

Syracuse University

SURFACE

Dissertations - ALL

SURFACE

5-14-2017

Evolution, epistasis, and the genotype-to-phenotype problem in *Myxococcus xanthus*.

Michael Bradley
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/etd>



Part of the [Life Sciences Commons](#)

Recommended Citation

Bradley, Michael, "Evolution, epistasis, and the genotype-to-phenotype problem in *Myxococcus xanthus*."
(2017). *Dissertations - ALL*. 674.
<https://surface.syr.edu/etd/674>

This Dissertation is brought to you for free and open access by the SURFACE at SURFACE. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Abstract

The complex social behavior of *M. xanthus* makes it an excellent model system to study the relationship between genotype and phenotype. Under nutrient rich conditions, a swarm of *M. xanthus* cells coordinate their movement outward in search of prey. When starved, cells condense into multicellular structures called aggregates. Taken together, these two aspects of the *M. xanthus* life cycle display several sub-traits that are used to describe its phenotype. Furthermore, the genome of *M. xanthus* is large, encoding a predicted 7,314 genes, many of which have been linked to aspects of its multicellular phenotype.

This work presented here addresses the genotype-to-phenotype (G2P) problem as it relates to the annotation of a biological process in a model system. The first project addresses G2P from a population genetics approach; we constructed a mutant strain library consisting of 180 single gene knockouts of the ABC transporter superfamily of genes to examine the distribution of mutant phenotypes among an entire group of genes. While the phenotype of only ~10% of mutants show extreme defects, more than three quarters of mutants are parsed into different categories of phenotypic deviation following our analyses. Our results demonstrate that strong mutant phenotypes are uncommon, but the majority of null mutants are phenotypically distinct from wild type in at least one trait. Thus, a more comprehensive understanding of the *M. xanthus* phenome will help elucidate the biological function of many uncharacterized genes.

The second part of this dissertation examines the evolution of *M. xanthus* as it has been studied as a model organism in different laboratories. Disrupting a gene, or mutating a single nucleotide, may have no discernable impact on the organism's phenotype by itself, but may still substantially affect the phenotypes of additional mutation through epistasis. This is an ongoing phenomena in *M. xanthus*; whole genome resequencing and phenotypic characterization of several inter-laboratory isolates of *M. xanthus* wild type DK1622 revealed genomic variation that has resulted in significant phenotypic variation. We demonstrate that the naturally occurring

genetic variants among wild type isolates is sufficient to mask the effect of a targeted mutation in one isolate that is significant in another. These results are the first to indicate that isolates of wild type *M. xanthus* DK1622 have evolved to a functionally significant degree.

**Evolution, epistasis, and the genotype-to-phenotype problem
in *Myxococcus xanthus*.**

Michael D. Bradley

B.S. Biology, Syracuse University 2008

DISSERTATION

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

Syracuse University

May 2017

Copyright © Michael D. Bradley 2017

All Rights Reserved

ACKNOWLEDGEMENTS

I extend my sincerest gratitude to my advisor, mentor, and friend, Dr. Roy Welch. I would not have been able to complete this program without your continuous support, both personally and professionally. Graduate school is a humbling experience and I am immensely grateful to be included in your lab.

I am hugely indebted to Dr. Jinyuan Yan for teaching me about our study system, and helping me transition into graduate school. I would also like thank Drs. Anthony Garza, John Belote, Dacheng Ren, Eleanor Maine, Jannice Friedman, and everyone in the Welch lab for helpful discussion over the years. Lastly, I would like to thank Syracuse University and The National Science Foundation for supporting my projects.

TABLE OF CONTENTS

Chapter 1. Introduction.....	1
1.1 Motivation.....	2
1.2 Aims of research	3
1.3 Contributions	3
1.4 Organization.....	4
 Chapter 2. Background & Significance	5
2.1 Introduction	6
2.2 The genotype-to-phenotype relationship	7
2.3 Evolutionary approaches to G2P	12
2.4 Obstacles of G2P studies.....	14
2.5 <i>Myxococcus xanthus</i> as a model organism	17
2.6 Summary.....	23
 Chapter 3. Phenotypic Profiling of a Large Mutant Library.....	31
3.1 Project summary	32
3.2 Materials and methods	32
3.3 Results	36
3.4 Discussion.....	40
 Chapter 4. Inter-laboratory Evolution of Wild Type Sublines.....	48
4.1 Project summary	49
4.2 Materials and methods	49
4.3 Results	51
4.4 Discussion.....	57

Concluding remarks	66
Appendices	67
Appendix I Primers to construct ABC transporter mutants (Chapter 3)	67
Appendix II Source of <i>M. xanthus</i> sublines (Chapter 4)	73
Appendix III Pairwise comparison of A-motility phenotypes (Chapter 4)	74
Appendix IV Pairwise comparison of S-motility phenotypes (Chapter 4)	75
Appendix V Pairwise comparison of aggregation phenotypes (Chapter 4)	76
Appendix VI Pairwise comparison of sporulation phenotypes (Chapter 4)	77
Appendix VII Summary of read mappings (Chapter 4)	78
Appendix VIII Primers used to generate DNA inserts for plasmids (Chapter 4)	79
Appendix IX Plasmids used to construct mutant strains (Chapter 4)	80
References	81
Vita	93

TABLES & FIGURES

Table 2.1 Notable sequenced organisms	24
Table 4.1 Summary of ANOVAs	60
Table 4.2 Summary of ANOVAs with outliers removed	61
Table 4.3 Subline variant screen.....	62
Figure 2.1 Experimental and evolutionary approaches to phenotypic profiling	25
Figure 2.2 Site-directed mutagenesis and homologous recombination	26
Figure 2.3 Schematic of positive and negative epistasis on fitness	27
Figure 2.4 Life cycle of <i>M. xanthus</i> (simplified)	28
Figure 2.5 Phenotypic metrics used to describe <i>M. xanthus</i>	29
Figure 2.6 Origin of wild type <i>M. xanthus</i>	30
Figure 3.1 Distribution of ABC transporters in the <i>M. xanthus</i> genome	43
Figure 3.2 Phenotypic traits and phenotypic assays	44
Figure 3.3 Distribution of phenotypic data.....	45
Figure 3.4 Correlation of phenotypic variables.....	46
Figure 3.5 Phenotypic traits of wild type and mutant aggregates during development	47
Figure 4.1 Characterization of DK1622 subline phenotypes in a common garden	63
Figure 4.2 Characterization of S8 mutant strains	64
Figure 4.3 Epistasis in S1, S8, and S9 mutant strains.....	65

CHAPTER 1
INTRODUCTION

1.1 Motivation

Despite the wealth of available sequencing data, our ability to link gene products to their functions remains poor. The most standard method to ascertain gene function is to analyze phenotypic data when the gene is inactivated. Mutating a gene and observing the biological process that is disrupted will sometimes provide clues towards gene function. The guiding assumption of all genotype-to-phenotype (G2P) studies is that genetic perturbation causes a discrete change in phenotype that can be used to infer the biological function of a mutated gene. This genotype-first method of annotation relates one genotype to one phenotype, sectioning the organism into smaller parts that attempt to explain the whole. Genetic reductionism is a perfectly acceptable approach to uncover the causes of phenotypic diversity, but is limited in two ways: First, it is difficult to assign function to the proportion of mutants that display no deviation in phenotype. Redundancy may account for a lack of detectable phenotypes within a large family of proteins, but is unlikely to explain absent phenotypes for orphaned genes. Second, the phenotypic effects of mutation may not be reproducible in closely related yet genetically distinct cell lines. How can these disconnects between genotype and phenotype be reconciled to create a substantive map to infer biological function from phenotypic data?

The “no phenotype” phenomena can be addressed from two perspectives: First, a G2P diagram is often inaccessible without a comprehensive view of an organism’s phenome (the sum of all possible phenotypes). Subtle phenotypes may be hidden by broad profiling that examine only a subset of traits such as growth and morphology. It would be experimentally cumbersome to test tens or hundreds of assay conditions for thousands of mutant strains; therefore, extracting more information from the existing characterization assays is the most feasible approach to a better understanding of the phenome, including the structures between phenotypic traits. “Phenotypic sensitivity” can be can be addressed by determining if the model organism has evolved while it has been studied. Most G2P experiments manipulate a wild type

strain to generate mutant strains that are scored for changes in certain traits. For experimental purposes, isolates of a wild type studied by a research community are assumed to be isogenic. However, the laboratory is a model organism's natural environment, and evolution continues during and between experiments. Heterogeneity of the genetic background results after decades of research. Superficially, isolates may be indistinguishable, displaying the phenotypic prerequisites of wild type, while at the same time accumulate mutations that significantly diversify their genotypes. Epistasis between naturally occurring and targeted mutations may influence the results of phenotypic screens and the subsequent annotations of gene products.

1.2 Aims of research

The work presented here examines the complex relationship between genotype and phenotype in *Myxococcus xanthus*, a model bacterium for biofilm development. *M. xanthus* is unique in that it has one of the most complex life cycles in the bacterial world, capable of self-organization and cellular differentiation in response to environmental cues. Our first aim is to better define the intersection between wild type and mutant phenotype through a rigorous characterization and quantitative analyses of several phenotypic traits among a large group of single gene mutants. Our second aim is to determine the extent of *M. xanthus* microevolution while it has been studied by the field. The history of wild type *M. xanthus* can be traced back 40 years because it has a single origin and a sequenced genome.

1.3 Contributions

1. The first large-scale statistical analysis of phenotypic traits in *M. xanthus*. Our approach was to create and characterize a library of mutant strains (the ABC Transporter superfamily) and profile the resulting phenotypes in search of new mutant phenotypes, and to identify correlations among traits.

2. The first whole genome sequence comparison of inter-laboratory isolates of wild type *M. xanthus* strain DK1622. Our approach was to resequence multiple laboratory isolates to link genetic variants with the observed phenotypic variations. Lastly, we compare the phenotypic response of multiple isolates to an identical targeted mutation.

1.4 Organization

Chapter 2 of this dissertation introduces the concepts of genotype and phenotype, the use of phenotypic screens for functional analysis, and the obstacles for interpreting G2P data. Also introduced in Chapter 2 is *M. xanthus* as a model system, including descriptions of the major genetic pathways. Chapters 3 and 4 describe the projects listed in **section 1.3**; each chapter consists of a project summary, materials and methods, results, and discussion.

CHAPTER 2
BACKGROUND & SIGNIFICANCE

2.1 Introduction

An ambitious goal for the post-genomics era is to assign a function to every gene in an organism's genome. The first step towards accomplishing this goal is to search for strong sequence homology among previously annotated genes; function can be inferred from a homolog. Genome-wide analyses across multiple organisms *in silico* allows researchers to classify categories of well-characterized gene products such as enzymes, transporters, and receptors [1–3]. Orthologs in phenotypically similar organisms, such as spore forming bacteria, provide putative annotations in lieu of experimentation [4–6]. However, annotation via sequence homology remains an imperfect tool; highly conserved proteins may have entirely different structures and biological functions [7,8], while partial sequence alignments suggest a conserved domain but offer little insight with respect to function. Further compounding the problem are genes with no sequence homology; the majority of bacterial and archaeal genomes are predicted to consist of 20-40% orphaned hypothetical genes [9–11]. The absence of sequence homology provides no basis to predict function.

Counterintuitively, a more powerful approach to determine function is to inactivate a gene and observe the phenotypic response of the organism; gene function can also be inferred by mutation. The genotype-to-phenotype (G2P) experimental approach compares the phenotypic state of a mutant strain to the phenotypic state of a wild type reference - a convenient shortcut for functional analyses. If a mutation causes a variation in some observable trait, then the function of the gene product is assumed to be involved in the manifestation of that trait (**Fig. 2.1a**). For example, in the pathogenic bacterium *Salmonella typhimurium*, a mutation in *flgJ* results in non-motile cells. The initial functional annotation of this gene could be “involved in motility” and would direct future experiments towards understanding the cellular and molecular function of FlgJ with respect to motility [12].

Both methods of functional annotation require an understanding of the articulation between genotype and phenotype. The following sections introduce the history of G2P, genome

sequencing and its applications, and examples of high-throughput phenotypic profiling of model organisms. Finally, I will describe the current obstacles that limit our ability to assign functional annotations using phenotypic data.

2.2 The genotype-phenotype relationship

The definition of a gene is relatively straightforward - a sequence of nucleotides that is replicated, transcribed, and operates a cellular function [13]. The term genotype is used broadly to describe the total set of genes and genetic material, or narrowly to specify loci of interest. Genotype may also refer to pseudogenes and noncoding elements such as small RNAs and junk DNA, which have been shown to regulate expression of nearby genes [14–17]. The definition of phenotype is more complex and is classified several ways: (1) the molecular phenotype describes the processes that are detected by technical instruments, such as transcripts and small molecule production (e.g. metabolites); (2) the cellular phenotype describes the empirical traits of individuals such as growth rate and morphology; and (3) the system phenotype describes the dynamic processes such as cell-cell signaling and emergent behavior of groups. The abstract concept of “extended phenotype” includes unique behaviors such as beaver dam and caddisfly house construction as a phenotype [18]. Clearly, the concept of phenotype is multidimensional.

Initially hinted at by Gregor Mendel’s experiments with hybrid plants, the genotype-phenotype distinction was formally introduced by Wilhelm Johannsen in the early 1900s [19] and has provided the conceptual framework for modern geneticists. For nearly a century, researchers have sought to link these two elements into a coherent map that can be used to predict one from the other. In other words, using genes and their sequences to predict an organism’s phenotype, and map *post hoc* changes in phenotype to specific genes. This is known as the G2P problem [20]. Understanding this relationship allows researchers to design or alter a biological system to exploit a desirable trait. The broader impacts of G2P studies are

widespread, including the synthesis of biomaterials, identifying new antibiotic targets, and understanding the genetics of human disease.

Genome sequencing and population genetics

To understand the biology of an organism, it is necessary to determine its entire genome sequence. The advent of next-generation sequencing technology has revolutionized the way researchers think about basic and clinical sciences. Indeed, genome sequencing has been an important tool in the fields of forensic science [21,22], biotechnology [23,24], and molecular biology [25–27].

The first complete genome sequence of a free-living organism was of the pathogenic bacterium *Haemophilus influenzae* in 1995 [28]. Since this initial project, thousands of other organisms have been sequenced, ranging from viruses [29–31] to the consortium-based human genome project [32] (see **Table 2.1** for a list of major sequencing projects). The availability of fully sequenced genomes has motivated numerous comparative studies between organisms of varying complexity. For example, Rubin *et al.* report >20% orthologous genes among *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*, which suggests evidence of a core genome among eukaryotes [33]. Similarly, it has been reported that ~75% of genes associated with human disease have a functional homolog in *D. melanogaster* [34]. Clearly, the availability of sequencing data has expanded the fields of evolutionary biology and phylogenetics, and is integral for future therapeutic discoveries.

Following the completion of the Human Genome and HapMap projects [32,35], thousands of Genome-Wide Association Studies (GWAS) have been performed to analyze genomic metadata in search of associations between common variants and disease. Human disorders are usually not as straightforward as sickle cell anemia or Huntington's disease, where one point mutation causes the disease phenotype [36,37]. For this reason, a GWAS assays hundreds of thousands of SNPs to identify significant associations with the phenotype of

interest. These phenotype-first studies are non-candidate-driven and therefore unbiased to a *priori* expectations. The first successful GWAS identified a SNP located within an intron of the gene for complement factor H that is associated with age-related macular degeneration [38]. Other GWAS projects have identified SNPs correlated with heart disease [39], HIV susceptibility [40], and resistance to hepatitis C treatment [41]. GWAS databases such as NCBI dbGaP and the NHGRI GWAS Catalog have been curated [42,43], and to date thousands of disease associated SNPs have been reported [44].

Finally, the enormous amount of sequencing data necessitates the need for bioinformatics tools. The Gene Ontology Project (GO) is a bioinformatics initiative developed to annotate gene sequences with consistent language: (1) the cellular component, which describes parts of the cell (*i.e.* the biological matter); (2) the molecular function, which describes events at the molecular level such as binding and catalytic activity; (3) the biological function, which describes organized events that have a defined beginning and end that describe a larger process, such as metabolism and cell division [45].

Phenomics and functional analysis

The term phenomics describes the characterization of mutation on an organism-wide scale [46]. Because phenomics projects are inherently large and require an extensive collection of mutant strains, bacteria represent a nearly ideal system of study because of their simple and tractable genomes and ease of cultivation. To construct large mutant libraries, an organism may be exposed to DNA-damaging agents such as UV radiation or reactive chemicals. Phenotypic screens following UV bombardment have been widely conducted [47–49], but this method is indiscriminate and typically result in multigene mutations. DNA mutagens have largely been replaced by transposons, which disrupt functional elements by incorporating a short piece of foreign DNA into the host chromosome. Transposons are generally less lethal, can be recovered to determine insertion sites, and some bacterial transposons carry selectable

markers. Tn5 transposon insertion-mutations have been successful in studies of virulence in *Bordetella pertussis* [50], identification of auxotrophic *Escherichia coli* [51], and bioluminescence of *Vibrio fischeri* [52]. In addition to bacterial studies, transposons have been used to construct genetically modified lines of *D. melanogaster* (*P* elements) [53–55] and *C. elegans* (*Mariner*) [56,57]. Transposon screens generate a large number of mutants; near-saturation mutant libraries have been constructed in the budding yeast *Saccharomyces cerevisiae* [58], pathogenic bacterium *Pseudomonas aeruginosa* [59,60], and the flowering plant *Arabidopsis thaliana* [61], but are generally limited by specificity and the size of the genome, even in overrepresented experiments. Site-directed mutagenesis is a method that inactivates target genes; a fragment of the target gene is PCR amplified, ligated into a double stranded plasmid, and introduced into the host chromosome via homologous recombination, resulting in complete loss-of-function of the gene product (**Fig. 2.2**). While less scalable than random mutagenesis, site-directed mutation allows the systematic disruption of single targets, which is necessary to confirm essential genes. Other methods such as RNA interference (RNAi) elucidate function by knocking down gene expression post-transcriptionally [62].

The transcriptome is the sum total of mRNAs expressed under a given environmental condition. The ability to quantify global transcript levels allows researchers to conceive of expression as a phenotype, and transcriptional events under various environmental or development conditions may lend insight to gene function. Presumably, a gene involved in a particular biological process is expressed to a higher degree during that process. Nucleic acid based microarrays have been developed to quantify mRNA levels in a high-throughput manner [63,64]. Fluorophore labeled targets hybridize to complementary probes on a microarray chip, and the signal is measured digitally to quantify relative abundance of molecules bound to each probe. Microarray technology has been instrumental in understanding the molecular basis of various cancers [65–67]. More recently, RNA sequencing by synthesis has emerged as a more precise tool to quantify expression; RNA-Seq has the advantages of being able to identify novel

genes, fusion transcripts, and sequence variants without high-intensity signal saturation. The promise of these technologies is that we will be able to group genes into functional pathways based on a similar temporal or developmentally regulated patterns of expression.

Similar to how DNA microarrays provide a high-throughput assay to quantify gene expression, a phenotype microarray (PM) allow researchers to test numerous growth phenotypes simultaneously. Unlike qualitative assays that interpret phenotypic traits as plus or minus, PM plates are scored quantitatively via colorimetric changes (indicator dyes) that are measured with commercial software. PM plates may contain hundreds of chemicals – carbon and nitrogen sources, ions, hormones, at various concentrations – to observe cellular respiration and metabolism under different environmental conditions [68]. Mutant libraries grown on PM plates have elucidated gene function [69–72], metabolic variations among evolved lines [73], pathogenicity [74,75], and culture conditions that trigger morphological and developmental changes [76].

Antibiotic susceptibility testing is a clinical method of phenotypic profiling. In these experiments, an inoculum of bacteria is overlaid with an antibiotic impregnated disc. As the antibiotic diffuses, the appearance of a visible zone of clearing around the disc indicates the susceptibility of the bacterium. These results are scored qualitatively as either susceptible, intermediate, or resistant by measuring the diameter of the zones, or quantitatively by titrating dilutions of the antibiotic into broth cultures to determine the minimum inhibitory concentration (MIC) that prevent visible growth [77,78]. Antibiotic resistance of the biofilm forming bacteria *Staphylococcus aureus* and *P. aeruginosa* has been extensively studied in an effort to understand the genetics of resistance [79–81].

Using these techniques, *E. coli* strain K-12 has been the subject of numerous mutation and phenotypic screens. As of 2009, the Keio Collection at the Nara Institute of Science and Technology (Japan) contain nearly 4,000 thousand single gene deletion mutants, accounting for ~93% of the K-12 genome [82,83]. Phenotypic profiling of this collection has identified growth

defects, provided putative annotations for previously uncharacterized genes, and revealed the functional relationships between genes and the genomic organization of *E. coli* [84,85]. The sheer number of G2P databases renders K-12 one of the best-known model organisms [86,87].

2.3 Evolutionary approaches to G2P

An important consideration in biology is how phenotypes are shaped by evolution; examining the genotypes of two closely related organisms may reveal genes that correlate with phenotypic variation (**Fig. 2.1b**). For example, a sequence comparison of *Bacillus anthracis* and *Bacillus cereus*, two bacteria that are widely considered to be the same species, has provided some evidence for the observed differences in pathology [88,89]. Similarly, a genetic comparison of two closely related enterobacteria revealed a diversity of molecular mechanisms relating to pathogenicity [90].

Recently, the paradigm of evolutionary genetics has extended beyond investigating the differences between extant natural organisms to investigating how model organisms evolve in the laboratory [91–96]. Long-term evolution experiments (LTEE) are carried out under controlled environments to explore the evolutionary dynamics with respect to gene function and adaptation. The utility of this experiment is that researchers can implement predetermined selective pressures to force an evolutionary response, and resequence the evolved lineage to correlate mutations and the adaptive phenotype. LTEEs track evolution in real time; bacterial systems have been exploited because their rapid generational times allow evolution to be observed on a reasonably short timeline.

The most famous example of an LTEE was started in 1988 by Richard Lenski [97]. Twelve parallel cultures of *E. coli* Bc251 have been growing aerobically under identical environmental conditions and limited to six generations per day in glucose-limited media for nearly 30 years. Samples are archived every 500 generations to create a “frozen fossil record” of Bc251 evolution for each independently evolving population [97]. Each culture is sequenced

at milestone generations to catalog genomic divergence [98–102]. At present, the twelve founding cultures are separated by over 64,000 generations of independent evolution, a timeline sufficiently long that every possible spontaneous point mutation could have occurred several times [103]. The results of this study have pioneered the sub-field of experimental evolution; first was the discovery of two mutator genes [104]. Genomic instability was observed in several cultures, and in each case resulted from spontaneous mutations in at least one mutator gene. Second is the parallel evolution of cell size and generation time across all cultures, a phenotype that is sustained through 60,000 generations, which may suggest perpetual and unbound evolution [105,106]. Third is the discovery of a novel phenotype; while all twelve cultures experienced similar evolutionary trajectories, one culture evolved the ability to metabolize citrate present in the growth media under aerobic conditions, a phenotype that normally differentiates *E. coli* from other *Enterobacteriaceae*. Trait acquisition was mapped back to mutations in the transporters CitT and DctA, and the upstream promoter *rnk* [103,107]. This finding is significant with respect to historical contingency and evolution; the accumulation of many neutral potentiating mutations prepared the organism to achieve a fitness gain from a future mutation. The longstanding impact of this ongoing study improves our understanding of the evolutionary mechanisms relating to mutation rates and fitness trajectories [104,106,108].

Experimental evolution studies have also been carried out in eukaryotic systems. *Drosophila melanogaster* has been extensively studied to understand the mechanisms of selection and adaptation [109,110]. One prominent example is the study by Haddad *et al.* that placed replicate lineages of flies into environments of low oxygen content for 200 generations. Genome resequencing revealed the genetics of tolerance, and to date 188 genes have been correlated with apoxia tolerance [111]. A second example is the ongoing “high runner (HR)” selective breeding experiment in mice. There have been 65 generations in this evolutionary lineage of mice that demonstrate increased aerobic capacity and overall endurance in a running-wheel experiment [112]. The genetics of this adaption are not yet understood.

2.4 Obstacles for G2P studies

Despite much being already known regarding the functions of individual genes, even in relatively simply bacterial systems there are still thousands genes without a known function. As the number of sequenced bacteria species quickly approaches 100,000 [113], it is doubtful that more sequencing will bring us any closer to resolving function via sequence homology. Gene inactivation is the now to most feasible tool to elucidate function.

Phenotypic constraints

Mutations in nonessential genes typically result in a small number of distinct phenotypes. Studies that have characterized the near-saturation mutant libraries of *P. aeruginosa* and *S. cerevisiae* report that only ~15% of mutant strains display changes in growth rate [59,114]. Whole genome RNAi knockdown experiments in *C. elegans* and *D. melanogaster* and T-DNA insertions in *Arabidopsis thaliana* yielded a similar percentage for defects relating to growth and development [115–117]. While many studies focus on simple characteristics such as growth rate and metabolism, others have focused on multicellular traits such as biofilm density in *P. aeruginosa* [118] and fruiting body formation in *Dictyostelium discoideum* [119]. However, the number of mutant strains with distinguishable phenotypes remained <10%.

One plausible explanation for the lack of obvious changes in phenotype is that too little information regarding the organism's phenotype is known to serve as a baseline to compare wild type and mutant phenotypes. In many cases, the assays that measure mutant phenotypes are qualitative or semi-quantitative, and the interpretation of these data vary between laboratories [120]. Furthermore, phenotypic defects are measured by the magnitude of change from the wild type strain, and studies that set arbitrary cutoffs to detect the most severe defects and may neglect more subtle defects. It has become necessary to refine and expand the current characterization techniques to better define the intersection of wild type and mutant phenotype.

A global view of genotype: strain evolution and the epistatic genome

The examples of experimental evolution described in **2.3** investigate natural selection and Darwinian adaptation by deliberately forcing evolution under preselected and controlled conditions. These fundamental concepts of these experiments can be applied to organisms under real laboratory conditions; the founding strain of a widely studied bacterial system has a documented origin and a sequenced genome, but is usually represented by several laboratory “sublines”. Each subline is an independent population that is capable of evolving across multiple organizational levels. Because they are in isolation, each subline can adopt different solutions to the selective pressures in a laboratory setting. Subline evolution that results in a sudden and drastic phenotypic shift would likely be observed, however, if the subline displays little or no change in phenotype it would go unnoticed. Over time, mutations may accumulate despite the wild type strain’s phenotype remaining relatively unchanged. This is the difference between proving evolution can happen in a model system, and proving that it has.

Several studies have begun to address the topic of subline evolution. Schacherer *et al.* characterized single nucleotide variation for seven experimentally relevant laboratory strains of *S. cerevisiae* using Affymetrix yeast tiling assays, and report as much as 0.36% genomic variation from the reference [121]. In *P. aeruginosa*, PAO1 is the major reference strain and has been distributed to laboratories worldwide. PAO1 has also been the subject of numerous mutation screens [123]; a comparison of three PAO1 sublines identified a 2.2Mb inversion, a 12Kb duplication, and numerous SNPs and deletions in protein coding regions [124,125]. Differential pathological potential has also been reported among these three PAO1 sublines [124]. These examples clearly demonstrate subline-specific genomic architecture. Without a detailed examination of how laboratory strains differ with respect to both genotype and baseline phenotype, the reproducibility of G2P studies may be jeopardized. Genome divergence following subline propagation presents a real possibility of conflicting phenotypes, particularly in the case of epistasis from multiple mutations.

The term epistasis was introduced by William Bateson to resolve the discrepancy between predicted segregation ratios and the observed phenotypic outcomes of dihybrid crosses [126]. Epistasis is defined as the interaction of mutations in two or more genes that produce a phenotype that is unequal to the sum of the individual mutations. In other words, one mutation alters the phenotypic effect of the other. Positive epistasis refers to a phenotype that is more fit than would be expected from the two mutations individually, through either enhancing the individual beneficial effects or alleviating the deleterious effect of one mutation (**Fig. 2.3**). For example, mathematical models predict a high fitness cost (e.g. reduced growth rate) of drug resistance-conferring mutations [127–129]. However, multiple studies report fitness gains (positive epistasis) from compensatory mutations in drug resistant strains. In *P. aeruginosa*, Ward *et al.* report a reduction in the cost of acquiring streptomycin resistance (StrepR) in genetic backgrounds that carry parallel resistance to rifampicin (RifS) [130]. Similarly, in *E. coli*, Trindade *et al.* report compensated fitness costs of StrepR when RifS is introduced concurrently [131]. Other studies have reported positive epistasis in bacteria that harbor multiple resistance conferring plasmids in the absence of the corresponding selective pressure [132]. Conversely, negative epistasis refers to a phenotype that is less fit than expected, where the net benefit of two individual mutations is reduced, or two deleterious mutations result in a greater than additive effect. Negative epistasis has been reported between the malaria-protective mutations in α and β globin genes in humans [133].

Epistatic interactions are also used to evaluate the function and organization of complex networks; epistasis is commonly studied between pairs or sets of relevant genes. In *S. cerevisiae*, nearly 80% of the genome is nonessential, at least in part because of genetic buffering [114,134,135]. A study by Tong *et al.* constructed pairwise double mutants for eight genes against the nonessential mutant library to reveal synthetic lethal phenotypes [136]. The results of this study generated a network of 291 interactions among 204 genes, some of which were previously uncharacterized. Larger studies have investigated hundreds of query strains

against the mutant library to further elucidate the *S. cerevisiae* interactome [137] In *C. elegans*, analyses of epistatic interactions have successfully ordered genetic pathways for traits such as sex determination [138], vulva development [139], and entry into dauer [140].

2.5 *Myxococcus xanthus* as a model organism

The Myxobacteria are members of the δ -proteobacteria and display a cooperative and coordinated social behavior that is more characteristic of a eukaryote. Multicellular behaviors include swarming/group predation, self-organization (aggregation), and cellular morphogenesis (sporulation). Because of the complex life cycle yet simple prokaryotic genome, the Myxobacteria are an attractive system to study the genetics and signaling associated with multicellular phenotypes.

Myxococcus xanthus is a gram-negative soil dwelling bacterium that displays a complex multicellular phenotype when several million cells are spotted as a dense swarm on an agar surface. If the agar is nutrient rich, the swarm will expand out from the point of inoculation through the coordination of two motility systems, adventurous (A-motility) and social (S-motility), in a process called swarming [141] (**Fig. 2.4**). Alternatively, if the agar contains no nutrients, the starving swarm will appear to contract in a process called development, where cells first move to form aggregates of approximately 1×10^5 cells each [142], and then a subset of cells within each aggregate differentiate into dormant and environmentally resistant spores that germinate when nutrients become available. *M. xanthus* myxospores are environmentally resistant, able to withstand temperatures and other stresses that are lethal to vegetative cells. Explained in this way, the life cycle of *M. xanthus* is divided into two distinct halves, swarming and development, both of which can be described by measuring different phenotypic traits, such as the rate at which a swarm expands on nutrient agar, or the number of aggregates and spores that form during development (**Fig. 2.5**). The phenotype of the wild type *M. xanthus* has always been described as a range of assay results that measure traits like these.

Genetics of major biological processes

Vegetative growth and swarming: Since *M. xanthus* contains no physical mechanism for propulsion in liquid, the entire life cycle of *M. xanthus* depends on solid substrate motility. Motility is a coordinated swarm, with some isolated individuals found along the edge of the swarm. Individual cells of an *M. xanthus* swarm glide along the direction of their long axis, occasionally stopping and reversing their direction of movement [143]. *M. xanthus* is a predatory bacterium, and coordinated motility allows the swarm to collectively prey on nearby microorganisms [144,145]. Gliding motility demonstrated by *M. xanthus* is regulated using the A- and S-motility engines [146]. Coordination of these engines is essential for directed movement, and reversing the intracellular localization of motility proteins allows individual *M. xanthus* cells to change directions.

S-motility is characterized by type IV pili (T4P) mediated movement and extracellular fibrils, and is similar to twitching motility observed in *P. aeruginosa* [147]. Forward locomotion of individual cells is achieved by the extension of pili from the leading pole, attachment to the solid substrate, followed by retraction of the pili thus pulling the cell forward towards the site of attachment [147]. T4P can extend up to 5µm from the leading pole, and retraction of a single pilus generates sufficient force to pull an individual cell forward.

PilA is the major subunit for *M. xanthus* T4P and is localized at the leading pole [148]. PilA is a 23-kDa protein monomer that is anchored to the inner membrane. PilA monomers are processed by PilD peptidases and polymerized into a growing pilus by PilB ATPases [149]. Polymerized PilA is secreted through PilQ outer membrane secretory channels [150]. For retraction, PilA disassembly is mediated by PilT, a PilB homolog [149]. Presumably, ATP bound to PilB is hydrolyzed to insert PilA monomers to the growing pilus from an inner membrane reservoir. Conversely, PilT catalyzes disassembly by returning monomers from the elongated pilus back to the inner membrane. While the other S-motility components are localized only at the leading pole, PilQ is localized to both poles to facilitate cell reversal; S-motility components

reverse their localizations from the leading pole to the lagging pole (now the leading pole) to allow for bidirectional movement. Studies investigating the *frz* system in *M. xanthus* have shown cells move in one direction only a for a few minutes before reversing the cellular localization of their motility machinery [151–153]. A second important extracellular appendage found on *M. xanthus* cells are fibrils. These branching, filamentous structures are 30-50nm in diameter [154–156] and are composed of nearly equal parts proteins and carbohydrates. Fibrils are considered essential for contact mediated cell-cell interaction, linking cells to the substrate as a collective unit [155,157,158]. Deficiencies in fibril production, as demonstrated by *dsp* and *dif* mutants, abolish S-motility [157,159–161]. Together, T4P and fibrils promote cohesion (cell-to-cell) and adhesion (cell-to-substrate).

The mechanism of A-motility in *M. xanthus* is not as well understood because it is not dependent on external structures. One model suggests the use of a “slime gun” at the lagging pole, a similar feature found in gliding Cyanobacteria [162]. In this model, an extruder systems acts as a thrusting motor by secreting slime to propel the cell forward [163], while also serving as a slime track for neighboring cells [164]. However, this model remains largely speculative because of the lack of genetic evidence. A second model of A-motility is the use of focal adhesion complexes distributed along the cell body. Identified through transposon screens, AglZ has been shown to be required for A-motility [165]. AglZ accumulates at the leading pole and is dispersed at regular intervals along the cell body. AglZ assemblies remain fixed in position relative to the substrate [166]. Presumably, a moving cell spirals around the AglZ focal adhesions until the adhesions reach the lagging pole, where they are disassembled. A motility mutants are grouped into two classes; motor proteins (Agl, adventurous gliding) and stimulation/protein exchange between contacting cells (Cgl, conditional gliding). Early work has demonstrated that *cgI* mutants are also defective for A-motility [167]. Recent work has shown that A-motility is restored by diluting *cgI* mutant strains with wild type [168]. The proposed mechanism is outer membrane exchange (OME); OME is the bidirectional exchange of

lipopolysaccharide and is thought to repair membrane damage caused by environmental hardship, or act as a mechanism to exchange toxic material to rival cells [169]. Presumably, *cgl* gene products are donated from wild type host cells via OME to restore A-motility in *cgl* mutant cells.

Nutrient limitation and development: One of the most distinguishing features of the myxobacteria is their response to nutrient deprivation. When organic material is exhausted, a swarm of *M. xanthus* cells cease to swarm outward, instead condensing into mound like structures called aggregates. Pre-aggregation (*i.e.* transcription of early development genes) occurs approximately two hours post starvation, while aggregation and maturation into three-dimensional stalk-like structures known as fruiting bodies occurs within 24 hours of starvation [150]. Within fruiting bodies are metabolic dormant and environmentally resistant myxospores that have the capacity to germinate when nutrients become available again. Sporulation represents the end of the cell cycle.

Developing cells assume different fates; autolysis, differentiation into peripheral rods, or differentiation into myxospores. Previous studies estimate nearly 80% of developing cells undergo autolysis, presumably to release nutrients that are cannibalized by the remainder of the population [170,171]. Peripheral rods represent approximately 5% of cells and are located outside of aggregate mounds. Peripheral rods are structurally identical to vegetative cells, but have an altered transcription profile and do not aggregate, sporulate, or divide [172]. Thus, peripheral rods represent the first category of differentiated cell types in *M. xanthus* development and are proposed to function either as nutrient sensors or in defense [172]. The remaining 1-15% of the starting population differentiate from rod shaped cells into ~1µm diameter quiescent myxospores that are resistant to environmental stress [173]. Tremendous progress at the molecular level has been made to identify the intracellular signals responsible for induction of development. Currently, there is genetic evidence of at least five signals,

annotated A- through E-signal, that are necessary to initiate and culminate development [150]. Mutant strains deficient in any of these signals show extreme developmental defects and an inability to fruit or sporulation. Previous work has shown that these phenotypic defects can be rescued by co-development with wild type [174,175]; however, only A- and C-signals have been extensively studied.

Starving cells can either slow their growth proportional to the level of available nutrients, or engage the development program. Fruiting body morphogenesis and sporulation require the synthesis of at least 30 proteins [176], and therefore development must proceed before nutrients are depleted entirely. Starving cells synthesize the alarmone guanosine penta-phosphate ((p)ppGpp) as an indicator of starvation in a RelA-dependent manner that is similar to that of *E. coli* [177,178]. Briefly, starvation of essential amino acids (leucine, isoleucine, valine) result in uncharged cognate tRNAs. Free tRNAs encountered at the ribosome halt translation, causing RelA to synthesize (p)ppGpp [179,180]. Accumulation of (p)ppGpp halts DNA and RNA synthesis and increases proteolysis [181]. (p)ppGpp above the threshold concentration is necessary and sufficient to induce development. *relA* mutants do not produce (p)ppGpp, and thus are defective for aggregation and sporulation [179]. Furthermore, ectopic expression of *relA* in *M. xanthus* initiates development, even in the presence of abundant nutrients [182].

A-signal is a quorum signal that functions in the pre-aggregation phase of development as an indicator of cell density [183]. A-signal can be recovered from conditioned media and separated into two fractions, heat stable and heat labile. The heat stable fraction contains roughly equal portions of amino acids and small peptides, and serves as a chemosensory signal [184]. The heat labile fraction contains two proteolytic enzymes that hydrolyze cell surface proteins [185]. The current model of A-signal is that RelA synthesizes (p)ppGpp in response to nutrient deprivation, which serves as a signal of starvation. Proteases then degrade cell surface proteins causing amino acids and peptides to accumulate in the extracellular space. A-signal

acts as an indicator of density because *M. xanthus* cells themselves are the substrate for proteolysis; the concentration of the heat stable fraction is directly proportional to the number of cells in the population. A-signal must accumulate to a threshold concentration to ensure that there are enough cells to enter the aggregation phase of development. While an A-signal receptor has not been identified, the *sas* three-component regulatory system is suggested to function in the A-signal transduction pathway [186,187]. However, five components of the A-signaling system have been identified; AsgA/D (hybrid HPKs), AsgB (putative DNA-binding protein), AsgC (sigma factor), and AsgE (putative amidohydrolase) [188–192].

Following the events of early development, the morphological events of late development (aggregation and sporulation) are governed by C-signal, a membrane associated signal that is exchanged via cell-cell contacts [193]. C-signal is a 25-kDa (p25) protein encoded by *csgA* that accumulates on the outer membrane [193]. p25 is cleaved into the 17-kDa protein (p17) active form by PopC six hours after nutrient depletion [181], and therefore is only expressed by starving cells. Mutations in *csgA* abolish fruiting body morphogenesis and sporulation. However, these events are restored by co-development with *csgA*⁺ cells or purified exogenous *csgA* gene product; low levels of C-signal restore aggregation and C-signal dependent gene expression, while higher levels restore sporulation [194]. Incremental addition of C-signal beyond a threshold concentration induces sporulation without the prerequisite aggregate formation [194]. It is proposed that, similar to A-signal, C-signal acts in a threshold-dependent manner to regulate developmental events sequentially (*i.e.* cells do not sporulate before aggregating) [193,195,196]. Presumably, the local density within an aggregate increases the efficiency of end-to-end cell contacts, and therefore increases C-signal transmission. Unlike endospore formation in *Bacillus* spp. which undergo binary fission in an asocial manner, the positive feedback mechanism of C-signal exchange in *M. xanthus* ensures that a sufficiently large number of cells sporulate and emerge from an aggregate to collectively feed.

Origin of wild type *M. xanthus* DK1622

The *M. xanthus* founding strain FB (DK100) was obtained from the Stanier Collection at the University of California, Berkeley in 1960 [197] (**Fig. 2.6**). Two distinct colony varieties were reported for FB; yellow and tan swarmers (YS/TS) [197]. DK101 is a derivative of DK100 that harbors a disruption in *pilQ1* [198,199]. DK101 was isolated because the spontaneous mutation retained fruiting body formation and allowed dispersed cell growth in liquid media [198]. The disruption of *pilQ1* abolishes S-motility [198]. DK320 is a derivative of DK101 generated by exposure to UV radiation [141]. *aglB1* is disrupted in DK320, rendering the strain defective for motility. *pilq1* was restored in DK320 via transduction from a YS donor to create DK1217 [198]. DK1217 was the recipient of second transduction from YS for *aglB1*, yielding the fully motile strain DK1622 [198]. DK1622 is commonly used as wild type because it reliably forms symmetric fruiting bodies, ripples (a predatory behavior), and develops in submerged culture [200]. DK1622 has been one of two *de facto* wild type strains (DZ2 is the other) since its initial isolation in 1979. DK1622 and DZ2 are phenotypically similar, capable of self-organization and differentiation, but differ by a 220Kb deletion in DK1622 [201].

2.6 Summary

High-throughput functional annotation of bacterial genomes is currently at a crossroad; some mutants have no discernable phenotype, while others do not reflect the anticipated outcome based on previous studies. There are two likely explanations for this disconnect: (1) phenotypes are present, but undiscovered in the current assays, and (2) phenotypes are influenced by variations in the genetic background. As the most severe phenotypes – the lowest hanging fruit – have been well studied, a more thorough characterization of mutant phenotypes is needed to identify the most subtle variations. Lastly, taking advantage of the decreasing costs of genome sequencing will resolve the genetics of discordant phenotypes.

CHAPTER 2

FIGURES & TABLES

Table 2.1 Notable sequenced organisms.

Organism (strain)	Type	Relevance	# Chr	# Genes	# Bp	Year	Ref.
<i>E. coli</i> (K-12)	Bacteria	Model organism	1	4,331	4.6 Mb	1997	Blattner, 1997
<i>M. tuberculosis</i> (H37Rv)	Bacteria	Pathogen (tuberculosis)	1	3,999	4.4 Mb	1998	Cole, 1998
<i>M. xanthus</i> (DK1622)	Bacteria	Model organism	1	7,331	9.1 Mb	2005	Goldman, 2006
<i>P. falciparum</i> (3D7)	Protist	Pathogen (malaria)	14	5,268	22.9 Mb	2002	Gardner, 2002
<i>S. cerevisiae</i> (S288C)	Fungi	Model organism	16	6,294	12.1 Mb	1996	Goffeau, 1996
<i>A. thaliana</i>	Plant	Model organism	5	25,498	119 Mb	2000	<i>Arabidopsis</i> Genome Initiative, 2006
<i>C. elegans</i> (Bristol N2)	Animal	Model organism	6	20,470	100 Mb	1998	<i>C. elegans</i> Sequencing Consortium, 1998
<i>D. melanogaster</i>	Animal	Model organism	4	15,682	123 Mb	2000	Adams, 2000
<i>H. sapiens</i>	Animal		23	18,826	3.2 Gb	2001, 2006	Venter, 2001; McPherson, 2001

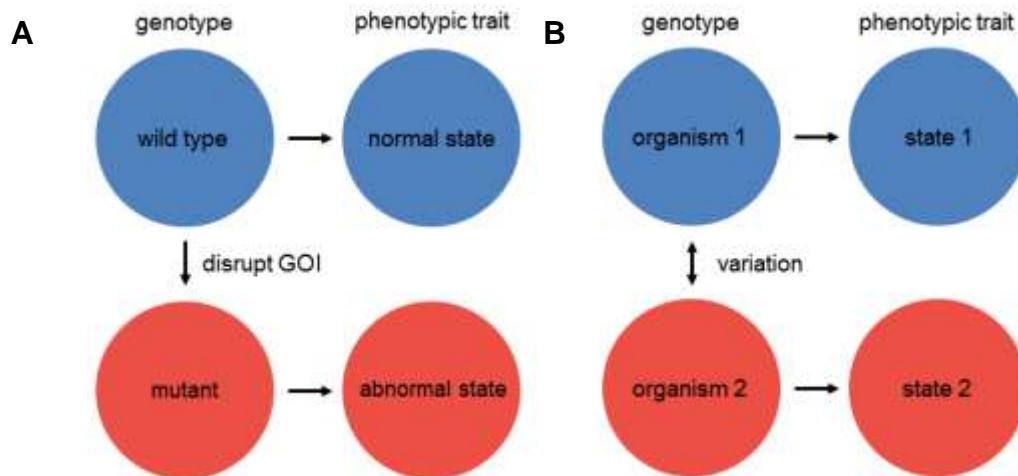


Fig. 2.1 Experimental and evolutionary approaches to phenotypic profiling. **(A)** Schematic of a forward genetics experimental approach. A gene of interest (GOI) is disrupted and the resulting abnormal phenotype is scored and compared to wild type. **(B)** Schematic of evolutionary approach. The genomes of two closely related organisms are compared to reveal genetic and phenotypic variation.

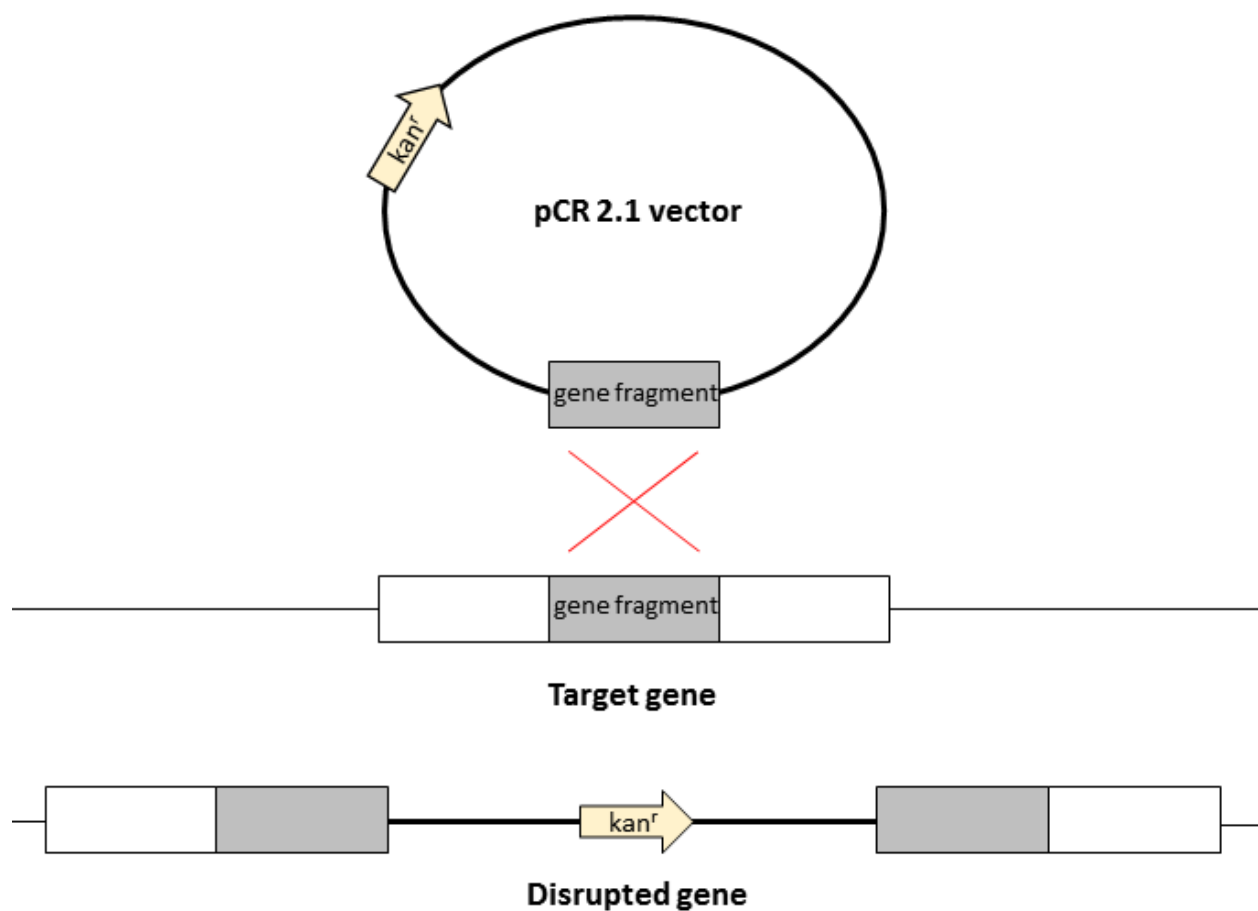


Fig. 2.2 Site-directed mutagenesis and homologous recombination. A fragment of the target gene (grey box) is ligated into a vector containing antibiotic resistance (yellow arrow). The plasmid is introduced into the host cell and recombination involving the crossover (red lines) between the plasmid and host chromosome disrupts the target gene.

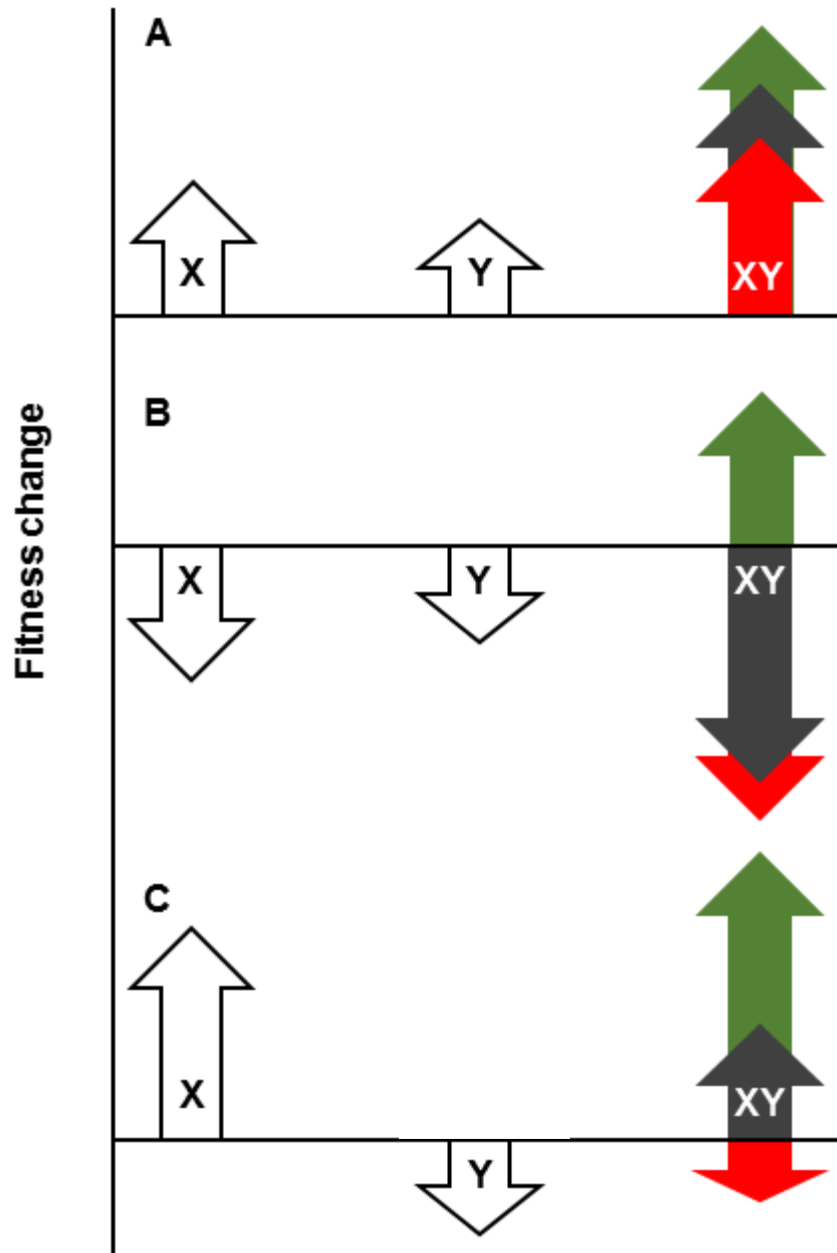


Fig. 2.3 Schematic of positive and negative epistasis on fitness. The phenotypic impact (fitness) of two individual mutations (X and Y) can have epistatic interactions. For individual mutations (X or Y, white arrows) the magnitude of change in fitness relative to wild type is represented by the direction and size of the arrow. For double mutants (XY), the expected change in fitness for non-interacting double mutants is shown in gray (the summation of the individual X and Y mutations). Positive epistasis (green arrow) exceeds the expected change in fitness from each individual mutation; negative epistasis (red arrow) results in a change in fitness that is less than the sum of the individual mutations, or a negative effect that is greater than the sum of two deleterious mutation. (A) Potential net effect of two beneficial mutations. (B) Potential net effect of two deleterious mutations. (C) Potential net effect of two mutations, one beneficial and one deleterious. Adapted from Ostman et al. (2012).

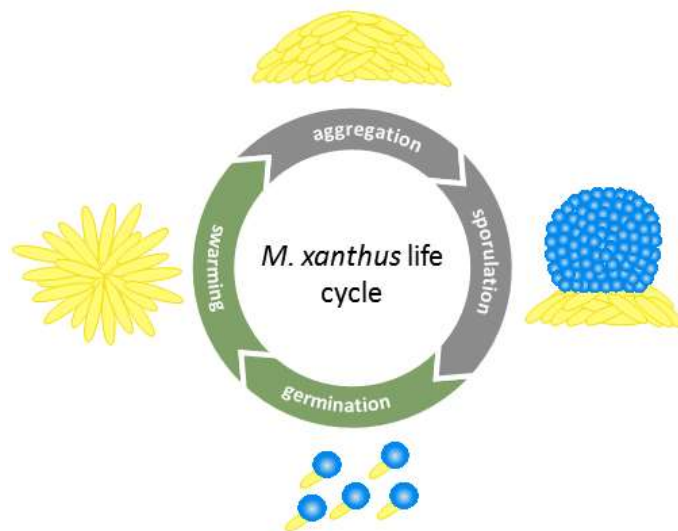


Fig. 2.4 Life cycle of *M. xanthus* (simplified). On a solid substrate with abundant nutrients, a group of *M. xanthus* cells feed, divide, and expand radially (swarming). On a solid surface that is nutrient depleted, swarming cells aggregate into mounds and three-dimension fruiting body structures (aggregation). Prolonged starvation induces a subset of the rod-shaped cells within fruiting bodies to differentiate into spherical spores that are metabolically dormant and resistant to environmental conditions (sporulation). When nutrients return, spores are released to germinate into vegetative rod shaped cells and re-enter the life cycle (germination). The life cycle of *M. xanthus* is divided into behaviors under nutrient rich conditions (green arrows) and nutrient deprived conditions (gray arrows).

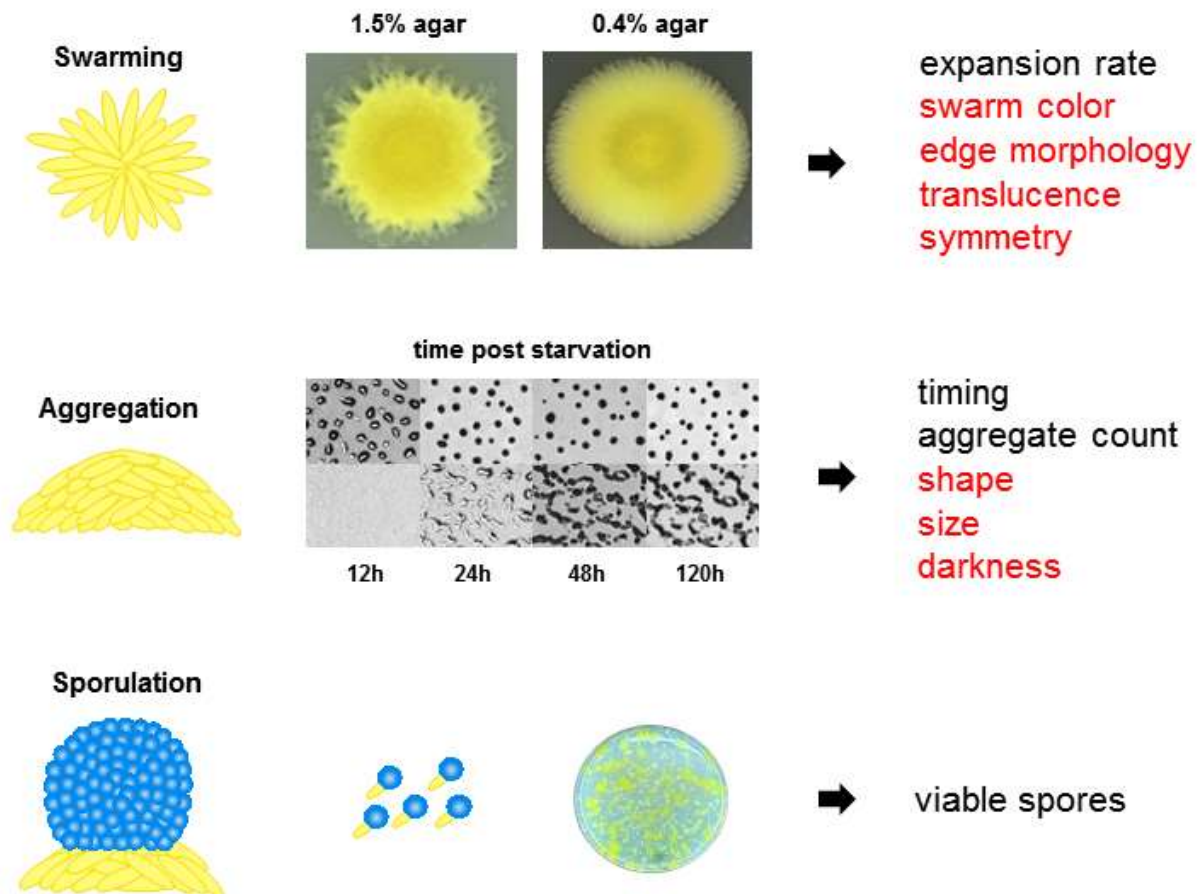


Fig. 2.5 Metrics used to describe the phenotype of *M. xanthus*. There are several qualitative and quantitative features of the life cycle at each stage (swarming, aggregation, and sporulation). Representative images at each corresponding stage (vegetative growth, starvation, germination) are shown. Quantitative (black) and qualitative (red) metrics used to describe phenotype are listed for each stage.

Roger Y. Stanier collection (University of California, Berkeley)

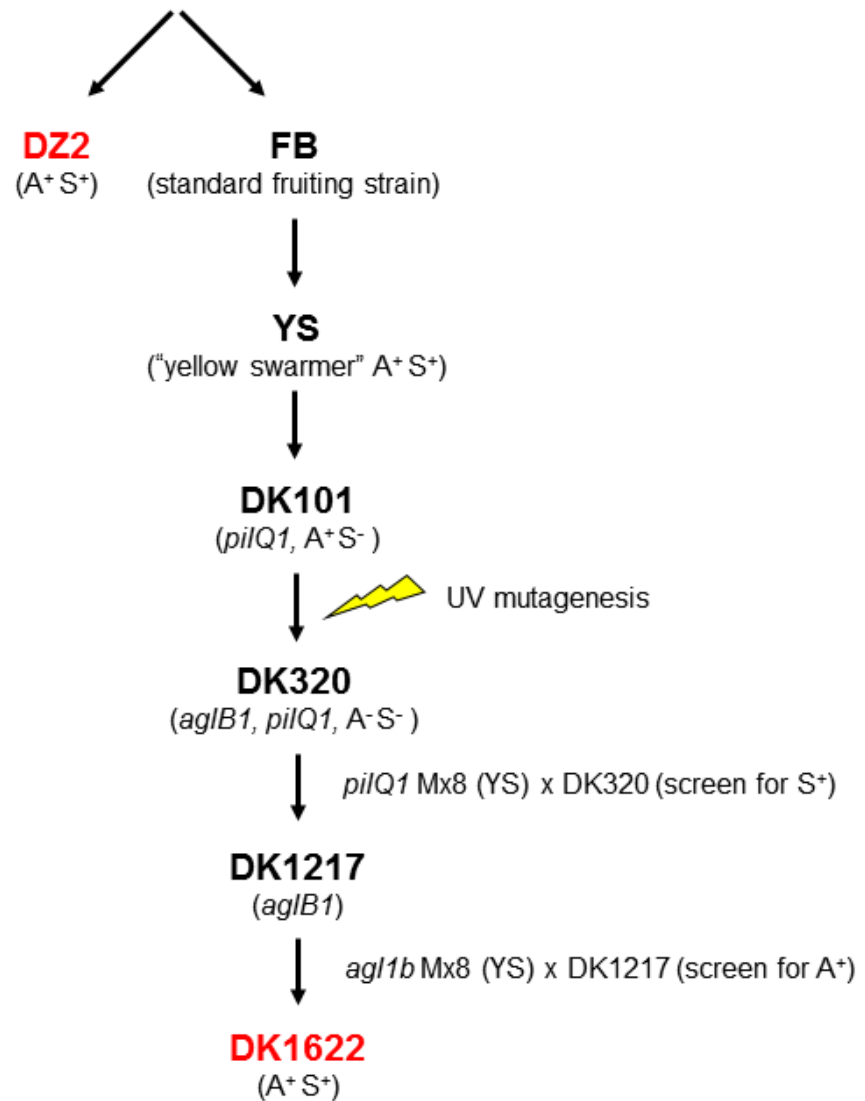


Fig. 2.6 Origin of wild type *M. xanthus*. Two strains, FB and DZ2 were acquired from the Stanier Collection. Strain FB has been the subject of several mutation and phenotypic screens to produce the fully motile (A⁺S⁺) strain DK1622. Wild type strains used by the research community are shown in red.

CHAPTER 3

PHENOTYPIC PROFILING OF A LARGE MUTANT LIBRARY

The material presented in this chapter has been published in:

Yan, J., **Bradley, M. D.**, Friedman, J. & Welch, R. D. Phenotypic profiling of ABC transporter coding genes in *Myxococcus xanthus*. *Front. Microbiol.* 5, 352 (2014).

3.1 Project summary

The ATP-binding cassette (ABC) transporters superfamily is large and diverse, present in both prokaryotic and eukaryotic systems. Consisting of two distinct domains, these ATP-dependent transmembrane proteins recognize and translocate nutrients, proteins, and small molecules when undergoing a conformational change. The mechanism of import and export maintains a variety of cellular processes including signal transduction, protein secretion and multi-drug resistance.

Using multiple bioinformatics tools, we have identified 192 Open Reading Frames (ORF) in *Myxococcus xanthus* that code for ABC transporters. We compiled an insertion-disruption mutant library containing each of the ORFs and characterized the resulting phenotype using three independent assays to determine cell motility, aggregation rate, and sporulation efficiency against wild type DK1622. The aim of this project is to examine the relationship between ORF sequence homology, gene function, and phenotype. To identify any correlations, we analyzed all of the *M. xanthus* ABC transporters as a group, to identify any abnormal phenotypes that might be overlooked if mutants were characterized individually.

3.2 Materials and methods

Annotation of ABC transporters: The sequence of *M. xanthus* was obtained from GenBank ([http:// www.ncbi.nlm.nih.gov/GenBank/](http://www.ncbi.nlm.nih.gov/GenBank/)) with the accession number NC_008095.1. Each predicted ORF in the genome was annotated using multiple databases. ABC transporter-associated genes in the *M. xanthus* genome were reviewed and identified mostly using the databases pfam [202-204] and COG [205]. Additional tools such as BLAST [206], InterPro [207], GenBank [208] and transmembrane prediction server TMHMM Server v. 2.0 [209] ([http://www.cbs.dtu.dk/services/ TMHMM/](http://www.cbs.dtu.dk/services/TMHMM/)) were used to assist in the selection. A manual curation was adopted to complete the annotation. Briefly, an ORF was deemed an ABC

transporter when at least two databases identified it as such. In order to ensure that all putative ABC transporter-associated ORFs were identified in the genome, a list of pfam accession IDs associated with these ORFs was compiled and used to search the rest of the genome. ORFs annotated as “hypothetical” that were located in the same operon with ABC transporter coding ORFs were manually checked using pfam and psi-blast. TMHMM was used to predict transmembrane domains in putative permeases. Using these methods, two additional ABC transporter-associated ORFs were identified, resulting in a total of 192 putative ABC transporter component coding ORFs.

Cultivation and development: *M. xanthus* strains were grown at 32°C in CTTYE broth [1.0% Casitone, 0.5% yeast extract, 10mM Tris-HCl (pH 8.0), 1mM KH₂PO₄, and 8 mM MgSO₄] or on plates containing CTTYE broth and 1.5% agar. CTTYE broth and plates were supplemented with 40µg/ml kanamycin sulfate as needed. Cells underwent development on TPM agar [10mM Tris-HCl (pH 8.0), 1mM KH₂PO₄, 8mM MgSO₄, and 1.5% agar] at 32°C for 5 days.

Mutagenesis: Primers for amplifying internal fragments of *M. xanthus* ORFs were selected using primer3 (<http://sourceforge.net/projects/primer3/>). The procedure for homologous recombination by plasmid insertion has been described previously [210]. Briefly, an internal fragment of 400-600bp was amplified using the polymerase chain reaction (PCR), and ligated into a linear plasmid pCR®2.1-TOPO (Invitrogen). Following ligation, the plasmid now includes the PCR product, and thus can be amplified in TOP10 *E. coli* cells. The plasmid was then isolated from *E. coli* and electroporated into *M. xanthus* cells (650V). The transformed plasmid was incorporated into the *M. xanthus* chromosome by homologous recombination [211], thus conferring kanamycin resistance on the cells. In order to confirm that the plasmid was successfully inserted into the desired location in the *M. xanthus* chromosome, we used PCR to

amplify across the upstream region of the target gene loci and TOPO vector, thereby generating an amplicon with size ~1.2Kb. Wild type DK1622 was used as negative control.

Phenotype assays: *M. xanthus* cells were inoculated in CTTYE broth, cultivated with vigorous agitation (300rpm) over night, and harvested at the density of $\sim 5 \times 10^8$ cells/ml. For motility assays, four spots of 2 μ l cells at a concentration of 5×10^9 cells/ml were placed on CTTYE plates containing 0.4 or 1.5% agar. Plates were incubated at 32°C for 3 days, and the diameters of colonies were then measured. For cell development, cells were washed once with TPM buffer and resuspended at a concentration of 5×10^9 cells/ml in TPM buffer. Spots of 20 μ l cell resuspension were spotted on TPM agar and incubated at 32°C for up to 5 days. The development of aggregates was observed and recorded at designated time intervals using 40 \times brightfield microscopy (Nikon) and SPOT imaging software. For the sporulation assay, cells were spotted on TPM agar and incubated at 32°C for 5 days to allow full development. Three sets of five spots were harvested and suspended in 500 μ l TPM buffer. The cells were then exposed to mild sonication (10% altitude, 10s \times 3 with 30s intervals, MISONIX, S-4000), followed by heat treatment at 50°C for 2h. Cells were then diluted to the desired concentration and plated with CTTSA [1.0% Casitone, 10mM Tris-HCl (pH 8.0), 1mM KH₂PO₄, and 8mM MgSO₄, 0.7% agar] onto CTTYE agar plates (supplemented with 40 μ g/ml kanamycin sulfate for insertion mutants). After 5 days of incubation at 32°C, viable spores germinated and grew into visible colonies, and the number of colonies was recorded and converted to the unit of cells/ml.

Image analysis of development: At each time point of development on TPM agar, we examined at least five spots of 10^8 cells. Brightfield images were taken for two of the five spots using a Nikon microscope and SPOT Insight camera (model #11.0 monochrome w/o IR) and imaging

software at 40x magnification, and images were saved as non-reduced .tiff files. A .tiff file was selected because it preserves image quality and is lossless. To avoid the effect that the edge of the spot would have on image analysis, a section of each image representing 25% of the original that did not include the spot edge was submitted for analysis. ImageJ (<http://rsbweb.nih.gov/ij/index.html>) [212] was used to analyze the features of aggregates. These images were threshold using the RenyiEntropy macro, and then manually corrected if necessary. After the area of each aggregate was selected for analysis, any aggregate that was less than 200 pixels (background noise) or that overlapped the edge of the image was excluded.

Data analysis: We analyzed the extent of phenotypic variation in the mutant strains by comparing the mean \pm *SD* across three independent replicates for each trait separately. For two of the traits we did not generate an *SD*; count was presented simply as the total number of aggregates observed, and timing, because it was recorded at five discrete intervals, was insufficient to resolve variation between replicates. To identify mutant phenotypes that are statistically distinct from wild type, we used a randomization test with 1000 iterations. This method generates *P* values regardless of the distribution of the raw data, and we adjusted the *P* values using the Benjamini, Hochberg, and Yekutieli method to control for any false discovery rate. To investigate whether phenotypic traits are correlated within mutant strains, we used a Spearman's rank correlation (both zero-order and partial correlations). We added the *SDs* from three phenotypic variables in order to exclude the possibility that a large variation is caused by large average values (e.g., large average fruiting body area will also have a larger variation). We normalized the *SD* using std/mean.

3.3 Results

Identification of ABC transporter ORFs in *M. xanthus*: We performed a sequence-level characterization of all ABC transporters in the *M. xanthus* genome. To accomplish this, we carefully examined the entire genome using a combination of pfam [202,203], COG [213] GO [214] and GenBank [208]. The results are included in **Figure 3.1**. Once we had established the number, type, and distribution of ABC transporter component ORFs, we performed insertion-disruption mutagenesis on all 192 of them. Among these were 12 that contain an ATPase domain and a TMC, and three that contain a TMC and an SBP. The rest contain only one of the three types of components, either an ATPase, a TMC, or an SBP. A total of 139 of the 192 ORFs were predicted to be coding for components of 57 complete ABC transporters, 20 importers and 37 exporters, based on our observation that they clustered within operons. The remaining 53 ORFs either form operons incomplete for an ABC transporter, or they are located alone in the genome.

Mutagenesis results and phenotypic assays: We succeeded in creating 180 ABC transporter insertion-disruption mutant strains. For the remaining 12, we made three independent attempts, and each time no viable colonies were produced. We labeled all of these 12 ABC transporter ORFs “putative” essentials, since we did not perform a standard complementation assay to confirm that they were essential. It is important to note that this number is close to previous estimations of the percent of essential genes in *M. xanthus* [215]. All of the assays we performed for this study used methods for measuring mutant phenotypes that are considered standard for *M. xanthus* laboratory research. Each assay focused on either swarming or development, and we examined a total of eight phenotypic traits: (1) and (2) the expansion rate of a swarm on both 0.4% (soft) and 1.5% (hard) agar surfaces, taken as a rough estimation of Social and Adventurous motility systems, respectively; (3) the time required for development

(timing); (4) the opacity of aggregates as an indication of density (grayness); (5) the circularity of aggregates (circularity); (6) the number of aggregates in one unit area (count); (7) the average size of aggregates (area); and (8) the efficiency of sporulation (sporulation). We refer to each phenotypic trait using the above name in parentheses. For three of these variables (circularity, grayness, area), we quantified what is typically reported as a qualitative observation, but each one measures an aspect of the *M. xanthus* phenotype that is frequently described [210]. **Figure 3.2** illustrates how each assay is related to the *M. xanthus* phenotype and lifecycle.

Distribution of phenotypic variation: For each of the phenotypic assays, data from wild type and the 180 mutant strains were listed in decreasing order according to their resultant means on the x-axes and the experimental values for each phenotypic trait on the y-axes (**Figure 3.3**).

Therefore, the order of mutant strains is different for each graph. The range of phenotypic variation is different across the traits: *e.g.*, sporulation ranges from 0 to 270% of wild type, soft and hard expansion are from 10 to 130% and 50 to 125% of wild type, respectively, and count is from 0 to 4-fold greater than wild type. Nonetheless, the distributions of all eight phenotypic traits in **Figure 3.3** have at least four notable features in common: (1) all display a continuous distribution; (2) the majority of mutant strains fall within a confidence interval one standard deviation about the mean; (3) wild type is always within this confidence interval, usually near the middle; (4) mutant strains with “outlier” phenotypes (*i.e.*, ones located where the slope sharply changes at either end) always represent a small percentage of the overall population. To identify mutant phenotypes that are statistically distinguishable from wild type, we used a randomization test. Because the distributions of each phenotypic trait are continuous, it is difficult to establish meaningful thresholds that can be used to distinguish a set of strains as having one or more “mutant phenotypes.” We used a resampling strategy to compare each strain to wild type, and used an adjusted *P* value to control for false discovery rates. Using this

method, we identified the number of mutant strains that are different from wild type for each phenotypic trait—area: 39, circularity: 91, grayness: 46, hard: 37, soft: 59, and sporulation: 62. A total of 86% of strains (154/180) exhibited at least one phenotypic trait that was statistically distinguishable from wild type.

Measuring pleiotropy among phenotypic traits: A total of 93 mutant strains (52%) exhibited some degree of pleiotropy, with at least two phenotypic traits statistically distinguishable from wild type; for example, strain MXAN_1097 exhibited both a slow rate of expansion on soft agar (soft) and reduced sporulation efficiency (sporulation). To characterize pleiotropic effects in *M. xanthus*, we examined correlations among phenotypic traits using Spearman's rank correlation coefficient for all 180 mutant strains. For several strains, we noticed a large variation in the area, grayness, and/or circularity of aggregates within the same swarm. We therefore analyzed means and standard deviations separately for these three traits. **Figure 3.4** shows the correlation between the now eleven traits (with area, circularity, and grayness reported as both average and standard deviation “_std”). Each trait exhibits several positive or negative correlations. Some support common sense hypotheses or confirm long-standing empirical observations. For example, soft and hard expansion exhibit a strong positive correlation with each other, and timing exhibits a strong negative correlation with both soft and hard expansion, thus indicating that slower swarm expansion is linked to slower development. Other correlations may seem less obvious; for example, sporulation exhibits no correlation with soft or hard expansion, or with most of the phenotypic traits associated with development, except for count and circularity. In addition to this zero-order correlation, we also calculated the partial correlation coefficients between each pair of phenotypic traits, and found that they decreased in some cases (**Figure 3.4, upper panels, inside parenthesis**). In particular, five pairs of phenotypic traits associated with development exhibit decreased partial correlations: timing and count,

timing and area, timing and grayness, timing and circularity_std, area and grayness_std. For each of the three phenotypic variables represented by both average and standard deviation (area, grayness and circularity), each average still exhibits a correlation with its corresponding standard deviation, and all three standard deviations are still correlated with each other. Sporulation correlates only with count, area, and area_std, but is no longer correlated with circularity. Interestingly, the partial correlations associated with soft and hard expansion are diminished significantly. Neither soft nor hard expansion is correlated with area or area_std. More surprisingly, soft and hard expansion are no longer correlated with each other, which indicates that they are more independent with respect to phenotype than may have been previously assumed. Furthermore, soft and hard expansions now exhibit different correlations; count correlates only with soft, while grayness and circularity correlate only with hard. To summarize, partial correlation analysis reveals that the phenotypic traits associated with development are more closely correlated with each other (11 of 28 with $P < 0.05$, or 36%) than with phenotypes associated with swarming (soft and hard expansion) (5 of 18 with $P < 0.05$, or 28%). Also, the correlation of other phenotypes with soft and hard expansion shows patterns different from their zero-order correlations, and they are also more different from each other, with soft exhibiting a correlation with only two of 10 phenotypic traits, namely timing and count.

A comparison of wild type and mutant development: A quantitative description of swarm patterns that form during development has been previously reported using time-lapse microcinematography images [216-218], but these kinds of observations have not been reported for a collection of mutants. Here, we used the four phenotypic traits extracted from our images (count, area, grayness, and circularity) to describe the dynamics of development for the 180 ABC transporter mutant strains over 5 days. Changes in each of these four traits are plotted in **Figure 3.5**, and they are compared to 24 independent replicates of wild type. For the five time

points of the four traits shown in **Figure 3.5**, the averages and standard deviations of all the mutant strains are actually very similar to the replicates of wild-type, although almost all of the outliers are mutant strains.

3.4 Discussion

From these data and analyses we have made several observations that apply to at least the 180 single gene disruption mutants in the *M. xanthus* ABC transporters. Our observations can be summarized in the following statements: (1) By combining several quantitative measurements and applying randomization tests, 155 out of 180, or 86% of mutant strains were observed to exhibit at least one phenotypic trait that could be statistically distinguished from wild type. (2) The average wild type phenotypic trait closely follows the average for all 180 mutant strains, so that wild type was always within the confidence interval about the mean, and never represented an outlier for any phenotypic traits. (3) Phenotypic traits are not independent, so that observing one changed trait in a mutant strain alters the probability that other traits will also be changed, and this significantly impacts what should be considered an improbable phenotype.

The impact of the first observation is the most obvious. The percent of strains that exhibit mutant phenotypes becomes higher when more than one phenotypic trait is included, the observation's comparison to wild type is quantitative, and setting the cutoff between what is wild type and what is mutant is not done completely arbitrarily, but instead is done using well-known statistical procedures. By altering our experimental design and analysis accordingly, we were able to increase the percent of distinguishable mutant strains in *M. xanthus* to 86%, and this was using only a subset of standard *M. xanthus* phenotypic assays. These findings directly address the question of why previous studies have detected few *M. xanthus* mutant strains with distinguishable phenotypes: in part, it is because the limitations of our assays and analysis did not capture more subtle phenotypic changes.

The second observation affects how we view each mutant strain's phenotypic traits and, by extension, the relationship between these standard laboratory assays and any real consideration of fitness. We disrupted 180 ORFs with no pre-selection for a specific phenotype, and found that between 3 and 26% of mutant strains performed “better” than wild type for any trait. This means that wild type is almost always somewhere toward the middle of the confidence interval about the mean, and this observation alone provides compelling evidence that fitness should not be defined through trait maximization, and assay results should be analyzed accordingly. Otherwise, by assigning wild type as “100%” of any phenotypic trait and then selecting only mutant strains that perform less than this arbitrary 100%, we will fail to distinguish a significant number of mutant phenotypes and inadvertently bias our results.

The impact of the third observation relates to the identification of outlier mutant phenotypes. If two phenotypic traits (A and B) exhibit a high degree of pleiotropy, so that for every mutant strain that exhibits a change in A there is always a change in B, then it is assumed that A and B must share some molecular underpinnings. In such a case, a strain that exhibits a large change in A with no change in B would be exceptional, and might be of particular interest to someone studying either trait. For example, *M. xanthus* aggregation and sporulation are correlated, so that mutant strains which fail to aggregate are much more likely to fail at sporulation. Therefore, a mutant strain such as MXAN_6671, that has been shown to sporulate without aggregating, represents a very interesting outlier mutant strain (Welch lab, unpublished data*). *The gene MXAN_6671 (sglK) has been previously disrupted using transposon insertion [219,220] and was described as having defects in aggregation and a lower sporulation efficiency than wild type (5%). In our laboratory, the disruption mutant displayed defects in aggregation and nearly wild type sporulation efficiency.

For a model organism such as *M. xanthus*, initial estimations of the dimensionality and scale of the phenome depend on the identification and characterization of outlier strains, since

sets of phenotypic traits exhibited by these exceptional strains represent practical boundary values. Current definitions for the phenome of an organism are imprecise, but if the word “phenome” has any valid scientific meaning, it cannot be defined as infinite. Perhaps it would be logical to think of boundary values for a phenome as sets of phenotypic traits that are so unlikely to occur that their probability approaches zero. Therefore, by compiling sets of traits using standard assays for hundreds of *M. xanthus* mutant strains, we are just beginning to populate a map of the *M. xanthus* phenome. In this map, each independent phenotypic trait represents one full “dimension” of the phenome, and two correlated traits represent more than one and less than two full dimensions.

CHAPTER 3

FIGURES

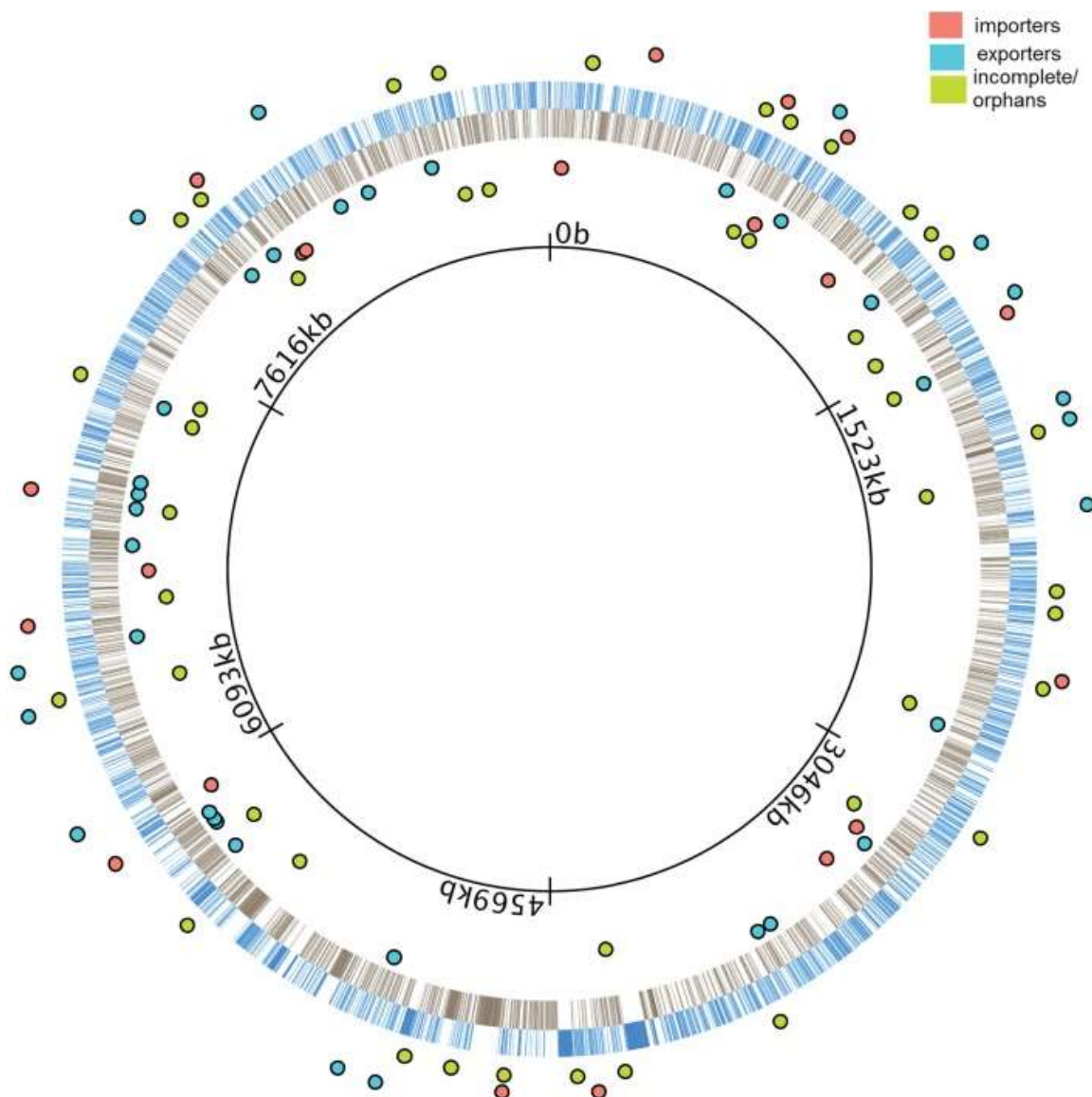


Fig. 3.1 Distribution of ABC transporters in the *M. xanthus* genome. The innermost black ring represents the *M. xanthus* chromosome. Blue bars in the outer ring represent the genes transcribed in the clockwise direction (+ strand) while gray bars in the inner ring represents the genes transcribed in the counterclockwise direction (- strand). Colored dots represent operons containing coding genes for ABC transporters. The location of dots indicates that they are either in the + strand (outside blue ring) or in the - strand (inside black ring). Cyan and salmon colored dots represent full operons coding for complete exporters and importers, respectively. Green colored dots represent incomplete and orphan operons.

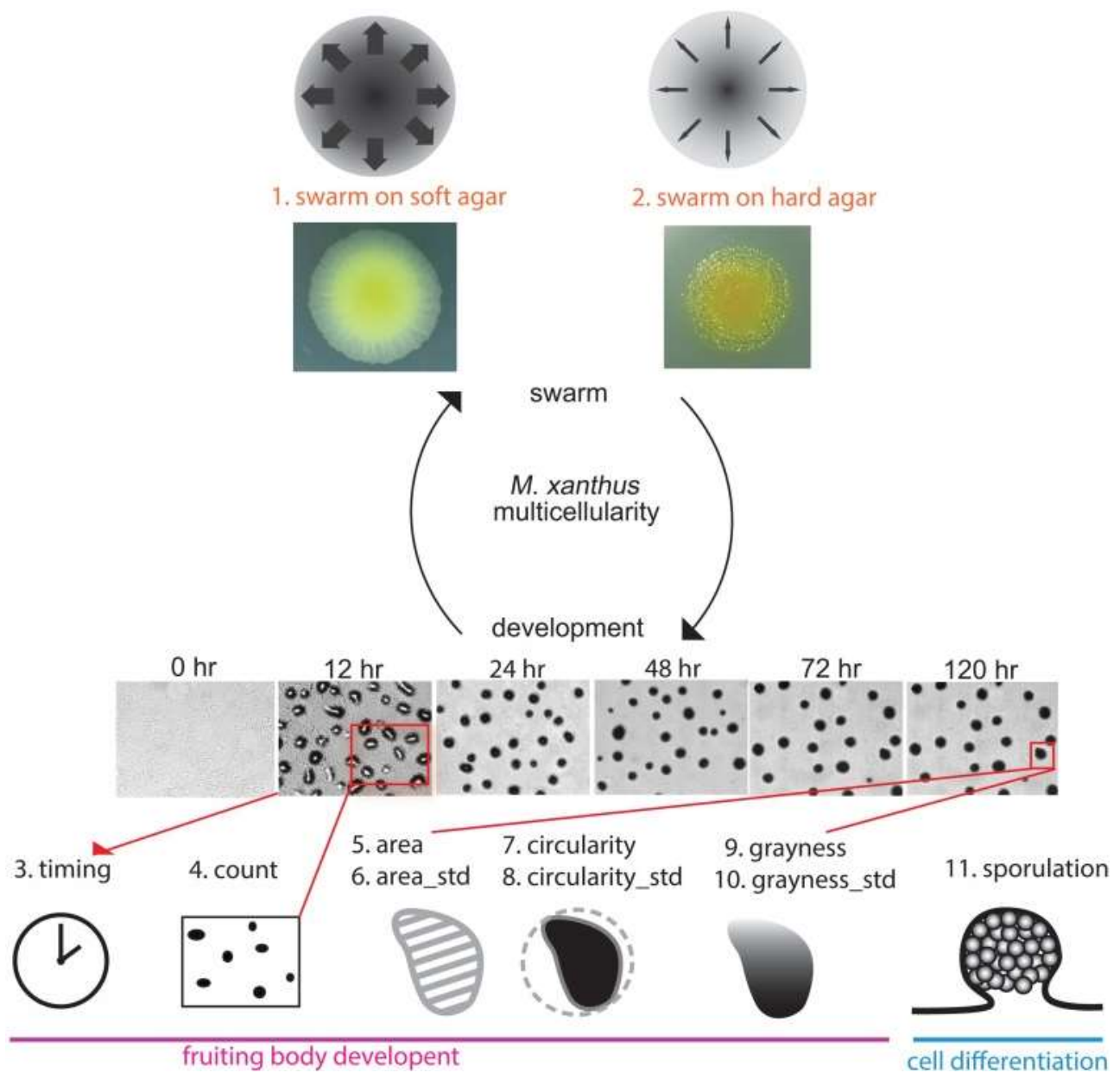


Fig. 3.2 Phenotypic traits and phenotypic assays. We tested eight phenotypic traits spanning the two parts of the *M. xanthus* life cycle: swarming (top) and development (bottom). Two data sets related to swarming were obtained under rich media using either soft (0.4%) or hard (1.5%) agar as rough measurements of S and A motility, respectively. Images of the two yellow colonies are swarms after 3 days on both agar concentrations. Six data sets related to development were obtained under nutrient starvation. The six panels with times listed above are images of wild type development. Sporulation is also related to development, as well as cell differentiation.

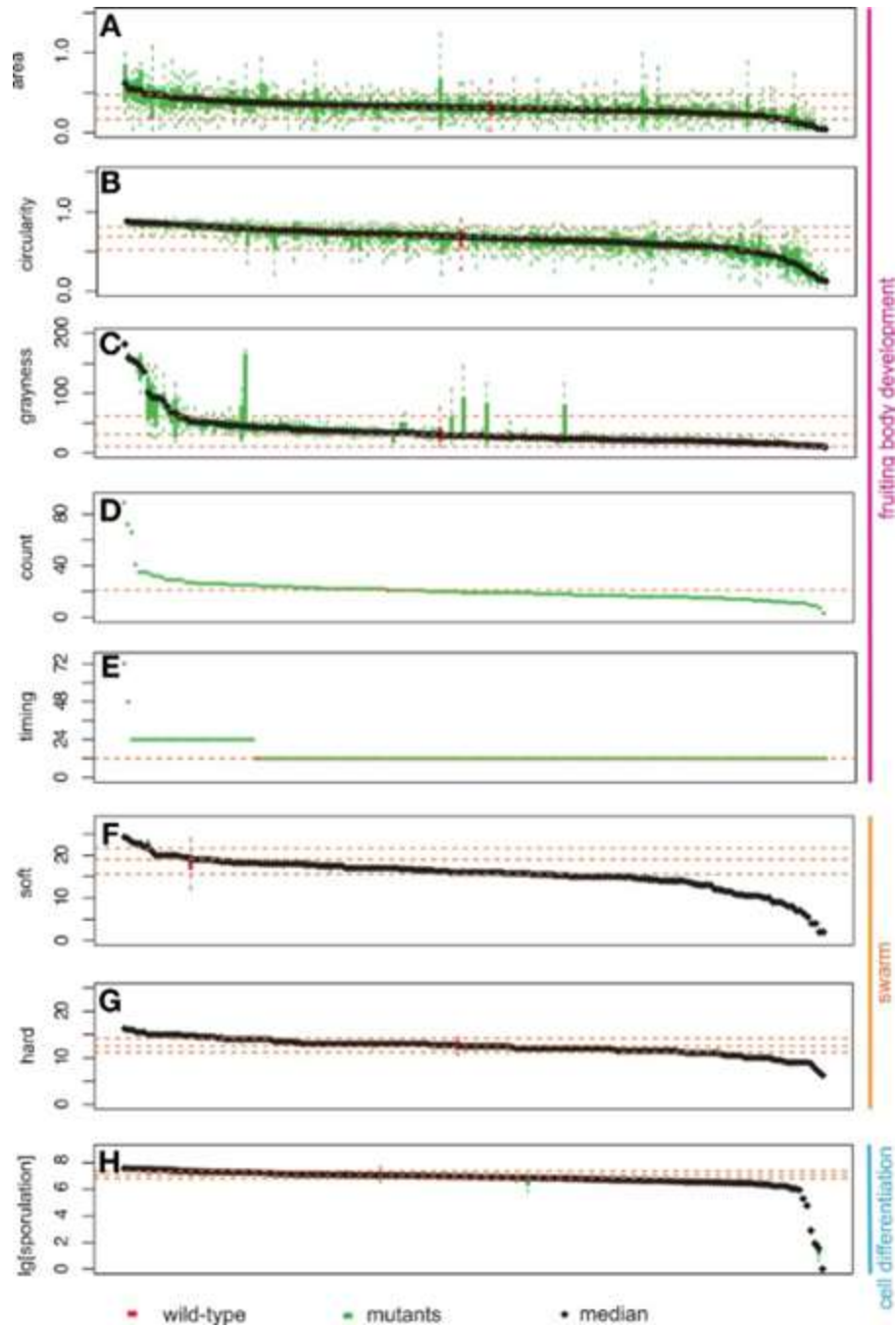


Fig. 3.3 Distribution of phenotypic data: area (A), circularity (B), grayness (C), count (D), timing (E), expansion on soft (F), and hard agar (G), and sporulation efficiency (H). The y-axis represents the measurement for each phenotypic assay. Bars on the x-axis represent 180 mutants and wild type. Green bars represent mutant strains, while red bars represent wild type. Box plots represent two middle quartiles. Error bars represent the top and bottom interquartile range for each strain. Yellow dash lines represent the mean \pm SD (top, bottom), and median for wild type. In (D,E), no error bars are present due to the nature of our measurements (for details, please see Materials and Methods).

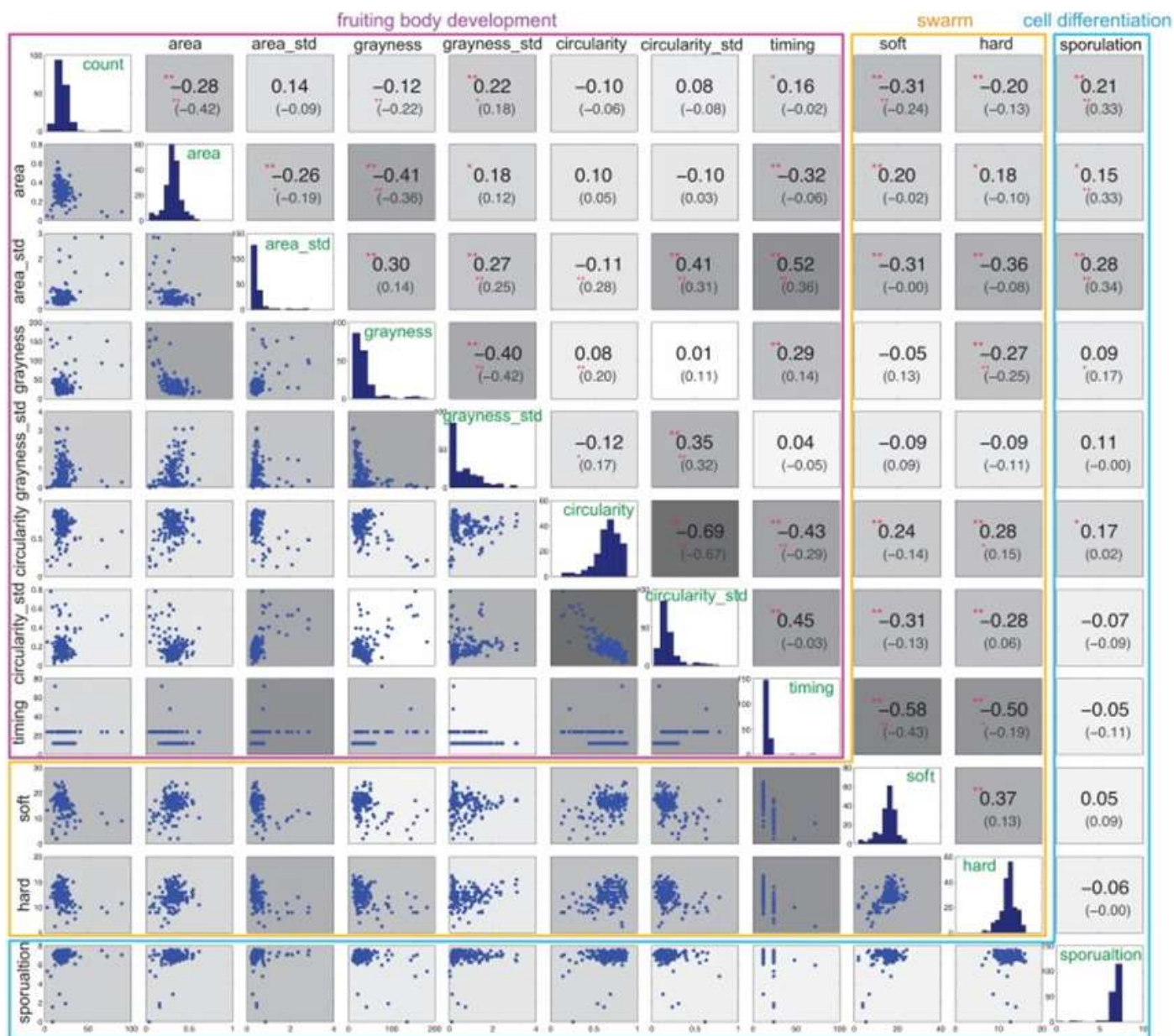


Fig. 3.4 Correlation of phenotypic variables. The histogram in the diagonal panels shows the distribution of each phenotypic trait. The value of Spearman's rank correlation is the first number shown above the diagonal. The number inside the parenthesis is the Spearman's rank partial correlation. The background color above and below the diagonal corresponds to the degree of correlation, where deeper gray indicates a higher correlation. Significance correlations are indicated with asterisks: **0.01, *0.05.

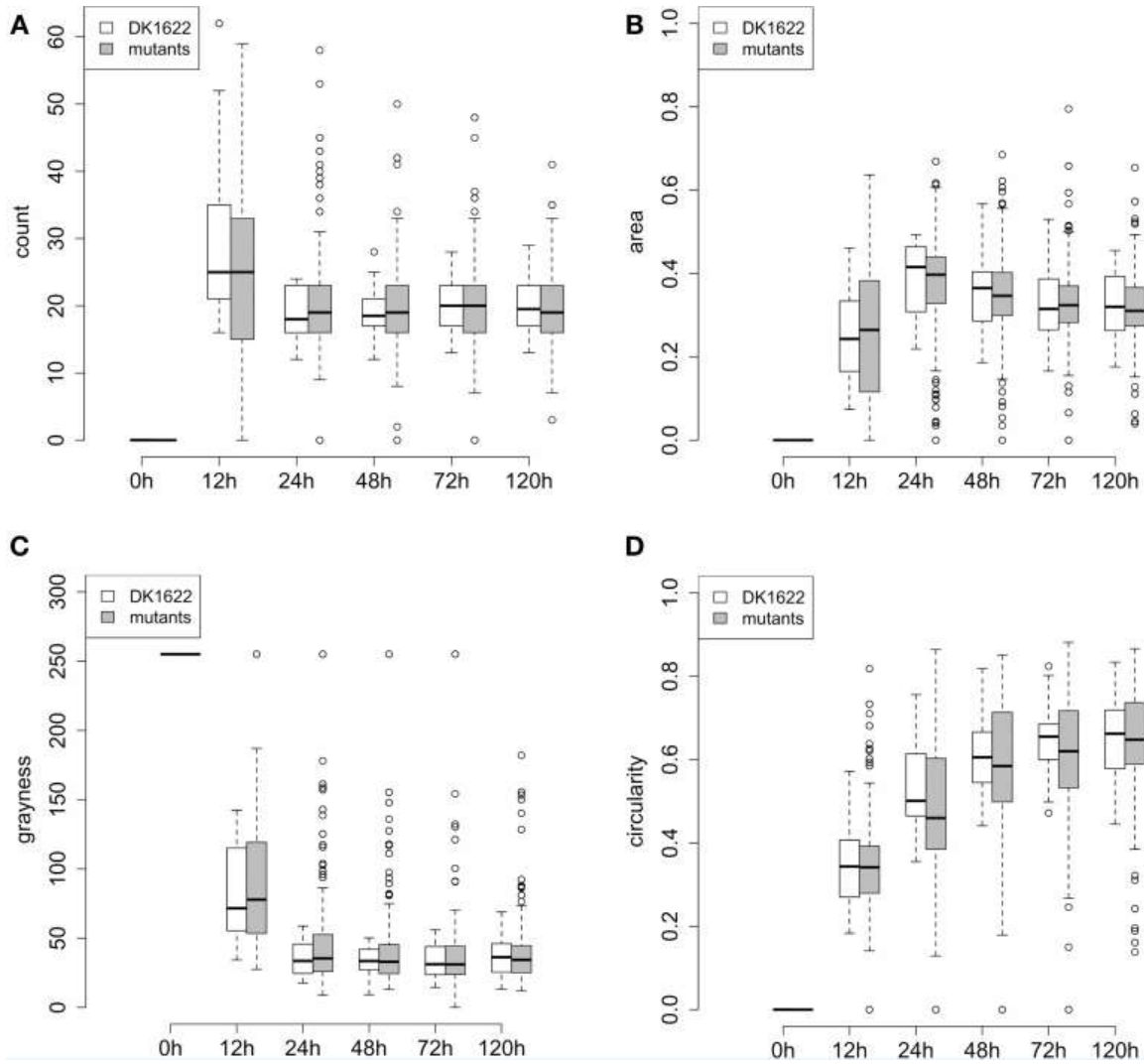


Fig. 3.5 Phenotypic traits of wild type and mutant aggregates during development. (A–D) Represent the quantitative traits for fruiting body development: **(A)** count, **(B)** area, **(C)** grayness, and **(D)** circularity. At each time point, the box represents the middle 50% of the data points, together with the median (thick line in each box) for the 180 mutants or DK1622 wild type replicates. Error bars represent the 1.5 interquartile ranges. Small circles above and below each error bar represent outliers.

CHAPTER 4

INTER-LABORATORY EVOLUTION OF WILD TYPE SUBLINES

The material in this section has been published in:

Bradley, M.D., Neu, D., Bahar, F., & Welch R.D. Inter-laboratory evolution of a model organism and its epistatic effects on mutagenesis screens. *Sci. Rep.* 6, 38001 (2016).

4.1 Project summary:

Laboratory evolution is inevitable for a model microbial system like *Myxococcus xanthus*.

Mutations that have a less than catastrophic impact on the wild type phenotype of *M. xanthus* will likely go unnoticed. Over time, these mutations might cause deviations in the phenotypes of inter-laboratory wild type sublines.

We have carefully measured the phenotypic profiles of DK1622 obtained from nine different laboratories and report considerable differences in aggregate number, aggregate size, and viable spore counts, with the developmental phenotype of two strains representing both extremes. Whole genome sequencing revealed few genomic differences among all sublines, supporting the hypothesis that a few naturally occurring mutations can have a significant impact on phenotype. Using resequencing and common garden characterizations, our first aim was to convert a strain with a mean developmental phenotype to resemble a subline with an extreme developmental phenotype by introducing the mutations identified by our variant screen. Our second aim was to identify an example of naturally occurring epistasis, so that constructing an identical mutation in two different sublines yields a significantly different phenotype.

4.2 Materials and Methods:

Wild type sublines: The first *M. xanthus* isolate moved into the laboratory was strain FB [221]. DK1622 is a derivative of FB that swarms on nutritive media and develops on starvation media [222]. In 2014, we received DK1622 from eight other laboratories that study *M. xanthus* as a model organism. Each subline was received on nutrient agar, grown in nutrient broth, concentrated, and preserved as a frozen stock. Our laboratory subline (S8) was cloned from the Kaiser Strain Archive at Stanford University in 2003.

Growth conditions: *M. xanthus* cells were cultured on CTTYE [223] + 1.5% agar plates and incubated at 32°C. Liquid cultures were prepared in agitating CTTYE liquid media. Media was

supplemented with 40µg/mL kanamycin sulfate for insertion-disruption mutants. Liquid cultures were harvested at a density of 5×10^8 cells/mL and concentrated tenfold for characterization assays. Cells were washed with 5mL TPM buffer [224] before performing development assays.

Strain characterization and analysis: DK1622 sublines were characterized as previously described [225]. Briefly, A and S expansion rates were measured by spotting sublines onto CTTYE + 1.5% (hard) and 0.4% (soft) agar plates. Growth rates were determined by dividing the swarm diameters by growth hours. Five temporally independent replicates were conducted. Aggregation assays were performed by spotting cells onto TPM + 1.5% agar. Images of resulting aggregates were captured after 24 hours using 20× brightfield microscopy and SPOT software (SPOT Business Systems). Resulting aggregates were manually counted. Sporulation assays were performed by spotting cells onto TPM + 1.5% agar, incubating for 120 hours, and scraping spore containing aggregates off the substrate. Cells were sonicated, diluted, and plated onto CTTYE. The resulting colonies are presumed to arise from a single germinated spore, and colony counts represent the number of viable spores (*i.e.* spores that survive heat and sonication). Three temporally independent replicates were conducted for both development assays.

Multiplex sequencing: Genomic DNA was extracted and purified using Zymo Universal Quick-DNA and DNA Clean & Concentrator miniprep kits (Zymo Research). Library preparation was performed using Nextera XT dual indexing kit (Illumina) according to the manufacturer's instructions. Fragmented DNA libraries were verified with a 2100 Bioanalyzer (Agilent) and sequenced on a NextSeq500 pyrosequencer (Illumina) at the University of Pittsburgh Children's Hospital Rangos Genomics Facility.

Genome assembly and variant detection: CLC Genomics Workbench (v8.0, Qiagen) was used to filter and assemble reads: Short reads (<60 base pairs), reads containing ambiguous nucleotides (“N”), low quality reads, duplicate reads (artificially inflates mapping coverage), and homopolymers were removed. Reads were assembled against the *M. xanthus* DK1622 reference genome [226] (NCBI accession number: NC_008095). Fixed variants were identified by restricting candidates to a frequency of $\geq 95\%$ and a minimum sequencing depth of 15x.

Mutant strain construction: Targeted insertion-disruption mutations were performed as previously described [5]. Briefly, fragments of target genes were ligated into a pCR2.1 TOPO vector (Thermo Fisher) containing a kanamycin resistance selective marker, and replicated in TOP10 *E. coli* host cells (Invitrogen). Plasmids were integrated into the *M. xanthus* chromosome via homologous recombination [227]. Plasmid integration was confirmed by PCR.

Data analysis: Statistical comparisons between subline phenotype data sets were made using a one-way ANOVA ($\alpha = 0.05$), followed by *post hoc* analysis with Tukey’s multiple comparisons test (TMC) with multiplicity adjusted *P* values. Aggregation data were log-transformed prior to analyses to achieve a normal distribution.

4.3 Results

Characterization of subline phenotypes in a common garden: Phenotype is considered the product of two variables, genotype and environment. To minimize environmental effects, all characterization experiments were performed under identical conditions in the same laboratory. All sublines were grown in aliquots from the same media preparations, each assay was performed on all sublines together using the same reagents and equipment, and images of each subline were acquired at the same time. These conditions defined our “common garden.”

A set of representative swarming images for each subline reveal clear differences in several qualitative aspects of phenotype (**Fig. 4.1a**); a side-by-side comparison of swarm expansion on hard agar (**Fig. 4.1a, top row**) reveals differences in swarm translucency and edge flare patterns. For example, sublines S5 and S6 are more translucent than S2 and S8, and the swarm edge flares are more pronounced in S2 and S8 than in S5 and S6. The same type of comparison on soft agar (**Fig. 4.1a, middle row**) reveals a range of swarm shapes, edge flare patterns, and color gradients extending from the center of the swarm to its edge. Sublines S3 and S5 are nearly circular in shape with a smooth swarm edge and a steep color gradient from dark yellow to translucent, while S4 and S8 are nearly circular in shape with a rough swarm edge consisting of numerous small and directional flares and a more subtle color gradient. Subline S6 is irregular in shape with a rough swarm edge consisting of a variety of flare shapes and an inconsistent color gradient, while S9 is circular in shape with long pronounced edge flares and no color gradient. A similar range of phenotypes is revealed by a side-by-side comparison of development on starvation agar (**Fig. 4.1a, bottom row**). Subline aggregates range from small (S5) to large (S4), some of the sublines appear to have a greater distribution of individual aggregate sizes (S1), and some appear to have a dense ring of aggregates at the outermost edge (S1, S2, S4, S5, S8).

This qualitative characterization of sublines in a common garden reveals obvious differences, but their description is subjective, and so it is impossible to rank them using only this information. To achieve such a ranking, we selected four quantitative assays to represent the phenotype of DK1622 at both stages of its life cycle (**Fig. 4.1b**). Two swarming assays are used to measure changes in swarm diameter on hard and soft nutrient agar, and are considered estimates of the expansion rate for A and S motility systems. Two development assays are used to measure the number of aggregates and the number of spores that form on starvation agar, and are considered tests of self-organization and cellular differentiation. These four assays are commonly used by the research community to compare the phenotypes of mutant *M. xanthus*

strains, using DK1622 as a wild type control. Because all of the strains in this study are DK1622 there can be no wild type control, and thus subline assay data are arranged by increasing means for each of these four traits.

The sublines differed significantly from one another in each of the four assays (**Table 4.1**), indicating that there are real and measurable differences between sublines. Because we observed a continuous distribution of means in each assay, we focused on the sublines at the phenotypic extremes, which we hereafter refer to as “outlier” sublines. We deemed a subline to be an outlier if its mean \pm *SD* falls beyond one standard deviation of the total population mean for that assay. Based on this criterion, we identified two outliers with respect to development traits: S1 is an outlier for sporulation, and S9 is an outlier for both aggregation and sporulation. S9 is also an outlier with respect to S motility. To determine if the non-outlier sublines are different from each other, we repeated the analyses with S1 and S9 excluded, and the results remain significant for all traits except S motility (**Table 4.2**).

Subline variant screen: Each subline was sequenced and assembled using the original closed DK1622 genome sequence [226] as a scaffold (NCBI accession number: NC_008095). An average of 8.1 million reads covering >99.4% of the scaffold genome were mapped for each subline. A total of 29 variants, consisting of 28 single nucleotide polymorphisms (SNPs) and one nucleotide deletion were identified among the nine sublines (**Table 4.3**). Any variant that occurred in two or more sublines was counted as one variant (i.e. the overlapping SNPs found in S2, S4, S5, and S8 were counted only once). Of the 28 SNPs, eight are transitions and 20 are transversions. Twenty-one variants (72%) are located within putative Open Reading Frames (ORFs), 11 of which are non-synonymous (i.e. they alter the protein coding sequence of their constituent ORF). Eight variants are found within noncoding regions. No evidence of chromosome structural variation was found in any of the sublines using the variant detection parameters described in Materials and Methods; a sampling of possible insertions and deletions

with scores below the stated threshold were examined, and all were confirmed to be false positives by PCR (data not shown).

One part of the subline variant screen is in agreement with the common garden characterizations and functions as a useful control; sublines S2 and S4 are identical with respect to genotype, and they do not vary to a significant degree with respect to the four quantitative assays used in this study ($P \geq 0.818$ for each trait). In addition, there are two results from these resequencing data that are notable, even though they are tangential to the primary focus of this study: First, Velicer *et al.* previously reported five variants in a derivative of the S3 subline in 2006 [228]. These five variants were independently identified in this study along with three more, which may indicate that S3 has continued to accumulate mutations since 2006 or that the higher sequencing depth in this study was able to identify three variants that were not identified in the previous study. Second, two variants, a thymine-to-guanine transversion at position 830180 and a thymine-to-cytosine transition at position 7101832 are in all of the sublines. Because some of these sublines have been isolated from each other for more than 30 years, while the reference genome sequence was completed just over ten years ago, the simplest explanation for these two variants is that they represent sequencing errors within the reference genome.

Targeted mutagenesis: S1 and S9 represent outlier sublines on different ends of the development rankings (**Fig. 4.1b**); S1 produces fewer spores than any other subline and is among the group of sublines (S1, S2, S3, and S4) that produce the fewest aggregates, whereas S9 produces more aggregates and more spores than any other subline. However, despite these differences, both sublines have functioned effectively as wild type controls for years in their respective laboratories.

The non-synonymous variants from either S1 or S9 are reasonable candidates for causing each subline's outlier phenotype because they alter a protein's sequence, and therefore

may negatively affect its function. If it does, then disrupting the ORFs that harbor these variants in a more “average” non-outlier subline may shift its phenotype to resemble the phenotype of the corresponding outlier subline. To test this, we selected S8 to represent the average subline; S8 has a nearly average aggregate count, and its spore count is significantly different from both S1 ($P < 0.001$) and S9 ($P = 0.023$).

We constructed mutant strains containing single ORF insertion-disruptions in S8 for each of the ORFs harboring the three non-synonymous variants specific to the outliers S1 (located in MXAN4601 and MXAN4672) and S9 (located in MXAN7041); these new mutant strains are hereafter referred to as S8_4601, S8_4762, and S8_7041 respectively. Aggregation and sporulation assays were performed on each of the strains, and results were compared to both the parent subline (S8) and the corresponding outlier subline (either S1 or S9). For each mutant strain, the change in phenotype is reported as a percent change compared to the parent subline: S8_4601 exhibits a 33% reduction in aggregate count, which is significantly lower than S8 ($P = 0.005$) and matches its corresponding outlier subline, S1 ($P = 0.863$) (**Fig. 4.2a**). S8_4762 exhibits a 36% reduction in spore count, which is significantly lower than S8 ($P = 0.021$) and is intermediate between S8 and its corresponding outlier subline, S1 (**Fig. 4.2b**). Spore count for S8_4601 and aggregate count for S8_4762 did not differ from S8 to a significant degree (data not shown). S8_7041 exhibits a 238% increase in aggregate count, which is significantly higher than S8 ($P < 0.001$) and matches its corresponding outlier subline, S9 ($P = 0.981$) (**Fig. 4.2c**). S8_7041 also exhibits a 70% increase in spore count, which is significantly higher than both S8 ($P = 0.001$) and its corresponding outlier subline, S9 ($P = 0.026$) (**Fig. 4.2c**).

It is important to note that while the sporulation phenotypes of S8_4762 and S8_7041 do not exactly match that of their corresponding outlier sublines, they are both different from their parent subline in a way that moves their phenotypes closer to their corresponding outlier sublines. In other words, S8_4762 produces significantly fewer spores, which is more like S1, and S8_7041 produces significantly more spores, which is more like S9. These data support the

idea that disrupting ORFs harboring these unique outlier variants in a more phenotypically average subline would shift its phenotype towards that of the outlier subline. These data also support the idea that the genetic variants in the outlier sublines are likely loss-of-function mutations, because if one of the variants were gain-of-function, then disrupting its corresponding ORF would likely have driven the phenotype of the average subline away from the phenotype of the outlier subline. In particular, the alanine-to-proline substitution in MXAN7041 of S9 almost certainly has a detrimental impact on its protein structure and function, due to severe conformational constraints imposed on it by proline.

Results from the *M. xanthus* subline resequencing and analysis provide strong evidence for microevolution, and the purpose of our mutant analysis thus far has been to identify the candidate sublines and candidate ORFs most likely to provide strong evidence of epistasis. For the candidate sublines, we chose the average subline S8, together with the two phenotypically opposite outlier sublines S1 and S9. For the candidate ORFs we chose MXAN4601 and MXAN7041 because, at least for S8, their disruption significantly changes the results of the two most common development assays in opposite directions, so differences in their impact are easy to distinguish. To test for epistasis, one of the variant ORFs specific to each outlier subline, in this case MXAN4601, which is specific to S1, and MXAN7041, which is specific to S9, was disrupted in the opposing outlier subline. In other words, MXAN4601 was disrupted in S9 to create the mutant strain S9_4601, and MXAN7041 was disrupted in S1 to create the mutant strain S1_7041. The development phenotypes of both these strains were then compared to the development phenotypes of the same ORF disrupted in S8 (**Fig. 4.3**).

Differences in phenotype between a mutant strain and its parent subline are reported as a percent change compared to the parent subline: S9_4601 exhibits a relatively large reduction in aggregate count (81%) and spore count (45%) when compared to S8_4601, which exhibits a small reduction in aggregate count (33%) and no significant change in spore count (**Fig. 4.3a**). Most notably, S1_7041 exhibits no significant change in either aggregate or spore count,

whereas S8_7041 exhibits a large increase in both (238% and 70%, respectively) (**Fig. 4.3b**). It is important to note that the same construct was used to disrupt MXAN7041 in sublines S1 and S8; S1_7041 and S8_7041 have the same disruption genotype, and only differ by the naturally occurring variants listed in **Table 4.1**. Clearly, the variants between sublines S1 and S8 are having an epistatic effect on the disruption of MXAN7041, enough that this ORF would be annotated as “involved in development” in S8, but not S1.

4.4 Discussion

Several previous studies have explored inter-laboratory microbial evolution. In 2007, Schacherer *et al.* identified nonrandom mutational events among several closely related laboratory sublines of *Saccharomyces cerevisiae* [229]. In 2008, Srivatsan *et al.* resequenced several *Bacillus subtilis* sublines and identified a previously unknown metabolism defect [230]. Finally, in 2010, Klockgether *et al.* identified discordant genotypes of the widely studied *Pseudomonas aeruginosa* strain PAO1 [231]. To the best of our knowledge, this is the first study to examine the epistatic impact of microevolution on a microbial model organism, and to demonstrate that it was sufficient to change the initial annotation of a gene with respect to its biological function.

There are two mechanisms that could affect the evolution of *M. xanthus* in the laboratory. The first is genetic drift, which would have a stochastic effect on each subline’s genome and is almost certainly responsible for some of the genetic variation between sublines. The second is selection, which probably varies between laboratories and affects each subline differently. Selection may occur when cells are grown in liquid culture, favoring faster growing cells. It may occur when cells are grown as swarms on nutrient agar plates because inoculants for liquid cultures are taken from the swarm edge, favoring highly motile cells that may be overrepresented there. It may occur when cell cultures are made into frozen stocks, favoring

cells that are better able to resist lysis when frozen. Certainly, purifying selection is always occurring, so that any cells with mutations that have a deleterious impact on growth or survival are removed. The variants in the nine *M. xanthus* sublines most likely were produced through some combination of these factors: drift, purifying selection, and selective pressures that were slightly different in each of the laboratories.

Resequencing of the nine *M. xanthus* DK1622 sublines clearly demonstrated microevolution, given that DK1622 has a single origin. Rather than randomly searching for an example of epistasis from that point, we decided to hedge our bets. This is why we singled out the two outlier sublines from the common garden, why we selected the three candidate ORFs we deemed most likely to be responsible for the sublines' outlier phenotypes, and why we finally settled on the two ORFs whose disruption caused the strongest opposite changes in phenotype. In this study, our goal was to identify a statistically significant and entirely unambiguous example of naturally occurring epistasis, and we believe that we identified at least one: S1_7041 versus S8_7041. Future studies that employ a broader mutagenesis approach will likely produce an "epistasis distribution", which may provide insight into the role of epistasis in the annotation of the *M. xanthus* genome.

The practical impact of epistasis on determining biological function in *M. xanthus* is evident in these results. Depending on which laboratory constructed the initial disruption, the ORF MXAN7041 may or may not have been identified as important for development. Disrupting MXAN7041 in S8 causes a more than 200% increase in aggregation and an almost 100% increase in sporulation, whereas disrupting that same ORF in S1 results in no change to either aggregation or sporulation. If a screen were performed for aggregation and sporulation mutants, S8_7041 would be identified as a gain-of-function mutation, whereas S1_7041 would not. This initial characterization would then guide all further experiments, as well as the annotation of this ORF with respect to its biological function.

Epistasis is a fundamental and frequently observed evolutionary phenomenon; thousands of examples have been identified [232-233], and yet our ability to predict when and how epistasis will manifest remains very poor, and shows no real sign of improving. Perhaps this is because established evolutionary principles, like epistasis, seem to contradict the current interaction-network-as-a-circuit functional genomics paradigm, and this has produced a form of cognitive dissonance. As a result, a concept like gradual microevolution and epistasis can seem both obvious and confounding.

It is important to note that our findings are not from a controlled evolution project designed to demonstrate that gradual microevolution and epistasis could occur in isolates of a model bacterium when separated by time and distance. Rather, gradual microevolution and epistasis has occurred in *M. xanthus* wild type DK1622 laboratory stocks whose genomes were assumed to be identical and static. For the past ten years, the interpretation of mutant *M. xanthus* phenotype data has been based on the implicit assumption that the 2005 published reference sequence was the genome sequence for DK1622 in every laboratory that studied *M. xanthus*. This assumption is false, at least for the sublines in this study, all of which have at least one or two variants that are different from the reference. It seems very likely that this occurrence in *M. xanthus* is one example of a common phenomenon that is also happening in other model organisms.

CHAPTER 4

TABLES & FIGURES

Table 4.1 Summary of ANOVAs.

Quantitative trait	<i>F</i> (DFn, DFd)	<i>P</i> value
Expansion rate (A motility)	3.877 (8, 36)	0.0022
Expansion rate (S motility)	3.050 (8, 36)	0.01
Aggregate count	111.4 (8, 18)	< 0.0001
Viable spore count	28.23 (8, 18)	< 0.0001

A one-way analysis of variance (ANOVA) was performed for each quantitative trait. Degrees of freedom (DF) are calculated from the number of sublines (nine, numerator) and replicate experiments (three or five, denominator). Aggregate count data were normalized by log-transforming prior to analysis. Significant differences in subline means are indicated by $P < 0.05$.

Table 4.2 Summary of ANOVAs with outliers removed.

Quantitative trait	<i>F</i> (DFn, DFd)	<i>P</i> value
Expansion rate (A motility)	4.340 (6, 28)	0.0032
Expansion rate (S motility)	1.776 (6,28)	0.1405
Aggregate count	78.81 (6,14)	< 0.0001
Viable spore count	13.34 (6,14)	< 0.0001

A one-way analysis of variance (ANOVA) was performed for each quantitative trait with sublines S1 and S9 removed from the analysis. Degrees of freedom (DF) are calculated from the number of sublines (nine, numerator) and replicate experiments (three or five, denominator). Aggregate count data were normalized by log-transforming prior to analysis. Significant differences in subline means are indicated by $P < 0.05$.

Table 4.3 Subline variant screen.

Subline	Position	ORF	Nt change	AA change	TIGR annotation
S1	2678143	Nc	C → A		
	3849765	Nc	G → T		
	5766255	MXAN_4601	G → C	G408R	non-ribosomal peptide synthase
	5964097	MXAN_4762	C → A	G102C	6,7-dimethyl-8-ribityllumazine synthase
	6146569	MXAN_4912	G → A	Syn	hypothetical protein
	8023083	MXAN_6505	C → A	Syn	sulfate permease
	9009601	MXAN_7384	T → A	Syn	conserved hypothetical protein
	9089098	MXAN_7464	C → A	Syn	hypothetical protein
S2, S4	3591006	MXAN_3061	T → C	V140A	α-L-glutamate ligases, RimK family
	5591111	MXAN_4513	G → T	L837I	conserved hypothetical protein
S3	1717943	MXAN_1458	G → T	Syn	methyltransferase domain protein
	2304200	MXAN_1970	T → Δ	L72Fs	transcriptional regulator, ArsR family
	4545964	MXAN_3780	G → T	Syn	patatin-like phospholipase family protein
	5258697	MXAN_4292	G → C	Q654E	polyketide synthase
	5391338	MXAN_4388	G → A	T199M	D-lysine 5,6-aminomutase, α subunit
	5893144	Nc	A → C		
	6287570	Nc	C → A		
S5	8333625	MXAN_6783	G → A	Syn	decarboxylase, group II
	2727430	Nc	C → G		
	3591006	MXAN_3061	T → C	V140A	α-L-glutamate ligases, RimK family
	5591111	MXAN_4513	G → T	L837I	conserved hypothetical protein
	6423875	MXAN_5143	C → T	Syn	tol-pal system protein YbgF
	6465733	MXAN_5175	T → G	B139A	hypothetical protein
	8966988	Nc	T → C		
S6	1523890	MXAN_1297	C → A	R288S	putative serine/threonine protein kinase
	1985177	Nc	G → T		
S7	2310245	Nc	C → G		
S8	3591006	MXAN_3061	T → C	V140A	α-L-glutamate ligases, RimK family
	4880522	MXAN_4000	G → T	Syn	non-ribosomal peptide synthetase
	5180405	MXAN_4219	G → A	Syn	α keto acid dehydrogenase complex
	5591111	MXAN_4513	G → T	L837I	conserved hypothetical protein
S9	8617589	MXAN_7041	C → G	A80P	cyclic nucleotide-binding domain protein
	8617592	MXAN_7041	C → T	V79M	cyclic nucleotide-binding domain protein

Nt, nucleotide; Nc, noncoding; AA, amino acid; Δ, deletion; Syn, synonymous; Fs, frameshift. Sublines S2 and S4 are listed together because they have identical variants.

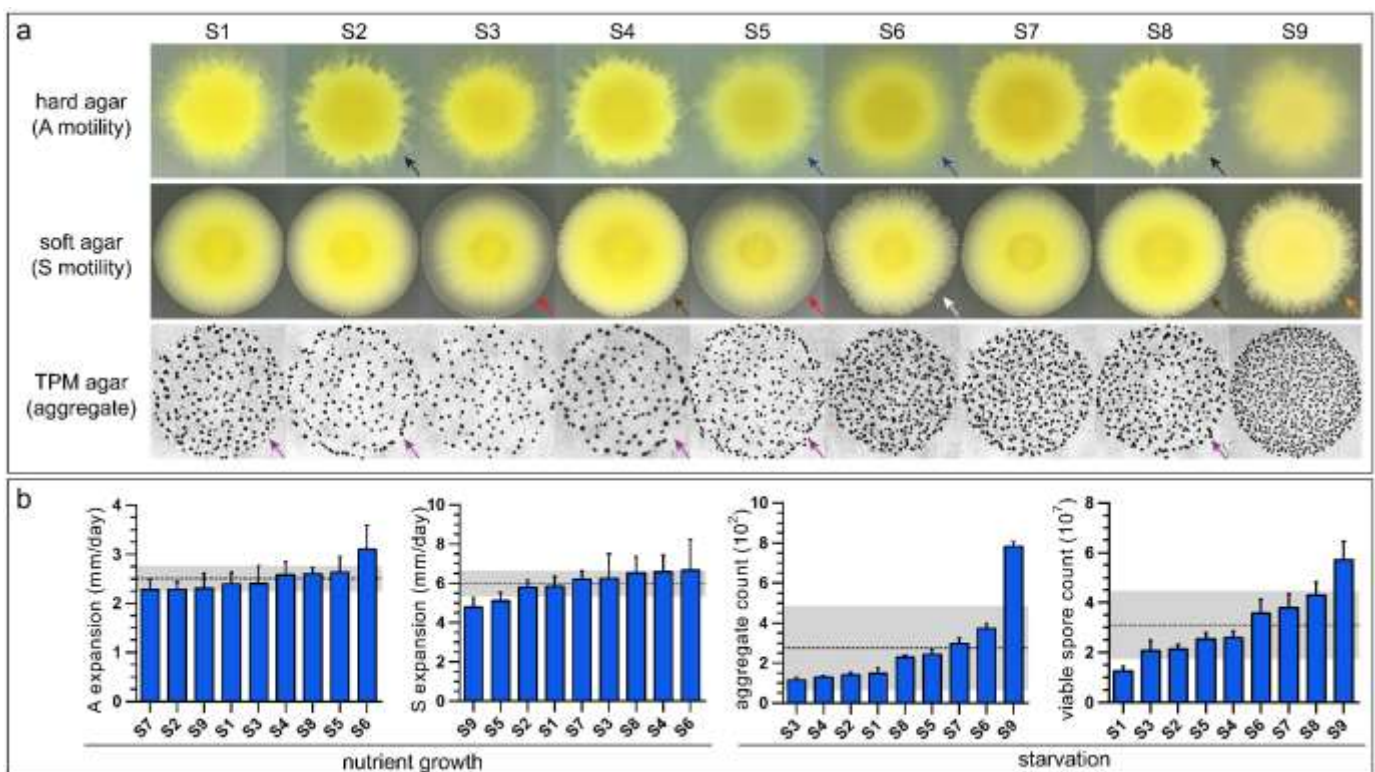


Fig. 4.1 Characterization of DK1622 subline phenotypes in a common garden. The nine sublines were characterized for growth and development: **(a)** Qualitative comparisons of A motility (top), S motility (middle), and aggregation (bottom); Black arrows indicate pronounced edge flares; blue arrows indicate stunted edge flares; red arrows indicate steep color gradients from the swarm center to the edge; brown arrows indicate directional edge flares; the white arrow indicates an irregular swarm shape; the orange arrow indicates long edge flares; purple arrows indicate sublines that have dense outer rings of aggregates. **(b)** Quantitative comparison of A & S motility, aggregation, and sporulation. The x-axes are ordered by increasing mean values. Error bars represent \pm SD for each subline. The dashed line represents the population mean. The gray bar represents \pm SD of the population mean.

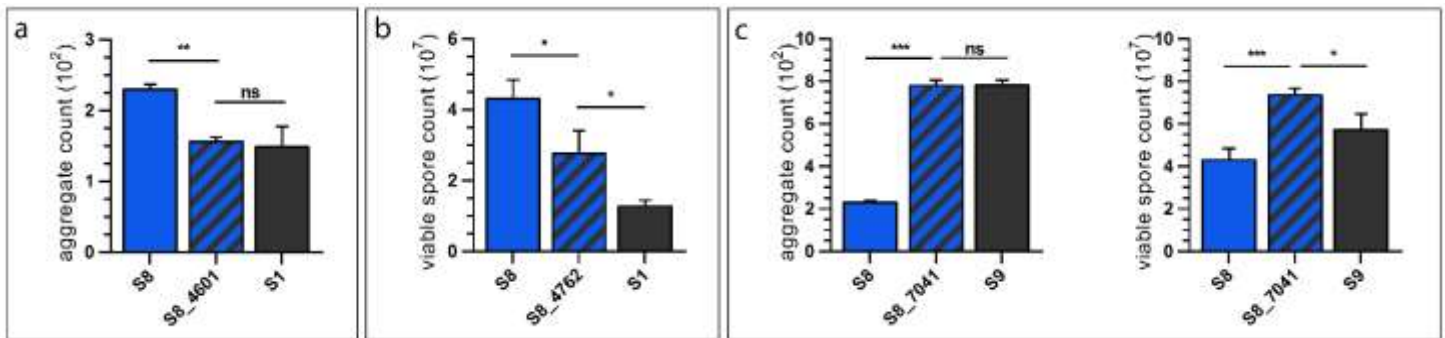


Fig. 4.2 Characterization of S8 mutant strains. Insertion-disruption mutations constructed in S8 targeting ORFs MXAN4601, MXAN4762, and MXAN7041: The outlier subline is shown in gray; the parent subline is shown in blue; the mutant strain is shown as a cross of blue and gray. **(a)** Aggregate counts for S1, S8, and the mutant strain S8_4601. **(b)** Spore counts for S1, S8, and the mutant strain S8_4762. **(c)** Aggregate and spore counts for S9, S8, and the mutant strain S8_7041. Significance was determined using Tukey's multiple comparison test: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

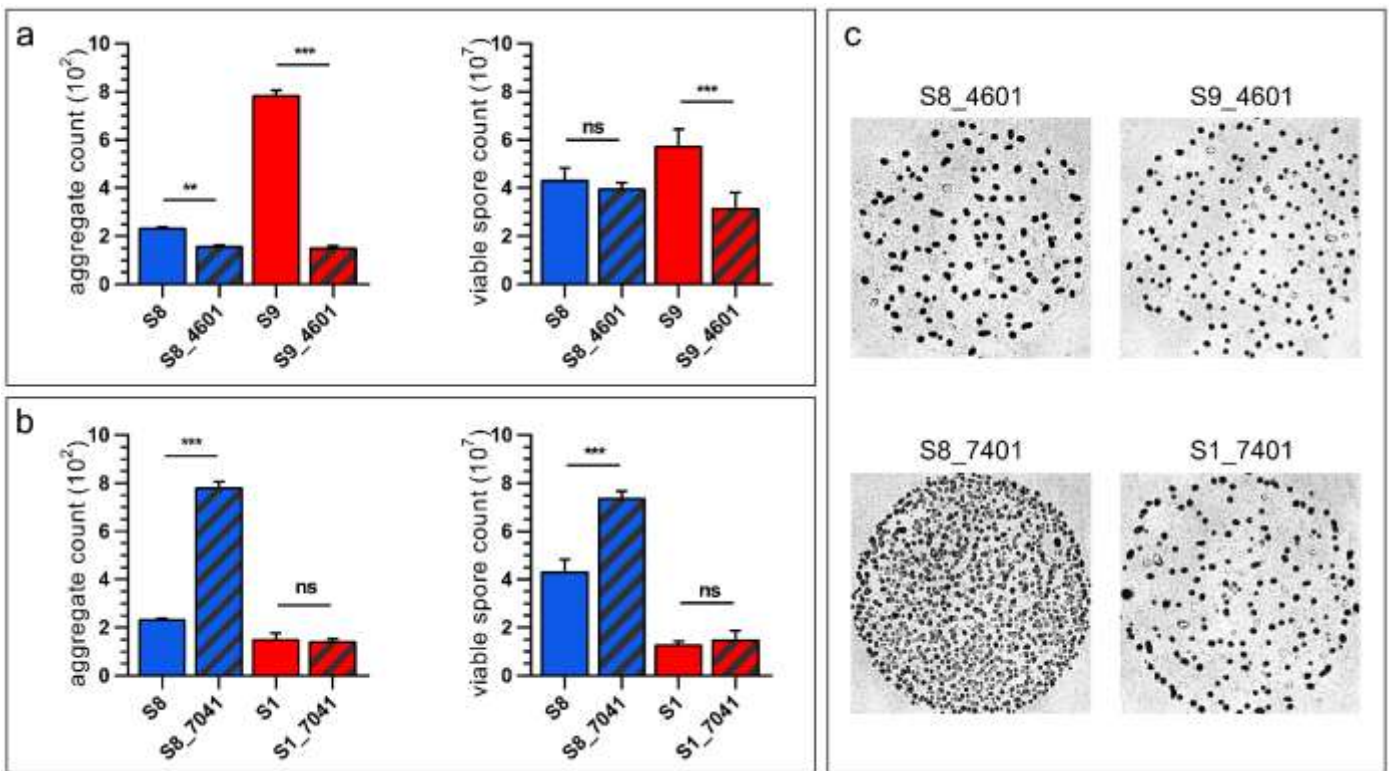


Fig. 4.3 Epistasis in S1, S8, and S9 mutant strains. Insertion-disruption mutations constructed in S1, S8, and S9 targeting ORFS MXAN4601 and MXAN7041: The parent sublines are shown as blue or red; the mutant strain specific to each parent subline is shown as a cross of blue and gray or red and gray. **(a)** Aggregate and spore counts for MXAN4601 disruptions in S8 and S9. **(b)** Aggregate and spore counts for MXAN7041 disruptions in S8 and S1. **(c)** Representative aggregate images of mutant strains. Significance was determined using Tukey's multiple comparison test: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

CONCLUDING REMARKS

Populating a G2P map is an important step towards understanding the most basic question in biology - what makes two organisms, A and B, different. Tremendous progress in the field of genetics has been made by studying genes in isolation, but genes by themselves do not cause a phenotype; genes require a suitable environment and multiple interactions between many other elements (such as gene-gene or network-network interactions) to produce an observable trait. With respect to phenotype, genes are best described as predictors rather than determinants.

The primary focus of this work has been to populate the genotype and phenotype spaces of single gene mutant strains and pseudo-isogenic wild type sublines of *M. xanthus*. Through a series of rigorous characterization experiments, we highlight the need for a statistically derived definition of phenotype combined with a more routine whole genome resequencing of genetic backgrounds. The conclusions presented here are possible because of our large mutant strain library, the availability of multiple reference strains, and a sequenced reference genome.

APPENDICES

Appendix I. Primers to construct ABC transporter mutants (Chapter 3).

ORF	Forward Primer	Reverse Primer
MXAN_0035	CTTCTTCCCCAACATCACCC	GTCTCCAGCACCTGTCGTGT
MXAN_0036	CTACCTGATGTCAGTGGCCC	GGAGACGAAGAGCAGCTCAC
MXAN_0037	GCTGCTGATGTTCCAGGAG	CATGGGCAGGTGGACTTC
MXAN_0107	TCTCGAAGTTCCCCTGGAC	GGTTGATGGCGAAGTCGT
MXAN_0108	TCCGACTGCTTCCAGGTG	CTCCTCCAACGGCTTCAG
MXAN_0146	CATCCAGGTCTCCGCCAC	AGGACGATGGGATTGAGTGC
MXAN_0249	CTCGTGTAACACACGGAGTC	CGATGATGAGTCCGAACACC
MXAN_0250	TATGACGACATGGACCTGGA	GTCACGACGATGGAAGTGC
MXAN_0251	GACGGGCATCGTCTTCTC	GTCTGGCCGTAGACGAG
MXAN_0553	CGTACCTGCACTTCATCGTG	GTTGAAGAGGCTGACGGTGT
MXAN_0554	GTCTTCGGCAAGGACCTG	CTCCACGTAGGTGAGGGTG
MXAN_0559	CAGGAGCCACAGCTCGAT	CTCTCCGACTTCTCCTCCAG
MXAN_0596	TGTGGCAAGTCAACACTGCT	ATGAACCGGGTGACGAAGT
MXAN_0597	TCGACGTGTACCCCGAGT	GCGTTGAGCTTCACCATGT
MXAN_0622	ATTGGCTACTTCAGCCAGGA	CTGCTGCTGCTTCTCGTTC
MXAN_0629	GCGAAGCTCTCTTTCGATTTGC	TTCATCAAGGAGCAGCACGTGG
MXAN_0684	ACCCTGCTGTCCATGATGA	ATGTGGTCGGAGTAGCAGGA
MXAN_0685	GTGCTCTTCCAGACGGTGAC	ATGTTGGCCACCAGCAAC
MXAN_0686	ATCATGCAGATGATTGCC	ATGACCATGCTGACCAGGTT
MXAN_0687	GACATGTTCCCCCAGCAC	CACCAATTCATTGTCCAGCA
MXAN_0696	ATGGCCCCAATTACCAGGACT	GTTGCCCCGTACCCGTCTC
MXAN_0721	GGAAGACGACCACCATCAAC	GACCTCCTCCACCTCGTACA
MXAN_0722	ATACAACTTCGTTCCGGTG	ACACGTACGAGGGTGGTAGC
MXAN_0748	GTTCTCTACGGACGGCTG	GACTTCACGGGCTGGTACA
MXAN_0751	CACCCCGGACAAGAAGAAG	GTCCCGCAGGTGGTAGTAGA
MXAN_0770	CACGTTGCTGGAGGAGGG	GTGTTCTTCGAGCCCTTGAG
MXAN_0771	CAAGACGTGGGGCCTGTT	ACCGTCACGGACACCAAC
MXAN_0772	GTGTGGCAAGACGACGAC	AGGAAGTACGCCACGAAGG
MXAN_0966	GTTCCCCACGGAGACATTC	GTCCAGCTTCAGGTGCTTCT
MXAN_0967	CCGTTTCATCCTCATGGT	GTCACCCAGCCATTGGAC
MXAN_0968	CGACGGACAGGACTTGCT	GGAAAGCACAGTTGATGCAG
MXAN_0995	CACACAAGACGGTGTTTCGAC	GACAAGGTGTTGTTACCCCC
MXAN_1060	GGAAGACGACCAGCTTCAAC	TGATGTACGCACGATCACAG
MXAN_1097	GTATGTGGTGGACCGCT	ATGAAGACGGAGGTCTGGAG
MXAN_1124	GGAAGACGGTGCTGATGAAG	GTGTTTCGATGATGCGTCCTT
MXAN_1151	CTCATCACCCAGGACAAGGT	GTCCGGCGAGTAGTGTTTC
MXAN_1153	GACTATCAGGCCATCCGCTA	GGAGTACTTGATGGTGGCGT

MXAN_1154	GTGACGGCTACGAGGTGAC	CTTGTCCGGGACGATGTAGT
MXAN_1155	GTGACGAGGCGTGGAAGTA	GAGCTTGTAGTGGTGGAGGC
MXAN_1262	GAGGGCTGTGGAGTGGTCTA	TCATTACGTAGCTGATGCC
MXAN_1286	CATCTTTCCCCGCTGGTC	ATAGGCGACCTCGTTCCAG
MXAN_1287	AGGTTTCAGGTCGTCGTTCC	CTCCGAACTCCAGCACGTC
MXAN_1288	GTTGGATGGGATTGCCTTCT	GTCGAACTGGAAGAAGTCCG
MXAN_1318	CTCCTGCCTCGTTTCGAGT	AAAATCATGATGGGCAGCTC
MXAN_1319	CCATCACCGAAACCGTCTT	CTGGCATGAGGATGAAGTCC
MXAN_1320	ACAAGCTGGAGTCCGTGC	CTTGAGGCGCTCGACATC
MXAN_1321	GTGCTTCTGGACGGCACT	ACTCCTGGGAAAGCACCTG
MXAN_1377	GGAAGTCCACCAGCATCG	CGGAGATGAGCTTCTCGAAC
MXAN_1547	CATGACGTGGTGAGCAATG	ACCTCCACCGAGCCAAAG
MXAN_1548	GGTGCTGTGCGAGTTCTTCT	GGGAAGTGCTGATTCCACAG
MXAN_1597	AGACGACGACCGTGGAAT	CTCCAACCTCGATGACCTGCT
MXAN_1598	GTTCCGGAATGACTCCCTG	CAATGTGAAGTAGCTGCCGA
MXAN_1604	CGAAGTCTCATCGAGAGCG	AGGAAGCTGCCCCGAGAC
MXAN_1605	CTACTTCAAGGAGCATCCCG	CCCAGCTTGTGCGCGTAG
MXAN_1695	GTTCTCGGACCCCAACTTC	CAAGCAGGTAGAGCGTGGAC
MXAN_1788	GTGCTCTCGCTGCTCAAAC	GTCGGCAATGAACGCTCC
MXAN_1789	GTCACCTTGATGAAGTCGC	GGGCGAACGGACCAATAG
MXAN_1790	ATCACGCCACTCTACCACG	CTTCAGCAGCGTTCCGAG
MXAN_1791	GAGCAGCTACTCGACCGC	TTGTTGTAGTCGATCTGCCG
MXAN_1792	ATTCTGGGACAAGATCCG	CACCAGCACGTACCCGAC
MXAN_2018	GTGACTTCGTCGTGGGCT	CAGGAAGTACGTGGACAGCA
MXAN_2019	CGTGGGAGAAGCTCTTTTCG	GTAGAGCTGGCCGAAGTACG
MXAN_2020	CCCTGTCCAGCACCTGTC	AAGAAGCTCACGGTGGACTC
MXAN_2078	CCGTATGCCTCCGTCATC	CCCTTCCTTCACCTGGCT
MXAN_2249	CTGCCTCAACCGCCTCAT	GTCCACGTGCTCCACGAAG
MXAN_2250	CCAGCTACGCCATCATCACGG	ACCACCATGGACAGCGACAGC
MXAN_2251	ATGACACATCTGGTCCAGGC	CGTGTGGATGGACTCCTTG
MXAN_2268	GACTTCTTCCCGATGCTGG	ACGGAGACTTCCCCCTCAC
MXAN_2407	GAGCTGTCCGTGTCCAG	GCCACTTCTGGGCAATC
MXAN_2428	CGGTCGTCACCGGCTATC	CCTCACGGTTGACGGTCT
MXAN_2429	TCGGATGACTCCAAGGACC	GAAGAACTTGACAGCAGCA
MXAN_2430	CCGGTGAGTTCATCTCCATC	GAGAAGATGAAGGTGGTGCC
MXAN_2654	TGTCGTCTACCTGGCCTACC	GTTTCAGCGTGTCCAGGAACT
MXAN_2783	GTGCAGGAGTCGGTGAGC	TCGTCGTGAGGTGGTCCT
MXAN_2795	TTTGTGCAACAGTTCATGCC	AGGTGGCTCGGGTGAAGAT
MXAN_2831	GAATTCGGCGGCTTCTATG	GTAGGTCCGCCACCCCTC
MXAN_2832	GAGGCCCTCTCGCTCCTC	GATCCGTGCGCAGTTGAC
MXAN_2833	ACATCTCCTTACCCCCGAC	CTTCGGAAATGGAGTGGGT
MXAN_2853	CTGTCTGGCCATGGTGCT	GACGTGAGCAGCTCGAAGA
MXAN_2948	TCCTGAGCGGTACGTGGT	CCGCAGCTCCAACATCAG

MXAN_2949	CAAGAGCACCTGGTTCGG	GGTCCACGAAGACGAGCAC
MXAN_2951	ACTTCGTGGATGCCAAGC	TGGTGTCCGGGTAGTAGCTC
MXAN_3208	GATGTGCGGCTACCTCTACC	GTGAGCTCCGGGATGATG
MXAN_3209	CTACAGCATGTTCTCCGCGT	AGAAGACGAACAGCAGGATGA
MXAN_3256	CAGCTACCTCATGAAGGCGA	GAACCAGAGGGTGAGGAACA
MXAN_3257	ACGCTGCTGATGTTCTCCTT	GTCTCCCTGAAGTACGAGCG
MXAN_3258	ACAACGGTTCGGGAAGAC	CTTGAGTTCGGCGATGATG
MXAN_3339	TGCAATACAGCCGCAAGG	TCAACAGCTCCTTCTCCTCC
MXAN_3452	GGTACTTCGAATCCGTGGC	CGTGAAGGTGCCCATGTAG
MXAN_3648	AGGTGTTACTGCGGCCGTT	AGCAGGCTCCTCCGTGAA
MXAN_3650	CCTACGAGGTGCTGCCTTC	GAGGTTTGCCAGGTTGACAC
MXAN_3717	AACTGGACTTCGTGGGTGTG	AAGCAGCAGATGAGGGACAC
MXAN_3718	ATGAAGCACATGATTGGCCT	ATGTCATGGCTGATGACCAC
MXAN_3719	GCGAAGGTCTGGTTGAAGAT	CAGCTTCTTGAGCGTCTCCT
MXAN_3745	CACTGACGGTCGTCCAATC	TTCTCCTTCTCCAAGTGCCT
MXAN_3773	GGTTAGTCTCGTCGAGGTGC	CTTTCATCGAGCGGAATGAC
MXAN_3908	GAAGACGACGCTGGTGAAG	CTGGGACACGTCCAGGAG
MXAN_3909	GCTGGTGCAATTCGGCTAC	GAAGATGCCCGTCTTCACC
MXAN_3910	CGTGTGGCTGGTGATGTC	GCCAGCGGAATGTAGAGC
MXAN_3911	ACGAAGGAGCTGACGTTTCAT	TAGTCGTGGGACACCACCAT
MXAN_3911	ACGAAGGAGCTGACGTTTCAT	GGCTCGTCGAAGAGGAGATA
MXAN_3912	AGGACATCCAGCTCCAACC	GAGGCATCGTCCAACAGC
MXAN_3986	AGGGACACCTTCAGTTCGAG	CGTCGTTGTCGTGTTTGTTT
MXAN_4074	GGCACCATGACCTTCGAC	GCGATGACCTCCAAGTGTTC
MXAN_4102	GGGTCCATCCGGGTACTG	CTCGACGACGTATTGTCCCT
MXAN_4172	ACCTTCCAGGTGACGAAGTG	GAACAGCTCCGGGGACTT
MXAN_4173	GATCAAGGGCGACTGGTG	CGGTTGATGGAGAAGTCGTC
MXAN_4174	GCAAGAGCACGATGATGAAC	ACAATCTCTCCGTCGCTCAG
MXAN_4175	ACGGTGAAGATCTCGTCCAG	AGTACGTCTGGCCCTCCAGT
MXAN_4176	AGGAGTTCACCACGGTTCC	CAGCTCCAGCGTCCTGAG
MXAN_4177	GGTCTGGTCAGCTGATGG	AAGGCCGATGAGGAAGATG
MXAN_4198	GAGCCTACCTCCCCAACCT	AAGGTGGACACGTCCTGG
MXAN_4199	ACCTTCATGAACCTCATCGG	GAAGAGGGCCATGATCTCCT
MXAN_4200	TGAAGGTCCACTACGACACG	GAAGAGGGTGACGTTCTGCT
MXAN_4201	GATGACCATCGACGACGC	GTCTCCTGCATGGTGTGAG
MXAN_4523	GGAGCATTCGGACATCGAC	CTCTCCGCCAGGGAAGTAT
MXAN_4586	ACTTCCGCTACGAGGACAAC	GCTGGGCAATCTTCTTCTTC
MXAN_4622	ACCATCCCCAAGGGTAAGAC	GAAATGGCCTCACGGTACTC
MXAN_4623	CTGGTGTACGTGCTGCTGTT	GTGCTCGTAGAAGATGGCCT
MXAN_4664	GTTTCGACGGACAGGAATC	ACAACAACGCCTGGGTGTA
MXAN_4665	GGAGAGGTGCGCTTCCAG	CGCGTACATCACCACCAC
MXAN_4716	GCTGGTGATGGTCATCGTT	AGCACCTCCATCAGCCCT
MXAN_4729	AGCACCTTCCTGCACGTC	GTCTCGTTGTGGGTGACGAC

MXAN_4730	GGTTCGAGCTCTTTGTCGC	GGATGTTCTTGGGCAGGTC
MXAN_4749	AGGTGGTGCAACTCTTCCC	CGTCTCCTCGAACGGAAC
MXAN_4750	ATATGACTTGTCGCGGACG	ACAACTCCTCCACCAGCG
MXAN_4787	GTCCTCATCGTCATGCTCAC	GACGTGGGTGTCGATGGT
MXAN_4788	CAACATCAAGGTCCAGGTCACCG	CGATCTCCTCGTTGCCCTTCTTC
MXAN_4789	CGCTCATCCTCATCATCGTC	CCATGAGGACAATCATCGTC
MXAN_4790	ATCTTCCCCGCCATCTTC	ATGAACTGGGAGTCCAGGTG
MXAN_4791	GAACGACCTCATCTCCGGCACC	GTGACGATGGCGATGGTGTACG
MXAN_4792	TCATCGTCCAGAGCATGAAG	ATGTAGACGACCATCTCCGC
MXAN_4818	GTGGGGGTGGTCTTCCAG	GTCATGCGGACGGACAAAG
MXAN_4819	TTCCAGAAGGGCTGGGAG	TTGTAGAGAATCTCCGCGTG
MXAN_4820	CTGAGCTGGGAGGAGTTCTG	ATGTTGCCGAGATGAAGAC
MXAN_4821	GTCATCGAAGGACTGCGG	GTCCGAATAGAGGTACGCCA
MXAN_4878	GTCCACCACCATGAAAATCC	TGGATGATGAGCACCTTCTG
MXAN_4879	CCCGGACTTCAACATGTACC	CTTGATCATCCCCTGCAACT
MXAN_4880	CAGCTTCTTCGCCACCAC	GACACCCGCGAGGTTCTAG
MXAN_5167	GGGGTGTGTTTCATCCAGTC	GTACCAGAAGACGAGCAGGG
MXAN_5168	GAGCGGAGCGAGACATCC	CTCAAGGGCCCCAGGTAG
MXAN_5183	GACACACCCTTGGAAGCAAC	CCAGGAAGGTGGCGTAGAG
MXAN_5275	GGCATCCACGTCATCAAGACCT	CCTGGAACACCATCGAGAGGTG
MXAN_5276	GATCAGAGCGATGTCCAACA	CAGGGTGCTTCCTGCGTA
MXAN_5315	TGCCTGTAAGAAGGAGGAGC	GAAGTACGGTTCTCGTTGC
MXAN_5316	CTCTTCTGGTTCGTCATCGG	CTCCACGGAGTAGCGCAT
MXAN_5317	AAGTCCACCACCTTCCAGG	CCTCCACGGAGAGGATGTC
MXAN_5377	GCGATGGTGTTCCAATCCT	GGACAGGGCAGGGTGAAG
MXAN_5378	GCCCGGACGTCTTCATCT	GGGATGTTCTGTTGCTC
MXAN_5379	GAAACCGTCACGCTGGAG	ACTCCACCTTGTGCTCGC
MXAN_5380	CTCTACTTCGGCCAGCAAGA	GAATACCCGACGAAGCATGT
MXAN_5419	CGTACCGGGGACTTCTACTG	AAGAGCTCGAAGCCCATCTT
MXAN_5502	GCATCAACGCCGTCTTCT	GTCCATCAGCAGCACCTCC
MXAN_5503	CGTTTCGTCCAGAACGACT	GTCCGTCCAGTCGACCAT
MXAN_5535	CATCGACCCACTGGTTTTCT	ACTACTGAAACCCGCTCCTG
MXAN_5583	CTCTTCTACGACCCACGCT	CTGGAGCAGGTAGTAGCCGA
MXAN_5584	GGGAAGACGACGACGGTG	GAGACAAAAGCGCTCCGC
MXAN_5698	GTCATCCGGGAGAAGGTCT	GTAGTTCTCCGAGTCCACGC
MXAN_5699	AAGTCCACGCTGCTCCAC	CACAACACGGTGAGTCCTTC
MXAN_5702	GGTGCTCATCGGCTTCAG	CTCCACGTCACCCACAG
MXAN_5711	CTACGAGGGCGGCAAGTA	CCCTTGAGATGAGGTCCTG
MXAN_5712	GGAGAACGTGGTGCTGGG	GATGCCGACGATTTACC
MXAN_5713	GTTCAAGGTGGGCCTGTTT	GTGTCAGCCACACCCACAC
MXAN_5714	GGCATGGAGGGGATGATG	CTGCTCGAAACGGTCCAGTA
MXAN_5747	GGTGACGGTGTATCTGTCGC	AACAGAAAGGCGAAGAGCG
MXAN_5748	GACGCTGCTGAAGCTCATCT	GTCCTCGTCGGACACAATCT

MXAN_5780	ATGAGCCTGCTGCTTGTCTT	GTCACCTCGATGCCATAGGT
MXAN_5781	CCATCAAGATTCTCATGGGG	GAATCCCAGAGACTGGAGCA
MXAN_5847	GCCCTCTACCTGTCTTCGC	GCCGTAGAACTCCAGCGAC
MXAN_5978	CGCCAAGGATTTGCTCTCT	TGAAGATGTCCTTCAGCAGC
MXAN_6000	TGTCCGCCTACAACAGCA	GTCTCCAGGGCCTTGTCTC
MXAN_6001	TGTTCTGCTGCTACGGTGAG	CGTACGTGAACACCATCGTC
MXAN_6002	GAGGCGGACTTCATCAGCTA	GA CTCCTCCCAGCCATAGG
MXAN_6003	ACGACGCTCAAGGTCCTCT	GTCCACTTCCTCCGACAGG
MXAN_6042	GTGCTGGGCTTCTCTGACTT	CCTCGTAGAGCTGGTCATCC
MXAN_6402	CAAGTCCAGCCTGATGAACA	CGTCCTTGAGCAGGAGTTTC
MXAN_6403	GCGCAATACATGTTCTCCAA	GACAGTTCAGCCACGTCTT
MXAN_6456	CACCAGCACGGACGTGAG	GTCTGGTCCGCAATCTCCT
MXAN_6474	CGTCTTCTGGGTGTGCTTC	TGCCCACCAGATAGAGGAAC
MXAN_6475	AGGTGCTACTGGACGGACAC	CAGTAGACGCCCCTGCTC
MXAN_6518	GTGAAGACCGTGGTGCAGT	TTCTTCTTCACCTTCGCCAG
MXAN_6551	CGTCTACGAGTGCCTTGA AA	CAGTCCACTTCCCCCTCCAC
MXAN_6552	CAGTACGCCACGTACCTGAA	G TAGCGCAGCATCACCTG
MXAN_6553	GTGATGGACCCACGGAG	GGCCCAGTAGAGGTTGGTG
MXAN_6554	GACGTGCTCACCCTGGAG	GGGTGTTGAGGAAGTCCG
MXAN_6568	CGTCCACCTGGGACTGCTA	ACGTCCTGTGAGAGCTGGTT
MXAN_6569	GGAGAAGTCCGCCTGCTG	G TAGAGCGTTTCCAGGGAGC
MXAN_6575	GAGCAGTCCATCTCCCTCAC	GGCGAGCCAGTAGAGGATTT
MXAN_6576	CTGTGCGGAGATGGTGCTGT	ACTTGTCTTGCCACGTC
MXAN_6643	GTGTTGTTGGACAGCGTTGC	CACCCACAACGACAACCC
MXAN_6644	GGTGGAGACGGTGCTGAC	AAGATGGCCAGGGAGAGC
MXAN_6645	CGGTGGAGTTCTCCCTGG	CTCAGGCACCTGGAAGGC
MXAN_6661	GCTGAAGGGTGTGTGCT	CAACACATAGCCGTAGTGCG
MXAN_6662	GTGTACCAGCCCACCCAG	CATGACCAGCTTCATGTCTG
MXAN_6663	GTCCTGTGCTGAACCTCA	AGCAGCGTGAGGAACGTG
MXAN_6664	GTCTACATGGGCTACGCCAC	CTTGAGGCCCACGTACAGG
MXAN_6665	CACCTTCGGCATCTCCAC	ATGCCCACCTCGCTGTAGTA
MXAN_6765	GCTGCTCAACCTCATCAGTG	CAGGTGTTGCCCTTCTGG
MXAN_6766	CGGTATACCAGGCGCTGC	ATGAGGAACATCCCCACCAC
MXAN_6826	CTGATGGCCTTCCTGGAG	CTTCAACCCAGGTCCTC
MXAN_6827	ATGCTCTACACGCTGGTGC	ATCACGTGGCTGGAGAACAC
MXAN_6934	CTGGTGGGGTTGGAGTTC	ATAGAAGGCGCCCAATGAG
MXAN_7114	ACGCTGGGTCTACACGTCC	CTCTCCACCTTCTGGAGCAC
MXAN_7115	CCTTCACGGTGCTGTTAC	GAAGGGACGGGACGACAG
MXAN_7144	TCATCTCCCTGAGGAACGTC	GGAAGACCTCCATGATTGTC
MXAN_7145	AGACCGAGGACGTGGAGAC	AGTTCCGTCAAGAGCTGCAT
MXAN_7146	CTCGAGGACATCTCCCTGTC	GAGAAGGACCAACCCACCTT
MXAN_7147	GAGTACATCCGCTGGCTCAC	CTTCAACACCTTCGCCAG
MXAN_7225	GGTGAACGACATGTCACCG	ACGAAGCCCCAACACCAC

MXAN_7226	GCTGATGAAGCACATCATGG	GTGCTGTGACTCACGGAAGG
MXAN_7293	CGTTCCTGACCTTCTTCCTG	GGAACGTCAGCGCATACAG
MXAN_7294	ACCGCTTCATCCTCGTCTTT	GTCAGCACGTAGATGGCGT
MXAN_7295	CTGGCGCACCAGTTCTTC	CCAGTTGCGCATGAACAC

Appendix II. Source of *M. xanthus* sublines (Chapter 4).

Subline	Source
S1	University of Georgia
S2	Augustana College
S3	Michigan St. University
S4	Syracuse University
S5	University of Iowa
S6	Wayne State University
S7	University of California, Davis
S8	Syracuse University
S9	University of Wyoming

Appendix III. Pairwise comparison of A-motility phenotypes (Chapter 4).

Tukey's MCT	Mean Diff.	95% CI of diff.	Summary	P Value
S1 vs. S7	0.1167	-0.4990 to 0.7323	ns	0.9993
S1 vs. S8	-0.2	-0.8157 to 0.4157	ns	0.9748
S1 vs. S9	0.09167	-0.5240 to 0.7073	ns	0.9999
S2 vs. S1	-0.1083	-0.7240 to 0.5073	ns	0.9996
S2 vs. S7	0.008333	-0.6073 to 0.6240	ns	> 0.9999
S2 vs. S8	-0.3083	-0.9240 to 0.3073	ns	0.7705
S2 vs. S9	-0.01667	-0.6323 to 0.5990	ns	> 0.9999
S3 vs. S1	1.907E-07	-0.6157 to 0.6157	ns	> 0.9999
S3 vs. S2	0.1083	-0.5073 to 0.7240	ns	0.9996
S3 vs. S7	0.1167	-0.4990 to 0.7323	ns	0.9993
S3 vs. S8	-0.2	-0.8157 to 0.4157	ns	0.9748
S3 vs. S9	0.09167	-0.5240 to 0.7073	ns	0.9999
S4 vs. S1	0.1833	-0.4323 to 0.7990	ns	0.9853
S4 vs. S2	0.2917	-0.3240 to 0.9073	ns	0.8182
S4 vs. S3	0.1833	-0.4323 to 0.7990	ns	0.9853
S4 vs. S5	-0.05	-0.6657 to 0.5657	ns	> 0.9999
S4 vs. S6	-0.5167	-1.132 to 0.09901	ns	0.1611
S4 vs. S7	0.3	-0.3157 to 0.9157	ns	0.795
S4 vs. S8	-0.01667	-0.6323 to 0.5990	ns	> 0.9999
S4 vs. S9	0.275	-0.3407 to 0.8907	ns	0.8605
S5 vs. S1	0.2333	-0.3823 to 0.8490	ns	0.9392
S5 vs. S2	0.3417	-0.2740 to 0.9573	ns	0.6631
S5 vs. S3	0.2333	-0.3823 to 0.8490	ns	0.9392
S5 vs. S7	0.35	-0.2657 to 0.9657	ns	0.6346
S5 vs. S8	0.03333	-0.5823 to 0.6490	ns	> 0.9999
S5 vs. S9	0.325	-0.2907 to 0.9407	ns	0.7184
S6 vs. S1	0.7	0.08433 to 1.316	*	0.016
S6 vs. S2	0.8083	0.1927 to 1.424	**	0.0033
S6 vs. S3	0.7	0.08433 to 1.316	*	0.016
S6 vs. S5	0.4667	-0.1490 to 1.082	ns	0.266
S6 vs. S7	0.8167	0.2010 to 1.432	**	0.0029
S6 vs. S8	0.5	-0.1157 to 1.116	ns	0.1918
S6 vs. S9	0.7917	0.1760 to 1.407	**	0.0042
S7 vs. S8	-0.3167	-0.9323 to 0.2990	ns	0.745
S7 vs. S9	-0.025	-0.6407 to 0.5907	ns	> 0.9999
S9 vs. S8	-0.2917	-0.9073 to 0.3240	ns	0.8182

Appendix IV. Pairwise comparison of S-motility phenotypes (Chapter 4).

Tukey's MCT	Mean Diff.	95% CI of diff.	Summary	P Value
S1 vs. S7	-0.3833	-2.146 to 1.380	ns	0.9982
S1 vs. S8	-0.7	-2.463 to 1.063	ns	0.9221
S1 vs. S9	1.042	-0.7213 to 2.805	ns	0.5869
S2 vs. S1	-0.04167	-1.805 to 1.721	ns	> 0.9999
S2 vs. S7	-0.425	-2.188 to 1.338	ns	0.9963
S2 vs. S8	-0.7417	-2.505 to 1.021	ns	0.8954
S2 vs. S9	1	-0.7630 to 2.763	ns	0.6373
S3 vs. S1	0.425	-1.338 to 2.188	ns	0.9963
S3 vs. S2	0.4667	-1.296 to 2.230	ns	0.9931
S3 vs. S7	0.04167	-1.721 to 1.805	ns	> 0.9999
S3 vs. S8	-0.275	-2.038 to 1.488	ns	0.9998
S3 vs. S9	1.467	-0.2963 to 3.230	ns	0.1689
S4 vs. S1	0.7583	-1.005 to 2.521	ns	0.8834
S4 vs. S2	0.8	-0.9630 to 2.563	ns	0.85
S4 vs. S3	0.3333	-1.430 to 2.096	ns	0.9993
S4 vs. S5	1.475	-0.2880 to 3.238	ns	0.1638
S4 vs. S6	-0.06667	-1.830 to 1.696	ns	> 0.9999
S4 vs. S7	0.375	-1.388 to 2.138	ns	0.9985
S4 vs. S8	0.05833	-1.705 to 1.821	ns	> 0.9999
S4 vs. S9	1.8	0.03700 to 3.563	*	0.0423
S5 vs. S1	-0.7167	-2.480 to 1.046	ns	0.912
S5 vs. S2	-0.675	-2.438 to 1.088	ns	0.9357
S5 vs. S3	-1.142	-2.905 to 0.6213	ns	0.467
S5 vs. S7	-1.1	-2.863 to 0.6630	ns	0.5163
S5 vs. S8	-1.417	-3.180 to 0.3463	ns	0.2024
S5 vs. S9	0.325	-1.438 to 2.088	ns	0.9995
S6 vs. S1	0.825	-0.9380 to 2.588	ns	0.8277
S6 vs. S2	0.8667	-0.8963 to 2.630	ns	0.7873
S6 vs. S3	0.4	-1.363 to 2.163	ns	0.9976
S6 vs. S5	1.542	-0.2213 to 3.305	ns	0.1269
S6 vs. S7	0.4417	-1.321 to 2.205	ns	0.9952
S6 vs. S8	0.125	-1.638 to 1.888	ns	> 0.9999
S6 vs. S9	1.867	0.1037 to 3.630	*	0.0311
S7 vs. S8	-0.3167	-2.080 to 1.446	ns	0.9995
S7 vs. S9	1.425	-0.3380 to 3.188	ns	0.1965
S9 vs. S8	-1.742	-3.505 to 0.02133	ns	0.055

Appendix V. Pairwise comparison of aggregation phenotypes (Chapter 4).

Tukey's MCT	Mean Diff.	95% CI of diff.	Summary	P Value
S1 vs. S5	-0.2194	-0.3456 to -0.09327	***	0.0003
S1 vs. S6	-0.406	-0.5321 to -0.2798	****	< 0.0001
S1 vs. S7	-0.3077	-0.4338 to -0.1815	****	< 0.0001
S1 vs. S8	-0.1961	-0.3222 to -0.06990	***	0.0009
S1 vs. S9	-0.7275	-0.8536 to -0.6013	****	< 0.0001
S2 vs. S1	-0.01648	-0.1426 to 0.1097	ns	> 0.9999
S2 vs. S5	-0.2359	-0.3621 to -0.1098	***	0.0001
S2 vs. S6	-0.4225	-0.5486 to -0.2963	****	< 0.0001
S2 vs. S7	-0.3241	-0.4503 to -0.1980	****	< 0.0001
S2 vs. S8	-0.2125	-0.3387 to -0.08638	***	0.0004
S2 vs. S9	-0.744	-0.8701 to -0.6178	****	< 0.0001
S3 vs. S1	-0.108	-0.2342 to 0.01816	ns	0.1288
S3 vs. S2	-0.09152	-0.2177 to 0.03464	ns	0.2759
S3 vs. S4	-0.05305	-0.1792 to 0.07311	ns	0.8536
S3 vs. S5	-0.3274	-0.4536 to -0.2013	****	< 0.0001
S3 vs. S6	-0.514	-0.6401 to -0.3878	****	< 0.0001
S3 vs. S7	-0.4157	-0.5418 to -0.2895	****	< 0.0001
S3 vs. S8	-0.3041	-0.4302 to -0.1779	****	< 0.0001
S3 vs. S9	-0.8355	-0.9616 to -0.7093	****	< 0.0001
S4 vs. S1	-0.05495	-0.1811 to 0.07121	ns	0.8296
S4 vs. S2	-0.03847	-0.1646 to 0.08769	ns	0.9719
S4 vs. S5	-0.2744	-0.4005 to -0.1482	****	< 0.0001
S4 vs. S6	-0.4609	-0.5871 to -0.3348	****	< 0.0001
S4 vs. S7	-0.3626	-0.4888 to -0.2365	****	< 0.0001
S4 vs. S8	-0.251	-0.3772 to -0.1249	****	< 0.0001
S4 vs. S9	-0.7824	-0.9086 to -0.6563	****	< 0.0001
S5 vs. S6	-0.1866	-0.3127 to -0.06039	**	0.0016
S5 vs. S7	-0.08824	-0.2144 to 0.03792	ns	0.3161
S5 vs. S9	-0.508	-0.6342 to -0.3819	****	< 0.0001
S6 vs. S9	-0.3215	-0.4476 to -0.1953	****	< 0.0001
S7 vs. S6	-0.09831	-0.2245 to 0.02785	ns	0.2045
S7 vs. S9	-0.4198	-0.5460 to -0.2936	****	< 0.0001
S8 vs. S5	-0.02337	-0.1495 to 0.1028	ns	0.9989
S8 vs. S6	-0.2099	-0.3361 to -0.08376	***	0.0004
S8 vs. S7	-0.1116	-0.2378 to 0.01455	ns	0.1075
S8 vs. S9	-0.5314	-0.6576 to -0.4052	****	< 0.0001

Appendix VI. Pairwise comparison of sporulation phenotypes (Chapter 4).

Tukey's MCT	Mean Diff.	95% CI of diff.	Summary	P Value
S1 vs. S2	-0.8673	-2.142 to 0.4070	ns	0.3475
S1 vs. S3	-0.8081	-2.082 to 0.4662	ns	0.4326
S1 vs. S4	-1.325	-2.599 to -0.05078	*	0.038
S1 vs. S5	-1.276	-2.550 to -0.001668	*	0.0496
S1 vs. S6	-2.311	-3.586 to -1.037	***	0.0002
S1 vs. S7	-2.529	-3.803 to -1.254	****	< 0.0001
S1 vs. S8	-3.038	-4.312 to -1.763	****	< 0.0001
S1 vs. S9	-4.452	-5.727 to -3.178	****	< 0.0001
S2 vs. S3	0.05922	-1.215 to 1.334	ns	> 0.9999
S2 vs. S4	-0.4578	-1.732 to 0.8166	ns	0.931
S2 vs. S5	-0.4087	-1.683 to 0.8657	ns	0.9625
S2 vs. S6	-1.444	-2.718 to -0.1696	*	0.0197
S2 vs. S7	-1.661	-2.936 to -0.3870	**	0.0057
S2 vs. S8	-2.17	-3.445 to -0.8961	***	0.0003
S2 vs. S9	-3.585	-4.859 to -2.311	****	< 0.0001
S3 vs. S4	-0.517	-1.791 to 0.7573	ns	0.8753
S3 vs. S5	-0.4679	-1.742 to 0.8064	ns	0.9229
S3 vs. S6	-1.503	-2.777 to -0.2288	*	0.0141
S3 vs. S7	-1.721	-2.995 to -0.4462	**	0.0041
S3 vs. S8	-2.23	-3.504 to -0.9553	***	0.0002
S3 vs. S9	-3.644	-4.918 to -2.370	****	< 0.0001
S4 vs. S5	0.04911	-1.225 to 1.323	ns	> 0.9999
S4 vs. S6	-0.9861	-2.260 to 0.2882	ns	0.211
S4 vs. S7	-1.204	-2.478 to 0.07078	ns	0.0728
S4 vs. S8	-1.713	-2.987 to -0.4383	**	0.0043
S4 vs. S9	-3.127	-4.401 to -1.853	****	< 0.0001
S5 vs. S6	-1.035	-2.310 to 0.2391	ns	0.1683
S5 vs. S7	-1.253	-2.527 to 0.02167	ns	0.0561
S5 vs. S8	-1.762	-3.036 to -0.4874	**	0.0032
S5 vs. S9	-3.176	-4.451 to -1.902	****	< 0.0001
S6 vs. S7	-0.2174	-1.492 to 1.057	ns	0.9994
S6 vs. S8	-0.7266	-2.001 to 0.5478	ns	0.5631
S6 vs. S9	-2.141	-3.415 to -0.8667	***	0.0004
S7 vs. S8	-0.5091	-1.783 to 0.7652	ns	0.8839
S7 vs. S9	-1.924	-3.198 to -0.6492	**	0.0013
S8 vs. S9	-1.414	-2.689 to -0.1401	*	0.0232

Appendix VII. Summary of read mappings (Chapter 4).

Library	Mapped read count	Average length (bp)
S1	6,897,065	126.88
S2	12,922,299	128.15
S3	6,872,636	128.98
S4	6,199,545	126.16
S5	7,266,263	129.7
S6	6,433,845	126.76
S7	8,677,798	126.24
S8	9,871,644	126.64
S9	8,199,433	127.55

Appendix VIII. Primers used to generate DNA inserts for plasmids (Chapter 4).

Target ORF	Primer Sequence	Plasmid
MXAN4601	F: CGAATACAACGTGGACCTCT	pS1.a
	R: AGGGAGTTGAACTGGAGGAT	
	C: GGAGAGCAAAGACTTTACCG	
MXAN4762	F: TTCGACGGTGACTTTCTACC	pS1.b
	R: GCCAAAAGTCACCGTCTTG	
	C: GACCAGATAGGGAAGTACGTG	
MXAN7041	F: ATCCTCTCGCAGGTGAGC	pS9.a
	R: ATTGAAGAGCGGGTTGCT	
	C: GACATGAAGGACCTCTACCAAC	

ORF, open reading frame; F, forward; R, reverse; C, confirmation.

Appendix IX. Plasmids used to construct mutant strains (Chapter 4).

Mutant strain	Parent subline	Plasmid	Source
S1_7041	S1	pS9.a	This study
S8_4601	S8	pS1.a	This study
S8_4762	S8	pS1.b	This study
S8_7041	S8	pS9.a	This study
S9_4601	S9	pS1.a	This study

REFERENCES

1. Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J. Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science*. 2003;299: 262–265.
2. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, et al. A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A*. 1988;85: 6622–6626.
3. Olsen LC, Aasland R, Wittwer CU, Krokan HE, Helland DE. Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA repair enzyme. *EMBO J*. 1989;8: 3121–3125.
4. Traag BA, Pugliese A, Eisen JA, Losick R. Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*. *J Bacteriol*. 2013;195: 253–260.
5. Miller MB, Bassler BL. Quorum sensing in bacteria. *Annu Rev Microbiol*. 2001;55: 165–199.
6. Gerth U, Krüger E, Derré I, Msadek T, Hecker M. Stress induction of the *Bacillus subtilis* clpP gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance. *Mol Microbiol*. 1998;28: 787–802.
7. Kosloff M, Kolodny R. Sequence-similar, structure-dissimilar protein pairs in the PDB. *Proteins*. 2008;71: 891–902.
8. Henikoff S, Greene EA, Pietrokovski S, Bork P, Attwood TK, Hood L. Gene families: the taxonomy of protein paralogs and chimeras. *Science*. 1997;278: 609–614.
9. Bork P. Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res*. 2000;10: 398–400.
10. Galperin MY, Koonin EV. Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol*. 2000;18: 609–613.
11. Nelson KE, Paulsen IT, Heidelberg JF, Fraser CM. Status of genome projects for nonpathogenic bacteria and archaea. *Nat Biotechnol*. 2000;18: 1049–1054.
12. Nambu T, Minamino T, Macnab RM, Kutsukake K. Peptidoglycan-hydrolyzing activity of the FlgJ protein, essential for flagellar rod formation in *Salmonella typhimurium*. *J Bacteriol*. 1999;181: 1555–1561.
13. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, et al. What is a gene, post-ENCODE? History and updated definition. *Genome Res*. 2007;17: 669–681.
14. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci*. 2012;69: 3613–3634.
15. Balakirev ES, Ayala FJ. Pseudogenes: Are They “Junk” or Functional DNA? *Annu Rev Genet*. 2003;37: 123–151.
16. Zheng Y, Josefowicz S, Chaudhry A, Peng XP, Forbush K, Rudensky AY. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. *Nature*. 2010;463: 808–812.
17. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10: 155–159.
18. Dawkins R. The Gene as the Unit of Selection. Available: <https://web.natur.cuni.cz/filosof/markos/Publikace/Dawkins%20extended.pdf>
19. Johannsen W. The genotype conception of heredity. 1911. *Int J Epidemiol*. 2014;43: 989–1000.
20. Benfey PN, Mitchell-Olds T. From genotype to phenotype: systems biology meets natural variation. *Science*. 2008;320: 495–497.

21. Holland MM, Fisher DL, Mitchell LG, Rodriguez WC, Canik JJ, Merrill CR, et al. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J Forensic Sci.* 1993;38: 542–553.
22. Kasai K, Nakamura Y, White R. Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic *Science. J Forensic Sci.* 1990;35: 1196–1200.
23. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005;437: 376–380.
24. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol.* 2008;26: 553–560.
25. Lovley DR. Cleaning up with genomics: applying molecular biology to bioremediation. *Nat Rev Microbiol.* 2003;1: 35–44.
26. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet.* 1999;21: 33–37.
27. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* 1998;393: 537–544.
28. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science.* 1995;269: 496–512.
29. Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, et al. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J Gen Virol.* 2000;81: 67–74.
30. Triques K, Bourgeois A, Vidal N, Mpoudi-Ngole E, Mulanga-Kabeya C, Nzilambi N, et al. Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. *AIDS Res Hum Retroviruses.* 2000;16: 139–151.
31. Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, et al. The Genome sequence of the SARS-associated coronavirus. *Science.* 2003;300: 1399–1404.
32. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409: 860–921.
33. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, et al. Comparative genomics of the eukaryotes. *Science.* 2000;287: 2204–2215.
34. Pandey UB, Nichols CD. Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery. *Pharmacol Rev.* 2011;63: 411–436.
35. International HapMap Consortium. The International HapMap Project. *Nature.* 2003;426: 789–796.
36. Wu DY, Ugozzoli L, Pal BK, Wallace RB. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proc Natl Acad Sci U S A.* 1989;86: 2757–2760.
37. Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, et al. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet.* 1993;4: 387–392.
38. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005;308: 385–389.
39. the CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet. Nature Research;* 2015;47: 1121–1130.

40. van Manen D, van 't Wout AB, Schuitemaker H. Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. *Retrovirology*. 2012;9: 70.
41. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*. 2009;461: 399–401.
42. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;39: 1181–1186.
43. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42: D1001–6.
44. Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. *BMC Med Genet*. 2009;10: 6.
45. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25: 25–29.
46. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet*. 2010;11: 855–866.
47. Kato T, Shinoura Y. Isolation and characterization of mutants of *Escherichia coli* deficient in induction of mutations by ultraviolet light. *Mol Gen Genet*. 1977;156: 121–131.
48. Begley TJ, Rosenbach AS, Ideker T, Samson LD. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Mol Cancer Res*. 2002;1: 103–112.
49. Hashimoto S, Ogura M, Aritomi K, Hoshida H, Nishizawa Y, Akada R. Isolation of auxotrophic mutants of diploid industrial yeast strains after UV mutagenesis. *Appl Environ Microbiol*. 2005;71: 312–319.
50. Weiss AA, Hewlett EL, Myers GA, Falkow S. Tn5-induced mutations affecting virulence factors of *Bordetella pertussis*. *Infect Immun*. 1983;42: 33–41.
51. Shaw KJ, Berg CM. *Escherichia coli* K-12 auxotrophs induced by insertion of the transposable element Tn5. *Genetics*. 1979;92: 741–747.
52. Engebrecht J, Neilson K, Silverman M. Bacterial bioluminescence: isolation and genetic analysis of functions from *Vibrio fischeri*. *Cell*. 1983;32: 773–781.
53. Griffiths AJF, Gelbart WM, Miller JH, Lewontin RC. Transposable Elements in Drosophila. W. H. Freeman; 1999.
54. Kaiser K, Goodwin SF. “Site-selected” transposon mutagenesis of Drosophila. *Proc Natl Acad Sci U S A*. 1990;87: 1686–1690.
55. Bellen HJ, O’Kane CJ, Wilson C, Grossniklaus U, Pearson RK, Gehring WJ. P-element-mediated enhancer detection: a versatile method to study development in Drosophila. *Genes Dev*. 1989;3: 1288–1300.
56. Emmons SW, Yesner L, Ruan KS, Katzenberg D. Evidence for a transposon in *Caenorhabditis elegans*. *Cell*. 1983;32: 55–65.
57. Liao LW, Rosenzweig B, Hirsh D. Analysis of a transposable element in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A*. 1983;80: 3585–3589.
58. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*. 2002;418: 387–391.
59. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, et al. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A*. 2003;100: 14339–14344.

60. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, et al. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci U S A*. 2006;103: 2833–2838.
61. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, et al. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*. 2003;301: 653–657.
62. Dorsett Y, Tuschl T. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov*. 2004;3: 318–329.
63. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270: 467–470.
64. Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*. 1996;6: 639–645.
65. Mischel PS, Cloughesy TF, Nelson SF. DNA-microarray analysis of brain cancer: molecular classification for therapy. *Nat Rev Neurosci*. 2004;5: 782–792.
66. Forozan F, Mahlamäki EH, Monni O, Chen Y, Veldman R, Jiang Y, et al. Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. *Cancer Res*. 2000;60: 4519–4525.
67. Hasegawa S, Furukawa Y, Li M, Satoh S, Kato T, Watanabe T, et al. Genome-wide analysis of gene expression in intestinal-type gastric cancers using a complementary DNA microarray representing 23,040 genes. *Cancer Res*. 2002;62: 7012–7017.
68. Bochner BR. Sleuthing out bacterial identities. *Nature*. 1989;339: 157–158.
69. Prüß BM, Campbell JW, Van Dyk TK, Zhu C, Kogan Y, Matsumura P. FlhD/FlhC Is a Regulator of Anaerobic Respiration and the Entner-Doudoroff Pathway through Induction of the Methyl-Accepting Chemotaxis Protein Aer. *J Bacteriol*. 2003;185: 534–543.
70. Jones J, Studholme DJ, Knight CG, Preston GM. Integrated bioinformatic and phenotypic analysis of RpoN-dependent traits in the plant growth-promoting bacterium *Pseudomonas fluorescens* SBW25. *Environ Microbiol*. 2007;9: 3046–3064.
71. Biswas S, Biswas I. Role of HtrA in surface protein expression and biofilm formation by *Streptococcus mutans*. *Infect Immun*. 2005;73: 6923–6934.
72. Loh KD, Gyaneshwar P, Markenscoff Papadimitriou E, Fong R, Kim K-S, Parales R, et al. A previously undescribed pathway for pyrimidine catabolism. *Proc Natl Acad Sci U S A*. 2006;103: 5114–5119.
73. Tracy BS, Edwards KK, Eisenstark A. Carbon and nitrogen substrate utilization by archival *Salmonella typhimurium* LT2 cells. *BMC Evol Biol*. 2002;2: 14.
74. Mukherjee A, Mammel MK, LeClerc JE, Cebula TA. Altered utilization of N-acetyl-D-galactosamine by *Escherichia coli* O157:H7 from the 2006 spinach outbreak. *J Bacteriol*. 2008;190: 1710–1717.
75. Durso LM, Smith D, Hutkins RW. Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157:H7 strains. *Appl Environ Microbiol*. 2004;70: 6466–6472.
76. Bochner BR, Giovannetti L, Viti C. Important discoveries from analysing bacterial phenotypes. *Mol Microbiol*. 2008;70: 274–280.
77. Ericsson HM, Sherris JC. Antibiotic sensitivity testing. Report of an international collaborative study. *Acta Pathol Microbiol Scand B Microbiol Immunol*. 1971;217: Suppl 217:1+.
78. Jorgensen JH, Turnidge JD. Susceptibility Test Methods: Dilution and Disk Diffusion Methods*. In: Pfaller MA, Richter SS, Funke G, Jorgensen JH, Landry ML, Carroll KC, et al., editors. Manual of Clinical Microbiology, 11th Edition. American Society of Microbiology; 2015. pp. 1253–1273.

79. Enright MC, Day NPJ, Davies CE, Peacock SJ, Spratt BG. Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. *J Clin Microbiol.* 2000;38: 1008–1015.
80. De Oliveira AP, Watts JL, Salmon SA, Aarestrup FM. Antimicrobial susceptibility of *Staphylococcus aureus* isolated from bovine mastitis in Europe and the United States. *J Dairy Sci.* 2000;83: 855–862.
81. Hilf M, Yu VL, Sharp J, Zuravleff JJ, Korvick JA, Muder RR. Antibiotic therapy for *Pseudomonas aeruginosa* bacteremia: outcome correlations in a prospective study of 200 patients. *Am J Med.* 1989;87: 540–546.
82. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.* 2006;2: 2006.0008.
83. Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, et al. Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol Syst Biol.* 2009;5: 335.
84. Inoue T, Shingaki R, Hirose S, Waki K, Mori H, Fukui K. Genome-wide screening of genes required for swarming motility in *Escherichia coli* K-12. *J Bacteriol.* 2007;189: 950–957.
85. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, et al. Phenotypic landscape of a bacterial cell. *Cell.* 2011;144: 143–156.
86. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, et al. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 2013;41: D605–12.
87. Zhou J, Rudd KE. EcoGene 3.0. *Nucleic Acids Res.* 2013;41: D613–24.
88. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, et al. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*--one species on the basis of genetic evidence. *Appl Environ Microbiol.* 2000;66: 2627–2630.
89. Ivanova N, Sorokin A, Anderson I, Galleron N, Candelon B, Kapatral V, et al. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature.* 2003;423: 87–91.
90. An R, Sreevatsan S, Grewal PS. Comparative in vivo gene expression of the closely related bacteria *Photorhabdus temperata* and *Xenorhabdus koppenhoeferi* upon infection of the same insect host, *Rhizotrogus majalis*. *BMC Genomics.* 2009;10: 433.
91. Ebert D. Experimental evolution of parasites. *Science.* 1998;282: 1432–1435.
92. Kassen R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J Evol Biol.* Blackwell Science Ltd.; 2002;15: 173–190.
93. Perron GG, Zasloff M, Bell G. Experimental evolution of resistance to an antimicrobial peptide. *Proc Biol Sci.* 2006;273: 251–256.
94. Garland T, Rose MR. Experimental Evolution. University of California Press; 2009.
95. Stearns SC, Ackermann M, Doebeli M, Kaiser M. Experimental evolution of aging, growth, and reproduction in fruitflies. *Proc Natl Acad Sci U S A.* 2000;97: 3309–3313.
96. Ratcliff WC, Denison RF, Borrello M, Travisano M. Experimental evolution of multicellularity. *Proc Natl Acad Sci U S A.* 2012;109: 1595–1600.
97. Lenski, R.E., Rose, M.R., Simpson, S.C., & Tadler, S.C. Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *Am Nat.* 1991;138: 1315–1341.
98. Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M. Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci U S A.* 1999;96: 3807–3812.

99. Cooper VS, Bennett AF, Lenski RE. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. *Evolution*. 2001;55: 889–896.
100. Lenski RE, Winkworth CL, Riley MA. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J Mol Evol*. 2003;56: 498–508.
101. Cooper VS, Lenski RE. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature*. 2000;407: 736–739.
102. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*. 1997;387: 703–705.
103. Blount ZD, Borland CZ, Lenski RE. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2008;105: 7899–7906.
104. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, et al. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A*. 2013;110: 222–227.
105. Philippe N, Pelosi L, Lenski RE, Schneider D. Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with *Escherichia coli*. *J Bacteriol*. 2009;191: 909–921.
106. Lenski RE, Wiser MJ, Ribick N, Blount ZD, Nahum JR, Jeffrey Morris J, et al. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proc R Soc B. The Royal Society*; 2015;282: 20152292.
107. Quandt EM, Deatherage DE, Ellington AD, Georgiou G, Barrick JE. Recursive genome wide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2014;111: 2217–2222.
108. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A*. 2012;109: E2774–83.
109. Rose MR. Artificial Selection on a Fitness-Component in *Drosophila melanogaster*. *Evolution*. [Society for the Study of Evolution, Wiley]; 1984;38: 516–526.
110. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature*. 2010;467: 587–590.
111. Zhou D, Udpa N, Gersten M, Visk DW, Bashir A, Xue J, et al. Experimental selection of hypoxia-tolerant *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2011;108: 2349–2354.
112. Swallow JG, Carter PA, Garland T Jr. Artificial selection for increased wheel-running behavior in house mice. *Behav Genet*. 1998;28: 227–237.
113. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics*. 2015;15: 141–161.
114. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*. 1999;285: 901–906.
115. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, et al. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*. 2003;421: 231–237.
116. Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, Koch B, et al. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science*. 2004;303: 832–835.
117. Bouché N, Bouchez D. Arabidopsis gene knockout: phenotypes wanted. *Curr Opin Plant Biol*. 2001;4: 111–117.
118. Pommerenke C, Müsken M, Becker T, Dötsch A, Klawonn F, Häussler S. Global genotype-phenotype correlations in *Pseudomonas aeruginosa*. *PLoS Pathog*. 2010;6: e1001074.

119. Sawai S, Guan X-J, Kuspa A, Cox EC. High-throughput analysis of spatio-temporal dynamics in *Dictyostelium*. *Genome Biol.* 2007;8: R144.
120. Green ECJ, Gkoutos GV, Lad HV, Blake A, Weekes J, Hancock JM. EMPReSS: European mouse phenotyping resource for standardized screens. *Bioinformatics.* 2005;21: 2930–2931.
121. Schacherer J, Ruderfer DM, Gresham D, Dolinski K, Botstein D, Kruglyak L. Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS One.* 2007;2: e322.
122. Srivatsan A, Han Y, Peng J, Tehranchi AK, Gibbs R, Wang JD, et al. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* 2008;4: e1000139.
123. Held K, Ramage E, Jacobs M, Gallagher L, Manoil C. Sequence-verified two-allele transposon mutant library for *Pseudomonas aeruginosa* PAO1. *J Bacteriol.* 2012;194: 6387–6389.
124. Klockgether J, Munder A, Neugebauer J, Davenport CF, Stanke F, Larbig KD, et al. Genome diversity of *Pseudomonas aeruginosa* PAO1 laboratory strains. *J Bacteriol.* 2010;192: 1113–1121.
125. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warren P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature.* 2000;406: 959–964.
126. Bateson W, Mendel G. Mendel's Principles of Heredity. Dover Publications; 2013.
127. Cohen T, Murray M. Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness. *Nat Med.* 2004;10: 1117–1121.
128. Dye C, Williams BG, Espinal MA, Raviglione MC. Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science.* 2002;295: 2042–2046.
129. Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol.* 2010;8: 260–271.
130. Ward H, Perron GG, Maclean RC. The cost of multiple drug resistance in *Pseudomonas aeruginosa*. *J Evol Biol.* 2009;22: 997–1003.
131. Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet.* 2009;5: e1000578.
132. Millan AS, Heilbron K, Craig MacLean R. Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. *ISME J. Nature Publishing Group;* 2013;8: 601–612.
133. Williams TN, Mwangi TW, Wambua S, Peto TEA, Weatherall DJ, Gupta S, et al. Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait. *Nat Genet.* 2005;37: 1253–1257.
134. Hartman JL 4th, Garvik B, Hartwell L. Principles for the buffering of genetic variation. *Science.* 2001;291: 1001–1004.
135. Wagner A. Robustness against mutations in genetic networks of yeast. *Nat Genet.* 2000;24: 355–361.
136. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science.* 2001;294: 2364–2368.
137. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. *Science.* 2004;303: 808–813.
138. Goodwin EB, Ellis RE. Turning clustering loops: sex determination in *Caenorhabditis elegans*. *Curr Biol.* 2002;12: R111–20.
139. Sternberg PW, Horvitz HR. The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in *C. elegans*. *Cell.* 1989;58: 679–693.

140. Thomas JH, Birnby DA, Vowels JJ. Evidence for parallel processing of sensory information controlling dauer formation in *Caenorhabditis elegans*. *Genetics*. 1993;134: 1105–1117.
141. Kaiser D. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proc Natl Acad Sci U S A*. 1979;76: 5952–5956.
142. Sliusarenko O, Zusman DR, Oster G. Aggregation during fruiting body formation in *Myxococcus xanthus* is driven by reducing cell movement. *J Bacteriol*. 2007;189: 611–619.
143. Spormann AM, Kaiser AD. Gliding movements in *Myxococcus xanthus*. *J Bacteriol*. 1995;177: 5846–5852.
144. Zusman DR, Scott AE, Yang Z, Kirby JR. Chemosensory pathways, motility and development in *Myxococcus xanthus*. *Nat Rev Microbiol*. 2007;5: 862–872.
145. Mauriello EMF, Mignot T, Yang Z, Zusman DR. Gliding motility revisited: how do the myxobacteria move without flagella? *Microbiol Mol Biol Rev*. 2010;74: 229–249.
146. Hodgkin J, Kaiser D. Genetics of gliding motility in *Myxococcus xanthus* (Myxobacterales): Genes controlling movement of single cells. *Mol Gen Genet*. Springer-Verlag; 171: 167–176.
147. Skerker JM, Berg HC. Direct observation of extension and retraction of type IV pili. *Proc Natl Acad Sci U S A*. 2001;98: 6901–6904.
148. Wu SS, Kaiser D. Genetic and functional evidence that Type IV pili are required for social gliding motility in *Myxococcus xanthus*. *Mol Microbiol*. 1995;18: 547–558.
149. Jakovljevic V, Leonardy S, Hoppert M, Søgaard-Andersen L. PilB and PilT are ATPases acting antagonistically in type IV pilus function in *Myxococcus xanthus*. *J Bacteriol*. 2008;190: 2411–2421.
150. Kaiser D. Signaling in myxobacteria. *Annu Rev Microbiol*. 2004;58: 75–98.
151. Sun H, Zusman DR, Shi W. Type IV pilus of *Myxococcus xanthus* is a motility apparatus controlled by the frz chemosensory system. *Curr Biol*. 2000;10: 1143–1146.
152. Blackhart BD, Zusman DR. “Frizzy” genes of *Myxococcus xanthus* are involved in control of frequency of reversal of gliding motility. *Proc Natl Acad Sci U S A*. 1985;82: 8767–8770.
153. Ward MJ, Zusman DR. Regulation of directed motility in *Myxococcus xanthus*. *Mol Microbiol*. 1997;24: 885–893.
154. Behmlander RM, Dworkin M. Extracellular fibrils and contact-mediated cell interactions in *Myxococcus xanthus*. *J Bacteriol*. 1991;173: 7810–7820.
155. Dworkin M. Fibrils as extracellular appendages of bacteria: their role in contact-mediated cell-cell interactions in *Myxococcus xanthus*. *Bioessays*. 1999;21: 590–595.
156. Kim S-H, Ramaswamy S, Downard J. Regulated Exopolysaccharide Production in *Myxococcus xanthus*. *J Bacteriol*. 1999;181: 1496–1507.
157. Arnold JW, Shimkets LJ. Cell surface properties correlated with cohesion in *Myxococcus xanthus*. *J Bacteriol*. 1988;170: 5771–5777.
158. Ramaswamy S, Dworkin M, Downard J. Identification and characterization of *Myxococcus xanthus* mutants deficient in calcofluor white binding. *J Bacteriol*. 1997;179: 2863–2871.
159. Dana JR, Shimkets LJ. Regulation of cohesion-dependent cell interactions in *Myxococcus xanthus*. *J Bacteriol*. 1993;175: 3636–3647.
160. Shimkets LJ. Role of cell cohesion in *Myxococcus xanthus* fruiting body formation. *J Bacteriol*. 1986;166: 842–848.

161. Yang Z, Geng Y, Xu D, Kaplan HB, Shi W. A new set of chemotaxis homologues is essential for *Myxococcus xanthus* social motility. *Mol Microbiol.* 1998;30: 1123–1130.
162. Hoiczky E, Baumeister W. The junctional pore complex, a prokaryotic secretion organelle, is the molecular motor underlying gliding motility in cyanobacteria. *Curr Biol.* 1998;8: 1161–1168.
163. Wolgemuth C, Hoiczky E, Kaiser D, Oster G. How myxobacteria glide. *Curr Biol.* 2002;12: 369–377.
164. Burchard RP. Trail following by gliding bacteria. *J Bacteriol.* 1982;152: 495–501.
165. Yang R, Bartle S, Otto R, Stassinopoulos A, Rogers M, Plamann L, et al. AglZ is a filament-forming coiled-coil protein required for adventurous gliding motility of *Myxococcus xanthus*. *J Bacteriol.* 2004;186: 6168–6178.
166. Mignot T, Shaevitz JW, Hartzell PL, Zusman DR. Evidence that focal adhesion complexes power bacterial gliding motility. *Science.* 2007;315: 853–856.
167. Hodgkin J, Kaiser D. Cell-to-cell stimulation of movement in nonmotile mutants of *Myxococcus*. *Proc Natl Acad Sci U S A.* 1977;74: 2938–2942.
168. Pathak DT, Wall D. Identification of the cglC, cglD, cglE, and cglF genes and their role in cell contact-dependent gliding motility in *Myxococcus xanthus*. *J Bacteriol.* 2012;194: 1940–1949.
169. Vassallo C, Pathak DT, Cao P, Zuckerman DM, Hoiczky E, Wall D. Cell rejuvenation and social behaviors promoted by LPS exchange in myxobacteria. *Proc Natl Acad Sci U S A.* 2015;112: E2939–46.
170. Wireman JW, Dworkin M. Developmentally induced autolysis during fruiting body formation by *Myxococcus xanthus*. *J Bacteriol.* 1977;129: 798–802.
171. Nariya H, Inouye M. MazF, an mRNA interferase, mediates programmed cell death during multicellular *Myxococcus* development. *Cell.* 2008;132: 55–66.
172. O'Connor KA, Zusman DR. Behavior of peripheral rods and their role in the life cycle of *Myxococcus xanthus*. *J Bacteriol.* 1991;173: 3342–3355.
173. Sudo SZ, Dworkin M. Resistance of vegetative cells and microcysts of *Myxococcus xanthus*. *J Bacteriol.* 1969;98: 883–887.
174. Hagen DC, Bretscher AP, Kaiser D. Synergism between morphogenetic mutants of *Myxococcus xanthus*. *Dev Biol.* 1978;64: 284–296.
175. Downard J, Ramaswamy SV, Kil KS. Identification of esg, a genetic locus involved in cell-cell signaling during *Myxococcus xanthus* development. *J Bacteriol.* 1993;175: 7762–7770.
176. Inouye M, Inouye S, Zusman DR. Gene expression during development of *Myxococcus xanthus*: pattern of protein synthesis. *Dev Biol.* 1979;68: 579–591.
177. Cashel M. The control of ribonucleic acid synthesis in *Escherichia coli*. IV. Relevance of unusual phosphorylated compounds from amino acid-starved stringent strains. *J Biol Chem.* 1969;244: 3133–3141.
178. Cashel M, Gallant J. Two compounds implicated in the function of the RC gene of *Escherichia coli*. *Nature.* 1969;221: 838–841.
179. Harris BZ, Kaiser D, Singer M. The guanosine nucleotide (p)ppGpp initiates development and A-factor production in *Myxococcus xanthus*. *Genes Dev.* 1998;12: 1022–1035.
180. Manoil C, Kaiser D. Accumulation of guanosine tetraphosphate and guanosine pentaphosphate in *Myxococcus xanthus* during starvation and myxospore formation. *J Bacteriol.* 1980;141: 297–304.
181. Konovalova A, Löbach S, Søgaard-Andersen L. A RelA-dependent two-tiered regulated proteolysis cascade controls synthesis of a contact-dependent intercellular signal in *Myxococcus xanthus*. *Mol Microbiol.* 2012;84: 260–275.

182. Singer M, Kaiser D. Ectopic production of guanosine penta- and tetraphosphate can initiate early developmental gene expression in *Myxococcus xanthus*. *Genes Dev.* 1995;9: 1633–1644.
183. Kuspa A, Kroos L, Kaiser D. Intercellular signaling is required for developmental gene expression in *Myxococcus xanthus*. *Dev Biol.* 1986;117: 267–276.
184. Kuspa A, Plamann L, Kaiser D. Identification of heat-stable A-factor from *Myxococcus xanthus*. *J Bacteriol.* 1992;174: 3319–3326.
185. Plamann L, Kuspa A, Kaiser D. Proteins that rescue A-signal-defective mutants of *Myxococcus xanthus*. *J Bacteriol.* 1992;174: 3311–3318.
186. Guo D, Wu Y, Kaplan HB. Identification and characterization of genes required for early *Myxococcus xanthus* developmental gene expression. *J Bacteriol.* 2000;182: 4564–4571.
187. Yang C, Kaplan HB. *Myxococcus xanthus* sasS encodes a sensor histidine kinase required for early developmental gene expression. *J Bacteriol.* 1997;179: 7759–7767.
188. Kuspa A, Kaiser D. Genes required for developmental signalling in *Myxococcus xanthus*: three asg loci. *J Bacteriol.* 1989;171: 2762–2772.
189. Plamann L, Davis JM, Cantwell B, Mayor J. Evidence that asgB encodes a DNA-binding protein essential for growth and development of *Myxococcus xanthus*. *J Bacteriol.* 1994;176: 2013–2020.
190. Cho K, Zusman DR. AsgD, a new two-component regulator required for A-signalling and nutrient sensing during early development of *Myxococcus xanthus*. *Mol Microbiol.* 1999;34: 268–281.
191. Garza AG, Harris BZ, Greenberg BM, Singer M. Control of asgE expression during growth and development of *Myxococcus xanthus*. *J Bacteriol.* 2000;182: 6622–6629.
192. Plamann L, Li Y, Cantwell B, Mayor J. The *Myxococcus xanthus* asgA gene encodes a novel signal transduction protein required for multicellular development. *J Bacteriol.* 1995;177: 2014–2020.
193. Kim SK, Kaiser D. C-factor: a cell-cell signaling protein required for fruiting body morphogenesis of *M. xanthus*. *Cell.* 1990;61: 19–26.
194. Kim SK, Kaiser D. C-factor has distinct aggregation and sporulation thresholds during *Myxococcus* development. *J Bacteriol.* 1991;173: 1722–1728.
195. Jelsbak L, Søgaard-Andersen L. The cell surface-associated intercellular C-signal induces behavioral changes in individual *Myxococcus xanthus* cells during fruiting body morphogenesis. *Proc Natl Acad Sci U S A.* 1999;96: 5031–5036.
196. Kroos L, Kaiser D. Expression of many developmentally regulated genes in *Myxococcus* depends on a sequence of cell interactions. *Genes Dev.* 1987;1: 840–854.
197. Dworkin M. Nutritional regulation of morphogenesis in *Myxococcus xanthus*. *J Bacteriol.* 1963;86: 67–72.
198. Wall D, Kolenbrander PE, Kaiser D. The *Myxococcus xanthus* pilQ(sglA) Gene Encodes a Secretin Homolog Required for Type IV Pilus Biogenesis, Social Motility, and Development. *J Bacteriol.* 1999;181: 24–33.
199. Bretscher AP, Kaiser D. Nutrition of *Myxococcus xanthus*, a fruiting myxobacterium. *J Bacteriol.* 1978;133: 763–768.
200. Kuner JM, Kaiser D. Fruiting body morphogenesis in submerged cultures of *Myxococcus xanthus*. *J Bacteriol.* 1982;151: 458–461.
201. Chen H, Keseler IM, Shimkets LJ. Genome size of *Myxococcus xanthus* determined by pulsed-field gel electrophoresis. *J Bacteriol.* 1990;172: 4206–4213.

202. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res.* 28, 263–266.
203. Finn, R. D., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J. E., et al. (2010). The Pfam protein families database. *Nucleic Acids Res.* 38, D211–D222.
204. Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28, 405–420.
205. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. *Science* 278, 631–637.
206. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
207. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., et al. (2000). InterPro—an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* 16, 1145–1150.
208. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2008). GenBank. *Nucleic Acids Res.* 39, D32–D37.
209. Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
210. Caberoy, N. B., Welch, R. D., Jakobsen, J. S., Slater, S. C., and Garza, A. G. (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.
211. Plamann, L., Davis, J. M., Cantwell, B., and Mayor, J. (1994). Evidence that asgB encodes a DNA-binding protein essential for growth and development of *Myxococcus xanthus*. *J. Bacteriol.* 176, 2013–2020.
212. Abramoff, M. D., Magalhaes, P. J., and Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics Int.* 11, 36–42.
213. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
214. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29.
215. Kroos, L., Kuspa, A., and Kaiser, D. (1986). A global analysis of developmentally regulated genes in *Myxococcus xanthus*. *Dev. Biol.* 117, 252–266.
216. Welch, R., and Kaiser, D. (2001). Cell behavior in traveling wave patterns of myxobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14907–14912.
217. Xie, C., Zhang, H., Shimkets, L. J., and Igoshin, O. A. (2011). Statistical image analysis reveals features affecting fates of *Myxococcus xanthus* developmental aggregates. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5915–5920.
218. Zhang, H., Angus, S., Tran, M., Xie, C., Igoshin, O. A., and Welch, R. D. (2011). Quantifying aggregation dynamics during *Myxococcus xanthus* development. *J. Bacteriol.* 193, 5164–5170.
219. Weimer, R. M., Creighton, C., Stassinopoulos, A., Youderian, P., and Hartzell, P. L. (1998). A chaperone in the HSP70 family controls production of extracellular fibrils in *Myxococcus xanthus*. *J. Bacteriol.* 180, 5357–5368.
220. Yang, Z., Geng, Y., and Shi, W. (1998). A DnaK homolog in *Myxococcus xanthus* is involved in social motility and fruiting body formation. *J. Bacteriol.* 180, 218–224.
221. Dworkin, M. Nutritional regulation of morphogenesis in *Myxococcus xanthus*. *J. Bacteriol.* 86, 67–72 (1963).
222. Kaiser, D. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proc. Natl. Acad. Sci. U. S. A.* 76, 5952–5956 (1979).

223. Kuner, J. M. & Kaiser, D. Introduction of transposon Tn5 into *Myxococcus* for analysis of developmental and other nonselectable mutants. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 425–429 (1981).
224. Bretscher, A. P. & Kaiser, D. Nutrition of *Myxococcus xanthus*, a fruiting myxobacterium. *J. Bacteriol.* **133**, 763–768 (1978).
225. Yan, J., Bradley, M. D., Friedman, J. & Welch, R. D. Phenotypic profiling of ABC transporter coding genes in *Myxococcus xanthus*. *Front. Microbiol.* **5**, 352 (2014).
226. Goldman, B. S. *et al.* Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15200–15205 (2006).
227. Caberoy, N. B., Welch, R. D., Jakobsen, J. S., Slater, S. C. & Garza, A. G. 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 (2003).
228. Velicer, G. J. *et al.* Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8107–8112 (2006).
229. Schacherer, J. *et al.* Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS One* **2**, e322 (2007).
230. Srivatsan, A. *et al.* High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* **4**, e1000139 (2008).
231. Klockgether, J. *et al.* Genome diversity of *Pseudomonas aeruginosa* PAO1 laboratory strains. *J. Bacteriol.* **192**, 1113–1121 (2010).
232. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008).
233. Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27**, 323–331 (2011).

VITA

NAME OF AUTHOR: Michael Douglas Bradley

DATE OF BIRTH: 28 May 1986

UNDERGRADUATE & GRADUATE SCHOOLS ATTENDED:

Syracuse University

DEGREES AWARDED:

Bachelor of Science, Biology (2008)

PROFESSIONAL EXPERIENCE:

Graduate research associate, Department of Biology, Syracuse University (2011-2016)

PUBLICATIONS:

1. **Bradley M.D.**, Neu D., Bahar F. & Welch R.D. Inter-laboratory evolution of a model organism and its epistatic effects on mutagenesis screens. *Sci. Rep.* 6, 38001 (2016).
2. Yan, J., **Bradley, M. D.**, Friedman, J. & Welch, R. D. Phenotypic profiling of ABC transporter coding genes in *Myxococcus xanthus*. *Front. Microbiol.* 5, 352 (2014).
3. Yan, J., Garza, A. G., **Bradley, M. D.** & Welch, R. D. A Clp/Hsp100 chaperone functions in *Myxococcus xanthus* sporulation and self-organization. *J. Bacteriol.* 194, 1689–1696 (2012).