Variability of the distribution of differentiation pathway choices regulated by a multipotent delayed stochastic switch

Andre S. Ribeiro, Xiaofeng Dai, Olli Yli-Harja

www.elsevier.com/locate/yjtbi

Variability of the distribution of differentiation pathway choices regulated by a multipotent delayed stochastic switch

Andre S. Ribeiro, Xiaofeng Dai, Olli Yli-Harja

PII: S0022-5193(09)00251-3
DOI: doi:10.1016/j.jtbi.2009.05.025
Reference: YJTBI 5581

To appear in: *Journal of Theoretical Biology*

Received date: 16 February 2009
Revised date: 25 May 2009
Accepted date: 26 May 2009

Cite this article as: Andre S. Ribeiro, Xiaofeng Dai and Olli Yli-Harja, Variability of the distribution of differentiation pathway choices regulated by a multipotent delayed stochastic switch, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2009.05.025

Variability of the distribution of differentiation pathway choices regulated by a multipotent delayed stochastic switch

Andre S. Ribeiro[a,b], Xiaofeng Dai[a], and Olli Yli-Harja[a,c]

[a] Computational Systems Biology Research Group, Dept. of Signal Processing, Tampere University of Technology, Finland.
[b] Centre for Computational Physics, Univ. Coimbra, P-3004-516 Coimbra, Portugal
[c] Institute for Systems Biology, 1441N 34th St, Seattle, WA, 98103-8904, USA.

*Contact information of corresponding author:
Mailing address
Andre S. Ribeiro
Department of Signal Processing, Tampere University of Technology
P.O Box 553
33101 Tampere, FINLAND
E-mail address: andre.sanchesribeiro@tut.fi
Tel. No.: +358 408 490 736
Fax No.: +358 331 154 989

ABSTRACT

We study the plasticity of a delayed stochastic model of a genetic toggle switch as a multipotent differentiation pathway switch, at the single cell and cell population levels, by observing distributions of differentiation pathways choices of genetically homogeneous cell populations. Assuming a model of stochastic pathway determination of cell differentiation that is regulated by the proteins of the switch, we vary the proteins' expression level and degradation rates, which cells are known to be able to regulate, to vary mean level, noise, and bias of the proteins' expression levels. It is shown that small changes in each of these dynamical features significantly and distinctively affects the dynamics of the switch at the single cell level and thus, the cell differentiation patterns. The regulation of these features allows cells to regulate their pluripotency and cell populations' distribution of lineage choice, suggesting that the stochastic switch has high plasticity regarding differentiation pathway choice regulation, thus providing adaptability to environmental stresses and changes.

Keywords: Gene regulatory network, stochastic, cell lineage, toggle switch

I. INTRODUCTION

It has long been hypothesized that some types of cell differentiation are based on bistable genetic sub-circuits controlling many downstream genes (Monod and Jacob, 1961). In this process, a stem cell turns into a 'stable' phenotype. The genetic decision circuit of differentiation must be (at least) bistable, to allow branching into distinct cell types, and reliable to prevent reversibility (Gardner et al., 2000). Recent evidences suggest that the neutrophil cell lineage has such bistable switch-like behavior (Chang et al., 2006). Single cell analysis of the expression kinetics of the differentiation marker CD11b (Mac-1) revealed an all-or none switch-like behavior in HL60 promyelocytic precursor cells that transit to the neutrophil cell lineage. The progression from the precursor to the differentiated state is a discrete transition between low- ($CD11b_{Low}$) and high- ($CD11b_{High}$) expressing subpopulations which are distinguishable in a bimodal distribution (Chang et al., 2006).

In the deterministic framework, it was hypothesized that cell types are attractors of the gene regulatory network dynamics (Kauffman, 1969). To be biologically plausible, the attractors need to be very confined patterns of gene expression (Kauffman, 1993). An important criticism to this framework (Aldana et al., 2002) is that noise renders attractors a poor model of cell types since the closure of an attractor in the discrete dynamics is delicate. To address this, the concept of 'ergodic set' was proposed as a set of states from which the gene network, once entering, does not leave for a very long time, given internal noise (Ribeiro and Kauffman, 2007). An ergodic set can be composed of one or several 'noisy attractors', which will be the set of states where the gene regulatory network spends most of the time when on that ergodic set. It was shown, using a delayed stochastic model of gene regulatory networks (Ribeiro et al., 2006), that there are stochastic gene networks that are able to remain in confined

regions of the state space for very long periods of time, in agreement with the hypothesis that ergodic sets corresponds to cell types (Ribeiro and Kauffman, 2007). In support of this hypothesis, it was observed that *Bacillus subtilis* transiently and probabilistically differentiates between two phenotypes (Suel et al., 2006). One phenotype is able to uptake DNA from the environment and appears mostly when the cells are under stress. This might be an example of noise-driven transitions between noisy attractors of the same ergodic set.

The toggle switch, a gene network of two mutually repressing genes, can be used by cells to adopt different phenotypes (Arkin et al., 1998)(Neubauerz and Calef, 1970)(Gardner et al., 2000) and as decision circuits of differentiation pathways (Huang et al., 2005). Recent evidences further support that stochastic genetic switches in cells determine alone, or within a network of bistable switches, cell differentiation lineages, and are also responsible for preventing reversibility (Wang et al, 2009).

The bistability of the toggle switch, i.e., having two noisy attractors (with one gene's expression level high and the other low), depends on its internal noise level. Given sufficient noise, the two genes can express simultaneously for long periods of time, thus, the toggle switch can be tri-stable (Ribeiro, 2008). Importantly, this 'unstable attractor' can be the region of the state space where cells spend most of its lifetime (Ribeiro, 2008).

Previously, assuming genes following a Boolean-like noise-free dynamics, proposals of intervention strategies to drive differentiation focused on the mechanisms that silence or over-express genes (Shmulevich et al, 2002). More accurate measures of gene expression showed that the dynamics of gene regulatory networks is 'richer' than the assumed on-off logic and that its inherent stochasticity cannot be neglected (Kaern et al, 2005)(Elowitz et al, 2002), especially in the process of determination of

differentiation pathways (Arkin et al, 1998) and as a source of phenotypic diversity (Blake et al, 2006)(Weinberger et al, 2006). The increased understanding of cell differentiation raises questions concerning the plasticity of the genetic circuits regulating differentiation pathway choices, i.e., besides the intervention strategies aforementioned, it might be possible to regulate differentiation by using the inherent stochasticity of gene networks. Here we address this issue and explore strategies to regulate patterns of cell differentiation.

Specifically, we model single cells, each with a toggle switch whose dynamics is driven by the delayed stochastic simulation algorithm (delayed SSA) (Roussel and Zhu, 2006), and investigate how the distribution of differentiation pathway choices of an initially undifferentiated cell population is determined and can be altered by the changes in the dynamics of the switch. We chose to vary proteins' degradation rates, since it has been shown that these vary widely from one protein to another and there are evidences that cells can regulate (and thus evolve) such rates to some extent (Cooper, 2000), and transcription initiation rates, which can be subject to regulation (Golding et al, 2005) and are, to some extent, sequence dependent (Herbert et al., 2006), thus evolvable.

We assume a simple scenario where in each cell, the proteins' levels of its toggle switch determine, at a moment in time determined stochastically, which one of four possible differentiation pathways does the cell follow. This choice is "binary" in the sense that it depends only on the "presence" or "absence" of the two proteins, but stochastic since, given the presence of proteins, they will compete towards 'activating' distinct differentiation pathways via a specific chemical reaction out of the four possible ones (modeling the 'activation' of different sets of downstream genes). We

study how changes in the dynamics of the toggle switch affect the ratios of cells opting for each pathway.

II. METHODS

Delayed Stochastic Gene Network Model

We model gene regulatory networks according to the modeling strategy proposed in (Ribeiro et al., 2007). We use the delayed SSA (Roussel and Zhu, 2006) to drive the cell gene network's dynamics since evidences suggest that relatively low levels of expression (Bon et al, 2006) and small changes in expression levels affect early cell fate outcomes (Heyworth et al, 1999). Namely, the number of molecules involved in the reactions controlling differentiation can be from one to a few (Arkin et al, 1998). Therefore, noise needs to be accurately accounted for in the differentiation process of multipotent cells (Bruno et al, 2004).

This modeling strategy also accounts for the time needed for transcription and translation to be completed, once initiated, which can be from a few seconds to several minutes, and thus, cannot be neglected (McClure, 1980)(Zhu et al, 2007) and was shown to affect significantly the dynamics of gene networks (Ribeiro, 2007).

Gene regulatory networks consist of genes coupled via protein-protein interactions and protein-operator sites interactions. Transcription and translation are independently modeled as multiple time delayed reactions. The dynamics is driven by a modified version of the original Stochastic Simulation Algorithm (SSA) (Gillespie, 1977), the 'delayed SSA' (Roussel and Zhu, 2006), which uses a waiting list to store delayed output events. The waiting list is a list of elements (e.g., proteins being produced and occupied promoter regions), each to be released after a certain time interval (also stored in the waitlist). The algorithm proceeds as follows:

1) Set t = 0, t_{stop} = stop time, read initial number of molecules and reactions, create empty waitlist *L*.

2) Do an SSA step for reacting events to get next reacting event R_1 and corresponding occurrence time $t + t_1$.

3) Compare t_1 with the least time in *L*, t_{min}. If $t_1 < t_{min}$ or *L* is empty, set: $t = t + t_1$. Update the number of molecules by performing R_1, adding delayed products (if existing) and the time delay they have to stay in *L* from the appropriate distribution.

4) If *L* is not empty and if t_1 ¸ t_{min}, set $t = t + t_{min}$. Update the number of molecules and *L*, by releasing the first element in *L*; otherwise go to step 5.

5) If $t < t_{stop}$, go to step 2; otherwise stop.

The delayed SSA was used to simulate the dynamics of a model of a repressed gene (Zhu et al, 2007) which matched the observed dynamics of RNA and proteins levels measured at the single molecule level (Yu et al, 2006).

Other stochastic algorithms have been proposed to simulate chemical reactions with time delays. In (Bratsun et al, 2005) was proposed a detailed delay SSA (from which the delayed SSA was later on developed) which allows explicit delays on the protein production. A similar algorithm was independently proposed in (Barrio et al, 2006). Both works present examples of gene regulatory networks where delays in proteins' production play a relevant role in the dynamics.

The algorithm proposed in (Roussel and Zhu, 2006) differs from the mentioned algorithms, in that it can handle more than one delayed generating event for one reacting event. Because the time delay on the promoter release has been shown to be as crucial in gene networks' dynamics as the proteins' delays (namely, it affects the toggling frequency of the toggle switch (Ribeiro, 2007)), we opted for using here the delayed SSA (Roussel and Zhu, 2006).

Model of the Toggle Switch

The model of toggle switch consists of reactions (1) to (6) (Ribeiro, 2008), where $i = 1, 2$ (when only the index i is present), or $i, j = 1, 2$ with $i \neq j$ (when both indices are present):

$$\mathrm{Pr}\,o_i + Rp \xrightarrow{k_t} \mathrm{Pr}\,o_i(\tau_1) + Rp(\tau_2) + R_i(\tau_1) \quad (1)$$

$$R_i + Rib \xrightarrow{k_{tr}} R_i(\tau_3) + Rib(t+\tau_4) + P_i(\tau_5, \tau_{5std}) \quad (2)$$

$$R_i \xrightarrow{d_{rbs}} \varnothing \quad (3)$$

$$P_i \xrightarrow{k_d} \varnothing \quad (4)$$

$$\mathrm{Pr}\,o_i + P_j \underset{kunrep}{\overset{krep}{\rightleftharpoons}} \mathrm{Pr}\,o_i P_j \quad (5)$$

$$\mathrm{Pr}\,o_i P_j \xrightarrow{k_d} \mathrm{Pr}\,o_i \quad (6)$$

Gene expression is modeled by multiple time-delayed reactions for transcription (1) and translation (2), where $\mathrm{Pro_i}$ is the promoter of gene i, Rp is an RNA polymerase, Rib is a ribosome, and $\mathrm{R_i}$ is the ribosome binding site of each RNA. The delays (τ_1 to τ_5) account for the duration of the processes in transcription and translation. When a product X has a delay τ, represented by $\mathrm{X}(\tau)$, it implies that when the reaction occurs, it takes τ seconds after that for X to appear in the system.

Reaction (2) for translation accounts for the variability of the time needed to complete a functional protein (translation, folding, activation, etc.) given that the delay of $\mathrm{P_i}$ follows a normal distribution, with a mean of τ_5 and a standard deviation of $\tau_{5,\mathrm{std}}$ (we choose the normal distribution since the distribution has not yet been experimentally assessed, only mean and variance have) (Zhu et al., 2007).

Each protein $\mathrm{P_i}$ represses the other gene's promoter. Reactions (5) model binding and unbinding of the repressor protein to the other gene's promoter, which defines the

toggle switch. Reactions (4) and (6) model the decay and degradation of proteins, while reaction (3) models decay and degradation of RNAs.

The rates (in s^{-1}) of reactions (1) to (6) are $k_t = 0.005$, $k_{tr} = 0.00042$, $d_{rbs} = 0.01$, $k_{rep} = 1$, $k_{unrep} = 0.1$, and $k_d = 0.0012$. Time delays (in seconds) are $\tau_1 = 40$, $\tau_2 = 90$, $\tau_3 = 2$, $\tau_4 = 58$, $\tau_5 = 420$, and $\tau_{5std} = 140$. The transcription rate k_t and the protein decay rate k_d are varied as discussed ahead. Each 'cell' is initialized with $P_1 = 0$, $P_2 = 0$, $R_1 = 0$, and $R_2 = 0$, and with one promoter of each gene ($Pro_1 = 1$, $Pro_2 = 1$), 40 RNA polymerases ($R_p = 40$), and 100 ribosomes (Rib = 100).

Values for the delays where extracted from measurements. One of the most detailed measurements of gene expression to date has been presented in (Yu et al., 2006), thus most values were here derived from this experiment. The length of the gene tsr-venus driven by a Lac promoter studied in *E. coli* is ~2500 nucleotides (Yu et al., 2006). Since the average rate of transcriptional elongation in *E. coli* is approximately 50 nt/s, it follows that τ_2 is, on average, 90 s ($\tau_2 = \tau_1 + 2500 / 50$). The post-translational protein assembly process was observed to take 420±140 s in (Yu et al, 2006), thus τ_5 and $\tau_{5,std}$ were set in accordance. The time of the RBS clearance in translation initiation, τ_3, is set to 2 s, according to experimental estimations (Draper, 1996). The average translation rate is 15 amino acids /s, thus $\tau_4 = \tau_3 + 2500$ nt/(45 nt/s) = 58 s. Finally, the transcription initiation process includes the transition from the closed promoter complex to the open promoter complex (accounted by τ_1). We set the value of τ_1 to 40 s, within the measured range from 10 s to several minutes for the lac promoter (McClure 1980).

All delays, except τ_5, were assumed constant (Zhu et al, 2007) since experimental measurements suggest that they are not highly variable between transcription events of a single gene (although vary widely from gene to gene) (Herbert et al, 2006)(McClure, 1980)(Yu et al, 2006)(Zhu et al, 2007). For example, regarding the promoter delay τ_1,

although it varies from one transcription event to another, measurements on an unrepressed Lac promoter (Lutz et al, 2001) suggest that it follows a Gaussian distribution with a mean of 40 s and std. of 4 s, but this varies widely from gene to gene (Ross and Gourse, 2009). We tested for the present case and in previous works (Ribeiro, 2007)(Ribeiro, 2008) the effects of introducing variable delays following a normal distribution with the same mean values and with standard deviations small compared to the mean values, and the dynamics of the toggle switch was not significantly different to the dynamics of model with fixed delays, i.e., qualitatively, our conclusions hold when using a model with variable delays.

Model of differentiation pathway choice

In the simplified model of cell differentiation here used, both the moments at which a cell commits to differentiate and the choice of differentiation pathway are stochastic. It is assumed that the choice of pathway is internally driven, i.e., not determined by external signals. Namely, it depends solely on the amounts of proteins of the toggle switch present at the moment of differentiation (reactions 8 to 11). The mechanism determining when differentiation occurs fits both the assumption of externally and internally induced differentiation (reaction 7) since it is regulated by the appearance of a protein X in the cell, which could be internally produced or be an external signal.

In order to have stochasticity in these two processes we model them via simple chemical reactions that are, as the previous ones, dynamically controlled by the delayed SSA, thus their occurrence are probabilistic events with a given propensity (Gillespie, 1977). It's noted that the cell is only allowed to differentiate once, and that this process is irreversible. The cell commits to differentiate after the protein 'X' becomes present in the cell, via reaction (7). Its rate, k_x, is set to 5.10^{-6} s^{-1} implying that,

on average, X will appear in the cell around t = 500.000 s (half the cell lifetime, here set to 10^7 s, as justified ahead):

$$\varnothing \xrightarrow{k_x} X \quad (7)$$

Reaction (7) can only occur once, at most, during a cell's lifetime. Once it occurs and X becomes present in the cell, reactions (8) to (11) can then occur. These need X as substrate, thus, are competing with one another, and only one can ever occur in a cell, since there is only, at most, one molecule X (k_{dif} is set to $10^{-5}s^{-1}$):

$$X + {*P_1} + cell_0 \xrightarrow{k_{dif}} cell_1 \quad (8)$$

$$X + {*P_2} + cell_0 \xrightarrow{k_{dif}} cell_2 \quad (9)$$

$$X + {*P_1} + {*P_2} + cell_0 \xrightarrow{k_{dif}} cell_3 \quad (10)$$

$$X + cell_0 \xrightarrow{k_{dif}} cell_4 \quad (11)$$

In these reactions, the notation "*P" indicates that substance P is not consumed in the reaction, acting only as an indirect activator. The need for the substrate "$cell_0$" in reactions (8) to (11) guarantees that only one of these four reactions will ever occur in a cell, since we initialize each cell with a single 'molecule' $cell_0$. The presence of "cell0" in the cell indicates that the cell is still undifferentiated.

From reactions (8) to (11) it is apparent how the amounts of P_1 and P_2 in a cell, at the moment of differentiating, will stochastically determine the differentiation pathway ($cell_i$, i =1,..,4, represent the possible four cell types that the initial cell can differentiate into). For example, when these reactions become active, the cell will most likely differentiate into cell type 1 (reaction 8) if only P_1 is present (it is possible to also differentiate into type 4 but with lower propensity). If both P_1 and P_2 are present with significant amount, the cell will probably differentiate into cell type 3 (reaction 10) and, if both proteins are absent, into cell type 4 (reaction 11).

Cell type 4 ($cell_4$) can, instead of being considered as a differentiated cell, be seen as an undifferentiated cell, i.e., a cell of the initial population that did not undergo any of the other three differentiation pathways, due to the absence of both transcription factors that activate the differentiation process. From this point onwards we refer to cells of type 4 as undifferentiated cells.

Figure 1 is a schematic representation of the stochastic genetic toggle switch, along with the four different differentiation pathways that can be chosen, depending on the proteins' levels of the mother cell during differentiation.

FIGURE 1

Given this model of differentiation, we next study how, starting with a population of cells genetically and phenotypically identical, the final resulting distribution of differentiated cells varies with changes in the internal properties of the toggle switch.

III. RESULTS AND DISCUSSION

To study the effects on the dynamics of single cells, and its consequences on the population's distribution of differentiation pathway choices, of varying a given internal parameter, we simulate in each case 9 cell populations with distinct parameter values. Each cell population consists of 1000 identical cells to provide enough sampling. This amount of cells for each population is sufficient for the obtained distributions to vary less than 5% between different simulations, in number of cells choosing each pathway.

All simulations are done by simulators SGNSim (Ribeiro and Lloyd-Price, 2007) and CellLine (Ribeiro et al, 2007). Besides the distribution of differentiated cells, we measure proteins' mean level, noise in expression levels, and toggling frequency of the toggle switch. In all cases, a cell's lifetime is set to 10^7 s and a cell's state is measured every 10^4 s. The unrealistically long lifetime allows better statistics without affecting

the results qualitatively (alternatively one could follow the dynamics of cell lines for several generations).

An initial transient of 10^4 s is discarded since cells are initialized without proteins. A cell's mean level of proteins is measured as the average value of (P_1+P_2) during its lifetime. The mean proteins' level of a cell population is the average of the means of each of its cells. The noise level of the proteins' time series is measured by the standard deviation over the mean of the time series of (P_1+P_2). The noise level of a cell population is the average of the noise of its cells. Finally, the toggling frequency of the toggle switch is measured from the time series of the two proteins, P_1 and P_2. We count a "toggling event" when, from moment *t* to moment (*t+1*) in the time series, the relationship between the amounts of P_1 and P_2 invert and, one has either P_1 or P_2 amounts (or both) changing more than 50% of its maximum amplitude observed, i.e., ($|P_i(t)-P_i(t+1)| > 0.5*P_{i,max}$), where i=1,2 and $P_{i,max}$, is the maximum value observed in the time series of $P_{i,max}$. This number of events is then divided by the number of time steps in the time series to obtain the frequency of toggling. This definition of toggling proved to be reliable, not mistaking toggling events of the toggle switch with stochastic fluctuations of a protein level around a mean value when on a 'noisy attractor'.

Next, we analyze the effects of varying proteins' mean levels, varying noise while maintaining the same mean levels, and varying the bias in the proteins' mean levels, separately, in the dynamics of single cells and in the cell populations' distribution of differentiation pathway choices.

Effects of varying the proteins mean levels

To vary the proteins' mean level we vary the proteins' decay rate, k_d, by multiplying it by a specific value, which is different for cells of each of the nine populations. The original k_d of proteins P_1 and P_2 is multiplied by the 0.1, 0.2, 0.3, 0.4,

0.5, 0.6, 0.7, 0.8, and 0.9 in cells of populations 1 to 9, respectively. These values for the multiplicative factors were found empirically to alter significantly the toggling dynamics of the switch (Ribeiro, 2008). Note that the smaller the multiplicative factor, the smaller is the decay rate and, thus, the higher is the increase of proteins' mean levels. There are other means by which the proteins' mean level can be varied, e.g., by varying the transcription initiation rate. Here we opted for varying the degradation rates since its value has a more linear relationship with the proteins' mean levels than the transcription rate, for example. As mentioned, proteins degradation rates can be, to some extent, controlled by cells (Cooper, 2000) thus, we consider it here to be a parameter that cells can tune to alter proteins' mean levels.

As shown in FIG 2, (P_1+P_2) varies nonlinearly with a linear variation of k_d, since transcription (1) and translation (2) are time-delayed reactions.

FIGURE 2

Interestingly, the differences in mean protein levels do not cause significant differences in proteins' noise level, which is similar in cells of all populations (FIG 3). To explain this we show in FIG 4 examples of the time series of proteins' levels in single cells of two cell populations, one from population 1 (cell A) and the other from population 9 (cell B). It is apparent why the noise levels are identical in all populations. In cell A (pop. 1), (P1+P2) mean level is ~500 with ~600 standard deviation, resulting in the observed noise level of ~1.2 (FIG 3A), while in cell B (pop. 9) the mean of (P1+P2) is ~50 and the standard deviation is ~60, resulting in an identical noise level. From pop. 1 to 9, the decrease in (P1+P2) mean is 'compensated' by a decrease in the standard deviation, causing the noise level to be identical in cells of all populations.

FIGURE 3

FIGURE 4

However, the differences in proteins mean levels do affect the cells' dynamics, i.e., the toggling frequency of the switch (FIG 5). Small decreases in proteins' mean levels cause a significant increase in the toggling frequency. This is somehow surprising since the switches in the switch are noise driven and the noise level is similar in cells of all populations (thus, one would expect similar toggling frequencies). The difference in toggling frequencies is caused by differing proteins' mean levels, which determine the average fraction of time that promoters are repressed. Higher mean levels cause longer occupation intervals of the promoters by their repressors, resulting in less switches.

FIGURE 5

FIGURE 6

Given the differences in the dynamics of cells of each population, we now observe the consequences in their differentiation process (FIG 6). The results show that the composition of the differentiated cell populations of all nine cell populations are very similar, thus one can conclude that changes in proteins' mean levels and toggling frequency alone do not affect the differentiation process at a population level.

The only significant difference in the differentiation patterns is that populations 1 and 2 have a higher number of cells differentiated into cell type 3 (the cell type more commonly chosen when both proteins are present in significant amounts). This is explained by the results shown in FIG 2. While not easily apparent, it is noted that in cells of populations 1 and 2, the protein level of the repressed gene is significantly higher (~10-50 proteins) than that of other populations (~1-10 proteins). Cells of populations 1 and 2 actually have the promoters of the repressed gene more strongly repressed than the others, however, their proteins' decay rate is sufficiently lower for proteins of the repressed gene to remain present for much longer periods of time, thus

explaining how the number of cells of these two populations that differentiate into cell type 3 is larger than in the other populations.

The appearance of cells of type 4 (assumed as undifferentiated cells) is rare (~0.1%) but identical in all populations, which further confirms that the toggling frequency does not affect significantly the distribution of differentiation pathway choice and that simulating 1000 cells per cell population provides sufficient sampling.

Effects of varying proteins' time series noise level while maintaining mean levels

We now test the effects of changing the noise level of the toggle switch. It was shown that the noise of this model of switch can be varied by multiplying the transcription rate k_t and decay rate k_d with a pair of factors, while maintaining approximately constant the mean level of (P_1+P_2) (Ribeiro, 2008). Using the same procedure, we study how the differentiation pathway choice, at the population level, depends on varying the toggle switch's noise level alone.

It's noted that the changes in k_t and k_d, while not affecting mean protein levels, do affect the dynamics of the toggle switch significantly. Until a certain noise level only the switching frequency increases but, beyond that (pop. 6 and beyond), the two genes start expressing simultaneously (when the transcription reactions (1) have higher propensity than the reactions causing the promoters' repression (5)) (Ribeiro, 2007).

To obtain cells with identical mean proteins' levels but distinct noise levels, the cells are altered by multiplying k_t and k_d with a pair of factors (v_i, w_i), where $i \in \{1,\ldots 9\}$, respectively, for each of the nine cell populations (originally, $k_t=10^{-2}$ and $k_d=10^{-3}$). These pairs of values cannot be varied by constant amounts each time to attain equal mean proteins' levels due to the delay τ_1 at each transcription event, which accounts for the promoter occupancy time by an RNA polymerase (McClure, 1980).

We first set the v's, where v = 1 is the 'baseline' and four inverse pairs are arranged on each side in ascending order; then w's are searched by a heuristic method so that the mean ($P_1 + P_2$) doesn't differ beyond 4% among cells of different populations. The nine pairs of multiplicative values of k_t and k_d are (0.01, 0.28), (0.1, 1.09), (0.25, 1.25), (0.5, 1.2), (1, 1), (2, 0.92), (4, 1.08), (10, 1.73), and (100, 3.46), respectively (Dai et al, 2009), with the average noise levels shown in FIG 7. Empirically, these pairs of values were found to allow significant variation in the noise of the proteins levels of the switch.

FIGURE 7

FIGURE 8

FIG. 8 shows the distribution of the choices for differentiation pathways of cells of the nine populations. The distinct dynamics of the toggle switch in cells of the various populations results in distinct distributions of differentiated cells. Cells of populations 1 to 3 are bistable and thus, the two most common differentiation pathways are towards cell types 1 and 2. Cells of populations 4 and 5 have three differentiation pathways almost equally probable (pathways 1, 2 and 3). In populations 6 to 9, the most common noisy attractor that these cells are on is the unstable one, i.e., 'both genes are on', due to the loss of bistability (Ribeiro, 2008) thus, the most common cell type found after differentiation is cell type 3. The time series of a cell of each of the populations is shown in FIG. 9, illustrating how the dynamics of the toggle switch changes with the change in noise, explaining the differences in differentiation distributions.

FIGURE 9

The cells with higher noise levels are those that produce the most uniformly distributed populations of differentiated cells, expressing the significant role of noise in the differentiation process. The noisier the toggle switch, the less deterministic is the

choice of differentiation pathway. Regulating the noise level (e.g., by tuning the production and degradation rates of proteins) is a viable method to gradually tune the fractions of cells that chose between differentiating into cell types 1 or 2 and into cell type 3.

Effects of biasing proteins' levels at each noisy attractor

We now test the effects of biasing the amounts of P_1 versus P_2 in the process of cell differentiation. For this, given the original cell model, we set the degradation rates of P_1 and P_2 to different values to bias their expected concentrations. This is done maintaining the same noise level, while allowing variation in proteins' mean levels.

The multiplicative factor pairs of k_d for P_1 and P_2 are (0.10, 0.107), (0.2, 0.209), (0.3, 0.314), (0.4, 0.431), (0.5, 0.535), (0.6, 0.647), (0.7, 0.759), (0.8, 0.89), (0.9, 1), respectively, for cell populations 1 to 9. These values for the multiplicative factors were found empirically to alter significantly the differentiation patterns, while still being within realistic intervals of parameter values. For comparison purposes, we also show a 10^{th} population whose proteins' decays is unbiased (FIG. 10). Note that the difference (bias) between the two decay rates increases from pop. 1 to 9 (FIG. 14).

FIGURE 10

As a result of setting identical noise levels in cells of all nine populations, their mean protein levels (P_1+P_2) differs (FIG. 11), which causes the toggling frequency of cells of different cell populations to differ, as seen in FIG. 12.

FIGURE 11

FIGURE 12

However, unlike the first case (of varying the proteins' mean level) where the noise level is uniform over all populations and the toggling frequency differs, resulting in similar distributions of differentiation pathway choices for the nine cell populations, in

the present case, although the noise is still uniformly distributed across all cell populations, the differentiation choices differ more pronouncedly (FIG. 13) due to a combination of two effects. First, there is a bias towards differentiating into cell type 1 in comparison to type 2 (type 1 is more probable since k_d of P_1 is smaller than k_d of P_2 in all populations) and, second, the amount of cells that undergo pathway 3 varies significantly across the populations (due to the differences in mean proteins' levels).

Regarding the fraction of cells undergoing pathway 3, the variation across the populations is similar to the first case studied, i.e., as the mean protein level decreases, the number of cells choosing this pathway decreases (FIG 13). This, as previously, is only clear in the first two cell populations, while the other populations do not differ significantly among each other.

To measure the effects of the bias in decay rates on choosing between pathways 1 and 2, we compare the ratio $(N_1-N_2)/N_{12}$ of each cell population, where N_1 and N_2 are the number of cells that differentiated into cell types 1 and 2, respectively, and $N_{12} = N_1+N_2$. This quantity accounts for the variation in the number of cells that are available to choose pathways 1 and 2 among the various cell populations, thus allowing a more precise comparison of the effect of different biases in proteins' decay rates. From FIG. 14, that shows both the bias in decay rates and the bias in choosing between pathway 1 or 2, one can confirm that the bigger the bias (k_{d1}-k_{d2}), the bigger the ratio $(N_1-N_2)/N_{12}$.

FIGURE 13

FIGURE 14

As before, the fraction of undifferentiated cells (pathway 4) is very small and similar in all populations, based on which one can conclude that biasing proteins' amounts doesn't change the fraction of cells opting for differentiating given the assumed control mechanism of differentiation pathway selection.

IV. CONCLUSIONS

One genetic circuit used by some cells for differentiation pathway selection is the toggle switch (Huang et al., 2005)(Wang et al, 2009). Here we investigated how changes in the dynamics of this genetic circuit affect the distribution of the stochastic choices of differentiation pathways of a population of genetically and, initially, phenotypically identical cells.

Given our stochastic model of cell differentiation, it is shown that the distribution of the differentiation pathway choices can be significantly altered by phenotypic changes in the cells, i.e., proteins' degradation rates and gene expression activities, thus suggesting that a differentiation mechanism based on a stochastic toggle switch can be 'easily' tuned to fit survival needs. Importantly, unlike previous models such as Boolean networks, the delayed stochastic model of gene regulatory networks has much greater plasticity as seen in how the distribution of differentiation pathway choice gradually changed due to gradual changes in some of the internal parameters of the toggle switch. In some cases we observed also that small changes in the proteins' levels dynamics of the switch caused significant changes in the switch's dynamics.

The results suggest, for example, that if some steps in neutrophil differentiation are regulated by bistable gene networks, e.g., toggle switches (Huang et al, 2005)(Chang et al, 2006)(Chang et al, 2008), the regulation of k_t and k_d would provide a mechanism to bias the differentiation pathway choices along the cell lineage according to the needs of the organism at each moment, which is of significant importance in the adaptability of the organism to the environmental stress caused by pathogens or other external stimuli.

It is noted as well that if the noise level of the toggle switch can be altered by interaction with other proteins (e.g. signaling stress) in the cell, or external factors, then

our results suggest how the differentiation patterns can change depending on external conditions, and not necessarily by mutation as assumed here.

Understanding how the dynamics of stochastic differentiation pathway determination determines the distributions of differentiated cell populations is needed to better understand the degree of plasticity of multi- and uni-cellular organisms in their adaptation to the environment. There are already known cases where plasticity is mostly conferred by stochastic mechanisms, such as in neutrophil differentiation (Huang et al., 2005), and there also evidences in bacterial populations of a correlation between phenotypic diversity and environmental stress and/or variability (Acar et al., 2008)(Suel et al, 2006)(Suel et al, 2007)(Samoilov et al, 2006).

We have shown that cells whose differentiation pathway determination is regulated by a bistable switch could be tuned to differentiate into cell types 1 to 3 in different ratios by tuning transcription rates and protein degradation rates. However, it is also worthwhile discussing how one could increase the fraction of cells that undergo the fourth possible differentiation pathway in the present model. Here, we have not explicitly shown such a case since it is assumed that this 'cell type' corresponds to the fraction of undifferentiated cells, since it is the case where both proteins of the switch are virtually absent in the cell during differentiation. Let us assume, instead, that the absence of both proteins would lead to a distinct cell type. To increase the fraction of cells opting for this pathway one could, e.g., decrease the proteins' degradation rate when, and only when, they are bound to the promoters (reaction 6). Assuming a sufficiently high repression strength, k_{rep}, such that $k_{rep} > k_d$ (as is the case here) and assuming that when the protein is bound to the promoter (forming the complex Pro_jP_i), it becomes "protected" from degradation (which could be achieved by decreasing the rate of reaction (6) alone), then even with very low amounts of P_1 and P_2 present in the

cell, both genes would remain repressed most of the time. This would therefore cause an increase in the fraction of cells undergoing the differentiation pathway corresponding to having both proteins absent.

We note that we did not vary all parameters and features of the toggle switch to our disposal, and we focused on two of the parameters known to vary from cell to cell in genetically identical cell populations. In the future, it would be of interest to study the consequences of varying the lengths of the time delays in transcription or translation, which are sequence dependent, such as the promoter time delay that is known to have significant effects on the dynamics of the toggle switch (Ribeiro, 2007). In eukaryotes, one could also study the potential role of methylation to this end. Here we focused on changes in the toggle switch that do not, necessarily, require mutations. It would also be of interest to analyze further the plasticity of stochastic genetic networks, such as the toggle switch, as regulators of differentiation pathway choice, especially within the context of cells' adaptability to environmental changes in long time scales.

ACKNOWLEDGMENTS

This work was supported by Tampere Graduate School in Information Sci. and Eng. (TISE) (XD), Academy of Finland, project no. 213462 (Finnish Centre of Excellence program 2006-11) (OY) and project no. 126803 (ASR), and the FiDiPro program of the Finnish Funding Agency for Technology and Innovation (40284/08) (ASR)(XD)(OY).

REFERENCES

Acar, M., Mettetal, J., and van Oudenaarden, A. , (2008), Stochastic switching as a survival strategy in fluctuating environments, Nature Genetics 40(4) 471-5.

Aldana, M., Coppersmith, S., and Kadanoff, L. , (2003) Boolean Dynamics with Random Couplings, in Perspectives and Problems in Nonlinear Science, Kaplan E., Marsden J.E., Sreenivasan K.R. eds., Springer Applied Mathematical Sciences Series.

Arkin, A., Ross, J., and McAdams, H., (1998) Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells., Genetics 149, 1633-1648.

Barrio, M., Burrage, K., Leier, A., Tian, T., (2006) Oscillatory Regulation of Hes1: Discrete Stochastic Delay Modelling and Simulation. PLoS Comput Biol 2(9): e117.

Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S., and Cohen, S. N. (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. PNAS 99, 9697-702.

Blake, W.J., Balázsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006) Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. Molecular Cell 24, 853-865.

Bon, M., McGowan, S. J., Cook, P. R. (2006). Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. FASEB J. 20: 1721-1723.

Bruno L, Hoffmann R, McBlane F, Brown J, Gupta R, Joshi C, Pearson S, Seidl T, Heyworth C, and Enver T. (2004) 56 Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro. Mol Cell Biol 24(2), 741-756

Bratsun, D., Volfson, D., Tsimring, L.S., Hasty, J., 2005. Delay-induced stochastic oscillations in gene regulation. Proc. Natl. Acad. Sci. USA 102, 14593-14598

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S., (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453(7194), 544-7.

Chang, H.H., Oh, P.Y., Ingber, D.E., Huang, S., (2006) Multistable and multistep dynamics in neutrophil differentiation. BMC Cell Biol. 7(1):11.

Cooper, G.M., (2000) "The Cell: A Molecular Approach", 2nd Ed., Sinauer Associates, Inc., Sunderland, Massachusetts USA.

Dai, X., Yli-Harja, O., and Ribeiro, A.S., (2009) "Determining the noisy attractors of a delayed

stochastic Toggle Switch from multiple data sources", Bioinformatics (submitted).

Draper, D.E., (1996). In Escherichia coli and Salmonella. ASM Press, Washington, DC, pp. 902–908.

Elowitz M.B., Levine A.J., Siggia E.D., and Swain, P.S. (2002) Stochastic gene expression in a single cell. Science 297(5584), 1183-1186.

Gardner, T., Cantor, C., and Collins, J., (2000) Construction of a genetic toggle switch in Escherichia coli, Nature 403, 339-342.

Gillespie, D.T., (1977) Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81, 2340-2361.

Golding, I., Johan, Paulsson, J., Zawilski, S.M., and Cox, E.C., (2005) Real-Time Kinetics of Gene Activity in Individual Bacteria. Cell 123: 1025–1036

Herbert K., La Porta A., Wong B., Mooney R., Neuman K., Landick R, Block S. (2006) Sequence resolved detection of pausing by single RNA polymerase molecules. Cell 125(6): 1083-94.

Huang, S., Eichler, G., Bar-Yam, Y., Ingber, D., (2005) Cell Fates as a High-Dimensional Attractor States of a Complex Gene Regulatory Network, Phys. Rev. Lett. 94, 128701.

Heyworth, C., Gale K., Dexter M., May G., and Enver, T. (1999) A GATA-2/estrogen receptor chimera functions as a ligand-dependent negative regulator of self-renewal. *Genes Dev.* 13, 1847-1860.

Kaern, M., Elston, T.R., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression. *Nature Reviews Genetics* 6, 451-464.

Kauffman, S.A.. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets, J. Theor. Biol. 22, 437-467.

Kauffman, S.A., (1993) The Origins of Order, Oxford Univ. Press, New York.

R. Lutz, R., Lozinski, T., Ellinger, T., and Bujard, H. (2001) Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator," Nucleic Acids Research 29(18), 3873–81.

McClure, W.R., (1980) Rate-limiting steps in RNA chain initiation, Proc. Natl. Acad. Sci. USA 77, 5634-5638.

Monod, J. and Jacob, F., (1961) Teleonomic mechanisms in cellular metabolism, growth, and differentiation, Cold Spring Harb Symp Quant Biol 26, 389-401.

Neubauerz, Z. and Calef, E., (1970) Immunity phase-shift in defective lysogens: non-mutational hereditary change of early regulation of lambda prophage, J. Mol. Biol. 51, 1-13.

Ribeiro, A.S., Zhu, R. and Kauffman ,S.A. , (2006) A general modelling strategy for gene regulatory networks with stochastic dynamics, J. Comput. Biol. 13(9), 1630-1639.

Ribeiro, A.S. and Kauffman, S.A., (2007) Noisy attractors and ergodic sets in models of gene regulatory networks, J. Theor. Biol. 247, 743-755.

Ribeiro, A.S. and Lloyd-Price, J., (2007) SGN Sim, a Stochastic Genetic Networks Simulator, Bioinformatics 23, 777-779.

Ribeiro, A.S., Charlebois, D., and Lloyd-Price, J., (2007) CellLine, a stochastic cell lineage simulator, Bioinformatics 23(24):3409-3411.

Ribeiro, A.S., (2007) "Dynamics of a two-dimensional model of cell tissues with coupled stochastic gene networks", Phys. Review E 76(5).

Ribeiro, A.S., (2008) Dynamics and evolution of stochastic bistable gene networks with sensing in fluctuating environments, Phys. Rev. E 78(6),061902.

Ross, W. and Gourse, R.L. (2009) Analysis of RNA polymerase-promoter complex formation, Methods 47, 13-24.

Roussel ,M. and Zhu, R., (2006) Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression, Phys. Biol. 3, 274-284.

Samoilov, M., Price, G., and Arkin, A., (2006) From Fluctuations to Phenotypes: The Physiology of Noise, Science STKE 366, re17.

Shmulevich, I., Dougherty, E.R., Zhang W. (2002) "Gene Perturbation and Intervention in Probabilistic Boolean Networks," Bioinformatics 18(10), 1319-1331.

Süel, G.M., Garcia-Ojalvo, J., Liberman, L.M., and Elowitz, M.B., (2006) An excitable gene regulatory circuit induces transient cellular differentiation, Nature 440, 545-550.

Süel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J., and Elowitz, M.B., (2007) Tunability and Noise Dependence in Differentiation Dynamics, *Science* 315, 1716-1719.

Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X., (2006) Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science* 311, 1600-1603

Wang, L., Walker, B., Iannaccone, S., Bhatt, D., Kennedy, P., and Tse, W. (2009) Bistable switches control memory and plasticity in cellular differentiation. *Proc. Natl. Acad. Sci. USA* 106(16), 6638-6643.

Weinberger, L., Burnett, J., Toettcher, J., Arkin, A., and Schaffer, D., (2005) Stochastic Gene Expression

in a Lentiviral Positive-Feedback Loop: HIV-1 Tat Fluctuations Drive Phenotypic Diversity, *Cell* 122, 169-82.

Zhu, R., Ribeiro, A.S., Salahub, D., and Kauffman, S.A., (2007) Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models, *J. Theor. Biol.* 246, 725-745.

FIGURE LEGENDS

Fig. 1

Schematic figure of the stochastic toggle switch along with the four different differentiation pathways that can be chosen, depending on the levels of the proteins during differentiation. Grey balls represent genes with high expression levels, and white balls represent genes with low expression levels. In the mother cell (cell type 0), the toggling behavior of the genes' expression levels is represented by a half-white, half-grey ball.

Fig. 2

Average of (P_1+P_2) mean level (y-axis) over all cells of each of the 9 cell populations (x-axis), differing in proteins' decay rates. Protein decay increases along the x-axis.

Fig. 3

Average proteins' time series noise level (y-axis) over all cells of each of the 9 cell populations (x-axis), differing in proteins' mean levels. Protein's mean level decreases along the x-axis.

Fig. 4

Protein time series of a single cell of (A) cell population 1 and (B) cell population 9. Notice the differences in the y-axis scales.

Fig. 5

Toggling frequency of the TS (y-axis) of cells of the 9 cell populations (x-axis). Toggling frequency is calculated across all the cells in each population.

Fig. 6

Distribution of the number of cells (y-axis) that are differentiated into each of the four cell types (x-axis) of the 9 cell populations (1000 cells per population), differing in proteins' mean levels. Protein's mean level decreases along the x-axis.

Fig. 7

Average proteins' noise level (y-axis) of the 9 cell populations (x-axis), differing in proteins' mean levels. Protein's mean level decreases along the x-axis.

Fig. 8

Distribution of the amount of cells (y-axis) that are differentiated into each of the four cell types (x-axis) of the 9 cell populations (1000 cells per cell population), differing in noise levels while maintaining the same protein mean level. Transcription rate increases along the x-axis.

Fig. 9

Protein time series of a single cell chosen from each cell population. Notice the differences in scales in the Y-axis.

Fig. 10

Average noise level (y-axis) of the cells of each of the 9 cell populations (x-axis), differing in the bias in proteins decays. Population '10', which contains no bias in protein decay rates, is shown for comparison.

Fig. 11

Average (P_1+P_2) mean level (y-axis) over 1000 cells of each of the 9 cell populations (x-axis), differing in the bias of proteins' decay.

Fig. 12

Toggling frequencies (y-axis) of 1000 cells of each of the 9 cell populations (x-axis). Toggling frequency is calculated across all the cells in each population.

Fig. 13

Distribution of the amount of cells (y-axis) differentiated into each of the four cell types (x-axis) of the 9 cell populations (1000 cells per cell population), differing in the bias of protein decay rates while keeping the noise fixed. Bias and proteins' decay rate increase along the x-axis.

Fig. 14

Bias (black line, y-axis) and differentiation change (grey line, y axis) of each cell population (x-axis). Bias and differentiation change are represented as (k_{d1}–k_{d2}) and (N_1-N_2)/$N_{12,}$ respectively, where N_1 and N_2 are the number of cells that are differentiated into cell type 1 and cell type 2, and N_{12} = N_1+N_2.

Fig. 1

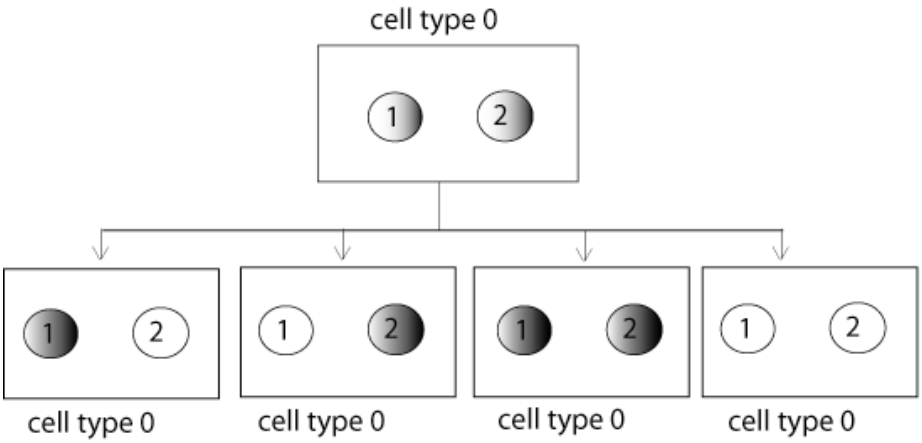

Fig. 2

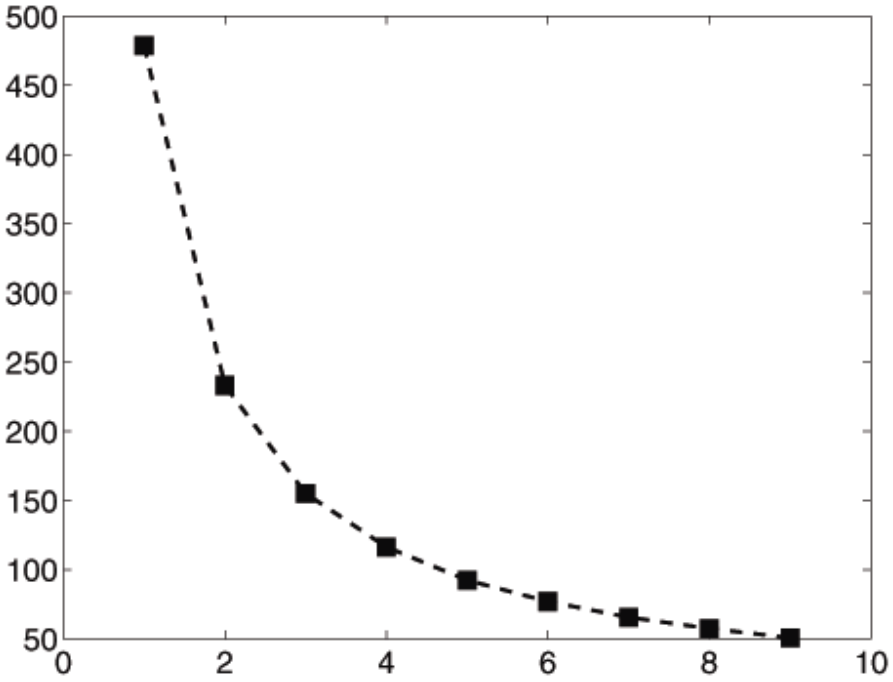

Fig. 3

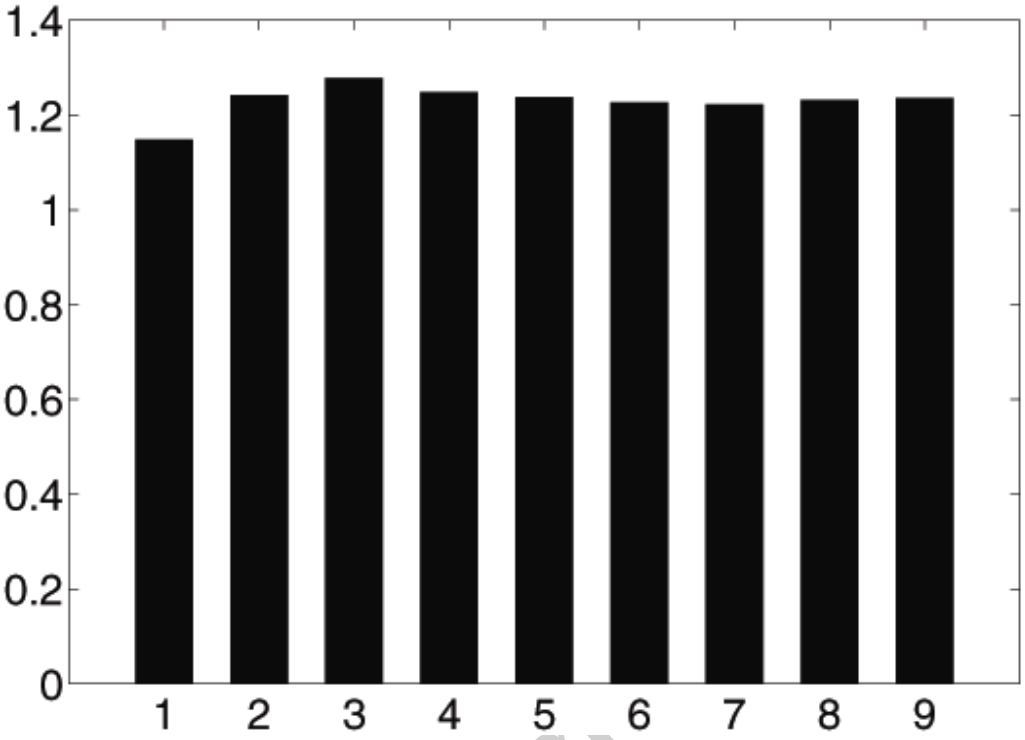

Fig. 4

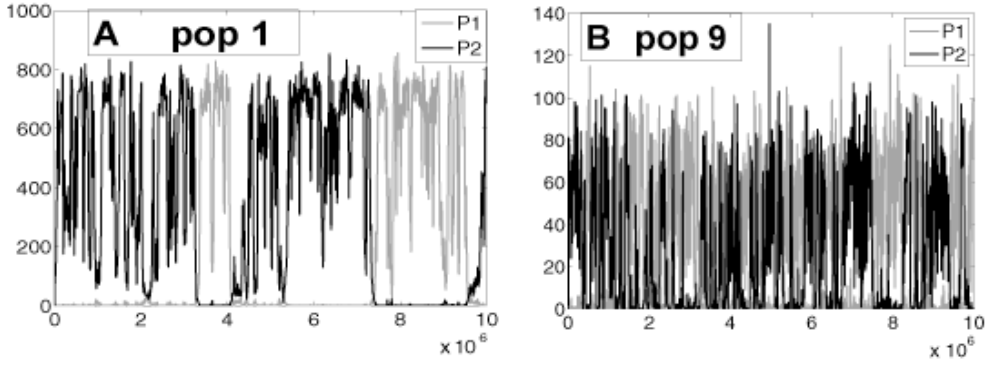

Fig. 5

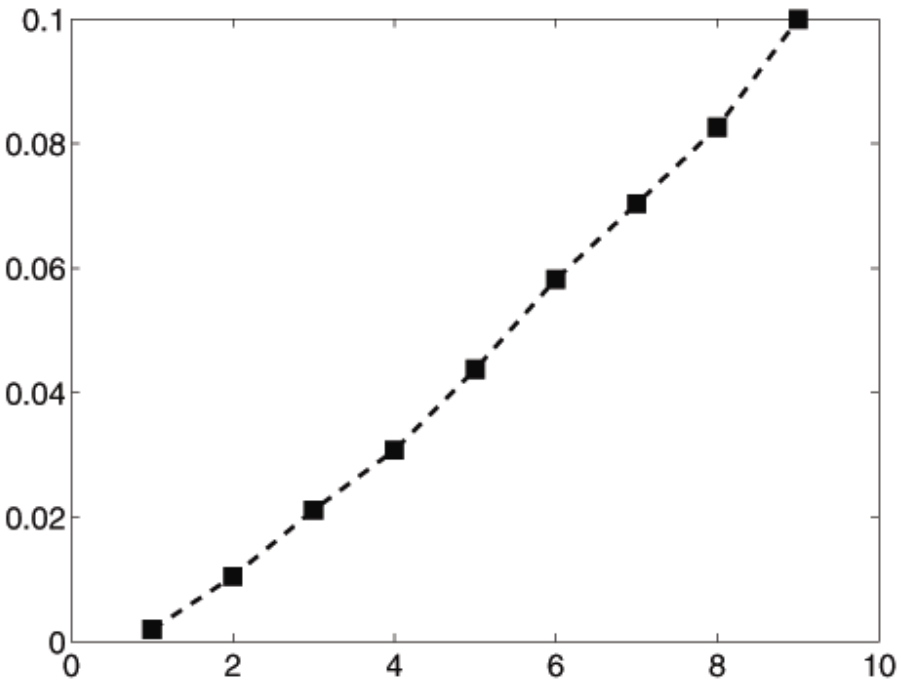

Fig. 6

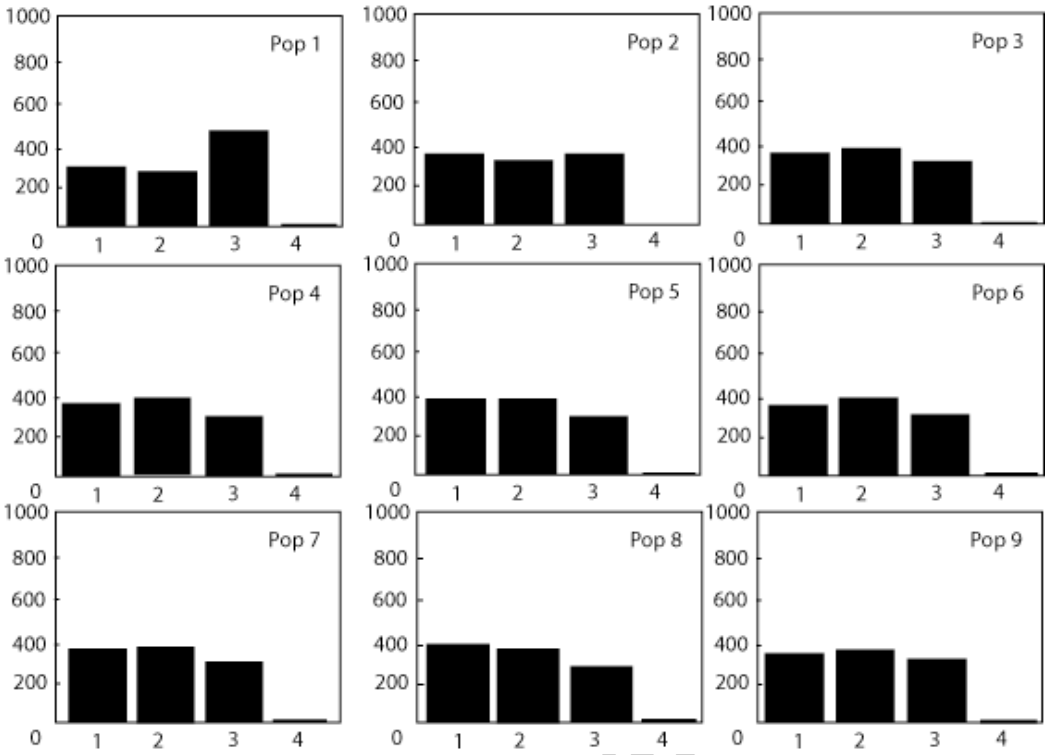

Fig. 7

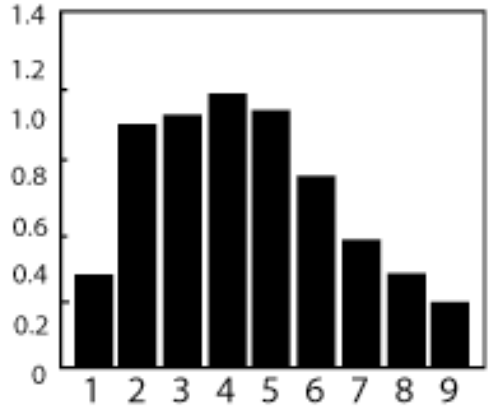

Fig. 8

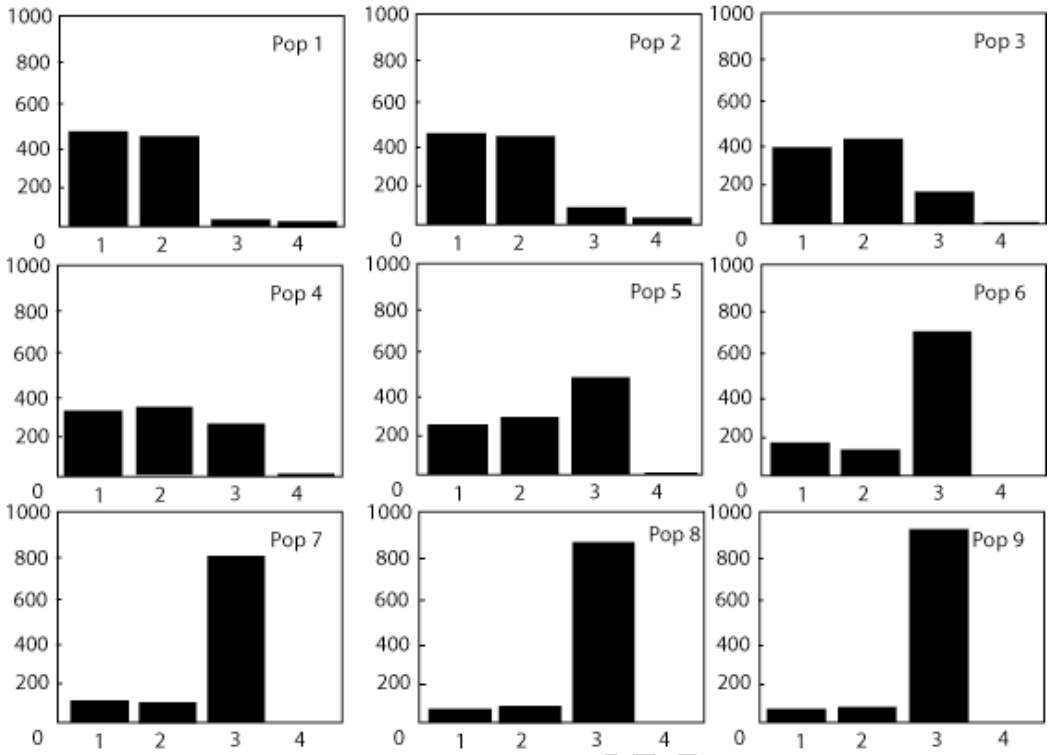

Fig. 9

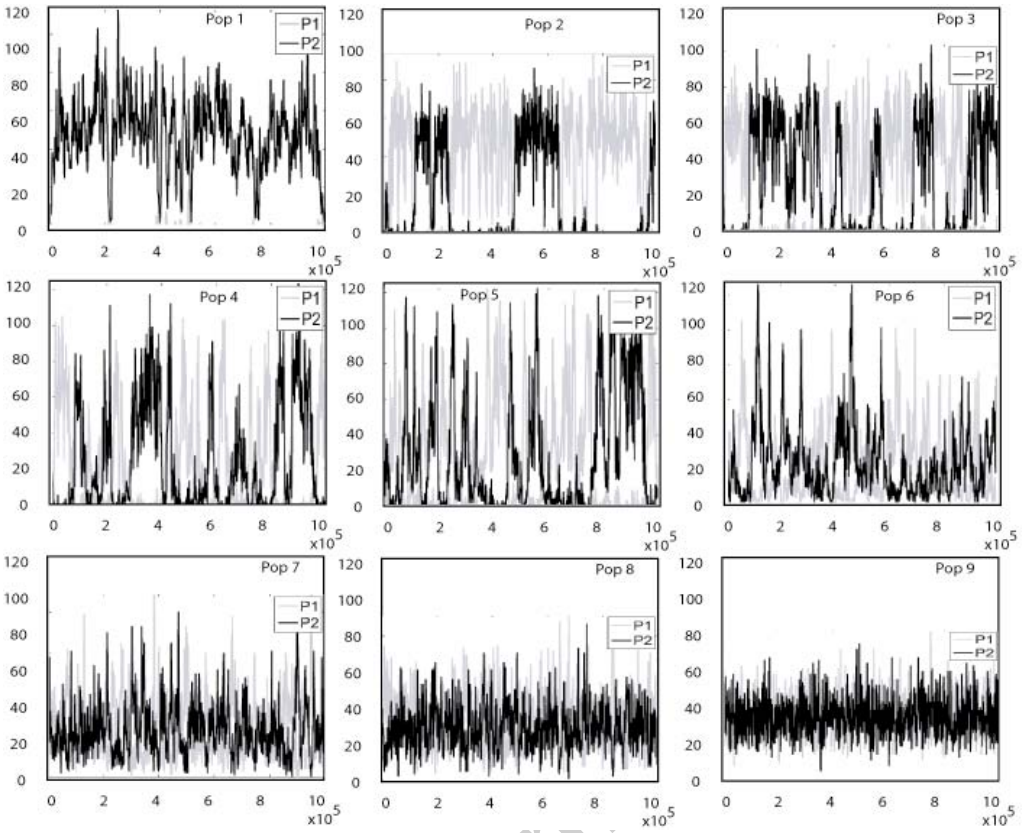

Fig. 10

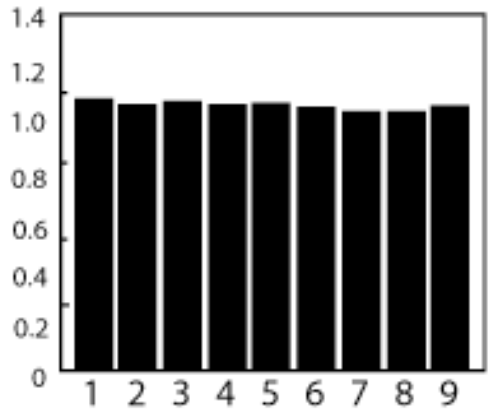

Fig. 11

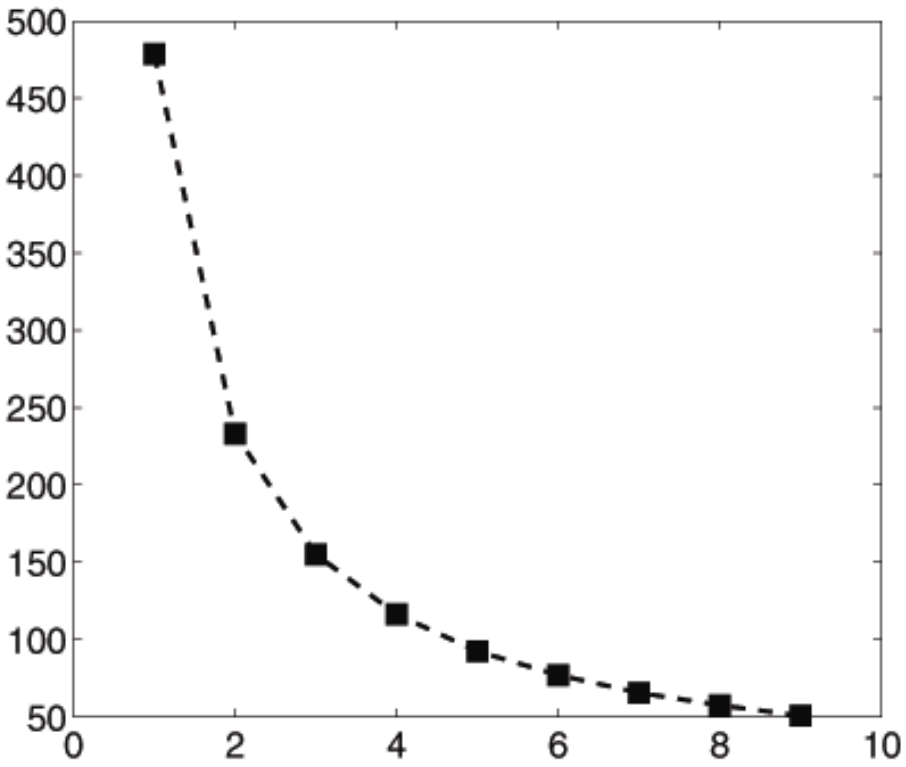

Fig. 12

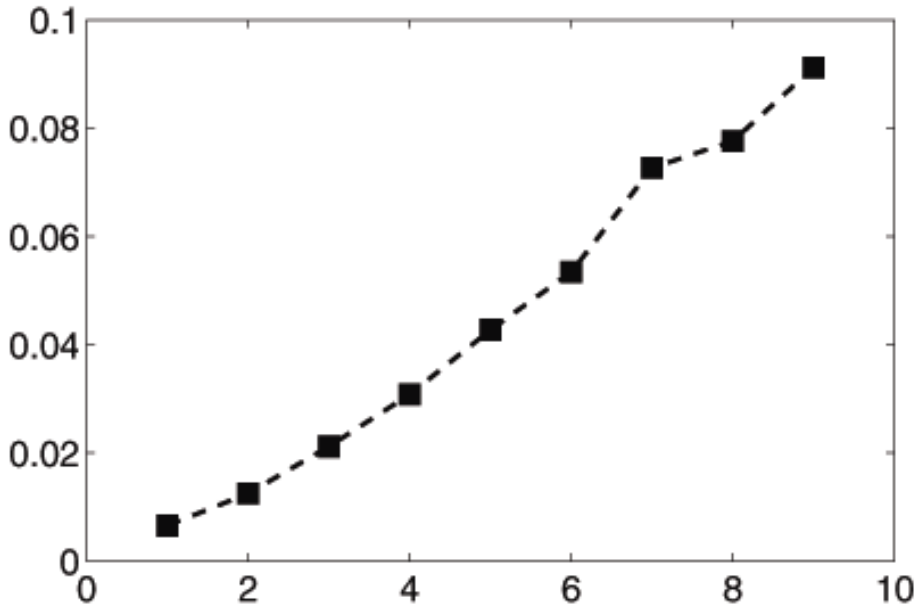

Fig. 13

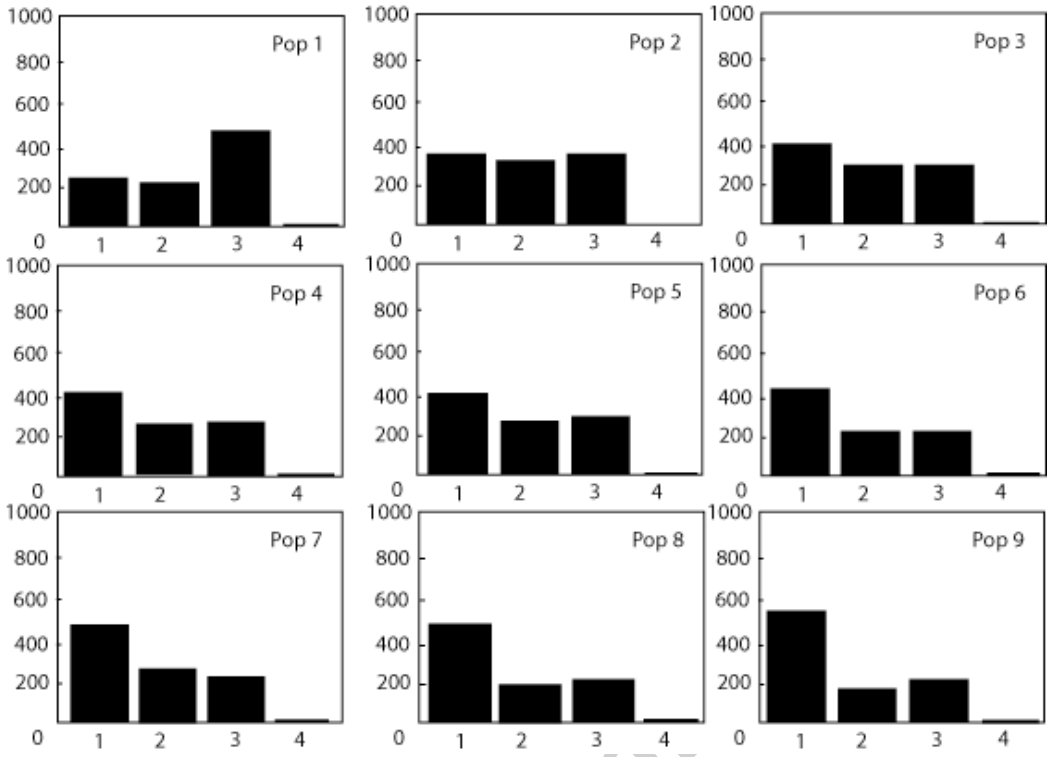

Fig. 14