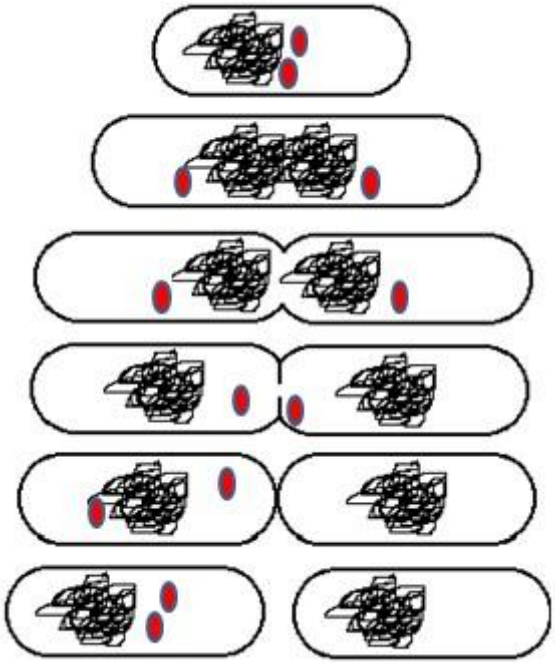
UPPSALA
UNIVERSITET

Partitioning error of transcription factors at cell division



M. Intakhar Ahmad

Degree project in Biology, 2012 (15 ECTS),
Biology Education Centre and Department of Engineering Sciences,
Uppsala University, Sweden.
Supervisor: Dr. Johan Elf & Dr. Andreas Grönlund

Abstract

Transcription factors (TFs) are molecules involved in regulating gene expression in cells. Many transcription factors in bacteria are present in less than 10 copies per cell. Inaccuracy in partitioning of these molecules at cell division leads to a large stochastic variation in the copy number between daughters with severe consequences for fidelity in gene regulation and other processes like metabolism. In this study, we have focused on the copy number variation of TFs in the light of previously published theoretical models for noise in cell division. We made a computer simulation in accordance of an experiment which was done to observe LacI (lactose operon repressor) molecule distribution at cell division in *Escherichia coli*. The experiment was performed in two different chemical conditions, with or without IPTG (Isopropyl β-D-1-thiogalactopyranoside). The model includes a coupled process of stochastic gene expression and partitioning of protein molecules. Parameters were fitted from experimental data. Three principles of distribution were considered in our simulation: equal, binomial and the one cased by chromosome hitchhiking mechanism. By modeling, the distribution of LacI molecules between two daughter cells over a long cell lineage, we revealed the extent of noise related to the three different distribution principles. We found that our experimental data fits best with our newly proposed "chromosome hitchhiking" model for transcription factor distribution.

Contents

List of abbreviations

E. coli	*Escherichia coli*
LacI	Lac operon Inhibitor or repressor
lacP	Lac promoter
lacO	Lac operator
IPTG	Isopropyl-β-D-thio-galactoside
CDF	Cumulative Distribution Functions

1. Introduction

Gene expression is a fundamental biological process that regulates specific sets of genes (genomic information) according to physiological or environmental conditions of the organism. Transcription factors (mostly protein in nature) govern the gene expression by binding or unbinding to the promoter of the particular gene, usually a DNA sequence "upstream" from the coding region of the gene. Gene regulation by transcription factors need to be highly specific to manage the cellar behavior properly [1]. Stochastic fluctuation in expression levels is unavoidable due to the low copy number of transcription factors which can bring about significant heterogeneity even in a isogenic populations of cells [2]. The extent of noise in the molecule number of a species is generally captured by the following mathematical term call "Fano" factor (F) [3].

$$F = \frac{\sigma_W^2}{\mu_W},$$

Here, σ_W^2 =variance, μ_W = mean of a random process in some time window w

The same fluctuation principle is also applicable to cell division and the error in equal resource partitioning between two daughter cells can be responsible for persistent fluctuation in all other partition dependent cellular processes [4].

Transcription factors are mostly bound to the chromosome, and this property makes them different from most other protein molecules in terms of distribution at cell division. Bound molecules travel to the newborn cells residing on the newly replicated chromosome (termed as “Chromosome Hitchhiking”). If it is so, the general belief of binomial partitioning of transcription factors could be incorrect.

With the recent experimental development in Elf lab at Uppsala University, it is now possible to count the numbers of LacI molecule (the transcription factor for *lac* operon in *E. coli*) that goes to each daughter cell at cell-division and preliminary results indicate that the partitioning accuracy is better than what can be expected from random partitioning between the daughter cells [5]. The experimentally generated data makes a

scope for a computer simulation of the particular cellular event for a very long lineage of cells to widen our view in understanding the extent of fluctuation with probable consequences.

Gillespie's stochastic simulation algorithm (SSA) [6] is very popular in modeling occurrence of stochastic or random events that follows a Markov process with continuous time and discrete-state space. LacI molecule production is a random event, which can be simulated with this algorithm. Rate of production can be determined from the experimental data. Three different recourse distribution principles can be taken into consideration. The molecules will divide equally, binomially or follow the hitch hiking model of distribution. The hitchhiking model depends on the number of binding sites for the transcription factor on the chromosome.

2. Model and Methods

2.1 Automatic detection of Cell division and generation time

Five minutes long time-lapse movies of monolayer of growing *E. coli* are made with Phase contrast microscopy. With the help of microbeTracker [7] aided with some in-house MATLAB scripts, each and every cell throughout all the frames is tracked assigning a specific number for each cell in a specific image frame. A cell division function called “Lambda function” is measured for each cell [Ullman G, personal communication] to distinguish divided and undivided cells. We have also computed that the generation time (T_g) of cells, which is around 24 minutes (the median value of the experimental data of 151 cells) in the provided conditions.

2.2 The *lac* Operon*:* our model genetic circuit

The lactose operon is the perfect genetic circuit for our study. It is the most studied and well characterized genetic circuit so far. We can easily control the different dynamic states of the repressor molecule (LacI) of the circuit by using IPTG (Isopropyl-β-D-thio-galactoside) [8]. LacI has low copy number (about 10 molecules per cell). Moreover, our strain, JE12 has a special feature; the repression has been increased by inserting an O1 in the place of O3 because O1 has better binding strength (here, O1, O2 and O3 represent repressor binding sites) [9]. This feature improves the in counting of LacI molecules having less copy number then the wild type. For our simulation the replication point of *lacI* gene is crucial for the rate of LacI production. The replication point (T_r) is calculated theoretically by the following equation. Published features of the cell cycle have been taken into consideration [9].

$$Tr = 2.T_g - Rep_i + C\left(\frac{L1}{L/2}\right)$$

Here, generation time (T_g) = 24min, replication initiation in mother cell before cell division (Rep_i) = 60min(a full cell cycle), replication time for the entire chromosome(C) = 40min, distance between *lacI* and *oriC*(L1) = 1082411bp, length of the whole genome(L) = 463967bp.

Thus we optain the replication point of *lacI* gene is about 6.8 minutes after birth.

2.3 Rate of lacI molecule production per cell per generation

Fluorescence images of the cells provide us the copy number of LacI molecules per cell. Lambda function (a principal components analysis regarding morphological features of cells) is used to determine the cell age and the division feature. The production rate doubles after the replication of *lacI* gene. The average molecule production per generation is needed to be divided into two parts.

Rate before transcription of LacI in the daughter cell, K1 = $\frac{Km}{\frac{Tr}{Tg}+2.\frac{Tg-Tr}{Tg}}$

Here, average rate of lacI production is Km, replication point of *lacI* gene (Tr) = 6.8 minutes after birth and generation time (Tg) =24 min.

Rate after transcription of LacI in the daughter cell, K2 = 2 x K1.

2.4 Time points of new LacI molecule production

The *lacI* gene is transcribed and the molecules are produced. This process does not depend on the preexisting number of LacI in the cell. So, production of a new LacI molecule is a memory less event.

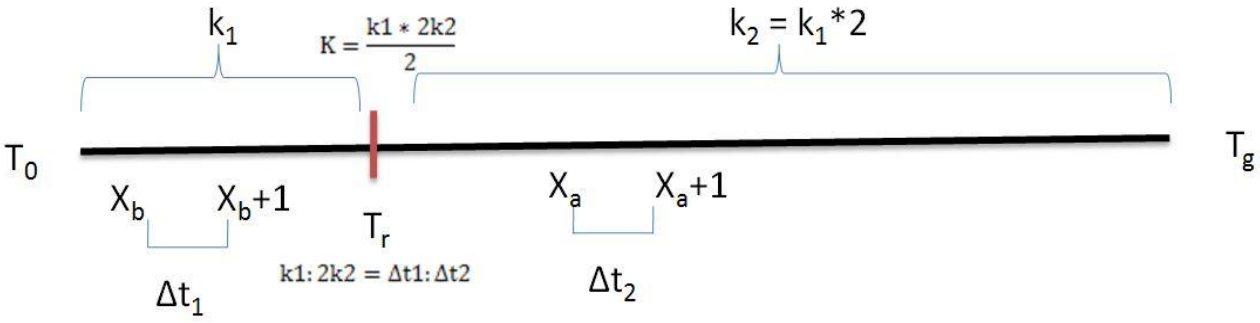


Figure 2.1 Schematic representation of LacI molecule production in *E.coli* Cells.

A cell is born at $t_{0,}$ new lacI molecules(X) are produced as a random event with the interval of Δt and rate $k_{1.}$ Production rate is doubled after new *lacI* replication Tr.

Memory less random time points (Δt) can be simulated by Gillespie algorithm [6] which obeys the following equation

$$\Delta t = -\frac{\ln(1-x)}{k}$$

x is probability mass (0> x <1) and k is the rate of production.

2.5 Principles of LacI molecule distribution at cell division

Three different distribution principles are taken into consideration for the simulation

a. Equal distribution: One half goes to one daughter cell and other half goes to another daughter cell (if we consider the division event is absolutely fair).

b. Binomial distribution: Each lacI molecule has 50% chance to any of the two daughter cells. This feature of low copy number molecule distribution during cell division was hypothesized previously [11].

c. Distribution following chromosome hitchhiking model: The hitchhiking model is dependent of the number of binding sites of the particular molecule on chromosome. Equal distribution is followed up to the number of available binds sites, and the rest of the molecules distribute binomially. For example, in our case, in a growing *E. coli* cell, there are around 4 binding sites for lac operon repressor (the number is dependent on the number of ongoing replication forks). That is why, up to four molecules, the distribution is equal and if there are more than 4 molecules, 4 of them distribute equally, and the rest of them follow binomial distribution. For *E. coli* cells the number of binding sites (N) in a growing cell can be calculated by the following equation.

$$N = n\,[\ln 2.\,(\frac{Tg}{C})]\,[2^{(C+D)/Tg} - 2^{D/Tg}]$$

Here, n=Number of binding sites per chromosome, Tg=generation time, C=C (replication time of total chromosome) period in *E. coli.* =D (after replication to cell division) period in *E coli*.

2.6 Representation of LacI molecule distribution variation between two daughter cells.

In our simulation we represent the LacI molecule distribution variation between two daughter cells with a variable (V). We have V values for our experimental data with two different conditions (cells are grown with or without IPTG).Later, the statistical significance of our experiential data is observed placing them on the simulation curves. The V value is calculated by the following equation

$$V = \frac{\sum_{i=1}^{n} \left[\frac{(D1-D2)^2}{M}\right]_i}{n}$$

Here, D1 = number of molecules in one daughter. D2= number of molecules in other daughter cell. M= number of molecules in mother cell what is distributed between two daughter cells. n=total Number of mother cells.

3. Results

3.1 Selection Generation time calculation

We have followed 151 cells in our experimental condition to measure the average generation time of cells in our experimental condition.

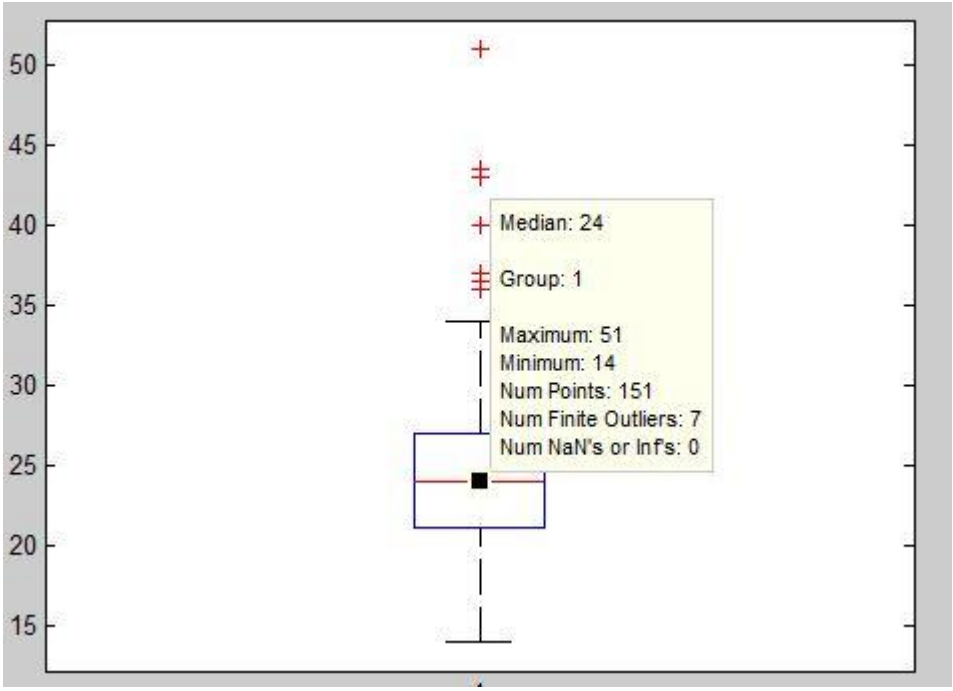


Figure 3.1 Box plot of generation times in minutes.

3.2 Selection of cells of interest

We have experimental data for about 500 cells, both in IPTG and Without IPTG condition. We have calculated the lambda value for each cell. We have selected cells of our interest (which are just divided) depending on the lambda value (all cells having lambda above 13.5). We sampled 105 and 103 cells with or without IPTG respectively.

3.3 Attributes of the selected cells

Attributes of the cells, necessary for the simulation have been calculated. The result is summarized in the following table.

Table 3.1: Calculated parameters for simulation

<table>
<tr><td>Average number of molecule production per cell per generation</td><td>Rate of molecule production(K)</td><td>Distribution variation(V)</td><td>Condition</td></tr>
<tr><td rowspan="2">2.4190</td><td>Before lacI replication(k1)=0.0587</td><td rowspan="2">V=0.3376</td><td rowspan="2">+IPTG</td></tr>
<tr><td>After lacI replication(k2)=0.1174</td></tr>
<tr><td rowspan="2">1.7864</td><td>Before lacI replication(k1)= 0.0434</td><td rowspan="2">V=0.3135</td><td rowspan="2">-IPTG</td></tr>
<tr><td>After lacI replication(k2)= 0.0867</td></tr>
</table>

Note: We use very precise values for better simulation.

3.4 Simulation results

We made 1 million simulations for each case. Thus, we optioned 1 million individual

points representing the distribution variation. We observed bell shaped curves plotting those points (figure 3.2)

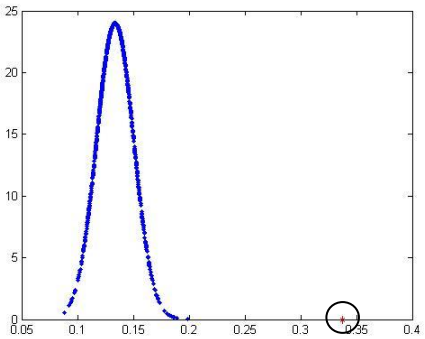
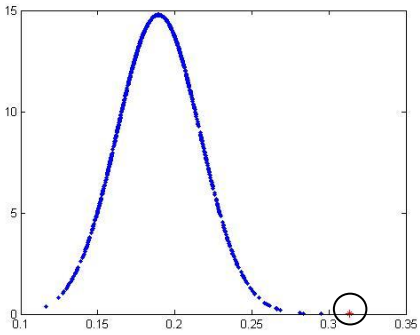
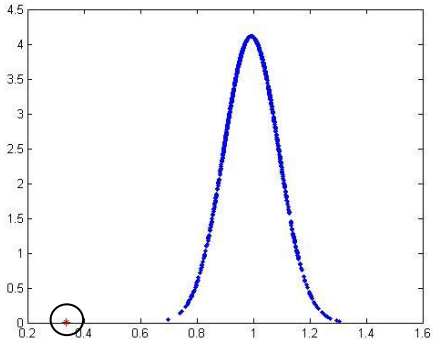
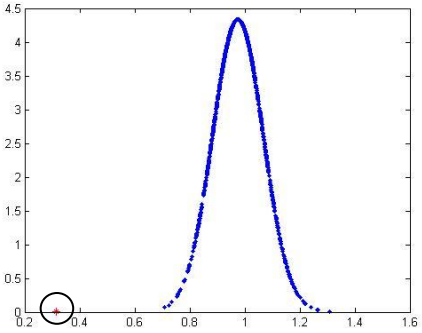
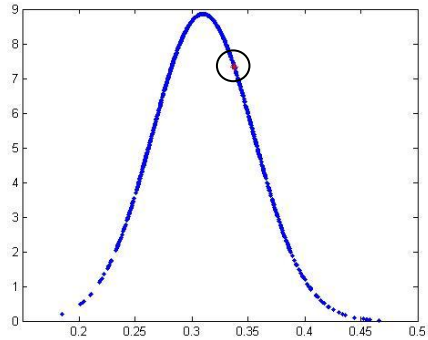
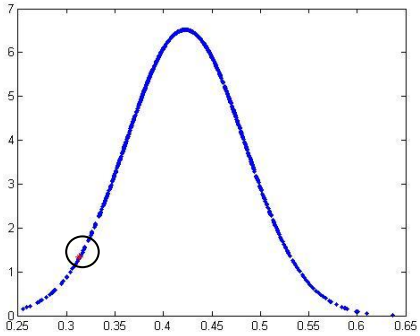
With IPTG	Without IPTG	Principle
		Equal Distribution
CDF=1(100%)	CDF=1(100%)	
		Binomial Distribution
CDF= 0(0%)	CDF= 0(0%)	
		Hitchhiking Model
CDF= 0.0587 (58%)	CDF= 0.0013(1.3%)	

Figure 3.2: Graphical representation of statistical significance of experimental data compared to the simulation. The curves represent the distribution variation (V) of 1000 simulated experiment (equivalent of 100 hours real experiment for each simulated

experiment, total=1000x100 hours).The values are normally distributed (blue line) and the red spots represent the experimental value on the simulated distribution results.

3.5 Normalization of simulation results

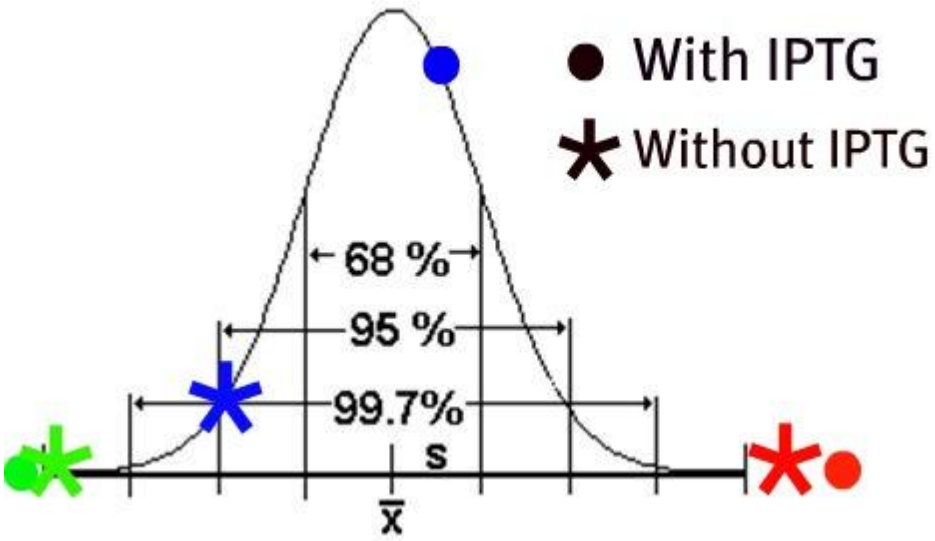


Figure 3.3 Normalization of the simulation result on a single bell curve. Statistical significance of experimental data regarding different hypothesis are presented with different colors and shapes. Equal distribution is represented with red (circle for IPTG+ and star for IPTG-). Similarly; Green represents Binomial distribution and, blue represents Hitchhiking model distribution.

4. Discussion

Stochasticity is everywhere in biological systems. Biochemical reactions that involve small numbers of molecules are intended to be noisy. Noise plays roles in various biological functions, including introduction of errors in replication process leading to mutation and evolution, noise drives divergence of cell-fate, noise can induce amplification or reduction of cellular signaling process, controls a switching point in a pathway and produces probabilistic pathway selection. In general, it helps to generate the qualitative individuality of cells. On the contrary, if the noise is above the threshold of tolerance, it becomes detrimental. However, the coherence of a biological system suggests that, there are some control mechanisms for noise; those are under evolutionary selective pressure. For example, inefficient translation is overcome by frequent transcription [12].

Recent publications suggest a new kind of noise that is molecule partitioning noise at cell division [4]. With some experimental data (observed at the single molecule level) and computer simulation, we shed light on the pattern and the extent of this noise. We fitted parameters for the simulation from experimental data that made the simulation more realistic. However, in our experiment, we had some limitation in terms of identifying single molecules with very high precision. In our experimental data set, we allowed some margin of error (confidence interval, CI = 90%).

We had additional difficulty with the data generated without IPTG in the medium. When there was no IPTG in the medium, LacI molecule are bound to the DNA. Because of the binding distance between two molecule is smaller than the resolving power of our optical system, we underestimated the number of molecules in cells.

In general, errors are randomly distributed. Despite of having some additional randomness form experimental limitations, the data suggested that the molecule distribution after cell division is remarkably equal, when there was no IPTG in the medium. Even in the other condition (with IPTG), the distribution had a good tendency

to distribute evenly. A plausible explanation; there could be some non-specific DNA binding event takes during cell division that helps them to distribute uniformly with chromosome itself.

The overall impression we get from the simulation is that, the cells must have some underlying mechanisms helping them to distribute the molecule more evenly than the random way. It is highly likely that the chromosome hitchhiking could be the reason or one of many reasons. In future, we need to improve our magnification capability of the optical system. Our sample size should be bigger to come up with a strong conclusion.

Another fascinating thing has been published recently claiming that bacteria can transfer molecules in a closed proximity from a very small membrane nano-tube serving as a mediator for intra-cellular molecule exchange [13]. If it is so, we need to redefine the cellular individuality in bacterial population.

Acknowledgements

A big thank to my supervisors, Johan Elf and Andreas Grönlund, for their overall support and constant guidance to complete this project. I am also grateful to Gustaf Ullman, Mats Wallden ,Anel Mahmutovic, Fredrik Persson and David Fange for their valuable comments and suggestions.

References

1. Alberts B. *Molecular Biology of the Cell*. New York: Garland Science, 2002. Print.
2. Kim KH, Sauro HM. Measuring retroactivity from noise in gene regulatory networks. *Biophys J*. 100(5):1167-1177 (2011).
3. Fano U. Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions. *Phys. Rev*. 72(1):26–29 (1947).
4. Huh D, Paulsson J. Non-genetic heterogeneity from stochastic partitioning at cell division.
Nat. Genet. 43(2):95-100 (2011).
5. Wallden M, Ullman G, Ahmad I, Marklund E, Mahmutovic A and Elf J. Active partitioning of low copy number transcription factors in E. coli cell division.(In preparation)
5. Gillespie DT. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J Comput Phy*. 22(4): 403-434(1976).
6. MicrobeTracker Suite. Available at http://emonet.biology.yale.edu/microbetracker/
7. Müller-Hill B. The Lac Operon: a Short History of a Genetic Paradigm. Berlin: Walter De Gruyter, 1996. Print.
8. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*. 316(5828):1191-1194 (2007).
9. Michelsen O, Teixeira de Mattos MJ, Jensen PR and Hansen FG. Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r. *Microbiol*.149:1001-1010 (2003).
10. Berg OG. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol.* 71(4):587-603 (1978).
11. Raser JM and O'Shea EK. Noise in Gene Expression: Origins, Consequences, and Control. *Science .* 309(5743):2010-2013 (2005).
12. Dubey GP and Ben-Yehuda S. Intercellular Nanotubes Mediate Bacteria Communication. *Cell* 144(4):590-600 (2011).