bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Multispecies phase diagram reveals biophysical principles of bacterial biofilm architectures

Hannah Jeckel,[1,2,3,*] Francisco Díaz-Pascual,[1,*] Dominic J. Skinner,[4,*] Boya Song,[4,*] Eva Jiménez Siebert,[1,2,3] Eric Jelli,[1,2] Sanika Vaidya,[1] Jörn Dunkel,[4,†] and Knut Drescher[1,2,3,†]

[1] *Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Str. 16, 35043 Marburg, Germany*
[2] *Department of Physics, Philipps-Universität Marburg, Karl-von-Frisch-Str. 16, 35043 Marburg, Germany*
[3] *Biozentrum, University of Basel, Spitalstrasse 41, 4056 Basel, Switzerland*
[4] *Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 01239, USA*

Bacterial biofilms are among the most abundant multicellular communities on Earth and play essential roles in a wide range of ecological, medical, and industrial processes. Recently developed imaging techniques offer unprecedented insights into the three-dimensional internal structure and external morphology of growing biofilms, but general ordering principles that govern the emergence of biofilm architecture across species remain unknown. Here, we combine experiments, simulations, and topological analysis to identify universal mechanical interaction properties that determine early-stage biofilm architectures of different bacterial species. Performing single-cell resolution imaging of *Vibrio cholerae*, *Escherichia coli*, *Salmonella enterica*, and *Pseudomonas aeruginosa* biofilms, we discovered that biofilm architectures up to a few thousand cells can be described by a two-dimensional phase diagram similar to nematic liquid crystals. Mechanistic simulations and experiments using single-species mutants for which the cell aspect ratio and the cell-cell adhesion are systematically varied, show that tuning these parameters reproduces biofilm architectures of different species. A topological analysis of biofilm architectures across species further reveals that cell neighborhood motifs can be described by a universal Tracy-Widom distribution. More generally, due to its generic mathematical formulation, the topological analysis framework enables a structural comparisons and classification of a wide range of multicellular life forms. Early-stage biofilm architectures of different species therefore display a universal topological structure, and their development is determined by conserved mechanical cell-cell interactions.

Bacterial biofilms are multicellular communities that grow on surfaces within a self-produced extracellular matrix [1, 2]. Biofilms are remarkably robust against mechanical and chemical perturbations, and they are one of the most abundant forms of microbial life on Earth [3]. Major research efforts over the past two decades [4–7] have established the ecological, biomedical and industrial importance of bacterial biofilms. Yet, it is not well understood how multicellular functions of biofilms, such as mechanical stability or antibiotic tolerance, arise from the collective growth and spatiotemporal self-organization of these communities. Recent advances in live imaging techniques now make it possible to observe the development of early-stage biofilms at single-cell resolution, starting from a single founder cell up to a few thousand cells [8–11]. Imaging-based studies have provided key insights into the importance of mechanical cell interactions [10, 12–18], cell surface attachment [14, 19–23], growth memory [11], external fluid flow [24–26], and external mechanical environment [27–29] on the emergent architecture in *V. cholerae* biofilms. Despite such progress, however, it remains an open question whether there exist common biophysical principles that govern biofilm development across species.

To tackle this longstanding problem, we report here a combined experimental and theoretical investigation of three-dimensional (3D) biofilm architectures for several bacterial species: *V. cholerae*, *E. coli*, *S. enterica* and *P. aeruginosa*. Each of these species displays different growth characteristics, extracellular matrix components, cell morphology, and different biofilm architectures [30, 31]. Building on recent tools for 3D biofilm imaging analysis [32], our data enables a comprehensive statistical characterization of both single-cell properties and emergent collective properties that determine biofilm architecture across species. To quantitatively compare and match experimentally observed biofilms and numerically simulated biofilms, we formulated an improved statistical metric framework based on Chebyshev representations of the experimentally measured parameter distributions, because previous methods based on the assumption of normally-distributed data [10] were unable to robustly detect and resolve the differences in the biofilm architectures of the different strains. Due to its generic mathematical formulation, the underlying methodology will be broadly applicable to characterize and compare other prokaryotic and eukaryotic multicellular systems.

* These authors contributed equally
† To whom correspondence should be addressed. E-mail: dunkel@mit.edu, knut.drescher@unibas.ch

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To enable a topological structure analysis of biofilm architectures, we also generalized a recently introduced topological metric framework [33], by extending a 2D planar flip graph construction [33, 34] to 3D Delaunay networks.

Through the quantitative biophysical and topological analysis methodology outlined above, we find that emergent architectural differences across biofilms correlate with variations in cell geometry and local cell density. As a result, early-stage biofilms from different species can be embedded within a two-dimensional phase diagram spanned by the cell aspect ratio and the cell density, which is reminiscent of classical liquid crystal systems [35, 36]. Using mutants in a single species and particle-based computational modeling, we can independently explore the biophysical phase space of early-stage biofilm architectures, and demonstrate that cell aspect ratio and the cell-cell attraction jointly determine the emergent biofilm structure across species. Our topological analysis of 3D biofilms architectures from different species further shows that neighborhood size distributions are consistent with a Tracy-Widom law [37–39], revealing an unexpected universal structural property of bacterial biofilms.

RESULTS

Quantifying early-stage biofilm architecture across species. To investigate the structural differences between biofilm architectures within and across bacterial species, we performed single-cell resolution imaging. For each of the four species *E. coli*, *V. cholerae*, *P. aeruginosa*, and *S. enterica*, 15 biofilms were grown in microfluidic flow chambers from a few surface-attached founder cells to around 2000 cells, and then imaged by confocal microscopy (Fig. 1A; Materials and Methods). Although all species form colonies, the biofilm architectures of the four species are qualitatively different (Fig. 1A). To quantify the observed differences in biofilm shape and structure between species, we segmented all individual cells in all biofilms following Ref. [10]. Using BiofilmQ [32] for each biofilm, we measured single-cell quantities such as cell length, cell diameter, and cell convexity, together with emergent collective properties, such as local cell number density and nematic order, resulting in a histogram for every one of the $m = 16$ measured observables (see SI Appendix section 1 for a complete list). Each biofilm is thus represented by a set of m histograms.

Using this collection of. histograms to quantify the architectural properties of each biofilm, we sought a systematic approach to compare sets of histograms, and thereby compare biofilm architecture. To this end, we represented each empirically measured histogram with a Chebyshev polynomial of degree $d = 20$ using kernel density estimation (SI Appendix section 1). Replacing ~ 2000 single cell measurements for each biofilm and each parameter with $d + 1 = 21$ polynomial coefficients allowed us to compress the experimentally observed data whilst retaining information about their distributions beyond mean values and variances. From a $(d+1) \times m$ matrix containing all the Chebyshev coefficients for a given biofilm, we constructed a Chebyshev dissimilarity (Cd) measure, to compare two such matrices and hence two biofilms (SI Appendix section 2). Mathematically, Cd provides an upper bound on the cumulative L_1-distance between collections of histograms. We found that, compared with previously used mean- and variance-based measures [10], Cd more robustly distinguishes biofilms from different species, as evident from the block structure in the upper right half of Fig. 1B. Similarly, taking a vector of Chebyshev coefficients constructed from a single property across all biofilms, allows us to apply Cd to compare similarities of measured observables (SI Appendix section 2). To prevent double-counting, in the case of highly correlated observables, as identified using Cd, only one was kept, leaving $p = 13$ essential properties which characterize biofilm architecture (Fig. 1B, left lower half).

Phase diagram of early-stage biofilm architecture. Principal component analysis (PCA) applied to the flattened $(d + 1) \times p = 21 \times 13$ dimensional vectors of Chebyshev coefficients representing each biofilm revealed that there are four distinct clusters corresponding to the four bacterial species (Fig. 1C). The information contained in the $p = 13$ distributions of measured parameters is therefore sufficient to capture the key architectural differences between species. To investigate which of the measured observables could be responsible for the inter-species variation, we examined the contributions of each observable to the first principal component (SI Appendix section 2). The feature that contributed the most is the local cell number density, defined as the number of neighbors a cell has within a 2 μm radius, while the second highest contributing feature was the cell aspect ratio, suggesting that variation in these parameters across biofilms could be responsible for variation in the observed architectures. To verify that these two observables provide a suitable biophysical phase diagram of biofilm architecture, we plot each biofilm in the mean cell number density vs. mean cell aspect ratio plane (Fig. 1D). The clear separation of the four species in this two-dimensional phase space shows that biofilm architectures can be efficiently characterized by these two parameters. We note that classical liquid crystals can also be characterized by an aspect ratio vs. number density phase diagram [35, 36], which highlights an interesting analogy between passive nematic structures, and growth-active nematic biofilms.

Altering biofilm architecture with cell aspect ratio and cell-cell adhesion mutants. The four species analyzed in Fig. 1 differ in a large number of

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

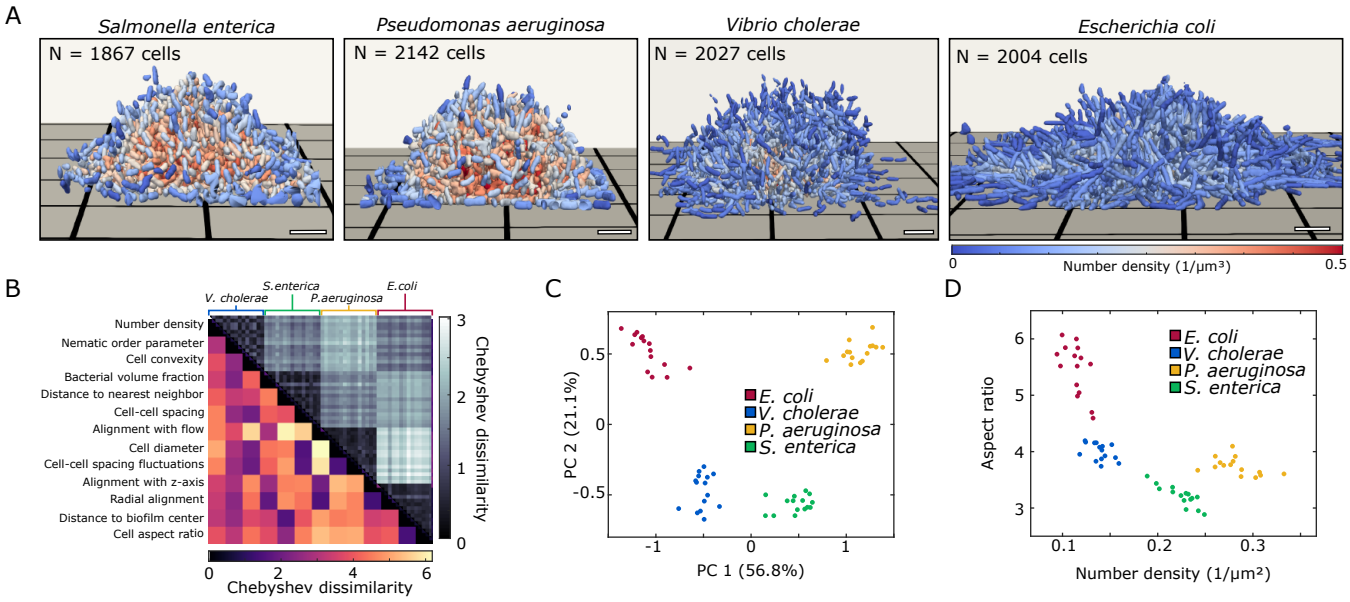

FIG. 1. Early-stage biofilm architectures of different bacterial species can be distinguished within a two-dimensional phase diagram obtained by a statistical analysis of single-cell resolution measurements. (A) Representative 3D biofilm architectures of four bacterial species reconstructed from segmented confocal microscopy images at comparable cell numbers ~ 2000. Each cell is colored according to the local density within a radius $2\,\mu$m. Scale bars, $5\,\mu$m. (B) By approximating the distributions of 13 relevant observables that were computed from single-cell measurements (SI Appendix section 1) for each biofilm with Chebyshev polynomials, a Chebyshev dissimilarity (Cd) measure for biofilms and observables is defined (SI Appendix section 2). The Chebyshev dissimilarity compares the absolute distance between distributions and enables the robust comparison and differentiation of biofilm architectures from different species. (C) Principal component analysis (PCA) based on the Chebyshev coefficient space (SI Appendix section 2) robustly distinguishes biofilms of the four different species *E. coli*, *V. cholerae*, *P. aeruginosa*, and *S. enterica*, revealing the cell aspect ratio and local number of neighbors as the key contributors to the first principal components (SI Appendix section 2). (D) Representing the experimental data in the mean aspect ratio vs. cell number density plane confirms that these two observables define a biophysically interpretable phase diagram to categorize biofilm architectures.

biological properties beyond cell aspect ratio and number density. To test if cell aspect ratio and local density not only correlate with but also determine the different biofilm architectures observed across the four species, we generated several mutants in a single species, *V. cholerae*. By analyzing the biofilm architectures that arise from mutants within a single species, it is possible to isolate the effects of cell aspect ratio and local density on the biofilm architecture. To this end, we generated mutations in *mreB*, following [40], which resulted in different aspect ratios compared to the parental strain (Fig. 2A). To control the cell density, we introduced mutations that alter the abundance of the cell-cell attraction-mediating matrix protein RbmA [8, 10]; specifically, we deleted the native *rbmA* gene from the chromosome, and re-introduced a copy of *rbmA* under the control of a promoter that is inducible by the monosaccharide arabinose (Materials and Methods). By growing the cells in the presence of different levels of arabinose, we can therefore tune the level of the cell-cell attraction, resulting in different cell number densities (Fig. 2B). We introduced this *rbmA* mutation in the parental *V. cholerae* strain, as well as in strains with smaller aspect ratios, to perform a comprehensive experimental scan over the space of cell sizes and densities (Fig. 2C-E). Within the experimentally accessible parameter space of cell aspect ratio and arabinose concentrations, biofilms with approximately the same cell number ($N_{\text{cell}} \sim 2000$) were found exhibit major architectural differences that can be recapitulated in simulations of a mechanistic biofilm model (Fig. 2C).

Computational model based on mechanical interactions reproduces experimental biofilm architectures. Cell aspect ratio and cell-cell attraction, which were systematically varied for *V. cholerae* experimentally (Fig. 2A-C), are key parameters for the mechanical cell-cell interactions. To test if the effect of these parameters on the biofilm architecture is primarily due to changes in mechanical cell-cell interactions, we compared the experimental measurements for the *V. cholerae* strains with a computational model in which cells only interact mechanically. In this model, which extends a previously introduced simulation framework [10, 24], individual cells are represented as growing, dividing ellipsoids which experience pairwise cell-cell interactions and

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

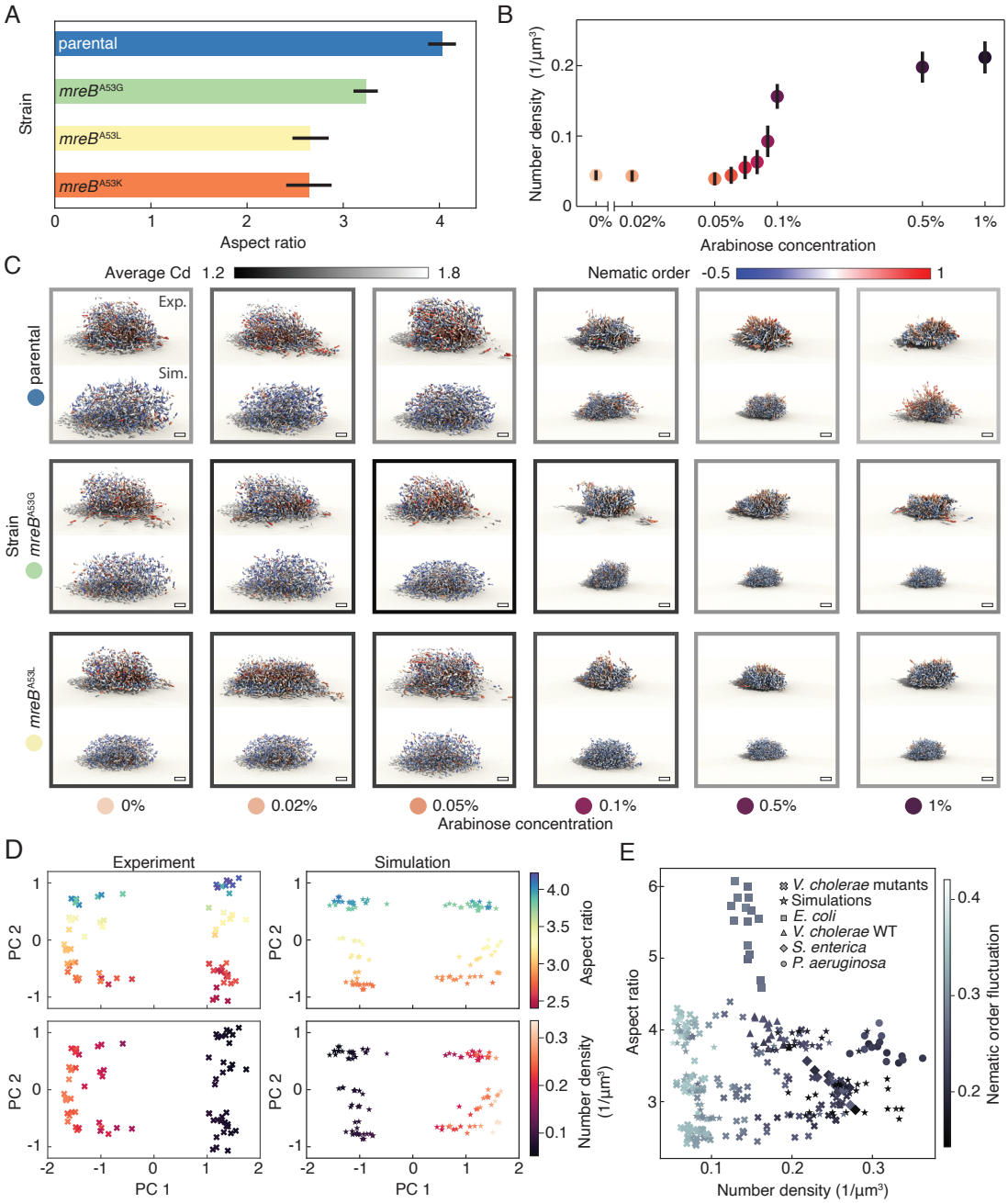

FIG. 2. Differences in biofilm architecture are driven by cell aspect ratio and cell-cell adhesion in *V. cholerae*. (A) Using different *V. cholerae* strains, the cell aspect ratio (left panel) and number density in the biofilm (right panel) can be experimentally varied independently. The cell aspect ratio was measured for $n = 30$ biofilms for each strain, including strains with point mutations in the *mreB* gene. Error bars indicate standard deviations. (B) The local number density was measured for biofilms that were grown in the presence of different arabinose concentrations ($n = 12$ biofilms for each arabinose concentration). Measurements were performed and averaged for $\Delta rbmA$ strains with the wild type *mreB*, $mreB^{\mathrm{A53K}}$, $mreB^{\mathrm{A53L}}$, and $mreB^{\mathrm{A53G}}$ harboring a plasmid with the arabinose-inducible P_{BAD}-*rbmA* construct. Data points are colored according to the arabinose concentration, corresponding to the horizontal axis in panel (C). Error bars indicate standard deviations. (C) Renderings of *V. cholerae* biofilms (top, $N_{\mathrm{cell}} \sim 2000$) and corresponding best-fit simulations (bottom, $N_{\mathrm{cell}} = 2000$) for combinations of three different mutants (arranged vertically) and six different arabinose concentration levels (arranged horizontally). Each cell in the biofilms is colored by the local nematic order around it. The outline of each grid panel is colored in grayscale by the average Chebyshev dissimilarity (Cd) between the corresponding experiments ($n = 3$) and the best-fit simulation (see Fig. S7 for the exact values). Scale bar, $5\,\mu$m. (D) Two-dimensional PCA embedding of the Chebyshev features of $n = 72$ *V. cholerae* mutant biofilms and a group of $n = 114$ simulations consisting of the top 5 best-fitting simulations for each grid point in panel (C). The PCA embedding is colored by average aspect ratio (top) and average local number density (bottom) of all the cells in each biofilm, confirming that these two parameters are principal determinants of biofilm architecture, consistent with Fig. 1C. (E) Joint phase diagram combining experimental biofilms from the four different species shown in Fig. 1C with the experimental and simulated biofilms for *V. cholerae* mutants from panels C and D, colored by the variance of the nematic order parameter.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cell-surface interactions that determine their overdamped positional and orientational dynamics. The cell-cell interactions account for both short-range steric repulsion together with RbmA mediated attraction [10, 24]. In addition to cell-surface steric repulsion [10, 24], our simulations now also include an effective cell-surface attraction to account for the surface attachment of *V. cholerae* before and during biofilm formation [41, 42]. To further refine the previously introduced minimal model [10, 24], we implemented strongly anisotropic friction effects to account for the fact that the matrix polymer network can suppress the transverse motions of cells [43–45] (SI Appendix section 3). We generally found that the inclusion of the cell anchoring to the bottom surface and the anisotropic matrix-mediated friction leads to a substantially improved agreement between experimentally observed and simulated biofilms (Fig. 2B), when comparing their architectural properties in terms of the Cd measure (SI Appendix section 3).

To compare the experimental biofilm architectures of the *V. cholerae* mutants with the computational model, we performed systematic parameter scans to identify the values of simulation parameters which correspond to a given experimental system. Specifically, we performed $> 6,000$ simulations to search the parameter space of cell length at the time of division, width of cell-cell repulsion force, range of cell-cell attraction force, and strength of the cell-cell attraction (SI Appendix section 3), with the remaining parameters determined from previous experimental biofilm calibration [10, 24] (see Table S1). The best-fitting parameter values for a given experiment were determined by taking the values with the smallest Cd between experiment and simulation (Fig. S6). Using the fitted parameter values, we see a qualitative agreement between the experiment and simulation across various combinations of cell aspect ratio mutants and arabinose concentration levels (Fig. 2C). This agreement between the biofilm architectures obtained from the experimental parameter scan and the simulation parameter scan indicates that changes in cell aspect ratio and cell-cell attraction cause changes in the biofilm architecture through their effects on mechanical cell-cell interactions.

To understand whether the natural phase diagram of biofilm architectures for the different *V. cholerae* mutants and the mechanical simulations is, like the phase diagram of the different species (Fig. 1C), also based on the cell aspect ratio and cell number density we again performed PCA. Applying PCA to the vectors of Chebyshev coefficients for each experimental biofilm (Fig. 2D, left) and for each simulated biofilm (Fig. 2D, right), and coloring the data points by aspect ratio (Fig. 2D, top) and number density (Fig. 2D, bottom) reveals that these parameters exactly correspond to the first two principal components of this embedding, in both experiments and simulations (Fig. 2D). Therefore, the appropriate phase diagram of biofilm architectures of *V. cholerae* mutants and the simulation results is spanned by the aspect ratio and number density, consistent with the results for the different species in Fig. 1C and D.

Biofilm architecture of one species can be transformed into architecture of another species by changing control parameters of the phase diagram. Inspired by the characterization of liquid-crystals, where the location of a system on the aspect ratio – density plane determines the liquid crystal phase and nematic order [35, 36], we plot our all the experimental biofilms for the different strains and mutants together with our simulation results for the mutants in the aspect ratio – density phase plane (Fig. 2E). This phase diagram, in which an emergent property of the biofilm architecture is color-coded, shows that the biofilm architecture of *V. cholerae* can be modified to reproduce the biofilm architecture of other species, by simply tuning the control parameters of this phase diagram (cell aspect ratio and cell number density). Each species therefore inhabits a particular architectural phase in this aspect ratio – cell density space. The particular molecular structure of the extracellular matrix, which differs widely for the different species, affects this phase diagram only indirectly through the number density.

DISCUSSION

Topological characterization of 3D biofilm architectures distinguishes biofilms of different species. Traditional biophysical characterizations of multicellular structures typically rely on geometric information such as cell shape and distances between cells. An interesting complementary challenge is to understand the extent to which structural differences between biofilm architectures are encoded in local neighborhood information. Topological characterizations of disordered materials have emerged as a successful method to not only classify cellular structures [34, 46] but also to identify parametric embeddings and reveal organizing principles [33]. However, recently developed methods [33] focused on 2D structures and are therefore not applicable to the 3D cell packings in biofilms and other multicellular systems.

To develop a more general framework that can be used to compare 3D cell packings within and across biological systems, we started from the 3D Delaunay tessellation, which is the dual of the Voronoi diagram, using the centroid of each cell [47]. The local Delaunay tessellation around a point, which we refer to as a motif, topologically encodes the local neighborhood structure, whilst being unaware of specific information such as local density or cell lengths (Fig. 3A). Moreover, this tessellation is invariant under small perturbations, only changing through discrete events known as topological transitions [47]. These discrete transitions define a natural distance between two motifs, as the minimum number of

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

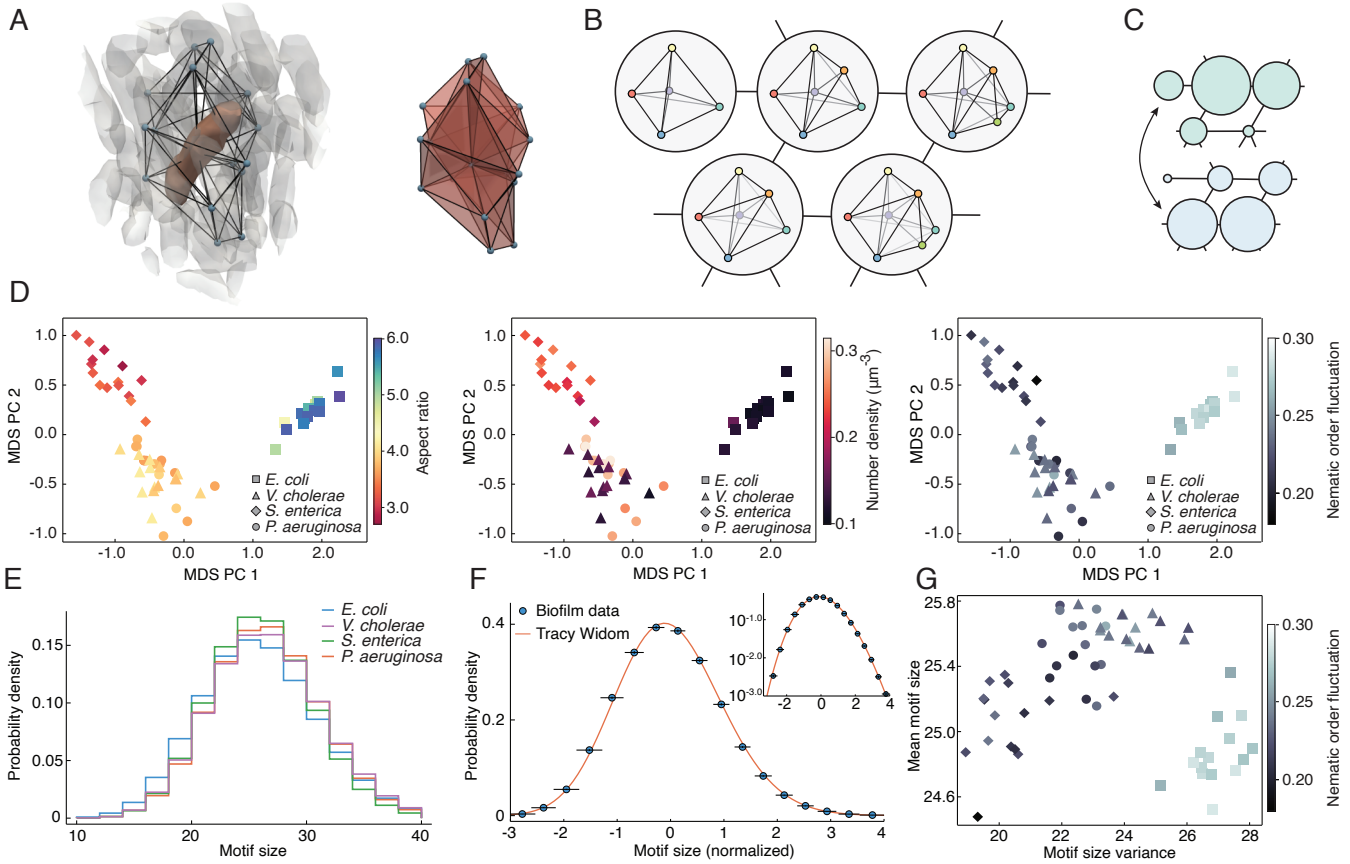

FIG. 3. Different bacterial species can be distinguished from local topological information alone, with cellular neighborhood motif sizes being well-described by a Tracy-Widom law. (A) For each cell (red, left), we compute the local Delaunay tesselation or motif (shown, right), which is the simplicial complex formed by all tetrahedrons of the Delaunay made from the cell and its neighbors (centroids of neighboring cells shown as blue spheres). (B) Motifs only change through discrete topological transitions, which naturally induces a graph structure where each vertex is a motif and vertices are connected if they are one transition apart, shown here for selected motifs. A single biofilm is characterized as a probability distribution of motifs over this graph. (C) Comparing two biofilms topologically requires comparing their two motif probability distributions. Example distributions are shown here with vertex size proportional to motif frequency. We compare topological motif distributions using a spectral distance that can be efficiently computed and which approximates the topological earth mover's distance [33] between two measured motif distributions (SI Appendix section 4). (D) Planar MDS embedding of biofilms from different species based on the spectral distance matrix that encodes the pairwise topological distance between experiments. The same embedding is colored differently in each of the three diagrams using mean cell aspect ratio (left), number density (middle), and nematic order fluctuations (right), respectively, showing that topological changes correlate strongly with changes in cell aspect ratio (left). (E) Histograms of motif sizes across different species have differing means and variances, but otherwise fall on a similar distribution. (F) Normalizing the motif-size distribution of each biofilm to have zero mean and unit variance, the combined histogram is well-described by a Tracy-Widom distribution (inset shows log scale plot). Circles represent the probability density of each bin, horizontal error bars the standard deviation. (G) Plotting each experiment in the motif-size mean–variance plane reveals that different species have distinct topological properties, and can therefore be distinguished by topological neighborhood information alone.

transitions to transform one motif into the other, which can also be interpreted as a minimum path distance on a graph of motifs where motifs are connected by an edge if they are a single transition apart (Fig. 3B). Since every cell in a biofilm has a motif associated to it, we can consider the topological encoding of a biofilm to be a probability distribution over the space of motifs, and comparing two biofilms requires comparing two probability distributions (Fig. 3C).

To perform this comparison for our experimental data, we used a spectral distance between two distributions [48], which can be interpreted as the number of transitions required to transform the first distribution into the second (SI Appendix section 4), similar to a topological earth mover's distance [33]. After calculating the pairwise distance between the 60 biofilms across the 4 species (Fig. 1), we constructed a low-dimensional embedding using Multi-Dimensional Scaling (MDS) [49].

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MDS is a generalized principal component analysis that embeds points from a general metric space, in our case the topological motif distributions representing the different biofilms, into Euclidean space by approximately preserving their distance structure (Fig. 3D). The embedding revealed that the topological structure of biofilms from the four species varies along a curve in the 2D embedding (Fig. 3D). Coloring the biofilm data in the 2D embedding according to different biofilm properties (Fig. 3D) reveals that the topological differences correlate strongly with changes in cell aspect ratio (Fig. 3D, left).

Topological motif-size distributions for biofilms of different species follow a universal Tracy-Widom law. Recent work [33] showed that topological motif sizes in 2D bacterial swarms display universal statistics that can be accurately described by a two-parameter Tracy-Widom distribution [37, 38]. It is currently unknown whether similar universal laws hold for the statistics of 3D cell packings. To explore this question, we measured the motif size for each cell within a given biofilm for each of the four bacterial species. The resulting motif-size distributions provide a condensed statistical characterization of the biofilm topology and appear to differ systematically across species (Fig. 3E). Rescaling all distributions so they have zero mean and unit variance, and plotting them in a common histogram, we see that they are well-approximated by the Tracy-Widom distribution (Fig. 3F). In particular, this observation suggests that the motif-size distributions in 3D biofilms can be characterized by their mean and variance. Indeed, plotting each biofilm in the plane spanned by the mean motif size and variance of motif size clusters the different species, showing that the topological local neighborhood structure contains sufficient information to identify biofilms of a particular bacterial species (Fig. 3G). The fact that the topological motif-size distribution for biofilms of all species investigated in this study can be captured by the same Tracy-Widom distribution indicates that there is a universal topological structure to the cellular packing in biofilms. To understand how biofilm growth achieves packings that follow this distribution, and whether 3D cell packings in other prokaryotic and eukaryotic systems [50] follow similar statistics poses an interesting challenge for future research.

Topological vs. geometric vs. geometric characterization of multicellular structures. The topological analysis in Fig. 3 shows that biofilms from different bacterial species can be distinguished based on the statistics of their local Delaunay motifs. Since Delaunay tesselations can be readily constructed for various types of 3D multicellular structures that can be imaged at single-cell resolution, including organoids [51] and tissues [52], the topo-statistical analysis framework introduced here presents a mathematically well-defined framework for comparing the 3D architectures of a wide range of eukaryotic and prokaryotic systems.

Complementing topological information with additional geometric information generally enables a finer distinction of multicellular structures. The Chebyshev dissimilarity (Cd) framework (SI Appendix section 4) utilized in Figs. 1 and 2 offers a systematic way of combining distributions of geometric quantities that have different units and ranges into a joint (dis)similarity measure. Similar to the Delaunay tesselation-based topo-statistical analysis, the Cd-based geometric approach can in principle be used to compare different biological systems whenever corresponding sets of single-cell and emergent collective observables (cell geometry, cell densities, orientational order parameters, etc.) are experimentally accessible.

CONCLUSIONS

By performing single-cell resolution imaging on early-stage bacterial biofilms of several bacterial species, we found that the emergent biofilm architecture correlates with differences in cell aspect ratio and local cell number density. By systematically varying the aspect ratio and cell-cell attraction using mutants of a single bacterial species, we then showed that these parameters determine the observed architectural differences. Extensive particle-based simulations of biofilm growth support this conclusion and further revealed that the impact of these parameters on the emergent biofilm architecture reflects the underlying effective mechanical cell-cell interactions. Our combined experimental and theoretical results show that bacterial biofilm architectures populate an aspect ratio – number density phase diagram, similar to classical liquid crystals. By changing the cell aspect ratio and number density of a particular species, this species can reproduce biofilm architecures of other species, even though the extracellular matrix composition can differ substantially between species. Furthermore, the generalization of a recently introduced topological analysis framework [33] to three dimensions led us to discover that the local neighborhood structure of biofilms of all analyzed bacterial species shares universal statistical properties. Together, our results suggest that despite the numerous biological differences between the species, general organizing principles of biofilm architecture development exist, and mechanical interactions play an important role; the full extent of these principles and their limitations have yet to be uncovered. From a more general perspective, neighborhood motif analysis could provide a quantitative route to comparing biofilms with other multicellular structures, which may reveal general topological ordering principles shared not only by prokaryotic, but also by eukaryotic multicellular systems.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

Bacterial strains and media. All *V. cholerae* strains used in this study are derivatives of a rugose variant of the O1 biovar El Tor wild type strain N16961 [53]. The *E. coli* strain used in this study (KDE2011) is a derivative of the AR3110 wild type [54], carrying a point mutation in the promoter of the gene *csgD*, which increases biofilm formation [55]. The *S. enterica* strain used here (KDS38) is a derivative of the UMR1 wild type [56], carrying a mutation in the promoter of *csgD* (formerly called *agfD* in *Salmonella*), which increases biofilm formation [57]. The point mutations in the *E. coli* and *S. enterica* strains were necessary to grow isolated biofilm colonies in our experimental conditions. The *P. aeruginosa* strain used here (KDP63) is a derivative of the PAO1 wild type [58] (obtained from Urs Jenal, Basel). The *V. cholerae*, *E. coli*, and *S. enterica* strains carried a plasmind with a P_{tac} promoter-driven production of sfGFP. The *P. aeruginosa* strain KDP63 carries a high-copy number plasmid producing the fluorescent protein YPet under the control of a pX2 promoter [59].

To engineer *V. cholerae* strains with a different cell length and width, amino acid 53 of the native MreB protein was replaced according to Monds *et al.* [40]. These modifications were introduced to the chromosome of *V. cholerae* by conjugation using the *E. coli* strain S17-1 λpir [60] and the pKAS32 suicide vector [61], containing *mreB* with the corresponding mutation and 500 bp upstream and 500 bp downstream from the codon that codes for amino acid 53 of MreB. To control the expression of *rbmA* in *V. cholerae*, inducible strains were created by conjugating a plasmid that contained P_{tac}-*sfGFP* and P_{BAD}-*rbmA* constructs. This plasmid enabled us to vary the production of RbmA by adding different concentrations of arabinose to the growth medium [10]. All strains, plasmids, and oligonucleotides that were used in this study are listed in the SI Appendix, Table S3, Table S4, and Table S5, respectively.

For overnight cultures or strain construction, cells were either grown in liquid Luria–Bertani–Miller broth (LB-Miller; 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, and 10 g L^{-1} NaCl) or LB-Miller without salt (10 g L^{-1} tryptone and 5 g L^{-1} yeast extract) with the corresponding antibiotic and shaking at 250 rpm, or using agar-solidified LB media (containing 1.5% agar). All *V. cholerae* biofilm experiments were performed in M9 minimal medium, with the following composition: M9 minimal salts (M6030, Sigma), 2 mM $MgSO_4$, 100 µM $CaCl_2$, MEM vitamins, 0.5% glucose, 15 mM triethanolamine (pH 7.1), and gentamicin (30 µg mL^{-1}). *E. coli* biofilm experiments were performed in tryptone broth (10 g L^{-1} tryptone) supplemented with kanamycin (50 µg mL^{-1}). *S. enterica* biofilm experiments were performed in tryptone broth supplemented with spectinomycin (100 µg mL^{-1}). *P. aeruginosa* biofilm experiments were performed in FAB medium, with the following composition: $CaCl_2$ (11 mg L^{-1}), $MgCl_2$ (93 mg L^{-1}), $(NH_4)_2SO_4$ (2 g L^{-1}), $Na_2HPO_4{\cdot}2H_2O$ (6 g L^{-1}), KH_2PO_4 (3 g L^{-1}), NaCl (3 g L^{-1}), glucose (25 ml L^{-1}), and the trace metals solution (100 ml L^{-1}). The trace metals solution consists of $CaSO_4{\cdot}2H_2O$ (2 mg L^{-1}), $FeSO_4{\cdot}7H_2O$ (2 mg L^{-1}), $MnSO_4{\cdot}H_2O$ (0.2 mg L^{-1}), $CuSO_4{\cdot}5H_2O$ (0.2 mg L^{-1}), $ZnSO_4{\cdot}7H_2O$ (0.2 mg L^{-1}), $CoSO_4{\cdot}7H_2O$ (0.1 mg L^{-1}), $NaMoO_4{\cdot}H_2O$ (0.1 mg L^{-1}), and H_3BO_3 (0.05 mg L^{-1}).

Flow chamber biofilm experiments. Biofilms were grown in microfluidic flow chambers made from polydimethylsiloxane bonded to glass coverslips using an oxygen plasma, with four to eight identical flow channels on a single coverslip. All flow rates were controlled using a syringe pump (PicoPlus, Harvard Apparatus). The microfluidic channels were 500 µm wide and 7 mm long. For *V. cholerae*, *E. coli*, and *S. enterica*, channels with height 100 µm were used, whereas for *P. aeruginosa*, channels with height 300 µm were used.

For *V. cholerae* biofilm growth, overnight cultures grown in liquid LB-Miller with gentamicin (30 µg mL^{-1}) at 28 °C were diluted 1:200 into fresh LB-Miller with gentamicin and grown for 2 h. Then, these cultures were adjusted to an optical density at 600 nm (OD_{600}) of 0.001 and used to inoculate a microfluidic channel. The cells were given 1 h at room temperature to attach to the glass surface without flow, before fresh M9 medium with gentamicin was flown through the channel at a rate of 50 µL min^{-1} for 45 s, to wash away the non-attached cells. Then, the flow rate was set to 0.5 µL min^{-1} for the remainder of the experiment, and the flow channel as incubated at 25 °C.

For *E. coli* biofilm growth, overnight cultures were grown in liquid LB-Miller with kanamycin (50 µg mL^{-1}) at 37 °C. These cultures were diluted 1:2000 into tryptone broth and used to inoculate a microfluidic flow chamber. The cells were given 1 h to attach to the substrate without flow, before washing away non-adherent cells using tryptone broth with kanamycin at a flow rate of 50 µL min^{-1} for 45 s. Then, the flow rate was set to 0.1 µL min^{-1} for the remainder of the experiment, and the flow channel was incubated at 25 °C.

For *S. enterica* biofilm growth, overnight cultures were grown at 37 °C in liquid LB-Miller without salt, supplemented with spectinomycin (100 µg mL^{-1}). The overnight cultures were diluted 1:2000 and used to inoculate a flow channel. After giving the cells 1 h to attach to the coverslip without flow, the non-attached cells were washed away with tryptone broth supplemented with spectinomycin for 45 s using a flow rate of 50 µL min^{-1}. The flow rate was then set to 0.1 µL min^{-1} for the remainder of the experiment, and the flow channel was incubated at 25 °C.

P. aeruginosa strains were grown overnight in 5 ml liquid LB-Miller with 30 µg mL^{-1} gentamicin at 37 °C with shaking. The overnight culture was back-diluted 1:200 in 3 mL LB-Miller and grown until $OD_{600} = 0.5$. This culture was subsequently diluted 1:1000 in FAB medium and used to inoculate microfluidic flow chambers. After allowing cells to attach to the glass coverslip for 1 h at 30 °C without flow, the cells were washed for 50 s using a flow rate of 200 µL min^{-1}. The flow rate was then set to 3 µL min^{-1} for the remainder of the experiment, and the flow channel was incubated at 30 °C.

Image acquisition. Biofilms were imaged using a electron-multiplying charge-coupled device camera (EM-CCD, iXon, Andor) and a Yokogawa confocal spinning disk unit mounted on a Nikon Ti-E inverted microscope, and an Olympus 100× silicone oil (refractive index = 1.406) objective with a 1.35 numerical aperture. The fluorescent protein sfGFP was excited using a 488 nm laser. Three-dimensional images were acquired during biofilm growth every 60 min, using a z-spacing of 400 nm. The hardware was controlled using Matlab (Mathworks). A live feedback between image acquisition, image analysis, and microscope control was used to automatically detect the biofilm and expand the imaging field during growth in 3D, as described by Hartmann *et al.* [10], to minimize the laser exposure of the growing biofilm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGEMENTS

We are grateful to Lucia Vidakovic for generating several strains that were used in this study, and to Urs Jenal and Benoit-Joseph Laventie for *P. aeruginosa* strains. This research was supported by grants from the Studienstiftung des deutschen Volkes and Joachim Herz Stiftung (to H.J.), a Mathworks Fellowship (to D.J.S.), the MIT Mathematics Robert E. Collins Distinguished Scholar Fund (J.D.), the European Research Council (StG-716734), Deutsche Forschungsgemeinschaft (DR 982/5-1), Minna James Heineman Foundation, Bundesministerium für Bildung und Forschung (TARGET-Biofilms), Max Planck Society, and the Swiss National Science Foundation NCCR "AntiResist" (to K.D.).

[1] H.-C. Flemming, J. Wingender, U. Szewzyk, P. Steinberg, S. A. Rice, and S. Kjelleberg, Nat. Rev. Microbiol. **14**, 563 (2016).

[2] L. Hall-Stoodley, J. W. Costerton, and P. Stoodley, Nat. Rev. Microbiol. **2**, 95 (2004).

[3] H. C. Flemming and S. Wuertz, Nat. Rev. Microbiol. **17**, 247 (2019).

[4] P. S. Stewart and M. J. Franklin, Nat. Rev. Microbiol. **6**, 199 (2008).

[5] C. Fux, J. Costerton, P. Stewart, and P. Stoodley, Trends Microbiol. **13**, 34 (2005).

[6] T. J. Battin, K. Besemer, M. M. Bengtsson, A. M. Romani, and A. I. Packmann, Nat. Rev. Microbiol. **14**, 251 (2016).

[7] J. K. Teschler, D. Zamorano-Sánchez, A. S. Utada, C. J. A. Warner, G. C. L. Wong, R. G. Linington, and F. H. Yildiz, Nat. Rev. Microbiol. **13**, 255 (2015).

[8] K. Drescher, J. Dunkel, C. D. Nadell, S. Van Teeffelen, I. Grnja, N. S. Wingreen, H. A. Stone, and B. L. Bassler, Proc. Natl Acad. Sci. U.S.A. **113**, E2066 (2016).

[9] J. Yan, A. G. Sharo, H. A. Stone, N. S. Wingreen, and B. L. Bassler, Proc. Natl Acad. Sci. U.S.A. **113**, E5337 LP (2016).

[10] R. Hartmann, P. K. Singh, P. Pearce, R. Mok, B. Song, F. Díaz-Pascual, J. Dunkel, and K. Drescher, Nat. Phys. **15**, 251 (2019).

[11] B. Qin, C. Fei, A. A. Bridges, A. A. Mashruwala, H. A. Stone, N. S. Wingreen, and B. L. Bassler, Science **369**, 71 (2020).

[12] A. Persat, C. D. Nadell, M. K. Kim, F. Ingremeau, A. Siryaporn, K. Drescher, N. S. Wingreen, B. L. Bassler, Z. Gitai, and H. A. Stone, Cell **161**, 988 (2015).

[13] P. Ghosh, J. Mondal, E. Ben-Jacob, and H. Levine, Proc. Natl Acad. Sci. U.S.A. **112**, E2166 (2015).

[14] F. Beroz, J. Yan, Y. Meir, B. Sabass, H. A. Stone, B. L. Bassler, and N. S. Wingreen, Nat. Phys. **14**, 954 (2018).

[15] B. Maier, Ann. Rev. Biophys. **50**, 401 (2021), pMID: 33637007.

[16] E. R. Oldewurtel, N. Kouzel, L. Dewenter, K. Henseler, and B. Maier, eLife **4** (2015), 10.7554/elife.10811.

[17] J. Taktikos, Y. T. Lin, H. Stark, N. Biais, and V. Zaburdaev, PLOS ONE **10**, 1 (2015).

[18] G. C. L. Wong, J. D. Antani, P. P. Lele, J. Chen, B. Nan, M. J. Kühn, A. Persat, J.-L. Bru, N. M. Høyland-Kroghsbo, *et al.*, Physical Biol. **18**, 051501 (2021).

[19] C. R. Armbruster, C. K. Lee, J. Parker-Gilham, J. de Anda, A. Xia, K. Zhao, K. Murakami, B. S. Tseng, L. R. Hoffman, F. Jin, *et al.*, eLife **8** (2019), 10.7554/elife.45084.

[20] C. K. Lee, J. Vachier, J. de Anda, K. Zhao, A. E. Baker, R. R. Bennett, C. R. Armbruster, K. A. Lewis, R. L. Tarnopol, *et al.*, mBio **11**, e02644 (2020).

[21] C. K. Lee, J. de Anda, A. E. Baker, R. R. Bennett, Y. Luo, E. Y. Lee, J. A. Keefe, J. S. Helali, J. Ma, K. Zhao, *et al.*, Proc. Natl Acad. Sci. U.S.A. **115**, 4471 (2018).

[22] B. R. Wucher, T. M. Bartlett, M. Hoyos, K. Papenfort, A. Persat, and C. D. Nadell, Proc. Natl Acad. Sci. U.S.A. **116**, 14216 (2019).

[23] B.-J. Laventie, M. Sangermani, F. Estermann, P. Manfredi, R. Planes, I. Hug, T. Jaeger, E. Meunier, P. Broz, and U. Jenal, Cell Host & Microbe **25**, 140 (2019).

[24] P. Pearce, B. Song, D. J. Skinner, R. Mok, R. Hartmann, P. K. Singh, H. Jeckel, J. S. Oishi, K. Drescher, and J. Dunkel, Phys. Rev. Lett. **123**, 258101 (2019).

[25] C. A. Rodesney, B. Roman, N. Dhamani, B. J. Cooley, P. Katira, A. Touhami, and V. D. Gordon, Proc. Natl Acad. Sci. U.S.A. **114**, 5906 (2017).

[26] M. K. Kim, F. Ingremeau, A. Zhao, B. L. Bassler, and H. A. Stone, Nat. Microbiol. **1**, 15005 (2016).

[27] K. N. Kovach, D. Fleming, M. J. Wells, K. P. Rumbaugh, and V. D. Gordon, *Langmuir*, Langmuir **36**, 1585 (2020).

[28] A. Cont, T. Rossy, Z. Al-Mayyah, and A. Persat, eLife **9** (2020), 10.7554/elife.56533.

[29] Q. Zhang, J. Li, J. Nijjer, H. Lu, M. Kothari, R. Alert, T. Cohen, and J. Yan, Proc. Natl Acad. Sci. U.S.A. **118** (2021), 10.1073/pnas.2107107118.

[30] A. Dragoš and A. T. Kovács, Trends Microbiol. **25**, 257 (2017).

[31] L. Hobley, C. Harkins, C. E. MacPhee, and N. R. Stanley-Wall, FEMS Microbiol. Rev. **39**, 649 (2015).

[32] R. Hartmann, H. Jeckel, E. Jelli, P. K. Singh, S. Vaidya, M. Bayer, D. K. H. Rode, L. Vidakovic, F. Díaz-Pascual, J. C. N. Fong, *et al.*, Nat. Microbiol. **6**, 151 (2021).

[33] D. J. Skinner, B. Song, H. Jeckel, E. Jelli, K. Drescher, and J. Dunkel, Phys. Rev. Lett. **126**, 048101 (2021).

[34] E. A. Lazar, J. K. Mason, R. D. MacPherson, and D. J. Srolovitz, Phys. Rev. Lett. **109**, 095505 (2012).

[35] P.-G. De Gennes and J. Prost, *The physics of liquid crystals*, Vol. 83 (Oxford university press, 1993).

[36] S. C. McGrother, D. C. Williamson, and G. Jackson, J. Chem. Phys. **104**, 6755 (1996).

[37] G. Makey, S. Galioglu, R. Ghaffari, E. D. Engin, G. Yıldırım, Ö. Yavuz, O. Bektaş, Ü. S. Nizam, Ö. Akbulut, *et al.*, Nat. Phys. **16**, 795 (2020).

[38] K. A. Takeuchi, M. Sano, T. Sasamoto, and H. Spohn, Sci. Rep. **1**, 34 (2011).

[39] C. A. Tracy and H. Widom, Commun. Math. Phys. **290**, 129 (2009).

[40] R. D. Monds, T. K. Lee, A. Colavin, T. Ursell, S. Quan, T. F. Cooper, and K. C. Huang, Cell Rep. **9**, 1528 (2014).

[41] A. S. Utada, R. R. Bennett, J. C. Fong, M. L. Gibiansky, F. H. Yildiz, R. Golestanian, and G. C. Wong, Nat. Commun. **5**, 4913 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.06.455416; this version posted August 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[42] L. Craig, K. T. Forest, and B. Maier, Nat. Rev. Microbiol. **17**, 429 (2019).

[43] T. Munk, F. Höfling, E. Frey, and T. Franosch, Euro. Phys. Lett. **85**, 30003 (2009).

[44] F. Höfling, E. Frey, and T. Franosch, Phys. Rev. Lett. **101**, 120605 (2008).

[45] Y. Han, A. Alsayed, M. Nobili, and A. G. Yodh, Phys. Rev. E **80**, 011403 (2009).

[46] E. A. Lazar, J. Han, and D. J. Srolovitz, Proc. Natl Acad. Sci. U.S.A. **112**, E5769 (2015).

[47] F. Aurenhammer, R. Klein, and D.-T. Lee, *Voronoi Diagrams and Delaunay Triangulations*, 1st ed. (World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2013).

[48] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher, ACM Trans. Graph. **33** (2014), 10.1145/2601097.2601175.

[49] I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, 2nd ed., Springer Series in Statistics (Springer, 2005) Chap. 12, pp. 201–268.

[50] T. C. Day, S. S. Höhn, S. A. Zamani-Dahaj, D. Yanni, A. Burnetti, J. Pentz, A. R. Honerkamp-Smith, H. Wioland, Sleath, *et al.*, bioRxiv (2021), 10.1101/2021.07.29.454238.

[51] J. F. Dekkers, M. Alieva, L. M. Wellens, H. C. R. Ariese, P. R. Jamieson, A. M. Vonk, G. D. Amatngalim, H. Hu, K. C. Oost, Snippert, *et al.*, Nat. Protocols **14**, 1756 (2019).

[52] R. Etournay, M. Popović, M. Merkel, A. Nandi, C. Blasse, B. Aigouy, H. Brandl, G. Myers, G. Salbreux, F. Jülicher, and S. Eaton, eLife **4**, e07090 (2015).

[53] K. L. Meibom, X. B. Li, A. T. Nielsen, C.-Y. Wu, S. Roseman, and G. K. Schoolnik, Proc. Natl Acad. Sci. U.S.A. **101**, 2524 (2004).

[54] D. O. Serra, A. M. Richter, G. Klauck, F. Mika, and R. Hengge, mBio **4** (2013), 10.1128/mBio.00103-13.

[55] N. Grantcharova, V. Peters, C. Monteiro, K. Zakikhany, and U. Römling, J. Bacteriol. **192**, 456 (2010).

[56] U. Römling, Z. Bian, M. Hammar, W. D. Sierralta, and S. Normark, J. Bacteriol. **180**, 722 (1998).

[57] U. Römling, W. D. Sierralta, K. Eriksson, and S. Normark, Mol. Microbiol. **28**, 249 (1998).

[58] J. G. Malone, T. Jaeger, C. Spangler, D. Ritz, A. Spang, C. Arrieumerlou, V. Kaever, R. Landmann, and U. Jenal, PLoS Pathog **6**, e1000804 (2010).

[59] J. Thibault, E. Faudry, C. Ebel, I. Attree, and S. Elsen, J. Biol. Chem. **284**, 15762 (2009).

[60] V. De Lorenzo and K. N. Timmis, Meth. Enzymol. **235**, 386 (1994).

[61] K. Skorupski and R. K. Taylor, Gene **169**, 47 (1996).