

Supporting Information

Lee et al. 10.1073/pnas.1720071115

SI Materials and Methods

Strains and Growth Conditions. *P. aeruginosa* PA14 WT and its isogenic strains carrying mutations in the $\Delta pilA$ (1), $\Delta pilT$, $\Delta pilJ$ (2), and $\Delta pilU$ genes were used in this study. For cAMP reporter experiments, WT and $\Delta pilU$ with the YFP/CFP reporter/control plasmid were used (3). Multigenerational tracking is at present not technologically feasible using a FRET sensor or a chromosomally tagged reporter. Details about strains and their construction are in *SI Materials and Methods, Construction of $\Delta pilT$ and $\Delta pilU$ Mutants and Construction of cAMP Reporter Strains*. Bacteria were plated onto LB agar plates and incubated at 37 °C overnight. The plates were then sealed with Parafilm and stored in a refrigerator at 4 °C. Individual colonies were swabbed from the plate and grown overnight to an $OD_{600\text{ nm}} \sim 1.5$ (~20 h) in an incubator at 37 °C shaking at 220 rpm. Overnight growth medium contained M63, which contained (per liter of deionized water) 3 g potassium phosphate monobasic, 7 g potassium phosphate dibasic, and 2 g ammonium sulfate, supplemented with 1 mM $MgSO_4$, 0.2% glucose, and 0.5% CAA (Sigma-Aldrich). For mutants with gentamicin-resistance or carbenicillin-resistance plasmids, LB agar plates and overnight media were also supplemented with 20 and 25 $\mu\text{g/mL}$ gentamicin or 300 and 250 $\mu\text{g/mL}$ carbenicillin (Sigma-Aldrich), respectively. The overnight culture was regrown under the same conditions as the overnight growth (without antibiotics) to an $OD_{600\text{ nm}} \sim 0.4$ at dilutions of 1:50, 1:100, and 1:200. Cultures were then diluted to an $OD_{600\text{ nm}} \sim 0.01$ (~0.1 for initial cAMP reporter intensity imaging) in flow-cell medium, which consisted of M63 supplemented with 1 mM $MgSO_4$, 0.05% glucose, and 0.125% CAA. The diluted culture was used for injection into the flow chamber.

Single-Channel Flow-Cell Experiments. The flow cell was purchased from the Department of Systems Biology, Technical University of Denmark, and prepared as previously described (4, 5). The prepared flow cell was connected to a syringe through a 0.22- μm filter (Fisher Scientific) using Silastic silicon tubing of inner diameter 1.57 mm and outer diameter 3.18 mm (Dow Corning) and a natural Kynar PVDF female Luer to 1.6-mm barb adapter (Value Plastics DBA Nordson Medical). The assembled system was flushed with 3% H_2O_2 at a volumetric flow rate of 20 mL/h using a syringe pump (KD Scientific or Harvard Apparatus) and allowed to sit for a total of 4 h including flushing time. The sterilized system was then flushed with autoclaved, deionized water at a flow rate of 5 mL/h using a syringe pump and allowed to sit overnight. Before inoculation of the bacteria into the flow cell, the flow-cell system was flushed with flow-cell medium at 30 mL/h. The diluted bacteria culture was injected into the flow cell and allowed to incubate for 10 min on the heating stage at 30 °C. Then, the flow cell was flushed with flow-cell medium at 30 mL/h to wash away cells that had not adhered to the surface. For these experiments, data were collected using a flow rate of 3 mL/h and total flow time of 10~40 h.

In all flow-cell experiments, no new cells are introduced into the system after the initial inoculation. However, not all cells that do not attach to the surface are flushed out of the system by flow. Despite the flow, some cells will remain in the liquid throughout the experiment. Cells that “replenish” the surface by attaching to it come from the liquid in the flow cell. These cells from the liquid are either cells that have never attached to the surface (surface-naive) or are cells that have previously attached to and then detached from the surface (surface-sentient).

Tandem-Channel Flow-Cell Experiments. To optimize for flow-cell design and data acquisition between flow-cell channels, the channels were joined in series by connecting the outlet tubing of one channel (FC1) and the inlet tubing of a second channel (FC2) using two T junctions and additional tubing cut as short as possible. The T junctions and tubing were introduced as close as possible to the inlet/outlet to minimize nonchannel volume and with enough spacing to allow for clamping of the tubing. Each channel of the combined flow cell was then sterilized as described for a single channel experiment. Before inoculation, FC1 was filled with flow-cell medium, FC2 had its liquid drained out, and the channel-to-channel connection tubing was clamped next to each T junction to prevent bacteria growth and contamination in FC2 while imaging FC1. FC1 was then inoculated and imaged following the single channel protocol. After the bacteria reached the required surface cell density at time T from initial inoculation, the harvesting procedure was performed. The outlet tubing of FC1 was flushed, FC2 was filled with flow-cell medium, the connection tubing was opened, and bacteria in FC1 were harvested into FC2 via the connection tubing and flow at 10 mL/h. Cells that were harvested into the liquid of FC2 came from the liquid of FC1. However, during the harvesting stage, some cells on the surface of FC1 could detach and join the liquid subpopulation of cells and subsequently get harvested into FC2. After a harvesting period of 10 min, the connection tubing was once again clamped to seal off FC1 and the connection tubing. FC2 was then imaged following the single channel protocol. This protocol can be extended to m number of flow-cell channels in series, treating channel $m - 1$ as FC1 and channel m as FC2.

Memory-Loss Flow-Cell Experiments. For these experiments, we captured cells exiting from FC1, simulated nonsurface conditions using the planktonic culturing conditions used for overnight growth before inoculation into FC1, then assessed in a new FC2 the surface colonization behavior of these bacteria transitioned back to planktonic growth. Experiments here were performed as described for a single-channel experiment for a period of $T = 30\sim 40$ h, after which the recording was ended and the flow cell was dismantled from the microscope. Approximately one milliliter of flow-cell liquid was extracted, 250 μL was resuspended in 1.75 mL of overnight medium, and then this culture was incubated with previous overnight growth conditions. This overnight culture was grown and sampled at various time points t_r . Each sample was prepared for flow-cell imaging experiments in a fresh flow-cell channel as described for a single-channel experiment (including regrowth, dilution, inoculation, and imaging).

Data Acquisition. Images were taken using an EMCCD camera (Andor iXon) with IQ software (Andor) on an Olympus IX81 microscope, which was equipped with a Zero Drift Correction autofocus system. Bright-field images were taken every 3 s (30 ms exposure time). For cAMP reporter experiments, fluorescence images were also taken every 15 min (100-ms or 150-ms exposure time) using a Lambda LS (Sutter Instrument) xenon arc lamp; this time interval was optimized for longitudinal studies lasting ~2 d. cAMP reporter experiments used YFP and either DAPI or CFP filters. Acquisition for a single flow-cell channel continued for a total recording time of about 10~40 h, which resulted in 12,000~48,000 bright-field images and 40~160 fluorescence images. The image size was $67 \times 67 \mu\text{m}^2$

(1,024 × 1,024 pixels). For imaging cAMP reporter intensities immediately after inoculation of FC1, a slightly modified imaging protocol was used where fluorescence images were taken every 30 s for 1–5 h (bright-field images were still taken every 3 s).

cAMP-Dependent lacP1-lacZ Reporter Activity Assays. The cAMP-dependent *lacP1-lacZ* reporter plasmid and the corresponding control plasmid were provided by M. C. Wolfgang, University of North Carolina at Chapel Hill, Chapel Hill, NC and described previously (6). PA14 strains were transformed with pR-P1-lacZ (cAMP-dependent reporter) or pR-lacZ (vector control) via electroporation. Transformants were selected on gentamicin. To perform the reporter activity assay, cultures were grown overnight in LB with 20 µg/mL gentamicin, diluted 1:100 in M8 broth supplemented with 1 mM MgSO₄, 0.2% glucose, and 0.5% CAA, and grown for 3 h at 37 °C with shaking. Cultures were then spread onto M8 medium-based 1% agar plates. After 5 h, cells were harvested from plates in 1× PBS. β-Galactosidase assays were performed as previously described (7). Significance was determined by one-way ANOVA followed by Dunnett's posttest comparison for differences relative to WT.

Multigenerational Cell-Tracking and Image Analysis. The image analysis algorithms and software are adapted from methods previously described (5, 8–12) and written in MATLAB R2015a (MathWorks). First, all raw bright-field images were processed using a sequence of background correction, intensity normalization, Gaussian and edge filters, and Otsu thresholding to generate binary images, which contained main features of bacteria. Then, a tracking routine (with a sequence of algorithms using maximum bacteria feature overlap and minimum displacement between consecutive frames and history-based segmentation of neighboring cells) was used to link cells in different frames in time. Cells that are related to one another by division are likewise incorporated in a “tree-like” data structure so that it is possible to generate a bacterial “family tree,” where each node in the tree is a single bacterium identity. To minimize errors generated in both recognizing and tracking cells, we manually validated each tracked family (minimum of three families per strain) to make sure that the final assignment of bacterial lineage in the family tree was correct [*SI Materials and Methods, Manual Data Validation Using a Custom Graphical User Interface (GUI)*]. Validated families were chosen by randomly sampling bacteria in an image of the dataset, tracking the families in which they appear, and repeating this for multiple images. An example video showing the final validated tracking data are shown in [Movie S1](#). A typical dataset contains 18,000–24,000 frames and up to 1 million bacteria images.

After the error correction stage, fluorescence information was integrated into the family tree. For cAMP reporter experiments, reporter intensity values were extracted from fluorescence images using the following sequence of algorithms. For each fluorescence image (reporter and control), we used the corresponding segmented bright-field image as a mask to extract the intensity values. In each of the fluorescence images that a given identity was present in, the mean intensity of all of the pixels comprising the bacterium was calculated for both fluorescence images (reporter and control) and then divided, such that the normalized reporter intensity was reported as a fold change of the control. To ensure that changes in the normalized reporter intensities did not result from control intensity fluctuations, we repeated the above algorithm, except we divided the reporter intensities by the mean background intensity level extracted using the inverse mask of the segmented bright-field image.

The final bacterial family trees were a database which contained a complete history for each cell, including motility, cell division, lineage, and normalized cAMP reporter intensities I_{cAMP} (if the strain had the cAMP reporter plasmid). These data were plotted using a family tree plot, where the lengths of vertical lines on the plots were proportional to time spent in each generation. Vertical lines that ended with downward arrows were detachment events, lines that intersected with a horizontal line were division events, and lines that ended without a marker were out-of-bound events where we lost track of the bacterium (moving out of the field of view or reaching the end of the recording, represented as moving outside the XYT limits of the dataset boundaries). Horizontal lines were arbitrarily spaced to show all of the descendants. White, outlined, upside-down triangles indicated potential division events that were missed because the bacterium was completely vertical while dividing, with the distal daughter cell out of the focal plane. Metrics were summarized as a single value per identity and then plotted as the colors of the vertical lines in the family tree plots. To quantitatively compare family trees, we used the proportion of “*n*-legged” division branching per family, where *n* = 2, 1, or 0 is the number of nondetached daughter cells postdivision, and the tree asymmetry value, where 0 is a perfectly symmetric tree and 1 is a perfectly asymmetric tree (*SI Materials and Methods, Metrics for Quantitative Comparisons of Family Trees*). MATLAB functions from the base installation of MATLAB R2015a, Statistics and Machine Learning Toolbox, Curve Fitting Toolbox, Image Processing Toolbox, Signal Processing Toolbox, and custom MATLAB functions were used for all analyses.

Manual Data Validation Using a Custom Graphical User Interface (GUI). All segmentation and cell-tracking algorithms have compromises that can lead to potential tracking errors that can propagate in the family trees. To ensure accuracy of our family tree tracking results, we validate all of our tracked results manually using a custom-built GUI that propagates corrections. Common corrections, using a custom GUI, were grouped into four categories: swapped identities, loss of identity, false division events, and inaccurately segmented bacterial regions.

First, switching of identities between two bacteria occurred when two overlapping bacteria regions quickly changed positions such that the tracking algorithm was unable to detect this event due to the image acquisition rate of one image every 3 s. To correct this error, the identities of both bacteria after the error were assigned to new families to avoid having multiple cells with the same identity. Then, the region of one of the bacteria before the error was selected and assigned to the proper region following the error. The same was done for the other bacterium.

Second, loss of identity of individual bacteria occurred when the center of mass of a bacterium moved in and out of the field of view or in and out of focus. To correct this error, the region of the bacterium before the error was selected and assigned to the first frame after the error in which either the bacterium's center of mass could be detected. To ensure that the correct bacterium was being tracked, the bright-field images were used to visualize the boundaries of the bacteria.

Third, false division events occurred due to several reasons. Standing up of the bacterium during a division event prevented the tracking algorithm from properly detecting the division due to the division's occurring out of the focal plane. To correct this error, the first frame in which the two daughter cells could be distinguished as two distinct regions was assigned as the frame at which the division took place. A second cause of false division events involved the moving of one cell over another cell, resulting in the detection of two regions in the bottom cell, which the tracking algorithm incorrectly assigned as a division event. To correct this error, the region of the falsely divided cell before the false division was selected. This region was then assigned to the identities of the two daughter cells for all

frames in which the original cell was present as two daughter cells until the two original cells were not overlapping, at which point the tracking algorithm combined the two daughter cells into one cell.

Fourth, inaccurate segmentation of a bacterium resulted in an incorrect region encompassing the bacterium. This error usually arose in the frames following one of the previous errors. To correct this error, the region of the bacterium in the first frame in which the error occurred was selected and a polygon was drawn around the incorrectly segmented bacterium to correctly define the region encompassing the bacterium. This process was repeated for every subsequent frame in which the same error occurred.

Metrics for Quantitative Comparisons of Family Trees. To quantitatively compare family trees, we used several metrics. Built-in MATLAB statistical test functions from the Statistics and Machine Learning Toolbox for one-way ANOVA, Kruskal–Wallis test, and χ^2 test and multiple comparison tests using Tukey’s honest significant difference criterion (with a P value less than 0.05 as the significance value) and the Benjamini and Yekutieli (13) procedure for controlling the false discovery rate of a family of hypothesis tests were used to compare strains with these metrics.

One metric is the proportion of “ n -legged” division branching per family, where $n = 2, 1$, or 0 is the number of nondetached daughter cells postdivision. Division events that result in any daughter cell’s moving out of the field of view are excluded because it is not known whether they detach or not. To avoid effects due to each family tree being different in size and number of bacteria, we aggregate the first six generations of three representative family trees for each strain. Another method of calculating and comparing this metric is to count the proportion of division-branching events observed for all families of a given strain and then to compare these proportions using the χ^2 test.

Another metric is tree asymmetry, adapted from previous work that defined and used this metric for studying neuronal branching patterns (14). To define tree asymmetry, several terms and parameters regarding our family trees must first be defined. Family trees are plotted on a vertical time axis starting from the top, and the horizontal axis is arbitrary. Each vertical line (parallel to the time axis), or branch, in each family tree represents a single bacterium, and each horizontal line (perpendicular to the time axis) represents a single division event. There are n terminal branches, which represent bacteria where we do not observe a division on the surface and are represented in the tree plot as vertical lines that do not end by intersecting a horizontal line. There are $n - 1$ subtrees, which represent bacteria that divide and all their descendants (not just immediate ones) and are represented in the tree plot as vertical lines that end by intersecting a horizontal line. The topmost branch in any subtree (the entire tree is also one) is called the founder branch. The number of subtrees is also equal to the number of division events (horizontal lines). The tree asymmetry calculation involves several steps and operates on all possible subtrees. The first step is to calculate branch asymmetry for the founder branch of every subtree:

$$\text{Branch Asymmetry: } A_b(L, R) = \frac{|L - R|}{L + R - 2}; A_b(1, 1) = 0.$$

The number of terminal branches to the left and right of the founder branch are L and R , respectively. The second step is to calculate subtree asymmetry using all of the branch asymmetries that are in the subtree:

$$\text{Subtree Asymmetry: } A_s = \frac{1}{n_s - 1} \sum_{b=1}^{n_s-1} A_b(L, R).$$

The total number of terminal branches in a subtree is n_s . The tree asymmetry value λ is then the subtree asymmetry value corresponding to the entire tree. To account for our experimental boundary conditions (finite temporal measurement period and finite spatial frame size), we include the concept of “true” and “pseudo” terminal branches. Cells that detach from the surface are considered true terminal branches and use the previously defined definition of terminal branch. Cells that extend outside the spatial or temporal boundaries of a dataset (moving out of the field of view or reaching the end of the recording) are considered pseudo terminal branches because there is an uncertainty in the branching pattern. We account for this uncertainty by making the least intrusive assumption that these branches continue to stay on the surface and divide for the next generation, and then count these daughter cells as true terminal branches.

To look at values of λ for early generations, we introduce a notation λ_G , where G is the number of generations used to calculate the parameter. If G is less than the total number of generations for a family, then that family is essentially truncated to only include generations $1 - G$ (the first G generations). In the main text, λ_6 was calculated using the exact same family data used to calculate the metric in Fig. 3C.

In principle, λ is scale invariant, since it averages across all subtrees and ranges between 0 (perfectly symmetric) and 1 (perfectly asymmetric). In practice, however, for finite-size and finite-time trees, λ can saturate at values that are not the theoretical limits of 0 and 1. In addition, in our datasets, there can be a big spread in λ even for perfectly symmetric and asymmetric trees. This is due to the spread in division times, which causes desyncing of later generations, combined with the temporal truncation of the tree due to the experimental boundary conditions. They cause different subtrees in a family tree to have different numbers of generations, thus a spread in the calculated λ , since the family tree size (number of total generations) can influence the actual range of λ .

We introduced a correction to λ to account for these complications. The λ values were first plotted against the number of total tracked generations. Family tree simulations for perfectly symmetric and asymmetric trees were also plotted on this diagram. The range of values for each simulation was then obtained by fitting a boundary shape around all of the points using the function “boundary.m” in MATLAB R2015a. Due to complications in simulating large symmetric trees, the upper boundary of the simulated symmetric tree values for five or more generations were set to the value obtained for five generations. The lower boundary of the simulated asymmetric tree values and the upper boundary of the simulated symmetric tree values formed the experimental upper and lower limits, respectively, of λ in a generation-dependent manner. Values of λ obtained from experimental data were then rescaled to be a fraction of this experimental range, and values above this range were set to 1. As a result, this corrected value, called “generation-corrected tree asymmetry” (λ_{GC}), still had the same limits as the original λ . Note that the experimental range overlapped for three or fewer generations, so the corrected value requires a minimum of four generations. The original values for the experimental data are shown in Fig. 3C, while the corrected tree asymmetry vs. number of generations plot is shown in Fig. S5.

Family Tree Simulations. Family trees were simulated using experimental input data, which were the probabilities of staying and dividing or detaching, the division time distribution, the distributions of founder cell and detached cell lifetimes, and the temporal boundaries of the dataset. After every division, daughter cells A and B had probabilities P_A and P_B of staying and dividing and probabilities $1 - P_A$ and $1 - P_B$ of detaching, respectively. These probabilities were held constant when simulating a single family to represent a “memoryless” state where the probabilities are time-independent. Division times were randomly generated from a Gaussian distribution fitted to the experimental data. The lifetimes of the founder cell and detached cells were randomly generated from a uniform distribution between 1 and the randomly generated division time, since we observed that these cells could land on and detach from the surface during any part of their growth cycle. To simulate the temporal boundaries of the dataset, total tracking times were randomly generated from a uniform distribution between the range of actual number of tracked frames in the experimental data. Any bacteria that had not divided or detached by the end of the simulation were marked the same as experimental tracking data. Perfectly symmetric family trees were simulated by setting P_A and P_B to 1, and perfectly asymmetric trees were simulated by setting P_A to 1 and P_B to 0. For perfectly symmetric family trees, the number of total divisions simulated increased exponentially as the number of generations increased, so the total simulation time and number of iterations were much smaller than that for perfectly asymmetric trees.

TFP Activity Metric. TFP-driven motility is known as “twitching” motility because of its characteristic motion: directional movement punctuated by sudden jittery changes of direction. Based on our previous work on quantitatively characterizing twitching activity in WT and flagellum deletion mutants, we designed a multimetric algorithm to recognize TFP activity automatically. These previous studies include work on the “motility signature” of TFP-driven motion (10) and flagellum-driven motion (9, 15, 16) for *P. aeruginosa* (WT and flagellum knockout mutants).

High-speed microscopy shows that TFP-driven twitching motility in *P. aeruginosa* consists mostly of translations for variable temporal durations, alternating with combined translation-rotation “pulses” that are much shorter in temporal duration (10, 17). This behavior contrasts with flagellum-generated motility based on “spinning” (9, 15, 16). Based on the temporal predominance of surface translations along the cell body axis in TFP-driven motion, combined with the fact that these translations generally produce trajectories that have superdiffusive mean squared displacements (MSDs), the multiparameter metric for TFP activity is defined as follows and summarized visually in Fig. S6. A bacterium has TFP activity during a given time point when it is “not spinning” and has nonzero displacement over this w frame moving window every $w/10$ frames. A value of 100 was chosen for w (values for w ranging from 10 to 200 were tested), which corresponds to a 5-min moving window every 30 s. A cell that is “not spinning” is defined as having the following characteristics during the w frame window: an MSD slope of greater than or equal to 0.9, having the maximum two-point distance of its trajectory (proportional to the radius of gyration) being greater than or equal to 50% of its maximum cell body length, and having the circularity of its visit map, defined as $4\pi(\text{Area})/(\text{Perimeter})$, less than 0.55. Displacement was defined using the visit map of the cell body [the cumulative trace of the body on the surface over time (5)] during the w frame window, finding the maximum two-point distance of the visit map, and then subtracting the maximum cell body length during this w frame window. The resulting metric output is a binary signal (presence or absence of TFP activity at each time point), which can be quantified by the proportion of events within a given time frame. Cells with high TFP activity have elongated visit maps and nonspinning behavior with nonzero displacement.

To examine the robustness of these biometric recognition algorithms for TFP activity, the analysis was repeated with an alternate definition of spinning and nonspinning behavior based on bacteria centroid and pole tracking instead of bacterial visit maps. This alternate definition was similar to the ones used in our previous studies (9, 10, 15). This coordinate algorithm compared the value of three parameters over time: the displacement of each pole and centroid between consecutive frames with a 21-frame moving window. Spinning and nonspinning behavior was classified based on whether the displacement of these three values was above or below an empirically set threshold, respectively. Both analyses converged on the same results, but we chose the visit map algorithm, instead of the coordinate algorithm (*SI Materials and Methods, Comparison of Visit Map vs. Coordinate Algorithms for TFP Activity Metrics*).

Comparison of Visit Map vs. Coordinate Algorithms for TFP Activity Metrics. Our previous studies on the “motility signature” of TFP-driven motion (10) and flagellum-driven motion (9, 15, 16) for *P. aeruginosa* on surfaces mainly utilized high-speed recording at sampling rates around 10~200 Hz for 5~45 min.

To perform long-term tracking (10~40 h), which is necessary for generation-dependent studies, lower sampling rates (<1 Hz) must be used because of hardware and software constraints related to data acquisition and analysis. For example, since each image takes up a set amount of storage space and the total space is finite, we must lower the sampling rate to image for a longer time without running out of space. Having a lower sampling rate introduces several complications in the tracking algorithms. One is the stroboscopic effect when detecting spinning behavior that occurs at frequencies higher than the sampling rate. The coordinate algorithm will fail to capture the spinning behavior, since the stroboscopic effect will produce a jumbled collection of points that can be analyzed only if the exact frequency of the spinning behavior is known. However, a spinning bacterium’s visit map will look circular (or a sector of a circle) if the w frame window is large enough (Fig. S6), unless it is spinning at an exact multiple of the sampling rate for the entire w frame window. Detecting and filtering out predominantly spinning behavior is important for characterizing TFP activity.

A lower sampling rate also results in a loss of data density, defined as all of the input spatial and temporal information that is entered into the tracking algorithm. Lower data densities are generally correlated with higher uncertainties in the tracking and are thus undesirable. The visit map algorithm retains more spatial information of a bacterium compared with the coordinate algorithm, which reduces a bacterium’s spatial information down to only its centroid and poles. Thus, using the visit map algorithm partially compensates for the loss of data density due to a lower sampling rate, resulting in a higher overall data density and thus lower uncertainties in the tracking compared with using the coordinate algorithm.

Family Correlation Analysis of cAMP Levels and TFP Activity. Input time series data used for cAMP levels were the normalized reporter intensity values I_{cAMP} extracted every 15 min; input data for TFP activity were the binary readouts (presence or absence of TFP activity) every 30 s using 5 min worth of data to generate a single readout. Representative input data for one lineage is shown in Fig. S7. The input data were first synchronized in time by applying the following algorithm on every tracked bacterium for a given family. For every cAMP readout interval (from just after the previous intensity readout to just before the current intensity readout), the proportion

of time points with presence of TFP activity was calculated. This step produced time-series data for each readout for every branch of a given family. To correlate the readouts for a family, time points with multiple readouts from different branches were averaged to obtain a single time series per readout. TFP activity levels of 0 were ignored in this average. Autocorrelations and cross-correlations were performed using the function “xcov.m” in MATLAB R2015a (which subtracts the mean from each time sequence and then calls the function “xcorr.m”).

cAMP–TFP Model. A minimal model is built based on the picture previously described (3, 18) with the following ingredients shown in Fig. 4A: PilA monomer $m(t)$, cAMP signal $s(t)$, and TFP activity $A(t)$. Dynamical equations relating these quantities are coupled and nonlinear, but they can be linearized around a fixed point to examine the fluctuations and response, following the spirit of the seminal work of Alan Turing to describe generic reaction–diffusion systems (19). Hence, for the concentrations, $\delta m(t)$ and $\delta s(t)$ are used. The processes described in Fig. 4A can be turned into the following linearized system of equations:

$$\delta \dot{s}(t) = -k_i \delta m(t) - k_s \delta s(t)$$

$$\delta \dot{m}(t) = -k_m \delta m(t) - k_t A(t)$$

$$\dot{A}(t) = g \delta m(t) + \sigma(t),$$

where $k_s, k_i, k_t, k_m,$ and g are rates with $k_i > 0, k_t > 0,$ and $g > 0$.

The term $-k_t A(t)$ describes the constraint that accompanies TFP activity in terms of the depletion of free pilin monomers when the pilus is elongating and reintroduction of pilin monomers when it retracts (represented by the double arrows). This activity is coupled to the sensing via a kinematic rule: When the TFP retracts, the resulting disassembled monomers will be released back to the periplasmic space and couple to the activity of the Pil-Chp complex; when it extends, the monomer pool in the periplasmic space will be depleted accordingly. The term $\sigma(t)$ is a noise term that is assumed to be the main driving force for fluctuations; this is assumed to be Gaussian white noise since the measurements for TFP activity have time scales much longer than the actual retraction/elongation process (minutes~hours vs. milliseconds~seconds) (20). We can relax this and build a telegraph process for the TFP activity drive $\sigma(t)$. The term $-k_i \delta m(t)$ represents the regulation of the activity of the Pil-Chp complex, which is controlled by its methylation state, by the free monomer concentration. The term $g \delta m(t)$ represents a positive feedback/regulation of TFP activity by the free monomers.

This system of equations can be solved, and correlation functions can be calculated to yield the autocorrelation of TFP activity $C_{AA}(t)$, autocorrelation of cAMP signal $C_{ss}(t)$, and cross-correlation between TFP activity and cAMP signal $C_{As}(t)$ (t here corresponds to time lag and not regular time):

$$C_{AA}(t) = \frac{1}{T} \int d\tau \langle A(\tau+t) A(\tau) \rangle$$

$$C_{AA}(t) = \begin{cases} \frac{D}{2k\omega_0} \left(\frac{\omega_0^2 + 5k^2}{\omega_0^2 + k^2} \right) [\omega_0 \cos(\omega_0 t) + k \sin(\omega_0 t)] e^{-kt}, & t > 0 \\ \frac{D}{2k\omega_0} \left(\frac{\omega_0^2 + 5k^2}{\omega_0^2 + k^2} \right) [\omega_0 \cos(\omega_0 t) - k \sin(\omega_0 t)] e^{kt}, & t < 0 \end{cases}$$

$$C_{ss}(t) = \frac{1}{T} \int d\tau \langle \delta s(\tau+t) \delta s(\tau) \rangle$$

$$C_{ss}(t) = \begin{cases} \frac{\omega_0 (k_s^2 + \omega_0^2 - 3k^2) \cos(\omega_0 t) + k (k_s^2 + 3\omega_0^2 - k^2) \sin(\omega_0 t)}{2k\omega_0 [\omega_0^2 + k^2] [(k_s^2 + \omega_0^2 - k^2)^2 + 4k^2\omega_0^2]} Dk_i^2 k_t^2 e^{-kt} \\ \quad + \frac{Dk_i^2 k_t^2 e^{-k_s t}}{k_s [(k_s^2 + \omega_0^2 - k^2)^2 + 4k^2\omega_0^2]}, & t > 0 \\ \frac{\omega_0 (k_s^2 + \omega_0^2 - 3k^2) \cos(\omega_0 t) - k (k_s^2 + 3\omega_0^2 - k^2) \sin(\omega_0 t)}{2k\omega_0 [\omega_0^2 + k^2] [(k_s^2 + \omega_0^2 - k^2)^2 + 4k^2\omega_0^2]} Dk_i^2 k_t^2 e^{kt} \\ \quad + \frac{Dk_i^2 k_t^2 e^{k_s t}}{k_s [(k_s^2 + \omega_0^2 - k^2)^2 + 4k^2\omega_0^2]}, & t < 0 \end{cases}$$

$$C_{As}(t) = \frac{1}{T} \int d\tau \langle A(\tau+t) \delta s(\tau) \rangle$$

$$\text{For } k_s < 0, C_{As}^-(t) = \begin{cases} \frac{\omega_0 [3k^2 + 2kk_s - \omega_0^2] \cos(\omega_0 t) + [k^2(k + k_s) - \omega_0^2(3k + k_s)] \sin(\omega_0 t)}{2k\omega_0 [\omega_0^2 + k^2] [\omega_0^2 + (k + k_s)^2]} Dk_i k_t e^{-kt} \\ - \frac{2Dk_i k_t (2k + k_s) e^{k_s t}}{[\omega_0^2 + (k + k_s)^2] [\omega_0^2 + (k - k_s)^2]}, t > 0 \\ - \frac{\omega_0 [\omega_0^2 + 5k^2 - 2kk_s] \cos(\omega_0 t) + [\omega_0^2(k + k_s) - 3k^2(k - k_s)] \sin(\omega_0 t)}{2k\omega_0 [\omega_0^2 + k^2] [\omega_0^2 + (k - k_s)^2]} Dk_i k_t e^{k_t t}, t < 0 \end{cases}$$

$$\text{For } k_s > 0, C_{As}^+(t) = \begin{cases} \frac{\omega_0 [3k^2 + 2kk_s - \omega_0^2] \cos(\omega_0 t) + [k^2(k + k_s) - \omega_0^2(3k + k_s)] \sin(\omega_0 t)}{2k\omega_0 [\omega_0^2 + k^2] [\omega_0^2 + (k + k_s)^2]} Dk_i k_t e^{-kt}, t > 0 \\ - \frac{\omega_0 [\omega_0^2 + 5k^2 - 2kk_s] \cos(\omega_0 t) + [\omega_0^2(k + k_s) - 3k^2(k - k_s)] \sin(\omega_0 t)}{2k\omega_0 [\omega_0^2 + k^2] [\omega_0^2 + (k - k_s)^2]} Dk_i k_t e^{k_t t}, \\ + \frac{2Dk_i k_t (2k + k_s) e^{k_s t}}{[\omega_0^2 + (k + k_s)^2] [\omega_0^2 + (k - k_s)^2]}, t < 0 \end{cases}$$

where $k \equiv k_m/2$ and $\omega_0 \equiv \sqrt{gk_t - k_m^2/4}$. These equations are then used to manually fit the experimental correlation data to extract the parameters $[k, \omega_0, k_s, D, k_i k_t]$.

Construction of $\Delta pilT$ and $\Delta pilU$ Mutants. *P. aeruginosa* in-frame deletion mutants were generated by allelic exchange as previously described (21). The following primer pairs were used to generate DNA fragments upstream and downstream of the *pilT* and *pilU* genes, respectively:

pilT KO P1 tgtaaacgacgcccagtgccaagcttgcctgGATCTTCGCGCGTTGCTCGG and
pilT KO P2 CGCGGCGGATCGGCGCCAGGAGGGACTCCCCAATTACAAGCA, and
pilT KO P3 TGCTTGTAATTGGGGAGTCCCTCCTGGCGCCGATCCGCCGCG and
pilT KO P4 ccatgattacgaattcgagctcggtaccgggatccTCAGCAGCACCTCGATCACC.
pilU KO P1 tgtaaacgacgcccagtgccaagcttgcctgGCAGCGCAAGGACTTCGAGG and
pilU KO P2 CGCTGGCCTACTGAAGACGGTGATGTTCTCGCTCACTCAGGG, and
pilU KO P3 CCCTGAGTGAGCGAGAACATCACCGTCTTCAGTAGGCCAGCG and
pilU KO P4 ccatgattacgaattcgagctcggtaccgggatccTGAGCCGCGAGTTAGCGGAG.

PCR products were cloned into pMQ30 via in vivo homologous recombination in *Saccharomyces cerevisiae* InvSci (Invitrogen) as previously described (21). Knockout constructs in pMQ30 were transformed into *Escherichia coli* S17 and introduced into *P. aeruginosa* via conjugation. Integrants were selected on LB medium containing gentamicin (20 μ g/mL) and nalidixic acid (20 μ g/mL) followed by sucrose counter selection. Resolved integrants were confirmed by PCR and DNA sequencing.

Construction of cAMP Reporter Strains. The cAMP reporter plasmid was provided by Z. Gitai, Princeton University, Princeton, NJ and described previously (3). The reporter plasmid contains a *yfp* transcriptional reporter for the cAMP-dependent *PaQa* operon and a *cfp* reporter fused to the *rpoD* promoter as an internal control. The reporter plasmid was introduced into PA14 WT and $\Delta pilU$ by electroporation.

Data and Code Availability. Binary images (generated from the bright-field images) and the fluorescence images (for datasets that use the cAMP reporter) are available at figshare.com (doi:10.6084/m9.figshare.5969335). These data are stored as MATLAB data files (.mat) containing variables that are 3-D matrices with dimensions [y,x,t]. Binary images are stored as a single variable per data file, where each variable contains images corresponding to a single fluorescence image. Fluorescence images are stored together in a single data file with 2 matrix variables corresponding to each fluorescence image channel. Each dataset is stored inside a .zip archive file. At present, the MATLAB code is in the alpha phase of development and not in a state ready for circulation.

SI Discussion

Surface sensing (22–34) has been used to describe diverse behaviors, including detection of surface proximity and the signaling cascade triggered by surface engagement. Examples include sensors based on flagellum rotation in *Vibrio parahaemolyticus* (22, 27, 29, 30) and analogous work in *P. aeruginosa* (24, 26, 28). In *P. aeruginosa* PAO1, surface sensing requires the clustering of the Wsp system to stimulate c-di-GMP synthesis (25, 32). The behavior of PAO1 can be different from that of PA14 due to differences in the repertoire of secreted EPS in early biofilm development. In other species, flagellum rotation can activate two-component systems that impact a range of behaviors (33, 34). Periplasmic stress pathways (Rcs and Cpx) (23, 31) activated via perturbation of the membrane and/or cell wall can also alert the microbe to cell-to-substratum contact. At present, it is not known why cells that contact the surface do not always

surface sense, or why surface sensing seems to be heterogeneous and enhanced with time, resulting eventually in irreversible attachment. Although similar observations have perplexed the field since the 1930s (35), the technological question of how to prevent bacterial surface adhesion (36) and the complementary scientific question of the bacterial mechanism(s) for adhering to surfaces, remain salient concerns.

- Kuchma SL, et al. (2010) Cyclic-di-GMP-mediated repression of swarming motility by *Pseudomonas aeruginosa*: The *pilY1* gene and its impact on surface-associated behaviors. *J Bacteriol* 192:2950–2964.
- Caiazza NC, O'Toole GA (2004) SadB is required for the transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas aeruginosa* PA14. *J Bacteriol* 186:4476–4485.
- Persat A, Inclan YF, Engel JN, Stone HA, Gitai Z (2015) Type IV pili mechanochemically regulate virulence factors in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 112:7563–7568.
- Tolker-Nielsen T, Sternberg C (2005) Growing and analyzing biofilms in flow chambers. *Current Protocols in Microbiology* (Wiley, New York).
- Zhao K, et al. (2013) Psl trails guide exploration and microcolony formation in *Pseudomonas aeruginosa* biofilms. *Nature* 497:388–391.
- Fulcher NB, Holliday PM, Klem E, Cann MJ, Wolfgang MC (2010) The *Pseudomonas aeruginosa* Chp chemosensory system regulates intracellular cAMP levels by modulating adenylate cyclase activity. *Mol Microbiol* 76:889–904.
- Miller JH (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
- Gibiansky ML, et al. (2010) Bacteria use type IV pili to walk upright and detach from surfaces. *Science* 330:197.
- Conrad JC, et al. (2011) Flagella and pili-mediated near-surface single-cell motility mechanisms in *P. aeruginosa*. *Biophys J* 100:1608–1616.
- Jin F, Conrad JC, Gibiansky ML, Wong GC (2011) Bacteria use type-IV pili to slingshot on surfaces. *Proc Natl Acad Sci USA* 108:12617–12622.
- Utada AS, et al. (2014) *Vibrio cholerae* use pili and flagella synergistically to effect motility switching and conditional surface attachment. *Nat Commun* 5:4913.
- Lee CK, et al. (2016) Evolution of cell size homeostasis and growth rate diversity during initial surface colonization of *Shewanella oneidensis*. *ACS Nano* 10:9183–9192.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188.
- Van Pelt J, Uyllings HBM, Verwer RWH, Pentney RJ, Woldenberg MJ (1992) Tree asymmetry—A sensitive and practical measure for binary topological trees. *Bull Math Biol* 54:759–784.
- Bennett RR, et al. (2016) Species-dependent hydrodynamics of flagellum-tethered bacteria in early biofilm development. *J R Soc Interface* 13:20150966.
- de Anda J, et al. (2017) High-speed “4D” computational microscopy of bacterial surface motility. *ACS Nano* 11:9340–9351.
- Zhang R, Ni L, Jin Z, Li J, Jin F (2014) Bacteria slingshot more on soft surfaces. *Nat Commun* 5:5541.
- Luo Y, et al. (2015) A hierarchical cascade of second messengers regulates *Pseudomonas aeruginosa* surface behaviors. *MBio* 6:1–11.
- Turing AM (1952) The chemical basis of morphogenesis. *Philos Trans R Soc Lond B Biol Sci* 237:37–72.
- Skerker JM, Berg HC (2001) Direct observation of extension and retraction of type IV pili. *Proc Natl Acad Sci USA* 98:6901–6904.
- Shanks RMQ, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA (2006) *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Appl Environ Microbiol* 72:5027–5036.
- McCarter L, Hilmen M, Silverman M (1988) Flagellar dynamometer controls swarmer cell differentiation of *V. parahaemolyticus*. *Cell* 54:345–351.
- Otto K, Silhavy TJ (2002) Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *Proc Natl Acad Sci USA* 99:2287–2292.
- Caiazza NC, Merritt JH, Brothers KM, O'Toole GA (2007) Inverse regulation of biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *J Bacteriol* 189:3603–3612.
- Güvener ZT, Harwood CS (2007) Subcellular location characteristics of the *Pseudomonas aeruginosa* GGDEF protein, WspR, indicate that it produces cyclic-di-GMP in response to growth on surfaces. *Mol Microbiol* 66:1459–1473.
- Merritt JH, Brothers KM, Kuchma SL, O'Toole GA (2007) SadC reciprocally influences biofilm formation and swarming motility via modulation of exopolysaccharide production and flagellar function. *J Bacteriol* 189:8154–8164.
- Ferreira RB, Antunes LC, Greenberg EP, McCarter LL (2008) *Vibrio parahaemolyticus* ScrC modulates cyclic dimeric GMP regulation of gene expression relevant to growth on surfaces. *J Bacteriol* 190:851–860.
- O'Toole GA (2008) How *Pseudomonas aeruginosa* regulates surface behaviors. *Microbe* 3:65–71.
- Gode-Potratz CJ, Kustusich RJ, Breheny PJ, Weiss DS, McCarter LL (2011) Surface sensing in *Vibrio parahaemolyticus* triggers a programme of gene expression that promotes colonization and virulence. *Mol Microbiol* 79:240–263.
- Ferreira RB, Chodur DM, Antunes LC, Trimble MJ, McCarter LL (2012) Output targets and transcriptional regulation by a cyclic dimeric GMP-responsive circuit in the *Vibrio parahaemolyticus* Scr network. *J Bacteriol* 194:914–924.
- Morgenstein RM, Rather PN (2012) Role of the Umo proteins and the Rcs phosphorelay in the swarming motility of the wild type and an O-antigen (*waal*) mutant of *Proteus mirabilis*. *J Bacteriol* 194:669–676.
- O'Connor JR, Kuwada NJ, Huangyutitham V, Wiggins PA, Harwood CS (2012) Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol Microbiol* 86:720–729.
- Cairns LS, Marlow VL, Bissett E, Ostrowski A, Stanley-Wall NR (2013) A mechanical signal transmitted by the flagellum controls signalling in *Bacillus subtilis*. *Mol Microbiol* 90:6–21.
- Guttenplan SB, Shaw S, Kearns DB (2013) The cell biology of peritrichous flagella in *Bacillus subtilis*. *Mol Microbiol* 87:211–229.
- Zobell CE, Allen EC (1935) The significance of marine bacteria in the fouling of submerged surfaces. *J Bacteriol* 29:239–251.
- Epstein AK, Wong T-S, Belisle RA, Boggs EM, Aizenberg J (2012) Liquid-infused structured surfaces with exceptional anti-biofouling performance. *Proc Natl Acad Sci USA* 109:13182–13187.

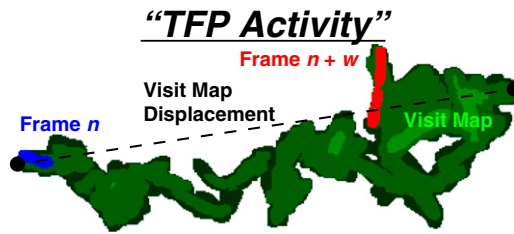


Fig. S8. Schematic illustration of our biometric signatures for TFP activity based on bacterial surface visit maps. We adapt previous studies on the “motility signature” of TFP-driven motion (1) and flagellum-driven motion (2, 3) for *P. aeruginosa* on surfaces to track generation-dependent TFP activity. High-speed microscopy shows that TFP-driven twitching motility in *P. aeruginosa* consists mostly of translations for variable temporal durations, alternating with combined translation-rotation “pulses” that are much shorter in temporal duration (1). Based on the predominance of surface translations along the cell body axis in TFP-driven motion, combined with the fact that these translations generally produce trajectories that have superdiffusive MSDs, the metric for TFP activity is defined in *SI Materials and Methods* and summarized visually here. Cells with high TFP activity have elongated visit maps and nonspinning behavior with nonzero displacement.

1. Jin F, Conrad JC, Gibiansky ML, Wong GC (2011) Bacteria use type-IV pili to slingshot on surfaces. *Proc Natl Acad Sci USA* 108:12617–12622.
2. Conrad JC, et al. (2011) Flagella and pili-mediated near-surface single-cell motility mechanisms in *P. aeruginosa*. *Biophys J* 100:1608–1616.
3. Bennett RR, et al. (2016) Species-dependent hydrodynamics of flagellum-tethered bacteria in early biofilm development. *J R Soc Interface* 13:20150966.

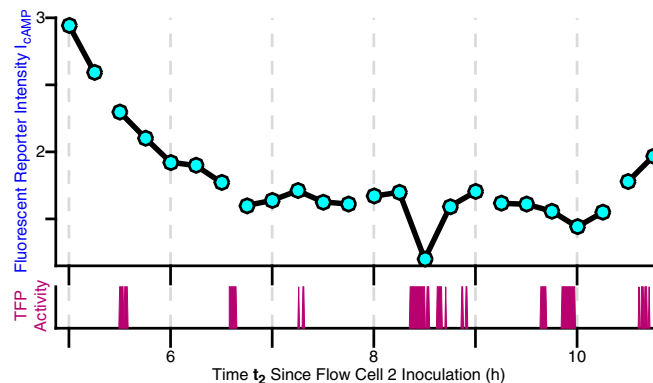


Fig. S9. Raw data for representative evolution of intracellular cAMP levels and of TFP activity for one lineage. The weak oscillations are obscured by the average trend and by noise. The raw data show a strongly damped trend: at most a few periods in TFP evolution and only about a period in the cAMP evolution. Such strongly damped behavior is consistent with the predictions of our Turing model (via the fits which include the damping constant values k , k_s , k_i , and k_t). Moreover, within the time frame of these experiments, these bacteria effectively become a sessile biofilm with essentially zero TFP motility, so the oscillations must eventually disappear if cAMP and TFP are coupled.

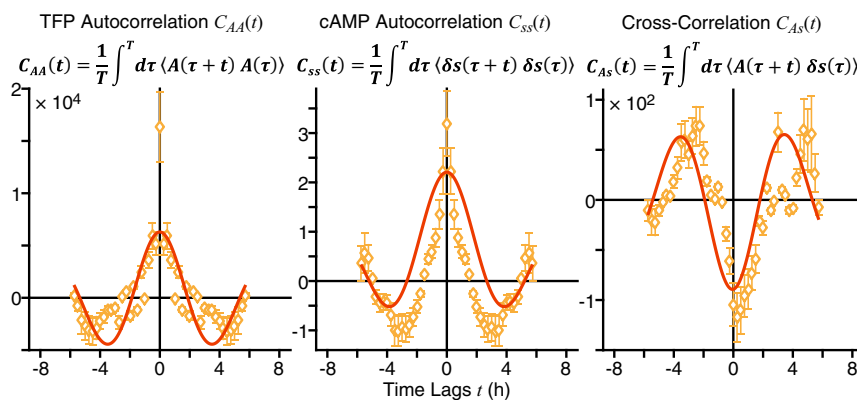


Fig. S10. cAMP-TFP correlation analysis for an additional family in WT FC2 ($T = 29.5$ h). Autocorrelations of TFP activity (*Left*), autocorrelations of cAMP reporter intensity (*Middle*), and cross-correlations between cAMP and TFP activity (*Right*) calculated from data for an additional family in WT FC2. Multi-generational traces of single branches were averaged before performing the correlations (calculation details are in *SI Materials and Methods*). Circles are experimental data; error bars indicate relative error of 1 divided by the square root of the number of points used in the correlation for that time lag. Solid lines indicate the model fit (parameters $[k, \omega_0, k_s, D, k_i k_t] = [0.1, 0.9, -0.25, 1200, 0.0125]$). There are negative cross-correlations at time lag = 0 h and positive cross-correlations at time lags of $\sim \pm 3$ h (*Right*).

