Mapping Stochastic Collective Behavior Distinguishes Subtle Mutations in Social Bacteria

Merrill E. Asp [1,2] Eduardo A. Caro [1,3] Roy D. Welch [1,3,*] and Alison E. Patteson [1,2,*]

[1] *BioInspired Institute, Syracuse University, Syracuse, New York 13244, USA*

[2] *Physics Department, Syracuse University, Syracuse, New York 13244, USA*

[3] *Biology Department, Syracuse University, Syracuse, New York 13244, USA*

(Received 22 August 2023; accepted 4 December 2023; published 21 December 2023)

The genotype-to-phenotype problem (G2P) for multicellular development asks how genetic inputs control collective phenotypic outputs. However, this is a challenging problem due to gene redundancy and stochasticity, causing mutations to have subtle phenotypic effects and replicates to display significant variation. We approach this problem using the model organism *Myxococcus xanthus*, a motile self-organizing bacterium that forms three-dimensional cell aggregates that mature into spore-filled fruiting bodies when under starvation stress. We develop a high-throughput imaging method using three-dimensional-printed microscopes to efficiently collect large phenotypic datasets. Our automated methods for analysis and visualization produce a map of phenotypic variation in *M. xanthus* development. We demonstrate that even subtle effects on developmental dynamics caused by mutation can be identified, discriminated, characterized, and given statistical significance, with implications for future gene annotation studies and the effect of environmental factors on G2P.

DOI: 10.1103/PRXLife.1.023011

I. INTRODUCTION

Development is an energetically expensive and complicated part of many organisms' life cycles. For a genetically specified multicellular phenotype to be realized, the process requires the stepwise self-organization of increasingly ordered states and an active and robust dampening response to disruptive forces, such as environmental stress and mutation [1–3]. The control systems for development involve nonlinear redundant branching and intersecting intracellular and intercellular signal transduction pathways that provide spatiotemporal coordination of the transcriptional, translational, and post-translational events required for the cellular collective to manifest a final phenotype [4,5]. This complicated genotype-to-phenotype problem, abbreviated as G2P, is the broad task of understanding the iterative process of genetic cause and phenotypic effect that eventually results in biological emergence.

We use the Gram-negative delta-proteobacterium *Myxococcus xanthus* as a model organism. In the laboratory on an agar substrate, *M. xanthus* exists as a single-species motile biofilm called a swarm. Under nutrient-rich conditions, a swarm will expand across an agar surface as its component cells grow and move (swarming). In contrast, under nonnutritive (starvation) conditions, a swarm will not expand, although its component cells are in fact moving around faster within it. Instead, a starving swarm will undergo a transformation over a period of approximately one day, during which swarm cells organize into a discrete number of mound-shaped aggregates distributed nonrandomly across the swarm area, with each aggregate harboring a bolus of thousands of cells. Over the next few days, the cells at the center of each aggregate differentiate into quiescent myxospores, at which point the aggregates are considered to have matured into a fruiting body. The entire process represents a rudimentary but robust form of multicellular development and can therefore be studied as an example of G2P [6]. We note that throughout this work we use the term "phenotype" to refer to the collective, rather than cellular, characteristics.

Development of a multicellular prokaryote, such as the formation of an *M. xanthus* fruiting body, may be less complicated than development of a eukaryote, such as the formation of a mouse, fly, or flatworm, but it is still complicated enough to involve hundreds of genes arranged in branching networks of intersecting pathways. Central to *M. xanthus* fruiting body development is the extracellular protein A-signal, a diffusible proxy for local cell density. The pathway for A-signal-mediated communication includes receptors, response regulators, and effectors that influence cell motility and cell-cell interactions [7]. The Frz pathway, another two-component system, also plays a crucial role in regulating cell movement and aggregation in response to external signals [8]. The crosstalk between the regulation of these genes and cell motility, which changes the rates and types of cell-to-cell signaling, creates feedback that further affects multicellular development. Additionally, the interaction between genetic regulatory pathways and environmental factors, including prey bacteria, further complicates the process [9]. Many of the genes known to be involved in these networks and pathways were first identified through mutation and phenotypic characterization [10]. It is a foundational protocol in developmental biology: first, a wild-type strain is selected and its development

*Corresponding authors: rowelch@syr.edu; aepattes@syr.edu

phenotype is characterized, then mutations are introduced into the wild-type genome to create new (mutant) strains, and their development phenotypes are characterized, then the phenotypes of the mutant strains are compared to wild-type, and, if a strain displays a significant deviation, the gene(s) and other genetic element(s) affected by the mutation are deemed more likely to be involved in development. This information can then be used for genome annotation and to guide future research.

A requisite condition to establish significance when characterizing and quantifying a developmental phenotype is a defined boundary that distinguishes the phenotype of wild-type from a near-wild-type mutant. There are at least three confounding factors that make this difficult. First, developing biological systems exhibit an inherent phenotypic stochasticity, and efforts at holding genome sequence and experimental conditions constant can only reduce the variation to a nontrivial baseline. Second, biological systems are also phenotypically robust to the impact of mutation because evolution guides the genes, networks, and pathways that control development to incorporate redundancies due, at least in part, to mechanisms such as duplication and divergence [11]. Third, developmental processes are often difficult to observe, record, and analyze in replicates sufficient to establish significance (see Fig. 1).

Statistical techniques that are useful in the face of highly stochastic events such as gene expression are a topic of active interdisciplinary research [12], and the inherent ability of living systems to submit to statistical study is an open epistemological question [13]. But, if a baseline variation can be determined for wild-type, any differences distinct from that baseline can distinguish phenotypes.

Although a developmental program is encoded in an organism's genome, having full information about that genome is often insufficient on its own to predict phenotype. A reciprocal approach to this problem is a systematic determination of an organism's developmental phenotype to infer the function of the complex genomic network. Understanding the range of developmental phenotypes adds new information in the quest to map incomplete genetic network data to (often incompletely described) developmental outcomes. The range of developmental outcomes thus warrants its own systematic investigation. Indeed, the collection, analysis, and interpretation of collective phenotypic data has been identified as a central challenge in understanding biological systems as a whole, known as phenomics [14,15]. Here, we approach that challenge in the context of multicellular bacterial development.

In this study, we describe the design, construction, and operation of an experimental setup for observing many instances of *M. xanthus* development and an analysis pipeline that quantifies distinctive features of development to display them on a phenotypic map. We use this map to demonstrate that even for a single strain (wild-type), with environmental factors held constant to the highest feasible degree, a multimodal distribution of developmental phenotypes emerges that tends to funnel towards a singular mature state—developed fruiting bodies.

We then employ this distribution of wild-type phenotypes to distinguish between wild-type *M. xanthus* development and four near-wild-type mutants, demonstrating that the impact of genotype in the G2P mapping is to shift these phenotype distributions. That such shifts are visible even in what may be considered minimal mutations shows that we can assess the phenotypic impact of what would otherwise be silent mutations. The perspective of phenotype as a distribution that can be perturbed can also inform future studies on the impacts of environmental variables, potentially uncovering multimodal developmental patterns in organisms beyond bacteria.

II. RESULTS

A time-lapse brightfield microcinematography assay of *M. xanthus* development is initiated upon the settling of cells from liquid suspension onto an agar surface. The entire development assay is recorded at 4× magnification with one image taken per minute over 24 h so that the final dataset consists of 1440 sequential images (frames). Initially, local cell density varies across the field of view, with denser areas appearing darker. Black spots are occasionally visible in the first hour, which is either small tight cell clumps transferred from suspension or small pieces of extracellular material. Movement becomes evident in subsequent hours as the gray areas move and change shape, and black spots disperse. Over the next several hours, the different gray areas fade, and the entire field of view becomes a more homogeneous population. The movement then appears to speed up and become more synchronous. Transient darker gray regions begin to emerge across the field of view, and the population appears effervescent. Subsequently, darker and less transient ridges appear, moving in an amoeboid or sluglike fashion, eventually consolidating into the initial population of aggregates. These initial aggregates are unstable, merging if they collide, while others occasionally divide in binary fission. Aggregates undergo coarsening when a significant subset of the aggregates distributed across the field of view shrink and disappear. The remaining aggregates become more stationary and stable. Throughout these aggregation dynamics, the space between aggregates is teeming with cells and cell movement. Once the stable field of aggregates is established, the cell activity seems to diminish.

In our study, we observed over 500 wild-type time series, with each fruiting body assay requiring a sealed chamber with sufficient temperature, oxygenation, and humidity for development to occur. The 96 independent microscopes of our image acquisition setup (Fig. 2) combined with a high-volume sample preparation protocol allow for dozens of time series to be collected simultaneously, maintaining environmental variables such as agar percentage, day-to-day cell culture variation, and ambient humidity to the greatest possible extent. Consistent temperature was also manually verified across samples. Cells are inoculated from liquid culture and sealed in each slide assembly, where the aggregates begin to form. We report the spectrum of development behaviors, which manifested to varying degrees. Coarsening varied, as did the length of the slug phase and the occurrence of aggregation. Rippling, previously described [16], appeared intermittently, and could dominate the entire field of view. We also observed pulsing, a distinct wavelike pattern where an aggregate seemingly propagates a signal throughout the field of view. Additionally, fruiting body merging, dividing, and other features varied in intensity and duration, even in replicates designed to be

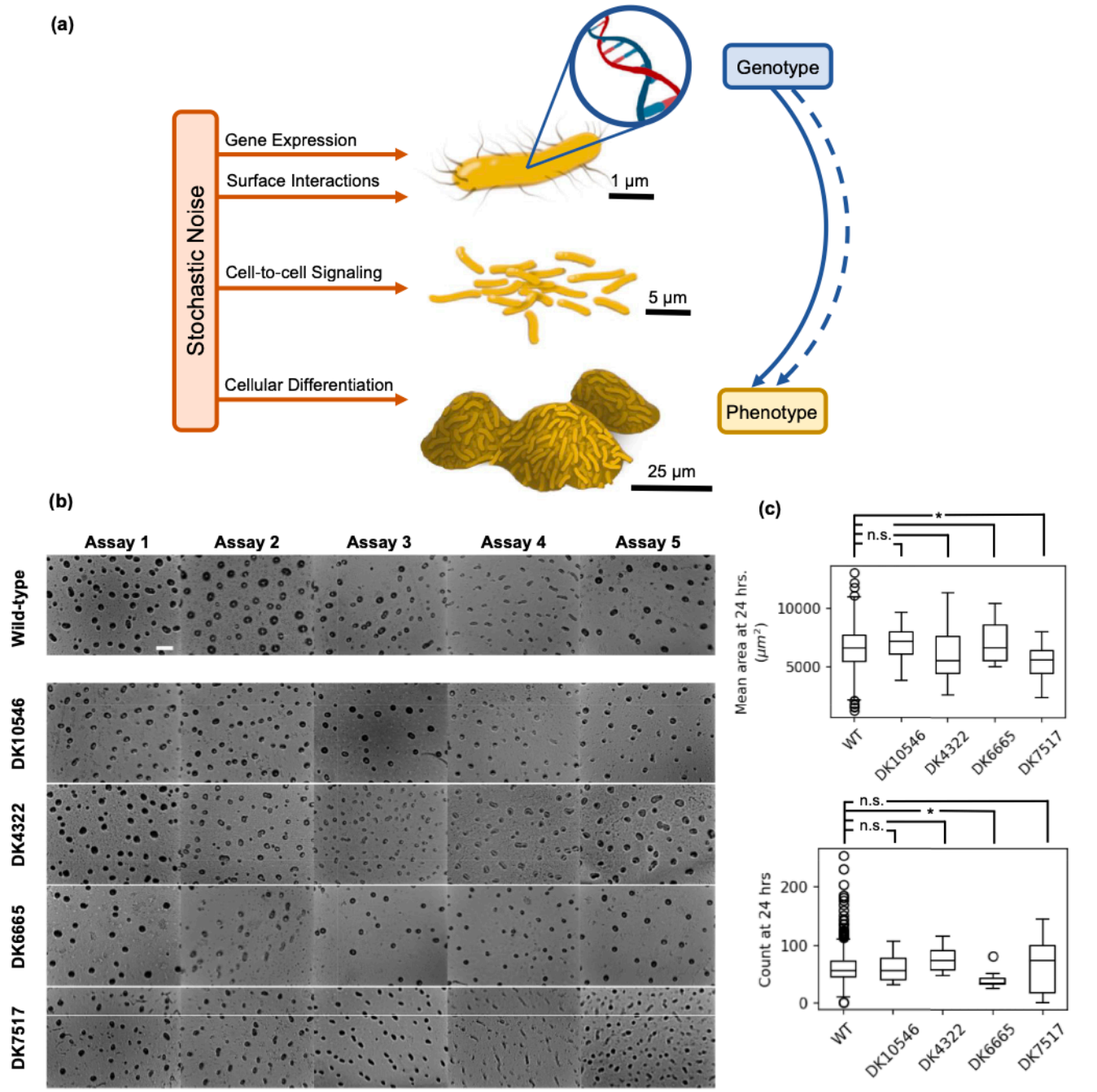

FIG. 1. Stochasticity is inherent to multicellular behaviors in social bacteria. (a) A bacterial colony undergoing fruiting body development is exposed to stochastic noise on multiple scales. At the cellular level, gene expression depends on thermally driven chemical events, and environmental factors such as variations in temperature and humidity introduce further uncertainty. Thus, both direct and indirect effects of genotype arrive at a final phenotype, via multiple developmental paths. (b) Images of final developmental phenotype for separate aggregation assays at 24 h post-inoculation. Pictured are a range of outcomes from the wild-type *M. xanthus* strain as well as the four mutant strains used in this study. Aggregates are visible as dark spots, seen from above. Scale bar, 250 μm. (c) The average final area and final count of wild-type (WT) aggregates and those for four mutant strains are reported with boxplots. Although there are some differences in these typical metrics of comparison, there is considerable overlap between wild-type and each of the mutant strains. $N > 500$ measurements for wild-type, taken over 25 different days; $N = 15$ measurements for each mutant strain, taken over two different days.

biologically and technically identical. Based on our observations of the video data, we note that the formation of *M. xanthus* aggregation tended to follow a sequence of events: an initiation of aggregation from the undifferentiated swarm, an increase in the area of the nascent aggregates, and eventual stability. The dynamics of each of these stages varied from video to video [Figs. 3(d) and 3(e)]. While it is not yet clear whether these are prescribed stereotyped events necessary for fruiting body formation, these stages are present in all our videos that form maturing fruiting bodies.

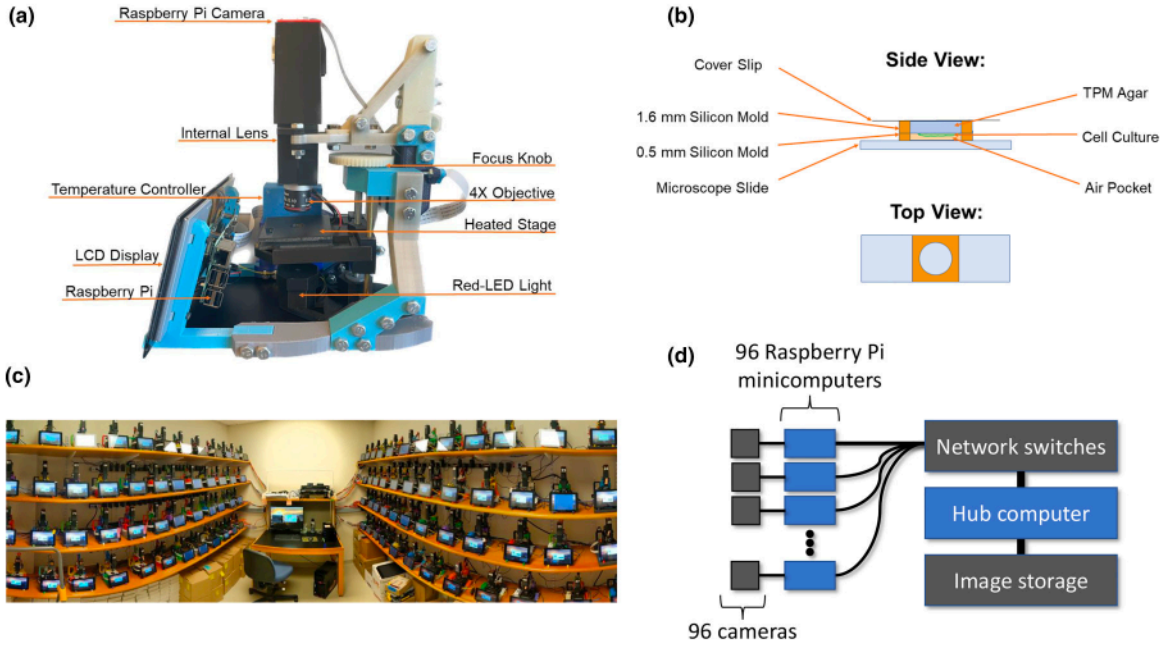

FIG. 2. High-throughput time series acquisition setup. (a) A single microscope, at 1.2 kg and 24 × 19 × 25 cm^3, with 3D-printed armature, 4× objective lens, light source, heated stage, camera, and Raspberry Pi microcomputer. (b) Slide assembly for each developmental experiment. Sandwiched between a glass coverslip and a glass slide, two silicone gaskets create a sealed enclosure containing a disk of non-nutritive agarose on which a colony of *M. xanthus* has been inoculated. Aggregate development is imaged over a 24-h period with one image taken each minute. (c) Panoramic photograph of full image acquisition setup including 96 microscopes and a central hub computer. (d) Basic network architecture for centralized image storage and control of all 96 microscopes.

Towards creating a map of developmental phenotype distribution in *M. xanthus*, we sought a set of metrics to quantify how aggregation proceeds. After observing many instances of fruiting body aggregation, we identified variations in development (timing, rates of area change, etc.) and chose metrics to capture as many types of variation as possible. The information-dense nature of time series data compelled us to identify metrics that could be reproducibly measured regardless of variations in image acquisition parameters such as lighting and camera properties. We then produced a dataset of wild-type aggregation time series recorded over 25 separate days, measuring the 10 quantitative metrics for each video. These 10 metrics consist of three metrics related to timing (start time—when aggregation begins, peak time—when the area occupied by aggregates is maximized, and stability time—when the number of aggregates becomes stable), as well as the mean and standard deviation in the average aggregate area at peak time and 24 h. The number of identifiable aggregates at peak time and 24 h was also measured, along with the fraction of aggregates that disperse after forming. Our method here distinguishes parameters that may be under a genetic influence and identifies new characteristics in phenotype that can be compared.

The specific formulas used to determine each phenotypic metric are detailed in the Supplemental Material (Table S1) [17]. These metrics are extracted with a custom Python image processing algorithm that identifies and measures each aggregate, as described in the Methods section. Values for these metrics across the wild-type dataset are shown in Fig. 3(a), with the distributions illustrating averages and variation for each metric. Peak time, aggregate count at peak time, and the fraction of aggregates that disperse exhibit bimodal distributions. Long-tailed distributions, such as the standard deviation (σ) of area and aggregate count (both at peak time and after 24 h), indicate the presence of abnormal phenotypes with extreme values in these metrics.

To visualize the wild-type dataset, we used these 10 phenotypic metrics to represent each time series as a point in a 10-dimensional phenotype space. Points closer together indicate more phenotypic similarity than points far apart. To simplify the dataset while maintaining its structure, we utilized principal component analysis (PCA) to reduce the ten dimensions to two dimensions, named PC1 and PC2. These two dimensions are each a single numerical measure, each a differently weighted mathematical composite of the 10 quantitative features. PC1 and PC2 show the most variation across the wild-type dataset when compared to any other linearly independent combination of input metrics. In fact, PC1 and PC2 account for 56% of the variance across the full wild-type dataset of over 500 time series. The resulting distribution of points in this 2D space represents the wild-type phenotype profile, a visualization of the distribution of developmental phenotypes for our model system, including variations in dynamics. The definition of this phenotype space in terms of PC1 and PC2 allows for quantitative phenomics in *M. xanthus* aggregation, which would otherwise be cumbersome in the 10-dimensional space.

To interpret the meaning of PC1 and PC2, we can consider how each developmental metric is weighted. The weights, which range from −1 to 1, indicate the significance of each metric within either PC1 or PC2, with larger absolute values indicating more strongly weighted metrics. Both PC1 and PC2

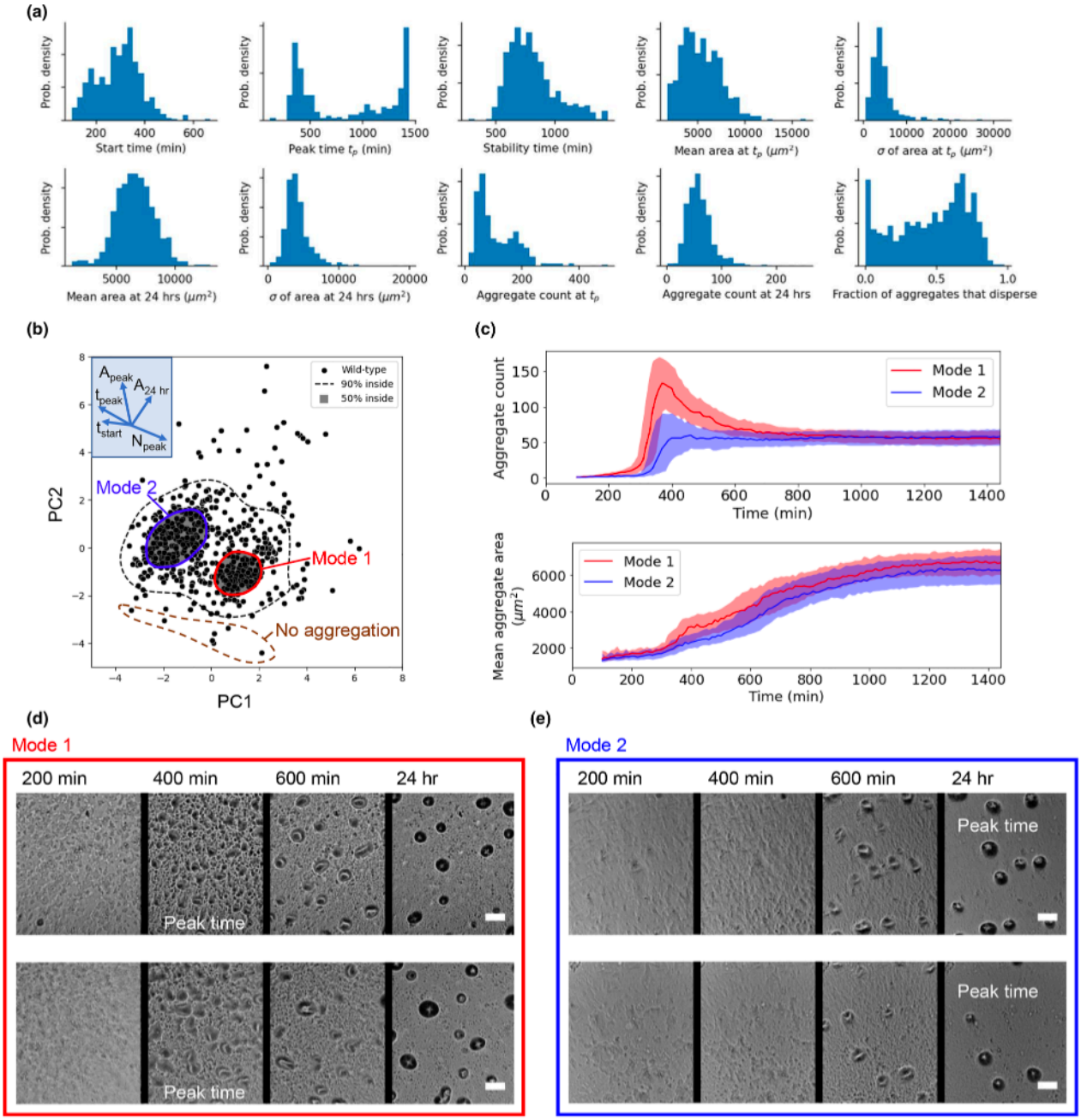

FIG. 3. Quantitative breadth of wild-type phenotype. (a) Histograms display the range of phenotypic metrics across over 500 wild-type aggregate development time series. Bimodal shapes in peak time (when total aggregate area is maximum), aggregate count at peak time, and fraction of aggregates that disperse reflect the two most common groupings of metrics. Long-tailed distributions, such as standard dev. (σ) of area and aggregate count (both at peak time and after 24 h) indicate the presence of abnormal phenotypes. All y-axes display probability density. (b) By using PCA to combine information from all 10 metrics, each wild-type time series is plotted as a single datapoint in a phenotypic feature space. PC1 primarily measures aggregate area, and PC2 correlates with the number and timing of aggregates. For example, while moving in the direction of the arrow labeled "N_{peak}", datapoints will have higher numbers of aggregates at peak time. Units of PC1 and PC2 are arbitrary, although the origin at (0,0) represents average behavior across the full wild-type dataset. A contour is drawn enclosing 90% of the datapoints, separating typical phenotypes from rare phenotypes. Within typical behavior, two separate clusters—Mode 1 and Mode 2—contain 50% of the wild-type datapoints. (c) Curves displaying the total number of aggregates over time (top) and mean area of aggregates over time (bottom) illustrate the developmental differences and similarities between the two wild-type modes. The central line represents the median at each time point, and the colored bands span the 25th to 75th percentiles at each time point, i.e., half the data about the median. In Mode 1, more aggregates develop at an earlier time, most of which disperse. The final number of aggregates is comparable for both modes. The rates of increase of mean area are also similar across the two modes. (d),(e) Two representative time series each for Mode 1 and Mode 2 phenotypes at three relevant time points. Mode 1 displays many dense aggregates that form early and then disperse. This causes an early peak time. Mode 2 displays aggregates that form later, most of which persist through the 24 h of development, slowly growing in area and darkening. This causes a late peak time. Scale bar is 100 μm.

contain a mix of all 10 metrics, with groups of metrics being more significant than any single metric. The top weighted metrics for PC1 are the number of aggregates at peak time, the fraction of aggregates that disperse, peak time, and start time. For PC2, the top metrics are mean area at peak time, standard deviation in area at peak time, standard deviation in area at 24 h, and mean area at 24 h. The full list of weights is provided in Table S2 in the Supplemental Material [17]. In summary, PC1 primarily reflects timing and the total number of fruiting bodies that form, which are negatively correlated in the wild-type dataset. That is, earlier aggregation and peak times tend to result in more fruiting bodies (especially at peak time), while later results in fewer. PC2, on the other hand, is independent of PC1 and characterizes area. Large aggregates can occur in large or small numbers and early or late relative to the average wild-type behavior. The precise variation of each metric in PC1, PC2 space is illustrated in the Supplemental Material (Fig. S1) [17]. We have observed that the metrics most capable of distinguishing developmental phenotypes are variation in timing and the dynamics of aggregate area, which emphasizes the general importance of including dynamics in phenotype quantification. We have also retained all metrics in the PCA to maximize the ability to distinguish developmental phenotypes across the full dataset.

Two primary modes of aggregate formation were observed, as illustrated by the shaded regions in Fig. 3(b). Significantly, both modes could be observed in time series acquired on the same day, with the same cell culture and environmental conditions, indicating a role for developmental stochasticity in determining which assay developed according to one mode or the other. "Mode 1" features aggregates that start forming and peak in total aggregate area sooner than other wild-type assays. Mode 1 aggregates are generally numerous, small, and dark at peak time, but a large fraction of them disappear before 24 h of development. These aggregates tend to be dynamic and lack a well-defined shape until after peak time [see Fig. 3(d)]. In contrast, Mode 2 aggregation is less mature early on, with either no visible aggregates or aggregates with fewer layers of cells able to block light [see Fig. 3(e)]. These aggregates are more static and form with more well-defined shapes, and more of them tend to persist through the 24 h of development. Because these aggregates tend to persist once they form, the time of peak total area is extremely late for Mode 2 when stable aggregates are still growing slowly. Although there are fewer Mode 2 aggregates at peak time than most wild-type assays, the mean number and size of these aggregates at 24 h is equal to that of Mode 1, as well as wild-type assays in general. Both modes demonstrate more consistently sized aggregates than other wild-type assays, both at peak time and at 24 h. The histograms of all 10 metrics for the two modes are presented in the Supplemental Material (Fig. S2) [17]. This analysis shows that the bimodal shape of three metrics [peak time, aggregate count at peak time, and fraction of aggregates that disperse; see Fig. 3(a)] is associated with the separation of the two wild-type aggregation modes. The two modes thus represent a confluence of metrics that arise from two distinct developmental pathways towards fruiting body aggregation, both of which can occur in the wild-type genetic background.

Exceptional phenotypes in our wild-type dataset were observed, including those that produce unusually large fruiting bodies. These occur by a variety of mechanisms, such as large aggregates forming either extremely early with defined shapes from initial formation or extremely late with shapes that only become visible towards the end of 24 h (time series in Supplemental Material, Fig. S3) [17]. These abnormal behaviors are present at the margins of PCA phenotype space because they represent a confluence of multiple abnormal metrics, revealing more information than standard statistical tests on one metric at a time. Some rare behaviors observed include failure to aggregate, which occurred in about 2% of wild-type assays, and failure for aggregates to stabilize after 24 h, which occurred in about 17% of wild-type assays.

To test the sensitivity of our methods, we selected four mutant strains with 60–80 replicates each over two to six separate days, which were developmentally similar to wild-type in final outcome. In preliminary experiments, all four strains produced three replicates each that were manually identified as "near-wild-type" with no significant difference in final aggregate size and number when compared to wild-type using a Student's t-test. Three of the mutant strains contained simple reporter genes, such as DK10546 constitutively producing GFP (Green Fluorescent Protein), to investigate whether introducing reporter genes into a prokaryotic genome would significantly impact cellular behavior or emergent phenotypes. Standard statistical tests distinguish one mutant strain, DK7517, from wild-type because it produces smaller than average aggregates (Fig. 1), but this difference is not highly pronounced. As the distribution of wild-type final mean areas is non-Gaussian, the Kolmogorov-Smirnov test was used instead of a Student's t-test to distinguish two distributions, which assumes normality of the underlying distributions. See the Methods section for more information on each strain.

To compare the mutant strains with wild-type, the developmental data were projected onto the same PC1 and PC2 axes that were calculated from the wild-type data. PCA was thus performed only once, giving a single standard phenotype space in which the mutant phenotype distributions could be directly compared to the wild-type phenotype distribution. This method allowed us to visualize and compare multiple metrics simultaneously. We drew two contours to capture the typical behavior and variability of each mutant strain's development. The first contour enclosed 50% of the datapoints and represented the median region in PCA space, while the second contour was wider and enclosed 90% of the datapoints, which served as a boundary for abnormal phenotypes. We compared the distribution of the mutant strain points with that of wild-type to calculate a p-value for the null hypothesis. This p-value was determined using bootstrapping, a nonparametric, data-driven statistical method that did not make any assumptions about the dataset beforehand.

All four mutant strains showed phenotype distributions that differed from the wild-type distribution in a subtle but measurable manner (see the Supplemental Material, Fig. S4) [17]. DK10546 and DK4322 exhibited a preference for Mode 1 behavior, whereas Mode 2 was rarely expressed. Figure 4(b) highlights some of the rare behaviors observed in certain replicates. DK10546 replicates showed more extreme versions of Mode 1 behavior, in which many small aggregates formed early but dispersed by 24 h. DK4322 replicates also showed a more extreme version of Mode 1 behavior, in which

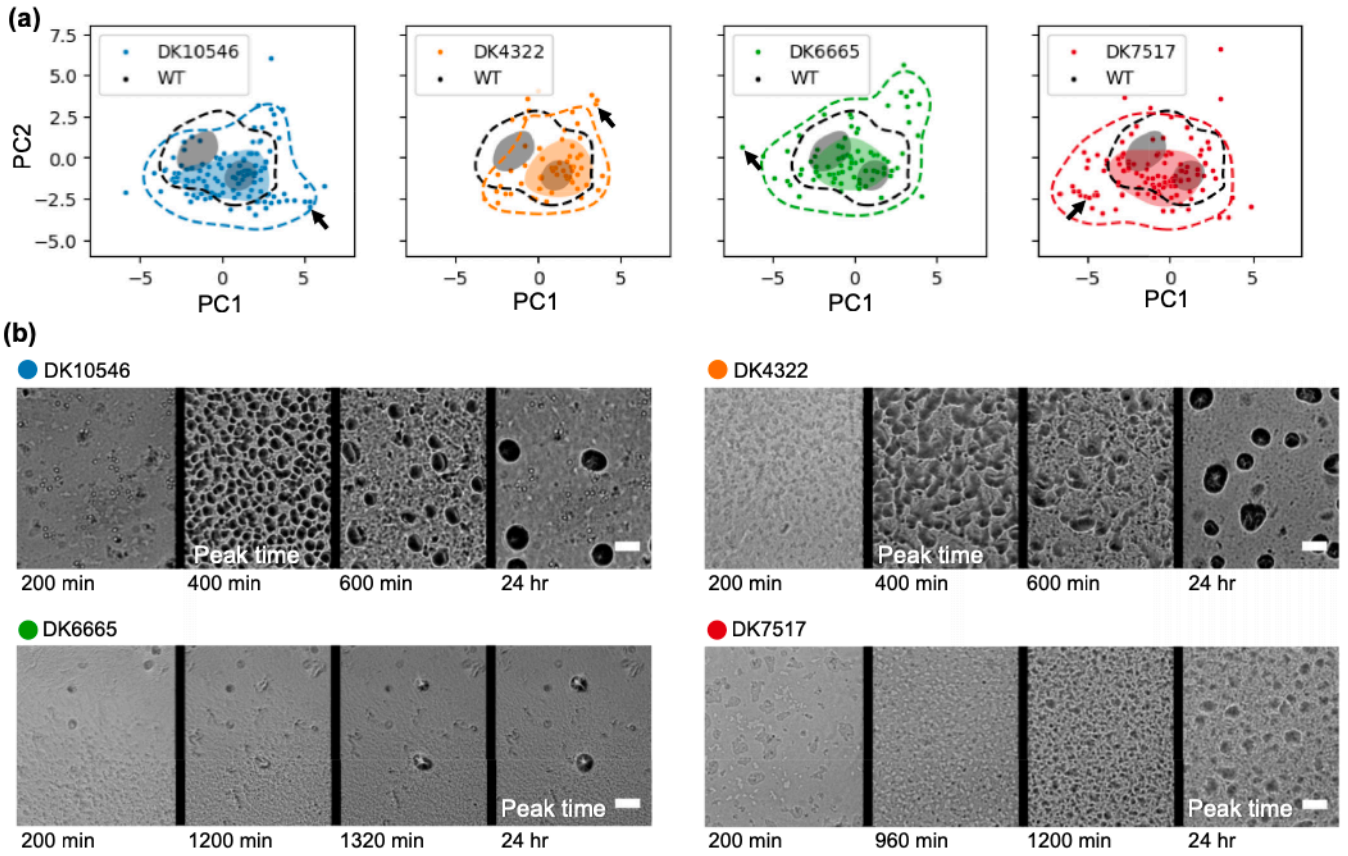

FIG. 4. Deviation of near-wild-type mutant strains from wild-type behavior. (a) Each mutant development time series is plotted as a single datapoint in phenotype space, as measured by the collective metrics PC1 and PC2. For each respective strain, dashed contours enclose 90% of the datapoints, and the shaded region(s) enclose 50% of the datapoints. The 90% and 50% contours for wild-type are shown for reference. The deviation of mutant phenotype from wild-type is determined by the departure of the mutant distribution from the wild-type distribution. Statistically significant departures from the wild-type distribution are measured for all four mutant strains, with p-values calculated for subsamples of only 15 replicates each. These p-values are calculated from many random samplings drawn from the wild-type dataset (Methods). Arrows point to time series shown in (b). Time series of phenotypes expressed rarely in wild-type are shown at three relevant time points for each mutant strain. Scale bar is 100 μm.

aggregates had irregular shapes at peak time. The final aggregates were slightly larger and more varied in area than those in typical wild-type assays. Some DK6665 replicates formed aggregates from sparse, small points that formed late and grew steadily over the 24-h period, a behavior rarely observed in wild-type. This strain also had difficulty dispersing random initial cell clumps present at inoculation, a behavior not observed in wild-type. DK7517 replicates displayed late aggregates that did not significantly darken, a noticeable deviation from even exceptional wild-type behavior. 2% of mutant strains and wild-type failed to aggregate, and 30–45% of mutant assays failed to stabilize after 24 h, significantly higher than the 17% observed in wild-type.

III. DISCUSSION

Each fruiting body formation event is a complex interplay of genetic background, controllable environmental factors such as temperature or substrate stiffness, and uncontrollable factors such as local variations in initial cell density or gene expression changes unique to a particular cell population. Genetic changes can affect the likelihood of multiple outcomes, and we have observed that most mutants behave similarly to the wild-type strain most of the time [18,19]. However, when broadening our understanding of phenotype as a distribution of developmental pathways, we can observe shifts in this distribution as a response to genetic changes. These observations show that the role of genotype in the G2P mapping is more generally to alter the distribution of phenotypes than to determine phenotype. This understanding of phenotype as a reshuffling of outcomes, rather than a guaranteed result, is a significant re-interpretation that aligns with the nature of living systems, which are adaptable and tuned by evolution to exhibit a range of behaviors. Biostatistical techniques, such as probabilistic latent variable models, support this perspective and complement the analysis presented in our study [20]. Additionally, the observation of the two developmental modes in the wild-type phenotype map underscores the ability of our approach to access information in the G2P mapping that otherwise would be hidden.

The statistical method presented in this study characterizes phenotype by supplementing mean behavior with information

about abnormal behavior, either locally or globally. Local abnormal behavior refers to groupings of behavior that fall within the scope of wild-type but are still considered outside the norm. Global abnormal behavior, on the other hand, refers to behaviors never expressed in the wild-type profile. Measuring mean behavior may not be sufficient to capture variations, and abnormality analysis is a promising area for identifying the effects of single-gene mutations, especially when mean behavior is not strongly affected due to the masking effects of developmental pathways that have evolved to be robust. This work demonstrates the value of measuring the boundaries of a phenotype distribution to observe changes from wild-type that would otherwise be washed out when considering only mean behavior. However, it requires sufficient replicates to reliably observe abnormal behavior [11]. Furthermore, an increased level of attention to abnormal phenotypes can suggest fruitful new research directions, possibly uncovering distributions of behaviors that are long-tailed, or even unexpected physics or new biological behaviors that were previously unexpected.

Our results suggest that the wild-type genome has a low failure rate in producing mature fruiting bodies, as observed using our methods. Previous reports have indicated that up to 10% of wild-type cell lines do not successfully germinate new colonies after development [21], but it is unclear if fruiting body morphogenesis is lacking in these failed sporulation assays or if biological factors such as disruptions in signaling pathways for cell differentiation into myxospores are the cause. Using time-lapse imaging, we found that only 2% of captured wild-type movies did not yield aggregation of fruiting bodies, indicating that factors other than physical aggregation contribute significantly to the failure of fruiting bodies to germinate. While the cause of the failure rate is unknown in cases where fruiting bodies mature under normal dynamics but remain unable to germinate wild-type colonies post-starvation, the quantification of this failure rate is crucial for mechano-physio modeling.

It is important to confirm that data deviating significantly from the mean in studies of stochastic systems are not just the result of inconsistent measurement conditions but instead represent genuine and rare behavior. To achieve this, the metrics used in this study were selected to minimize dependence on imaging setup. Additionally, manual vetting was conducted on data points that fell on the margins of PCA phenotype space. This approach emphasizes the significance of evaluating exceptional data to ensure accurate and reliable results.

In this work, we present methods that can be applied to any stochastic system that has many comparable replicates and multiple relevant metrics measured for each replicate. This approach can be compared to machine learning, which is powerful but lacks transparency in how categorization is achieved and is often sensitive to small variations in image acquisition that can vary from day to day or lab to lab. In cases in which image processing can measure clearly apparent aspects of a system, our method greatly simplifies the unprocessed image data to preserve phenotypic relevance over image acquisition noise. Further reduction of the phenotypic dataset from ten to two dimensions avoids the problem of the "curse of dimensionality" [22], reproducibly preserving relevant similarities and differences between datapoints to reveal overall structure and guide new investigations. Other types of time series image data are especially amenable to analysis with our methods, particularly with large datasets.

In choosing the metrics to quantify developmental phenotype, we have observed that the multimodal behavior of wild-type was revealed when considering both aggregation timing and the size of aggregates at multiple developmental timepoints. We found that the choice of developmental metrics was best guided by attempting to find metrics that include not just final phenotypic state but also snapshots of development at various times, revealing differing developmental pathways. We also observed the importance of choosing metrics that are robust to changes in imaging parameters by not relying on pixel-specific information, but rather on concrete quantities such as aggregate size. These principles also improve the likelihood of replication in other laboratories.

Although this study analyzed 60 or more replicates for each mutant strain, it is possible to differentiate a strain from wild-type with fewer replicates. By examining the distribution of many subsamples of wild-type aggregation in PCA space, we discovered that only 15 replicates collected over two days are required to detect deviations from wild-type behavior for each of the mutant strains mentioned ($p < 0.05$). Notably, with this same sample size, conventional statistical tests that rely on individual metrics (such as average aggregate area after 24 h) can only distinguish DK7517 as different from wild-type, and with less statistical power than the method utilized in this study. Since the distribution of final mean areas is not Gaussian, as determined by a Shapiro-Wilk normality test, we selected the Kolmogorov-Smirnov test as the standard test for distinguishing two distributions, rather than a Student's t-test, which assumes normality of the underlying distributions.

Our study found that the use of common reporters has indirect effects on fruiting body formation dynamics in *M. xanthus*. Reporter genes, such as GFP and Tn5 lac, are frequently used to mark successful transfection and for quantitative assays. However, the attachment of these tags to a molecule of interest can cause additional molecular noise and nontarget effects [23–25], leading to stochastic variation in the overall cell population [26]. The Tn5 lac reporter gene, which was designed to identify gene expression during *M. xanthus* fruiting body morphogenesis, can promote adjacent deletions [27] and disrupt regulatory regions, leading to changes in phenotype [28]. While the use of reporter genes is valuable, their impact should be assessed, particularly in biophysical studies that focus on developmental dynamics where differences may be more apparent.

The process of identifying gene function by the presence of a strong mutant phenotype, although extensively used, has fundamental limitations. It is generally exceptional for a single gene disruption to undeniably affect phenotype, and the effects of most genes will be more subtle or masked. Indeed, the process of canalization, in which evolution tends to push towards developmental phenotypes that are robust to genetic and environmental variation, was early observed as a universal phenomenon across multiple domains of life [29]. This fundamental robustness is a central part of G2P, and moving beyond a qualitative description to a quantitative framework capable of measuring incremental rather than abrupt phenotypic changes is a necessary step we take in this work.

The results suggest a gene annotation method that is sensitive enough to detect the impact of single gene mutations that would otherwise go unnoticed. Each mutant strain has a signature distribution in a PCA space, which can be used to identify strains with similar functions based on their impact on development. These signatures can be compared to those of well-understood genes to assign a function to unknown genes. The signatures are independent of any specific biological model and rely solely on visually observable traits. As more experiments are done using this method, a quantitative phenome for fruiting body development can be developed, which creates a common language for gene function comparison. Uncommon developmental phenotypes can expose the regulatory mechanisms of fruiting body formation, and studying these "exceptions" can provide valuable insights into the "rule."

We anticipate that slight changes in experimental protocol can have subtle yet detectable impacts on behavior. In our experiments, we discovered a new variable that has not traditionally been controlled for in *M. xanthus* culture: the age of the agar plate from which colonies are harvested for liquid bacterial culture. While it is expected that reintroducing bacteria to liquid culture will reset their metabolic state, our analysis found that colonies from agar plates at least three days old produce fewer fruiting bodies (see the Supplemental Material, Fig. S5) [17]. This suggests that colonies "remember" the age of the plate they were harvested from, although the underlying mechanism is unknown. We expect that other protocol variables may also have measurable phenotypic effects.

The methods presented in this study are sensitive and can be used to measure how organisms respond to changes in environmental variables. Like running replicates of single-gene mutants, running replicates in different environments can reveal corresponding changes in phenotype. In the future, fruiting body aggregation can be observed on agar surfaces with a range of stiffness values. The resulting shift in the phenotypic distribution would reveal the extent to which this aspect of the environment affects aggregation, such as the likelihood of Mode 1 or Mode 2 behavior, and the similarity of this shift to that observed in mutant strains can be analyzed to test potential mechanisms. This approach can address the missing environmental information of the genotype-phenotype problem and broaden the understanding of "wild-type behavior" as a function of environmental conditions. By identifying subregions in phenotype space with a root cause, the contour bounding abnormal wild-type behavior can provide a measure of the relevant physical and biological mechanisms that have yet to be characterized and help in organizing knowledge gained from future experiments.

IV. CONCLUSIONS

In this study, we have shown that wild-type DK1622 *M. xanthus* exhibits significant variation in fruiting body morphogenesis, with a consistent bimodal pattern of dynamics even held under the most stringent and established experimental conditions. The observed bimodal variation suggests the existence of differences in collective cellular fates that may be influenced by unknown stimuli. Our findings emphasize the importance of controlling experimental variables when characterizing cellular behavior, particularly when investigating single-gene mutations.

Furthermore, our investigation of modified strains carrying reporter genes revealed that the bimodal distribution was suppressed, indicating that the wild-type distribution is genotype-specific. These results suggest that specific developmental stages and requirements may be necessary prior to phenotype bifurcations. The process of fruiting body morphogenesis in wild-type *M. xanthus* is not a clockwork example, but rather it expresses significant variation and genotype-specific distributions. While some aspects may be random, much of the process appears to be biologically regulated and dependent on specific developmental stages and requirements.

Our method of analysis proved effective, with reliable results obtained from a small number of replicates. The visual and quantitative nature of our approach makes it an ideal language for data presentation, with potential applications in gene annotation and investigations of the genotype-phenotype problem under different environmental conditions. Overall, our study provides a foundation for understanding the impact of single-gene disruptions on *M. xanthus* development and lays the groundwork for future investigations into the complex interplay between genes, environment, and cellular behavior. The identification of extreme phenotypes exhibited during our analysis can potentially provide insight into the state of cellular processes during morphogenesis.

V. METHODS

A. Imaging setup

An array of compact microscopes controlled by a central computer can simultaneously collect time series images for 96 experiments. Each microscope is equipped with a single 4× objective lens, a Peltier device for maintaining stage and sample temperature, a red-light source, and a camera controlled by a Raspberry Pi. The microscopes are held in place by a 3D-printed armature and assembly hardware, which also provides a focus knob for higher image quality. Details on microscope design are included in the Supplemental Material, with additional required hardware detailed in Table S3 [17].

To ensure uniform control and central storage of all microscopes' time series output, each Raspberry Pi unit is networked via ethernet and two 64-port network switches to a central hub computer. The computer runs Piserver software that allows software to be changed and updated for all Raspberry Pi units simultaneously. Custom Python software with a convenient GUI is used to control image acquisition from each camera via SSH and organize output in a centralized image storage location. All custom software is available via Github at github.com/masp01/3D-scope-myxo-tracking.

B. Cell culture

To recover long-term stock cultures, nutrient rich CTTYE media agar containing 1% Casein Peptone, 0.5% Bacto Yeast Extract, 10 mM Tris (pH 8.0), 1 mM $KH(H_2)PO_4$ (pH 7.6), and 8 mM $MgSO_4$ was used. The harvested cells were used to inoculate broth cultures in CTTYE with vigorous shaking at 32 °C and grown to an approximate density of 4×10^8 cells/mL (100 Klett or 0.7 A_{550}).

The cells were centrifuged and washed with TPM buffer [10 mM Tris (pH 7.6), 1 mM $KH(H_2)PO_4$, 8 mM $MgSO_4$] before resuspension to a final concentration of 4×10^9 cells/mL. For the development assay, approximately 4×10^7 cells (10 μL aliquots) were spotted onto a TPM agar slide, a nutrient-limited medium, and incubated on the microscope stage at 32 °C for 24 h. The TPM slides were prepared as previously described [30].

The strains used in this study are as follows:

Strain	Description
DK1622	To establish a stable baseline for fruiting body development assays, the nominal wild-type strain used in our experiments was genetically modified from a naturally occurring *M. xanthus* isolate [31]. This modification was necessary as strains directly isolated from soil have a high rate of developmental failure in a laboratory setting. It is important to note that the genetic modification was only performed to establish a more reliable baseline for our experiments, and it did not affect any other aspects of the strain's behavior or characteristics.
DK10546	The constitutively expressing GFP labeled strain is used to track motility and cell dynamics during development. Used as an experimental control for fluorescence microscopy, the construct was generated fusing a copy of the *pilA* promoter to the coding sequence of GFP (*pilAp*-GFP), which was then reinserted into the *M. xanthus* chromosome [32]. This study showed an increased likelihood of early aggregation with many dispersing aggregates over wild-type for this strain.
DK6665	The Tn5 Ω6658 sasB7 mutant was generated from a mutation created in suppressor developmental gene *sasB* [33]. Previous work observed no visible phenotypic impact on the mutant as the strain can still proceed through development via other regulatory channels. This study observed initial cell clumps having an unusually high impact on final aggregates due to a lack of dispersal of initial cell clumps relative to wild-type.
DK4322	The spiA::Tn5-lacZ strain is a reporter fusion for the developmental gene *spi* with lacZ for β-galactosidase assays. The *spi* gene has been shown to be induced at 2 h into development and is developmentally regulated by C signal pathways. In previous work, the transposon insertion was characterized as not interfering with development or affecting spore production [34]. This study showed a higher likelihood of irregular aggregate shapes at early times for this strain than wild-type.
DK7517	Generated via a Tn5-LacZ insertion into a TA synthesis gene, as a reporter gene involved in toxin and antitoxin production also for β-galactosidase assays [35] to isolate regulatory mutants. This reporter fusion was shown to be expressed during vegetative growth while peaking during lag phase. In this study, late and immature aggregates were more likely to develop in this strain than wild-type.

C. Image processing pipeline

Phenotype was automatically quantified for each fruiting body aggregation assay in this study by running 144 individual TIFF images (10 min between each frame over 24 h of total development) from each time series through a custom Python image processing and analysis pipeline to identify in each frame which pixels could belong to a fruiting body, based on their gray value. The analysis code is available via Github at github.com/masp01/3D-scope-myxo-tracking. The information for the position and geometry of each aggregate was filtered to remove noise and spurious aggregates. This detailed data summary for each time series then had a list of 10 specific numbers extracted from it, each of which captures one overall feature, such as the time at which aggregation began or the average size of the final fruiting bodies. The values of these 10 metrics together were then used in further analysis. The full details of all phenotypic metrics are available in the Supplemental Material, Table S1 [17].

D. Statistical methods

To calculate p-values that test the null hypothesis of mutant development datapoints in PCA space being drawn from the same distribution as the wild-type development datapoints, we first generate the contours for the wild-type PCA data [Fig. 3(c)] by starting with Gaussian kernel density estimation (KDE) and using standard root-finding techniques to draw contours from the density estimate that capture 50% and 90% of the PCA datapoints. An appropriate kernel size for the KDE is validated by using 75% of the wild-type dataset, and ensuring that, across many subsamples of the remaining 25% (verification data), the distribution of enclosed points is centered on the appropriate percentage. When this distribution is skewed, it indicates overfitting of the original contour. With these contours drawn, we use a data-driven statistical technique based on bootstrapping. Given a sample size N, 10 000 random samples of that size are drawn from the wild-type dataset, which contains 569 datapoints sampled over 25 different days. Each subsample has a characteristic pair of numbers, (n_{50}, n_{90}), which corresponds to the number of points in the sample that fall inside the 50% and 90% contours, respectively. Once the distribution of these pairs for wild-type data is known, n_{50} and n_{90} are calculated for a sample of mutant PCA datapoints of size N. The fraction of wild-type videos that have both n_{50} and n_{90} greater than the mutant sample's values of n_{50} and n_{90} gives

the p-value, or the probability that a sample of wild-type data of size N would exhibit the same distribution. Contours for mutant strains are shown in Fig. 4 for visualization only, and they do not figure into the calculation of the p-values.

ACKNOWLEDGMENTS

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

M.E.A. and E.A.C. contributed equally to this work.

[1] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, Stochasticity in gene expression: From theories to phenotypes, Nat. Rev. Genet. **6**, 451 (2005).

[2] J. Paulsson, O. G. Berg, and M. Ehrenberg, Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation, Proc. Natl. Acad. Sci. USA **97**, 7148 (2000).

[3] M. Thattai and A. van Oudenaarden, Stochastic gene expression in fluctuating environments, Genetics **167**, 523 (2004).

[4] J. L. Spudich and D. E. Koshland, Non-Genetic individuality: Chance in the single cell, Nature (London) **262**, 467 (1976).

[5] M. Thattai and A. van Oudenaarden, Intrinsic noise in gene regulatory networks, Proc. Natl. Acad. Sci. USA **98**, 8614 (2001).

[6] J. Muñoz-Dorado, F. J. Marcos-Torres, E. García-Bravo, A. Moraleda-Muñoz, and J. Pérez, Myxobacteria: Moving, killing, feeding, and surviving together, Front. Microbiol. **7**, 1 (2016).

[7] L. Kroos, Highly signal-responsive gene regulatory network governing myxococcus development, Trends Genet. **33**, 3 (2017).

[8] D. R. Zusman, A. E. Scott, Z. Yang, and J. R. Kirby, Chemosensory pathways, motility and development in *Myxococcus xanthus*, Nat. Rev. Micro. **5**, 862 (2007).

[9] P. G. Livingstone, A. D. Millard, M. T. Swain, and D. E. Whitworth, Transcriptional changes when *Myxococcus xanthus* preys on *Escherichia coli* suggest myxobacterial predators are constitutively toxic but regulate their feeding, Microb Genom **4**, e000152 (2018).

[10] L. Jelsbak and L. Søgaard-Andersen, Pattern formation: Fruiting body morphogenesis in *Myxococcus xanthus*, Curr. Opin. Microbiol. **3**, 637 (2000).

[11] G. Diss, D. Ascencio, A. DeLuna, and C. R. Landry, Molecular mechanisms of paralogous compensation and the robustness of cellular networks, J. Exp. Zool. B Mol. Dev. E **322**, 488 (2014).

[12] H. Vuong, K. Shedden, Y. Liu, and D. M. Lubman, Outlier-based differential expression analysis in proteomics studies, J Proteomics Bioinform. **4**, 116 (2011).

[13] E. R. Dougherty and I. Shmulevich, On the limitations of biological knowledge, Curr. Genomics **13**, 574 (2012).

[14] J. C. S. McCoy, J. I. Spicer, Z. Ibbini, and O. Tills, Phenomics as an approach to comparative developmental physiology, Front. Physiol. **14**, 1229500 (2023).

[15] R. M. Bilder, F. W. Sabb, T. D. Cannon, E. D. London, J. D. Jentsch, D. S. Parker, R. A. Poldrack, C. Evans, and N. B. Freimer, Phenomics: The systematic study of phenotypes on a genome-wide scale, Neuroscience **164**, 30 (2009).

[16] R. Welch and D. Kaiser, Cell behavior in traveling wave patterns of myxobacteria, Proc. Natl. Acad. Sci. USA **98**, 14907 (2001).

[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PRXLife.1.023011 for fabrication details of the high-throughput microscope array; details on metric distributions across the wild-type dataset and for the two wild-type modes; time series of extreme developmental phenotypes; p-values for deviation from wild-type behavior calculated from the phenotype distributions; phenotype distributions as a result of differences in culture age at time of inoculation; full formulas for each developmental metric used to quantify phenotype; and weights for each metric as used in the PCA.

[18] N. B. Caberoy, R. D. Welch, J. S. Jakobsen, S. C. Slater, and A. G. Garza, Global mutational analysis of NtrC-Like activators in *Myxococcus xanthus*: Identifying activator mutants defective for motility and fruiting body development, J. Bacteriol. **185**, 6083 (2003).

[19] J. Yan, M. D. Bradley, J. Friedman, and R. D. Welch, Phenotypic profiling of ABC transporter coding genes in *Myxococcus xanthus*, Front. Microbiol. **5**, 1 (2014).

[20] K. Sankaran and S. P. Holmes, Latent variable modeling for the microbiome, Biostatistics **20**, 599 (2019).

[21] M. D. Bradley, D. Neu, F. Bahar, and R. D. Welch, Inter-Laboratory evolution of a model organism and its epistatic effects on mutagenesis screens, Sci. Rep. **6**, 1 (2016).

[22] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, When is "nearest neighbor" meaningful?, in *Database Theory—ICDT'99*, edited by C. Beeri and P. Buneman (Springer-Verlag, Berlin, Heidelberg, 1999), pp. 217–235.

[23] U. N. Singh, Polyribosomes and unstable messenger RNA: A stochastic model of protein synthesis, J. Theor. Biol. **25**, 444 (1969).

[24] O. G. Berg, A model for the statistical fluctuations of protein numbers in a microbial population, J. Theor. Biol. **71**, 587 (1978).

[25] M. S. H. Ko, A stochastic model for gene induction, J. Theor. Biol. **153**, 181 (1991).

[26] F. M. F. Nunes, A. C. Aleixo, A. R. Barchuk, A. D. Bomtorin, C. M. Grozinger, and Z. L. P. Simões, Non-Target effects of green fluorescent protein (GFP)-Derived double-stranded RNA (DsRNA-GFP) used in honey Bee RNA Interference (RNAi) Assays, Insects **4**, 90 (2013).

[27] R. A. Jilk, J. C. Makris, L. Borchardt, and W. S. Reznikoff, Implications of Tn5-associated adjacent deletions, J. Bacteriol. **175**, 1264 (1993).

[28] A. M. Kierzek, J. Zaim, and P. Zielenkiewicz, The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression, J. Biol. Chem. **276**, 8165 (2001).

[29] T. Flatt, The evolutionary genetics of canalization, Q. Rev. Biol. **80**, 287 (2005).

[30] R. G. Taylor and R. D. Welch, Recording multicellular behavior in *Myxococcus xanthus* biofilms using time-lapse microcinematography, J. Visualized Exp. **42**, e2038 (2010).

[31] D. Kaiser, Social gliding is correlated with the presence of pili Myxococcus xanthus, Proc. Natl. Acad. Sci. USA **76**, 5952 (1979).

[32] D. Wall and D. Kaiser, Alignment enhances the cell-to-cell transfer of pilus phenotype, Proc. Natl. Acad. Sci. USA **95**, 3054 (1998).

[33] H. B. Kaplan, A. Kuspa, and D. Kaiser, Suppressors that permit a-signal-independent developmental gene expression in Myxococcus xanthus, J. Bacteriol. **173**, 1460 (1991).

[34] L. Kroos, A. Kuspa, and D. Kaiser, A global analysis of developmentally regulated genes in *Myxococcus xanthus*, Dev. Biol. **117**, 252 (1986).

[35] M. Varon and E. Rosenberg, Transcriptional regulation of genes required for antibiotic TA synthesis in *Myxococcus xanthus*, FEMS Microbiol. Lett. **136**, 203 (1996).