and other environmental factors on σ factor expression and biofilm dynamics in *M. xanthus* and relate those changes.

Finally, our study highlights the importance of using common garden approaches to investigate bacterial social behavior and developmental plasticity. Common garden experiments allow for controlled manipulation of environmental variables and rigorous statistical analysis of developmental phenotypes. This approach is particularly useful for studying bacterial development, which is highly sensitive to environmental cues. By employing a common garden method, future research can continue to make significant contributions to our understanding of bacterial social behavior and developmental plasticity.

J. MATERIALS AND METHODS

Media used for growth, motility assays, and developmental assays. *M. xanthus* strains were grown at 32°C in CTTYE broth containing 1.0% Casitone (Difco Laboratories), 0.5% yeast extract

(Difco Laboratories), 10.0 mM Tris-HCl (pH 8.0), 1.0 mM KH₂PO₄, and 8.0 mM MgSO₄ or on plates containing CTTYE broth and 1.5% Difco Bacto Agar. The motility of *M. xanthus* strains was assayed at 32°C on CTTYE plates containing 1.5 or 0.4% Difco Bacto Agar. CTTYE broth and plates were supplemented with 40 g of kanamycin sulfate (Sigma). *E. coli* Top10 was grown at 37°C in Luria broth (LB) containing 1.0% tryptone (Difco), 0.5% yeast extract (Difco), and 0.5% NaCl or in plates containing LB and 1.5% Difco Bacto Agar. LB and LB plates were supplemented with 40 g of kanamycin (Sigma). Fruiting body development was conducted at 32°C on plates containing TPM buffer (10.0 mM Tris-HCl [pH 8.0], 1.0 mM KH₂PO₄, and 8.0 mM MgSO₄) and 1.5% Difco Bacto Agar. Sporulation was allowed on CTTYE with cells separated in CTTSA [1.0% Casitone, 10 mM Tris-HCl (pH 8.0), 1 mM KH₂PO₄, and 8 mM MgSO₄, 0.4% agar] soft agar, and spores developed over five days.

Generating a mutant library. Internal fragments of each gene were generated using PCR as described in the Invitrogen TOPO TA cloning kit (thermostable DNA polymerase and gene-specific primers were used in the PCRs). Each PCR fragment was cloned (using the procedure specified by the manufacturer [Invitrogen]) into a pCR2.1-TOPO plasmid, generating plasmids designated pEC1–pEC58. To confirm that the plasmid was successfully inserted into the desired location in the *M. xanthus* chromosome, we used PCR to amplify across the upstream region of the target gene loci and TOPO vector, thereby generating an amplicon with a size of ~1.2 kb. Wild type DK1622 was used as a negative control. Plasmids containing fragments of each σ factor (pEC1–pEC58) were electroporated as previously described.⁵⁸ Following electroporation, cells were placed into flasks containing 1.5 ml of CTTYE and incubated at 32°C for 12 to 24 h with vigorous agitation. Aliquots (500 μ l) of these cultures were added to 5.0 ml of CTTSA (0.4

% soft agar) and poured onto CTTYE plates containing 40 mg/mL kanamycin. Chromosomal DNA was isolated from Kan^r colonies and used for PCR screening analysis. A single band of the appropriate size identified transformants that carry a single copy of a pEC plasmid integrated into the target σ factor locus by homologous recombination. After confirmation that the transformants carried the appropriate insertions, they were scored for growth, motility, and viable spore count.

***M. xanthus* growth curves.** Cell growth was monitored with a Klett-Summerson colorimeter. Results are shown from the initial inoculation of liquid cultures to the portion of the growth curve during which cells left the log phase and entered the stationary phase. Next, 3 (20 ml) cultures were inoculated with 100 microliters of 5×10^9 starter culture. The density of each culture was then read every 6 hours from inoculation until a stationary phase or decay was established.

***M. xanthus* development.** *M. xanthus* strains were inoculated into flasks containing CTTYE broth, and the cultures were incubated at 32°C with vigorous swirling. After each culture reached a density of 5×10^8 cells/ml, the cells were pelleted, the supernatant was removed, and the cells were resuspended in TPM buffer to a density of 5×10^9 cells/ml. Aliquots of 20 microliters were spotted on TPM agar plates and incubated at 32°C.

Viable spore count. To determine the sporulation efficiency of each *M. xanthus* strain, developing cells were harvested from TPM agar plates after five days. Three sets of 5 spots were harvested and suspended in 500 μ l TPM buffer from the TPM agar plates. The cells were then exposed to mild sonication (10% altitude, 10 s \times 3 with 30 s intervals, MISONIX, S-4000),

followed by heat treatment at 50°C for 2 h. Cells were then diluted to the desired concentration and plated with CTTSA onto CTTYE agar plates (supplemented with 40 µg/ml kanamycin sulfate for insertion mutants). After 5 days of incubation at 32°C, viable spores germinated and grew into visible colonies, and the number of colonies was recorded.⁵³

Motility assays. In this study, the motility of *M. xanthus* cells with σ factor gene insertions was examined using the swarm expansion assay.^{54,55} The cells were grown to mid-log phase in CTTYE broth and resuspended in CTTYE broth to a density of 5×10^9 cells/ml. The cells were then spotted on CTTYE plates containing either 1.5% or 0.4% agar and incubated at 32°C for three days. The mean diameter of four swarms formed by each strain was measured and compared to the mean diameter of four swarms formed by wildtype strain DK1622. Mutants with statistically significant differences in mean swarm diameter on either 1.5% or 0.4% agar plates compared to the wildtype strain were classified as motility mutants.

Timelapse Microscopy. We set up simultaneously collecting time-series images for 96 experiments using an array of compact and identical microscopes controlled by a central computer. Each of the 96 microscopes is equipped with a single 4X objective lens, a Peltier device that maintains stage and sample temperature, a red-light source, and a camera controlled by a Raspberry Pi, a single-board minicomputer. The 3D-printed armature and assembly hardware serve to keep all components firmly in place and provide a focus knob for higher image quality.

To ensure uniform control of all microscopes and central storage of their time series output, each Raspberry Pi unit is networked via ethernet and two 64-port network switches to a central

hub computer. This computer runs Piserver software, which boots each Raspberry Pi from a single operating system image, allowing software to be changed and updated for all Raspberry Pi units simultaneously. Custom software written in Python provides a convenient GUI to control image acquisition from each camera via SSH and organize output in a centralized image storage location.

Quantitative PCR (qPCR). To examine the expression of developmental genes in *M. xanthus*, total RNA was isolated from both wild-type and mutant cells at different time points during vegetative growth and development from starvation media agar. RNA was isolated using the RNeasy Protect Bacteria Reagent and the RNeasy Mini Kit. Glass beads were added to help lyse developmental cells, and the cell suspensions were vigorously shaken. Pooled RNA samples were subsequently used to generate cDNA, and qPCR analysis was performed using the CFX Connect real-time PCR detection system. The CT value for each gene was repeated in triplicate and averaged, and relative fold changes in mRNA levels were calculated using the reference gene RpoD and the $\Delta\Delta CT$ method.⁵⁶ Seven independent wild-type and mutant samples were analyzed for each time point (0, 1, 2, 4, 6, 12, 24, and 48 hours). The gene expression data obtained from this study was used to identify the role of different σ factors in regulating gene expression during *M. xanthus* development.

Image processing pipeline. The phenotype was automatically quantified for each fruiting body aggregation assay in this study by running 144 individual TIFF images (ten minutes between each frame over 24 hours of total development) from each time series through a custom Python image processing and analysis pipeline to identify in each frame which pixels could belong to a fruiting body, based on their gray value. The information for the position and

geometry of each aggregate was filtered to remove noise and spurious aggregates. This detailed data summary for each time series then had a list of ten specific numbers extracted from it, each of which captures one overall feature, such as the time at which aggregation began or the average size of final fruiting bodies. The values of these ten metrics together were then used in further analysis.

Statistical methods. We first generated the contours for the wild-type PCA data by starting with Gaussian kernel density estimation (KDE) and using standard root-finding techniques to draw contours from the density estimate that capture 50% and 90% of the PCA data points. With these contours drawn, we then use a data-driven statistical technique like bootstrapping. Given a sample size N , 10,000 samples of that size are drawn from the wild-type dataset. Each subsample has a characteristic pair of numbers, (n_{50}, n_{90}) , which corresponds to the number of points in the sample that fall inside the 50% and 90% contours, respectively. Mutant strains are shown for visualization only and do not figure into the calculation of the p-values from wild-type but are visualized by the blue points in PCA space.

SUPPLEMENTARY DATA

TABLE S1: Strains and Their Genotypes Used in This Study

NCBI GENE ANNOTATION	ANNOTATION FOR THIS STUDY	PCR VERIFICATION PRIMER
MXAN_1210 putative RNA polymerase sigma-70 factor	MX01	GTGAGTGATGAGTTTC GCTGAC
MXAN_1510 hypothetical protein	MX02	GCATATGTGTGGAGT CGAAC
MXAN_1709 putative DNA-binding regulatory	MX03	GAGGAAGGTCACCGA GAAG
MXAN_1875 Hypothetical Protein	MX04	GTGGAGTCCCTCTCT CTCAAC
MXAN_2204 RNA polymerase sigma-70 factor	MX05	CGACTCTTCATCAGC GAAG
MXAN_2437 RNA polymerase sigma factor (FliA)	MX06	AAGACGTAGGACGGA GGTC
MXAN_6681 RNA polymerase sigma-70 factor	MX07	GACTACGCCTCACGC TTC
MXAN_3357 sigB; RNA polymerase sigma-B factor	MX08	ACAGTTCTGAAGCTCC TACAGC
MXAN_4088 carQ; RNA polymerase sigma factor CarQ	MX09	CCCTCTCACCAATTCC ATC
MXAN_4949 RNA polymerase sigma-70 factor	MX10	CGTAGTACTTCGCCTC GTC
MXAN_5101 putative ECF	MX11	GAGGAGGCAAGACAC AGC
MXAN_5204 rpoD; RNA polymerase sigma factor RpoD; σ A	MX12	CTCTTTGAATCCTCGG AAGC
MXAN_6209 sigC; RNA polymerase sigma-C factor	MX13	CGAAGAGCTGGAAGA GGTC
MXAN_7083 Putative DNA-binding regulatory	MX14	CGAGACGTTCCACAC CTC
MXAN_7161 Putative DNA-binding regulatory	MX15	GTGTCCCCCATGGAC TTAT
MXAN_0203 RNA polymerase sigma-70 factor	MX16	CTTGGAGGAAGTGAT GGATG
MXAN_0681 RNA polymerase sigma-70 factor	MX17	GTACCTGCTGATGATC TACGAG
MXAN_0947 RNA polymerase sigma-70 factor	MX18	GTAGCGGGTGAAGAA CTCC

MXAN_1514 conserved hypothetical protein	MX19	AGTTCTCCCAGAGCC TCAC
MXAN_1661 RNA polymerase sigma-70 factor	MX20	GTTGTGGGCCATTGT CTC
MXAN_2184 RNA polymerase sigma-70 factor	MX21	GAAGTTGACGCTCGA TTACG
MXAN_2395 RNA polymerase sigma-70 factor	MX22	GTTGAGCTCCAGGGG ATAC
MXAN_5245 conserved hypothetical protein	MX23	TCCGGTTCTCAAGCAA CT
MXAN_1879 Hypothetical Protein	MX24	GGTATTGAGGGTTGT GGAGA
MXAN_6010 Putative DNA-binding regulatory	MX25	GGGTTTGGAGGTCCA CTT
MXAN_2030 RNA polymerase sigma-70 factor	MX26	GGAGGAGCAGTAGCA GATGT
MXAN_2500 RNA polymerase sigma-70 factor	MX27	CATCTCCTTCGACCAG GA
MXAN_2929 RNA polymerase sigma-70 factor	MX28	AGATGTGGCTCACGT TGG
MXAN_3426 RNA polymerase sigma-70 factor	MX29	GATGAACCAGATGGG ACAC
MXAN_3686 RNA polymerase sigma-70 factor	MX30	GGCCTTCCTTGAAGA CATC
MXAN_3959 RNA polymerase sigma-70 factor	MX31	GGCTGTTCCACTTCG ATTAC
MXAN_4309 RNA polymerase sigma-70 factor	MX32	AGCTCCACCAGGTAG CATC
MXAN_4316 RNA polymerase sigma-70 factor	MX33	GACTACCGGCTCGCA TTC
MXAN_4662 RNA polymerase sigma-70 factor	MX34	GATGAGTACAACCGC ACCTAC
MXAN_4733 RNA polymerase sigma-70 factor	MX35	CCTGTACGAACTGGA GATGAAG
MXAN_4987 RNA polymerase sigma-70 factor	MX36	CTCTCACAGGTAATG GTCATCC
MXAN_5263 RNA polymerase sigma-70 factor	MX37	CGGGAAGCTCTACGT GTC
MXAN_5410 RNA polymerase sigma-70 factor, ECF subfamily	MX38	GAGAGAGCCACCATG AAGAT
MXAN_5506 RNA polymerase sigma-70 factor	MX39	GTCCATGGTGCTCAA GGT

MXAN_5731 RNA polymerase sigma-70 factor, ECF subfamily	MX40	CAAGTCCCCGAGCAA CTC
MXAN_6058 RNA polymerase sigma-70 factor	MX41	GAGCTGACGGACTGG TTC
MXAN_6173 RNA polymerase sigma-70 factor	MX42	CTCCTGACAGGGAGA GACAG
MXAN_6461 RNA polymerase sigma-70 factor	MX43	CGTGGCGTACGTCTA TCTG
MXAN_7214 RNA polymerase sigma-70 factor	MX44	GACCTGTTCCACAACA TCG
MXAN_7289 RNA polymerase sigma-70 factor	MX45	CAGATCTTGCCTCGT GTGT
MXAN_7326 RNA polymerase sigma-70 factor	MX46	CAGGTAGTAGAGCGG GAAGAT
MXAN_7454 RNA polymerase sigma-70 factor	MX47	GCATCACGTGGTGTG TGTA
MXAN_6759 rpoE; RNA polymerase sigma-E factor	MX48	CCTGTGGGTGATGGA AGC
MXAN_0233 RNA polymerase sigma-70 factor	MX49	ACGGAGATGACGTTC TCCT
MXAN_0068 putative DNA-binding regulatory protein	MX50	GTCACCATGTGTATG GCATC
MXAN_0785 RNA polymerase sigma-32 factor family protein	MX51	GACGTTCTTCAGCTC GTTC
MXAN_1061 rpoN; RNA polymerase sigma-54 factor	MX52	GGTATCTGTGATCGT GCGTA
MXAN_2737 putative DNA-binding regulatory protein	MX53	CTCCTCCAGAAGGCG AAC
MXAN_2957 sigD; RNA polymerase sigma-D factor	MX54	TATCTTTCGCTGGAGC AGTC
MXAN_4147 rpoE1; RNA polymerase sigma factor RpoE1	MX55	GTGCGATGAACACCT ACAAC
MXAN_4535 putative RNA polymerase sigma-70 factor	MX56	GAGGGATGGGTCTGT CTTTAC
MXAN_4696 Hypothetical Protein	MX57	GTAGTCCTCCAAGGG AATGAC
MXAN_5870 sigE; RNA polymerase sigma-E factor	MX58	CTCTGACAGGTTCATG TCGTC

TABLE S2: Summary of Developmental Genes Screened

SDEK (MXAN1014): SDEK is a transcriptional regulator that plays a critical role in regulating social motility in <i>M. xanthus</i> . SDEK expression is induced during early development and peaks at 4-6 hours of development, after which it declines. It has been shown to regulate the expression of several genes involved in cell motility and adhesion, including fibril genes and Type IV pilus genes.
NLA28 (MXAN1167): NLA28 is a peptidoglycan hydrolase that is involved in regulating cell division in <i>M. xanthus</i> . Its expression is induced during early development and peaks at around 6-8 hours of development. NLA28 is required for proper cell division and its deletion leads to aberrant cell division and reduced sporulation efficiency.
MrpC (MXAN5125): MrpC is a transcription factor that is involved in regulating the expression of genes required for fruiting body formation and sporulation in <i>M. xanthus</i> . Its expression is induced during early development and peaks at around 8-10 hours of development. MrpC regulates the expression of several genes involved in sporulation, including spore coat genes and genes involved in germination.
ActB (MXAN3214): ActB is an EBP that is involved in regulating cell shape and motility in <i>M. xanthus</i> . Its expression is induced during early development and peaks at around 8-10 hours of development. ActB forms part of the bacterial cytoskeleton and is involved in regulating the localization of proteins and organelles within the cell.
FruA (MXAN3117): FruA is a transcription factor that is involved in regulating cell-cell signaling and fruiting body formation in <i>M. xanthus</i> . Its expression is induced during early development and peaks at around 12-16 hours of development. FruA regulates the expression of several genes involved in fruiting body formation, including genes encoding extracellular matrix proteins and proteases.
MXAN4899: MXAN4899 is a hypothetical protein that is conserved in <i>M. xanthus</i> and other myxobacteria. Its expression is induced during early development and peaks at around 8-10 hours of development. Its function is currently unknown, but it has been suggested to be involved in regulating cell motility or adhesion.
MspC (MXAN6969): MspC is a surface protein that is involved in regulating cell-cell adhesion in <i>M. xanthus</i> . Its expression is induced during early development and peaks around 24-48 hours. MspC is required for proper fruiting body formation, and its deletion leads to cell aggregation and sporulation defects.

D. REFERENCES

1. Donlan RM, Costerton JW. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev.* 2002;15(2):167-193. doi:10.1128/CMR.15.2.167-193.2002
2. Kostakioti M, Hadjifrangiskou M, Hultgren SJ. Bacterial biofilms: development, dispersal, and therapeutic strategies in the dawn of the postantibiotic era. *Cold Spring Harb Perspect Med.* 2013 Apr 1;3(4): a010306. doi: 10.1101/cshperspect.a010306. PMID: 23545571; PMCID: PMC3683961.
3. Mah TF, O'Toole GA. Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol.* 2001 Jan;9(1):34-9. doi: 10.1016/s0966-842x(00)01913-2. PMID: 11166241.
4. Høiby, Niels et al. "The clinical impact of bacterial biofilms." *International Journal of oral science* vol. 3,2 (2011): 55-65. doi:10.4248/IJOS11026
5. Ritchie LJ, Curtis ER, Murphy KA, Welch RD. Profiling *Myxococcus xanthus* Swarming Phenotypes through Mutation and Environmental Variation. *J Bacteriol.* 2021 Nov 5;203(23):e0030621. doi 10.1128/JB.00306-21. Epub 2021 Sep 20. PMID: 34543101; PMCID: PMC8570273.
6. Ma, Muqing, et al. "Identifying the Gene Regulatory Network of the Starvation-Induced Transcriptional Activator Nla28." *Journal of bacteriology* vol. 204,12 (2022): e0026522. doi:10.1128/jb.00265-22

7. Zhang, Huan, et al. "Myxococcus xanthus as a Model Organism for Peptidoglycan Assembly and Bacterial Morphogenesis." *Microorganisms* vol. 9,5 916. 24 Apr. 2021, doi:10.3390/microorganisms9050916
8. de Villemereuil, Pierre, et al. "Common Garden experiments in the genomic era: new perspectives and opportunities." *Heredity* 116.3 (2016): 249-254.
9. Caberoy NB, Welch RD, Jakobsen JS, Slater SC, Garza AG. Global mutational analysis of NtrC-like activators in Myxococcus xanthus: identifying activator mutants defective for motility and fruiting body development. *J Bacteriol.* 2003;185(20):6083-6094. doi:10.1128/JB.185.20.6083-6094.2003
10. Kianianmomeni, Arash. "Potential impact of gene regulatory mechanisms on the evolution of multicellularity in the volvocine algae." *Communicative & integrative biology* vol. 8,2 e1017175. 29 Apr. 2015, doi:10.1080/19420889.2015.1017175
11. Agirrezabala X, Fernández IS, Kelley AC, Cartón DG, Ramakrishnan V, Valle M. The ribosome triggers the stringent response by RelA via a highly distorted tRNA. *EMBO Rep.* 2013 Sep;14(9):811-6. doi: 10.1038/embor.2013.106. Epub 2013 Jul 23. PMID: 23877429; PMCID: PMC3790049.
12. Caberoy NB, Welch RD, Jakobsen JS, Slater SC, Garza AG. Global mutational analysis of NtrC-like activators in Myxococcus xanthus: identifying activator mutants defective for motility and fruiting body development. *J Bacteriol.* 2003;185(20):6083-6094. doi:10.1128/JB.185.20.6083-6094.2003

13. Lancero, Hope, et al. "Characterization of a *Myxococcus xanthus* mutant that is defective for adventurous motility and social motility." *Microbiology* 150.12 (2004): 4085-4093.
14. Lu, Ann, et al. "Exopolysaccharide biosynthesis genes required for social motility in *Myxococcus xanthus*." *Molecular microbiology* 55.1 (2005): 206-220.
15. Kroos, Lee. "Highly signal-responsive gene regulatory network governing *Myxococcus* development." *Trends in genetics* 33.1 (2017): 3-15.
16. Konovalova, Anna, Tobias Petters, and Lotte SØgaard-Andersen. "Extracellular biology of *Myxococcus xanthus*." *FEMS microbiology reviews* 34.2 (2010): 89-106.
17. Sun, Hong, and Wenyan Shi. "Analyses of mrp genes during *Myxococcus xanthus* development." *Journal of Bacteriology* 183.23 (2001): 6733-6739.
18. Sun, Hong, and Wenyan Shi. "Genetic studies of mrp, a locus essential for cellular aggregation and sporulation of *Myxococcus xanthus*." *Journal of Bacteriology* 183.16 (2001): 4786-4795.
19. Robinson, Mark, et al. "Transcription factor MrpC binds to promoter regions of hundreds of developmentally-regulated genes in *Myxococcus xanthus*." *BMC Genomics* 15 (2014): 1-23.
20. Ogawa, Mari, et al. "FruA, a putative transcription factor essential for the development of *Myxococcus xanthus*." *Molecular microbiology* 22.4 (1996): 757-767.
21. Giglio, Krista M., Jessica Eisenstatt, and Anthony G. Garza. "Identification of enhancer-binding proteins important for *Myxococcus xanthus* development." *Journal of Bacteriology* 192.1 (2010): 360-364.

22. Kroos, Lee. "Highly signal-responsive gene regulatory network governing Myxococcus development." *Trends in genetics* 33.1 (2017): 3-15.
23. Davis, Maria C., et al. "The essential activities of the bacterial sigma factor." *Canadian Journal of Microbiology* 63.2 (2017): 89-99.
24. Weber, Harald, et al. "Genome-wide analysis of the general stress response network in Escherichia coli: σ S-dependent genes, promoters, and sigma factor selectivity." *Journal of bacteriology* 187.5 (2005): 1591-1603.
25. Paget, Mark S. "Bacterial sigma factors and anti-sigma factors: structure, function, and distribution." *Biomolecules* 5.3 (2015): 1245-1265.
26. Helmann, John D., and Michael J. Chamberlin. "Structure and function of bacterial sigma factors." *Annual review of biochemistry* 57.1 (1988): 839-872.
27. Davis, Maria C., et al. "The essential activities of the bacterial sigma factor." *Canadian Journal of Microbiology* 63.2 (2017): 89-99.
28. Österberg, Sofia, Teresa del Peso-Santos, and Victoria Shingler. "Regulation of alternative sigma factor use." *Annual review of microbiology* 65 (2011): 37-55.
29. Fimlaid, Kelly A., and Aimee Shen. "Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes." *Current Opinion in Microbiology* 24 (2015): 88-95.
30. Mearls, Elizabeth B., et al. "Transcription and translation of the sigG gene is tuned for proper execution of the switch from early to late gene expression in the developing Bacillus subtilis spore." *PLoS Genetics* 14.4 (2018): e1007350.

31. Gonzalez, Alfonso, et al. "A tripartite, hierarchical sigma factor cascade promotes hormogonium development in the filamentous cyanobacterium *Nostoc punctiforme*." *Msphere* 4.3 (2019): e00231-19.
32. Serrano, Mónica, et al. "A Negative Feedback Loop That Limits the Ectopic Activation of a Cell Type–Specific Sporulation Sigma Factor of *Bacillus subtilis*." *PLoS Genetics* 7.9 (2011): e1002220.
33. Elliot, Marie A., et al. "BldD is a direct regulator of key developmental genes in *Streptomyces coelicolor* A3 (2)." *Molecular microbiology* 40.1 (2001): 257-269.
34. Gotoh, Yasuhiro, et al. "Two-component signal transduction as potential drug targets in pathogenic bacteria." *Current opinion in microbiology* 13.2 (2010): 232-239.
35. Karthikeyan, Rajapandian SivaGanesa, et al. "Host response and bacterial virulence factor expression in *Pseudomonas aeruginosa* and *Streptococcus pneumoniae* corneal ulcers." *PloS one* 8.6 (2013): e64867.
36. Inouye, S. U. M. I. K. O. "Cloning and DNA sequence of the gene coding for the major sigma factor from *Myxococcus xanthus*." *Journal of Bacteriology* 172.1 (1990): 80-85.
37. Apelian, D. A. V. I. D., and S. U. M. I. K. O. Inouye. "A new putative sigma factor of *Myxococcus xanthus*." *Journal of Bacteriology* 175.11 (1993): 3335-3342.
38. Ueki, Toshiyuki, and Sumiko Inouye. "SigB, SigC, and SigE from *Myxococcus xanthus* Homologous to sigma³² are Not Required for Heat Shock Response but for Multicellular Differentiation." *Journal of molecular microbiology and biotechnology* 3.2 (2001): 287-294.

39. Viswanathan, Poorna, Mitchell Singer, and Lee Kroos. "Role of σ^D in regulating genes and signals during *Myxococcus xanthus* development." *Journal of Bacteriology* 188.9 (2006): 3246-3256.
40. Yoder, Deborah R., and Lee Kroos. "Mutational analysis of the *Myxococcus xanthus* $\Omega 4400$ promoter region provides insight into developmental gene regulation by C signaling." *Journal of Bacteriology* 186.3 (2004): 661-671. Yoder, Deborah R., and Lee Kroos. "Mutational analysis of the *Myxococcus xanthus* $\Omega 4499$ promoter region reveals shared and unique properties in comparison with other C-signal-dependent promoters." *Journal of bacteriology* 186.12 (2004): 3766-3776.
41. Taylor, Rion G., and Roy D. Welch. "Recording multicellular behavior in *Myxococcus xanthus* biofilms using time-lapse microcinematography." *JoVE (Journal of Visualized Experiments)* 42 (2010): e2038.
42. Ratcliff, William C., et al. "Experimental evolution of multicellularity." *Proceedings of the National Academy of Sciences* 109.5 (2012): 1595-1600.
43. Higgs, Penelope I., et al. "EspA, an orphan hybrid histidine protein kinase, regulates the timing of expression of key developmental proteins of *Myxococcus xanthus*." *Journal of Bacteriology* 190.13 (2008): 4416-4426.
44. Sonnhammer, Erik LL, Sean R. Eddy, and Richard Durbin. "Pfam: a comprehensive database of protein domain families based on seed alignments." *Proteins: Structure, Function, and Bioinformatics* 28.3 (1997): 405-420.
45. Apweiler, Rolf, et al. "InterPro—an integrated documentation resource for protein families, domains and functional sites." *Bioinformatics* 16.12 (2000): 1145-1150.

46. Finn, Robert D., et al. "The Pfam protein families database." *Nucleic acids research* 38.suppl_1 (2010): D211-D222.
47. Tatusov, Roman L., Eugene V. Koonin, and David J. Lipman. "A genomic perspective on protein families." *Science* 278.5338 (1997): 631-637.
48. Altschul, Stephen F., et al. "Basic local alignment search tool." *Journal of molecular biology* 215.3 (1990): 403-410.
49. Mulder, Nicola J., and Rolf Apweiler. "Tools and resources for identifying protein families, domains and motifs." *Genome biology* 3.1 (2001): 1-8.
50. Camon, Evelyn, et al. "The gene ontology annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro." *Genome research* 13.4 (2003): 662-672.
51. Martin, David, et al. "GOToolBox: functional analysis of gene datasets based on Gene Ontology." *Genome biology* 5.12 (2004): 1-8.
52. Waterhouse, Andrew, et al. "SWISS-MODEL: homology modelling of protein structures and complexes." *Nucleic acids research* 46.W1 (2018): W296-W303.
53. Bradley, Michael D., et al. "Inter-laboratory evolution of a model organism and its epistatic effects on mutagenesis screens." *Scientific Reports* 6.1 (2016): 38001.
54. Kaiser, Dale, and Cathy Crosby. "Cell movement and its coordination in swarms of *Myxococcus xanthus*." *Cell Motility* 3.3 (1983): 227-245.
55. Berleman, James E., et al. "Predataxis behavior in *Myxococcus xanthus*." *Proceedings of the National Academy of Sciences* 105.44 (2008): 17127-17132.

56. Rao, Xiayu, et al. "An improvement of the $2^{-\Delta\Delta CT}$ method for quantitative real-time polymerase chain reaction data analysis." *Biostatistics, bioinformatics, and biomathematics* 3.3 (2013): 71.
57. Ueki, Toshiyuki, et al. "SigF, a new sigma factor required for a motility system of *Myxococcus xanthus*." *Journal of bacteriology* vol. 187,24 (2005): 8537-41.
doi:10.1128/JB.187.24.8537-8541.2005
58. Kashefi K., Hartzell P. L. 1995. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* frzF- defect. *Mol. Microbiol.* 15:483–494

CHAPTER 4: CONCLUSIONS

A. SUMMARY

The significance of biofilms as complex, multicellular bacterial communities that can play various roles in nature, including nutrient cycling, bioremediation, and disease, cannot be overstated. Characterizing behaviors from biofilms like *Myxococcus xanthus* is of immense biological importance because it can help us understand how complex intracellular and intercellular signaling models work and compare them to higher-complexity multicellular models. *M. xanthus*, a social bacterium that forms highly organized fruiting bodies and moves through its environment in a coordinated manner, provides an excellent model organism for studying biofilm formation and cellular differentiation.

The genotype-to-phenotype relationship is a crucial aspect of understanding how genes and genetic variation relate to observable traits in biofilms. To address this challenge, researchers have developed comparative models for assessing the genetic mechanisms leading to different phenotypes for the same mutation in different genetic backgrounds. These models can help us understand how genes and genetic variation contribute to biofilm phenotype and identify how biofilms are regulated. The significance of genotype-to-phenotype modeling in biofilms lies in its potential to improve our understanding of the complex relationship between genes and observable traits. Recognizing and quantifying indistinguishable, or otherwise identical, phenotypes are important for genome annotation and understanding. For example, in the context of human health, incomplete penetrance refers to a phenomenon where a person with a genetic mutation that predisposes them to a particular disease may not actually develop the disease, despite carrying the mutation. Variable expressivity refers to the phenomenon

where the same genetic mutation may present to different degrees in different individuals. By identifying genetic mechanisms underlying biofilm formation and maintenance, researchers can develop more effective strategies for biofilm assembly and understand more about the complex genotype-to-phenotype associations.

The discussion presented in this thesis focuses on three unique approaches to characterizing dynamic behaviors in biofilms. Project 1 highlights how detailed characterization of stochastic variation in phenotype can be used to differentiate otherwise identical appearing phenotypes due to different gene mutations. Project 1 characterizes the wild-type DK1622 *M. xanthus* strain which is vital for gene annotation-based studies and incorporates more parameters for phenotypic comparison. Project 2 investigates a traditional approach to understanding emergent behaviors in *M. xanthus* by studying transcriptional regulators and highlighting their role during development. Project 2 characterizes 58 mutant stress response genes and distinguishes biologically significant mutants using quantitative phenotype assays.

In conclusion, the study of biofilm dynamics in *Myxococcus xanthus* is significant due to its implications for genetic and data modeling research. Understanding the dynamics of biofilm formation in this organism can shed light on the genetic mechanisms involved, including how the expression of genes affects biofilm development. Furthermore, data modeling approaches can be used to identify key variables and parameters that influence biofilm formation, enabling the development of predictive models for biofilm behavior. Overall, the findings from this research can have important implications for fields such as microbiology, biotechnology, and biomedical engineering, where biofilms play a critical role in a variety of biological processes and disease states. By advancing our understanding of how biofilm genomes regulate complex

behaviors, we can pave the way for more targeted and effective interventions in these fields, leading to better health outcomes and improved quality of life for individuals and communities.

Eduardo A. Caro

www.linkedin.com/in/eduardo-a-caro

eacaro@syr.edu

SUMMARY

A research scientist well-versed in characterizing differences in biofilm growth, specializing in microbiological and molecular lab techniques. Additional, expertise in managing, teaching, and planning scientific projects for data-driven modeling.

EDUCATION

SYRACUSE UNIVERSITY, SYRACUSE, NY

Ph.D. in Biology

- Microbiology, Systems Biology, and Molecular Biology expertise
- Seminar Courses taken: Statistical Analysis, Science Outreach in Biology, Infectious Diseases, Molecular Genetics, Public Speaking for Graduate Researchers
- Teaching Assistant for Introduction to Biology, Cellular Biology, Microbiology, and Biotechnology courses

UNIVERSITY OF CENTRAL FLORIDA, ORLANDO, FL

Bachelor in Biomedical Sciences, Minor in Business Administration

- Honor's College Alumni and Honor's College Team Leader

ACADEMIC PROJECTS

A 'Common Garden' Approach to σ Factors important for *Myxococcus xanthus* Development

- Presented research findings at 'The 46th International Meeting on the Biology of the Myxobacteria', Houston, TX
- Performed Plasmid preparation, Ligation, Transformation, Genomic DNA extraction, Total RNA extraction, RT-PCR, PCR, Site-directed mutagenesis, DNA sequencing using CLC workbench, Gel Electrophoresis

Investigating Fruiting Body Morphogenesis in *Myxococcus xanthus* as a Genetic Tool for Improving Machine Learning using StyleGan2

- Managed a team of 3 Undergraduate researchers to generate a 10 TB phenotype data library
- Created a workflow for mass data collection to efficiently share and analyze with collaborators
- Conducted Microscopy and Imaging, Aseptic technique, Quality Control, Development Assays
- Transcriptome analysis using rRNA depletion and Illumina Sequencing

GEAR Program: Group Effort Applied Research ([doi: 10.1002/bmb.20802](https://doi.org/10.1002/bmb.20802))

- Collaborated as 1 of 6 for a pilot study to improve Undergraduate research strategies
- Research skills: Report writing, Data collection, Analysis of information from different sources, Critical thinking, Collaborative project planning, Conflict Resolution

PROFESSIONAL EXPERIENCE

BIOLOGY GRADUATE STUDENT SEMINAR COORDINATOR, SYRACUSE UNIVERSITY

SYRACUSE, NY [AUGUST 2020 - MAY 2021]

- Facilitated professional development with 6 guest speakers for ethics, research, and data management
- Provided mentorship and support for 12 incoming graduate students for classes and program requirements

- Maintained quality and consistency of instruction through research presentation feedback and team building for a class of 35

LAB TECHNICIAN, SGS LABS

ORLANDO, FLORIDA [AUGUST 2016 - MAY 2017]

- Delivered effective team management and quality control for a team of 5 students
- Prepared, logged, and processed daily operational duties of quality assurance, verification, and proficiency test samples containing biological agents
- Provided data entry utilizing a computerized system for daily sample collection and processing using LIMS

PUBLICATIONS

Unraveling a Bacterial Starvation Response Through the Direct Targets of a Starvation-Induced Transcriptional Activator. Muqing Ma, Ting Li, David J. Lemon, Eduardo A. Caro, Linnea Ritchie, Charles Ryan, Victoria M. Spearing, Roy D. Welch, Anthony G. Garza, Kimberly A. Murphy doi: <https://doi.org/10.1101/2021.11.28.470282>

Stochastic Bounds of Aggregation Dynamics Distinguish Near-wild-type from Wild-type strains in social bacteria. Merrill E. Asp*, Eduardo A. Caro*, Roy D. Welch, Alison E. Patteson (*Co-First Author) doi: <https://doi.org/10.1101/2023.02.26.530117>

OTHER TECHNICAL SKILLS

Teaching skills: Classroom Instruction, Management, Syllabus Organization, 3D Printing, Soldering, Computer Building

Other: Spanish (Native), Microsoft Office, Minitab, Graphpad, Bioinformatics using Kegg, NCBI, and UniProt, Fiji, FTP operating WinSCP, Public Speaking, Teamwork