Biophysical Journal, Volume 98

Supporting Material

Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell

Shu-Wen Teng, Yufang Wang, Kimberley C. Tu, Tao Long, Pankaj Mehta, Ned S. Wingreen, Bonnie L. Bassler, and N. Phuan Ong

Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell

Shu-Wen Teng[*], Yufang Wang[*], Kimberly C. Tu[†], Tao Long[*], Pankaj Mehta[†], Ned S. Wingreen[†], Bonnie L. Bassler[†,‡], N. P. Ong[*]

[*]Department of Physics, [†]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA, [‡]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

1. *V. harveyi* strain construction

To demonstrate that the LuxR-*m*Cherry fusion retains functionality, we measured activity from promoter-*gfp* fusions to LuxR-controlled target genes in wild-type *V. harveyi* and two isolates of the same *V. harveyi* strain carrying the LuxR-*m*Cherry fusion. Both LuxR-*m*Cherry fusions activated fluorescence similarly to WT LuxR (Panel A of Fig. S1). Conversely, when fluorescence is repressed by WT LuxR, similar repression by theLuxR-*m*Cherry fusions is observed (Panels B and C).

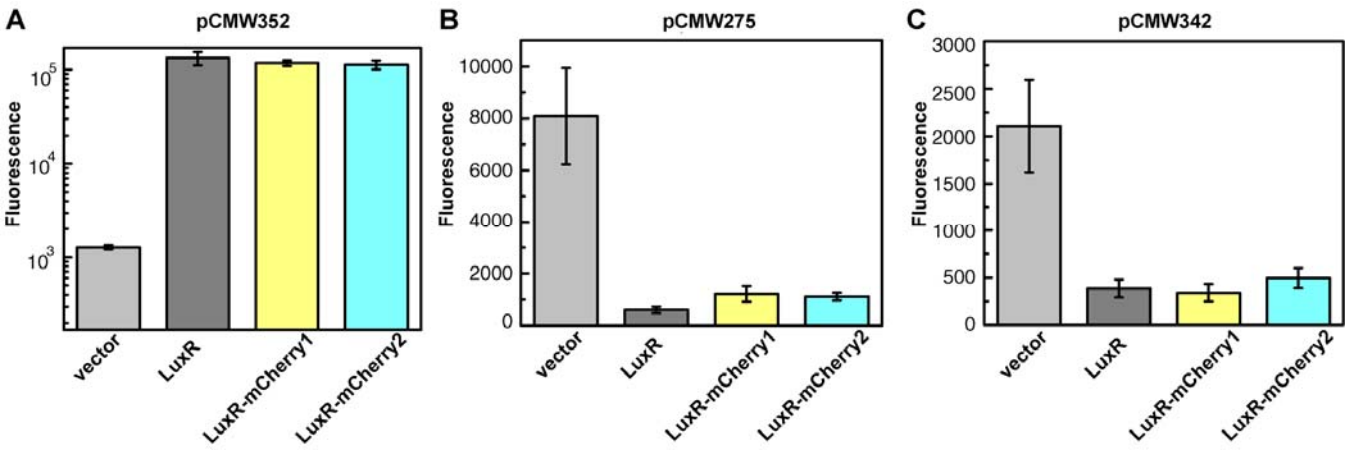

Fig. S1. Comparison of activation and repression by WT LuxR and LuxR-*m*Cherry fusions. A plasmid encoding a vector, *V. harveyi* WT LuxR protein, or the *V. harveyi* LuxR-*m*Cherry fusion was transformed into *E. coli*. Plasmids containing promoter-*gfp* fusions to direct targets of LuxR were transformed into the various strains (1). Fluorescence production was measured using a flow cytometer. (A) The target pCMW352 is activated by LuxR and LuxR-*m*Cherry. (B) The target pCMW275 is repressed by LuxR and LuxR-*m*Cherry. (C) The target pCMW342 is also repressed by LuxR and LuxR-*m*Cherry. Each sample was assayed in triplicate and error bars denote the standard deviation of the mean.

2. Growth conditions and Experimental set-up

V. harveyi strains were grown overnight in AB (autoinducer bioassay) medium (0.3 M NaCl, 0.05 M $MgSO_4$, 0.2% vitamin-free casamino acids, 0.01M $K_xH_yPO_4$, 0.01M arginine, 1% glycerol, pH7.5). Overnight cultures were subsequently diluted (1:2000) into fresh AB medium and grown for 12 hours (until OD_{600} reached 0.4). One-half micro-liter of culture was spotted on a clean No. 1.5 glass bottom Petri-dish (Willco Wells) and covered with a 1% agarose pad made of the same medium. As shown in Fig. S2, a small piece of coverslip was placed on top of the agarose pad, and the annular space between the 2 coverslips surrounding the pad was filled with mineral oil to prevent evaporation. To obtain the real boundary of cells, we stained the colony with FM4-64, a fluorescence dye that is known to accumulate in the cytoplasmic membrane.

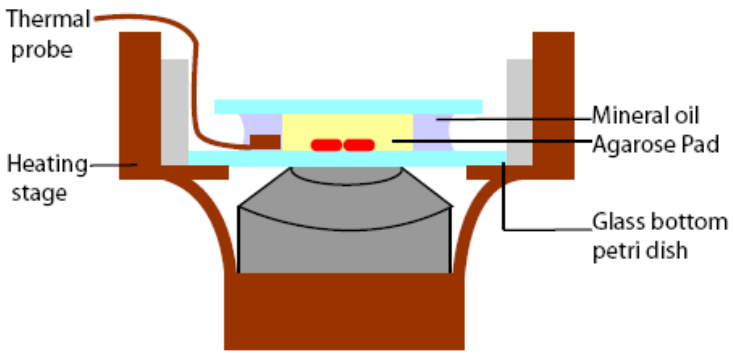

Fig. S2. Schematic of the experimental set-up. *V. harveyi* cells (red ovals) were grown under an agarose pad (yellow rectangle) placed between a coverslip and the bottom of a Petri dish (light blue strips). The space was surrounded by mineral oil as indicated. A thermistor continuously monitored the temperature of the experimental space.

3. Image analysis: Area determination

Custom software was developed using MATLAB (The MathWorks, Natick, MA) to estimate the areas of the individual cells in each frame of the time-lapse movies. The *V. harveyi* microcolony grows with a dense-packed morphology. In the phase-contrast images, each pixel was broadened by the point-spread function as well as the interference halo. However, because of the dense packing, the broadening severely affected the boundaries of the cells at the edges of the colony inferred from phase-contrast images leading to an overestimate of their area (by 20-30%). The enhanced distortion of the edges of cells was detected when we compared the phase-contrast images with the fluorescence images of test colonies used for calibration. Specifically, we stained live *V.*

harveyi with F4-64, a fluorescent dye that accumulates in the cytoplasmic membrane. The high-intensity fluorescence image produced by the stained membrane accurately located the true microcolony boundary. Thus, to avoid overestimating area, stained images and the phase-contrast images of the same microcolony were captured in rapid succession and compared with one another.

In Fig. S3A, the upper and lower insets show such (grey) phase-contrast images and (red) stained fluorescence images, respectively. The fluorescence profiles at the regions indicated by vertical yellow lines are plotted in Figs. S2B and S2C for the phase-contrast (F_p vs. y) and stained (F_s vs. y) images, respectively. Comparing the profiles of F_p and F_s, we note that their peaks agree well in the interior. Hence, this method can be used to define the edges of the interior cells. However, the images disagree significantly at the boundaries. The strong peaks in F_p at the right and left edges (Fig. S3B) are shifted outward compared with the true cell boundaries (located by the small peaks of F_s at the edges in Fig. S3C). After examining large numbers of such sections, we found an empirical, iterative method to accurately locate the true boundary using the phase contrast trace F_p vs. y. As a starting approximation, we used the midpoint y_{mid} between the first maximum and the first minimum in F_p to approximate the true cell boundary.

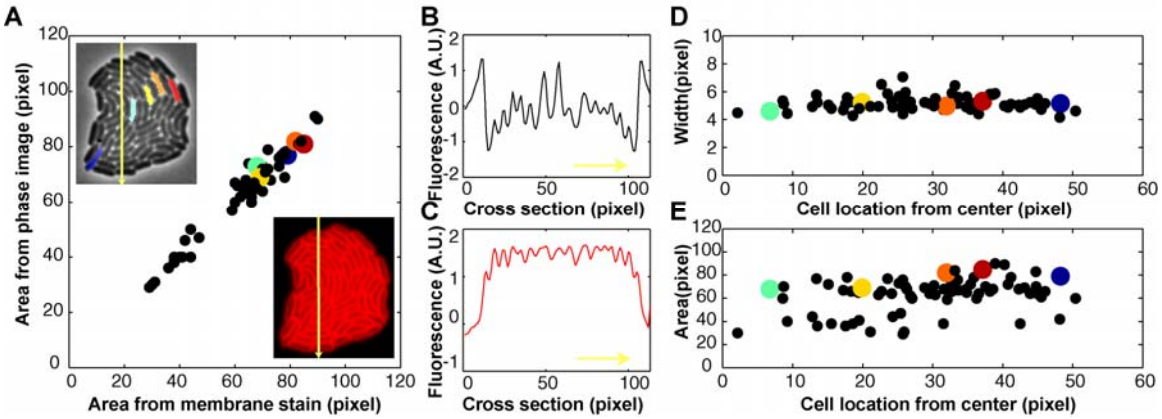

Fig. S3. Comparison of phase-contrast and membrane-stained images. (A) Comparison of individual cell areas A_k measured with phase-contrast microscopy of a microcolony (upper inset) and A'_k measured using stained-membrane fluorescence images of the same microcolony (lower inset). Each symbol represents a cell. The linear array of the symbols confirms that A_k agrees well with A'_k. (B) Fluorescence signal profile I_p vs. y along the vertical yellow line shown in the phase-contrast image in A. Peaks correspond to bright areas in the phase-contrast image. (C) Fluorescence profile I_s vs. y along the same section using the membrane-stained fluorescence image. The small peaks at the right and left edges show the true boundaries of the colony. (D) Plot of the cell width w_k of cell k (obtained from the phase-contrast image) versus its distance from the colony center. (E) Plot of cell area A_k versus the cell distance from the colony center. The flat profiles in Panels D and E confirm that there are no spurious correlations between the two quantities compared. The five colored cells in the upper inset of Panel A correspond to the same-color symbols plotted in the three panels A, D and E.

An improved estimate was subsequently obtained by shifting y_{mid} inward by one pixel. Hence, in the trace of F_p, the true boundary was located at $y_0 = y_{mid} \pm 1$ (where the correct sign is the one that shifts y_0 toward the interior). By incorporating these algorithms, the program automatically traces out the boundaries of both interior and

exterior cells of the microcolony and computes the cell areas A_k. The quality of the image processing was subsequently examined, and poorly segmented cells were corrected by hand. As a verification, we have plotted A_k of all the cells in the microcolony, determined from the phase-contrast image, against A'_k, the corresponding areas determined from the stained-membrane image only (Fig. S3A). The linear correlation confirms the expected, strictly linear scaling between A_k and A'_k. The colored symbols correspond to the five cells shown with the same colors in the upper inset. Figs. S3D and S3E plot the measured cell widths w_k and areas A_k, respectively, versus their positions from the center of the microcolony. A spurious enhancement of either quantity at the edges of the colony would be immediately apparent as an increasing curve. Clearly, the horizontal arrays in both panels show that these spurious effects are negligible.

4. Scaling between area and volume

In the confined space of the experimental set-up, cross sections of the growing *V. harveyi* cells were significantly distorted from circular cross-sections. The distortion results from both vertical compression (the "low-ceiling" effect) and horizontal compression (dense packing). Hence, we assumed that the measured area A_k scales linearly with the volume V_k. By contrast, for a circular cross-section, the observed A_k should scale as $\sqrt{V_k}$ (ignoring small end-corrections). To test this assumption, we examined how the measured (areal) fluorescence density F_k/A_k varied with A_k over a large population. Our test relies on the observed constancy of the concentration of LuxR protein over the 5-hour experiment. Since LuxR concentration is strictly proportional to the *volume* fluorescence density F_k/V_k, we expect F_k/V_k to be independent of V_k. Hence, if A_k is indeed proportional to V_k, we should observe F_k/A_k to be independent of A_k. By contrast, if A_k

varies as $\sqrt{V_k}$, we would instead have $F_k/A_k \sim (F_k/V_k)\sqrt{V_k} \sim \sqrt{V_k}$. Equivalently, F_k/A_k should increase *linearly* with A_k.

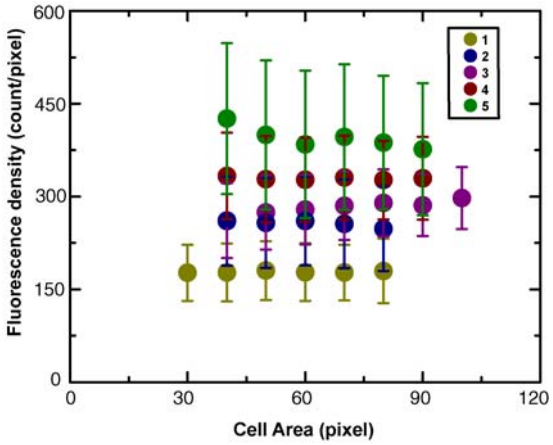

Fig. S4. A test for linear scaling between A_k and V_k. The observed (areal) fluorescence density F_k/A_k is plotted against the cell area A_k in each of the five samples investigated. In each sample (represented by a different color), the areas A_k are distributed over six or seven bins (the bin size is 10 area-pixels). For example, in Sample 1 (bottom row), A_k varies from 30 to 80 area-pixels. The horizontal pattern implies that A_k scales linearly with V_k. Indeed, if A_k scaled as $\sqrt{V_k}$, F_k/A_k should increase linearly with A_k (see text for details).

In each experiment, the observed areas A_k of the cells (~200-250 in population) varied over a substantial range during their cell cycle (from 30 to 90 area-pixels). Fig. S4 plots the fluorescence density F_k/A_k of all the cells for each of the five samples versus the individual areas A_k (for convenience, we have binned A_k into six or seven bins of width 10 area-pixels). As shown in the figure, F_k/A_k is independent of A_k, which confirms the assumption that A_k is linear in V_k. The flat variation also argues against other fractional powers V_k^{μ} ($\mu = ⅓, ¼$).

5A. Image analysis: cell lineage

Individual cell boundaries were used to collect data from fluorescent images, and the area of each cell at each time point t was recorded as the number of pixels inside the

boundary. The sum of the fluorescent counts of these pixels was recorded as fluorescence in the fluorescence channel. Background values were subtracted from the fluorescence channel. This algorithm was applied to six samples to create six ensembles. The quality of the image processing was subsequently verified for each frame. A tracking algorithm was applied to the time series of segmented images to obtain a time course for each cell and its descendant lineage. Tracking is based on the fact that there is little cell movement between frames. We therefore assume that the cell that occupies the location of the previous cell is the same cell or its descendant. This tracking analysis was also checked manually.

5B. Errors from pixelation and defocusing

Uncertainties in determining A_i and A^0_i caused by pixelation (CCD camera digitization) and errors associated with slight defocusing occur following an automated stage translation. The cell areas $A(t)$ were recorded every two minutes. In an average life cycle, this corresponds to ~25 measurements of the trace of $A(t)$ vs. t. A second-order, linear regression fit to the 25 points gives the "best fit" $A_{fit}(t) = at^2+bt+c$, from which both A_i and A^0_{i+1} may be found. To display the relative fluctuations of the 25 measurements about $A_{fit}(t)$, we plot in Fig. S5 the trace of the measured $A(t)$ normalized to $A_{fit}(t)$ [this is the first 120 minutes of the lineage (index 90) shown in Fig. 2D]. The fluctuations around 1 correspond to a standard deviation σ_x of 4%, which we identify with the standard deviation of each measurement. By a standard result (application of the Central-

Limit Theorem) in error analysis (2), the standard deviation of the mean σ_{xm} equals $\sigma_x/\sqrt{25} = 0.8\%$ (σ_x and σ_{xm} are unrelated to σ_A and σ_N in the main text). Hence, we estimate that the uncertainties in A_i (or A^0_i) are roughly 0.8%. This is the maximum error (caused by pixelation and defocusing) in locating the x-coordinate of each of the M (~250) points in the scatter plot in Fig. 3A. In the MLE process of finding σ_N, the standard deviation of the mean involves a further reduction of $\sqrt{250}$, which renders this source of error insignificant. We discuss a more important source of error in determining N_0 in Sec. 8 (MLE).

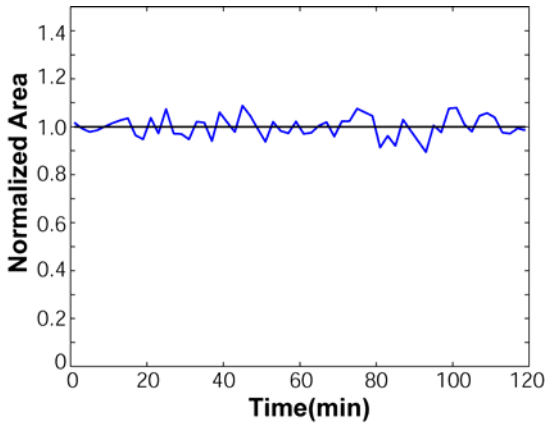

Fig. S5. Trace of the normalized area $A(t)/A_{fit}(t)$ (blue curve) in the first 120 minutes of the lineage (index 90) shown in Fig. 2D, where $A_{fit}(t)$ is the linear regression fit. The normalized quantity randomly fluctuates about 1 (black line) with a standard deviation of 4%. This implies that, on average, a single measurement of $A(t)$ has an uncertainty of 4%. Because the values of A_i are extracted from $A_{fit}(t)$ which is based on 25 measurements within a cell cycle, the uncertainty in A_i is given by the standard deviation of the mean, or $4\%/\sqrt{25} = 0.8\%$.

6. Distribution functions as joint probability

We describe in more detail the derivation of Eq. 1 of the text. The primary measured quantity in our experiment is the fluorescence-partition distribution $P_F(y)$, which changes in a reproducible, quantifiable way as the LuxR concentration is varied over a large range. By contrast, the area-partition distribution $P_A(x)$, fixed by biological and physical mechanisms outside of our control, does not change with LuxR concentration, and may be considered as given by the normal distribution Eq. 1, with fixed standard deviation σ_A. We consider the subset of events in which the mother-cell area divides in the ratio $1\text{-}x : x$ (with $0<x<½$). These events fall within the interval x dx in the scatter plot (Fig. 3A). The N_0 molecules distribute between the daughter cells according to this ratio, but the process is stochastic. Thus, our method is analogous to estimating the number N of "heads" in N_0 tosses of a coin with bias x ("heads" means the protein ends up in the cell with relative area x). The conditional probability is the binomial distribution $P(N \mid x) = \binom{N_0}{N} x^N (1-x)^{N_0-N}$.

For a binomial distribution, the mean $\langle N \rangle = N_0 x$, while the variance $\Sigma_N^2 = \langle (N-\langle N \rangle)^2 \rangle = Nx(1\text{-}x)$, where $\langle N \rangle = \sum_N N P_N$, *etc*. Consequently, the distribution of N is a bell-shaped curve that peaks at $N_0 x$ with a width $\sqrt{[N_0 x(1\text{-}x)]} \approx \sqrt{N_0}/2$. The peak increases linearly with N_0 while the width broadens as $\sqrt{N_0}$.

For our purpose, it is preferable to regard the measurement of the fluorescence as constituting N_0 attempts to measure the relative area x. This implies that, instead of N, we take the ratio $y = N/N_0$ as the sampling variable. Transforming the binominal distribution P_N given above to the y axis, it is clear that the distribution of y becomes a bell-shaped curve centered at $y = x$ with a width parameter σ_N equal to $\Sigma_N/N_0 = 1/(2\sqrt{N_0})$. With increasing N_0, the uncertainty in estimating x decreases as $1/(2\sqrt{N_0})$. Thus, when

the sampling count is very large ($\sigma_N \ll \sigma_A$, as in Sample 6), the fluorescence partitioning faithfully determines x without adding measurably to the uncertainty (whence $\sigma_F \approx \sigma_A$). Conversely, if the sampling count is low (Samples 1-5), the measurements y add an additional uncertainty ($\sigma_N > \sigma_A$) which reflects small-number fluctuations. This results in an enhanced total width σ_F for the fluorescence-partition distribution. We have exploited this additional broadening to determine N_0. In the limit N, $N_0 \gg 1$, the bell-shaped curve is well-approximated by a Gaussian function. Finally, multiplying $P(y|x)$ by $P_A(x)$ we obtain the joint-probability density $P(x,y)$ for observing a point (x,y), as given by Eq. 2.

7. Maximum Likelihood Estimation (MLE)

In MLE, we postulate an analytic function to describe a set of measurements. The function is characterized by a few parameters $(a,b,\ldots)$ whose values are unknown. The best estimates of the parameters are obtained by maximizing a "likelihood" function $L(a,b,\ldots)$ (3). In our experiment, the measurements are the set $\{x_i, y_i\}$ in an ensemble ($i = 1,\ldots, M$). As shown in Fig. 3 and Fig. S6A, these events are plotted in the x-y plane. The distribution of the points is postulated to be given by the probability density

$$P(x,y) = \frac{1}{2\pi\sigma_A\sigma_N} e^{-(y-x)^2/2\sigma_N^2} e^{-(x-x_0)^2/2\sigma_A^2} \qquad \text{(S2)}$$

where $x_0 = ½$, and (σ_A,σ_N) are the two parameters to be determined. The likelihood function is the joint probability that all M measurements are described by Eq. S2 with the *same* values of (σ_A,σ_N), *viz.*

$$L(\sigma_A,\sigma_N)=\left(\frac{1}{2\pi\sigma_A\sigma_N}\right)^M\prod_i^M e^{-(y_i-x_i)^2/2\sigma_N{}^2}e^{-(x_i-x_0)^2/2\sigma_A{}^2} \quad . \qquad \text{(S3)}$$

To maximize $L(\sigma_A,\sigma_N)$, it is convenient to take derivatives of $\log_e L(\sigma_A,\sigma_N)$. Setting to zero the derivatives with respect to σ_A and σ_N, we find that the optimal values are given by

$$\sigma^*{}_A{}^2=\frac{1}{M}\sum_{i=1}^M\left(x_i-\tfrac{1}{2}\right)^2, \qquad \sigma^*{}_N{}^2=\frac{1}{M}\sum_{i=1}^M\left(y_i-x_i\right)^2. \qquad \text{(S4)}$$

A measure of how likely the postulate is to be correct is obtained by plotting the contours of $L(\sigma_A,\sigma_N)$ in the σ_A-σ_N plane. The existence of saddle points or local maxima in close proximity would imply that the starting postulate is in doubt. However, as shown in Fig. S6B, the contour plot of Eq. S3 gives a single sharp maximum.

Error in finding N_0

In the MLE method, the contours of $\log_e L$ near its peak provide an estimate of the total uncertainties in fixing the optimal values of σ_A and σ_N. The contour representing the value of $\log_e L(\sigma_A,\sigma_N)$ one unit less than its maximum value (the smallest oval in Panel B of Fig. S6) gives the uncertainties in fixing σ_A and σ_N. We used the latter to define our error bars for N_0 plotted in Fig. 4C (values reported in Table I). For each sample, the largest source of this uncertainty is the fluctuation of the peak fluorescence $F^0{}_i$ about its ensemble average F^0 (as explained in Sec. 6B, errors from pixelation and defocusing contribute insignificantly to the uncertainties in σ_A and σ_N).

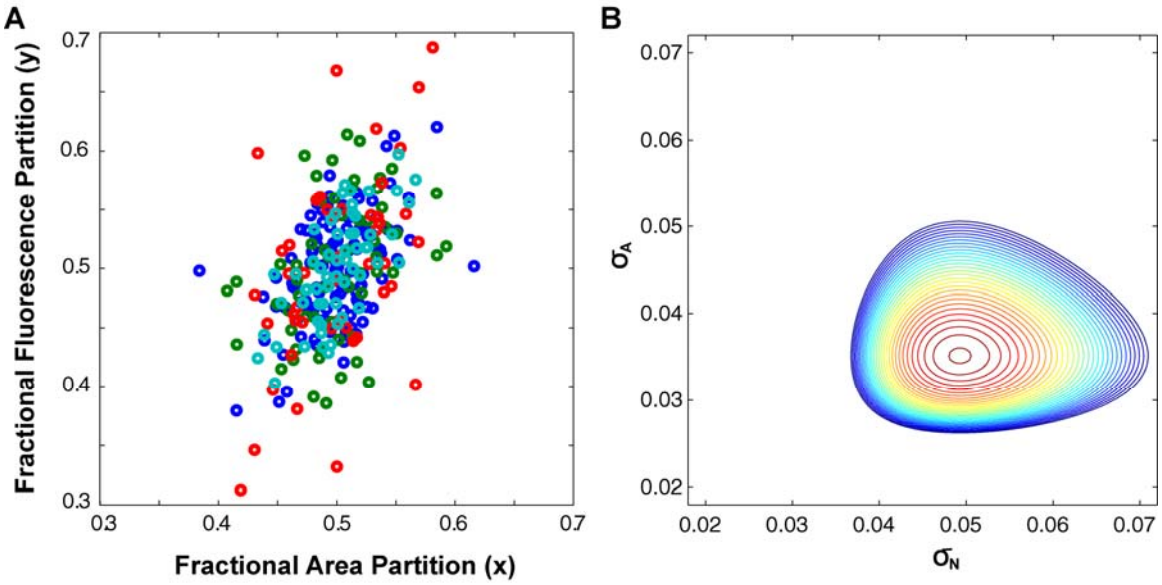

Fig. S6. Maximum likelihood estimation applied to the ensemble in Sample 2 (grown with AI = 0 nM). (A) Cluster plot of the set $\{x_i, y_i\}$ (i = 1,…, M) plotted in the x-y plane. As explained in the main text, $x_i = A_i/A^0_i$ is the fractional area, and $y_i = F_i/F^0_i$ is the fractional fluorescence for the cell-division event i. (B) Contour plot of the likelihood function $L(\sigma_A,\sigma_N)$ in the σ_A-σ_N plane defined in Eq. S3. Adjacent contours differ by a log unit of $L(\sigma_A,\sigma_N)$. The function $L(\sigma_A,\sigma_N)$ displays a single sharp peak at the optimal values $\sigma^*_A \approx 0.035$ and $\sigma^*_N \approx 0.048$. The smallest closed-loop contour determines the uncertainty in determining σ_A and σ_N.

8. Linearity between protein fluorescence signal and incident power

At high concentrations of LuxR (Sample 6), the high fluorescence intensity leads to a reduction in the viability of the *V. harveyi* colonies. This is apparent in the significant

lengthening of the cell division time and increased cell death which we suspect arises from photon toxicity or local heating of the cells. We eliminated these problems by sharply reducing the incident beam intensity in high-concentration samples. In order to compare signals across samples taken at different incident powers, we needed to verify that the fluorescence response of the LuxR-*m*Cherry is linear for the power levels employed.

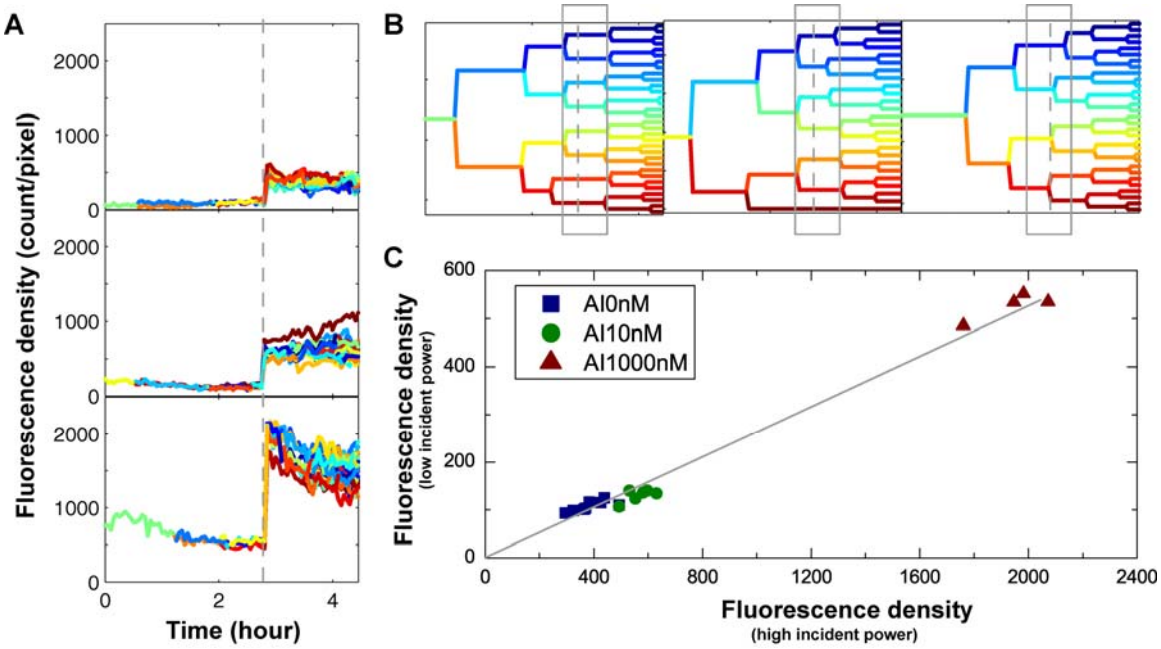

Fig. S7. Examination of linearity between the observed *m*Cherry fluorescence density $F(t)/A(t)$ and the incident light power at wavelength λ_{ex}=570 nm in three colonies containing dramatically different LuxR protein levels. (A) Curves of the fluorescence density observed in three colonies grown with AI = 0 (top panel), 10 nM (middle), and 1000 nM (bottom panel). At time t_c = 2.75 hours (dashed line), the incident power was increased by a factor of 4, causing $F(t)/A(t)$ to increase abruptly by a factor of ~4. (B) The lineage tree of the three colonies with AI =0, 10, 1000 nM are in the left, middle, and right panels, respectively. Dashed lines mark $t = t_c$. The three boxes identify the cells exposed to the 4-fold power increase. (C) Plot of fluorescence density $F(t_1)/A(t_1)$ (low power) versus $I(t_2)/A(t_2)$ (high power) measured in the three colonies, where $t_1 < t_c$ and $t_2 > t_c$. The

linear correlation between the symbols verifies that the observed fluorescence density is strictly proportional to the incident power. Moreover, the linearity holds up even with large LuxR concentrations corresponding to very high AI concentrations (i.e., 1000 nM).

To verify the linearity, we grew three colonies in three micro-chambers containing dramatically different AI concentrations (0, 10 and 1000 nM). At time t_c, the incident power was increased 4-fold, causing the fluorescence signal to increase proportionately. Comparing the measured F after the step increase with that before, we found that the increase in F is also 4–fold in all three samples, verifying the linearity of the response to the incident power (Fig. S7).

8. Protein Distribution Data Acquisition Analysis

For the static snapshot technique, overnight cultures were rediluted 10^6-fold in AI free and AI saturated AB media and grown to OD600~0.05. A volume 1ml of the culture was pelleted by centrifugation, re-suspended in ~10 μL of new media, and ~1μL placed between 1% agarose pad and a glass cover slip. By automating the stage control in the x-y directions and the focusing control in the z direction, we can search and measure the area and fluorescence in ~3000 cells in ~6 mins. Data analysis was performed using MATLAB (The Mathworks, Natick, MA). The phase contrast images were used to identify cell boundary and the corresponding pixels from the fluorescence image used to calculate the integrated cell fluorescence intensity, normalized by cell-size, to contruct histograms for single cell snapshot analysis. Objects with green fluorescence (internal standard) smaller than 0.5% of the mean were discarded. Matlab was used to calculate the variance and mean of the distribution function and for fitting to proposed distributions, e.g., the Gamma function.

Within the colony of 3,000 cells measured by the snapshot technique, the volume (or observed area) varies by a factor of ~2, reflecting different stages of the cell cycle. Since we are interested in intrinsic fluctuations of the protein fluorescence, we should factor out the cell-to-cell variation in volume. For each cell, we measured the observed area A as well as the total fluorescence signal from the cell. This allows the protein concentration p to be computed as fluorescence count per unit area. Thus the distribution function $G(p)$ does not include the trivial volume fluctuation factor. Knowing the scaling factor $\nu = F/N$ from the time-lapse experiments allows us to calibrate the protein concentration p in the plot of $G(p)$, which is plotted in Figs. 5C and 5D. In our experimental set-up, the scaling factor ν is determined to be ~175 counts per copy. The results show that ~40 photons per copy number per sec are collected by the CCD camera. At the set illumination level, each molecule's emission is estimated as ~400 photons/s before bleaching sets in. [This is computed from the spectrum of the xenon lamp, the transmission of the excitation filter, the reflection of the dichroic mirror and the fluorescence quantum yield of *m*Cherry (4)]. The value of ν implies that only ~5% of the photons emitted from each mCherry molecule are collected. This seems reasonable if we take into account the strong scattering inside the cell, the numerical aperture of the objective, the transmission of the emission filter and the dichroic mirror, and the sensitivity of the camera.

For the purpose of computing the Fano factor, however, it is convenient to express p as the dimensionless number $N_p = p\langle A\rangle$, where $\langle A\rangle$ is the mean area over the whole sample. Hence N_p is effectively the copy number per cell in the hypothetical case

that all cell areas are equal to $\langle A \rangle$. Expressing p as N_p allows us to read off the (dimensionless) Fano factor $\langle \delta N_p^2 \rangle / \langle N_p \rangle$ from the variance $\langle \delta N_p^2 \rangle$ and the mean $\langle N_p \rangle$.

Supplementary references:

1. Waters, C. M. and B.L. Bassler 2006. The vibrio harveyi quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers. Genes Dev. 20, 2754-2767.

2. Taylor, J. R. 1997. An introduction to error analysis : the study of uncertainties in physical measurements. University Science Books, Sausalito Calif.

3. Gelman, A.,J. B. Carlin,H. S. Stern and D. B. Rubin 1995. Bayesian data analysis. Chapman & Hall, London New York.

4. Shaner, N. C., R.E. Campbell, P.A. Steinbach, B.N.G. Giepmans, A.E. Palmer and R.Y. Tsien 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from discosoma sp red fluorescent protein. Nat. Biotechnol. 22, 1567-1572.