8-26-2022

REDUNDANCY AND DEVELOPMENTAL COORDINATION IN MYXOCOCCUS XANTHUS: INSIGHTS FROM THE PHENOME

Jessica A. Comstock
Syracuse University, jacomsto@syr.edu

ABSTRACT

Genotype-to-phenotype mapping can typically involve disrupting the function of a gene and observing the impact of mutation on phenotype. While this can be a powerful tool for uncovering gene function, complicating factors such as influences between genes and the environment, epistatic interactions with other genes, and genetic redundancy could all potentially mask the phenotype of a mutation such that functional inferences cannot be made. On the level of the single gene, this may not be particularly informative, but it is possible that studying phenotype in this way at a genome scale might allow for the observation of patterns between genes and their associated mutant phenotypes that can inform the genotype-to-phenotype space. Phenotypic profiling, involving high-throughput phenotyping across different genetic backgrounds and environments, has the potential to inform our understanding of how different genes interact with the environment and each other.

Myxococcus xanthus is a soil bacterium with a relatively large genome that lives a remarkably multicellular lifestyle for a prokaryote. Under starvation, cells of *M. xanthus* aggregate into clusters that eventually mature into spore-filled fruiting bodies. This is a complex phenotype for which we can observe and quantify multiple features, providing a landscape of features that can be used to measure the effect of mutation on phenotype. In this thesis, I first explore how high-throughput phenotypic observations reveal a pattern of widespread genetic redundancy by demonstrating that mutant strains within the same gene family are more phenotypically similar to one another than they are to those outside of their gene family. I further show that genotypic similarity and phenotypic similarity do not correlate well on a finer scale, indicating that genotype alone is not a good predictor of genetic redundancy.

Next, I characterize a wave phenomenon that is observed during time-lapse movies of *M. xanthus* fruiting body formation. The oscillatory behavior seen within the swarm, which we here call pulsing, was striking and was initially thought to be a rare mutant phenotype until it was observed frequently within a database of thousands of time-lapse movies of different mutant strains. It was then quantified through image analysis and found to exist in wild-type strains as well, at the same frequency that we observed pulsing in genetic mutants. We found that pulses are waves that originate at early fruiting bodies and propagate through the starving swarm, causing individual cells to suppress reversals and travel more persistently for a longer duration. This serves as a potential mechanism to aid in the rate of aggregation and represents possible inter-aggregate communication.

Finally, I explore cell behaviors related to the coarsening phase of *M. xanthus* aggregation wherein larger aggregates remain stable and mature into fruiting bodies and smaller aggregates disperse back into the swarm. I collected cell tracking data that will inform a data-driven model to improve understanding of the individual cell behaviors that lead to small aggregate dispersal during the coarsening phase. These experiments culminated in an approximately 85% loss of fluorescent cells used for tracking that did not occur in samples without fruiting bodies. This cell loss co-occurs with an increase in propidium iodide staining of fruiting bodies, indicating an increase in extracellular DNA that could be associated with cell lysis.

Notably, the phenomena described in all three chapters were observed from the same set of high-throughput phenotype data. Characterizing the phenome by collecting this type of data can improve our understanding of biological patterns.

REDUNDANCY AND DEVELOPMENTAL COORDINATION IN MYXOCOCCUS

XANTHUS: INSIGHTS FROM THE PHENOME

by

Jessica A. Comstock

B.S., Westfield State University, 2013

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biology.

Syracuse University

August 2022

ACKNOWLEDGEMENTS

It is not an exaggeration to say that I would not be where I am today without the support of so many people that have come to be a huge part of my life. Firstly, I'd like to thank my advisor, Dr. Roy Welch, for being a wonderful mentor and sharing my interest in exploring all things weird and cool. He provided me with the latitude to pursue my interests and to constantly challenge myself, for which I will be forever grateful. I have thoroughly enjoyed my time working in his lab and appreciate how his mentorship has made me a more creative, independent thinker and communicator of science. For encouraging healthy scientific debate, respecting my ideas and opinions, and being an unfailing source of professional advice and support, thank you.

I have been fortunate during my time at Syracuse University to work with several fantastic collaborators. I am privileged to have had the opportunity to work with Dr. Alison Patteson and Merrill Asp in the Physics Department at Syracuse, two wonderful collaborators whose efforts and quantitative expertise were invaluable to advancing the work in this thesis. Dr. Oleg Igoshin and his lab members Patrick Murphy and Jiangguo Zhang at Rice University have also been wonderful collaborators. I want to thank them for their great work and exciting conversations about the intersection of biology, physics, and engineering, and for teaching me so much about the power of computational modeling. These collaborations have pushed me out of my research comfort zone and have really influenced the direction of my work, and for that I am very grateful.

I want to thank my research committee, Dr. Eleanor Maine, Dr. Steve Dorus, and Dr. Scott Erdman, for their support, ideas, and suggestions over the past years. Each has provided a unique perspective on my research and provided critical assessments of my work that have

years. Thanks to Chelsea Hoffman, Trang Nguyen, and Jessica O'Connell, who have participated in troubleshooting the early stages of these projects and have all gone on to do wonderful things.

To my family, thank you all for your unwavering support over the years. I would not be where I am today without my parents, who for my entire life have modeled the work ethic and stubbornness that got me through my PhD. Thank you, Sarah, for being my constant cheerleader and a supportive ear for my frustrations. Thank you, Dylan, for your unfailing support and belief in me during times when I didn't believe in myself. And thank you to Tina and Vicky, who have been dear friends for as long as I can remember and have been endless providers of support. You all mean the world to me and I cannot thank you enough.

Finally, I would like to acknowledge all the people within the scientific community with whom I have interacted, both positively and negatively, that have informed the kind of scientist, mentor, and person that I want to be. Science and social justice are inextricably linked, and you cannot separate the science that you do from who you are as a person. Some interactions have reinforced that there are ways to practice equitable and inclusive science while valuing diverse opinions, and some interactions have taught me that there is still a lot of work to be done within the scientific community to achieve that goal, and that this work must be intentional. Both types of interactions have shaped the way that I will move forward in my career as a scientist and mentor to future generations of scientists.

TABLE OF CONTENTS

LIST OF ILLUSTRATIVE MATERIALS

CHAPTER 1: INTRODUCTION

PHENOMICS AND THE GENOTYPE TO PHENOTYPE PROBLEM

The Genotype-to-Phenotype Problem

Genome sequencing is a powerful tool for improving our understanding of the way that genes contribute to phenotype. Continually improving technologies and decreasing costs have made sequencing data more accessible than ever. The promise of the unexplored within in sequenced genomes is enticing, but the amount of sequencing data we can collect is greatly outpacing our ability to attribute phenotype to specific genes. The prevalence and availability of sequence data is an invaluable tool for manipulating genomes, searching for homology among better-studied orthologs, and finding associations between alleles and human diseases in populations (1). However, functional annotation of every gene in the genome, one intended goal of pioneering genome sequencing studies (2), seems a far-off goal given the difficulties in mapping genotype to phenotype.

In model systems, we predominantly ascribe function to a gene through mutation, deleting or modulating its expression and recording the impact on phenotype in reference to a wild-type strain. Any deviation in phenotype has the potential to inform the function of that gene. While this is a powerful approach when an obvious phenotype results, there are complicating factors such as redundancy, epistasis, pleiotropy, and environmental effects that can obscure the effect of mutation.

Frequently, mutant phenotypes are indistinguishable from wild-type and therefore uninformative for functional annotation. This is particularly evident when measuring fitness effects of mutations. For example, a yeast deletion study measuring growth showed that only 20% of genes were required for viability (3). Similar and increasingly complex studies that use double and triple mutants to map the functional genome have found extensive digenic and

trigenic interactions that can reduce the likelihood that a single mutation would cause a discernible phenotype, and increase the complexity of any inferences about one specific gene made based on an observable phenotype (4, 5). This pattern of extensive genetic interaction is also applicable to prokaryotes, though they are often thought to be less complex (6, 7). Further, epistatic interactions complicate our understanding of the role of mutations in phenotypes ranging from human disease (8) to laboratory bacterial strains (9).

Genetic redundancy is also likely a major factor in the ubiquity of mutational robustness that is widely distributed through most characterized genomes. Many proteins are part of larger families with multiple paralogs that could potentially contribute to functional redundancy at the molecular level. Though gene duplication is one of the mechanisms by which organisms obtain redundant genes, horizontal gene transfer is also common among microbial systems (10). Redundant genes can increase fitness by providing "backup" mechanisms for essential processes or virulence factors. For example, the *fepA, ihaA, and iroN* outer membrane iron receptors of *E. coli,* which are virulence factors that allow for the colonization of the urinary tract, form an array on the outer membrane and work in tandem such that a deleterious deletion in one does not impact fitness in the urinary tract environment (11). Since gene duplication and horizontal gene transfer are so common in bacteria (12–15), they are a good model system for studying redundancy, especially since they lack the complications of tissue-specific expression, for example the potential for differential spatiotemporal expression of redundant genes in different cell types, found in higher organisms.

Influences of environmental factors and developmental contexts also play a significant role in the manifestation of phenotype. Particularly for bacteria, the expression of some genes is context-dependent, for example the presence of lactose in the environment triggers induction of

the genes required for lactose digestion (16). Therefore, a phenotype would only be observed if cells were relying on lactose as a primary nutrient source. This kind of phenotyping requires a prior hypothesis for the function of the gene, which is not always obvious. Bacteria are constantly sensing and responding to their environments, so they have a bounty of genes that are expressed only in explicit environmental conditions, necessitating that the phenotype of mutations is observed across a number of environments (17, 18). Characterizing all genes in all relevant conditions presents obvious challenges to the parameters of experimental design.

Toward Characterizing the Phenome

The above-mentioned issues with mapping genotype to phenotype shed light on a need to supplement single-gene mutation studies with more data, setting observed phenotypes into the context of the phenome, or the collective set of phenotypes expressed by an organism in different environments and genetic backgrounds (19). Studies attempting to attribute function to a single gene or pathway would be greatly improved by observing the resulting phenotypes within the broader context of the phenome (or at least a growing collection of phenotype data that is working toward that goal). For example, in Chapter 2 of this thesis, I present a phenotype that occurs in over half of the mutants assayed, which could have been falsely attributed to a particular genetic pathway without the sufficiently large library of mutant phenotype data that we have collected and quantified.

At its heart, phenomics requires amassing large quantities of data, finding effective ways to quantify it, and creating public databases for sharing of phenome data. Almost immediately after the Human Genome Project was published in 2003, the idea of the Human Phenome Project was generated to map similarities in phenotype and genotype across populations (19). PhenGenI, the Phenotype-Genotype Integrator, is a phenotype-based tool for both researchers and clinicians

to explore variants detected in genome-wide association studies with their corresponding phenotypes (20) Similarly, the Mouse Phenome Database, Flybase, and the Saccharomyces Genome Database link genotype to phenotype (21–23). Predominantly the goal of these projects, aside from genotype-to-phenotype mapping, is to identify associations between mutations and disease, or developmental phenotypes. In bacteria, phenomics studies are starting to be employed more frequently. Synthetic genetic array methods developed using the yeast model system *S. cerevisiae* have now been adapted to *E. coli* to study gene interactions and begin to construct functional network maps (6, 24). Phenotype microarrays can detect metabolic usage of nutrient sources and can detect sensitivities to particular drugs or chemicals in a high-throughput fashion (25), and phenotypic profiling of strains, particularly in differing environmental contexts, is becoming more common (17, 18, 26–28).

An additional benefit to large phenotypic datasets is the ability to use them to observe broader biological patterns and answer larger questions. For example, observing cell behavior across a wide range of conditions can reveal patterns in the distribution of phenotypes. In the soil bacterium *M. xanthus*, several studies that quantified phenotypes of mutant libraries revealed that the lab wild-type strain falls consistently within the middle of the phenotype distribution for most assays (17, 28), which brings up interesting questions about the costs associated with increased motility rate or sporulation efficiency, for example. Though we might expect that wild-type strains have evolved optimal fitness, phenomics studies reveal that this is often not the case, and very recent work has shown that wild-type alleles are often less fit but more mutationally robust (29). In this way, phenotype can be a powerful tool not only for functional annotation, but also for observing and uncovering broader biological patterns.

M. XANTHUS AS A MODEL ORGANISM

Though bacteria are unicellular and were long thought to exist predominantly as individuals, research in recent decades has shown that a significant fraction of terrestrial, aquatic, and pathogenic organisms exist in multicellular communities called biofilms (30, 31). The canonical definition of a biofilm as a surface-attached community of bacteria was coined by J. William Costerton in the late 1970s, spurring an influx of research into bacterial biofilms. Within these communities, there are localized microenvironments that trigger different cellular responses (32), specialized cell types that take on various roles (33), and spatial and structural organization mediated by polysaccharide matrices that facilitate interconnectivity between cells (34). In 1988, Shapiro published a foundational paper drawing parallels between developing multicellular organisms and these properties of biofilms, which has since shifted the perspective in the field of microbiology toward thinking of communities of bacteria as multicellular organisms (35). Though it had been known for quite some time that microbes such as Myxobacteria*, Proteus,* and *Streptomyces*, existed in multicellular communities, they were thought to be rare exceptions to the rule of bacterial individuality (36). Studies of bacterial social interactions now dominate the field, with an understanding that anywhere from 50-80% of known bacterial species exist as biofilms at some point in their life cycle or under certain environmental stressors (32).

Myxococcus xanthus has emerged as a popular model system for studying bacterial social interactions and self-organization into multicellular structures, as it almost exclusively exists as a multicellular community (37). *M. xanthus* is a δ-Proteobacterium that thrives in nutrient-rich soil environments by saprophytically feeding on decaying plant material and other nitrogenous waste products across temperate and tropical environments, but has also been isolated from the harsher

environments of wet rocky surfaces and sand (38). In vegetative conditions where nutrients are

abundant, *M. xanthus* exists as a swarm where cells are tethered together by pili and the

polysaccharide fibrils that comprise their extracellular matrix (ECM), facilitating group motility

(39, 40). These swarms are also predatory, and upon encountering prey bacterial colonies such as

E. coli and other soil microbes, they secrete enzymes that lyse prey cells and scavenge their

nutrients (41).

Life in the soil can be harsh due to frequent fluctuations in moisture, temperature, local

chemical concentrations, and most significantly, nutrient availability. *M. xanthus* cells undergo a

developmental program upon sensing nutrient stress. Cells transition away from behaviors that

lead to swarm expansion and begin to aggregate into mounds of approximately 10^5 cells (42). As

development continues, the initial mounds of cells mature into fruiting bodies wherein cells meet

one of two fates: programmed cell death, where cells lyse to release nutrients into the aggregate,

supporting the transition of the remaining cells into the other cell fate, environmentally-resistant

spores (43, 44). The majority of cells in a fruiting body are thought to undergo lysis, while only

an estimated 10% differentiate into spores (45). Though only a small fraction, this represents

hundreds to thousands of spores that are bundled together in fruiting bodies, protected from

harsh environmental conditions until local nutrient conditions become more favorable, or until

the adhesive fruiting body is transported by an insect, animal, or streams of water to a nutrient-

rich location, where germination into vegetative cells can occur. Notably, due to the inherent

multicellular nature of *M. xanthus*, fruiting bodies present a mechanism by which groups of

spores can germinate together to form a vegetative swarm, increasing chances of survival once

nutrients are plentiful (46). An additional cell fate, peripheral rods, never join fruiting bodies,

show low expression levels of developmental regulatory genes, and never differentiate into

spores (47). The cells are often compared to persister cells, subpopulations that employ a strategy of dormancy to survive antibiotic treatment (48), though the mechanistic basis is different. A proposed role of these persister-like peripheral rods is to sense and quickly respond to incoming nutrients, potentially signaling the remainder of the swarm to halt development or initiate germination (49).

Because of the extent of multicellularity in its lifecycle, *M. xanthus* serves as a model system for self-organization, helping to uncover both the physical and biological mechanisms that drive cells into ordered multicellular patterns. The following sections will summarize what is currently known about the biology of *M. xanthus* that brings about these emergent behaviors, with a particular emphasis on the genetic networks that lead to the development of fruiting bodies in response to nutrient stress.

THE *M. XANTHUS* GENOME

The complexity of molecular interactions that leads to the multicellular nature of *M. xanthus* is likely encoded into the genome. The myxobacterial lineage has roughly double the genome size relative to other δ-Proteobacteria (50), and much of this expansion is thought to have arisen from lineage specific duplications (51). The distribution of duplicated genes throughout the genome is asymmetric with respect to functional roles, as duplication of genes with roles in signal transduction and transcriptional regulation were found to be significantly greater than expected (51). The selective pressure to maintain duplications leading to these expanded gene families in *M. xanthus* implies their potential roles in the increased complexity of its multicellular lifestyle compared to less complex δ-Proteobacterial ancestors.

Specifically, the *M. xanthus* genome encodes more two-component signal transduction systems and fewer one-component signal transduction systems than expected for the size of its

genome (51). Bacteria overwhelmingly encode a greater number of one-component systems for

signal transduction (52), but the multi-component signal transduction networks that seem to

dominate in *M. xanthus* are more reminiscent of eukaryotic signal transduction and

transcriptional networks that regulate multicellular organisms (53). These multi-component

networks often form parallel pathways and are capable of integrating signals and directing

transcription in response to a combination of inputs. Transcriptional regulation in *M. xanthus*

also occurs through extracytoplasmic function (ECF) sigma factor systems, which canonically

respond to extracellular signals and direct RNA polymerase to transcribe specific genes (54).

Large gene families such as these, and the ATP-binding cassette (ABC) transporter family which

translocate substrates across the periplasmic space, demonstrate the ability of *M. xanthus* to

sense and respond to conditions in both its internal and external environment. I review each of

these gene families below given their relevance to the research presented in Chapter 2.
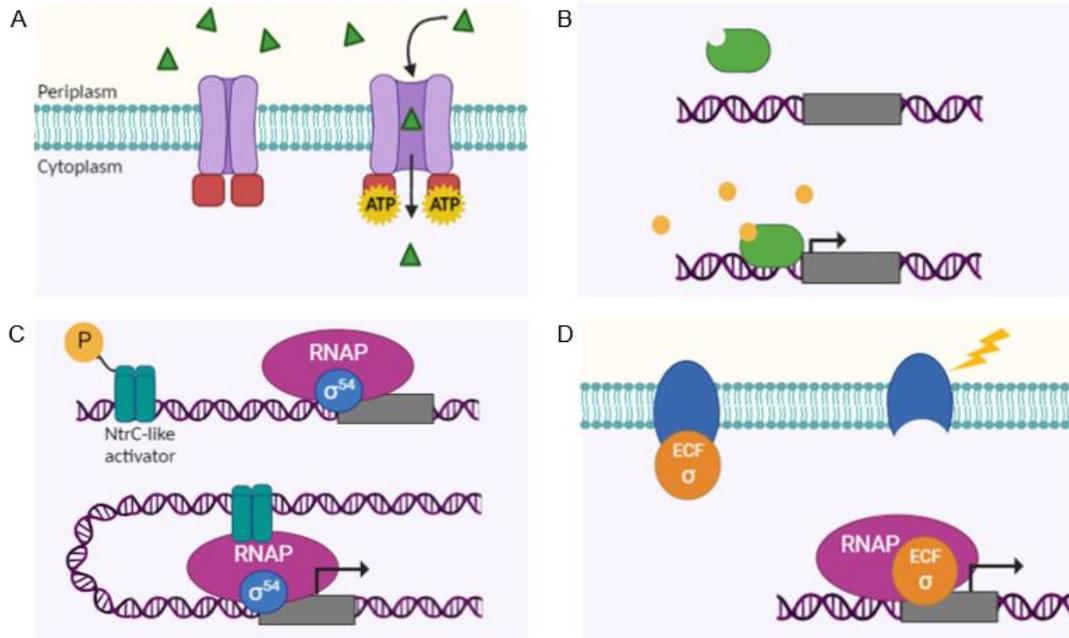
Figure 1: Gene families in *M. xanthus* relevant to the work presented in Chapter 2. A) ATP-binding cassette (ABC) transporters are positioned in the inner membrane of Gram-negative bacteria and serve as ATP-dependent transporters of substrates across membranes. These can be importers or exporters. These transporters are also sometimes associated with substrate-binding proteins that shuttle substrates to the transporter. B) One Component signal transduction proteins (green) are typically ligand binding and will act as transcription factors to initiate expression of genes in response to binding ligands. C) NtrC-like activators (teal) are activated via phosphorylation by a kinase partner in response to an intracellular cue. They form a looping structure in the DNA upstream of σ^{54} promoters to form an open promoter complex, allowing RNA polymerase to initiate transcription at these promoters. D) Extracytoplasmic function (ECF) sigma factors (orange) are typically sequestered at the membrane via anti-sigma factors (blue) and are released upon reception of an extracellular signal. These free sigma factors can then direct transcription of genes with the corresponding promoters.

One-Component Signal Transduction Systems

Primarily, one-component signal transduction systems (OCSTSs) involve directing transcription in response to small molecule sensing. Structurally, they tend to be composed of an input sensor domain that is ligand-binding and a DNA-binding output response domain, encoded on the same protein (Fig. 1B) (52). These proteins are often referred to as transcriptional activators rather than signal transduction systems given that they a single component "system". Well-studied examples of OCSTSs are AraC in *E. coli*, which activates transcription of the *araBAD* operon in the presence of arabinose (55, 56), and TraR in *A. tumefaciens,* which

activates expression of its tumor-inducing plasmid in response to sensing a population density-dependent inducer (57). Relatively little is known about the role of this family of proteins in *M. xanthus*, aside from associations based on orthologs in other systems. Though it encodes around 300 of these genes, that is approximately half the number expected based on its genome size (51).

Two-Component Signal Transduction Systems

Traditional two-component signal transduction systems (TCSTSs) consist of a sensor histidine kinase (HK) and a cognate response regulator (RR) partner. Upon sensing a particular signal via its ligand-binding (or other sensory) domain, autophosphorylation of the HK occurs (58). Following this activation, phosphotransfer to the receiver domain of the RR, activating a downstream effector. The EnvZ/OmpR system is a well-studied example of a TCSTS. The EnvZ HK is situated in the membrane and autophosphorylates due to high osmolality or pH in the surrounded medium. It phosphorylates its RR partner OmpR, which initiates transcription of *ompC*, a porin that helps to regulate flux (59). Another classic example is the CheA/CheY system involved in chemotaxis in *E. coli* and many other bacteria. The binding of a repellent to the methyl-accepting chemotaxis protein in the membrane triggers autophosphorylation of the HK CheA, which initiates phosphotransfer to the RR CheY, triggering a reversal in direction of the flagellar motor (60).

This type of signaling system, particularly the NtrC-like activator family, has been the subject of investigation in *M. xanthus* (61–63). NtrC-like activators are a class of RR that are responsible for activating gene regulation at σ^{54} promoters (Fig. 1C). NtrC-like activators, also known as enhancer binding proteins (EBPs), bind to enhancer elements upstream of σ^{54} promoters after activation by their HK partner, and form a looping mechanism with σ^{54}-RNAP

bound at the promoter site (64, 65). ATP hydrolysis by the ATPase domain of the NtrC-like

activators completes the transition from a closed to open promoter construct, allowing the

transcription of σ^{54}-dependent genes (65). Mutational analysis of these genes in *M. xanthus*

revealed many with a role in development (61), and suggests that these genes act in regulatory

cascades (63). Nla4 and Nla18 regulate the transition into starvation (62, 66), including by

initiating transcription of Nla6 and Nla28, the next genes in the cascade. These genes seem to

regulate one another, autoregulate, and initiate transcription of ActB and MXAN_4899, both of

which are NtrC-like activators that have roles in aggregation (63).

Extracytoplasmic Function Sigma (ECF) Factors

Sigma factors are responsible for directing transcription in bacteria by forming a

holoenzyme with the core RNAP for proper promoter recognition (67). The most predominant

sigma factors in *M. xanthus* are σ^{70} and σ^{54}, which are in this system both necessary for growth

(68). This is in contrast to many other bacterial systems where σ^{70} is considered the regulator of

"housekeeping" genes and σ^{54} exclusively directs accessory and/or stress response genes.

M. xanthus also contains 38 ECF sigma factors, a class of alternative sigma factors that

direct transcription in response to extracellular cues (69). ECF sigma factors are often

sequestered by a membrane-bound anti-sigma when inactive, and upon activation by an external

signal, the ECF sigma factor is released into the cellular environment, where it can initiate its

transcriptional response (Fig. 1D) (70). Though this is the most common regulatory paradigm for

controlling the activation of ECF sigma factors, there are others that follow a different

mechanism. Some ECF sigma factors respond to cytoplasmic signals, necessitating a

cytoplasmic anti-sigma that releases the sigma factor through a conformational change, and

others are not associated with anti-sigmas and are likely regulated through mechanisms that have

yet to be characterized (54). Well-studied ECF sigma factors in *M. xanthus* include the CarQ system, which responds to light induction, the CorE and CorE2 which are activated in the presence of metal ions (71, 72).

ATP-Binding Cassette (ABC) Transporters

The cell membrane is relatively impermeable to many molecules that are crucial for cellular processes. Yet others need to be actively transported across concentration gradients. ABC Transporters are often responsible for moving these substrates across membranes, acting as either importers or exporters, and have been shown to transport substrates of varying sizes and biochemical properties (Fig. 1A) (73). For example, in *E. coli*, FecBDCE transports ferric ions (74), BtuCDE imports vitamin B12 (75), and HlyB exports virulent hemolysin (76), demonstrating that substrates range from ions to small molecules to proteins. However, ABC Transporters are not limited to bacteria, though ABC importers do appear to be prokaryote-specific, and are in fact one of the largest and most widespread gene superfamilies (77). They are present in organisms from all orders of life, from bacteria to yeast to higher organisms such as worms, *Drosophila,* and humans (78–81).

Studies into the molecular organization of these proteins reveals that they are primarily made up of four subunits: 2 transmembrane domains, and 2 cytoplasmic nucleotide-binding domains that bind and hydrolyze ATP to provide energy necessary for active transport (82). Some prokaryotic transporters also have substrate-binding domains that are thought to play roles in substrate specificity. *M. xanthus* encodes 192 ABC Transporter ORFs, comprising a total of 57 complete transporters. Generally each subunit is encoded on its own gene, but some transmembrane/nucleotide-binding and transmembrane/substrate binding hybrids exist (28). The remaining ABC Transporter ORFs are part of incomplete operons or orphans. Though this might

indicate reduced selective pressure eliminating these genes, orphan ABC Transporter genes have been shown to interact as a component of transporters in different operons or have other functional roles (83, 84).

The specific roles of ABC Transporters in *M. xanthus* are largely unknown. AbcA, a transmembrane/nucleotide-binding hybrid protein, was shown to interact with FrzZ in a yeast-two-hybrid assay and was confirmed to produce a phenotype consistent with mutants of the Frz system that controls reversals (85). A homolog to a macrolide antibiotic transporter, Mac-1, was shown to be deficient in sporulation (86). It was originally suspected PilH formed a complex that is important for Type IV pilus assembly, as mutants were deficient in social motility and pilus production (87), and though this is now not known to be a major component of translocation of pilin during pilus assembly (88), its phenotype suggests that it does play an accessory role in social motility. Notably, a study by Yan et. al generated a mutant library of all ABC Transporter ORFs and characterized the phenotypes associated with these mutations (28). Rather than attributing the phenotypes to specific genetic pathways, this was presented an exploration of the correlation between metrics that we use for phenotyping strains. In all, though the ABC Transporter ORFs make up nearly 3% of the genome, we presently have relatively poor understanding of the specific roles of each transporter.

A genome such as *M. xanthus* that has undergone lineage-specific expansion due to duplication and divergence has families of dozens of genes that are sequence similar and thus have potential functional overlap. This source of functionally redundant genes could provide *M. xanthus* with the redundancy needed to insulate a crucial process such as sporulation from the potential dangers of mutation. We attempt to understand the extent of functional redundancy that occurs within these gene families in *M. xanthus*.

MOTILITY SYSTEMS

M. xanthus has two distinct motility systems for translocating across surfaces: Type IV Pilus-driven Social motility (S-motility), and Adventurous gliding motility (A-motility). These motility systems work in tandem to move cells along their long axis via motility engines located at the cell poles. S-motility is so named because the pili and exopolysaccharide (EPS) facilitate group movement whereby cells extend their Type IV Pili (TFP), attach to either the substrate or other cells, and retract the pili to pull themselves forward. Thus, the S-motility engine is localized to the leading pole of the cell. TFP-mediated motility is sometimes known as twitching motility, referring to the extension, tethering, and retraction of pili in contrast to flagellated movement in swimming bacteria (90). The A-motility engine propels the cell forward via a combination of focal-adhesion complexes distributed throughout the length of the cell and polysaccharide "slime" that mediates temporary adhesion to the substrate (91). It is referred to as adventurous motility because it does not require other the presence of other cells and is the mechanism by which individual cells leave the swarm edge. These motility systems are genetically distinct in that they are only connected by the polarity switching mechanism, and because cells with only one system are still motile, they can be studied independently. Because *M. xanthus* multicellularity is driven by motility, these systems are crucial for predation and fruiting body formation.

Social Motility

Cryo-electron tomography has solved the structure of the TFP in *M. xanthus* (92), depicting every component and its functional organization. Thus, the mechanistic understanding of the TFP-driven motility is very clear. The entire complex spans both the inner and outer membrane, with the PilQ secretin forming the outer pore that is anchored to the membrane via

the inner membrane ring structure comprised of other protein complexes. PilA monomers that will polymerize to form the pilus structure reside in the periplasmic space. PilB and PilT ATPases reside at the base of the pilus machinery and rotate in an ATP-dependent manner to polymerize or disassemble the PilA subunits, extending or retracting the pilus, respectively (92).

S-motility requires that cells are in close contact, as evidenced by mutants with only S-motility entering a nonmotile state when the distance between cells was greater than one cell length (39, 93). S-motility also requires secretion of exopolysaccharide (EPS) which coats the surfaces of cells, providing a point of attachment for the pili of neighboring cells (40, 94–96).

Adventurous Motility

Whereas S-motility coordinates movements of groups of cells, A-motility regulates the behavior of groups of cells at the edge of swarms. As these cells move away from the swarm edge they deposit a "slime trail" which other cells can follow (97, 98). The "slime" extrusion from the lagging pole of the cell was historically thought to propel the cell forward, but this has been experimentally disproven, and slime is now proposed to be an adhesion mechanism by which A-motile cells anchor themselves temporarily to their substrate via the more recently discovered mechanism discussed below (99–101).

The A-motility machinery involves focal adhesions which form between the cell and the substrate. Interestingly, these focal adhesions were regularly spaced and remained fixed in position with respect to the substrate as cells translocated (101). Focal adhesions are made up of clusters of Agl-Glt protein complexes which are directed around the helical cytoskeletal track of MreB, guiding the regular spacing and location of the focal adhesion clusters (102). The molecular motors that drive this rotation are the Glt proteins; they are distributed through the outer membrane, periplasm, inner membrane, and cytoplasm (99, 103–109). The periplasmic

components of the motor (GltD, GltE, and GltF) are thought to mediate connection between the outer membrane components of the motor (GltK, GltB, GltA, GltC, and GltH) and the inner membrane components (GltG and GltJ), creating a complex that spans the entire membrane. In the cytoplasm, GltI coordinates the association of the cytoskeletal MreB with the membrane-spanning complex (103, 106) via cytoplasmic proteins AglZ and MglA, which localize to the cytoplasmic side of this motor complex. The final component of the motor complex is the molecular motor itself, the AglRQS channel in the inner membrane that mediates the proton motive force (PMF) required for A-motility (99, 109). Several of these complexes are located along the length of the cell, along the MreB helical track, where they anchor to the slime on the substrate.

Together, the proposed mechanism is that MglA first localizes to the leading pole of the cell which is followed by MglA-AglZ localization at the focal adhesion-motor complexes along the length of the cell. The adhesive properties of the slime are thought to temporarily anchor the complex to the substrate while the PMF drives the cell forward with respect to these complexes, rotating around MreB (99, 100). Once the lagging pole of the cell catches up with a focal adhesion complex, it disassembles at the lagging pole, allowing for reassembly toward the leading pole and continued forward motion.

Maintaining Cell Polarity

Though these motility mechanisms are genetically distinct, a mutation in MglA disrupts both A- and S- motility (102). Its direct role in S-motility has not been elicudated, but its localization first to the leading pole in A-motility suggests it is important for maintaining cellular polarity, which could indirectly affect S-motility as well. MglA is suspected to play a role in the proper sorting of the pilus machinery to the correct poles(110) MglA localization to the leading

pole of the cell seems to be driven by the competitive binding and localization of MglB to the lagging pole, excluding the binding of MglA at that location and thus establishing polarity (111, 112). At the lagging pole, MglB disassembles the focal adhesion complex (103). RomR helps to localize MglA to the leading pole and MglB to the lagging pole through an unknown mechanism, contributing to establishment of cell polarity (113, 114).

Switching the Polarity Axis Through the Frz Chemosensory System

M. xanthus uses a Che-like Frz chemosensory system to reverse the polarity of the motility machinery, inducing reversals in the direction of motility (115). This is the primary mechanism by which cells can control directional motility; any turning seems to be a biproduct of collisions and interactions with other cells as well as the substrate. In *E. coli*, the outcome of the Che system is direct interaction with the flagellar motor to switch between clockwise and counterclockwise rotation in response to chemoattractants or repellants, the Frz system of *M. xanthus* does not interact directly with either motility mechanism. Rather, it appears to switch the polar localization of the MglA/RomR and MglB/RomR complexes, thus inducing a reversal (114, 116, 117).

The Frz system is best understood through its homology to the Che system in *E. coli*, reviewed by Sourjik and Wingreen (118). A methyl-accepting chemotaxis protein (MCP) in cytoplasmic membrane associates with the CheA-CheY HK-RR two-component system that directs flagellar motility. When not in the presence of a chemoattractant, CheA associates with the MCP, autophosphorylates, and signals CheY~P to bind to the FliM flagellar motor protein. This stimulates clockwise rotation of the flagellar motors, leaving the cell in a tumbling state. When the MCP binds a chemoattractant (or repellent), it inhibits its association with CheA, which reduces CheY phosphorylation, which is enhanced through the phosphatase CheZ,

reversing direction of the flagellar motor to counterclockwise, establishing the run mechanism. Together these comprise the "run and tumble" classical behaviors observed in *E. coli* chemotaxis.

Adaptation is a required mechanism for chemotaxis that allows cells to continually sense an increasing chemical gradient (119). This is achieved through the methylation state of the MCP. CheR is a constitutively active methyltransferase that methylates the MCP, which in turn influencing the action of CheA (120). Increasing the methylation state of the MCP stimulates its interaction with CheA, leading to its autophosphorylation, activation of CheY, and subsequent reversal of the flagellar motor to the tumbling state. This resets the baseline level of response so that cells can sense an even higher gradient. CheB, the methylesterase required for demethylation of the MCP is a second RR of CheA and is phosphorylated in response to activated CheA. Its requirement of activation before demethylation creates a lag in response time, allowing CheY to switch direction of the flagellar motor before demethylation of the MCP occurs. This mechanism provides cells a "memory" for chemical gradients over time, allowing continual taxis toward even higher chemical gradients. If one does not exist, then tumbling persists, keeping the cell in the area of highest chemoattractant concentration.

Table 1. Che system proteins and homologs in *M. xanthus*

E. coli **protein**	*M. xanthus* **homolog**	**Function**
Tar (MCP)	FrzCD	Senses attractants
CheA	FrzE	Histidine Kinase
CheY	(FrzZ), ?	Response Regulator
CheB	FrzG	Methylesterase
CheR	FrzF	Methyltransferase

In *M. xanthus,* there are some differences compared to the canonical system. Most notably, *M. xanthus* does not have flagella, activation of FrzE leads to switching of the polarity systems through an unknown mechanism. Another major difference is that FrzCD, the MCP for the Frz system, is cytoplasmic rather than anchored in the inner membrane (121), implying its role in sensing signals internal to the cell and not environmental chemoattractants. Interestingly, though FrzZ was originally suspected to be the main target of FrzE for stimulating reversal (122), the picture is not clear. FrzZ does get phosphorylated by FrzE, but it seems that FrzZ is mainly responsible for regulating the autophosphorylation of FrzE, as FrzZ is not required for reversals in FrzE mutants that lack the domain for autophosphorylation (123). It is proposed that RomR, which has a receiver domain for phosphorylation, may be an additional target of the FrzE kinase that acts to stimulate reversal (114), but as this direct interaction has not yet been demonstrated, it seems more likely that FrzE phosphorylates another downstream RR which then interacts with the RomR RR to stimulate reversal.

Relatively little is known about the signals that lead to polarity switching as well. Changes in reversal frequency in response to the level of C-signaling (see below) indicates that Frz is responding to C-signal through an unknown mechanism (124, 125). Levels of C-signaling

have been proposed to influence FrzCD methylation via its activity on FruA, but again, a direct mechanism remains to be seen.

M. xanthus additionally encodes seven other chemosensory systems that have been far less-studied. The *Dif* chemosensory system is proposed to have a role in EPS production as well as lipid chemotaxis (126). The Che3 chemosensory system, which has homologs for other Che proteins but not CheY, was not found to direct motility but did have an effect on the timing of entry into development (127). The Che4 system, which has a hybrid CheA-CheY protein, seems to affect the TFP, as mutations in the *che4* operon showed S-motility phenotypes in an A-S+ background (128). The chemosensory systems of *M. xanthus* are an interesting area for future exploration, as much of the functions and mechanistic basis, even within the Frz system which has been studied for decades, remains to be elucidated.

SIGNALING ALONG THE DEVELOPMENTAL TIMELINE

Sensing Starvation

Nutrient limitation triggers the onset of the *M. xanthus* developmental program, which ultimately culminates in spore-filled fruiting bodies. Development initially follows the paradigm of the stringent response in *E. coli*. Specifically, a dwindling supply of amino acids leads to ribosome stalling during translation due to uncharged tRNAs (129). Ribosome-associated protein RelA then catalyzes the transfer of phosphate from ATP to GTP to generate (p)ppGpp, an alarmone that is key to downstream signaling in the stringent response (130). This alarmone has several effects in *M. xanthus*, the first of which is to signal transcription factor DksA to downregulate genes for vegetative growth and upregulate developmental genes (131). (p)ppGpp also initiates a phosphorelay at the Asg locus, which ultimately culminates in the expression of A-signal proteases which *M. xanthus* cells respond to in a density-dependent manner (49, 132,

133). Specifically, trypsin is thought to be secreted into the extracellular environment, where it degrades surface proteins, releasing amino acids; both trypsin and six amino acids were shown to collectively comprise A-signal (134, 135).

Sensing of A-signal is mediated by the SasS/SasR two-component system, wherein SasS has two N-terminal transmembrane domains and a C-terminal HK domain that ostensibly autophosphorylates upon sensing A-signal and activating its = cognate RR, SasR (49, 136, 137). SasR is an NtrC-like σ^{54} activator (Nla) protein, and though direct targets of phosphorylated SasR have not been identified, A-signal leads to expression of developmental genes including Nla28, an important two-component system that is part of the EBP cascade discussed below (138, 139). The production of A-signal occurs about 2 hours into development, and its levels are directly proportional to cell density, so it serves as a quorum-sensing mechanism that transmits starvation on the individual cell level to a population-level signal which can be sensed by other cells to initiate developmental decisions.

The MrpC Module

Starvation also induces activation of the major developmental transcription factor MrpC. In vegetative cells, MrpC is inactivated via phosphorylation by a protein kinase cascade (140). Upon starvation, this protein kinase cascade is inhibited, and proteolytic cleavage of the by the LonD protease prevents future phosphorylation of MrpC and increases the affinity of the processed protein, MrpC2, for its binding sites (141). Though MrpC has been shown to bind to hundreds of genes that are developmentally regulated (142), its induction of FruA transcription has the most documented effect on development. The FruA module will be discussed below.

MrpC is necessary for aggregation and sporulation, and dysregulated activation of MrpC causes premature initiation of sporulation prior to complete fruiting body formation (143). In

addition, the levels of MrpC in cells may help to control cell fate, as its expression is greater

within aggregates and lesser in peripheral rods that do not aggregate or sporulate (47). MrpC is

suspected to regulate the *dev* locus which is necessary for the transition of rod-shaped cells to

spores within fruiting bodies (144). It has also been shown to block activity of the MazF toxin

that is predicted to lead to the programmed cell death of cells within fruiting bodies that are not

destined to become spores (145). Taken together, the history of work in MrpC has shown that it

is a master regulator of developmental genes and aids in the transition of cells into one of three

fates: sporulation, lysis, or peripheral rod formation.

C-signal

Perhaps the most controversial signaling system in *M. xanthus*, C-signal is initiated in

response to starvation in parallel with reception of A-signal by starving cells. The major player in

this pathway is CsgA, a short-chain alcohol dehydrogenase family protein (146). Most of what is

known about the mechanism of action of this protein is that its deletion can be complemented by

co-culture with wild-type to rescue its severe developmental phenotype, suggesting that its

function is driven by interactions with other cells (147). CsgA appears to be regulated via a

positive feedback loop such that its expression gets continually amplified throughout

development, and the level of CsgA appears to correlate with cell behaviors such there are lower

levels during rippling, moderate during aggregation, and levels are highest during sporulation

(125, 148).

In one model developed by Kaiser, Søgaard-Andersen, and colleagues, CsgA is translated

as a 25kDa protein (p25) that gets processed to a p17 form shortly after starvation by protease

PopC, and it then gets translocated to the outer membrane (149). It is proposed that this p17 form

of CsgA is the C-signal, and that cells have membrane-associated receptors for this, which have

currently not been identified, that interact with the signal and initiate responses, one of which is to amplify the signal through positive feedback. This model relies on the mechanism of contact-dependent signaling, and specifically end-to-end contacts (150). The other model, primarily proposed by Shimkets and colleagues, is associated with CsgA's function as a short-chain alcohol dehydrogenase, which converts cardiolipin and phosphatidylglycerol from the inner membrane into diacylglycerols that serve as the C-signal (146). This model is supported by the fact that upregulating SocA, another short-chain alcohol dehydrogenase, can rescue the phenotype of the CsgA mutant (151). Notably, both of these mechanisms require cells to be in close proximity, either to make end-to-end contacts or perceive short-range diffusion, which is consistent with data about local cell density increasing C-signaling (152, 153). Whichever of these mechanisms is functionally responsible for C-signaling, the outcomes are relatively clear. CsgA causes activation (by an unknown mechanism) of FruA , an important transcription factor whose expression is initiated by MrpC, thus connecting these two pathways. They are thought to co-regulate certain developmental genes (see below).

FruA Module

Little is known about how FruA is activated in response to C-signaling. It was originally proposed to be phosphorylated (154), however recent sequence analysis reveals that it lacks residues required for phosphorylation, and was not activated upon interactions with phosphodonors (155). It is likely activated through another post-translational modification. Through an unknown mechanism, FruA also interacts with the Frz chemosensory system that directs cellular reversals. It is thought that the positive feedback of C-signaling leads to a consistent increase in FruA activation such that the levels of activated FruA control the frequency of reversals. Low levels of FruA activation lead to increased reversal frequency and

rippling, and higher levels lead to suppression of reversals necessary to trigger aggregation (124). FruA also activates the *dev* operon governing entry to sporulation and formation of the spore coat (156).

Notably, FruA transcription is partially reliant on the A-signaling pathway and FruA activation requires C-signaling (157). The convergence of these pathways in co-regulating the activity of FruA suggests a mechanism to ensure that there are a sufficient number of cells that have individually sensed starvation combined with a sufficient local cell density before directing motility in such a way that leads to aggregation and activating sporulation genes. Additionally, MrpC, which is C-signal independent, and FruA cooperatively bind upstream of many developmentally regulated genes (155, 158, 159), likely ensuring active C-signaling before these genes are expressed.

Enhancer Binding Protein Cascade

Giglio et. al laid the groundwork for our understanding that each major step in the developmental process is also co-regulated by enhancer binding proteins (EBPs), also known as NtrC-like regulators (63). While little is known about exactly which genes these transcription factors are regulating, mutational studies provide clues to the processes that they are involved in. Most EBPs activate the next in the cascade, as well as positively autoregulating their own expression. For example, *nla4* and *nla18* mutants do not appropriately accumulate (p)ppGpp in response to starvation (62, 66). These two EBPs then activate Nla6, which positively autoregulates itself as well as the A-signal pathway in early starvation (63). Activation of the A-signal pathway initiates Nla28, which contributes to the positive amplification at this step. Both Nla6 and Nla28 in combination activate the EBP ActB. ActB responds to C-signal by positively regulating the expression of CsgA, contributing to its amplification (148). Finally, MXAN4899

is activated by ActB and seems to contribute to the activation of FruA by C-signaling (160). Though there is not currently evidence that Nla24 acts as part of this cascade, it has been recently shown to be required for accumulation of EPS which is necessary for aggregation and sporulation (161). Each point in the EBP cascade seems to act as an additional checkpoint for each regulatory module, ensuring proper signal integration.

Sporulation

The transition from vegetative cells to spores requires both a rewiring of gene regulatory networks as well as a physical mechanism to change from a rod-shaped cell to a spherical spore covered in a protective spore coat. First, several gene regulatory modules including C-signal, FruA, and MrpC, act together to ensure the initiation of sporulation only within fruiting bodies and is enacted at the appropriate time. The *dev* operon is activated by FruA, and though it also has a site for cooperative binding of MrpC, it seems that extensive activation by FruA drives transcription of the *dev* operon such that it commits the cell to sporulation (162).

Cells within fruiting bodies (including those destined to undergo autolysis) develop lipid bodies that are thought to originate from cell membrane lipids (163). Lipid bodies appear to be the source of energy that sporulating cells utilize from the cells that undergo lysis, and the exogenous addition of the fatty acids in the lipid bodies triggers cell shortening and the transition to sporulation (164, 165) Actin-like MreB cytoskeletal proteins are suspected to aid in the rearrangement of the rod-shaped cell into a spherical spore by directing the breakdown of the dense network of crosslinked peptidoglycan that normally protects the cell shape (166, 167). The Exo protein, part of the polysaccharide export system, aids in secretion of the thick polysaccharide spore coat, and Nfs proteins use the Agl motor proteins to deposit Exo-secreted polysaccharide around the outside of the spore (107, 168). Finally, the chromosome is replicated

prior to sporulation such that each spore has two copies, presumably to increase chances of survival upon germination (169). Linfong et. al propose that this may be mediated by DnaA, which is downregulated during the *E. coli* stringent response but is upregulated during development in *M. xanthus*. Together, the sporulation process involves the induction of the *dev* operon, utilization of lipid bodies of lysed cells for energy, restructuring of cell shape mediated by MreB, deposition of the spore coat by Exo and Nfs proteins, and replication of the chromosome.

M. XANTHUS AS A SYSTEM FOR STUDYING REDUNDANCY AND DEVELOPMENTAL COORDINATION

The introductory chapter to this thesis provided the context to understand the aspects of the genome and lifestyle of *M. xanthus* that make it a suitable model organism to study my primary research questions. First, in a genome rich with large homologous gene families what is the extent of functional redundancy? We attempt to address this question by quantifying phenotype in four homologous families and assessing phenotypic similarity as an indirect measure of functional redundancy. A complex, multifaceted behavior that is directly tied to biological fitness, such as the formation of fruiting bodies, provides a robust phenotype with which we can begin to address this question on a larger scale than has been previously investigated using double and triple mutants of paralogs.

Though many of the mechanisms that lead to the perception of starvation, aggregation, and sporulation have been elucidated, there remains much more to be discovered about the way *M. xanthus* cells communicate, coordinate their behavior, and direct motility to achieve aggregation and sporulation. My second research question stemmed entirely from observing time-lapse movies from the dataset used to answer my first question, highlighting the importance of large phenotypic collections. What is the functional role of the oscillations in swarm activity

that we observe so frequently during development? We attempt to quantify the behavior, explore the distribution of this phenotype within gene families, and determine the effect of pulses on individual cells. We propose a mechanism by which cells are synchronizing their behavior after nascent aggregates have formed during development, potentially directing cells into fruiting bodies more efficiently.

Finally, it has long been noted through observing *M. xanthus* development movies that more aggregates are initially formed than persist to become mature fruiting bodies. Several studies have come to the same macro-scale conclusion, which is that small aggregates disperse, and larger aggregates remain stable to become mature fruiting bodies. However, it is unknown whether this is driven exclusively by the physical principles of coarsening/Ostwald ripening, or if there is a genetic basis for coarsening. What are the behaviors on scale of the individual cell that lead to aggregate dispersal versus stability, and can these be tracked to any known genetic mechanism? I supplied coarsening-phase cell tracking data to collaborators who extracted key metrics from these experimental cell behaviors, such as bias toward aggregates over time, and will apply these behaviors to data-driven models to simulate the coarsening phase of aggregation.

In all, the three chapters presented in this thesis would not be possible without observations made from extensive collections of phenotype data. Profiling the phenome of an organism in this way is important for observing biological patterns that might normally be missed or misattributed to a single gene or pathway.

REFERENCES

1. Uffelmann E, et al. (2021) Genome-wide association studies. *Nat Rev Methods Prim* 1(1):59.
2. Hughes TR, Robinson MD, Mitsakakis N, Johnston M (2004) The promise of functional genomics: completing the encyclopedia of a cell. *Curr Opin Microbiol* 7(5):546–554.

3. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C (2009) Systematic mapping of genetic interaction networks. *Annu Rev Genet* 43:601–625.

4. Tong AH, et al. (2004) Global mapping of the yeast genetic interaction network. *Sci (New York, NY)* 303(5659):808–813.

5. Kuzmin E, et al. (2020) Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science (80-)* 368(6498). doi:10.1126/science.aaz5667.

6. Butland G, et al. (2008) eSGA: E. coli synthetic genetic array analysis. *Nat Methods* 5(9):789–795.

7. Jacobs MA, et al. (2003) Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* 100(SUPPL. 2):14339–14344.

8. Nagel RL (2005) Epistasis and the genetics of human diseases. *C R Biol* 328(7):606–615.

9. Bradley MD, Neu D, Bahar F, Welch RD (2016) Inter-laboratory evolution of a model organism and its epistatic effects on mutagenesis screens. *Sci Rep* 6(November):1–8.

10. Arnold BJ, Huang IT, Hanage WP (2022) Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* 20(4):206–218.

11. Garcia EC, Brumbaugh AR, Mobley HLT (2011) Redundancy and specificity of Escherichia coli iron acquisition systems during urinary tract infection. *Infect Immun* 79(3):1225–1235.

12. Bonham KS, Wolfe BE, Dutton RJ (2017) Extensive horizontal gene transfer in cheese-associated bacteria. *Elife* 6:e22144.

13. Evans DR, et al. (2020) Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. *Elife* 9:e53886.

14. Abe K, Nomura N, Suzuki S (2020) Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. *FEMS Microbiol Ecol* 96(5):fiaa031.

15. Hoffmann GR (1985) Genetic duplications in bacteria and their relevance for genetic toxicology. *Mutat Res Mol Mech Mutagen* 150(1):107–117.

16. Stoebel DM, Dean AM, Dykhuizen DE (2008) The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. *Genetics* 178(3):1653–1660.

17. Ritchie LJ, Curtis ER, Murphy KA, Welch RD (2021) Profiling Myxococcus xanthus Swarming Phenotypes through Mutation and Environmental Variation. *J Bacteriol*

203(23). doi:10.1128/JB.00306-21.

18. Rivera-Yoshida N, et al. (2019) Plastic multicellular development of Myxococcus xanthus: genotype-environment interactions in a physical gradient. *R Soc open Sci* 6(3):181730.

19. Freimer N, Sabatti C (2003) The human phenome project. *Nat Genet* 34(may):15–21.

20. Ramos EM, et al. (2014) Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur J Hum Genet* 22(1):144–147.

21. Bogue MA, Grubb SC, Maddatu TP, Bult CJ (2007) Mouse Phenome Database (MPD). *Nucleic Acids Res* 35(Database issue):D643-9.

22. Marygold SJ, Crosby MA, Goodman JL (2016) Using FlyBase, a Database of Drosophila Genes and Genomes. *Methods Mol Biol* 1478:1–31.

23. JM C, et al. (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40(Database issue):D700-5.

24. Gagarinova A, Babu M, Greenblatt J, Emili A (2012) Mapping bacterial functional networks and pathways in Escherichia coli using synthetic genetic arrays. *J Vis Exp* (69):1–11.

25. Bochner BR, Gadzinski P, Panomitros E (2001) Phenotype Microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* 11(7):1246–1255.

26. Griffin JE, et al. (2011) High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism. *PLOS Pathog* 7(9):e1002251.

27. Athamneh AIM, Alajlouni RA, Wallace RS, Seleem MN, Senger RS (2014) Phenotypic profiling of antibiotic response signatures in Escherichia coli using Raman spectroscopy. *Antimicrob Agents Chemother* 58(3):1302–1314.

28. Yan J, Bradley MD, Friedman J, Welch RD (2014) Phenotypic profiling of ABC transporter coding genes in Myxococcus xanthus. *Front Microbiol* 5(July):1–12.

29. Gabzi T, Pilpel Y, Friedlander T (2022) Fitness landscape analysis of a tRNA gene reveals that the wild type allele is sub-optimal, yet mutationally robust. *Mol Biol Evol*. doi:10.1093/molbev/msac178.

30. O'Toole G, Kaplan HB, Kolter R (2000) BIOFILM FORMATION AS MICROBIAL DEVELOPMENT. *Annu Rev Microbiol* 54:49.

31.   Costerton JW, Geesey GG, Cheng K-J (1978) How Bacteria Stick. *Sci Am* 238(1):86–95.

32.   Penesyan A, Paulsen IT, Kjelleberg S, Gillings MR (2021) Three faces of biofilms: a microbial lifestyle, a nascent multicellular organism, and an incubator for diversity. *npj Biofilms Microbiomes* 7(1):1–9.

33.   Gestel JVAN, Vlamakis H, Kolter R (2014) Division of Labor in Biofilms : the Ecology of Cell Differentiation. *Microbiol Spectr* 3(2). doi:10.1128/microbiolspec.MB-0002-2014.Correspondence.

34.   Flemming HC, Wingender J (2010) The biofilm matrix. *Nat Rev Microbiol* 8(9):623–633.

35.   Zhang H, et al. (2012) The Mechanistic Basis of Myxococcus xanthus Rippling Behavior and Its Physiological Role during Predation. *PLoS Comput Biol* 8(9). doi:10.1371/journal.pcbi.1002715.

36.   Claessen D, Rozen DE, Kuipers OP, Søgaard-Andersen L, Van Wezel GP (2014) Bacterial solutions to multicellularity: A tale of biofilms, filaments and fruiting bodies. *Nat Rev Microbiol* 12(2):115–124.

37.   Zhang Y, Ducret A, Shaevitz J, Mignot T (2012) From individual cell motility to collective behaviors: Insights from a prokaryote, Myxococcus xanthus. *FEMS Microbiol Rev* 36(1):149–164.

38.   Reichenbach H (1999) The ecology of the myxobacteria. *Environ Microbiol* 1(1):15–21.

39.   Shimkets LJ (1986) Correlation of energy-dependent cell cohesion with social motility in Myxococcus xanthus. *J Bacteriol* 166(3):837–841.

40.   Li Y, et al. (2003) Extracellular polysaccharides mediate pilus retraction during social motility of Myxococcus xanthus. *Proc Natl Acad Sci U S A* 100(9):5443–5448.

41.   Muñoz-Dorado J, Marcos-Torres FJ, García-Bravo E, Moraleda-Muñoz A, Pérez J (2016) Myxobacteria: Moving, killing, feeding, and surviving together. *Front Microbiol* 7(MAY):1–18.

42.   Zusman DR, Scott AE, Yang Z, Kirby JR (2007) Chemosensory pathways, motility and development in Myxococcus xanthus. *Nat Rev Microbiol* 5(11):862–872.

43.   Kroos L (2007) The bacillus and myxococcus developmental networks and their transcriptional regulators. *Annu Rev Genet* 41:13–39.

44.   Wireman JW, Dworkin M (1977) Developmentally Induced Autolysis During Fruiting Body Formation by Myxococcus xanthus. *J Bacteriol* 129(2):796–802.

45.   O'Connor KA, Zusman DR (1991) Development in Myxococcus xanthus involves differentiation into two cell types, peripheral rods and spores. *J Bacteriol* 173(11):3318–3333.

46.   Gardner A, West SA (2006) Demography, altruism, and the benefits of budding. *J Evol Biol* 19(5):1707–1716.

47.   Lee B, Holkenbrink C, Treuner-Lange A, Higgs PI (2012) Myxococcus xanthus developmental cell fate production: Heterogeneous accumulation of developmental Regulatory proteins and reexamination of the role of MazF in developmental lysis. *J Bacteriol* 194(12):3058–3068.

48.   Lewis K (2007) Persister cells, dormancy and infectious disease. *Nat Rev Microbiol* 5(1):48–56.

49.   Kaiser D (2004) Signaling in myxobacteria. *Annu Rev Microbiol* 58:75–98.

50.   Shimkets L, Dworkin M, Reichenbach H (2006) The Myxobacteria. *Prokaryotes*, pp 31–115.

51.   Goldman BS, et al. (2006) Evolution of sensory complexity recorded in a myxobacterial genome. *Proc Natl Acad Sci U S A* 103(41):15200–15205.

52.   Ulrich LE, Koonin E V, Zhulin IB (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* 13(2):52–56.

53.   Levine M, Davidson EH (2005) Gene regulatory networks for development. *Proc Natl Acad Sci U S A* 102(14):4936–4942.

54.   Marcos-torres FJ, Moraleda-muñoz A, Contreras-moreno FJ, Muñoz-dorado J, Pérez J (2022) Mechanisms of Action of Non-Canonical ECF Sigma Factors. *Int J Mol Sci* 23(7):1–13.

55.   Wallace LJ (1979) Regulation of the L-arabinose operon in strains of Escherichia coli containing ColE1-ara hybrid plasmids. *Mol Gen Genet MGG* 173(3):323–331.

56.   Miyada CG, Stoltzfus L, Wilcox G (1984) Regulation of the araC of Escherichia coli: Catabolite repression, autoregulation, and effect on araBAD expression. *Proc Natl Acad Sci U S A* 81(13 I):4120–4124.

57.   Luo ZQ, Farrand SK (1999) Signal-dependent DNA binding and functional domains of the quorum-sensing activator TraR as identified by repressor activity. *Proc Natl Acad Sci U S A* 96(16):9009–9014.

58. Stock AM, Robinson VL, Goudreau PN (2000) Two-component Signal Transduction. *Annu Rev Biochem* 69:183–215.

59. Kenney LJ, Anand GS (2020) EnvZ/OmpR Two-Component Signaling: An Archetype System That Can Function Noncanonically. *EcoSal Plus* 9(1). doi:10.1128/ecosalplus.ESP-0001-2019.

60. Sourjik V, Berg HC (2002) Binding of the Escherichia coli response regulator CheY to its target measured in vivo by fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A* 99(20):12669–12674.

61. Caberoy NB, Welch RD, Jakobsen JS, Slater SC, Garza AG (2003) Global mutational analysis of NtrC-like activators in myxococcus xanthus: Identifying activator mutants defective for motility and fruiting body development. *J Bacteriol* 185(20):6083–6094.

62. Faisury O, et al. (2007) The Myxococcus xanthus Nla4 Protein Is Important for Expression of Stringent Response-Associated Genes, ppGpp Accumulation, and Fruiting Body Development. *J Bacteriol* 189(23):8474–8483.

63. Giglio KM, Caberoy N, Suen G, Kaiser D, Garza AG (2011) A cascade of coregulating enhancer binding proteins initiates and propagates a multicellular developmental program. *Proc Natl Acad Sci U S A* 108(32):E431–E439.

64. Su W, Porter S, Kustu S, Echols H (1990) DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter. *Proc Natl Acad Sci U S A* 87(14):5504–5508.

65. Morett E, Segovia L (1993) The σ54 bacterial enhancer-binding protein family: Mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol* 175(19):6067–6074.

66. E. DM, et al. (2006) Nla18, a Key Regulatory Protein Required for Normal Growth and Development of Myxococcus xanthus. *J Bacteriol* 188(5):1733–1743.

67. Bae B, Feklistov A, Lass-Napiorkowska A, Landick R, Darst SA (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* 4:e08504.

68. Keseler IM, Kaiser D (1997) sigma54, a vital protein for Myxococcus xanthus. *Proc Natl Acad Sci U S A* 94(5):1979–1984.

69. Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 46:47–110.

70. Treviño-Quintanilla LG, Freyre-González JA, Martínez-Flores I (2013) Anti-Sigma Factors in E. coli: Common Regulatory Mechanisms Controlling Sigma Factors Availability. *Curr Genomics* 14(6):378–387.

71. Browning DF, Whitworth DE, Hodgson DA (2003) Light-induced carotenogenesis in Myxococcus xanthus: functional characterization of the ECF sigma factor CarQ and antisigma factor CarR. *Mol Microbiol* 48(1):237–251.

72. Marcos-Torres FJ, Pérez J, Gómez-Santos N, Moraleda-Muñoz A, Muñoz-Dorado J (2016) In depth analysis of the mechanism of action of metal-dependent sigma factors: characterization of CorE2 from Myxococcus xanthus. *Nucleic Acids Res* 44(12):5571–5584.

73. Higgins CF (1992) ABC Transporters: From microorganisms to man. *Annu Rev Cell Biol* 8:67–113.

74. Staudenmaier H, Van Hove B, Yaraghi Z, Braun V (1989) Nucleotide sequences of the fecBCDE genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in Escherichia coli. *J Bacteriol* 171(5):2626–2633.

75. DeVeaux LC, Kadner RJ (1985) Transport of vitamin B12 in Escherichia coli: cloning of the btuCD region. *J Bacteriol* 162(3):888–896.

76. Kanonenberg K, Smits SHJ, Schmitt L (2019) Functional Reconstitution of HlyB, a Type I Secretion ABC Transporter, in Saposin-A Nanoparticles. *Sci Rep* 9(1):8436.

77. Rees DC, Johnson E, Lewinson O (2009) ABC transporters: The power to change. *Nat Rev Mol Cell Biol* 10(3):218–227.

78. Paumi CM, Chuk M, Snider J, Stagljar I, Michaelis S (2009) ABC transporters in Saccharomyces cerevisiae and their interactors: new technology advances the biology of the ABCC (MRP) subfamily. *Microbiol Mol Biol Rev* 73(4):577–593.

79. Sheps JA, Ralph S, Zhao Z, Baillie DL, Ling V (2004) The ABC transporter gene family of Caenorhabditis elegans has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. *Genome Biol* 5(3):R15.

80. Wu C, Budding M, Griffin M, Croop J (1991) Isolation and characterization of Drosophila multidrug resistance gene homologs. *Mol Cell Biol* 11(8):3940–3948.

81. Vasiliou V, Vasiliou K, Nebert DW (2009) Human ATP-binding cassette (ABC)

transporter family. *Hum Genomics* 3(3):281–290.

82. Zolnerciks JK, Andress EJ, Nicolaou M, Linton KJ (2011) Structure of ABC transporters. *Essays Biochem* 50(1):43–61.

83. Benaroudj N, et al. (2013) Structural and functional characterization of an orphan ATP-binding cassette ATPase involved in manganese utilization and tolerance in Leptospira spp. *J Bacteriol* 195(24):5583–5591.

84. Thomas GH (2010) Homes for the orphans: utilization of multiple substrate-binding proteins by ABC transporters. *Mol Microbiol* 75(1):6–9.

85. Ward MJ, Mok KC, Astling DP, Lew H, Zusman DR (1998) An ABC transporter plays a developmental aggregation role in Myxococcus xanthus. *J Bacteriol* 180(21):5697–5703.

86. Kimura Y, Yamanishi Y, Tokumasu Y, Teresaka H, Yoshinobu J (2001) Characterization of the mac-1 Gene Encoding a Putative ABC Transporter from Myxococcus xanthus1. *J Biochem* 129(3):351–356.

87. Wu SS, Wu J, Cheng YL, Kaiser D (1998) The pilH gene encodes an ABC transporter homologue required for type IV pilus biogenesis and social gliding motility in Myxococcus xanthus. *Mol Microbiol* 29(5):1249–1261.

88. Craig L, Forest KT, Maier B (2019) Type IV pili: dynamics, biophysics and functional consequences. *Nat Rev Microbiol* 17(7):429–440.

89. Ohno S (1970) *Evolution by Gene Duplication* (Springer Berlin Heidelberg).

90. Mattick JS (2002) Type IV pili and twitching motility. *Annu Rev Microbiol* 56:289–314.

91. Yu R, Kaiser D (2007) Gliding motility and polarized slime secretion. *Mol Microbiol* 63(2):454–467.

92. Chang Y-W, et al. (2016) Architecture of the type IVa pilus machine. *Science* 351(6278):aad2001.

93. Kaiser D, Crosby C (1983) Cell movement and its coordination in swarms of Myxococcus xanthus. *Cell Motil* 3(3):227–245.

94. Yang Z, Lux R, Hu W, Hu C, Shi W (2010) PilA localization affects extracellular polysaccharide production and fruiting body formation in Myxococcus xanthus. *Mol Microbiol* 76(6):1500–1513.

95. Hu W, et al. (2016) Interplay between type IV pili activity and exopolysaccharides secretion controls motility patterns in single cells of Myxococcus xanthus. *Sci Rep*

6(November 2015):17790.

96. Black WP, Xu Q, Yang Z (2006) Type IV pili function upstream of the Dif chemotaxis pathway in Myxococcus xanthus EPS regulation. *Mol Microbiol* 61(2):447–456.

97. Islam ST, Mignot T (2015) The mysterious nature of bacterial surface (gliding) motility: A focal adhesion-based mechanism in Myxococcus xanthus. *Semin Cell Dev Biol* 46(November):143–154.

98. Burchard RP (1982) Trail following by gliding bacteria. *J Bacteriol* 152(1):495–501.

99. Beiyan N, et al. (2011) Myxobacteria gliding motility requires cytoskeleton rotation powered by proton motive force. *Proc Natl Acad Sci* 108(6):2498–2503.

100. Adrien D, Marie-Pierre V, Fabrice M, Tâm M, Olivier T (2012) Wet-surface–enhanced ellipsometric contrast microscopy identifies slime as a major adhesion factor during bacterial surface motility. *Proc Natl Acad Sci* 109(25):10036–10041.

101. Tâm M, W. SJ, L. HP, R. ZD (2007) Evidence That Focal Adhesion Complexes Power Bacterial Gliding Motility. *Science (80-)* 315(5813):853–856.

102. Mauriello EMF, et al. (2010) Bacterial motility complexes require the actin-like protein, MreB and the Ras homologue, MglA. *EMBO J* 29(2):315–326.

103. Treuner-Lange A, et al. (2015) The small G-protein MglA connects to the MreB actin cytoskeleton at bacterial focal adhesions. *J Cell Biol* 210(2):243–256.

104. Jakobczak B, Keilberg D, Wuichet K, Søgaard-Andersen L (2015) Contact- and Protein Transfer-Dependent Stimulation of Assembly of the Gliding Motility Machinery in Myxococcus xanthus. *PLoS Genet* 11(7):e1005341.

105. Luciano J, et al. (2011) Emergence and Modular Evolution of a Novel Motility Machinery in Bacteria. *PLOS Genet* 7(9):e1002268.

106. Nan B, Mauriello EMF, Sun I-H, Wong A, Zusman DR (2010) A multi-protein complex from Myxococcus xanthus required for bacterial gliding motility. *Mol Microbiol* 76(6):1539–1554.

107. Wartel M, et al. (2013) A Versatile Class of Cell Surface Directional Motors Gives Rise to Gliding Motility and Sporulation in Myxococcus xanthus. *PLOS Biol* 11(12):e1001728.

108. Beiyan N, et al. (2013) Flagella stator homologs function as motors for myxobacterial gliding motility by moving in helical trajectories. *Proc Natl Acad Sci* 110(16):E1508–E1513.

109. Mingzhai S, Morgane W, Eric C, W. SJ, Tâm M (2011) Motor-driven intracellular transport powers bacterial gliding motility. *Proc Natl Acad Sci* 108(18):7559–7564.

110. Schumacher D, Sogaard-Andersen L (2017) Regulation of Cell Polarity in Motility and Cell Division in Myxococcus xanthus. *Annu Rev Microbiol* 71:61–78.

111. Leonardy S, et al. (2010) Regulation of dynamic polarity switching in bacteria by a Ras-like G-protein and its cognate GAP. *EMBO J* 29(14):2276–2289.

112. Zhang Y, Franco M, Ducret A, Mignot T (2010) A bacterial Ras-like small GTP-binding protein and its cognate GAP establish a dynamic spatial polarity axis to control directed motility. *PLoS Biol* 8(7):e1000430.

113. Leonardy S, Freymark G, Hebener S, Ellehauge E, Søgaard-Andersen L (2007) Coupling of protein localization and cell movements by a dynamically localized response regulator in Myxococcus xanthus. *EMBO J* 26(21):4433–4444.

114. Keilberg D, Wuichet K, Drescher F, Søgaard-Andersen L (2012) A response regulator interfaces between the Frz chemosensory system and the MglA/MglB GTPase/GAP module to regulate polarity in Myxococcus xanthus. *PLoS Genet* 8(9):e1002951.

115. Kaimer C, Berleman JE, Zusman DR (2012) Chemosensory signaling controls motility and subcellular polarity in Myxococcus xanthus. *Curr Opin Microbiol* 15(6):751–757.

116. Guzzo M, et al. (2015) Evolution and Design Governing Signal Precision and Amplification in a Bacterial Chemosensory Pathway. *PLoS Genet* 11(8):1–26.

117. Zhang Y, Guzzo M, Ducret A, Li Y-Z, Mignot T (2012) A dynamic response regulator protein modulates G-protein-dependent polarity in the bacterium Myxococcus xanthus. *PLoS Genet* 8(8):e1002872.

118. Sourjik V, Wingreen NS (2012) Responding to chemical gradients: bacterial chemotaxis. *Curr Opin Cell Biol* 24(2):262–268.

119. Kehry MR, Dahlquist FW (1982) Adaptation in bacterial chemotaxis: CheB-dependent modification permits additional methylations of sensory transducer proteins. *Cell* 29(3):761–772.

120. Kirby JR, et al. (2008) Chemosensory Signal Transduction Systems in M. xanthus. *Myxobacteria: Multicellularity and Differentiation*, ed Whitworth DE, pp 135–147.

121. Bustamante VH, Martínez-Flores I, Vlamakis HC, Zusman DR (2004) Analysis of the Frz signal transduction system of Myxococcus xanthus shows the importance of the

conserved C-terminal region of the cytoplasmic chemoreceptor FrzCD in sensing signals. *Mol Microbiol* 53(5):1501–1513.

122. Inclán YF, Vlamakis HC, Zusman DR (2007) FrzZ, a dual CheY-like response regulator, functions as an output for the Frz chemosensory pathway of Myxococcus xanthus. *Mol Microbiol* 65(1):90–102.

123. Kaimer C, Zusman DR (2016) Regulation of cell reversal frequency in Myxococcus xanthus requires the balanced activity of CheY-like domains in FrzE and FrzZ. *Mol Microbiol* 100(2):379–395.

124. Lars J, Lotte S-A (2002) Pattern formation by a cell surface-associated morphogen in Myxococcus xanthus. *Proc Natl Acad Sci* 99(4):2032–2037.

125. Kruse T, Lobedanz S, Berthelsen NMS, Søgaard-Andersen L (2001) C-signal: A cell surface-associated morphogen that induces and co-ordinates multicellular fruiting body morphogenesis and sporulation in Myxococcus xanthus. *Mol Microbiol* 40(1):156–168.

126. Bonner PJ, et al. (2005) The Dif chemosensory pathway is directly involved in phosphatidylethanolamine sensory transduction in Myxococcus xanthus. *Mol Microbiol* 57(5):1499–1508.

127. Kirby JR, Zusman DR (2003) Chemosensory regulation of developmental gene expression in Myxococcus xanthus. *Proc Natl Acad Sci U S A* 100(4):2008–2013.

128. Vlamakis HC, Kirby JR, Zusman DR (2004) The Che4 pathway of Myxococcus xanthus regulates type IV pilus-mediated motility. *Mol Microbiol* 52(6):1799–1811.

129. Cashel M, Potrykus K (2013) *Stringent Response* (Elsevier Inc.) doi:10.1016/B978-0-12-374984-0.01486-8.

130. Harris BZ, Kaiser D, Singer M (1998) The guanosine nucleotide (p)ppGpp initiates development and A-factor production in Myxococcus xanthus. *Genes Dev* 12(7):1022–1035.

131. Perederina A, et al. (2004) Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell* 118(3):297–309.

132. Plamann L, Li Y, Cantwell B, Mayor J (1995) The Myxococcus xanthus asgA gene encodes a novel signal transduction protein required for multicellular development. *J Bacteriol* 177(8):2014–2020.

133. Kaplan HB, Plamann L (1996) A Myxococcus xanthus cell density-sensing system

required for multicellular development. *FEMS Microbiol Lett* 139(2–3):89–95.

134. Plamann L, Kuspa A, Kaiser D (1992) Proteins that rescue A-signal-defective mutants of Myxococcus xanthus. *J Bacteriol* 174(10):3311–3318.

135. Kuspa A, Plamann L, Kaiser D (1992) Identification Heat-Stable A-Factor from Myxococcus xanthus. *J Bacteriol* 174(10):3319–3326.

136. Guo D, Wu Y, Kaplan HB (2000) Identification and characterization of genes required for early Myxococcus xanthus developmental gene expression. *J Bacteriol* 182(16):4564–4571.

137. Yang C, Kaplan HB (1997) Myxococcus xanthus sasS encodes a sensor histidine kinase required for early developmental gene expression. *J Bacteriol* 179(24):7759–7767.

138. Kuspa A, Kroos L, Kaiser D (1986) Intercellular signaling is required for developmental gene expression in Myxococcus xanthus. *Dev Biol* 117(1):267–276.

139. Sarwar Z, Garza AG (2012) The Nla28S/Nla28 two-component signal transduction system regulates sporulation in Myxococcus xanthus. *J Bacteriol* 194(17):4698–4708.

140. Nariya H, Inouye S (2005) Identification of a protein Ser/Thr kinase cascade that regulates essential transcriptional activators in Myxococcus xanthus development. *Mol Microbiol* 58(2):367–379.

141. Nariya H, Inouye S (2006) A protein Ser/Thr kinase cascade negatively regulates the DNA-binding activity of MrpC, a smaller form of which may be necessary for the Myxococcus xanthus development. *Mol Microbiol* 60(5):1205–1217.

142. Robinson M, Son B, Kroos D, Kroos L (2014) Transcription factor MrpC binds to promoter regions of hundreds of developmentally-regulated genes in Myxococcus xanthus. *BMC Genomics* 15(1):1123.

143. Higgs PI, Jagadeesan S, Mann P, Zusman DR (2008) EspA, an orphan hybrid histidine protein kinase, regulates the timing of expression of key developmental proteins of Myxococcus xanthus. *J Bacteriol* 190(13):4416–4426.

144. Thony-Meyer L, Kaiser D (1993) devRS, an autoregulated and essential genetic locus for fruiting body development in Myxococcus xanthus. *J Bacteriol* 175(22):7450–7462.

145. Nariya H, Inouye M (2008) MazF, an mRNA Interferase, Mediates Programmed Cell Death during Multicellular Myxococcus Development. *Cell* 132(1):55–66.

146. Boynton TO, Shimkets LJ (2015) Myxococcus CsgA, Drosophila Sniffer, and human

HSD10 are cardiolipin phospholipases. *Genes Dev* 29(18):1903–1914.

147. Kim SK, Kaiser D (1991) C-factor has distinct aggregation and sporulation thresholds during Myxococcus development. *J Bacteriol* 173(5):1722–1728.

148. Gronewold TMA, Kaiser D (2001) The act operon controls the level and time of C-signal production for Myxococcus xanthus development. *Mol Microbiol* 40(3):744–756.

149. Lobedanz S, Søgaard-Andersen L (2003) Identification of the C-signal, a contact-dependent morphogen coordinating multiple developmental responses in Myxococcus xanthus. *Genes Dev* 17(17):2151–2161.

150. Kaiser D (2004) Signaling in myxobacteria. *Annu Rev Microbiol* 58:75–98.

151. Lee K, Shimkets LJ (1994) Cloning and characterization of the socA locus which restores development to Myxococcus xanthus C-signaling mutants. *J Bacteriol* 176(8):2200–2209.

152. K. KS, Dale K (1990) Cell Alignment Required in Differentiation of Myxococcus xanthus. *Science (80-)* 249(4971):926–928.

153. Kroos L, Hartzell P, Stephens K, Kaiser D (1988) A link between cell movement and gene expression argues that motility is required for cell-cell signaling during fruiting body development. *Genes Dev* 2(12A):1677–1685.

154. Ellehauge E, Nørregaard-Madsen M, Søgaard-Andersen L (1998) The FruA signal transduction protein provides a checkpoint for the temporal co-ordination of intercellular signals in Myxococcus xanthus development. *Mol Microbiol* 30(4):807–817.

155. Mittal S, Kroos L (2009) A combination of unusual transcription factors binds cooperatively to control Myxococcus xanthus developmental gene expression. *Proc Natl Acad Sci U S A* 106(6):1965–1970.

156. Kroos L (2017) Highly Signal-Responsive Gene Regulatory Network Governing Myxococcus Development. *Trends Genet* 33(1):3–15.

157. Bretl DJ, Kirby JR (2016) Molecular Mechanisms of Signaling in Myxococcus xanthus Development. *J Mol Biol* 428(19):3805–3830.

158. Campbell A, et al. (2015) Combinatorial regulation of the dev operon by MrpC2 and FruA during Myxococcus xanthus development. *J Bacteriol* 197(2):240–251.

159. Jun-seok L, Bongjun S, Poorna V, M. LP, Lee K (2011) Combinatorial Regulation of fmgD by MrpC2 and FruA during Myxococcus xanthus Development . *J Bacteriol*

193(7):1681–1689.

160. Lars J, Michael G, Dale K (2005) Enhancer-binding proteins with a forkhead-associated domain and the σ54 regulon in Myxococcus xanthus fruiting body development. *Proc Natl Acad Sci* 102(8):3010–3015.

161. Skotnicka D, et al. (2016) A Minimal Threshold of c-di-GMP Is Essential for Fruiting Body Formation and Sporulation in Myxococcus xanthus. *PLoS Genet* 12(5):1–27.

162. Saha S, Patra P, Igoshin O, Kroos L (2019) Systematic analysis of the Myxococcus xanthus developmental gene regulatory network supports posttranslational regulation of FruA by C-signaling. *Mol Microbiol* 111(6):1732–1752.

163. Bhat S, Boynton TO, Pham D, Shimkets LJ (2014) Fatty Acids from Membrane Lipids Become Incorporated into Lipid Bodies during Myxococcus xanthus Differentiation. *PLoS One* 9(6):e99622.

164. Swapna B, et al. (2014) Two Lipid Signals Guide Fruiting Body Development of Myxococcus xanthus. *MBio* 5(1):e00939-13.

165. Hoiczyk E, et al. (2009) Lipid body formation plays a central role in cell fate determination during developmental differentiation of Myxococcus xanthus. *Mol Microbiol* 74(2):497–517.

166. Müller FD, Treuner-Lange A, Heider J, Huntley SM, Higgs PI (2010) Global transcriptome analysis of spore formation in Myxococcus xanthus reveals a locus necessary for cell differentiation. *BMC Genomics* 11(1). doi:10.1186/1471-2164-11-264.

167. Zhang H, Venkatesan S, Nan B (2021) Myxococcus xanthus as a Model Organism for Peptidoglycan Assembly and Bacterial Morphogenesis. *Microorganisms* 9(5). doi:10.3390/microorganisms9050916.

168. Müller FD, Schink CW, Hoiczyk E, Cserti E, Higgs PI (2012) Spore formation in Myxococcus xanthus is tied to cytoskeleton functions and polysaccharide spore coat deposition. *Mol Microbiol* 83(3):486–505.

169. Linfong T, Mitchell S (2005) DNA replication during sporulation in Myxococcus xanthus fruiting bodies. *Proc Natl Acad Sci* 102(40):14428–14433.

CHAPTER 2: PHENOTYPIC SIMILARITY IS AN INDICATOR OF FUNCTIONAL REDUNDANCY WITHIN HOMOLOGOUS GENE FAMILIES

Jessica A. Comstock[1], Merrill E. Asp[2], Fatmagül Bahar[1], Isabella Lee[2], Alison E. Patteson[2], Roy D. Welch[1*]

[1] Department of Biology, Syracuse University, Syracuse, NY 13210, USA
[2] Department of Physics, Syracuse University, Syracuse, NY 13210, USA
[*] Corresponding Author, Email: rowelch@syr.edu

(Mutant movie collection primarily by FAB with some data collection by me. Manual phenotyping by me based on phenotypic categories designated by FAB. Conceptualization of experimental hypotheses about phenotypic similarity by RDW and me. Quantification of phenotypes by MEA and AEP, with my input on features to quantify. Writing by me, MEA, AEP, and RDW).

ABSTRACT

Robustness to the phenotypic impact of mutation is encoded into the genome through duplication and divergence, producing gene families with varying degrees of functional overlap. To better understand the extent to which homologous gene families buffer the impact of mutation, we explored phenotypic similarity as an indicator of functional redundancy. We hypothesize that, since redundancy is more likely to exist among homologs, knock-out strains with mutations in members of the same gene family will be more likely to display similar phenotypes. To test this, we created 265 mutant strains of the bacterium *Myxococcus xanthus*, each with a disruption in a gene belonging to one of four families of homologs. We used time-lapse microscopy to generate movies of multicellular development for each mutant strain and developed an image analysis pipeline to compare phenotypic features among different strains. We demonstrate that mutant strains cluster according to gene family in phenotypic feature space using principal component analysis, and we argue that this supports our hypothesis that the impact of mutation can be distributed broadly across large redundancy networks.

Keywords: functional redundancy; robustness; homologous gene families; genotype-to-phenotype, *Myxococcus xanthus;*

INTRODUCTION

A reverse genetics approach to characterizing a gene often begins by disrupting or deleting the gene and observing the resulting phenotype. Differences between the mutant and wild-type phenotypes can provide invaluable insights regarding gene function(s), but in practice many single-gene knockouts, even those in genes predicted to be important based on previously studied homologs, yield phenotypes that are relatively minor or indistinguishable from the wild-type organism (Diss *et al*, 2014; Giaever *et al*, 2002). This robustness to the phenotypic impact

of genetic mutation is an important part of an organism's phenotype and has implications for fitness.

Robustness is commonly attributed, at least in part, to functional redundancy, or the tendency for functionally similar genes to compensate for the role of a disrupted gene (Ohno, 1970). Functional redundancy can arise through many mechanisms including duplication and divergence, where reduced selective pressure can cause paralogs to accumulate mutations and take on new, slightly different functions over time (Vandersluis *et al*, 2010; Krakauer & Plotkin, 2002). Paralogs that are maintained over long timescales often retain some of their ancestral function in addition to their diverged function (Kuzmin *et al*, 2022; Dean *et al*, 2008), thus building in redundancy. Repeated gene duplication events can give rise to large gene families wherein genes have a range of biochemically similar but specialized functions. Though many homologs in a gene family may be capable of performing a similar function, due to divergence it is difficult to predict which genes might be able to compensate for the function of others. The most recent duplicates within a gene family are not always capable of being functionally redundant while some older and more diverged paralogs are (Baker *et al*, 2021). Sequence similarity alone is not enough to predict functional redundancy, and the extent to which duplicates contribute to robustness varies across organisms (Hannay *et al*, 2008). For these reasons, it is unclear to what extent families of homologs are contributing to the functional redundancy that gives rise to robustness in biological systems.

Many studies attempting to elucidate functional redundancy in the genome involve the creation of single and double knockouts of paralogs to probe for synthetic lethality (Thomaides *et al*, 2007; diCenzo & Finan, 2015; Butland *et al*, 2008). While this method is effective in assessing functional redundancy in pairs of closely related genes, it is limited in its power to

explore larger networks of redundancy, as may exist in expanded gene families. A double mutant that does not show a more significant phenotype than each of the corresponding single mutants could imply either that the genes are non-redundant, or that they are part of a larger redundancy network that has a strong buffering capacity and therefore has decreased fragility in the face of genetic perturbation (Lehár *et al*, 2008). In this way, phenotype is often the readout for assessing redundancy and robustness within biological systems. The phenotypic impact of mutation reveals information about robustness, and we can investigate the mechanisms that lead to robustness by considering gene sequence, so understanding how redundancy affects robustness is a crucial genotype-to-phenotype question.

Any given gene product processes the flow of information from precursors, producing outputs that feed into other networks or cellular functions. In a simple case of non-redundancy, a gene produces one protein with a primary function, and when this gene is intact, expresses a wild-type phenotype (Fig. 1A). A mutation in this gene would severely impact the fitness of the organism. However, if a given gene is part of a network of structurally similar genes which each have their own primary function but also retain some ancestral function, as in gene families that arise from duplication, the impact of a mutation can be diffused through the other members of its network, producing a relatively minor deviation in phenotype. Redundancy networks (Fig. 1B), which we here define as the group of two or more genes whose products can compensate for the loss of function of one another, allow for the rerouting of information through alternative pathways so that the end result has a minimal impact on fitness. As shown in Figure 1C, an additional byproduct of this buffering effect is that knocking out one member of a redundancy network should produce a similar phenotype as knocking out any other member of that group, because the entire set of genes is affected no matter which component of the network is disrupted

by mutation. In this way, phenotypic similarity may be an important indicator of functional redundancy within homologous gene families and may provide insights into the level of robustness in a genome. Further, because a gene's redundancy network likely overlaps significantly with its family of homologs due to the relationship between protein structure and function (Fig. 1D), we predict that mutations in genes from within the same gene family will be more phenotypically similar.

To test this, we phenotypically characterized over 250 single-gene mutations in *Myxococcus xanthus*, a soil bacterium with a large genome containing multiple homologous gene families (Goldman *et al*, 2006) and examined the relationship between gene family and phenotype. Under nutrient stress, swarming cells of *M. xanthus* undergo development, aggregating into multicellular fruiting bodies wherein populations of cells will differentiate into spores (Bretl & Kirby, 2016) (Fig. 2A). Since the ability of *M. xanthus* to form fruiting bodies and sporulate is directly tied to its fitness, it is likely a robust biological process that involves many functionally redundant genes. We generated a library of microscopic time-lapse movies (time series) showing the development of 265 knockout strains of *M. xanthus* belonging to four different gene families (102 ABC transporter genes, 45 NtrC-like activators, 80 One component signal transduction genes, and 38 ECF sigma factors; see references Yan *et al*, 2014, Caberoy *et al*, 2003, and Abellón-Ruiz *et al*, 2014 for previous work on some of these genes in *M. xanthus)*. We made qualitative observations of the ways in which resulting phenotypes differed from wild-type and used these observations to inform a novel image processing and phenotypic analysis pipeline that automates quantitative measurements of phenotype that are explicitly defined. Although previous studies have used image processing to extract phenotypic features of aggregate formation (Xie *et al*, 2011), this work has applied these tools to the largest library of

time series of which we are aware, necessitating a new pipeline and analysis methods. Finally, we compared the similarity of phenotypes across gene families using principal component analysis (PCA). We found that, just as mutant strains within a gene family cluster by sequence similarity through multidimensional scaling (Fig. 1E), they also cluster by gene family in the phenotypic feature space with a statistically significant sharpness (i.e. small cluster size) and separation of clusters, indicating large networks of redundancy within these gene families.
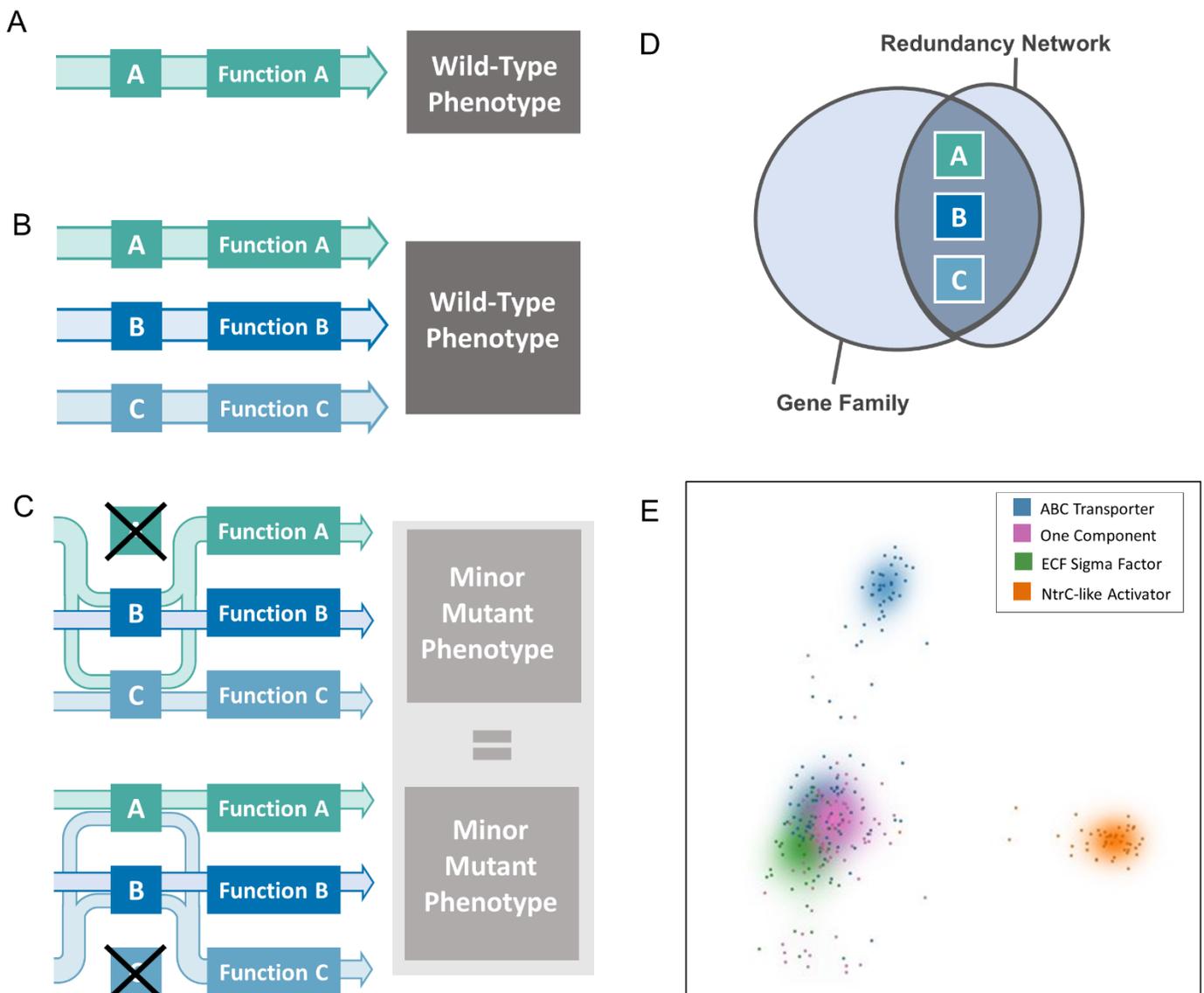
Figure 1. Functional redundancy resulting in phenotypic similarity. (**A**) In a pathway with no redundancy, Gene A contributes to Function A. Any mutation that renders Gene A nonfunctional would produce a severe phenotype or lethality if Function A is essential. (**B**) Genes A, B, and C belong to the same redundancy network, meaning each gene can compensate for the loss of function of one member of its network. In the scenario where all three genes are functional and operating optimally, each gene contributes to its primary function (for example, Gene A is responsible for most of the contribution to Function A), producing the wild-type phenotype. (**C**) When a mutation occurs that renders Gene A nonfunctional (top), the input to Gene A gets rerouted through Genes B and C such that Function A can still occur, but in a slightly reduced capacity (indicated by thickness of arrows compared to panel B). Since Genes B and C are processing more input from A, Functions B and C are also affected and operate at a reduced capacity. The slight reduction in function of all three network components produces a phenotype that is relatively minor and may be indistinguishable from wild-type. A mutation in Gene C (bottom) would result in a similar phenomenon, where the input that normally feeds into Gene C is processed by Genes A and B, resulting in overall decreased output from each. In this model, a mutation in one member of a redundancy network affects the output from all components regardless of which gene contains the mutation (indicated by the similar output arrow size of top and bottom of panel C), and we predict that mutations in members of the same redundancy network will produce similar phenotypes. (**D**) Though not every member of a gene family is functionally redundant, and there may be redundant genes that do not belong to the same gene family, the relationship between structure and function of proteins dictates that genes in the same redundancy network are likely to come from the same gene family. If redundancy networks come primarily from the same gene family, and components of redundancy networks show similar mutant phenotypes, then members of the same gene family would be more likely to produce the same mutant phenotype. (**E**) Similarity of the protein sequences used in this study by multidimensional scaling. Each point represents one gene, with a Gaussian kernel density estimate to guide the eye. Proteins that are more similar in sequence, belonging to the same gene family, cluster together. Each gene family forms a single cluster with the exception of the ABC Transporters, which form two major clusters due to the different subunits. The highly conserved ATP-binding domains (Rees *et al*, 2009) separate very distinctly from the periplasmic and substrate-binding domains. We predict that mutations in genes that belong to the same gene family will be more phenotypically similar than they will be to mutant phenotypes in other paralogous gene families. Thus, we expect phenotype to cluster by gene family.

RESULTS

Manual Characterization of Development Phenotypes

Under starvation, a swarm of *M. xanthus* cells will execute a developmental program during which millions of rod-shaped cells coordinate their movements and self-organize into dome-shaped multicellular aggregates. Some nascent aggregates destabilize and disperse, but

most persist and continue to grow; when the persisting aggregates become large enough, cells in the middle of each differentiate to form a cluster of spores, at which point they are considered mature fruiting bodies (Fig. 2A). Capturing this process with time-lapse brightfield microscopy results in a time series of grayscale images where initial aggregates appear roughly circular with irregular boundaries, somewhat darker than the background swarm. Later in the time series, dispersing aggregates shrink and disappear, and the persisting ones grow and darken, with boundaries that become stable and clearly defined (Fig. 2B). Image features such as these can be leveraged to compare development phenotypes between wild-type and mutant *M. xanthus* strains.

For this study we recorded 24-hour time series for wild-type and a set of 265 single gene knockout mutant strains (Appendix 2, Table 1), with an average of three replicates per strain. Due to their important roles in signal transduction, transport, and transcriptional regulation, we predict that genes within these families will be part of redundancy networks. We compared the mutant phenotypes to our wild-type strain with an emphasis on aggregate composition and dynamics. Wild-type aggregation initiated at 9.2±1.6 hours and formed uniformly dark circular aggregates with stable and clearly delineated boundaries within 24 hours (Fig. 2B). Mutant strains that consistently initiated aggregation either before or after wild-type were designated "early" or "late", respectively. Mutant strains that consistently initiated aggregation at the same time as wild-type but had aggregates that failed to darken and/or form clearly delineated boundaries were designated "immature" (Fig. 2C). Mutant strains that initiated aggregation at the same time as wild-type but then all the aggregates dispersed within 24 hours were designated "fall apart" (Fig. 2D). Mutant strains that initiated aggregation at the same time as wild-type but then all the aggregates dispersed and then re-aggregated within 24 hours were designated

"aggregate-reaggregate" (Fig. 2D). Mutant strains that consistently matched all aggregation

criteria and were indistinguishable from wild-type were designated "Like Wild-Type" (LWT).

Finally, Mutant strains where the replicates displayed different developmental classifications

were designated "variable".

Distribution of Manual Development Phenotypes within each Gene Family

Of the mutant strains characterized in this study, less than 10% failed to initiate

aggregation at all, and 62% consistently produced fruiting bodies that were qualitatively

comparable to wild-type by the end of the 24-hour window. An additional 20% of mutants were

able to initiate aggregation, but aggregates remained immature; some of these strains may have

formed mature aggregates if the time series extended longer than 24 hours.

We hypothesize that the relatively high success rate of aggregation in these mutants is

due, at least in part, to *M. xanthus* development being a robust phenotype. If redundancy

networks are contributing to functional redundancy to produce this robustness, then, according to

the hypothesis portrayed in Fig. 1, mutants within the same gene family will be more

phenotypically similar. As an initial test of our hypotheses, we sorted the mutant strains into their

gene families and visualized the proportional representation of our developmental phenotype

classifications (Fig. 2E). The distribution of some phenotypes did seem to favor specific gene

families. For example, LWT strains made up over half of the ABC Transporter family, the early

aggregating strains compose nearly half of the ECF sigma factor family, and about one third of

the One component family produce variable phenotypes in different replicates.

The manual categorization of development phenotypes presented here serves two

purposes. First, it provides support for our hypothesis that *M. xanthus* development is as robust a

phenotype as we expected, making it a suitable phenotype for observing the extent of functional

redundancy networks in gene families. Second, though we do not claim these data alone provide

sufficient evidence for the existence of redundancy networks, as the data show only the most

obvious associations between gene family and phenotype, these qualitative observations do

provide information about the various ways in which phenotype can differ during development.

This was used to inform a more systematic, quantitative, and multidimensional characterization

pipeline to test our remaining hypothesis about phenotypic similarity among families of paralogs:

if redundancy networks contribute to robustness, and if those networks are comprised primarily

of genes within the same family, then a grouping of mutant strains based on phenotypic features

should also group the strains according to gene family.
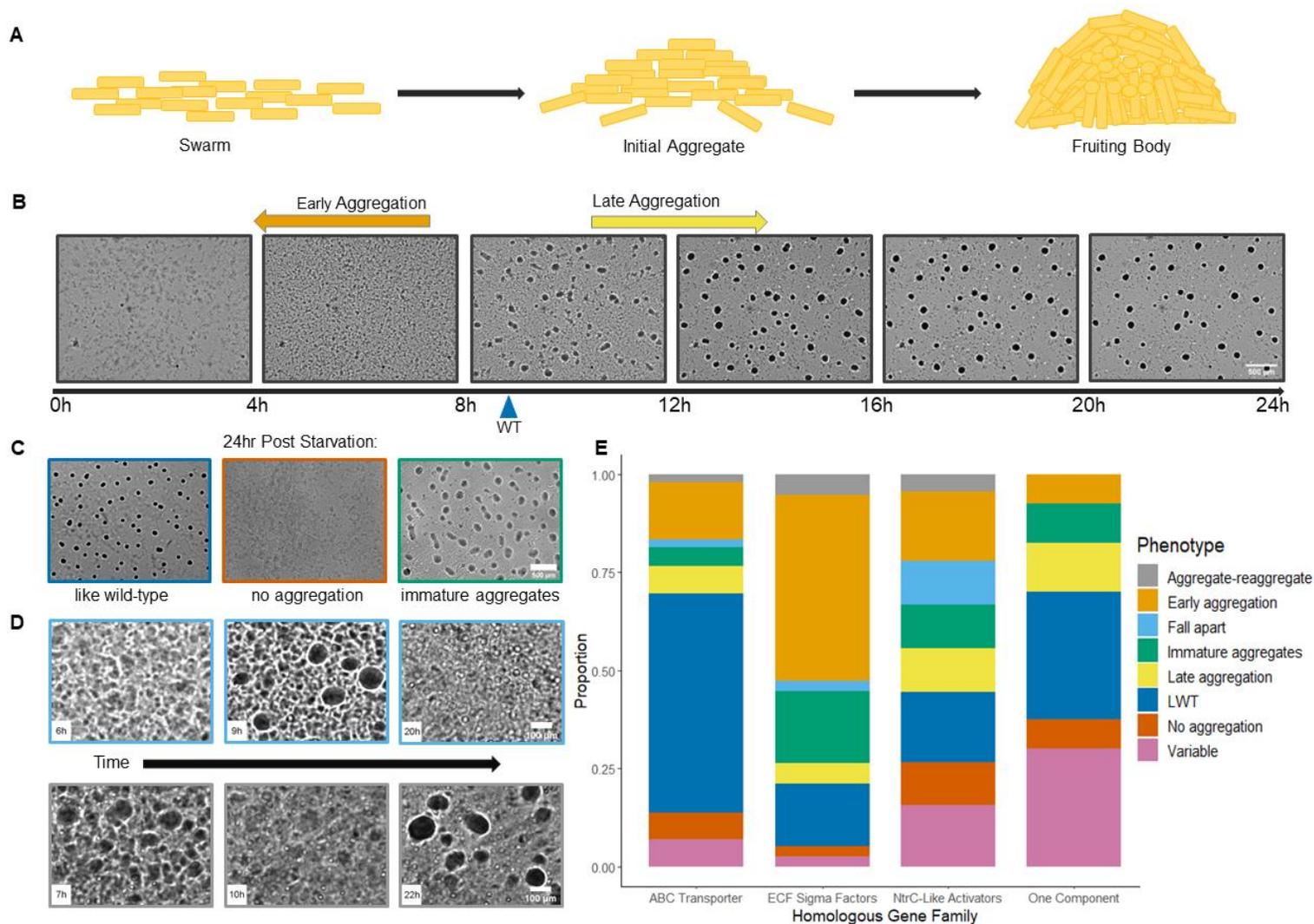
Figure 2: Manual categorization of *M. xanthus* development. (A) Upon sensing nutrient stress, vegetative *M. xanthus* cells undergo a developmental process that culminates in spore-filled fruiting bodies. **(B)** Wild-type *M. xanthus* cells on TPM agar begin to cluster into early aggregates after 9 hrs of starvation (blue arrow), and as more cells join the premature aggregates over the course of 24 hours, the aggregates mature into fruiting bodies that appear round and dark with conventional brightfield microscopy. Mutant strains that show initial aggregation either before or after the average time for wild-type are assigned the early aggregation (orange arrow) and late aggregation (yellow arrow) phenotypes, respectively. Scale bar 500 µm. **(C)** Like wild-type (LWT) mutants that produced dark, circular fruiting bodies on a timeline similar to wild-type (left), non-aggregating mutants (center), and mutants that produced immature aggregates (right). Scale bar 500 µm. **(D)** Some mutants formed initial aggregates that eventually shrank and fell apart (top). Other mutants formed initial aggregates that fell apart before re-aggregating into mature fruiting bodies (bottom). Scale bar 100 µm. **(E)** Distribution of development classifications within each gene family.

Automated Characterization of Phenotypes

We developed and implemented an automated image processing pipeline in Python (see Methods, SI). Using it, we were able to identify and track every aggregate in all the time series, recording changes in aggregate number, position, size, shape, and gray value. In total, our pipeline captured the developmental dynamics of more than 150,000 aggregates, both dispersing and persisting. These data were analyzed to determine the timing and position of significant changes in swarm dynamics, such as the initial onset of aggregation, the average aggregate growth rate, and the rate of change in aggregate gray value; these quantitative features serve as an unbiased and more accurate replacement for the manual phenotypes "early", "late", "immature", "LWT", and "variable".

We identified 18 quantitative features (Fig. 3) to represent and measure the variation observed in the wild-type and mutant strains. For each time series, we calculated a list of these 18 numbers, mapping it to a single point in an 18-dimensional feature space. Distance between points in this feature space is a measure of phenotypic dissimilarity. To reduce the complexity of these data and improve visualization, we used principal component analysis (PCA), a deterministic method with no additional input parameters, to reduce the feature space to two dimensions, PC1 and PC2. The distribution of points on a two-dimensional map of PC1 versus PC2 captures the phenotypic features that vary the most across the dataset, while discarding those combinations of features that vary less.
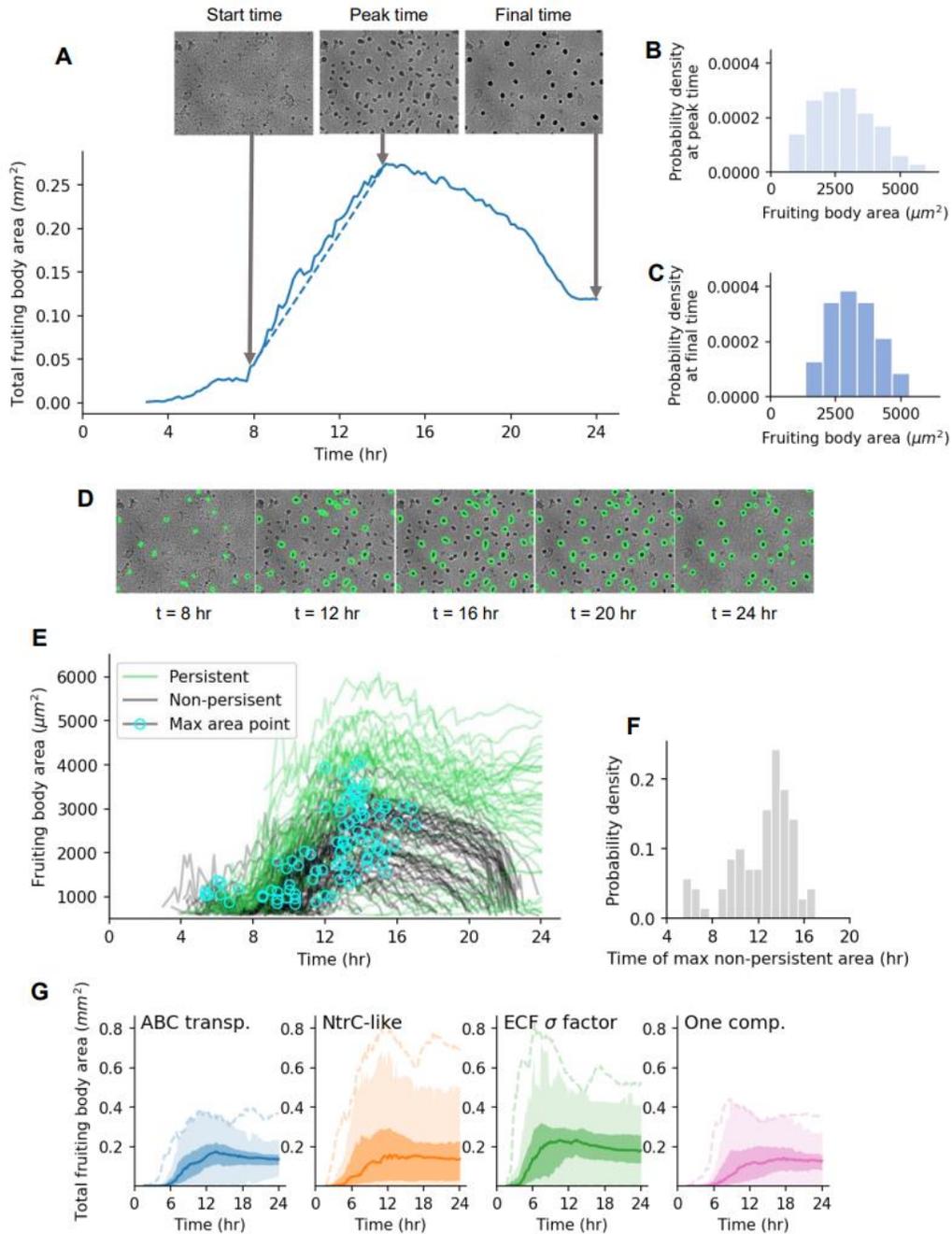
Figure 3: **Automated quantification of fruiting body formation phenotypes. (A-C)** Features related to global fruiting body development **(D-F)** Features related to fruiting body fate **(A)** A representative curve showing total fruiting body area over time in a 7.2 mm² field size. Images are shown of aggregation at start time, peak time, and final time (24 hours), all measured as time elapsed since inoculation (t=0). The slope of the dotted line in **(A)** gives the average growth rate, a key phenotypic feature. **(B)** Representative histograms from the same time series of average fruiting body area at peak time and **(C)** final time. The mean and standard deviation of these distributions are key phenotypic features. **(D)** Five representative time lapse images show fruiting body fate, either to persist (green) or disappear after 24 hours of development. **(E)** Area

versus time curves for each identifiable fruiting body in the same time series. For non-persistent fruiting bodies, the point of peak area is marked with a cyan circle. Two key features are the fraction of total identifiable fruiting bodies that persist (in this case, 32%, or 42 of 132), and **(F)** the standard deviation of the time at which non-persistent fruiting bodies peak in size (temporal coherence). Developmental dynamics can distinguish between time series of different homologous groups, as illustrated in **(G)** the curves for median total fruiting body area over time (quartiles bound the shaded regions, and outliers are bounded by the dotted lines). These variations are captured by 18 phenotypic features, with quantitative definitions given in Appendix 3, supplementary methods.

The most significant phenotypic features are revealed by the makeup of the first two principal components, PC1 and PC2 (Table 1). These two components together account for 43% of the total variance. The constituent parts of both principal components represent a broad array of many different features, with no single outstanding feature. However, there are significant differences between PC1 and PC2. PC1 primarily represents growth rate, mean and standard deviation in fruiting body area at peak time, and mean and standard deviation in fruiting body area at final time. PC2, while sharing mean and standard deviation in area at peak time with PC1, also represents features involved with timing, including growth time, peak time, and temporal coherence. These developmental features with definitions are illustrated in Fig. 3. Although PC2 shares some key features with PC1, its correlations are different. For example, a high value in PC2 indicates high standard deviation in aggregate area and low growth rate, whereas a high value in PC1 indicates high standard deviation in fruiting body area and high growth rate.

Table 1: Makeup of PC1 and PC2 by phenotypic feature. Each primary component is a direction or vector in the 18-dimensional phenotype space, with its makeup shared to varying degree by each feature, with either a positive (blue) or negative (red) correlation. PC1 captures the direction of greatest variance in the overall dataset, and PC2 is the direction perpendicular to PC1 that captures the next greatest amount of variance. The features most strongly represented in each primary component are those that have the greatest potential to distinguish time series phenotypically across the dataset. Each feature is numbered according to its prevalence in PC1.

PC1 Feature Name	Value		PC2 Feature Name	Value	
01) Final area mean	0.363		07) Peak time	0.426	
02) Final area std	0.325		05) Peak area std	0.375	
03) Growth rate	0.322		13) Growth time	0.328	
04) Peak area mean	0.321		04) Peak area mean	0.322	
05) Peak area std	0.297		16) Temporal coherence	0.305	
06) Maturation rate	0.279		02) Final area std	0.298	
07) Peak time	-0.256		17) Stability time	0.229	
08) Gray value % change	0.254		03) Growth rate	-0.195	
09) Max N falloff	-0.221		15) Start time	0.193	
10) Fraction of evaporators	0.217		11) Lifetime std	0.186	
11) Lifetime std	0.209		01) Final area mean	0.172	
12) Max N	0.205		06) Gray value rate	-0.165	
13) Growth time	-0.205		18) Lifetime mean	-0.141	
14) Max mean area falloff	-0.118		08) Gray value % change	-0.133	
15) Start time	-0.106		10) Fraction of evaporators	-0.132	
16) Temporal coherence	-0.077		09) Max N falloff	0.086	
17) Stability time	-0.062		14) Max mean area falloff	-0.058	
18) Lifetime mean	-0.045		12) Max N	-0.007	

Each of the four homologous gene families used in this study has a different number of genes, and therefore each family is not represented by an equal number of mutant strains. This presents a potential bias towards over-represented gene families if PCA were to be performed on the entire dataset. To address this, we performed the PCA multiple times on random samplings of the time series such that each gene family is always equally represented (see Methods). We found that across all such samplings of the full dataset of over 1000 time series, the gene families always form clusters in distinct regions of two-dimensional phenotype space (Fig. 4C). Points within each cluster represent a set of values for the phenotypic features in the PCA; these values represent the "average" or "typical" phenotype for each cluster. There is overlap between the

clusters, so that the differences between clusters only become apparent when using a sufficiently large sample size to visualize an estimate of the probability density. The PCA analysis therefore agrees with the manually derived developmental classifications presented in Fig. 2, in that mutant strains from each gene family display a full spectrum of phenotypes, but there are specific phenotypes that each family exhibits with higher frequency.

There are two important measures of the way in which phenotype clusters form for each gene family. First, the separation between clusters indicates that mutations in each homologous group affect phenotype in a distinct way. Second, the size of each cluster's individual peak points to how often each similar phenotype is expressed within the homologous group. Clusters in the PCA output were shown to be statistically significantly separated and small compared to randomized control grouping rather than grouping by gene family (with p-values of 10^{-5} and 0.0083 respectively) (Fig. S1).

Two of the phenotypic clusters, the ABC transporter mutants and ECF sigma factor mutants, are both centered in a region of PCA space with feature values that denote mature aggregates, but they occupy different parts of that region. This means that, although the "typical" ABC transporter mutant and the "typical" ECF sigma factor mutant do not produce a severe phenotype, the subtle differences in their phenotypes are discernable. The difference in features associated with these two gene families may explain the trends observed in the manual classifications for each homologous family (Fig. 2), such as the tendency for the ECF sigma factor mutants to be classified as "early aggregation" mutants.

The One component cluster is centered in a region that indicates immature aggregates, and the One component mutants also have a disproportionately high number that were manually classified as "variable". In fact, the classification "variable" can be more accurately redefined in

PCA feature space. Instead of being manually classified as a separate phenotype, the variability between replicates can be measured in the PCA as the distance between replicates for the same strain. Representative visualizations and a basic measure of the spread of strain replicates are presented in Fig. S2

The NtrC-like mutants are centered in a region that indicates a severe phenotype with few to no aggregates. The boundaries of this region are outlined in red in Fig. 4B. The initial decision to focus exclusively on features of aggregates when selecting phenotypic features for the PCA may have some negative impact on its resolving power within this non-aggregating region. However, the separation of the NtrC-like activator mutants from the One component mutants supports the idea that the PCA is still capable of distinguishing between subtle differences in mutant strains' phenotypes, even if those strains ultimately do not form aggregates. Details on each typical phenotype and the metric values that distinguish them are available in the SI.

These data support our hypothesis that phenotypic similarity and protein sequence similarity are positively correlated within the *M. xanthus* genome, but the correlation is not strong enough to make specific pairwise predictions of functional redundancy with any kind of accuracy. Pairwise sequence similarity correlates only weakly with phenotypic similarity (Fig. S3), which is consistent with previous findings (Hannay *et al*, 2008). The phenotype clusters for each gene family are populated by replicates of many mutant strains in that family, both the genetically similar and dissimilar. For the ABC transporters, 56% of strains (20 of 36) have replicates within the phenotypic cluster. This is also true for 24% (8 of 34) of NtrC-like activator strains, 39% (15 of 38) of ECF sigma factor strains, and 64% (16 of 25) of One component strains. Fine-grained genetic similarity is not necessary for redundancy to exist, and redundant gene networks may be quite large.
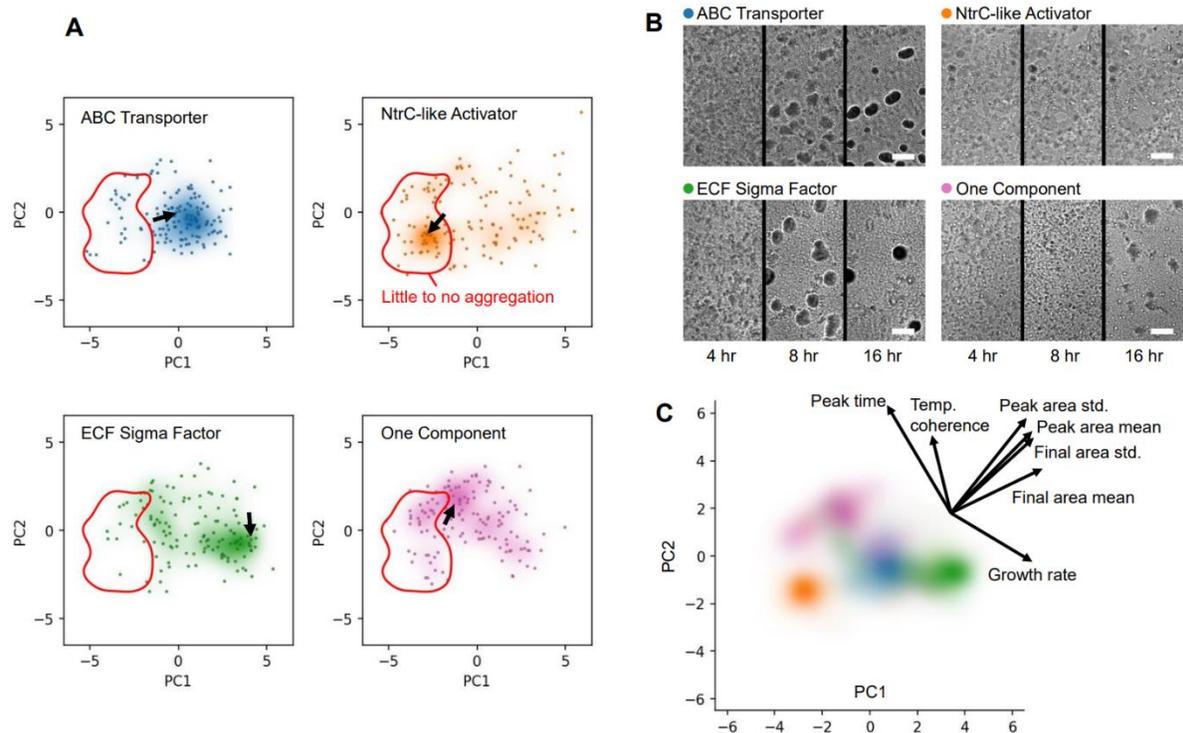
Figure 4: PCA reveals typical phenotypic features for each homologous family. (A) Each point represents a single time series, placed by phenotype according to values of PC1 and PC2. Units for PC1 and PC2 are arbitrary, but (0,0) represents average behavior. Behind points is displayed an estimation of the probability distribution function, using Gaussian kernel density estimation. Higher probability is plotted with higher opacity, revealing phenotypic clusters in each gene family. Outlined in red is a phenotypic zone containing only time series that exhibit little to no aggregation, a severe phenotype. Outside the red zone are successful fruiting body formation time series. An arrow points to a time series typical of the cluster, shown in: **(B)** The typical phenotype in each gene family cluster, illustrated with three frames from a representative time series, taken at 4 hours, 8 hours, and 16 hours after inoculation. Scale bars 100 μm. **(C)** Only the probability distribution estimates for each gene family are shown, illustrating both separation and overlap in phenotypic behavior. The directions of seven key phenotypic features are shown to indicate the coupled meaning of PC1 and PC2. Values of each feature increase in the direction of each respective arrow, with the length of the arrow indicating how much motion in feature space is caused by a fixed increase in the value of that feature, i.e. how significantly the feature is expressed by the two principal components.

DISCUSSION

Goldman et al proposed that the expansion of the *M. xanthus* genome can be attributed primarily to duplication and divergence, which has led to an enrichment of some gene families,

especially those involved in signaling and transcriptional regulation, over others (Goldman *et al*, 2006). This asymmetry of enrichment is notable because it suggests that the expansion of specific gene families holds some evolutionary advantage. We propose at least part of that advantage is to create functional redundancy networks that act as buffers to stabilize *M. xanthus* development (i.e. creating robustness). In this study we confirm that *M. xanthus* development meets the criteria of a robust phenotype: among 265 mutant strains with disruptions in genes that are part of four large homologous families, severe developmental phenotypes are very rare. We then provide support for the existence of large redundancy networks by quantifying the phenotypes of these mutant strains and using PCA to map the phenotypic feature space, which clusters the mutant strains according to the four homologous gene families.

Prior large-scale studies of genetic interactions and redundancies have examined their impact on phenotypic robustness and fitness (Kuzmin *et al*, 2021; Gagarinova *et al*, 2012). Other studies have sought to disentangle the relationship between subsets of multigene families and their roles in redundancy (Johnstun *et al*, 2021). We seek to define the scale and distributive nature of redundancy networks that overlap with large gene families and demonstrate that redundancy networks are not necessarily limited to a few of the closest paralogs but may include dozens of genes. Rather than trying to quantify the direct effect of mutations on fitness by measuring a single variable such as growth, we chose to measure multiple aspects of a complex development process. While our method requires the collection and analysis of time-lapse movies rather than just a few static timepoints, it has the ability to detect more subtle phenotypes that may not have a detectable fitness cost but can still inform studies of redundancy. Since many single gene disruptions have such subtle phenotypes, and since we propose that extensive redundancy networks protect an organism from the fitness costs of mutation, we chose

phenotypic similarity, rather than overall effect on fitness, to assess the extent of functional redundancy within gene families.

Proteins produced by paralogs within a gene family may share similar molecular mechanisms, but they are expected to have different biological functions. For example, the ABC transporters in *M. xanthus* all perform active transport across membranes, but they are expected to transport different substrates. This would mean that a disruption of any one ABC transporter would cause a change in developmental phenotype specific to its substrate. There is no obvious reason why mutant strains of the ABC transporters would display similar changes in phenotype unless there is significant functional redundancy between transporters. There is also no obvious reason why the phenotypic similarities would include a plurality of a large homologous gene family unless the functional redundancy is widely distributed.

When a group of functionally redundant genes mitigates the effect of one member's disruption with low overall stress on the system, the impact on phenotype can be subtle. The ABC Transporter and ECF sigma factor gene families exemplify this, as there are very few single gene knockouts that result in severe phenotypes (Fig. 2E). The phenotypes of these mutant strains cluster by homologous family in the PCA feature space (Fig. 4A), meaning that both gene families display a "typical" phenotype that is different from the other. There are several plausible molecular explanations for this redundancy. Many ABC Transporters, due to varying homology in periplasmic and substrate-binding domains across the gene family, may be able to transport similar and/or overlapping substrates (Orelle *et al*, 2019; Durmort & Brown, 2015), mitigating the effect of many of the mutations in this gene family and most often producing phenotypes close to wild-type. Similarly, robustness has been shown to be encoded in transcriptional regulatory networks by alternative pathways (Wagner & Wright, 2007), and though some studies

suggest that alternative sigma factors display minimal crosstalk (Rhodius *et al*, 2013), it is not unprecedented for there to be overlap in the regulation of genes by multiple ECF sigma factors, creating networks of integrated regulation (Mascher *et al*, 2007; Luo & Helmann, 2009; Paget, 2015). Our data indicate that *M. xanthus* may use such networks of crosstalk among ECF sigma factors to coordinate transcription in response to extracellular signals, and that this may involve integration from many redundant or parallel pathways, ultimately leading to earlier aggregation initiation time and faster fruiting body growth rate than we see in wild-type for the majority of ECF sigma factor mutants.

In contrast, NtrC-like activator mutants show more severe phenotypes and cluster in a region where strains do not form fruiting bodies (Fig. 4). Though it could be argued that a non-aggregating strain indicates a lack of redundancy for the mutated gene, this seems unlikely given that the NtrC-like activators fail to produce aggregates in a way that is distinct from non-aggregating mutants in other gene families (Fig. 4A—region outlined in red). This again points to the idea of networks of redundancy, but highlights that there can be a cost to redundancy in some situations. Extensive research has shown that kinase-response regulator pairs tend to be very insulated with limited crosstalk, and that this feature rapidly evolves in newly-duplicated two-component systems (Capra *et al*, 2012; Laub & Goulian, 2007). NtrC-like activators and other bacterial DNA-binding response regulators have high affinity interactions with their cognate kinases, and crosstalk generally increases noise and decreases the overall response of the system to the incoming signal (Rowland & Deeds, 2014). The specificity of response regulators for phosphorylation by their cognate kinases is governed primarily by molecular recognition, though these proteins can be very sequence similar, and by maintaining a relatively high abundance of response regulator relative to its cognate kinase within a cell to prevent unwanted

phosphorylation (Skerker *et al*, 2008; Capra *et al*, 2012; Podgornaia & Laub, 2013). Taken together, this indicates that mutations to response regulators, like those that we have introduced in the NtrC-like proteins in this study, might lead to a situation where there is a high concentration of phosphorylated kinase in the absence of its highest affinity interaction partner, allowing the cognate kinase to phosphorylate structurally similar non-target response regulators and inappropriately initiate those signaling cascades. This model would explain why so many of the NtrC-like activator mutations produced severe phenotypes that fail to form fruiting bodies in the same way and highlights that redundancy due to gene duplication can have negative consequences without proper insulation.

We do not propose that phenotypic similarity always serves as a strong indicator of functional redundancy. There are almost certainly insignificant associations in the PCA feature space. For example, there are a small number of ABC transporter mutants positioned within the NtrC-like activator cluster. We do not propose that these genes act in similar developmental processes, but we might suggest that they are less likely to be functionally redundant with those in the ABC transporter cluster. It is also possible that any group of completely unrelated genes could have some degree of functional redundancy, but this represents a background level or lower threshold of observable redundancy. We have shown that the redundancy we observe is significantly above that background by measuring the phenotypic clustering of random groupings of genes from the various homologous groups. We are sure there are many forms and many degrees of functional redundancy that are not represented by this PCA, but it does reveal a widely distributed functional redundancy above a background threshold.

The relatively high number of mutant strains with "variable" phenotypes is particularly interesting. Many of these strains manifest a phenotype indistinguishable from wild-type in one

replicate, and then form immature aggregates or fail to aggregate entirely in the next. Though this could be explained by slight variation in experimental conditions across replicates, our observation that this variability occurs more often in the One component family suggests some kind of genetic underpinning (Fig. S2). Given that development occurs in noisy environments, the ability to maintain phenotypic stability in the face of noise contributes to robust development. Networks of redundant genes may be one way that organisms ensure this robustness and contribute to fitness. Perhaps the impact of mutation on some redundancy networks increases an organism's sensitivity to stochastic fluctuations that can influence phenotype toward one developmental fate or another, thus decreasing stability and leading to the "variable" phenotype. If true, redundancy and robustness may be inseparable components of any genotype-to-phenotype problem, and quantifying subtle phenotypic changes in response to mutations may be an effective way to assess their extent within large gene families.

Our results highlight the importance of considering the nature and extent of redundancy when making claims regarding the relationship between genotype and phenotype. Gene families can have high degrees of functional entanglement that may mitigate the impact of mutation, so that quantifying even minor deviations in phenotype may allow for the recognition of patterns; if mutations within a redundancy network produce similar phenotypes, then subtle changes in phenotype have the potential to inform annotation. For example, a gene of unknown function displaying a subtle phenotype similar to that of genes of known function could provide evidence that the unknown gene is part of a redundancy network. Our image analysis pipeline can be extended to future studies of *M. xanthus*, even under differing experimental conditions, for automated extraction of phenotypic features. Further, our dataset can be used to probe whether there are patterns in amino acid sequence homology that lead to functional redundancy by

comparing the sequences of genes that are located within the family cluster on the PCA to those that are located outside the cluster and are presumably non-redundant. Underscoring all these results is the observation that without a sufficiently large collection of mutants and replicates, functional redundancy does not present itself clearly enough to be recognized.

CONCLUSIONS

This work provides evidence for the existence of large networks of redundant genes as a means by which an organism such as *Myxococcus xanthus* can execute complex multicellular social behaviors robust to perturbations to gene function. We observe subtle deviations in phenotype, a distinct set for each homologous gene family, that present when knocking out any one gene within these redundancy groups. These subtle deviations are measurable due to the large number of time series included in our full dataset and the quantitative detail of the extracted phenotypic information, which in combination necessitate the automated analysis pipeline we have developed.

MATERIALS AND METHODS

Strains and Culture Conditions

Myxococcus xanthus strain DK1622 was used as the wild-type for this study. All 265 mutant strains in the ABC Transporter, ECF sigma factor, NtrC-like activator, and One Component Signal Transduction System families (Appendix 2, Table 1) were created using plasmid insertion via homologous recombination as previously described (Caberoy *et al*, 2003; Plamann *et al*, 1995) and modified by Yan et al, 2014. Briefly, 400-600bp internal fragments of each gene were PCR amplified and ligated into pCR®2.1-TOPO [Invitrogen]. The plasmids were amplified in *E. coli* before isolation and electroporation into *M. xanthus* DK1622, where the

plasmid incorporates into the *M. xanthus* genome via the homologous region on the plasmid. PCR verification was used to confirm the location of each insertion.

Cells were grown overnight in CTTYE (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 $KH(H_2)PO_4$ (pH 7.6), 8 mM $MgSO_4$) with vigorous shaking at 32°C. Cultures of mutant strains were supplemented with 40µg/mL kanamycin. Cells were centrifuged to remove the nutrient broth, washed in TPM buffer (10 mM Tris (pH 7.6), 1 mM $KH(H_2)PO_4$, 8 mM $MgSO_4$), and resuspended to a final concentration of $5x10^9$ cells/mL. For development assays, approximately $2.5x10^7$ cells were spotted onto TPM agar slide complexes, as previously described (Taylor & Welch, 2010).

Imaging

Development assays for wild-type and mutant strains were carried out on TPM starvation agar slide complexes for 24 hours, with approximately three replicates per strain. Though it can take multiple days for cells within fruiting bodies to fully differentiate into spores, we generated time series of only the first 24 hours of development because wild-type cells show little to no observable change in fruiting body morphology, count, or behavior following this period at the magnification used. Time-lapse grayscale images were captured every 60 seconds under 4x magnification with a Nikon Eclipse E-400 microscope [Nikon Instruments] and SPOT Insight camera. ImageJ was used for processing the .TIFF images into time series for analysis.

Multidimensional scaling of gene sequence dissimilarity

Amino acid sequences for the four homologous families were retrieved from NCBI and imported into the Multiple Sequence Alignment tool in Clustal Omega (Madeira *et al*, 2022), generating a percent identity matrix for all 265 proteins. This was then converted to a percent

dissimilarity matrix and used as the input for the Classical Multidimensional Scaling package in R to generate plotting coordinates in two dimensions. Then Gaussian kernel density estimation was used to plot an estimate of the probability distribution function (plotted with increased opacity to represent higher probability) to guide the eye in identifying sub-clusters of similar genes within each paralogous group.

Manual phenotyping

Manual preliminary phenotyping of the mutant strains in this study was performed using the time series described above. We will refer to mature aggregates as fruiting bodies for simplicity, though we did not test sporulation efficiency in this study. First, strains that failed to produce fruiting bodies at all within 24 hours across all replicates were labeled "no aggregation" mutants. Strains that formed initial aggregates that disassociated completely before the 24-hour mark were labeled as "fall apart". Some strains, labeled "aggregate-reaggregate", formed aggregates that initially fell apart, but new aggregates were formed that persisted and looked similar to wild-type by the endpoint of the time series.

To qualitatively determine the start time of aggregation, the time series were observed in sliding windows of 25 minutes to identify the window where initial aggregates were first formed. The average start time of 22 wild-type replicates was used for comparison, and mutant strains that had start times outside of one standard deviation of the mean of wild-type were considered either "early" or "late" aggregation mutants.

Wild-type fruiting bodies at 24 hours appear almost black in color and are roughly circular in brightfield images. Any strains that appeared to have these characteristics and initiated aggregation within the same window as wild-type were classified like-wild-type (LWT). Strains that initiated aggregation at a normal time but didn't develop aggregates that were as dark in gray

value as wild-type were labeled as "immature aggregation" mutants. Finally, mutants that did not display consistent phenotypes across replicates were classified as "variable". A table of all mutant strains used in the study, as well as their manually assigned phenotype, can be found in Appendix 2, Table 1.

Automated phenotyping

Phenotype was automatically quantified for the mutant strains 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 information was then collected for the entire time series to track individual fruiting bodies over time, revealing their fate and dynamics. This detailed data summary for each time series then had a list of eighteen specific numbers extracted from it, each of which captures one overall feature, such as average growth rate or the average size of final fruiting bodies. The values of these eighteen metrics together (a phenotypic vector) constitute the phenotype profile for that time series. The full details of the image processing pipeline and all phenotypic metrics are available in the SI.

A selection of 133 mutant time series were chosen at random from each paralogous group so as not to weigh any paralogous group more than the other. The phenotypic vector for each time series was calculated, and the values of each metric were shifted by a constant amount and scaled by a constant factor so that across the dataset, each metric had a mean of zero and a variance of one. This ensured that one metric would not supersede the others simply due to the

magnitude of its units. PCA was performed on this normalized dataset to extract the two combinations of metrics, PC1 and PC2, that captured the most variation across the dataset.

The phenotypic clusters were revealed by plotting each time series as a point in the PC1 vs. PC2 phenotype space and then estimating the probability density for each homologous group via Gaussian kernel density estimation. Essentially, a Gaussian blur was applied to the points, and areas of greater overlap were colored with higher opacity, as shown in Fig. 4. The width of the smoothing kernel was chosen to be the smallest value that could preserve the shape of the probability density for different equally sized subsamples from each homologous group.

The statistical test used to generate the p-values for average cluster separation and average cluster size was a form of bootstrapping which started with the PC1 and PC2 coordinates of each point shown in the data sample of Fig. 4. Each point was randomly reassigned one of four arbitrary families in such a way that replicates of the same strain were all assigned the same family. A new Gaussian kernel density estimation was performed to approximate the probability density of each family in PCA phenotype space. The contour representing 75% of the maximum value of the estimated probability distribution function was then extracted, with cluster separation being the average across all pairings of families of the centroid-to-centroid distance between contours, and cluster size quantified by the average across families of the radius of gyration of each contour, i.e. the root mean square distance of each contour's points from its centroid. Each p-value was calculated as the fraction of the random groupings that had a greater average separation or smaller average size than that of the original data grouped by the actual gene families.

Acknowledgements

REFERENCES

Abellón-Ruiz J, Bernal-Bernal D, Abellán M, Fontes M, Padmanabhan S, Murillo FJ & Elías-Arnanz M (2014) The CarD/CarG regulatory complex is required for the action of several members of the large set of Myxococcus xanthus extracytoplasmic function σ factors. *Environ Microbiol* 16: 2475–2490

Baker EA, Gilbert SPR, Shimeld SM & Woollard A (2021) Extensive non-redundancy in a recently duplicated developmental gene family. *BMC Ecol Evol* 21: 1–17

Bretl DJ & Kirby JR (2016) Molecular Mechanisms of Signaling in Myxococcus xanthus Development. *J Mol Biol* 428: 3805–3830

Butland G, Babu M, Díaz-Mejía JJ, Bohdana F, Phanse S, Gold B, Yang W, Li J, Gagarinova AG, Pogoutse O, *et al* (2008) eSGA: E. coli synthetic genetic array analysis. *Nat Methods* 5: 789–795

Caberoy NB, Welch RD, Jakobsen JS, Slater SC & Garza AG (2003) Global Mutational Analysis of NtrC-Like Activators in Myxococcus xanthus: Identifying Activator Mutants Defective for Motility and Fruiting Body Development. *J Bacteriol* 185: 6083–6094

Capra EJ, Perchuk BS, Skerker JM & Laub MT (2012) Adaptive mutations that prevent crosstalk

enable the expansion of paralogous signaling protein families. *Cell* 150: 222–232

Dean EJ, Davis JC, Davis RW & Petrov DA (2008) Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet* 4

diCenzo GC & Finan TM (2015) Genetic redundancy is prevalent within the 6.7 Mb Sinorhizobium meliloti genome. *Mol Genet Genomics* 290: 1345–1356

Diss G, Ascencio D, Deluna A & Landry CR (2014) Molecular mechanisms of paralogous compensation and the robustness of cellular networks. *J Exp Zool Part B Mol Dev Evol* 322: 488–499

Durmort C & Brown JS (2015) Streptococcus pneumoniae Lipoproteins and ABC Transporters. *Streptococcus Pneumoniae Mol Mech Host-Pathogen Interact*: 181–206

Gagarinova A, Babu M, Greenblatt J & Emili A (2012) Mapping bacterial functional networks and pathways in Escherichia coli using synthetic genetic arrays. *J Vis Exp*: 1–11

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-danila A, Anderson K, Arkin AP, *et al* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature*: 387–391

Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Monning CM, Barbazuk WB, Blanchard M, Field C, *et al* (2006) Evolution of sensory complexity recorded in a myxobacterial genome. *Proc Natl Acad Sci U S A* 103: 15200–15205

Hannay K, Marcotte EM & Vogel C (2008) Buffering by gene duplicates: An analysis of molecular correlates and evolutionary conservation. *BMC Genomics* 9

Johnstun JA, Shankar V, Mokashi SS, Sunkara LT, Ihearahu UE, Lyman RL, Mackay TFC & Anholt RRH (2021) Functional Diversification, Redundancy, and Epistasis among Paralogs of the Drosophila melanogaster Obp50a-d Gene Cluster. *Mol Biol Evol* 38: 2030–2044

Krakauer DC & Plotkin JB (2002) Redundancy, antiredundancy, and the robustness of genomes. *Proc Natl Acad Sci U S A* 99: 1405–1409

Kuzmin E, Rahman M, VanderSluis B, Costanzo M, Myers CL, Andrews BJ & Boone C (2021) τ-SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. *Nat Protoc* 16: 1219–1250

Kuzmin E, Taylor JS & Boone C (2022) Retention of duplicated genes in evolution. *Trends Genet* 38: 59–72

Laub MT & Goulian M (2007) Specificity in two-component signal transduction pathways. *Annu*

Rev Genet 41: 121–145

Lehár J, Krueger A, Zimmermann G & Borisy A (2008) High-order combination effects and biological robustness. *Mol Syst Biol* 4: 1–6

Luo Y & Helmann JD (2009) Extracytoplasmic function σ factors with overlapping promoter specificity regulate sublancin production in Bacillus subtilis. *J Bacteriol* 191: 4951–4958

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A & Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*: 1–4

Mascher T, Hachmann AB & Helmann JD (2007) Regulatory overlap and functional redundancy among Bacillus subtilis extracytoplasmic function σ factors. *J Bacteriol* 189: 6919–6927

Ohno S (1970) Evolution by Gene Duplication Springer Berlin Heidelberg

Orelle C, Mathieu K & Jault JM (2019) Multidrug ABC transporters in bacteria. *Res Microbiol* 170: 381–391

Paget MS (2015) Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules* 5: 1245–1265

Plamann L, Li Y, Cantwell B & Mayor J (1995) The Myxococcus xanthus asgA gene encodes a novel signal transduction protein required for multicellular development. *J Bacteriol* 177: 2014–2020

Podgornaia AI & Laub MT (2013) Determinants of specificity in two-component signal transduction. *Curr Opin Microbiol* 16: 156–162

Rees DC, Johnson E & Lewinson O (2009) ABC transporters: the power to change. *Nat Rev Mol Cell Biol* 10: 218–27

Rhodius VA, Segall-Shapiro TH, Sharon BD, Ghodasara A, Orlova E, Tabakh H, Burkhardt DH, Clancy K, Peterson TC, Gross CA, *et al* (2013) Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol Syst Biol* 9: 1–13

Rowland MA & Deeds EJ (2014) Crosstalk and the evolution of specificity in two-component signaling. *Proc Natl Acad Sci U S A* 111: 5550–5555

Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M & Laub MT (2008) Rewiring the Specificity of Two-Component Signal Transduction Systems. *Cell* 133: 1043–1054

Taylor RG & Welch RD (2010) Recording Multicellular Behavior in Myxococcus xanthus

Biofilms using Time-lapse Microcinematography. *J Vis Exp*: 1–6

Thomaides HB, Davison EJ, Burston L, Johnson H, Brown DR, Hunt AC, Errington J & Czaplewski L (2007) Essential bacterial functions encoded by gene pairs. *J Bacteriol* 189: 591–602

Vandersluis B, Bellay J, Musso G, Costanzo M, Papp B, Vizeacoumar FJ, Baryshnikova A, Andrews B, Boone C & Myers CL (2010) Genetic interactions reveal the evolutionary trajectories of duplicate genes. *Mol Syst Biol* 6: 1–13

Wagner A & Wright J (2007) Alternative routes and mutational robustness in complex regulatory networks. *BioSystems* 88: 163–172

Xie C, Zhang H, Shimkets LJ & Igoshin OA (2011) Statistical image analysis reveals features affecting fates of Myxococcus xanthus developmental aggregates. *Proc Natl Acad Sci U S A* 108: 5915–20

Yan J, Bradley MD, Friedman J & Welch RD (2014) Phenotypic profiling of ABC transporter coding genes in Myxococcus xanthus. *Front Microbiol* 5: 1–12
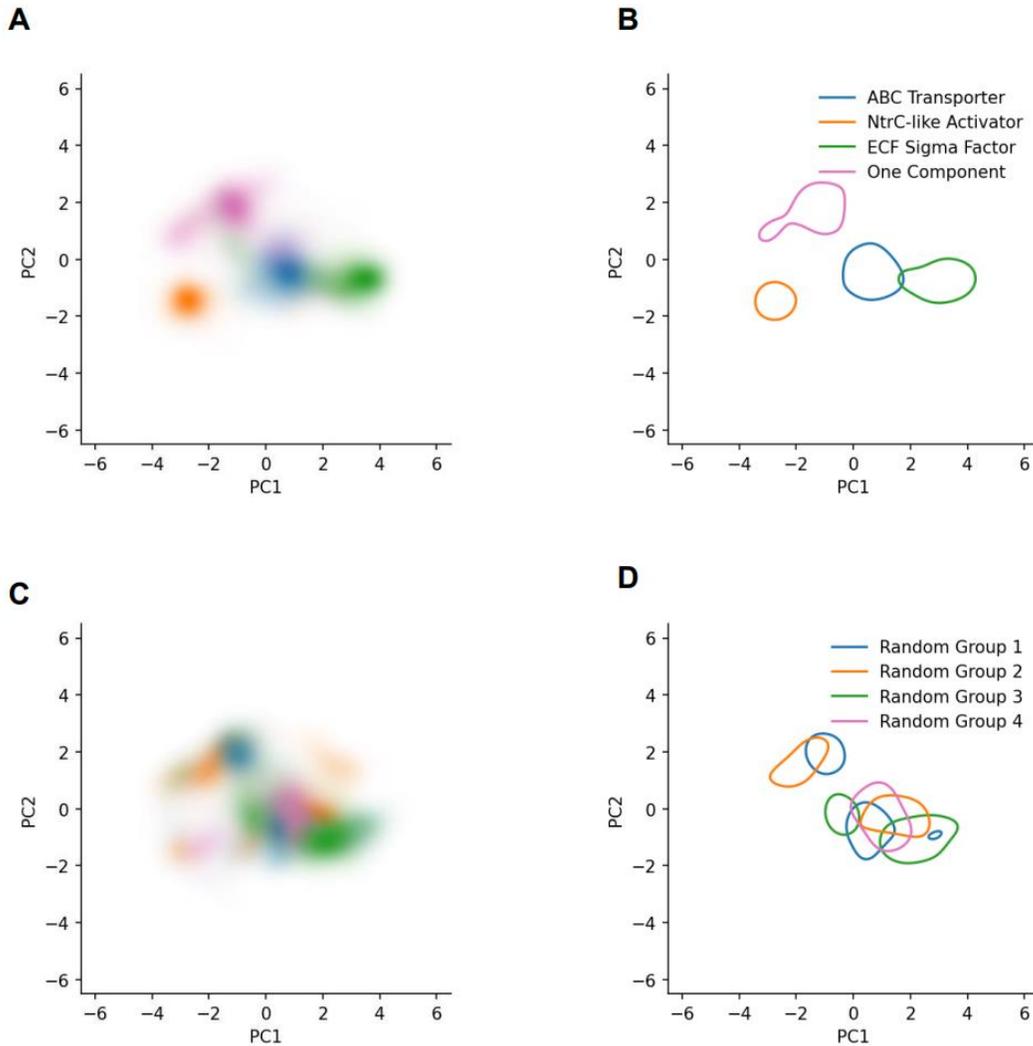
Figure S1: Phenotypic clusters arise robustly from homologous gene families as compared to random groupings of mutant strains. **(A)** Reproduced from Figure 4, each of the four gene families produces a distinct phenotypic cluster when plotting the estimated probability distribution function for that family (using Gaussian kernel density estimation) in phenotype space. **(B)** A contour is shown for each gene family where the estimated probability distribution function is at 75% of its maximum value, and the geometry of that contour is used to quantify the width of the cluster and its separation from other clusters. **(C)** The same data used in Figure 4 was regrouped into four random groupings, and the PCA and probability density function estimates were repeated, showing much more incoherent phenotype clusters. **(D)** The corresponding contours for the four random groupings show less separation and less sharpness than phenotypic clusters based on homologous gene groups. Both (C) and (D) come from one representative random grouping, many of which were made to calculate the p-values for cluster separation and sharpness reported in Results.
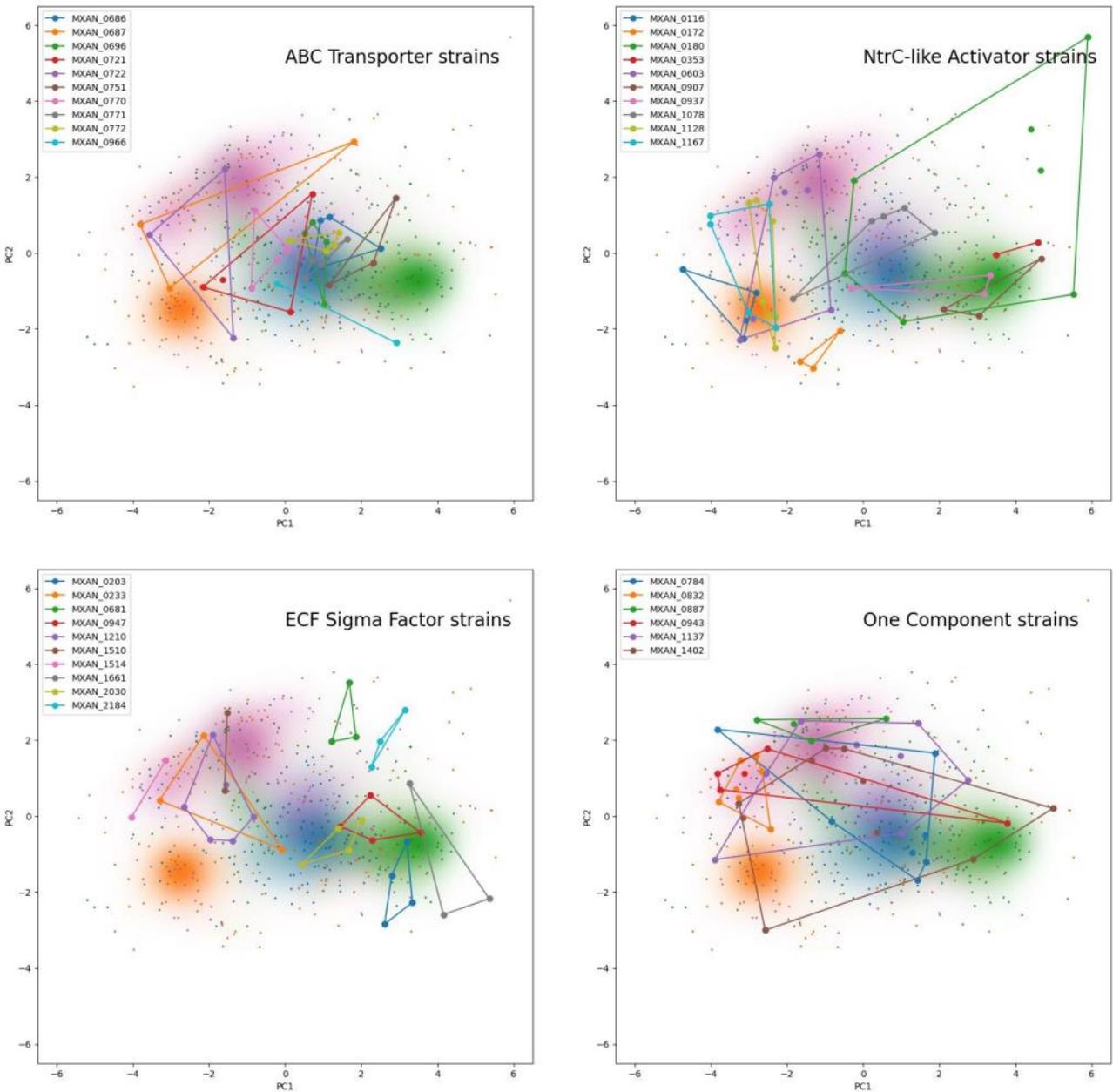
Figure S2: Replicates of the same strain can vary in phenotype. Reproduced for context from Figure 4 are the phenotypic scatterplot resulting from the PCA (where each point is a time series, plotted nearby other time series that are phenotypically similar) and the superimposed probability distribution functions for each of the homologous gene families: ABC transporters in blue, NtrC-like activators in orange, ECF sigma factors in green, and One component in pink. Each subplot includes all replicates of a few representative strains, where the replicates of each strain are represented in a single color and drawn with a bounding polygon to aid the eye. Replicate-to-replicate variation is larger or smaller depending on strain and to which homologous family the strain belongs.

A metric for replicate-to-replicate phenotypic spread of a specific strain is the sample standard deviation s generalized to two dimensions

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left((x_i - \bar{x})^2 + (y_i - \bar{y})^2\right)}$$

where n is the number of replicate points, x_i and y_i are the coordinates of the ith replicate point, and $\bar{x}$ and $\bar{y}$ are the means of each coordinate across the replicate points, i.e. the centroid coordinates. In this case, the x and y coordinates are the value of PC1 and PC2 respectively. Table S1 summarizes the mean replicate-to-replicate phenotype spread averaged over strain for each gene family using the metric s, with errors given by the standard error of the means.

Table S1: A summary of the average replicate-to-replicate spread for each homologous gene family, with errors given by the standard error of the means. This spread is illustrated for some representative

Gene family	Mean replicate-to-replicate spread s
ABC Transporters	1.42 ± 0.12
NtrC-like Activators	1.57 ± 0.14
ECF Sigma Factors	1.31 ± 0.09
One Component	1.94 ± 0.18

This indicates a statistically significant difference for replicate-to-replicate phenotype spread between One Component strains and ABC Transporter strains ($p = 0.024$), and between One Component strains and ECF Sigma Factor strains ($p = 0.004$) according to a two-sided Welch's t-test.
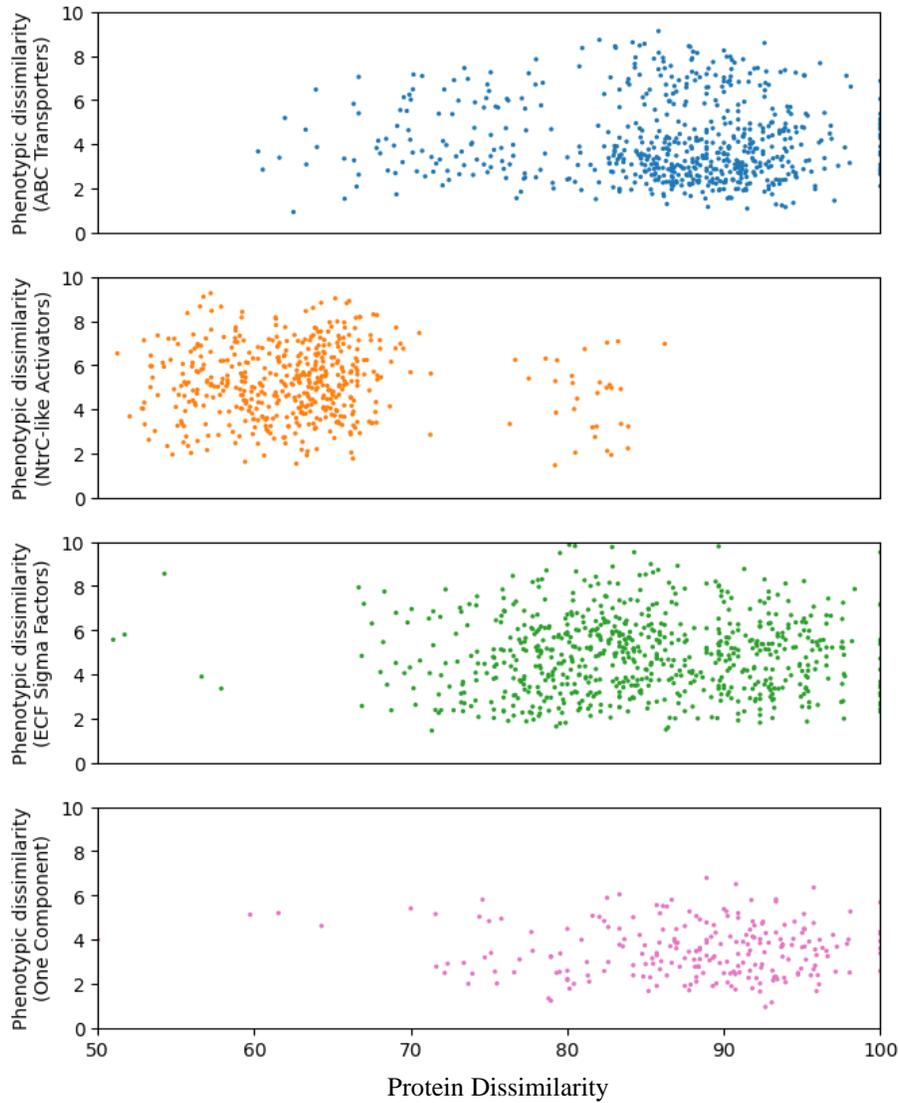
Figure S3: Protein similarity is not an effective predictor of phenotypic similarity within a homologous gene family. For each of the four gene families analyzed, each point represents a unique pairing of two strains. Phenotypic dissimilarity is quantified by Euclidean distance in 18-dimensional feature space, where feature values are represented by averages over all replicates for that strain. Protein dissimilarity is quantified by comparison of base pairs using Clustal Omega Multiple Sequence Alignment. Within each of the four gene families, phenotypic dissimilarity and protein sequence dissimilarity do not correlate.

CHAPTER 3: TRAVELING WAVES SYNCHRONIZE CELL MOTILITY DURING M. XANTHUS DEVELOPMENT

Jessica A. Comstock[1], Fatmagül Bahar[1], Patrick A. Murphy[2], Oleg A. Igoshin[2,3], and Roy D. Welch[1]

[1] Department of Biology, Syracuse University, Syracuse, NY, USA
[2] Department of Bioengineering, Rice University, Houston, TX, USA
[3] Center for Theoretical and Biological Physics, Rice University, Houston, TX, USA

(formatted for submission to *Proceedings of the National Academy of Sciences*)

(Data collection was predominantly by me with some time-lapse movies originating from FAB's data collection in Chapter 2. Conceptualization and data analysis done by me, with cell tracking performed by PAM in the lab of OAI.)

ABSTRACT

Waves play many varying yet critical roles in the development of organisms. Early mitotic divisions, nerve impulses, cardiac contractions, would healing, and chemotactic mechanisms are a small fraction of the numerous biological processes that rely on traveling chemical waves. Though the study of biological waves has traditionally been focused on higher organisms, bacteria also exhibit numerous types of waves that play roles in long-range communication, chemotaxis, and spatial patterning. In this work, we provide evidence for a previously uncharacterized wave phenomenon in the soil bacterium *Myxococcus xanthus*, which self-organizes into spore-containing fruiting bodies under starvation as a component of its stringent response. These waves, which we call pulses, occur in both wild-type and genetic mutant strains and are distinct from rippling waves, a well-characterized emergent behavior expressed by *M. xanthus* swarms during predation. Pulses originate from early aggregates in a developing swarm and cause cells to move more persistently by suppressing reversals that results in synchronized motility across the swarm. This increase in persistence in the presence of slime trails could be sufficient to move more cells into aggregates, which would drive aggregation forward. Pulses represent a new emergent behavior pattern discovered through observing a collection of over 1,000 time-lapse movies of development which highlights the importance of high-throughput phenotypic quantification.

INTRODUCTION

Across all orders of life, key biological processes must be coordinated to ensure proper multicellular development. At the molecular level, this means that signals must be able to travel great distances, such as the length of a developing organism, over short periods of time to synchronize biological functions and cellular organization. Diffusion alone is not sufficient to

coordinate over such distances, requiring many developing organisms to make use of more expedient methods for transduction of signals and messages. Chemical waves are one such mechanism that developing organisms use for spatiotemporal coordination.

Examples of chemical waves acting to coordinate multicellular development and self-organization are widespread in biology. Upon fertilization, the timing of early embryonic cell divisions are governed by waves of Cdk1 activity (1, 2). Calcium waves also play a role in mitosis (3), oocyte activation (4), and synchronized contractions of cardiac tissue (5). Chemical waves can also act to direct cell motility and migration. Traveling waves of actin are responsible for locomotion in fibroblasts, neutrophils, keratocytes, and many other cell types (6). Propagating waves of cAMP help direct the slime mold *Dictyostelium discoidem* into aggregates via chemotaxis during its starvation response (7). In bacteria, waves can also help establish spatial patterning, as in the Min-protein waves that help *E. coli* determine the cell center prior to cell division (8), and contribute to long-range communication and multicellular organization into biofilms (9–11). Given the prevalence of self-organization and patterning in biological systems, phenomena which are often governed by waves, it is likely that waves play many undiscovered roles in different systems.

The soil bacterium *Myxococcus xanthus* exhibits an almost entirely multicellular lifestyle (12). It lives in soil environments and can saprophytically feed on decaying organic matter to obtain nutrients (13). Upon contact with prey bacteria such as *E. coli* or other soil microbes (14), *M. xanthus* acts as a predatory swarm, secreting enzymes to lyse prey cells and scavenge their nutrients (15). The *M. xanthus* predation strategy involves characteristic rippling waves caused by crests of high cell density and troughs of low cell density that reflect off one another and reverse direction upon collision (16, 17). Nutrient stress (in particular nitrogen starvation)

triggers a second multicellular behavior called development, where cells in a swarm transition away from swarming and begin to cluster into mounds. These nascent aggregates continue to grow as starvation persists, accumulating more cells, some of which will differentiate and form a mass of spores at the center of the aggregate as it matures into a fruiting body (18). Such self-organization requires individual cells to sense starvation and then to communicate this information at a population level, promoting massive changes in gene expression, motility, and initiating signaling cascades (19) across relatively large spatial scales in their environment.

We observed *M. xanthus* development with time-lapse microcinematography (20), capturing the process of fruiting body formation under different genetic and environmental conditions, using image analysis to better understand developmental dynamics. In grayscale brightfield images, swarm activity is perceived as changes in the gray value of the individual pixels in each frame, as cell movement causes fluctuations in the level of light coming through to the camera. In essence, movement is captured through changes in pixel gray value. Changes in pixel value were measured from subsequent frames in our time-lapse image series as a proxy for swarm activity, generating plots of swarm activity that showcase a previously uncharacterized oscillatory phenomenon during *M. xanthus* development that we call pulsing. We define the characteristics of a pulse and document its effects on fruiting bodies within a developing swarm. We provide evidence that pulses are traveling chemical waves emitted by nascent fruiting bodies, and these pulses help to synchronize cell motility during development.

RESULTS

Observing common developmental patterns with time-lapse microcinematography.
M. xanthus cells on starvation agar sense nutrient limitation and initiate a developmental process

where the once vegetatively swarming cells aggregate into mounds which eventually mature to become spore-filled fruiting bodies (Figure 1A). As part of an ongoing effort to understand the full phenotypic range of *M. xanthus* development, we generated time-lapse movies of 265 mutant strains containing insertion-disruption mutations in genes with putative roles in important biological processes (signal transduction, gene regulation, and transport of ions and small molecules) (Appendix 2, Table 1) (strain generation: (20, 21)). We specifically recorded the first 24 hours of development on starvation agar to observe patterns that resulted from the mutations. This large-scale analysis revealed a recurring pattern seen in many different mutant strains involving a periodic disturbance in the stability of early aggregates, causing temporary changes in average aggregate area, gray value, and circularity, before the aggregate eventually stabilizes. Cells appear to move in pulses, spreading temporarily outward in groups from the aggregate centers (Figure 1B), causing the aggregate to increase in area and become more irregularly shaped and less dense, before coalescing into a more dense and circular aggregate again (Figure 1C).
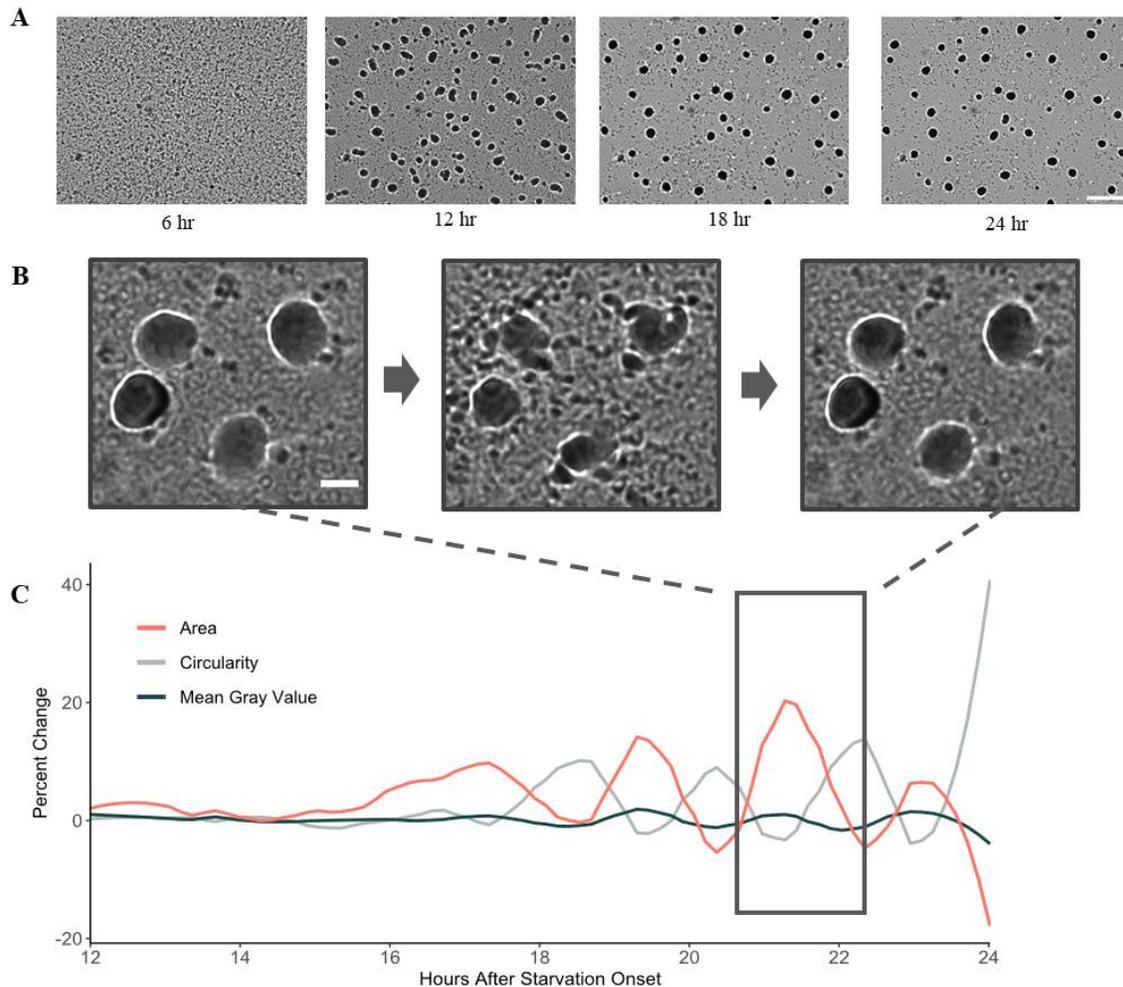
Figure 1. Periodic disturbances affect the stability of early aggregates. A) *Myxococcus xanthus* development on TPM starvation agar at 6hr timepoints following the onset of starvation (0hr). Initial aggregates in wild-type movies appear between 8-10hrs and cells accumulate into the aggregates and eventually mature into fruiting bodies. Scale bar 500µm. B) After initial formation of aggregates, there is often a disturbance in the swarm that causes a periodic and temporary instability in early aggregates of cells where stable, growing aggregates (left) transition to unstable, branched aggregates (middle), before coalescing back into rounded aggregates again (right). These images were each taken 45 minutes apart. Scale bar 100µm. C) Percent change in area, circularity, and mean gray value of early aggregates that occurs during the observed periodic instability. Values were averaged from multiple manually segmented fruiting bodies (n=7) at 20-minute intervals from 12-24hrs post-starvation. Periodic increases in fruiting body area are accompanied by decreases in circularity and slight increases in pixel gray value as aggregates get larger, more irregular in shape, and brighter.

Activity plots improve visualization of swarm dynamics. Brightfield images rely on light passing through a sample, with less translucent objects appearing darker. Within an *M. xanthus* swarm this means that regions of higher cell density appear darker. For example,

developing aggregates become darker as they accumulate more cells and mature into fruiting bodies. For this study, *M. xanthus* development was recorded as stacks of 1,440 images taken at one-minute intervals, with each image a 1600x1200 pixel array, and each pixel assigned a gray value from 0 to 255 (black to white). The dynamics of development were therefore recorded in the changing pixel values between sequential images, which represent local regions of cell density shifting as cells move. To quantify and reduce the complexity of these dynamics, we generated a Difference Stack Image Sequence for each brightfield time-lapse movie that reflects only what is changing from frame to frame. If a pixel remains the same from frame *n* to frame *n+1,* it has a difference of zero and is black. If a pixel changes from frame *n* to frame *n+1*, it will vary on a scale of 1 (gray) to 255 (white) depending on the magnitude of the difference. Thus, we viewed only what is changing from one frame to the next, or the activity of the swarm. We then plotted the average difference in pixel gray value for each frame over time to show a plot of swarm dynamics, or the Activity Plot.

The Activity Plot for a swarm undergoing development is seen in Figure 2. In phase 1, cells experience a short initial lag time with no movement while motility machinery is assembled, followed by a sharp increase in activity as cells perceive starvation and initiate their response. In phase 2, marked by the highest swarm activity, cells organize themselves into initial aggregates. Finally, phase 3 contains three key features: a) maturation of initial aggregates, defined by an increase in gray value and often a slight decrease in area as fruiting bodies grow taller, b) disappearance of some aggregates as previously described in the literature (20, 28), and c) previously undescribed and uncharacterized pulses of swarm activity that are consistent with the nature of the pulses observed by eye in brightfield movies. Because of differences in wave period and timing of the onset of pulsing, we have chosen to display the activity plot of one

representative time-lapse movie rather than reporting an average, and others can be seen in

Figure 4.
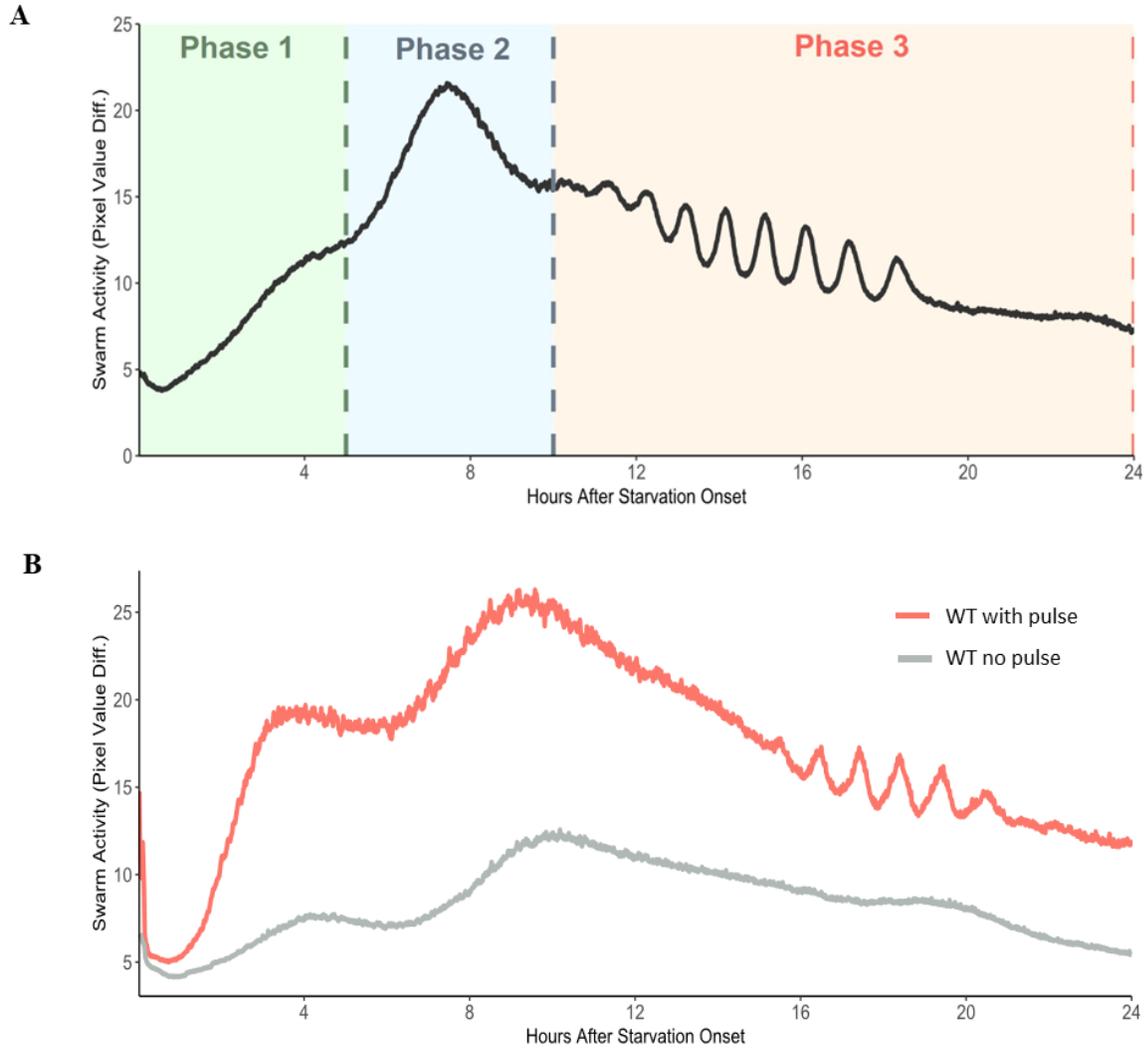
Figure 2. Activity plots represent swarm dynamics during development. A) Average pixel gray value differences represent swarm activity in each frame, with a greater value representing more swarm activity. In phase 1, the pre-aggregation phase, there is a steady increase in swarm activity. Phase 2 represents the peak of activity, during or after which fruiting bodies begin to form. Finally, maturing of fruiting bodies occurs in phase 3, where synchronized oscillations of high and low swarm activity occur, corresponding to the timing of pulses observed in the brightfield image series. This plot represents one mutant movie that shows pulses. For additional plots of other strains, see Figure 4. B) Representative wild-type activity plots both with and without synchronized pulses across the swarm.

Pulsing is common and is a component of wild-type development. Generating activity plots for time-lapse movies of each mutant strain in triplicate revealed that 56.6% of mutants exhibited pulsing under standard development assay conditions in at least one replicate. More strains, particularly those that aggregate late compared to wild-type, are likely capable of pulsing if the assay were carried out longer than 24 hours. The distribution of pulsing mutants within each gene family from Ch. 2 was relatively equal; however, the distribution of pulsing strains within each phenotypic category was unequal as determined by a chi-square test ($p=2.2e^{-16}$, df=7). Pulsing is significantly overrepresented in both early aggregating and like wild-type phenotypic categories as determined by Fisher's exact tests ($p=2.9e^{-5}$ and $2.1e^{-10}$, respectively). Over 80% of mutant strains within these categories demonstrate pulsing. Pulsing also occurs in about 50% of the mutants that produce variable phenotypes in different replicates. Significantly fewer of the late and immaturely aggregating strains showed pulsing, and no pulsing was detected in any mutants that did not at least initiate aggregation.

Interestingly, activity plots of wild-type movies also revealed that pulsing happens during wild-type development as well, only more subtly and with less obvious instability of the aggregates (Figure 3). An activity plot analysis of 22 replicates of wild-type movies showed that low-level pulsing occurs in 59% of wild-type replicates, suggesting that pulsing is a normal component of wild-type development that is exaggerated for the genetic mutants in which we first observed the behavior. Additionally, pulses seem to be provoked by certain environmental conditions. In wild-type, we have observed pulsing that destabilizes aggregates in a similar fashion as the mutants (Figure 4) when manipulating the concentration of potassium phosphate in the buffer, but not by altering other buffer components.

The representative activity plots for pulsing samples shown in Figure 3 display some of the variation that we observed from sample to sample. Most notably, there were differences in the onset of pulses, the earliest being three hours after starvation, and some beginning just before the end of the 24-hour time series. The only commonality was that pulses always occur after initial aggregation. Even when pulses happen very early in a movie, as in MXAN_1189 (Figure 4), aggregation precedes their onset. There were also differences in the number of pulses as well as in the period of the oscillations. Importantly, these differences were not driven by genetic differences between strains, as replicates of the same mutant strain varied widely in terms of these features. The period of pulses in both high and low potassium phosphate environments are much longer than what is observed in wild-type or the mutants (Figure 4), but fruiting bodies seem to destabilize in a similar fashion to those in genetic mutant pulsing samples. The way that the pulse affects the fruiting bodies also varied in different pulsing samples. For example, as in wild-type, some mutants such as MXAN_4316 (Figure 4) show pulsing without any obvious destabilizing of the aggregate, while others, such as MXAN_2407 or MXAN_4523, pulses cause a much more discernable disturbance within the early fruiting bodies. Again, this cannot be attributed to genetics alone, as there is within-strain variation in the extent of destabilization, albeit determined qualitatively. Note that we did not attempt to quantify and compare pulse amplitude in different replicates, as this metric would be more indicative of the initial contrast present in the brightfield time-lapse, which varies with each movie and holds no biological significance.
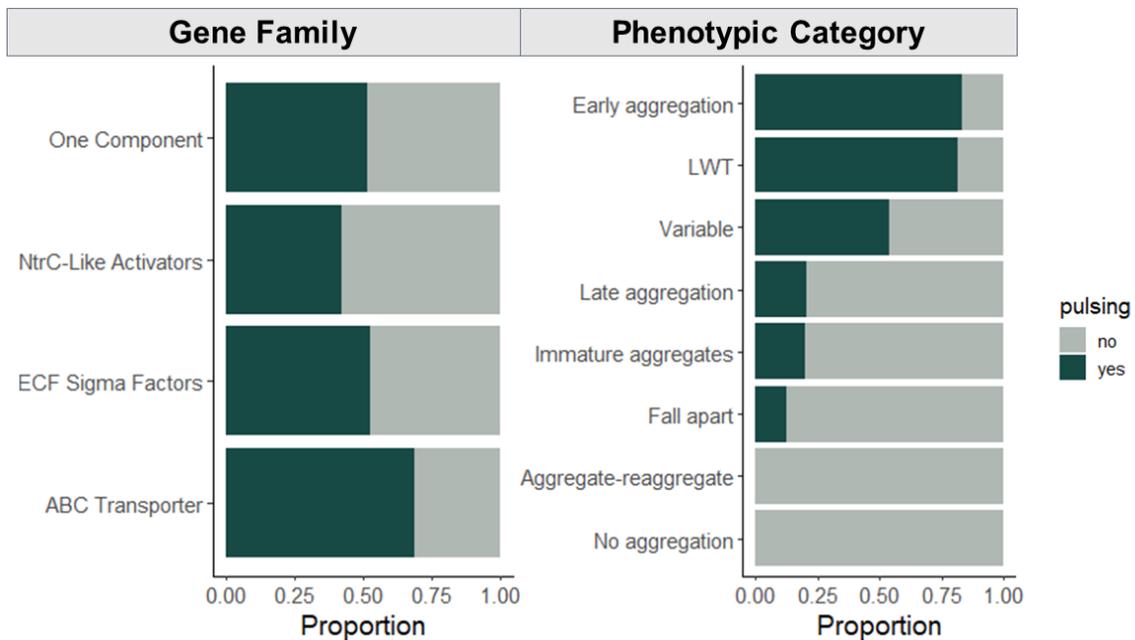
Figure 3. Distribution of pulsing mutants. A distribution of pulsing mutants within the gene families and phenotypic categories presented in Chapter 2. Strains were considered pulsing if at least one of three replicates showed pulsing on the activity plot. Note that strains considered non-pulsing may demonstrate pulsing with a lower frequency and may be able to pulse if additional replicates were collected. LWT, like wild-type. Broken down by gene family, 70 of 102 (69%) of ABC Transporters, 20 of 38 (53%) of ECF sigma factors, 19 of 45 (42%) of NtrC-like activators, and 41 of 80 (51%) of One Component signaling genes show pulsing. The distribution of pulsing mutants within each phenotypic category is more unequal. 39 of 47 (83%) of early aggregating strains, 79 of 97 (81%) of LWT strains, 5 of 24 (21%) of late aggregating strains, 5 of 25 (20%) of immature aggregating strains, 21 of 39 (54%) of variable phenotype strains, and 1 of 8 (12.5%) of fall apart strains show pulsing. Pulsing was not seen in any replicates of 0 of 6 aggregate-reaggregate strains and 0 of 19 non-aggregating strains.
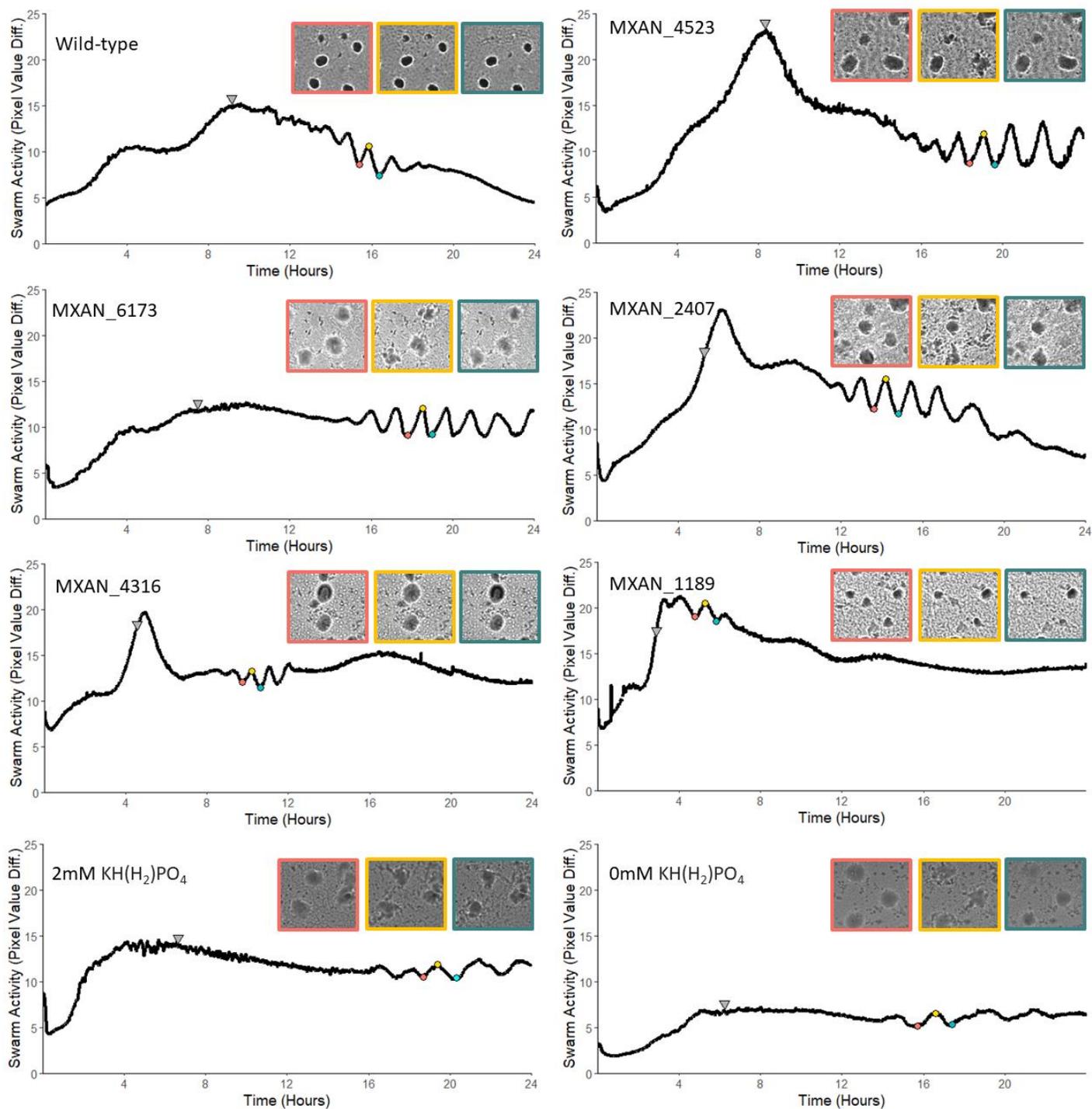
Figure 4. Representative activity plots for a subset of strains and conditions that provoke pulsing. All plots show the characteristic activity oscillations associated with pulsing, but the onset of pulsing as well as the pulse period appear to be different across samples. Curves for wild-type, four different mutants, as well as low and high potassium samples are shown. Inset images represent fruiting bodies before, at the peak, and after the pulse, corresponding to the colored points indicated on the curve. Gray arrows indicate the onset of aggregation. Insets each represent a 322x548 µm fraction of a larger field of view.

Pulses are waves that are distinct from rippling. To test whether the pulses happen simultaneously in the population of early fruiting bodies or whether they travel across the swarm, we measured the timing of the oscillations in activity in different regions of the field of view and found that there is a lag between the pulses from one region of the field of view to the other, indicating a directional wave (Figure 4). The waves traveled through the swarm with an average speed of 54.5±15µm/min (from n=30 measurements from different movies). For comparison, individual *M. xanthus* cells were reported to move an average of approximately 5 µm/min (29). These findings suggest that the waves that we see in the stack difference time-lapse movies are not the result of cells moving across the swarm with the same speed as the pulse, but rather a signal moving through the swarm causing a temporary increase in cell activity as it passes. This evidence suggests that pulses are a type of wave entirely distinct

Figure 5. Calculation of wave speed. The pulse is a wave that travels across the field of view. This developing population of *M. xanthus* shows the periodic pulses beginning at 10 hours post-starvation. There appear to be five clear pulses that cause aggregate instability. Boxes indicate two regions of interest for activity plots shown below, and the arrow indicates the direction of the wave seen in the Difference Stack movie, sweeping from the lower left to the upper right of the field of view. Using cross-correlation, we detect a lag of 15 min across a distance of 891 µm, indicating a wave speed of 59.4 µm/min. The histogram shows distribution of wave speeds (n=30, bins=16).

from rippling, the other type of wave behavior exhibited by starving *M. xanthus* populations. We

have also observed swarms where both rippling and pulsing co-occur, both wave types

seemingly unaffected by the other, and very different in terms of scale and speed.
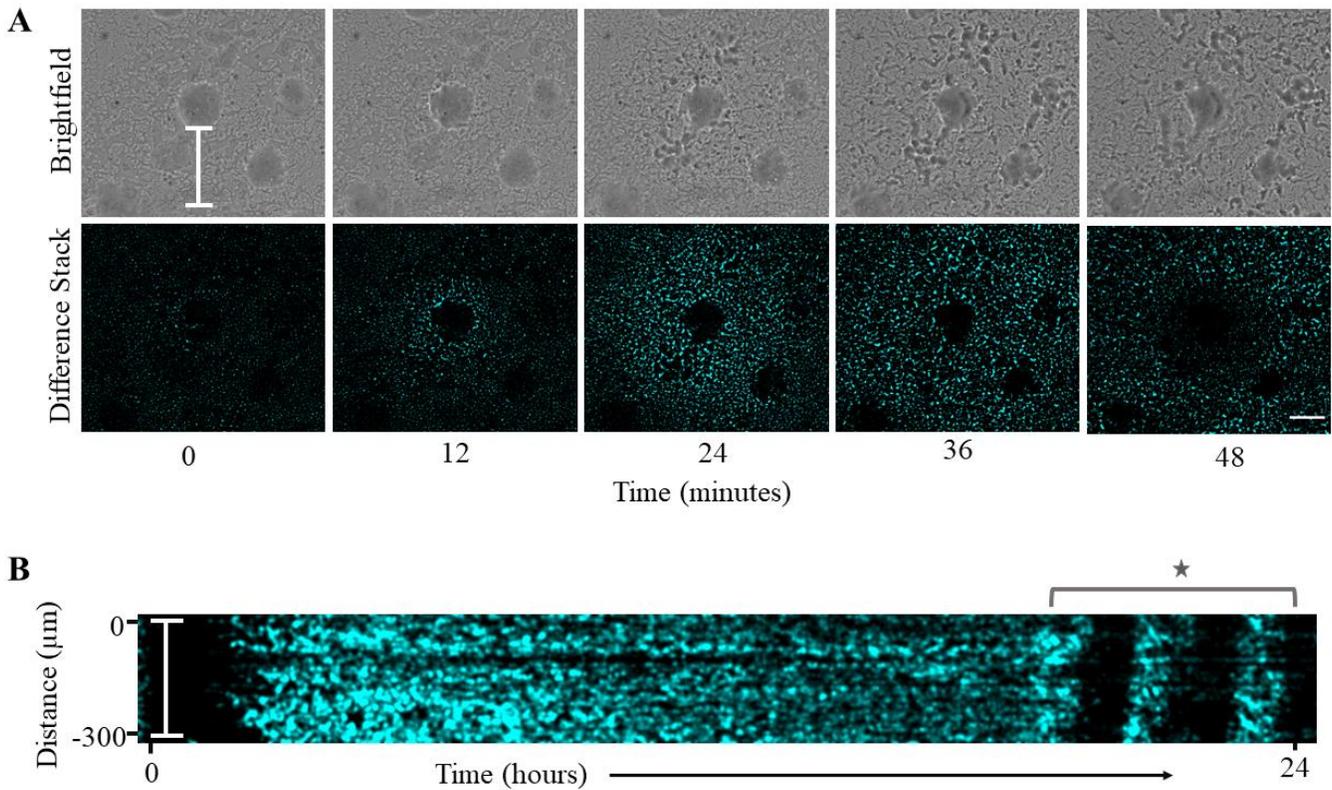
Figure 6. Pulses originate at aggregates. A) Representative brightfield images and Difference Stack Image Sequence showing the origin of a pulse. Over the course of 48 minutes, the burst of activity originates at the aggregate and propagates radially outward. Corresponding aggregate instability can be seen in the brightfield image. Scale bar 100 µm. B) Kymograph taken from the Difference Stack Image Sequence in A, over a distance of 295 µm from the aggregate, as indicated by white bar. The bracket indicated by the star highlights the three distinct pulses that occur in this time series, with three repeated pulses emanating from the aggregate. Kymograph is smoothed using a Gaussian blur with a radius of 2 pixels to eliminate noise and improve visualization.

Pulses originate at aggregates. We have never observed pulsing in movies where no

aggregates have formed. Additionally, the onset of pulses seems to be dictated, at least in part, by

the timing of aggregation initiation, since pulsing can happen earlier than it does on average in

mutant strains that aggregate much earlier than wild-type (Figure 4). We therefore hypothesized that pulses originate from aggregates. We searched the collection of time-lapse development movies for instances where the pulse seemed to originate within the field of view rather than traveling into the field of view from elsewhere in the swarm. In all instances where we saw the origin of the pulse, the oscillations in swarm activity indicative of the pulse began at one aggregate and radiated outward repeatedly (Figure 5). We have also observed instances where pulses initiate from multiple aggregates in the field of view, which adds a layer of complexity to the swarm dynamics. The fact that multiple fruiting bodies can initiate a pulse seems to suggest that a condition within the local environment of the fruiting body triggers a pulse, and this can happen in multiple regions within the swarm.

Cells become more persistent and increase in alignment during pulses. Though the pulses travel too quickly to be exclusively due to cells traveling the same distance and speed as the wave front, it is clear that the pulses have some detectable effect on cell behavior. We used fluorescently labeled cells, diluted 1:800 into the swarm, to track the behavior of individual cells and show that in a swarm undergoing multiple distinct pulses, cells oscillate between long runs where cells move persistently without reversing, and periods of high reversal with little net movement. As the pulse occurs, cells suppress reversals, traveling greater distances without changing direction. Corresponding peaks in run duration and distance traveled without reversal can be seen in Figure 6. Following this, reversal frequency increases and the oscillations cause a corresponding increase in alignment of neighboring cells (Figure 6). Taken together, these data indicate that the pulses are waves, likely of a chemical signal, that cause cells to suppress reversal in a synchronized manner. This synchronized increase in distances traveled by cells is followed by periods of high reversal which increase alignment of neighboring cells, providing a

mechanism by which a greater number of cells will move persistently in the same direction during the next pulse.
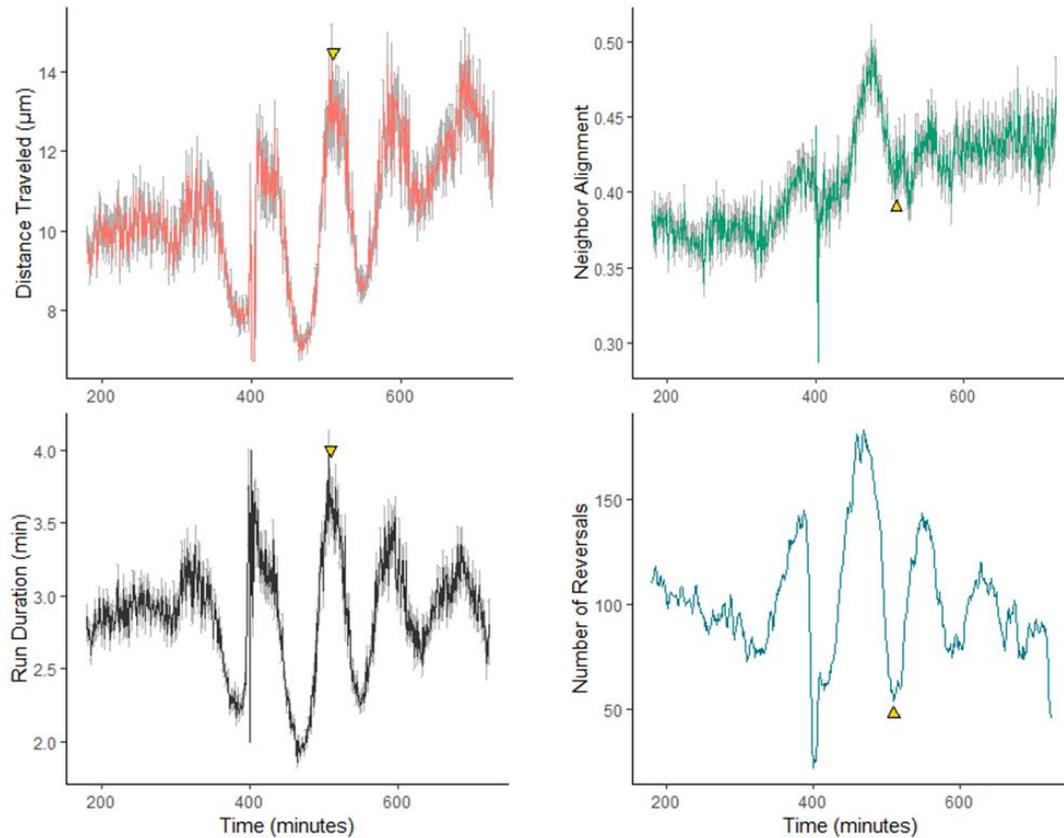
Figure 7. The effect of pulses on cell behavior. Data extracted from automated tracking of tdTomato-labeled cell trajectories. Figures show: distance traveled before reversal, run duration (the length of time cells move persistently without reversing), alignment of cells with its neighbors, and number of reversals. Plots (except for number of reversals) reflect trajectory means for each time frame, with grey bars showing standard error. All means are smoothed over a window of ±5 minutes. Gold arrows indicate the same timepoint on each of the four plots, showing that distance traveled and run duration are inversely related to alignment and number of reversals.

DISCUSSION

The data presented here demonstrate the prevalence and characteristics of pulses during *M. xanthus* development. There are several lines of evidence to suggest that pulses are a component of wild-type development, not an exclusively mutant phenotype, and that an

93

environmental condition during development triggers pulses to be more exaggerated. Given the number of mutant strains that show pulsing, we do not think that the genes included in the study are part of the genetic pathway that allows pulsing to occur. Rather, we suggest that the large number of mutations that lead to strong pulsing might be producing a similar overall effect on development such that pulsing is provoked.

First, over half of the wild-type replicates show pulsing. This highlights the importance of visualizing and quantifying pulses, as these pulses in wild-type do not affect the fruiting bodies as strongly as they do in mutants and therefore the pulses are not easily observed in the brightfield movies without the aid of the activity plots. Second, we found that genotype is a poor predictor of a strain's ability to pulse. As shown in Figure 3, there is roughly equal representation of pulsing within each homologous gene family. There is slight overrepresentation in the ABC Transporters and underrepresentation in the NtrC-like activators, but this is more likely driven by the distribution of phenotypes within these gene families. As discussed in Chapter 2, the ABC Transporters are primarily made up of like-wild-type strains, and many of the NtrC-like activators mutants fail to aggregate. We have shown here that the vast majority of like wild-type strains pulse, whereas no pulsing is seen in non-aggregating mutants (Figure 4). Together, these associations are likely responsible for the lower representation of pulsing mutants within the NtrC family and the higher representation within the ABC Transporters.

The average speed of the pulses points to a mechanistically distinct type of wave as compared to rippling, a well-studied wave behavior in *M. xanthus*. Ripples are caused by waves of cells that reverse direction upon collision (17), and the speed of rippling is therefore constrained by the maximum speed at which cells can move. Prior work has shown that during predation, rippling encourages swarm expansion and reduces the mean square displacement of

individual cells, which both promote prolonged contact with prey cells (30, 31). The cellular reversals that allow for ripple formation are mediated by the Che-like Frz chemosensory system, where contact between two countermoving waves of cells stimulates collective reversal (32, 33). In contrast, pulses move through the swarm with an average speed of 55 µm/min, or 0.9 µm/s, about 10-fold higher than the average speed of an *M. xanthus* cell. The full range of observed pulse speeds in this study is 0.5-1.6 µm/min. The speed of pulses is in much better agreement with the speed of traveling chemical waves (34). Waves of Cdc42 activation, for example, travel through *Dictyostelium* cells at an average speed of 1.1 µm/s to direct motility during chemotaxis (35). Apoptotic signals in *Xenopus* eggs (36), thrombin localization to wound sites (37), and Min-protein waves in *E. coli* (38) all travel with a velocity within the range we observed in this study.

We have also observed both *M. xanthus* wave behaviors occurring simultaneously; when a pulse passed through an area of the swarm that was also rippling, each wave type was seemingly unaffected by the other, and rippling continued as the pulse passed over it. Additionally, the two waves are distinct in scale, with the rippling wavelength being much shorter. While both rippling and pulsing involve cellular reversal, the simultaneous occurrence of both wave types indicates either that the two waves are mechanistically different in the genes that they target to trigger reversal, or that all cells within a swarm are not responsive to the pulse at the same time. Subpopulations of cells with varying gene expression profiles are well documented within the *M. xanthus* developmental scheme. Persister-like peripheral rods remain outside of fruiting bodies even during late stages of aggregation and they express very low levels of key developmental genes (39, 40), so perhaps this subpopulation does not have the means to respond to pulses based on its transcriptome during development.

These data presented here suggest that pulses are chemical waves that originate from early aggregates and synchronize cell motility at the population level during development. We propose a model where pulses serve as a mechanism to increase the cell density within aggregates, which would likely lead to greater sporulation efficiency and therefore greater fitness. Once initial aggregates reach sufficient density, cells within the aggregate release a chemical signal to which other cells respond by triggering the release of that chemical signal, thus creating a propagating wave. Additionally, the signal causes a shift in the motility patterns of the cells, temporarily increasing run distance and duration and suppressing reversal (Figure 7). We make no claims about this signal causing a bias in cell direction toward the fruiting bodies, only that the signal transiently suppresses cell reversal frequency and therefore might increase the likelihood of a cell to get to an aggregate. The ability for *M. xanthus* to chemotax up a chemical gradient is a subject of controversy. It has been theorized that, because *M. xanthus* moves so slowly relative to diffusion over short distances, chemotaxis would not be possible. This is further supported by early work that failed to detect a bias in cell direction toward common chemoattractants, including major nutrient sources casein and yeast extract (41). However, other work has shown attraction of *M. xanthus* cells to certain lipids (42), and there is additional evidence that chemotaxis may be an emergent property of a swarm rather than a behavior of individual cells (43).

A more likely scenario at play in our study is slime-trail following as a mechanism for directed cell movement toward aggregates. As *M. xanthus* moves through its environment, its gliding motility machinery extrudes an extracellular matrix "slime trail" from the lagging pole of the cell (44, 45). Gliding bacteria are known to follow slime trails laid by other cells (46), and slime trail following in *M. xanthus* has been experimentally observed (47) as well as

theoretically modeled to show that it facilitates cell movement into groups (48). Further data-driven modeling indicates the existence of biased random walks toward aggregate centers, indicative of a chemosensory mechanism; however the authors note that increased run duration in cells outside of aggregates can compensate for the absence of the observed biased movement (25). Taken together, these data suggest that even in the absence of a chemotactic signal, if a pulse causes cells to increase run duration and distance traveled before reversal, as we have observed through cell tracking, cells may be more likely to move into aggregates via slime trails. Every cell that has joined an aggregate leaves a slime trail in its wake, essentially creating a roadmap to aggregates that other cells can follow during their persistent state.

Notably, during wild-type development the pulse does not seem to impact the cells within the fruiting bodies, at least not to the extent that we see in mutants. Whereas mutant aggregates often destabilize during a pulse, wild-type aggregates for the most part remain stable. This suggests that wild-type pulsing primarily affects cells outside of aggregates, in support of our model, but also indicates that cells within developing wild-type fruiting bodies may not be responding to the pulse in the same way. We can speculate about what might cause this. Perhaps cells within aggregates are hindered by local cell density and/or the ECM that provides scaffolding and structural support to fruiting bodies, and mutants that show exaggerated pulsing form lower density fruiting bodies or have reduced ECM complexity. Perhaps there is a genetic mechanism that prevents wild-type cells within fruiting bodies from responding to the pulse, for example downregulation of a particular receptor, and exaggerated pulsing represents a dysregulation in the genetic coordination of that mechanism such that pulses occur before cells in the fruiting body are stabilized.

The prevalence of pulsing in wild-type and mutants brings up a key question: why does pulsing fail to happen about 40% of the time? Perhaps the strength of the pulse is related to the proportion of cells that remain outside of the fruiting body when pulses initiate. If most cells are within fruiting bodies, then the signal cannot be propagated, and pulsing would not be detected. However, if many cells still remain outside of fruiting bodies when initial aggregates form, the signal can be amplified by more cells. Or, perhaps pulses do occur during every instance of development, but are sometimes less synchronized and therefore less easily detected when calculating the mean over the entire swarm at a given time.

Our model proposes a new emergent behavioral pattern that was discovered through collecting and observing a large dataset of approximately 1,500 time-lapse movies. Without sufficient data collection, pulsing could be ascribed to one mutant or a small group of mutants, rather than observing its prevalence in wild-type development. The pulsing phenotype would also likely be overlooked entirely if we had taken images at 4-to-6-hour timepoints, as is traditional for observing *M. xanthus* mutant development. This study highlights the importance of collecting large, comprehensive datasets to build and characterize an organism's phenome. It also indicates a potential mechanism of inter-aggregate communication and warrants future study into chemical signals that may be causing the pulse.

METHODS

Strain Generation and Culture Conditions

Myxococcus xanthus DK1622 was used as the wild-type. Mutant strains used in this study were generated previously (21–23) and are listed in Table 1 of Appendix 2. Cells were grown overnight in CTTYE (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 $KH(H_2)PO_4$, 8

mM MgSO$_4$) with vigorous shaking at 32°C. Cultures were supplemented with 40µg/mL

kanamycin for mutant strains. Cells were centrifuged to remove the nutrient broth, washed in

TPM starvation buffer (10 mM Tris (pH 7.6), 1 mM KH(H$_2$)PO$_4$, 8 mM MgSO$_4$), and

resuspended in TPM buffer to a final concentration of $5x10^9$ cells/mL. For development assays, a

droplet of approximately $2.5x10^7$ cells was transferred onto TPM starvation agar slide

complexes, as previously described (24). For low and high potassium development assays, TPM

was prepared without the addition of KH(H$_2$)PO$_4$, or with 2mM KH(H$_2$)PO$_4$, respectively.

TdTomato-expressing strain LS3908 (25) was used for cell tracking experiments. This

strain was supplemented with 10 µg/mL oxytetracycline for selection and 1mM isopropyl β-D-1-

thiogalactopyranoside (IPTG) for induction of tdTomato expression. These cells were diluted

1:800 into DK10547 GFP-expressing wild-type cells (26) and prepared for development assays

as described above. LS3908 cells were used for tracking and DK10547 cells were used for

aggregate segmentation and assessment of local cell densities.

Time-lapse Image Capture

Slide complexes prepared as above were incubated at 32°C during imaging with the use

of a heated stage. Microscopic images were captured every 60 seconds for 24 hours on a Nikon

Eclipse E400 microscope using a SPOT Insight Camera and SPOT Imaging Software (SPOT

Imaging^{TM}). The .TIFF images were then compiled into time-lapse movies and further analyzed

in FIJI (27). For cell tracking experiments using fluorescence, imaging was performed on the

same Nikon Eclipse E400 microscope with a pco.panda 4.2 sCMOS camera and NIS-Elements

software. LS3908 samples were imaged with 400 ms exposure with a Sola LED light source at

75% intensity, and DK10547 samples were imaged with 200 ms exposure at 35% intensity.

Control of the fluorescent filter wheel and autofocus mechanism was managed with a MAC6000 system filter wheel controller and focus control module (Ludl Electronic Products, Ltd.).

Image Analysis and Calculations

Initial calculations of changes in circularity, gray value, and area were performed in Fiji. Seven fruiting bodies were randomly selected from the same movie and manually traced every 20th frame from 12-24 hours to obtain circularity, area, and mean gray value data for each fruiting body over time. Values for each of the seven fruiting bodies were averaged and percent change over time was plotted for a representative movie in Figure 1C.

The Difference Stack Image Sequences were created using the Stack Difference tool in Fiji. This results in a sequence of images that are a 1600x1200 array of pixels with an assigned gray value based on the difference pixel values from one frame to the next, or |(pixels in frame n+1) – (pixels in frame n)|, where any pixel that did not change is 0 and therefore black, and any pixel that did change will be a range from 1 to 255 (gray to white), where a higher value represents a greater magnitude change and therefore greater swarm activity. Any images of Difference Stacks presented in this work contain false coloring through cyan or magenta LUTs to provide additional contrast for better visualization. Activity plots represent the average pixel value difference for each frame in the Difference Stack Image Sequence over time, with a greater average pixel value difference indicating greater swarm activity. Any images of Difference Stacks presented in this work contain false coloring through cyan or magenta LUTs. Wave speed was calculated for 30 different movies where one clear directional wave could be seen. Two ROIs were selected that were perpendicular to the traveling front of the wave, and activity plots were generated for each ROI. We divided the distance between the two ROIs by the time shifts

in the peak of oscillations for the activity plots corresponding to each ROI to approximate wave speed.

For cell tracking image stacks, we constructed cell trajectories according to procedures laid out by Cotter et. al (25) and performed by Patrick Murphy at Rice University. We used their model to segment cell trajectories into different motility states, either persistent or non-persistent. For each persistent cell, a trajectory records the distance covered and duration of time before a reversal (state change), when a new trajectory begins. In this way, any trajectory ends at a reversal, so we can also calculate the number of reversals happening at each timepoint. For each trajectory, alignment to neighboring cells was also calculated by averaging cosines of the angles of nearby cells. Specifically, neighbor alignment was calculated for a window of ±5 min and a radius of 15 µm around the start coordinate of a run using the equation:

$$< \Omega_n > = \frac{1}{N} \sum_{i \in window} \cos\left(2(\chi_n - \chi_i)\right)$$

where nematic alignment (neighbor alignment) Ω_n is influenced by N, the number of runs within a window, and χ, the angle between a run and the x axis. We report the absolute value of this neighbor alignment metric so that a 0 indicates complete misalignment at a 1 indicates perfect alignment. Means over a moving window of 10 minutes were calculated for these datasets (with the exception of number of reversals) for plotting.

REFERENCES

1. Deneke VE, Melbinger A, Vergassola M, Di Talia S (2016) Waves of Cdk1 Activity in S Phase Synchronize the Cell Cycle in Drosophila Embryos. *Dev Cell* 38(4):399–412.

2. Chang JB, Ferrell JE (2013) Mitotic trigger waves and the spatial coordination of the Xenopus cell cycle. *Nature* 500(7464):603–607.

3. Tombes RM, Borisy GG (1995) THE ESSENTIAL ROLES O F CALCIUM DURING MITOSIS. *Adv Mol Cell Biol Mol Cell Biol* 13:69–87.

4.   Kaneuchi T, et al. (2015) Calcium waves occur as Drosophila oocytes activate. *Proc Natl Acad Sci U S A* 112(3):791–796.

5.   Bers DM (2002) Cardiac excitation–contraction coupling. *Nature* 415(January):198–205.

6.   Allard J, Mogilner A (2013) Traveling waves in actin dynamics and cell motility. *Curr Opin Cell Biol* 25(1):107–115.

7.   Loomis WF (2014) Cell signaling during development of dictyostelium. *Dev Biol* 391(1):1–16.

8.   Wettmann L, Kruse K (2018) The min-protein oscillations in Escherichia coli: An example of self-organized cellular protein waves. *Philos Trans R Soc B Biol Sci* 373(1747). doi:10.1098/rstb.2017.0111.

9.   Humphries J, et al. (2017) Species-Independent Attraction to Biofilms through Electrical Signaling Article Species-Independent Attraction to Biofilms through Electrical Signaling. *Cell* 168:200–209.

10.   Prindle A, et al. (2015) Ion channels enable electrical communication in bacterial communities. *Nature* 527(7576):59–63.

11.   Liu J, et al. (2015) Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature* 523:550–554.

12.   Muñoz-Dorado J, Marcos-Torres FJ, García-Bravo E, Moraleda-Muñoz A, Pérez J (2016) Myxobacteria: Moving, killing, feeding, and surviving together. *Front Microbiol* 7(MAY):1–18.

13.   Shimkets LJ (1984) Nutrition, Metabolism, and the Initiation of Development. *Myxobacteria*, pp 91–107.

14.   Mendes-Soares H, Velicer GJ (2013) Decomposing predation: Testing for parameters that correlate with predatory performance by a social bacterium. *Microb Ecol* 65(2):415–423.

15.   Thiery S, Kaimer C (2020) The Predation Strategy of Myxococcus xanthus. *Front Microbiol* 11(January). doi:10.3389/fmicb.2020.00002.

16.   Berleman JE, Chumley T, Cheung P, Kirby JR (2006) Rippling is a predatory behavior in Myxococcus xanthus. *J Bacteriol* 188(16):5888–5895.

17.   Sliusarenko O, Neu J, Zusman DR, Oster G (2006) Accordion waves in Myxococcus xanthus. *Proc Natl Acad Sci U S A* 103(5):1534–1539.

18.   Zusman DR, Scott AE, Yang Z, Kirby JR (2007) Chemosensory pathways, motility and development in Myxococcus xanthus. *Nat Rev Microbiol* 5(11):862–872.

19.   Bretl DJ, Kirby JR (2016) Molecular Mechanisms of Signaling in Myxococcus xanthus Development. *J Mol Biol* 428(19):3805–3830.

20.   Curtis PD, Taylor RG, Welch RD, Shimkets LJ (2007) Spatial organization of Myxococcus xanthus during fruiting body formation. *J Bacteriol* 189(24):9126–9130.

21.   Yan J, Bradley MD, Friedman J, Welch RD (2014) Phenotypic profiling of ABC

transporter coding genes in Myxococcus xanthus. *Front Microbiol* 5(July):1–12.

22. Ritchie LJ, Curtis ER, Murphy KA, Welch RD (2021) Profiling Myxococcus xanthus Swarming Phenotypes through Mutation and Environmental Variation. *J Bacteriol* 203(23). doi:10.1128/JB.00306-21.

23. Comstock JA, et al. (2022) Phenotypic similarity is a measure of functional redundancy within homologous gene families. *bioRxiv*. doi:10.1101/2022.07.25.501402.

24. Taylor RG, Welch RD (2010) Recording Multicellular Behavior in Myxococcus xanthus Biofilms using Time-lapse Microcinematography. *J Vis Exp* (42):1–6.

25. Cotter CR, Schüttler H, Igoshin OA, Shimkets LJ (2017) Data-driven modeling reveals cell behaviors controlling self-organization during Myxococcus xanthus development. *Proc Natl Acad Sci*. doi:10.1073/pnas.1620981114.

26. Welch R, Kaiser D (2001) Cell behavior in traveling wave patterns of myxobacteria. *Proc Natl Acad Sci U S A* 98(26):14907–14912.

27. Schindelin J, et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9(7):676–682.

28. Xie C, Zhang H, Shimkets LJ, Igoshin OA (2011) Statistical image analysis reveals features affecting fates of Myxococcus xanthus developmental aggregates. *Proc Natl Acad Sci U S A* 108(14):5915–20.

29. Spormann AM, Kaiser D (1999) Gliding mutants of Myxococcus xanthus with high reversal frequencies and small displacements. *J Bacteriol* 181(8):2593–2601.

30. Zhang H, et al. (2012) The Mechanistic Basis of Myxococcus xanthus Rippling Behavior and Its Physiological Role during Predation. *PLoS Comput Biol* 8(9). doi:10.1371/journal.pcbi.1002715.

31. Berleman JE, Scott J, Chumley T, Kirby JR (2008) Predataxis behavior in Myxococcus xanthus. *Proc Natl Acad Sci U S A* 105(44):17127–17132.

32. Mauriello EMF, Astling DP, Sliusarenko O, Zusman DR (2009) Localization of a bacterial cytoplasmic receptor is dynamic and changes with cell-cell contacts. *Proc Natl Acad Sci U S A* 106(12):4852–7.

33. Kaimer C, Berleman JE, Zusman DR (2012) Chemosensory signaling controls motility and subcellular polarity in Myxococcus xanthus. *Curr Opin Microbiol* 15(6):751–757.

34. Deneke VE, Di Talia S (2018) Chemical waves in cell and developmental biology. *J Cell Biol* 217(4):1193–1204.

35. Yang HW, Collins S, Meyer T (2016) Locally excitable Cdc42 signals steer cells during chemotaxis. *Nat Cell Biol* 18(2):191–201.

36. Cheng X, Ferrell JE (2018) Apoptosis propagates through the cytoplasm as trigger waves. *Science (80-)* 361(6402):607–612.

37. Dashkevich NM, et al. (2012) Thrombin activity propagates in space during blood

coagulation as an excitation wave. *Biophys J* 103(10):2233–2240.

38. Loose M, Fischer-friedrich E, Ries J, Kruse K, Schwille P (2008) Spatial Regulators for Bacterial Cell Division Self-Organize. *Science (80-)* 320(May):789–792.

39. Julien B, Kaiser a D, Garza a (2000) Spatial control of cell differentiation in Myxococcus xanthus. *Proc Natl Acad Sci U S A* 97(16):9098–9103.

40. Müller FD, Treuner-Lange A, Heider J, Huntley SM, Higgs PI (2010) Global transcriptome analysis of spore formation in Myxococcus xanthus reveals a locus necessary for cell differentiation. *BMC Genomics* 11(1). doi:10.1186/1471-2164-11-264.

41. Dworkin M, Eide D (1983) Myxococcus xanthus does not respond chemotactically to moderate concentration gradients. *J Bacteriol* 154(1):437–442.

42. Kearns DB, Shimkets LJ (1998) Chemotaxis in a gliding bacterium. *Proc Natl Acad Sci* 95(20):11957–11962.

43. Taylor RG, Welch RD (2008) Chemotaxis as an emergent property of a swarm. *J Bacteriol* 190(20):6811–6816.

44. Yu R, Kaiser D (2007) Gliding motility and polarized slime secretion. *Mol Microbiol* 63(2):454–467.

45. Wolgemuth C, Hoiczyk E, Kaiser D, Oster G (2002) How myxobacteria glide. *Curr Biol* 12(5):369–377.

46. Burchard RP (1982) Trail following by gliding bacteria. *J Bacteriol* 152(1):495–501.

47. Islam ST, Mignot T (2015) The mysterious nature of bacterial surface (gliding) motility: A focal adhesion-based mechanism in Myxococcus xanthus. *Semin Cell Dev Biol* 46(November):143–154.

48. Balagam R, Igoshin OA (2015) Mechanism for Collective Cell Alignment in Myxococcus xanthus Bacteria. *PLoS Comput Biol* 11(8):1–20.

CHAPTER 4: UNCOVERING CELL BEHAVIORS ASSOCIATED WITH COARSENING DURING DEVELOPMENT

(All experimental data presented in this chapter were collected by me, and I was responsible for much of the experimental design and conceptualization. Cell tracking experiments will contribute to data-driven simulations designed by Patrick Murphy and Oleg Igoshin at Rice University. Conceptualization and experimental work for PI staining data presented herein is my own.)

ABSTRACT

The biological patterns formed by the soil bacterium *Myxococcus xanthus* make it an interesting system for modeling the cellular behaviors that lead to self-organization and multicellularity. Many previous models, even those that reproduce the initiation of aggregation in a very similar manner to experimental data, fail to capture aggregate dynamics during the coarsening phase of aggregation, namely the dispersal of smaller aggregates. Very little is known about the mechanistic basis for aggregate dispersal. Data-driven modeling is advantageous in this scenario, as experimental cell behavior data can be applied to simulated cells without needing to understand the mechanisms directing that behavior, and can improve our understanding of which cellular behaviors are necessary for a population of aggregates to coarsen. This chapter lays out the initial data extracted from cell tracking experiments demonstrating that cell bias toward aggregates temporally depends on aggregate size, data which will inform coarsening simulations. We also investigate cell loss that occurs during coarsening, demonstrating that up to 80% of cells are lost during the coarsening phase. This cell loss within fruiting bodies correlates with an increase in extracellular DNA, suggestive of cell lysis, and aggregates that persist through coarsening have a greater eDNA density, a mechanism that may be used to provide stability and increase stress resistance.

INTRODUCTION

Myxococcus xanthus has long been of interest to those looking to model the behaviors that lead to self-organization of multicellular communities and organismal development. *M. xanthus* exhibits two interesting biological patterns under distinct environmental conditions. In the presence of prey cells, it forms ripples that are crests of high cell density that reverse upon collision, and models contributed to the understanding of ripples both mechanistically and

physiologically (1–5). Under starvation, cells coordinate to form multicellular fruiting bodies in a dynamic process known as development. Modeling of aggregation has also contributed to our understanding of the features required for fruiting body formation, as well as the behaviors of individual cells that lead to this emergent behavior (5–12).

   M. xanthus development into fruiting bodies is a very dynamic process that involves coordinated timing, aggregates of finite size ranges and spatial organizations, and behaviors at the level of the aggregates such as growing, shrinking, merging, splitting, and dispersing. Previous models of myxobacterial multicellularity have reproduced many of the features of aggregation, but most fail to capture the aggregate dispersal that we see so often in experimental data. For example, though the traffic jam model (wherein colliding streams of cells nucleate to form aggregates driven by a reduction in cell motility in high-density areas (4, 13)) reproduces aggregates with a similar quantity and spatial distribution as is observed in experimental data, the model does not fully capture the dynamics, including dispersal, that occur during development. Previous studies have estimated that only 30-50% of nascent aggregates mature into fruiting bodies (6, 14). Because the aggregation phase resembles the behavior of droplets on a surface, we refer to the later stage dynamics of aggregates as coarsening, where some aggregates get larger at the expense of others.

   Coarsening follows the initial developmental phases of aggregation and growth, and represents the period of time where aggregates either remain stable and mature into fruiting bodies or shrink and disperse, and coarsening ends with aggregates reaching a stable state. Though we can predict, based only on size, which aggregates will disperse (14, 15), we know very little about the mechanisms and biological significance of aggregate dispersal. Data-driven models, where simulated cell behavior is based on data from experimental cell tracking, have the

potential to inform the mechanistic basis of coarsening. In this chapter, I lay out the initial data extracted from the fluorescent images that I contributed to this project and show that the bias toward aggregates during coarsening is dependent on the size of the aggregate, which confirms what we believe to be true about the dynamics of coarsening based on aggregate size. These behaviors, in addition to run time, run duration, and cell alignment, will be applied to simulated cells in the data-driven model currently being developed by our collaborators Patrick Murphy and Oleg Igoshin to understand the minimum set of cell behaviors that are required for coarsening.

During coarsening in the tracking datasets, we observed a notable decrease in the number of trackable cells from initial aggregation through the end of coarsening. This loss of cells was reproducible throughout many datasets with fluorescently labeled cells. We sought to understand whether this phenomenon was a technical problem with imaging high cell density fruiting bodies with extensive polysaccharide extracellular matrices (16), phototoxicity due to repeated fluorescent imaging (17), or a biological phenomenon related to the reported lysis of cells before sporulation (18). Though this was not an initial intended goal of this study, the co-occurrence of coarsening with the loss of fluorescent cells makes it an intriguing area of additional exploration. I sought to determine the cause of the observed cell loss and to investigate whether there was any connection between cell loss in fruiting bodies and aggregate fate. We determined through Propidium Iodide (PI) staining that the most likely cause of cell loss was autolysis during the *M. xanthus* developmental program, and that unstable aggregates that eventually disperse have a lower PI signal density that is likely the result of decreased autolysis in those fruiting bodies.

METHODS

Strains and Culture Conditions

Myxococcus xanthus DK1622 was used as the wild-type. For fluorescent image series, LS3908 cells (8) used for tracking were diluted 1:800 into a background of DK10547 cells (19), used for aggregate segmentation and assessment of local cell densities. DK11316 (20) was the nonmotile control. Cells were grown overnight in CTTYE (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 $KH(H_2)PO_4$, 8 mM $MgSO_4$) with vigorous shaking at 32°C. Cultures were supplemented with 10 μg/mL oxytetracycline for selection and 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for induction of tdTomato expression.. Cells were centrifuged to remove the nutrient broth, washed in TPM buffer (10 mM Tris (pH 7.6), 1 mM $KH(H_2)PO_4$, 8 mM $MgSO_4$), and resuspended in TPM buffer to a final concentration of $5x10^9$ cells/mL. For development assays, a droplet of approximately $2.5x10^7$ cells was transferred onto TPM agar slide complexes, as previously described (21).

For live/dead staining assays, DK1622 cells were grown overnight and prepared for development assays as above. Five microliter droplets containing $2.5x10^7$ cells were spotted onto 1.5% TPM agar plates, with one plate having four droplets for each sample. Once dry, the plates were placed in a 32°C incubator until the corresponding timepoint. Using the LIVE/DEAD® BacLight™ Bacterial Viability Kit [Invitrogen], Components A and B were each mixed to a concentration of 0.5% into TPM buffer. Once each timepoint was ready for imaging, 20 μl of the dye solution was pipetted gently over the top of the developing spot of cells on the TPM agar plate. The dye was left to incubate in the dark at room temperature for 10 minutes before imaging. Each plate was carefully marked before imaging so that the same field of view would be imaged at 12 hours and again at 24 hours.

Imaging

Imaging was performed on a Nikon Eclipse E400 microscope with a pco.panda 4.2 sCMOS camera and NIS-Elements software. For cell tracking experiments, LS3908 samples were imaged with 400ms exposure with a Sola LED light source at 75% intensity, and DK10547 samples were imaged with 200ms exposure at 35% intensity. Control of the fluorescent filter wheel and autofocus mechanism was managed with a MAC6000 system filter wheel controller and focus control module (Ludl Electronic Products, Ltd.). Images in the phase contrast and tdTomato channels were captured every 60 seconds over 24 hours, and GFP every 15 minutes to track the position of aggregates and changes in local cell densities.

For live/dead staining assays, both channels were imaged with 100ms exposure and at 50% intensity. The data in Fig. 4B came from images captured at either 0hr or at a 2hr increment from 12 to 24 hr. For Fig. 4C, samples were imaged at 12 hours (pre-coarsening) with GFP and Texas Red to assess live and dead signal. A phase contrast image was captured as well for manually tracking aggregate position and fate. The plates were then placed back in the incubator for another 12 hours for development to continue and the same field of view was imaged at 24 hours with only phase contrast to determine aggregate fate. Images of cells stained with only one dye were captured with the opposite channel to assess the level of bleedthrough of each dye to the other channel. These values were averaged for three plates and subtracted from the corresponding channel during image processing

Image Analysis

Cell tracking was automated following Cotter et. al (8) and performed by Patrick Murphy at Rice University. Cells trajectories were divided into categories representing their different motile states, persistent and nonpersistent. A new trajectory began when a cell either changed

state from persistent to nonpersistent, or vice versa, or when it reversed. Thus, there can be hundreds of trajectories for each cell, depending on how many times it changes state. Metrics could then be extracted from the trajectories, such as run duration (the length of time before persistent cells reverse), alignment toward or away from the aggregate, local cell alignment, and cell velocity. Equations for how these metrics were calculated for each trajectory can be found in the original paper (8). The data presented in this chapter relied on calculating the bias toward the aggregate, which involved run duration and angle toward the nearest aggregate. The cosine of the angle to the nearest aggregate centroid was taken, given a range of values from -1 to 1, where 0 is perfectly antiparallel, 1 is perfectly aligned toward the aggregate, and -1 is perfectly aligned away from the aggregate. Bias is then equivalent to:

$$\frac{\left(\mathrm{T}_{towards} - \mathrm{T}_{away}\right)}{\mathrm{T}_{all}}$$

where $\mathrm{T}_{towards}$ is equivalent to time spent moving toward the aggregate (cosine of angle to the aggregate is positive), T_{away} is the time spent moving away from the aggregate (cosine of angle to the aggregate is negative), and T_{all} is the total run duration for that trajectory. Additionally, the position and number of aggregates at each timeframe were also automatically tracked using a code similar to that for cell tracking. Aggregate size categories for Fig. 1C were determined by size relative to the sample mean for each replicate, where any aggregate where the area was greater than the mean was considered large and any with an area less than the mean was considered small. However, bias data in Fig. 2 reflect three bins of aggregate sizes, where over 14,000 μm^2 was considered large, less than 5,918 μm^2 was considered small, and medium fell in between that range.

For live/dead staining data, the fluorescence intensity of the propidium iodine dead stain was measured at each timepoint for twelve fruiting bodies that were manually traced using FIJI.

Mean bleedthrough of the SYTO 9 stain into the PI channel was subtracted before quantifying

fluorescence. The PI signal density metric reflects the corrected total cell fluorescence (ctcf) for

each fruiting body by area. Specifically, we used:

$$\frac{Integrated\ Density\ _{FB} - (Area_{FB} \times Mean_{Background})}{Area_{FB}}$$

Since ctcf is a summed intensity, we divided by area to get the signal density per µm^2.

RESULTS

M. xanthus development is a dynamic multicellular process; early aggregates of cells

grow, shrink, merge, split, and disappear, necessitating visualization with time-lapse movies to

gain a better understanding of the dynamics. Consistent with prior experimental observations

(14, 15), in our time-lapse movies of DK1622 development, a significant fraction of early

aggregates disperse during the coarsening phase of aggregation (14, 22) (Fig. 1A &B).

Approximately 35% of initial aggregates remain stable throughout aggregation. Aggregate size is

a good predictor of aggregate fate, as 75% of large aggregates remain stable compared to about

15% of smaller aggregates, and unstable aggregates on average are significantly smaller (Fig.

1C). Given that there are differences in aggregate fate based on size, we sought to understand

what behaviors at the scale of the individual cell drive these distinct outcomes.
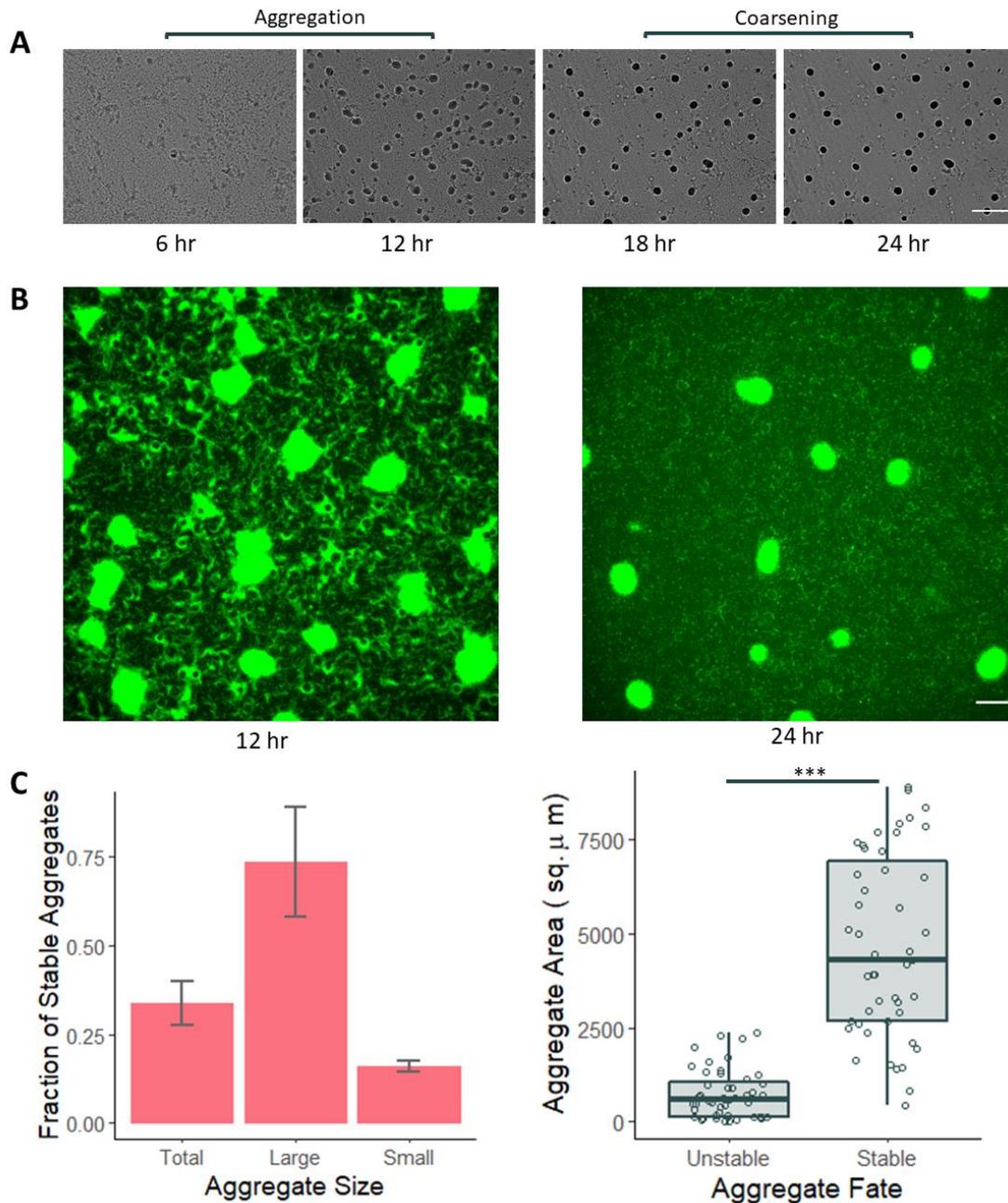
Figure 1. Aggregates disperse during the coarsening phase. A) Brightfield images showing the dispersal of aggregates. Initial aggregation takes place over the first 12 hours, followed by the coarsening phase (12-24 hours) where some aggregates disperse. Scale bar 500 µm. B) Fluorescent images showing the same phenomenon at higher magnification, with the pre-coarsening image taken just before coarsening (*left*) having more fruiting bodies. Scale bar 100 µm. C) *(left)* Fraction of each aggregate category that remains stable throughout the movie (n=3 movies, error bars show standard error). Large and small aggregate designation is with respect to mean aggregate size. *(right)* Aggregate fate by aggregate area, showing that stable aggregates are larger, (p=6.9e-15 in student's t-test for n=94 fruiting bodies).

113

To attempt to elucidate the individual cell behaviors that lead to the emergent patterns of aggregate fate, we tracked fluorescently labeled cells during development through the coarsening phase. We automated the extraction of three key metrics from cell trajectories: run duration (or the period of time over which a cell moves without reversing), run distance (the total distance a cell travels before reversing), and alignment to the nearest aggregate. Using the alignment metric, we were able to parse cell trajectories into two categories: run duration moving towards the nearest aggregate ($T_{towards}$), and run duration moving away from the nearest aggregate (T_{away}). We then calculated bias using the equation:

$$\frac{\left(T_{towards} - T_{away}\right)}{T_{all}}$$

A positive bias metric indicates that a cell was moving toward the nearest aggregate for longer than it was moving away during that run, and conversely, a negative bias metric indicates that a cell was moving away from the nearest aggregate for longer than it was moving toward it.

The data presented in Fig. 2 represent thousands of cell trajectories from a wild-type movie where initial aggregates can be seen at approximately 7.5 hours, and the first aggregate disappearances during coarsening occur at 11.5 hours. We found that cells showed an average and consistent bias toward aggregates that remained stable throughout the duration of the movie. Cells were initially positively biased toward unstable aggregates, but shift to a negative bias during coarsening (about 1 hour before the first aggregates begin to disappear), consistent with cells leaving unstable aggregates (Fig. 2A). When we binned aggregates into three explicit size categories, with aggregates greater than 14,000 μm^2 considered large, aggregates less than 5,918 μm^2 considered small, and aggregates between that range considered medium, we observed differences in bias based on the size of the nearest aggregate. Cells were consistently biased toward large aggregates (Fig. 2B), in agreement with our observation that the majority of large

aggregates remain stable throughout aggregation. Cells have an initial weak bias toward small aggregates, but an overall negative bias following initial aggregation, again consistent with our observations in Fig. 1. Midsize aggregates represent a category wherein some aggregates disperse and some remain stable. Before coarsening, cells are biased towards midsize aggregates, but during coarsening, unstable midsize aggregates shrink such that they fall into the small aggregate category, and the remaining midsize aggregates are stable, so the bias becomes positive again. Throughout coarsening, this can happen numerous times, giving rise to the oscillations in bias that is seen later in the medium aggregate curve (Figure 2B). This indicates that coarsening does not occur as one simultaneous event but rather that there are different waves of coarsening throughout the duration of the coarsening phase.

The cell trajectory data will be used to inform a data-driven model that simulates the coarsening phase of development. Similar data-driven models have been employed to simulate the initiation of aggregation in response to starvation, and have shown that cell bias toward aggregate centers, decreased run duration for cells inside aggregates, and alignment of cells toward aggregate centers all contribute to successful wild-type aggregation (8). However, no data-driven models have been used to examine behaviors on the cellular level might lead to coarsening of aggregates. Data-driven models extract probabilities of cell behavior based on the experimental cell's local environment from tracking data, creating a database of behaviors from which models can sample based on a simulated cell's local environment. By employing data-driven models, we hope to recapitulate the dynamics of coarsening, determining if the parameters that we are measuring from experimental data (run time, run duration, neighbor alignment, bias, and alignment toward the nearest aggregate) are sufficient to cause the same

fraction of aggregate dispersal and distribution of aggregate sizes that we see in experimental

data.
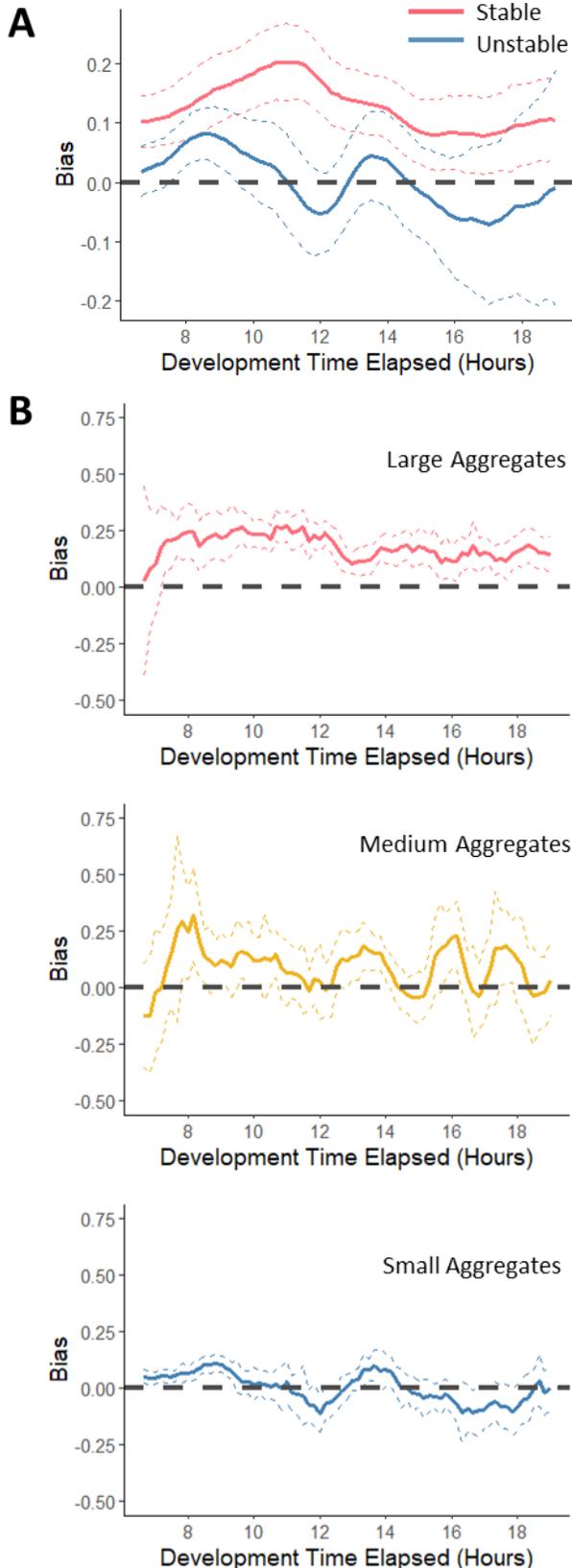
Figure 1. Aggregate properties affect cell bias. Bias was calculated as $(T_{toward} - T_{away}) / T_{all}$, where T_{toward} is the duration of time during a trajectory that a cell spent moving toward the nearest aggregate, T_{away} is the duration of time that a cell spent moving away from the nearest aggregate, and T_{all} is the total run duration for that trajectory. Dashed lines represent the bootstrapped 95% confidence interval of the means plotted in solid lines. Positive bias values indicate bias toward the nearest aggregate, and negative values indicate a bias away from the nearest aggregate. A) Cell bias for aggregates that are categorized as either stable (those that persist for the full length of the movie) and unstable (those that disperse during coarsening). There is a persistent positive bias for cells moving toward stable aggregates, and the bias for cells moving toward unstable aggregates becomes negative as coarsening proceeds and unstable aggregates disperse. Values were smoothed over a running window of 10 minutes. B) Bias of cells moving toward large (pink), medium (gold), and small (blue) aggregates. The bias for cells moving towards large aggregates remains positive. Coarsening has much more of an effect on the bias of cells toward medium and small aggregates. There is only a weak initial bias toward small aggregates that becomes negative early in coarsening. C) Midsize aggregates are made up of both stable and unstable aggregates, so the bias is a bit more dynamic. After an initial bias of cells toward these aggregates, coarsening causes oscillations in the strength of the bias toward these aggregates. See text for an explanation of these oscillations.

The collection of cell tracking data also allowed us to visualize other observable patterns that occur along the developmental timeline. In the majority of time-lapse fluorescent movies we generated, there was a significant decrease in the fraction of fluorescently labeled cells from the beginning of the movies to the end (Fig. 3). Notably, this loss of fluorescent cells was not observed when we diluted the same tdTomato-expressing cells into a background of DK11316, a nonmotile *M. xanthus* strain that does not make fruiting bodies, and imaged the cells under the same conditions. This indicated that the loss of trackable cells was not due to photobleaching or any negative effects of imaging. This also served as a control for the tdTomato-expressing strain that we use for cell tracking**,** showing that the loss of cells is context dependent and not related to the specific strain. It appears either that the developing three-dimensionality of growing fruiting bodies was affecting our ability to visualize the cells after a certain period, or that cells were dying within fruiting bodies.

We first tested the idea that cells were simply moving out of the focal plane as the fruiting body was growing taller**.** Imaging along a ±40 µm z-axis about the focal plane did not capture any additional cells that may have moved vertically in space, indicating that the depth of the focal plane was sufficient to capture labeled cells within the entirety of the fruiting body (Fig. S1). Another possibility is that dense nature of the fruiting bodies, or possibly even the extracellular matrix (ECM) secreted by cells therein, was obscuring the signal as cells formed new layers on top of one another. We tested this theory by observing cells leaving unstable aggregates during coarsening. Cells that disappeared during the early stages of coarsening in unstable aggregates did not reemerge as they lost three-dimensionality and dispersed. If the density of the fruiting bodies was obscuring the fluorescent signal, we should have recovered at least a fraction of the lost cells during the dispersal of unstable aggregates.
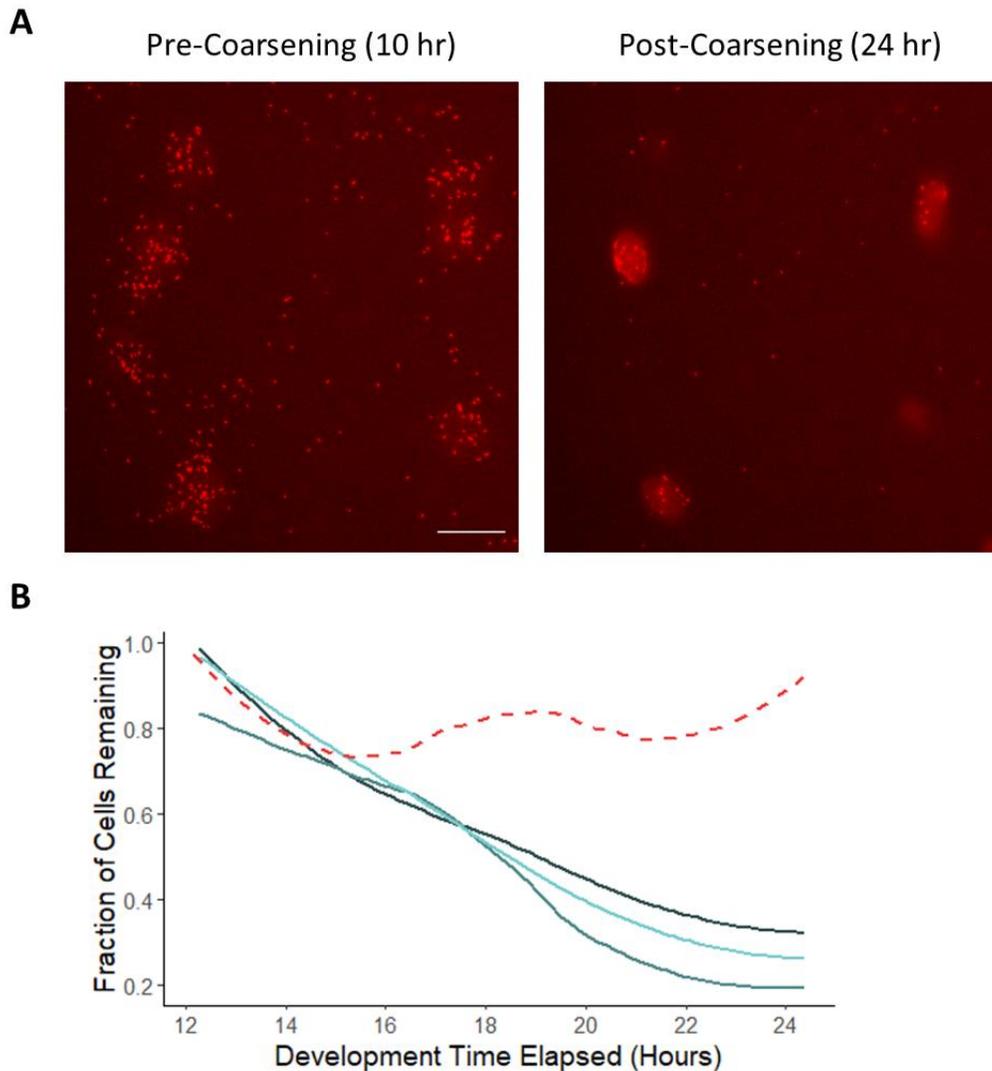
Figure 3: The population of trackable cells decreases throughout development. A) In the pre-coarsening image (*left*), cells labeled with tdTomato are numerous and clearly visible, but look more diffuse and much fewer in number post-coarsening (*right*). Scale bar 100 µm. B) Number of fluorescent cells, reported as a fraction of fluorescent cells at 12 hours, for three replicates of wild-type (blue), where only 20-40% of trackable cells remained. Pink dashed line represents tdTomato cells in a background of nonmotile DK11316 as a control that does not make fruiting bodies, and approximately 90% of cells remain in that sample. Loess smoothing was applied to these values to eliminate noise due to flickering of the camera and temporary lost cells due to slipping of the focal plane.

To test whether trackable cell loss was due to cell death, we stained with SYTO9, a

nucleic acid stain that can cross cell membranes to stain intact cells, and Propidium Iodide (PI), a

cell impermeable nucleic acid stain only binds DNA of cells with disrupted membranes, like

those that have undergone programmed cell death or lysis. Due to nonspecific staining of PI in intact cells over longer periods of time, we did not generate time-lapse movies of live-dead stained samples. We first collected timepoint images of live-dead staining during development in 6hr timepoints (Fig. 4A). The dead signal due to PI staining increased steadily increased as fruiting bodies persisted and was predominantly present inside fruiting bodies compared to the background. Within individual fruiting bodies, the density of PI staining increases from 12 to 16 hours, and then the rate of increase slows from 16 to 24 hours (Fig. 4B). This suggests that cell death might be the reason for the loss of tracked cells during fluorescence, and that this presents a biologically meaningful area of exploration rather than a technical issue with imaging three-dimensional structures.

Because the increase in signal from PI staining was highest during coarsening, we wanted to test the relationship between cell death and aggregate fate. To that end, we allowed cells to develop for 12 hours, stained with PI and imaged to get a picture of cell death mid-aggregation before unstable aggregates began disappearing, then allowed them to incubate for another 12 hours. We then imaged the same field of view to capture which fruiting bodies had dispersed during coarsening and which had remained so that we could look for an association between PI staining and aggregate fate. Stable aggregates showed nearly 3-fold grater PI staining density than unstable aggregates (Fig. 4C). Taken together, these data indicate that cell death may be a contributing factor to aggregate stability during coarsening.
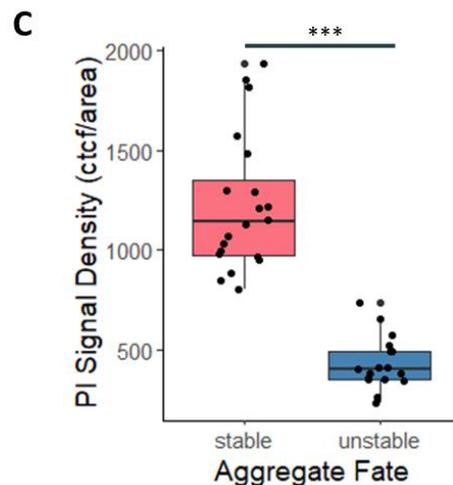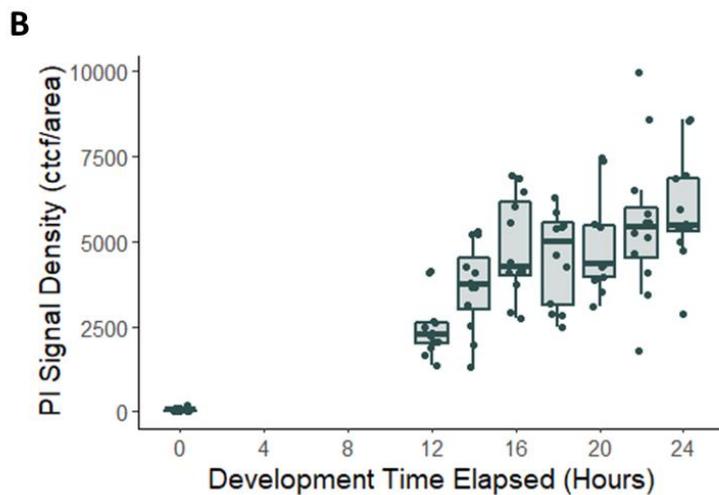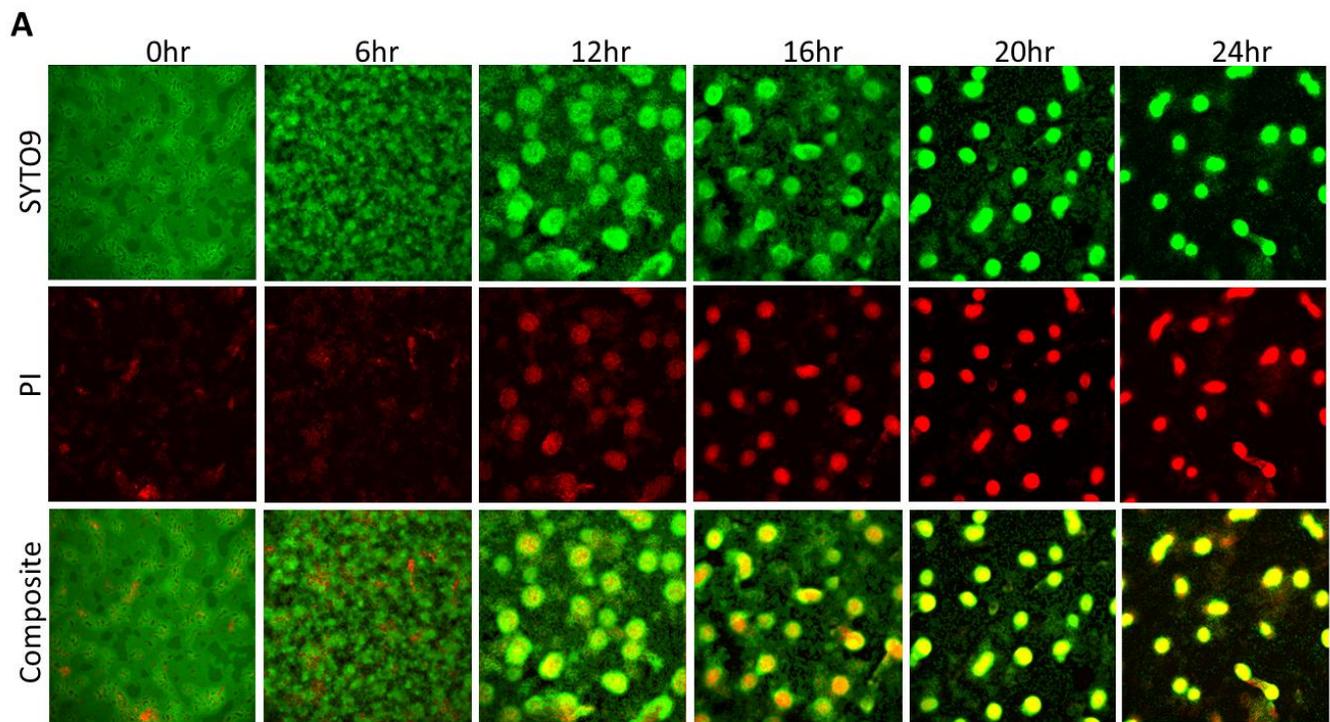
Figure 4: Extracellular DNA accumulates during coarsening. A) SYTO 9 and Propidium Iodide staining of fruiting bodies on agar at timepoints during development. Cell permeable SYTO 9 stains the nucleic acids of all cells, and PI is cell impermeable and stains only extracellular nucleic acids. B) Quantification of PI signal density in fruiting bodies at 4 hr timepoints beginning at 12 hours where fruiting bodies have clearly defined boundaries. 0hr timepoints reflect signal density over randomly chosen areas that fall into the size range for fruiting bodies, for comparison when cells are evenly distributed throughout the agar. (n=12 fruiting bodies per timepoint, signal density calculated as ctcf for each fruiting body, corrected by fruiting body area). C) PI signal density at 12 hours for aggregates that are either stable or unstable at 24 hours. (n=20 stable fruiting bodies and n=17 unstable fruiting bodies originating from 3 separate replicate plates; p=6.5e-10 in student's t-test).

DISCUSSION

Overall, the data presented in this chapter serve two major purposes. First, trajectory data from cell tracking experiments will be applied to data-driven models to simulate the coarsening phase of aggregation, with the goal of determining which individual cell behaviors are sufficient to drive the stability of some aggregates and the dispersal of others. Second, we have shown a link between PI signal density and aggregate fate that would explain both the trackable cell fraction lost during our fluorescent imaging time-lapse experiments and suggest a mechanism by which aggregates may stabilize before coarsening begins.

In the field, there are conflicting theories about the mechanistic basis of coarsening. It has been proposed that aggregates behave as droplets of liquid on a surface, in a similar manner to Ostwald ripening, with the mass of a shrinking droplet transferred to the mass of a nearby growing droplet, driving the system toward equilibrium (15, 23). Notably, this model implies that nearby aggregates exert influence over one another. Another model proposed by Holmes et. al contends that as aggregation proceeds, all early aggregates are competing for a limited population of cells, and that one aggregate in too close proximity to another may destabilize because the local cell density is not sufficient to support the growth of both (7). This model again relies on the influence of nearby aggregates. Contrasting this, Xie et. al have shown through experimental data and statistical image analysis that the size of an aggregate is the only reliable predictor of its dispersal, and that characteristics of neighboring aggregates, such as size and proximity, do not affect the probability of dispersal (14).

Thus, it is an open question whether aggregate dispersal is driven mainly by physical principles (which are of course still governed by the genetics of motility machinery in living systems), or if it is genetically controlled. For example, cells may be able to use short-range

diffusible signals to determine the density of an aggregate and change their behavior accordingly (14). Importantly, whether aggregate dispersal operates by a physical mechanism such as Ostwald ripening or through a signal-mediated change in cellular behavior that reflects an underlying biological mechanism, data about cell trajectory bias alone is not sufficient to provide evidence for one mechanism over the other. If live cells are leaving shrinking aggregates by any mechanism, the bias toward the aggregate will eventually be negative, and we can only concretely say that aggregates disperse because cells are moving out of it and not due to cell death or another mechanism. Therefore, it will be imperative that our simulations are based on the probabilities of cell behavior from experimental data, including metrics such as frequency of reversals, run duration, and velocity which are genetically controlled in addition to metrics like alignment of neighboring cells or bias toward or away from aggregates, which could be governed by physical principles. There are also genetic mutants from the collection in Chapter 2 that show altered coarsening phenotypes, where either a greater or lesser fraction of fruiting bodies disperse during development (Fig. S2). Observing individual cell behaviors that might differ between wild-type and the strains that have coarsening phenotypes may provide an additional source of information. Using this method, we hope to parse the physical and genetic effects underlying coarsening and provide a working model for why some aggregates remain stable and some disperse.

We explored an additional phenomenon during the coarsening phase and observed a reproducible loss of fluorescently labeled cells during tracking experiments that co-occurred with coarsening. We associate this cell loss with a corresponding increase in PI signal within fruiting bodies as development proceeds, indicating there is more extracellular DNA (eDNA) near later-stage fruiting bodies. It is important to note that, while this increase in eDNA may be the result

of cell lysis, we are not attempting to directly correlate the level of PI signal to a number of dead cells, as there are documented issues with using PI staining for cell density estimations. Particularly in environments like biofilms that are rich with eDNA, using a standard curve to approximate dead cell count often overestimates cell nonviability (24), as live cells can be coated in eDNA while their membranes and chromosomes are intact. This scenario seems unlikely in controlled laboratory settings where the sources of eDNA would not be nearly as rich as its soil environment might be, but should still be considered. An increase in PI signal density could mean that cells are lysing, and the resulting DNA is being released to the environment, or that *M. xanthus* cells are transporting extracellular DNA to their environment without undergoing lysis. We will consider these possibilities within the context of what is known about *M. xanthus* biology during development.

Wild-type cells under starvation differentiate into numerous different cell types based on expression of key developmental regulatory genes (25). Of the cell types that are present in fruiting bodies, some are programmed to differentiate into spores while an estimated 80-90% of cells undergo programmed cell death, purportedly via the MazF toxin-antitoxin system (26, 27). MazF is an endoribonuclease that cleaves single-stranded RNA sequences, halting protein synthesis and indirectly inducing lysis (28). The role of MazF as the primary driver of cell lysis may be strain specific, as its deletion in some wild-type backgrounds, including DK1622 used in this study, has only a minor effect on levels of lysis (25); however lysis appears to occur across all wild-type backgrounds to varying extents and is thought to be required for sporulation (18).

Autolysis is proposed to support cells in biofilms undergoing energetically costly differentiation into spores by providing essential nutrients in an otherwise nutrient-limited environment, an idea that has been proposed in several species (29). In *M. xanthus*, the

remaining 10-20% of cells within fruiting bodies will differentiate into spores, presumably making use of the contents released by lysed cells. The extent of cell loss that we observed within late-stage fruiting bodies roughly matches these estimates, with a range of about 20-35% (Fig. 3B). Further, autolysis has been shown to be developmentally regulated and not exclusively a consequence of prolonged starvation (25, 30). We similarly show that the same strain that loses 80-90% of tracked cells in a wild-type background shows less than 10% reduction in cell count when diluted into a nonmotile mutant that cannot form aggregates. Our data in the context of *M. xanthus* biology suggest that autolysis is responsible for the increased eDNA presence that leads to greater PI signal density.

We observe that aggregates with decreased PI signal density when stained pre-coarsening have a greater probability of dispersing during coarsening (Fig. 4C). eDNA is not as extensively studied in *M. xanthus* as it is in other biofilm species, but one study speculated that autolysis is the source of the eDNA observed during development, likely because lysis has long been a documented component of *M. xanthus* development (31). More recently, however, eDNAs have been observed in vegetative swarms of *M. xanthus* that are not developing and therefore not presumed to be undergoing autolysis (32), so it is not entirely clear whether autolysis is the only mechanism by which eDNAs enter the extracellular environment. Whatever the mechanism, two studies have shown that a role of eDNA in *M. xanthus* biofilms is to provide structure and support to the extracellular matrix, which ultimately improve stress resistance (32, 33). eDNA co-localizes with exopolysaccharide (EPS) within the ECM, decreasing sensitivity to mechanical forces. eDNA is proposed to have roles in stability and cell adhesion within the biofilm matrix in several other species (34–36).

It is unclear at this point whether coarsening and autolysis are genetically linked, or if coarsening aggregates are a byproduct of asymmetric autolysis in different fruiting bodies. Considering that small aggregates are both more likely to be unstable and are surrounded by less eDNA, it is tempting to think that the mechanism for lysis is induced by a quorum signal, and that small aggregates do not achieve sufficient density to trigger autolysis and therefore do not release as much eDNA into the ECM, reducing stability. In contrast, a quorum may be reached in larger aggregates such that autolysis is triggered, releasing eDNA which provides a more protective scaffolding for the aggregates, preventing the loss of cells during coarsening. It is also possible that these are two unrelated behaviors, and whether an aggregate disperses during coarsening is a matter of timing. MrpC, the antitoxin for the MazF protein, is a transcription factor that will not initiate transcription of MazF, and therefore not induce lysis, until it is activated by a signaling cascade induced by fruiting body formation (37). Perhaps, if coarsening begins before autolysis on the developmental timeline, smaller aggregates would be more affected because they have fewer cells and would disperse before autolysis could provide stability.

Additional experiments would help to clarify the relationship between coarsening and autolysis. Examining PI signal density in genetic mutants that display very little coarsening may indicate whether eDNA is playing a role in stabilizing cells during coarsening. The data-driven modeling of coarsening may also aid in determining which genetic pathways, if any, might be influencing coarsening, and in turn could inform studies of autolysis as well. Cell loss could also potentially be incorporated into simulations to determine whether it is necessary to reproduce the coarsening patterns that we see in experimental data. While time-lapse movies of PI staining are a poor indicator of cell death due to nonspecific staining over longer periods, examining PI

signal density over more frequent timepoints than we did in this study might help to disentangle where autolysis falls on the developmental timeline relative to the onset of coarsening.

REFERENCES

1.   Sliusarenko O, Neu J, Zusman DR, Oster G (2006) Accordion waves in Myxococcus xanthus. *Proc Natl Acad Sci U S A* 103(5):1534–1539.

2.   Zhang H, et al. (2012) The Mechanistic Basis of Myxococcus xanthus Rippling Behavior and Its Physiological Role during Predation. *PLoS Comput Biol* 8(9). doi:10.1371/journal.pcbi.1002715.

3.   Igoshin OA, Mogilner A, Welch RD, Kaiser D, Oster G (2001) Pattern formation and traveling waves in myxobacteria: Theory and modeling. *Proc Natl Acad Sci U S A* 98(26):14913–14918.

4.   Igoshin OA, Kaiser D, Oster G (2004) Breaking symmetry in myxobacteria. *Curr Biol* 14(12):459–462.

5.   Igoshin OA, Welch R, Kaiser D, Oster G (2004) Waves and aggregation patterns in myxobacteria. *Proc Natl Acad Sci U S A* 101(12):4256–4261.

6.   Zhang H, et al. (2011) Quantifying aggregation dynamics during Myxococcus xanthus development. *J Bacteriol* 193(19):5164–5170.

7.   Holmes AB, Kalvala S, Whitworth DE (2010) Spatial simulations of myxobacterial development. *PLoS Comput Biol* 6(2). doi:10.1371/journal.pcbi.1000686.

8.   Cotter CR, Schüttler H, Igoshin OA, Shimkets LJ (2017) Data-driven modeling reveals cell behaviors controlling self-organization during Myxococcus xanthus development. *Proc Natl Acad Sci*. doi:10.1073/pnas.1620981114.

9.   Zhaoyang Z, et al. (2020) Data-Driven Models Reveal Mutant Cell Behaviors Important for Myxobacterial Aggregation. *mSystems* 5(4):e00518-20.

10.   Liu G, et al. (2019) Self-Driven Phase Transitions Drive Myxococcus xanthus Fruiting Body Formation. *Phys Rev Lett* 122(24):248102.

11.   Hendrata M, Yang Z, Lux R, Shi W (2011) Experimentally Guided Computational Model Discovers Important Elements for Social Behavior in Myxobacteria. *PLoS One* 6(7):1–11.

12.   Sliusarenko O, Zusman DR, Oster G (2007) Aggregation during fruiting body formation in Myxococcus xanthus is driven by reducing cell movement. *J Bacteriol* 189(2):611–619.

13. Igoshin OA, Goldbeter A, Kaiser D, Oster G (2004) A biochemical oscillator explains several aspects of Myxococcus xanthus behavior during development. *Proc Natl Acad Sci U S A* 101(44):15760–15765.

14. Xie C, Zhang H, Shimkets LJ, Igoshin OA (2011) Statistical image analysis reveals features affecting fates of Myxococcus xanthus developmental aggregates. *Proc Natl Acad Sci U S A* 108(14):5915–20.

15. Bahar F, Pratt-Szeliga PC, Angus S, Guo J, Welch RD (2014) Describing Myxococcus xanthus Aggregation Using Ostwald Ripening Equations for Thin Liquid Films. *Sci Rep* 4:6376.

16. Behmlander RM, Dworkin M (1991) Extracellular fibrils and contact-mediated cell interactions in Myxococcus xanthus. *J Bacteriol* 173(24):7810–7820.

17. Icha J, Weber M, Waters JC, Norden C (2017) Phototoxicity in live fluorescence microscopy, and how to avoid it. *Bioessays* 39(8). doi:10.1002/bies.201700003.

18. Wireman JW, Dworkin M (1977) Developmentally Induced Autolysis During Fruiting Body Formation by Myxococcus xanthus. *J Bacteriol* 129(2):796–802.

19. Welch R, Kaiser D (2001) Cell behavior in traveling wave patterns of myxobacteria. *Proc Natl Acad Sci U S A* 98(26):14907–14912.

20. Fontes M, Kaiser D (1999) Myxococcus cells respond to elastic forces in their substrate. *Proc Natl Acad Sci U S A* 96(14):8052–8057.

21. Taylor RG, Welch RD (2010) Recording Multicellular Behavior in Myxococcus xanthus Biofilms using Time-lapse Microcinematography. *J Vis Exp* (42):1–6.

22. Zhang H, et al. (2011) Quantifying aggregation dynamics during Myxococcus xanthus development. *J Bacteriol* 193(19):5164–5170.

23. Glasner K, Otto F, Rump T, Slepčev D (2009) Ostwald ripening of droplets: The role of migration. *Eur J Appl Math* 20(1):1–67.

24. Rosenberg M, Azevedo NF, Ivask A (2019) Propidium iodide staining underestimates viability of adherent bacterial cells. *Sci Rep* 9(1):1–12.

25. Lee B, Holkenbrink C, Treuner-Lange A, Higgs PI (2012) Myxococcus xanthus developmental cell fate production: Heterogeneous accumulation of developmental Regulatory proteins and reexamination of the role of MazF in developmental lysis. *J Bacteriol* 194(12):3058–3068.

26. Berleman JE, Chumley T, Cheung P, Kirby JR (2006) Rippling is a predatory behavior in Myxococcus xanthus. *J Bacteriol* 188(16):5888–5895.

27. Nariya H, Inouye M (2008) MazF, an mRNA Interferase, Mediates Programmed Cell Death during Multicellular Myxococcus Development. *Cell* 132(1):55–66.

28. Zhang Y, et al. (2003) MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. *Mol Cell* 12(4):913–923.

29. E. G-PJ, C. HE, Richard L (2003) Cannibalism by Sporulating Bacteria. *Science (80-)* 301(5632):510–513.

30. Janssen GR, Dworkin M (1985) Cell-cell interactions in developmental lysis of Myxococcus xanthus. *Dev Biol* 112(1):194–202.

31. Rosenbluh A, Rosenberg E (1993) Developmental lysis and autocides. *Myxobacteria II*, eds Dworkin M, Kaiser D, pp 213–234.

32. Hu W, et al. (2012) DNA Builds and Strengthens the Extracellular Matrix in Myxococcus xanthus Biofilms by Interacting with Exopolysaccharides. *PLoS One* 7(12):1–12.

33. Wang Y, et al. (2022) Physicochemical and Biological Insights Into the Molecular Interactions Between Extracellular DNA and Exopolysaccharides in Myxococcus xanthus Biofilms. *Front Microbiol* 13(April). doi:10.3389/fmicb.2022.861865.

34. Grande R, et al. (2015) Helicobacter pylori ATCC 43629/NCTC 11639 Outer Membrane Vesicles (OMVs) from Biofilm and Planktonic Phase Associated with Extracellular DNA (eDNA). *Front Microbiol* 6:1369.

35. Sena-Vélez M, Redondo C, Graham JH, Cubero J (2016) Presence of Extracellular DNA during Biofilm Formation by Xanthomonas citri subsp. citri Strains with Different Host Range. *PLoS One* 11(6):e0156695.

36. Izano EA, Shah SM, Kaplan JB (2009) Intercellular adhesion and biocide resistance in nontypeable Haemophilus influenzae biofilms. *Microb Pathog* 46(4):207–213.

37. Nariya H, Inouye S (2006) A protein Ser/Thr kinase cascade negatively regulates the DNA-binding activity of MrpC, a smaller form of which may be necessary for the Myxococcus xanthus development. *Mol Microbiol* 60(5):1205–1217.
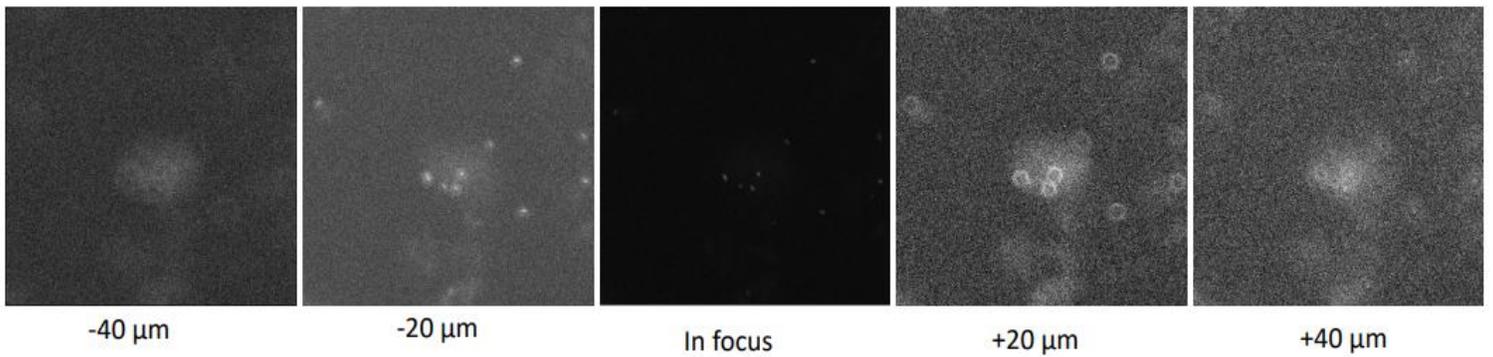
Figure S1: Capturing cells in a fruiting body through a Z-stack. Imaging ±40 µm from the primary focal plane of the movie did not reveal additional cells that had moved vertically in space. Additionally, all cells appear to lose focus equally throughout the Z-stack, indicating that the depth of the focal plane is sufficient to capture visible cells within the entirety of the fruiting body.

CHAPTER 5: DISCUSSION, CONCLUSIONS, AND FUTURE DIRECTIONS

INTRODUCTION

In the post-genomic era, we often make sense of genomic, transcriptomic, and proteomic data by observing an organism's phenome. We initially characterize organisms by observing what they can do and how they interact with their environments. We make inferences about individual gene function by measuring the impact of genetic mutation on phenotype. We identify candidate genes to study through gene homology, where mutations in orthologs provoked phenotypes of interests in other model systems. Random mutagenesis screens for phenotypes of interest have long served as another impactful method to identify candidate genes. In short, phenotype is the readout for the interaction between genes and the environment, and studying an organism's phenome under a variety of genetic and environmental backgrounds is essential for building a robust genotype-to-phenotype map.

The work presented in this thesis highlights the importance of observing and quantifying phenotype across a wide range of conditions. Associations between single gene mutations and phenotype are important and have assigned function to many genes; however, this paradigm can limit our understanding of genotype-phenotype relationships. In Chapter 2, I discuss 265 single gene mutations, the vast majority of which have minor phenotypes, and show that redundancy among genes in large homologous families is very widespread, limiting our understanding of gene function by introducing mutations. In Chapter 3, I demonstrate that pulsing, a phenotype that was initially believed to be rare and therefore only associated with a small group of genes, is actually very widespread among genetic mutants and occurs in wild-type as well. Without a sufficient collection of mutant and wild-type time-lapse data, this phenotype may have been falsely attributed to a single genetic pathway. In Chapter 4, I explore coarsening, a late-stage aggregation phase where some fruiting bodies persist and some disperse, which was discovered

through decades of observations and recordings of *M. xanthus* behavior in movies. Tracking the behaviors associated with this aggregation phase led to another observation, cell loss during coarsening, that prompted further study. In all, the work in this thesis emphasizes that studying the phenome holds both answers to current questions as well as the potential to introduce new and exciting questions.

ESTABLISHING THE EXTENT OF FUNCTIONAL REDUNDANCY WITHIN A GENOME

Assigning function to genes through knockout studies is complicated by the fact that the majority of mutations fail to yield an observable phenotype that differs from wild-type. Though there are multiple mechanisms by which this could happen, functional redundancy is expected to be one of the most common. Particularly in organisms with large families of paralogs that arose through duplication, there is increased potential for functional redundancy. In Chapter 2, I presented a large-scale mutant analysis of paralogs belonging to four different gene families in *M. xanthus* and demonstrate that mutants within the same gene family are more phenotypically similar, providing evidence to support the notion that redundancy networks are large and contribute to mutational robustness.

This idea can help to shape our view of the relationship between genotype and phenotype. Gene networks, particularly those involving components of homologous gene families, are highly complex, and though each gene is predicted to have one specific function through sequence homology, there is likely functional overlap of varying extents with other components in the same family. When we make mutations in one gene, we are not just affecting that gene pathway, but all the pathways through which the flow of information from the mutated gets rerouted. Mutations are created within the landscape of an entire cell, and we should

consider the potential for redundancy and the impact on other networks when we make inferences about gene function based on phenotype.

We developed an image analysis pipeline that automated the extraction of phenotypic metrics that quantify small-scale phenotypic detail. Rather than observing the direct effects of mutation on fitness by measuring growth and sporulation, we sought to provide meaning to the minor phenotypic deviations that occur more commonly. We propose that quantifying these minor phenotypes may not impact fitness was a more suitable method for detecting the higher order redundancies that we expected within the homologous families in our study. This presents the largest collection of bacterial mutant phenotype data of which we are aware that measures multiple aspects of a complex phenotype rather than using growth as a phenotypic readout. Our image analysis pipeline can be further extended to other studies of *M. xanthus* phenotype in other genetic backgrounds and environments to continue building the phenome.

It would be informative to supplement these phenotypic studies with transcriptomic studies to help determine the mechanism of the redundancy that we observe. For example it has been documented that some genes negatively regulate the expression of a redundant counterpart, and that deleting one paralog leads to upregulation of the other (1). Other redundant proteins may differ in their temporal regulation (2). Overall, transcriptomics could support the idea that there are large networks of functional redundancy by determining whether or not different mutants that express the same phenotype have similar transcriptional profiles.

We demonstrated that, while genotypic similarity on the order of gene family seems to drive phenotypic similarity, homology on a finer scale is not sufficient to predict redundancy. This suggests that the position or biochemistry of amino acid sequence changes is more important than the overall extent of sequence similarity. Our dataset could be used to compare

the amino acid sequences of genes that fall within the phenotypic cluster for their gene family and those that fall outside of the phenotype cluster to look for patterns associated with functional redundancy. If such patterns exist, for example conservation of certain residues within domains in a particular gene family, this type of data may be able to aid in prediction of functional redundancy bioinformatically. Additional genotype and phenotype data for other homologous groups would improve this kind of study.

CHARACTERIZING TRAVELING WAVES THAT SYNCHRONIZE MOTILITY DURING DEVELOPMENT

Understanding the signals that govern collective cell migration will inform the principles of multicellularity. Though they are inherently single-celled organisms, bacterial species that form biofilms often serve as important model systems for studying the cell-cell communication that leads to self-organization. The data presented in Chapter 3 present a characterization of a wave-like phenomenon that coordinates a shift to a persistent motility state within a developing *M. xanthus* swarm. We propose that a signal travels through the swarm via a reaction-diffusion mechanism, whereby cells within a developing fruiting body release a signal which is perceived by nearby cells, triggering those cells to release a signal, and the wave propagates through the swarm. The signal appears to cause cells to repress reversal, and slime trail following in combination with persistent motion could increase the probability that a cell makes it into an aggregate to either contribute to autolysis or differentiate into a spore.

The widespread distribution of this phenotype within mutant strains and wild-type suggests that it is a fundamental aspect of development that may aid in the collective migration of cells toward aggregates. If this is true, it has important implications for the fitness of the swarm and thus should be further investigated. The most interesting and obvious area for further exploration is determining the signal to which the cells are responding. Early studies of *M.*

xanthus tested whether cells could direct their motility up concentration gradients of common chemoattractants, such as cAMP that drives aggregation in *Dictyostelium*, and found that *M. xanthus* exhibited no chemotactic response toward any of the tested chemicals (3). Notably, this study specifically searched for directed motility up a concentration gradient in vegetative swarms. It is likely that their experimental methods would not detect a general increase in persistent movement due to reversal suppression. Further, I conducted a preliminary study where I deposited a newly starving population next to an already developing population that was exhibiting pulsing, and pulses did not extend into the swarm that was just exposed to starvation. Though this experiment was by no means comprehensive, it may suggest that the response to the signal might be developmentally regulated such that vegetative cells are not capable of responding to it.

In light of this, starving *M. xanthus* populations should be retested for their ability to respond to chemical signals. Fluorescence microscopy could be used to identify changes in persistence and reversal frequency when exposed to potential signals at differing concentrations. Identifying the signal would be particularly informative given that it could then help to identify a genetic mechanism by which the signal is sensed and propagated. Knocking out genes involved in pulsing would allow for visualization of development in the absence of this signaling, providing insight into what pulses might be doing to aid in development. Additionally, the tools that we developed in Chapter 2 of this thesis can be used to quantify differences in developmental scenarios that lead to pulsing and those that do not, potentially helping us understand what conditions are triggering pulsing in some situations but not others.

Kearns and Shimkets have shown that *M. xanthus* did exhibit directed motility toward phosphatidylethanolamine gradients, which suggests that under some circumstances it can

exhibit chemotactic responses (4). They further show that the suppression of reversals during this response is independent of the *frz* chemosensory system which is thought to govern the majority of reversals in *M. xanthus*. However, there are seven other chemosensory pathways and multiple other orphan chemoreceptors that may be involved in directing motility and other cell behaviors in response to specific signals (5). These present good targets for disrupting the mechanism that is responding to the pulse and would be interesting candidates for further study.

WORKING TOWARD DEFINING CELL BEHAVIORS NECESSARY FOR COARSENING

Computational modeling of *M. xanthus* multicellular behaviors is beneficial because simulations can test which behaviors are necessary and sufficient to drive a particular behavior in the absence of knowledge of the underlying causal mechanisms. These simulations indicate the physical mechanism by which cells self-organize, for example by slowing down cell velocity inside aggregates (6), which can then help to formulate testable hypotheses about what genetic systems may be involved. The data in Chapter 4 present an initial depiction of the types of data that we can extract from cell tracking experiments and apply to simulated cells to recapitulate coarsening during late aggregation. Data-driven models have already improved our understanding of the requirements for initial aggregation (7, 8), and the hope is that by applying the same methods to later-stage aggregation, we can achieve a better understanding of the principles that govern coarsening. Currently, models are being fine-tuned such that the behavior of simulated cells is based on the run duration, run distance, local alignment, and angle toward the nearest aggregate, with respect to local density, extracted from experimental data that I have contributed to the collaboration.

Collecting multiple replicates of fluorescent cell tracking movies required for the data-driven models revealed that the majority of trackable cells were lost during the coarsening phase.

I was interested in exploring this phenomenon, and found that the loss of cells was associated with an increase of eDNA in the immediate area of the fruiting bodies, and given what is known about cell fates in *M. xanthus* development, it is likely that the loss of cells and increase in eDNA density are both the result of autolysis, a programmed cell death mechanism where the majority of cells lyse, releasing nutrients that the remaining cells can use to differentiate into spores. Given that eDNA has been found to stabilize the ECM of biofilms across different species, including *M. xanthus*, I attempted to look for a relationship between eDNA concentration and aggregate stability and found that when stained in the pre-coarsening phase, unstable aggregates had lesser eDNA signal density when compared to stable aggregates that persisted and matured into fruiting bodies.

Future work focusing on the relationship between aggregate fate and autolysis during coarsening should attempt to disentangle where each falls on the developmental timeline. If coarsening consistently begins before autolysis, then the fact that smaller aggregates disperse more during coarsening may just be a matter of timing; too great a fraction of the cells in small aggregates may have left the aggregate before autolysis could stabilize it. However, if autolysis occurs consistently before coarsening begins, it may be a density-dependent mechanism for larger aggregates to remain stable and smaller aggregates, which may not meet the density requirements to sporulate efficiently, to disperse.

The role of MazF in aggregate stabilization should also be investigated. Though it is controversial whether this ribonuclease toxin is the primary driver of autolysis in all wild-type strains (9), it would be interesting to observe whether a *mazF* deletion has a different coarsening phenotype than is seen in wild-type, and if there are differences in eDNA content in these mutants. The *mazF* mutant created in DK1622, the wild-type background used in this study,

showed no significant difference in the amount of cell lysis compared to the parent strain in prior work (9), which indicates that other mechanisms of autolysis should be investigated. Though *mazF* does not appear to have any closely-related homologs in *M. xanthus*, one could search for homologs of the numerous other toxin-antitoxin systems that exist in bacteria (10). Developmental autolysis is still occurring in these strains, just perhaps via other mechanisms, and is potentially playing a role in the stability of aggregates.

EXPLORING THE INTERPLAY BETWEEN PULSING, COARSENING, AND AUTOLYSIS

Pulsing in many genetic mutants causes an instability of aggregates, allowing cells to temporarily extend outward from the aggregates in branching groups. However, in wild-type, cells within aggregates seem to be relatively unaffected by pulses and aggregates rarely destabilize during a pulse. Every instance of pulsing observed in our study has occurred after initial aggregation and seems to co-occur with the timing of coarsening. Further, fluorescent samples that show synchronized pulses seem to have less extensive cell loss than similar samples that do not pulse. Given this and the data presented in Chapter 4, it is tempting to speculate about a relationship between pulsing, coarsening, and autolysis.

Perhaps aggregates are stabilized during pulses in wild-type samples by autolysis such that the eDNA bound to the ECM mechanically prevents cells from responding to the pulse. If developmental autolysis is dysregulated in pulsing samples, cells within fruiting bodies might lack the stability of wild-type aggregates, and when the pulse passes over the aggregate, cells within the fruiting bodies would also become more persistent, causing the dramatic pulsing phenotype that we see in the mutants. The proposed autolysis pathway in *M. xanthus* involves the interaction and convergence of many signaling pathways, which could potentially explain the prevalence of pulsing (11, 12). PI staining of pulsing samples may be able to answer some of

these initial questions. If pulsing aggregates have an overall lesser PI signal density, that may indicate that autolysis is happening to a lesser extent. It would also be interesting to test whether pulses that destabilized aggregates occurred more frequently in *mazF* mutants (or other toxin-antitoxin system mutants) which might inherently have decreased stability.

CONCLUSIONS

In this thesis, I present three major contributions to the field. First, the quantification of minor deviations in phenotype among homologous gene families depicts a cellular environment with extensive networks of redundant genes that provide robustness to mutation. This study has implications for future studies that ascribe function to genes based on the effects of mutation and suggests that the potential for redundancy be considered. Second, we contribute to the knowledge of self-organization in *M. xanthus* fruiting bodies by identifying and characterizing a wave that synchronizes cell behavior. Future studies into the mechanisms that drive pulsing can contribute to our understanding of cell-cell interactions and multicellularity in bacteria. Finally, I contribute experimental cell tracking data to build computational models of the coarsening phase of aggregation, which will help to determine whether genetic mechanisms are involved in coarsening or if it is driven exclusively by physical principles. I also identify a connection between coarsening, autolysis, and also potentially pulsing, that provide an area of future exploration.

Overall, these data highlight the importance of collecting thorough phenotype data that capture the dynamics of living systems. Phenomics can help pave the way to a better understanding of genotype-to-phenotype relationships. Properly collecting, quantifying, and analyzing phenotype data on this scale requires interdisciplinary approaches. This work emphasizes the importance of collaboration amongst biologists, physicists, computer scientists,

and engineers to come up with the best approaches for measuring and interpreting phenotypes within the context of the entire genome.

REFERENCES

1. Kafri R, Springer M, Pilpel Y (2009) Genetic Redundancy: New Tricks for Old Genes. *Cell* 136(3):389–392.

2. Dean EJ, Davis JC, Davis RW, Petrov DA (2008) Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet* 4(7). doi:10.1371/journal.pgen.1000113.

3. Dworkin M, Eide D (1983) Myxococcus xanthus does not respond chemotactically to moderate concentration gradients. *J Bacteriol* 154(1):437–442.

4. Kearns DB, Shimkets LJ (1998) Chemotaxis in a gliding bacterium. *Proc Natl Acad Sci* 95(20):11957–11962.

5. Zusman DR, Scott AE, Yang Z, Kirby JR (2007) Chemosensory pathways, motility and development in Myxococcus xanthus. *Nat Rev Microbiol* 5(11):862–872.

6. Sliusarenko O, Zusman DR, Oster G (2007) Aggregation during fruiting body formation in Myxococcus xanthus is driven by reducing cell movement. *J Bacteriol* 189(2):611–619.

7. Cotter CR, Schüttler H, Igoshin OA, Shimkets LJ (2017) Data-driven modeling reveals cell behaviors controlling self-organization during Myxococcus xanthus development. *Proc Natl Acad Sci*. doi:10.1073/pnas.1620981114.

8. Zhaoyang Z, et al. (2020) Data-Driven Models Reveal Mutant Cell Behaviors Important for Myxobacterial Aggregation. *mSystems* 5(4):e00518-20.

9. Lee B, Holkenbrink C, Treuner-Lange A, Higgs PI (2012) Myxococcus xanthus developmental cell fate production: Heterogeneous accumulation of developmental Regulatory proteins and reexamination of the role of MazF in developmental lysis. *J Bacteriol* 194(12):3058–3068.

10. Jurėnas D, Fraikin N, Goormaghtigh F, Van Melderen L (2022) Biology and evolution of bacterial toxin–antitoxin systems. *Nat Rev Microbiol* 20(6):335–350.

11.   Nariya H, Inouye M (2008) MazF, an mRNA Interferase, Mediates Programmed Cell Death during Multicellular Myxococcus Development. *Cell* 132(1):55–66.

12.   Janssen GR, Dworkin M (1985) Cell-cell interactions in developmental lysis of Myxococcus xanthus. *Dev Biol* 112(1):194–202.

APPENDIX 1: QUANTIFICATION OF *MYXOCOCCUS XANTHUS* AGGREGATION AND RIPPLING BEHAVIORS: DEEP-LEARNING TRANSFORMATION OF PHASE-CONTRAST INTO FLUORESCENCE MICROSCOPY IMAGES

Author Contributions: Conceptualization, J.Z., **J.A.C**., A.B.P., R.D.W., and O.A.I.; methodology, J.Z., **J.A.C**., A.B.P., R.D.W., and O.A.I.; software, J.Z., C.R.C., W.N., and A.B.P.; validation, J.Z., A.B.P., and O.A.I.; formal analysis, C.R.C. and J.Z.; investigation, A.B.P., R.D.W., and O.A.I.; data curation, **J.A.C.** and R.D.W.; writing—original draft preparation, J.Z. and **J.A.C**.; writing—review and editing, J.Z., **J.A.C**., C.R.C., P.A.M., W.N., A.B.P., R.D.W., and O.A.I.; visualization, J.Z.; supervision, A.B.P., R.D.W., and O.A.I.; All authors have read and agreed to the published version of the manuscript.

ABSTRACT

Myxococcus xanthus bacteria are a model system for understanding pattern formation and collective cell behaviors. When starving, cells aggregate into fruiting bodies to form metabolically inert spores. During predation, cells self-organize into traveling cell-density waves termed ripples. Both phase-contrast and fluorescence microscopy are used to observe these patterns but each has its limitations. Phase-contrast images have higher contrast, but the resulting image intensities lose their correlation with cell density. The intensities of fluorescence microscopy images, on the other hand, are well-correlated with cell density, enabling better segmentation of aggregates and better visualization of streaming patterns in between aggregates; however, fluorescence microscopy requires the engineering of cells to express fluorescent proteins and can be phototoxic to cells. To combine the advantages of both imaging methodologies, we develop a generative adversarial network that converts phase-contrast into synthesized fluorescent images. By including an additional histogram equalized output to the state-of-the-art pix2pixHD algorithm, our model generates accurate images of aggregates and streams, enabling the estimation of aggregate positions and sizes, but with small shifts of their boundaries. Further training on ripple patterns enables accurate estimation of the rippling wavelength. Our methods are thus applicable for many other phenotypic behaviors and pattern formation studies.

RESULTS AND DISCUSSION

I created 14 microscopic time-lapse movies of *M. xanthus* development using tdTomato-expressing cells, taking concurrent images in the phase contrast and fluorescent channels every minute for 24 hours. Using these data, my collaborators demonstrated that the pixel value of phase contrast images does not correlate with cell density the way it does in the fluorescent channel (Fig. 1). This discrepancy is caused by the artefacts introduced by phase contrast imaging, which is optimized for enhancing the contrast of transparent cells on a transparent agar substrate, but introduces artefacts such as shade-off and halos around large objects (1, 2). This makes it difficult to estimate cell density, and thus to segment and quantify the properties of aggregates, based on these images. Pixel value in fluorescent images does correlate with cell density, but engineering fluorescent strains is time consuming, and extensive exposure to the high-intensity light needed to excite fluorophores can be phototoxic (3). Fluorescent images also capture the important interaggregate streaming behaviors more accurately (Fig. 2); these behaviors are thought to be significant for aggregate formation (4, 5).

In light of this, my collaborators developed a machine learning algorithm to transform phase contrast images into synthetic fluorescent images, such that the quantitative properties of real fluorescent images can be extracted from phase contrast images without the drawbacks of fluorescent microscopy. After appropriate training, the pix2pixHD-HE model that our collaborators built can construct synthetic fluorescent images from phase contrast images, where aggregates are very similar in size and morphology when compared to real fluorescent images. Without additional training, our model was capable of generating accurate synthetic fluorescent images at higher cell densities, where there are differences in aggregate size and morphology. Additionally, by using this pretrained model and limited additional training through transfer

learning, the model was able to generate accurate synthetic fluorescent images of rippling from

phase contrast time-lapse movies that I created, generating a distinctly different biological

pattern while maintaining its ability to reproduce fruiting bodies. This shows the generalizability

of the model under different experimental conditions with only minimal training.
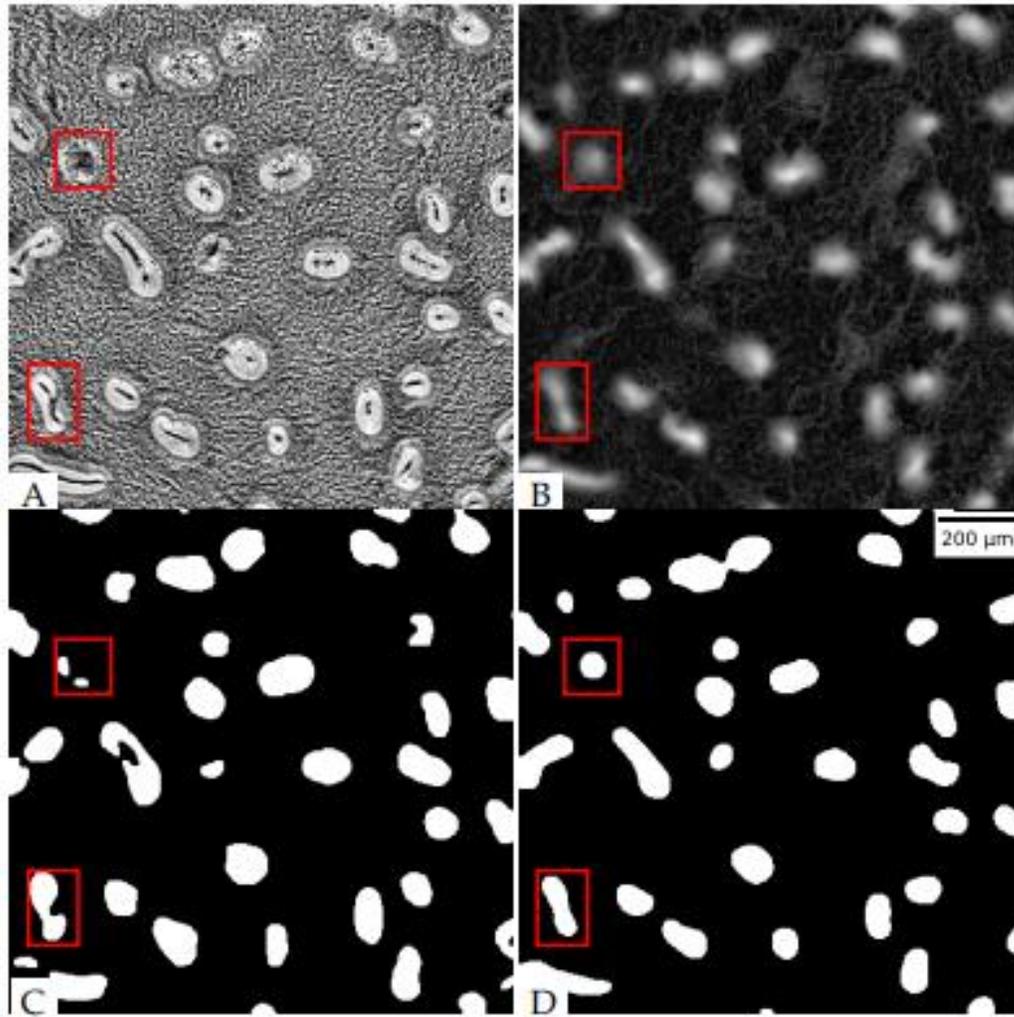
Figure 1: Aggregate segmentation from phase contrast and fluorescent images. A) Normalized phase contrast image, B) Normalized tdTomato fluorescent image, C) Aggregates segregated from phase contrast image, D) Aggregates segregated from fluorescent image. Note that aggregate segmentation from the phase contrast channel leaves large holes and/or irregularly shaped fruiting bodies compared to the fluorescent channel.
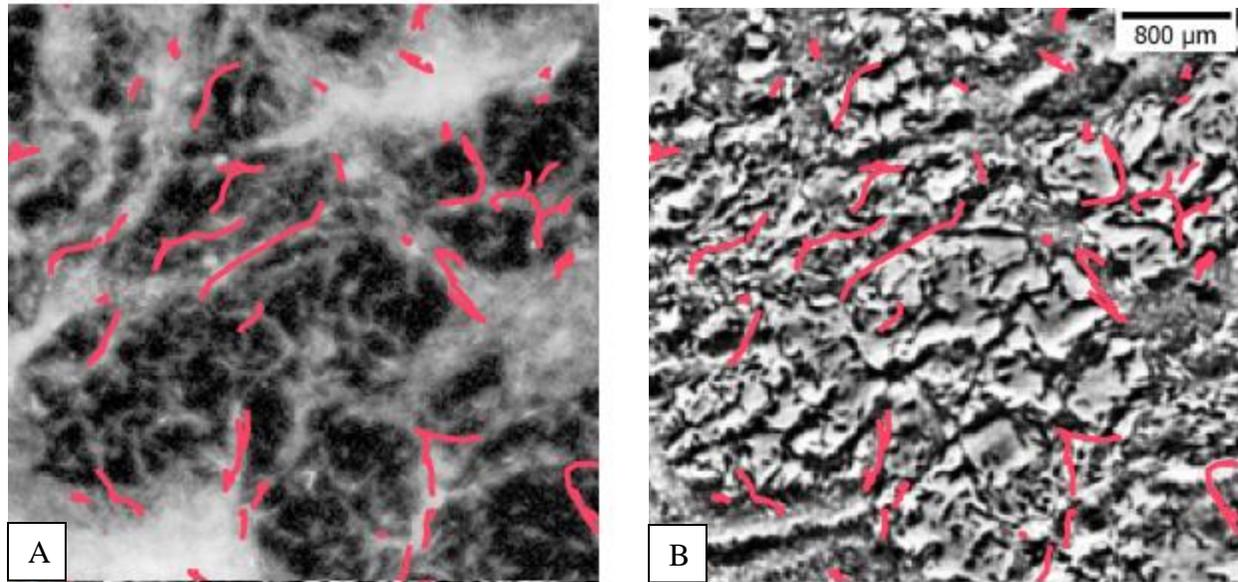
Figure 2: Fluorescent images better represent the location of interaggregate streams. Cell trajectories over 15 minutes of tdTomato-labeled tracked cells in a background of GFP-labeled cells. Cell trajectories are superimposed over the A) GFP fluorescent image, and B) phase contrast image.

MATERIALS & METHODS

Strains and Culture Conditions

M. xanthus tdTomato-expressing strain LS3908 (6) and GFP-expressing strain DK10547 (7), and E. coli K12 were used in this study. M. xanthus strains were grown overnight at 32°C with vigorous shaking in CTTYE broth (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 $KH(H_2)PO_4$, 8 mM $MgSO_4$) supplemented with 10 µg/mL oxytetracycline and 1 mM isopropyl β-D-1-thiogalactopyrano-side (IPTG) for LS3908 or with 40 µg/mL kanamycin for DK10547. For development assays, mid-log phase cells were harvested, resuspended in TPM starvation buffer (10 mM Tris (pH 7.6), 1 mM $KH(H_2)PO_4$, 8 mM $MgSO_4$) to a concentration of 5×10^9 cells/mL (or 1×10^{10} cells/mL for high-density and 2.5×10^9 cells/mL for low-density samples) and plated on a microscope slide chamber prepared, as previously described (8),

containing 1% agarose TPM media with 1 mM IPTG added. To track cells in streams during development, LS3908 cells were diluted 1:800 into DK10547 and plated on a microscope slide chamber as above. To induce rippling, *E. coli* K12 cells were grown overnight in LB broth (Sigma, St. Louis, MO, USA) in a 37 ◦C incubator with vigorous shaking, harvested and washed in TPM buffer, and plated on 1% or 0.6% agarose microscope slide chambers containing TPM supplemented with 1 mM IPTG. Once cell spots of *E. coli* were dry, LS3908 cells from an overnight culture were prepared in TPM as above, 3 μL were plated in the center of the *E. coli* spot, and the slide was incubated in the dark at 32 ◦C for 8–10 h before imaging to provide time for rippling to initiate.

Time-Lapse Imaging

Microscope slide chambers were placed on a stage warmer (20/20 Technologies, Wilmington, NC, USA) set to 32 ◦C on a Nikon Eclipse E-400 microscope (Nikon Instruments, Melville, NY, USA). A pco.panda 4.2 sCMOS camera and NIS-Elements software were used for automated time-lapse imaging, capturing a phase contrast and fluorescent image every 60 s for a total of 24 h for development movies and 8 h for rippling movies. Phase-contrast images were taken with 70 ms exposure, and transmitted light was manually shuttered with a Uniblitz VMM-D1 shutter (Uniblitz Electronics, Rochester, NY, USA) when not actively imaging. Fluorescent tdTomato-expressing samples were imaged with 400 ms exposure with a Sola LED light source (Lumencore) at 75% intensity, and GFP-expressing samples were imaged with 200 ms exposure at 35% intensity. A MAC6000 system filter wheel controller and focus control module (Ludl Electronic Products, Ltd., Hawthorne, NY, USA) were used for control of the fluorescent filter wheel and the autofocus feature.

REFERENCES

1. Otaki T (2001) Halo Reduction Technique in Phase Contrast Microscopy. *Opt Rev* 8(4):284–286.

2. Maurer C, Jesacher A, Bernet S, Ritsch-Marte M (2008) Phase contrast microscopy with full numerical aperture illumination. *Opt Express* 16(24):19821–19829.

3. Icha J, Weber M, Waters JC, Norden C (2017) Phototoxicity in live fluorescence microscopy, and how to avoid it. *Bioessays* 39(8). doi:10.1002/bies.201700003.

4. Yu R, Kaiser D (2007) Gliding motility and polarized slime secretion. *Mol Microbiol* 63(2):454–467.

5. Kaiser D, Welch R (2004) Dynamics of Fruiting Body Morphogenesis Dynamics of Fruiting Body Morphogenesis. 186(4):919–927.

6. Cotter CR, Schüttler H, Igoshin OA, Shimkets LJ (2017) Data-driven modeling reveals cell behaviors controlling self-organization during Myxococcus xanthus development. *Proc Natl Acad Sci*. doi:10.1073/pnas.1620981114.

7. Welch R, Kaiser D (2001) Cell behavior in traveling wave patterns of myxobacteria. *Proc Natl Acad Sci U S A* 98(26):14907–14912.

8. Comstock JA, et al. (2022) Phenotypic similarity is a measure of functional redundancy within homologous gene families. *bioRxiv*. doi:10.1101/2022.07.25.501402.

APPENDIX 2: LIST OF GENES USED IN CHAPTERS 2 & 3

Table 1: List of strains used in studies in Ch. 2&3, listed by MXAN number, followed by the manual phenotype classification associated with Ch. 2, gene family, citation for creation of the specific strain used. The final column indicates whether pulsing was observed in at least one replicate (Ch. 3).

Strain	Phenotypic Classification	Gene Family	Strain Creation	Pulsing Observed
DK1622	Wild-type	N/A		yes
MXAN_0035	No aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_0036	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0037	Late aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_0069	Early aggregation	One Component	This study	yes
MXAN_0079	Variable	One Component	(Ritchie *et al*, 2021)	no
MXAN_0090	Late aggregation	One Component	This study	no
MXAN_0107	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0108	LWT	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_0116	No aggregation	NtrC-Like Activators	This study	no
MXAN_0172	Fall apart	NtrC-Like Activators	(Ritchie *et al*, 2021)	no
MXAN_0180	LWT	NtrC-Like Activators	This study	yes
MXAN_0203	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_0213	Variable	One Component	(Ritchie *et al*, 2021)	no
MXAN_0214	Variable	One Component	This study	yes
MXAN_0233	Immature aggregates	ECF Sigma Factors	This study	no
MXAN_0250	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_0251	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0353	Early aggregation	NtrC-Like Activators	This study	yes
MXAN_0387	Late aggregation	One Component	This study	no
MXAN_0445	Early aggregation	One Component	This study	yes
MXAN_0502	Variable	One Component	(Ritchie *et al,* 2021)	yes
MXAN_0553	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0554	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0556	LWT	One Component	This study	yes
MXAN_0559	Variable	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_0596	Immature aggregates	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0597	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0603	Late aggregation	NtrC-Like Activators	(Ritchie *et al,* 2021)	no
MXAN_0622	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_0627	Early aggregation	One Component	This study	yes
MXAN_0629	Immature aggregates	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_0654	LWT	One Component	This study	no
MXAN_0665	Variable	One Component	(Ritchie *et al,* 2021)	yes

MXAN_0681	Aggregate-reaggregate	ECF Sigma Factors	(Ritchie *et al,* 2021)	no
MXAN_0684	Late aggregation	ABC Transporter	(Yan *et al,* 2014)	no
MXAN_0685	No aggregation	ABC Transporter	(Yan *et al,* 2014)	no
MXAN_0686	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0687	Late aggregation	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0696	Early aggregation	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0707	Variable	One Component	This study	yes
MXAN_0721	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0722	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0748	LWT	One Component	This study	yes
MXAN_0751	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0770	LWT	ABC Transporter	(Yan *et al,* 2014)	no
MXAN_0771	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0772	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0832	Variable	One Component	This study	no
MXAN_0887	Immature aggregates	One Component	This study	no
MXAN_0907	Early aggregation	NtrC-Like Activators	This study	yes
MXAN_0937	LWT	NtrC-Like Activators	This study	yes
MXAN_0943	Variable	One Component	This study	yes
MXAN_0947	LWT	ECF Sigma Factors	This study	yes
MXAN_0966	Early aggregation	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0967	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_0968	Late aggregation	ABC Transporter	(Yan *et al,* 2014)	no
MXAN_0995	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_1078	Immature aggregates	NtrC-Like Activators	(Ritchie *et al,* 2021)	no
MXAN_1097	LWT	ABC Transporter	(Yan *et al,* 2014)	no
MXAN_1124	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_1128	Variable	NtrC-Like Activators	(Ritchie *et al,* 2021)	no
MXAN_1137	Variable	One Component	This study	yes
MXAN_1151	Variable	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_1153	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_1154	LWT	ABC Transporter	(Yan *et al,* 2014)	no
MXAN_1155	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_1167	Late aggregation	NtrC-Like Activators	(Ritchie *et al,* 2021)	no
MXAN_1189	Early aggregation	NtrC-Like Activators	This study	yes
MXAN_1210	Late aggregation	ECF Sigma Factors	This study	no
MXAN_1245	Immature aggregates	NtrC-Like Activators	(Ritchie *et al,* 2021)	no
MXAN_1262	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_1286	Aggregate-reaggregate	ABC Transporter	(Yan *et al,* 2014)	no
MXAN_1319	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_1320	Early aggregation	ABC Transporter	(Yan *et al,* 2014)	yes

MXAN_1321	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_1345	LWT	NtrC-Like Activators	This study	yes
MXAN_1376	Late aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_1377	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_1402	Variable	One Component	This study	yes
MXAN_1510	LWT	ECF Sigma Factors	This study	yes
MXAN_1514	Immature aggregates	ECF Sigma Factors	This study	no
MXAN_1547	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_1548	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_1565	Variable	NtrC-Like Activators	(Ritchie *et al*, 2021)	no
MXAN_1575	Variable	One Component	This study	yes
MXAN_1597	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_1598	Variable	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_1604	Variable	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_1605	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_1654	LWT	One Component	This study	yes
MXAN_1661	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_1667	LWT	One Component	This study	yes
MXAN_1677	Variable	One Component	This study	yes
MXAN_1683	LWT	One Component	This study	yes
MXAN_1695	Immature aggregates	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_1711	No aggregation	One Component	This study	no
MXAN_1719	LWT	One Component	This study	yes
MXAN_1726	Variable	One Component	This study	yes
MXAN_1746	Variable	One Component	This study	yes
MXAN_1757	Variable	One Component	This study	yes
MXAN_2018	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2019	No aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_2020	Immature aggregates	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2030	Early aggregation	ECF Sigma Factors	(Ritchie *et al*, 2021)	yes
MXAN_2078	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2128	Immature aggregates	One Component	This study	no
MXAN_2145	Late aggregation	One Component	This study	no
MXAN_2159	Early aggregation	NtrC-Like Activators	This study	yes
MXAN_2184	Aggregate-reaggregate	ECF Sigma Factors	This study	no
MXAN_2204	Immature aggregates	ECF Sigma Factors	This study	no
MXAN_2230	Late aggregation	One Component	This study	no
MXAN_2234	Immature aggregates	One Component	This study	no
MXAN_2249	Late aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_2250	LWT	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_2251	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes

MXAN_2268	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2395	Early aggregation	ECF Sigma Factors	This study	no
MXAN_2407	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2428	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2429	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2430	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2437	LWT	ECF Sigma Factors	This study	no
MXAN_2500	Early aggregation	ECF Sigma Factors	This study	no
MXAN_2501	Aggregate-reaggregate	NtrC-Like Activators	This study	no
MXAN_2516	Immature aggregates	NtrC-Like Activators	This study	no
MXAN_2654	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2711	Variable	One Component	(Ritchie *et al*, 2021)	no
MXAN_2783	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2794	Immature aggregates	One Component	This study	no
MXAN_2795	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2831	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2832	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2833	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2853	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_2896	LWT	One Component	This study	no
MXAN_2929	LWT	ECF Sigma Factors	This study	no
MXAN_2949	Aggregate-reaggregate	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_2951	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_3095	Early aggregation	NtrC-Like Activators	This study	yes
MXAN_3142	Immature aggregates	One Component	This study	no
MXAN_3151	Immature aggregates	One Component	This study	no
MXAN_3208	Immature aggregates	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_3209	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_3214	Fall apart	NtrC-Like Activators	(Ritchie *et al*, 2021)	yes
MXAN_3240	LWT	One Component	This study	yes
MXAN_3256	No aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_3257	No aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_3258	No aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_3333	Variable	NtrC-Like Activators	This study	no
MXAN_3339	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_3381	Aggregate-reaggregate	NtrC-Like Activators	This study	no
MXAN_3418	LWT	NtrC-Like Activators	This study	yes
MXAN_3426	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_3429	Early aggregation	One Component	This study	yes
MXAN_3443	Immature aggregates	One Component	This study	yes
MXAN_3648	Variable	ABC Transporter	(Yan *et al*, 2014)	no

MXAN_3650	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_3686	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_3692	No aggregation	NtrC-Like Activators	This study	no
MXAN_3702	No aggregation	One Component	(Ritchie *et al*, 2021)	no
MXAN_3711	Immature aggregates	One Component	This study	no
MXAN_3717	No aggregation	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_3718	Fall apart	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_3773	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_3811	LWT	NtrC-Like Activators	This study	yes
MXAN_3908	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_3909	Early aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_3959	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_3986	LWT	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_4042	Immature aggregates	NtrC-Like Activators	This study	yes
MXAN_4060	LWT	One Component	This study	yes
MXAN_4072	Late aggregation	One Component	This study	no
MXAN_4110	Late aggregation	One Component	This study	no
MXAN_4173	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_4196	No aggregation	NtrC-Like Activators	(Ritchie *et al*, 2021)	no
MXAN_4199	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_4240	LWT	NtrC-Like Activators	This study	yes
MXAN_4247	Late aggregation	One Component	This study	yes
MXAN_4252	Late aggregation	NtrC-Like Activators	This study	no
MXAN_4261	LWT	NtrC-Like Activators	This study	yes
MXAN_4263	LWT	One Component	This study	yes
MXAN_4309	Early aggregation	ECF Sigma Factors	This study	no
MXAN_4316	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_4339	LWT	NtrC-Like Activators	This study	no
MXAN_4356	LWT	One Component	This study	yes
MXAN_4471	LWT	One Component	This study	no
MXAN_4523	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_4580	Early aggregation	NtrC-Like Activators	This study	yes
MXAN_4622	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_4662	Immature aggregates	ECF Sigma Factors	This study	no
MXAN_4665	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_4716	Fall apart	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_4733	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_4749	LWT	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_4750	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_4785	Late aggregation	NtrC-Like Activators	This study	yes
MXAN_4790	Variable	ABC Transporter	(Yan *et al*, 2014)	no

MXAN_4899	Fall apart	NtrC-Like Activators	This study	no
MXAN_4949	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_4977	Early aggregation	NtrC-Like Activators	This study	yes
MXAN_4983	Variable	NtrC-Like Activators	This study	no
MXAN_4987	LWT	ECF Sigma Factors	This study	yes
MXAN_5029	No aggregation	One Component	This study	no
MXAN_5041	Immature aggregates	NtrC-Like Activators	This study	no
MXAN_5048	Late aggregation	NtrC-Like Activators	This study	yes
MXAN_5101	Early aggregation	ECF Sigma Factors	(Ritchie *et al*, 2021)	yes
MXAN_5124	Fall apart	NtrC-Like Activators	(Ritchie *et al*, 2021)	no
MXAN_5128	LWT	One Component	This study	yes
MXAN_5153	Early aggregation	NtrC-Like Activators	(Ritchie *et al*, 2021)	no
MXAN_5245	Late aggregation	ECF Sigma Factors	This study	no
MXAN_5263	No aggregation	ECF Sigma Factors	(Ritchie *et al*, 2021)	no
MXAN_5271	No aggregation	One Component	This study	no
MXAN_5276	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_5305	LWT	One Component	This study	yes
MXAN_5356	Early aggregation	One Component	This study	no
MXAN_5379	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_5410	Early aggregation	ECF Sigma Factors	(Ritchie *et al*, 2021)	no
MXAN_5480	Early aggregation	One Component	This study	yes
MXAN_5492	LWT	One Component	This study	yes
MXAN_5503	LWT	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_5506	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_5545	Variable	One Component	This study	yes
MXAN_5547	LWT	One Component	This study	yes
MXAN_5584	LWT	ABC Transporter	(Yan *et al*, 2014)	no
MXAN_5680	Variable	NtrC-Like Activators	(Ritchie *et al*, 2021)	yes
MXAN_5731	Immature aggregates	ECF Sigma Factors	This study	no
MXAN_5777	Variable	NtrC-Like Activators	(Ritchie *et al*, 2021)	no
MXAN_5853	Variable	NtrC-Like Activators	This study	no
MXAN_5879	Fall apart	NtrC-Like Activators	(Ritchie *et al*, 2021)	no
MXAN_5894	Variable	One Component	(Ritchie *et al*, 2021)	no
MXAN_6000	Late aggregation	ABC Transporter	(Yan *et al*, 2014)	yes
MXAN_6058	Variable	ECF Sigma Factors	This study	yes
MXAN_6149	LWT	One Component	This study	yes
MXAN_6157	LWT	One Component	This study	no
MXAN_6161	LWT	One Component	This study	no
MXAN_6167	Variable	One Component	This study	yes
MXAN_6173	Early aggregation	ECF Sigma Factors	(Ritchie *et al*, 2021)	yes
MXAN_6206	Variable	One Component	This study	no

MXAN_6251	Variable	One Component	This study	no
MXAN_6402	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_6426	No aggregation	NtrC-Like Activators	(Ritchie *et al,* 2021)	no
MXAN_6461	Fall apart	ECF Sigma Factors	This study	no
MXAN_6468	Variable	One Component	This study	no
MXAN_6475	Early aggregation	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_6479	LWT	One Component	This study	no
MXAN_6486	LWT	One Component	This study	yes
MXAN_6518	Variable	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_6549	Late aggregation	One Component	This study	no
MXAN_6551	LWT	ABC Transporter	(Yan *et al,* 2014)	yes
MXAN_6646	LWT	One Component	This study	yes
MXAN_6653	No aggregation	One Component	This study	no
MXAN_6759	Immature aggregates	ECF Sigma Factors	This study	no
MXAN_6833	Variable	One Component	This study	no
MXAN_6889	No aggregation	One Component	(Ritchie *et al,* 2021)	no
MXAN_6967	Late aggregation	One Component	This study	no
MXAN_7072	Variable	One Component	This study	yes
MXAN_7078	LWT	One Component	This study	no
MXAN_7214	Early aggregation	ECF Sigma Factors	This study	yes
MXAN_7289	Immature aggregates	ECF Sigma Factors	This study	yes
MXAN_7312	LWT	One Component	This study	yes
MXAN_7316	LWT	One Component	This study	yes
MXAN_7322	Late aggregation	One Component	This study	no
MXAN_7326	LWT	ECF Sigma Factors	This study	yes
MXAN_7440	No aggregation	NtrC-Like Activators	(Ritchie *et al,* 2021)	no
MXAN_7454	Early aggregation	ECF Sigma Factors	This study	yes

APPENDIX 3: SUPPLEMENTARY METHODS FOR CHAPTER 2

Supplementary Methods

Image processing

Custom Python code was written for this analysis, available on Github (https://github.com/masp01/SU-myxo-aggregate-tracking). Using the Python implementation of OpenCV (Bradski, 2000), each raw frame is put through the following image processing steps to identify the size, shape, and position of each fruiting body:

1. Non-local means denoising (cv2.fastNlMeansDenoising)
   To remove background noise, a 7 pixel wide template window is moved over the image to find regions that visually match (typically uniform, noisy regions). This search is done within a 21 pixel distance of each patch of the image. The gray value of each pixel is replaced with the average gray value of pixels in matching regions, smoothing over noise while keeping boundaries distinct. The smoothing strength was chosen at a constant value of 70 after manually testing parameter values for many images. Template window and search sizes are standard and were not tuned.

2. Adaptive thresholding (cv2.adaptiveThreshold)
   To identify locally dark regions that should belong to fruiting bodies, the gray value of each pixel is compared to the average gray value of its neighbors within a block 101 pixels (145 µm) wide. 8-bit pixels (gray value from $0 - 255$) that have a gray value at least 20 below this local average are marked white. All other pixels are marked black, creating a binary image. Parameters were chosen after manually testing with many images and are robust enough to be used across the entire dataset.

3. Morphological opening (cv2.morphologyEx)
   A circular kernel 5 pixels (7.2 µm) wide is moved over the binary image. Any feature covered entirely by the kernel is removed. This reduces single-pixel noise.

4. Contour identification (cv2.findContours)
   Contiguous regions of white pixels are automatically identified in the binary image. A list is compiled of the x,y coordinates of the pixels on the boundary of each such region. This gives both a count of total candidate fruiting bodies and the geometry of their boundary.

At this point, a list of features has been identified, some of which are genuine fruiting bodies, and some of which are noise or spurious aggregates. The contour of each feature is measured for the x,y coordinates of its center, its area *A*, perimeter *P*, and average gray value. The circularity $4\pi A/P^2$, is also calculated. It captures the elongation of the fruiting bodies and ranges from 0 (completely flat) to 1 (perfectly circular).

Tracking fruiting bodies and filtering

Once all the frames of a time series have been processed, the Python package Trackpy (Allan, D. B., Caswell, T., Keim, N. C. & van der Wel, C. M. *trackpy: Trackpy v0.4.2* doi:10.5281/zenodo.4682814). is used to assign an ID to each feature that tracks it over time. It is at this point that filtering is done to remove spurious features:

1. Minimum area filter
   Features that are smaller than 576 µm² are ignored. This is the smallest fruiting body size that is distinguishable from noise at 4X magnification

2. Maximum gray value filter
   Features with an average gray value above 200 (max 255) are considered too bright to be fruiting bodies and are ignored.
3. Formation time filter
   Features that appear before 100 minutes have elapsed are incidental initial aggregates, and not genuine fruiting bodies that have assembled over time. In no time series did a new aggregate form in less than 100 minutes. These incidental aggregates are tracked over time and ignored in all frames in which they appear.
4. Category filter
   The area dynamics of each remaining fruiting body are considered to see if the fruiting body persists to the end of the time series (persistors) or if it vanishes smoothly (evaporators). Persistors with an average circularity below 0.5 are typically noise and are ignored. Smoothly vanishing is defined as starting with an area less than max area and then decreasing from maximum area by at least 25% by the final frame of the time series. Evaporators with centers within (14.4 µm of the edge of the frame or with an average circularity below 0.5 are considered noise and ignored. Any feature that cannot be categorized as a persistor or evaporator is assumed to be spurious or contain dynamics errors and is ignored.

Feature extraction

The data for each time series is then analyzed to measure the following quantitative features, each a single number summarizing one aspect of the time series. Measurements are taken over a 7.2 mm² field size.

Table S2: Enumeration of all quantitative features used in the automated phenotype analysis

	Feature name	**Description**	**Formula**
1	Start time	Elapsed time from inoculation to the beginning of visible aggregation	When at least 10 fruiting bodies grow larger than 1000 µm² in area
2	Peak time	Elapsed time from inoculation to the peak of visible aggregation	When total fruiting body area reaches a maximum
3	Stability time	Elapsed time from inoculation to overall aggregate stability	When the number of fruiting bodies changes by less than three per hour (24 hours maximum)
4	Growth time	Duration of initial growth phase	Peak time minus start time
5	Growth rate	Average rate of total area increase during growth phase	Change in total area divided by change in time between start time and peak time
6	Peak average area	The average fruiting body area at peak time	

7	Peak area std	The standard deviation of fruiting body area at peak time	
8	Final average area	The average fruiting body area at the moment 24 hours after inoculation	
9	Final area std	The standard deviation of fruiting body area 24 hours post-inoculation	
10	Gray value % change	Percent difference between minimum and maximum average gray value (only for persistent fruiting bodies)	
11	Maturation rate	Maximum slope of gray value vs. time curve for persistent fruiting bodies	
12	Temporal coherence	How closely in time each evaporating fruiting body reaches its maximum area before starting to shrink	The standard deviation of the distribution of the time of maximum area for evaporators
13	Fraction of evaporators	Total number of evaporators divided by total number of evaporators plus persistors	
14	Maximum number	Total number of fruiting bodies at peak time	
15	Average lifetime	The elapsed time between an evaporator's first and final moment above the minimum area threshold, averaged over all evaporators	
16	Std lifetime	Standard deviation of the elapsed times between each evaporator's first and final moment above the minimum area threshold	
17	Maximum average area falloff		Most negative slope of average area vs. time curve
18	Maximum number falloff		Most negative slope of number vs. time curve

REFERENCES

Bradski G (2000) The OpenCV Library. *Dr Dobb's J Softw Tools*

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A & Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*: 1–4

Ritchie LJ, Curtis ER, Murphy KA & Welch RD (2021) Profiling Myxococcus xanthus Swarming Phenotypes through Mutation and Environmental Variation. *J Bacteriol* 203

Yan J, Bradley MD, Friedman J & Welch RD (2014) Phenotypic profiling of ABC transporter coding genes in Myxococcus xanthus. *Front Microbiol* 5: 1–12

VITA

Author: Jessica A. Comstock   **Date of Birth:** 09 June 1991

Undergraduate Education:

    Westfield State University, Bachelor of Science, Biology   2013

Awards and Honors:

    Outstanding TA Award, Syracuse University   2019

    IGERT Traineeship, Syracuse University   2015-2017

Publications:

Comstock JA, Asp ME, Bahar FA, Lee I, Patteson AE, and Welch RD. 2022. Phenotypic Similarity as an Indicator of Functional Redundancy Within Large Homologous Groups. (*submitted to Molec Systems Biol)*

Comstock JA, Bahar F, Murphy PA, Igoshin OA, and Welch RD. 2022. Traveling Wave Synchronizes Cell Behavior During *Myxococcus xanthus* Development. (*in prep*)

Zhang J, **Comstock JA**, Cotter CR, Murphy PA, Nie W, Welch RD, Patel AB, and Igoshin OA. 2021. Quantification of *Myxococcus xanthus* Aggregation and Rippling Behaviors: Deep-learning Transformation of Phase-contrast into Fluorescence Microscopy Images. *Microorganisms.* doi: 10.3390/microorganisms9091954