

# Metabolic Variability in Micro-Populations

Yuval Elhanati<sup>1,3</sup>, Naama Brenner<sup>2,3\*</sup><sup>1</sup> Department of Physics, Technion, Haifa, Israel, <sup>2</sup> Department of Chemical Engineering, Technion, Haifa, Israel, <sup>3</sup> Network Biology Research Lab, Technion, Haifa, Israel

## Abstract

Biological cells in a population are variable in practically every property. Much is known about how variability of single cells is reflected in the statistical properties of infinitely large populations; however, many biologically relevant situations entail finite times and intermediate-sized populations. The statistical properties of an ensemble of finite populations then come into focus, raising questions concerning inter-population variability and dependence on initial conditions. Recent technologies of microfluidic and microdroplet-based population growth realize these situations and make them immediately relevant for experiments and biotechnological application. We here study the statistical properties, arising from metabolic variability of single cells, in an ensemble of micro-populations grown to saturation in a finite environment such as a micro-droplet. We develop a discrete stochastic model for this growth process, describing the possible histories as a random walk in a phenotypic space with an absorbing boundary. Using a mapping to Polya's Urn, a classic problem of probability theory, we find that distributions approach a limiting inoculum-dependent form after a large number of divisions. Thus, population size and structure are random variables whose mean, variance and in general their distribution can reflect initial conditions after many generations of growth. Implications of our results to experiments and to biotechnology are discussed.

**Citation:** Elhanati Y, Brenner N (2012) Metabolic Variability in Micro-Populations. *PLoS ONE* 7(12): e52105. doi:10.1371/journal.pone.0052105

**Editor:** Alexandre J. Kabla, University of Cambridge, United Kingdom

**Received:** July 28, 2012; **Accepted:** November 9, 2012; **Published:** December 27, 2012

**Copyright:** © 2012 Elhanati, Brenner. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Funding provided by Israeli Science Foundation personal grant 2011. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: nbrenner@technion.ac.il

## Introduction

Biological cell populations exhibit a broad distribution of phenotypes, even if genetically homogenous. Such variability has been observed in practically every single-cell phenotypic property that was measured, including cell size, protein content, division rate and more. Although this is an issue of general interest in biology, populations of microorganisms have provided important model systems for developing novel experimental single-cell techniques and applying theoretical analysis of the problem. Some of the major topics of recent work have been understanding the mechanisms underlying phenotypic variability [1–5], modeling the distribution of phenotypes in populations of cells [6–8], and revealing the role of phenotypic variability in the fate and survival of microbial populations [9–14].

In the standard approach to the problem, from single cell properties that have some stochastic components emerge the statistical properties of a large (practically infinite) population. Within this approach, variability in division rate seems to be a most important property for population dynamics, and its implications have been widely studied, especially in the context of fluctuating environments [15–19].

However, a population of growing and dividing cells is not a statistical ensemble of independent individuals. The environment always plays an important role in real population growth; even without direct interactions, as cells grow they continually modify their environment and thus indirectly affect other population members [20,21]. These considerations become even more important in many biologically relevant situations where growth is physically or biochemically constrained by the environment and the limit of an infinite population is never reached. Inheritance

and memory in the population can then cause a delicate dependence on initial conditions that can be more important over finite growth times than the asymptotic dynamics.

When the environment is taken into account, another single-cell property becomes important in addition to the division rate - the yield with respect to growth resources. For example, in growth under flow in a well-mixed environment, such as that described by an ideal chemostat, although the fastest-growing cells take over, the yield determines the population density at equilibrium [22]. Spatial structure in the environment can result in coexistence of high-yield and high-growth-rate phenotypes [23], whereas models for density-dependent selection show that high carrying capacity, which is analogous to high yield, can be selected under some conditions [24,25]. The picture that emerges from these works is that the role of yield with respect to growth resource is strongly dependent on the details of the environment and how exactly it constrains the population.

Recent advances in microfluidic and microdroplet technology have enabled to grow “micro-populations”, intermediate-sized populations that grow over intermediate lengths of time [26–30]. These experimental systems enable to follow the temporal dynamics of the populations and to accumulate statistics of a large number of such populations with well controlled initial and environmental conditions. While obviously promising many important practical applications, these advances at the same time define a new regime for studying cell populations and highlight some questions of fundamental nature: What is the variability between different micro-populations? What is the dependence on initial conditions and how does it decay with time and system size? One is led to consider the statistical properties of an ensemble of

finite populations, rather than those of single cells in an infinite population. For example, the dependence on initial conditions in constrained micro-chemostat growth was recently studied [31]. Using a model of growth-rate variability of single cells, it was found that sometimes the effect of initial conditions persists indefinitely, while in other circumstances it can take a very long time until this effect decays.

In this work we study theoretically the effects of variability in division rate and in yield on population dynamics in limited environments such as finite droplets. In many cases of biological relevance the yield is inversely correlated with division rate at the single cell level as both vary across the population [23,32–34]. Experimental evidence supporting such a “metabolic tradeoff” was reviewed in [33] and in the Supplement of [34]. Recent theoretical work has shown that accounting for metabolic tradeoffs between these two cellular properties can give rise to nontrivial behavior at the population level, even in a deterministic model in a simple chemostat-like environment [34]. Previous work suggested that high-yield growth can be considered a cooperative strategy that models evolutionary dilemmas [23,33], while here we are interested primarily in the statistical effects and in long-term traces of initial conditions on finite populations. We develop a stochastic model with emphasis on tracking histories of individual equally prepared micro-populations to characterize their distribution at the end of growth. While this work is of basic and theoretical nature, there are implications to experimental measurements and to practical biotechnology that arise and are discussed briefly at the end.

## Model

A single metabolic network defined by its enzymes, metabolites and biochemical reactions can support many different modes of metabolic flux at steady state growth [35,36]. These modes are generally associated with different cellular properties such as division rate and yield with respect to growth resources. Considering the large variability observed in practically any physiological property that was measured at the single cell level in microorganisms, it is only natural to assume that the metabolic state, and with it the yield, are also variable properties in such a population. This assumption is supported by experiments on bacteria grown in continuous culture, showing a broad range of operational modes of the metabolic system [37]. In addition to their direct measurements, which is relatively difficult, the implications of yield variability can possibly be measured in terms of the population dynamics and statistics under constrained conditions as described below.

We consider a population of microorganisms with a range of metabolic states, each characterized by a different yield towards a growth-limiting resource. In the model the population will thus be represented as a mixture of several different metabolic states, each characterized by a division rate  $\mu_j(S)$  and a yield  $Y_j$ . A single growth-limiting resource, a nutritional substrate  $S$ , describes the environment. The state space of the entire system is characterized by the number of cells of each phenotype and the amount of resource at a given time  $(\{n_j\}, S)$ . Cells of state  $j$  can divide with probability  $\mu_j(S)$  per unit time, in which case the resource is depleted by an amount  $1/Y_j$ . The different states supported by the metabolic network are generally of finite lifetime and therefore cells can change their metabolic state. These changes can be described by transitions from state  $j$  to  $i$  with probability  $b_{ij}$  per unit time, in which case the substrate remains unchanged. The full model is completely specified by the following master equation:

$$\begin{aligned} & \frac{\partial P(n_1, n_2, \dots, n_k, S, t)}{\partial t} \\ &= \sum_{j=1}^k P(n_1, n_2, \dots, n_j - 1, \dots, n_k, S + 1/Y_j, t) \mu_j(S) \\ & \quad - \sum_{j=1}^k P(n_1, n_2, \dots, n_j, \dots, n_k, S + 1/Y, t) n_j \mu_j(S) \\ & \quad + \sum_{i,j} P(n_1, n_2, \dots, n_j - 1, \dots, n_i + 1, \dots, n_k, S, t) (n_i + 1) b_{ji} \\ & \quad - P(n_1, n_2, \dots, n_j, \dots, n_i, \dots, n_k, S, t) \sum_{i,j} n_i b_{ji}. \end{aligned}$$

In the present analysis, however, we will be interested in timescales over which transitions can be neglected and they will therefore not be included in the model. A metabolic tradeoff is represented at a phenomenological level, by imposing some relation between the growth rate and yield at the single cell level.

## Results

### Deterministic (Average) Dynamics

A homogeneous population growing in a microdroplet, starting with initial conditions of cell number and resource  $(N_0, S_0)$ , will grow to a stopping time  $t^*$  when the resource is depleted. The stopping time and the final number of cells  $N(t^*)$  is determined by the number of divisions and the amount of resource taken up for each division:

$$S_0 - \frac{N(t^*) - N_0}{Y} = 0 \quad (1)$$

where  $Y$  is the uniform yield. For typical droplets the final population size is on the order of  $10^5 - 10^6$  cells, and the starting initial number  $N_0$ , typically in the range  $= 1 - 1000$ , is relatively very small. The final population size is therefore dominated by  $YS_0$ , and its dependence on  $N_0$  negligible (see Figure 1; dotted line). However, if the population is heterogeneous in its properties, generally its composition will change over time and with it the averaged yield. This will affect both the stopping time and the final number of cells. To see this, consider the deterministic approximation to the model dynamics: as long as there is still growth resource, these dynamics are described by the equations

$$\frac{dS}{dt} = - \sum_{j=1}^k \frac{\mu_j(S) n_j}{Y_j} \quad (2)$$

$$\frac{dn_i}{dt} = \mu_i(S) n_i \quad (3)$$

where  $(n_i(t), S(t))$  is the average trajectory of the system over time. The stopping condition,  $S=0$ , analogous to Eq. (1) can then be written, taking into account this variability:

$$S_0 - \sum_{j=1}^k \frac{1}{Y_j} \int_0^{t^*} \mu_j(S(t)) n_j(t) dt = \int_0^{t^*} N(t) \langle \frac{\mu}{Y} \rangle_t dt = 0 \quad (4)$$

Here  $N(t) = \sum_{j=1}^k n_j(t)$  is the total number of cells at time  $t$ , and we have defined the average of any property  $x$  over the population at time  $t$  as  $\langle x \rangle_t = \sum_{j=1}^k x_j \frac{n_j(t)}{N(t)}$ . The stopping time and the final cell number are therefore seen to depend on the average resource uptake rate  $\langle \frac{dS}{dt} \rangle_t$  along the trajectory, which makes them generally history-dependent quantities. Note that if cells differ only by their division rate but have all the same yield, the relation (1) is recovered; although the population composition changes with time, the mean uptake rate remains the same.

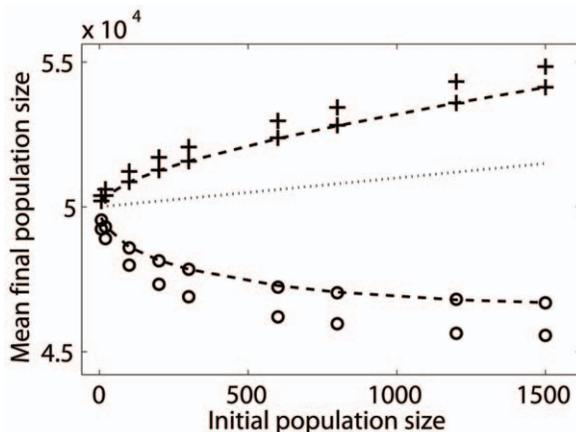
It is instructive to consider a simplified population with only two phenotypes and with a step-function dependence of the division rate on the substrate:  $\mu_i(S) = \mu_i$  when  $S > 0$  and  $\mu_i(S) = 0$  when  $S = 0$ . We assume initial conditions to be distributed around a symmetric mixed population so on average the initial amount of each phenotype will be half of the total initial population. Eq. (4) then reads.

$$S_0 - \frac{N_0}{2Y_1} [e^{\mu_1 t^*} - 1] - \frac{N_0}{2Y_2} [e^{\mu_2 t^*} - 1] = 0. \quad (5)$$

This equation can generally be solved numerically for the stopping time and from that the final number of cells can be found. For a special choice of division rates  $\mu_1 = 1$ ,  $\mu_2 = 2$  one may solve analytically the resulting quadratic equation in  $\lambda = e^{t^*}$ . Defining the ratio between the yields of the two cell states as  $r = \frac{Y_2}{Y_1}$  the quadratic equation reads.

$$\lambda^2 + r\lambda - r - 1 - 2\frac{Y_2 S_0}{N_0} = 0$$

revealing a dependence on one additional dimensionless parameter,  $\varepsilon = \frac{N_0}{Y_2 S_0}$ , which represents the initial population size relative



**Figure 1. Average final vs. initial population size in micro-populations grown to saturation of resource.** Dotted line: a population with a uniform yield. Symbols: Monte Carlo results for two-state populations with variability in yield and in growth rate. Dashed lines: analytic approximations relevant only for special parameter values. Crosses: Monte Carlo simulation for “metabolic tradeoff”  $r = Y_2/Y_1 = 2$ ,  $\rho = \mu_1/\mu_2 = 2$  (lower crosses),  $\rho = 1.8$  (upper crosses). Circles: variable yield positively correlated with division rate  $r = 0.5$ ,  $\rho = 2$  (upper circles),  $\rho = 1.8$  (lower circles). doi:10.1371/journal.pone.0052105.g001

to the approximate number of divisions supported by the medium. From the physical solution of this equation,  $\lambda$ , we find an explicit expression for the final number of cells as a function of the initial number:

$$N(t^*) = \frac{N_0}{2} (e^{t^*} + e^{2t^*}) = \frac{N_0}{2} (\lambda + \lambda^2) = \frac{N_0}{2} \phi(r, \varepsilon) \quad (6)$$

where the function  $\phi$  gives a generally nonlinear dependence on the initial number of cells  $N_0$ . In many experimental cases of interest  $\varepsilon$  is very small; droplets are started with a small number of cells and the amount of resource allows a large number of cell divisions before saturation. Expanding in this small parameter we find the approximation

$$N(t^*) \approx N_0 + Y_2 S_0 + Y_2 S_0 (1-r) \sqrt{\frac{\varepsilon}{2}}. \quad (7)$$

Now imagine an experiment in which the initial droplet cell content  $N_0$  is varied while other parameters are held fixed. Then the correction to the final cell number induced by metabolic variability will depend nonlinearly on the initial number, to leading order as the square root of  $N_0$ , with coefficients that can result in an increasing or decreasing function depending on the ratio between yields. Recall that we have marked by 2 the faster-growing phenotype, and therefore the ratio between yields,  $r = \frac{Y_2}{Y_1}$ , defines the correlation between division rate and yield in single cells. A ratio smaller than one,  $r < 1$ , describes a metabolic tradeoff, namely cells that grow faster have a smaller yield. Such a negative correlation between growth and yield has been observed in several cases for microorganisms (reviewed in [33,34]). Then, the final number of cells is an increasing function of the initial conditions (Figure 1, upper increasing dashed line). By contrast if the yield is positively correlated with the division rate,  $r > 1$ , a decreasing curve is found (Figure 1, lower decreasing dashed line). These analytical results are relevant for a special choice of division rates which render the equation solvable, however we expect that the qualitative nature of the solutions to be insensitive to the exact values of division rates. For comparison Figure 1 (symbols) shows also the results of a Monte-Carlo simulations of the two-state population that grows in a microdroplet, giving the final cell number directly without any approximation. For a growth-rate ratio of 2, the analytic results are recovered (symbols overlapping with dashed lines); for a different non-integer ratio, the qualitative features remain. (+, metabolic tradeoff; o, yield and division rate positively correlated).

### Stochastic Analysis by Phenotype-space Trajectories

The results for the averaged dynamics show that yield variability can significantly modify the dependence of the final mean number of cells on the initial number. It is expected that the variability among populations will show an even more pronounced dependence; to investigate this variability we consider the stochastic dynamics of the model. Here the object of interest is the probability distribution of the system among its possible states,  $P(n_1, n_2, \dots, n_k, S, t)$  which obeys the Master Equation (see Eq. (1)), so that for a given initial condition the explicit time dependence of the distribution can be computed. However, to describe the ensemble of populations grown to saturation we need much less information and one can formulate the problem in a simpler way. The key observation is that in order to characterize the distribution of populations at the end of this finite-time growth

one needs only the sequence of divisions that make up the trajectories, and not the time points at which they occurred. Therefore one may study the geometry of random trajectories in phenotype-space; we develop this approach for the toy two-state model discussed above, but the basic idea can be generalized to multiple states.

The growth of a micro-population with two cell states is pictured as a random walk on a discrete two-dimensional phenotype plane  $(n_1, n_2)$  with the axes representing the number of cells of each type. Time units of this random walk will be taken as times between divisions (these time units are not constant but vary randomly). The initial condition is  $(n_1(0), n_2(0))$ , and at each point  $(n_1, n_2)$  there is a probability  $p_1$  for the next division to be of type 1, in which case the trajectory will advance one step along the  $n_1$  direction; and a probability  $p_2 = 1 - p_1$  for it to be of the other type, in which case it will advance along the  $n_2$  direction. To find  $p_1$  we approximate the divisions as Poisson processes with probabilities per unit time  $\mu_1, \mu_2$  to divide. The probability that the next division is of type 1 and occurs after exactly time  $t$  is given by.

$$p(1, t) = e^{-[n_1\mu_1 t + n_2\mu_2 t]} \cdot n_1\mu_1 dt$$

Therefore the probability for the next division to be of type 1, regardless of the time it occurred, is just the integral over  $t$  of the above expression,

$$p_1 = \int_0^\infty e^{-[n_1\mu_1 + n_2\mu_2]t} \cdot n_1\mu_1 dt = \frac{n_1\mu_1}{n_1\mu_1 + n_2\mu_2} = \frac{\rho n_1}{\rho n_1 + n_2}.$$

In contrast to a simple random walk, where the probability to move in each direction is the same everywhere, here it depends on the position in the plane, namely on the number of cells of each type. It also depends globally on the ratio  $\rho$  between their division rates. To describe in phenotypic plane the stopping condition at depletion of the resource, we note that the two cell types consume a quantity  $\frac{1}{Y_1}$  or  $\frac{1}{Y_2}$  respectively of this resource in each division. Therefore the geometric place of all points  $(n_1, n_2)$  that correspond to the stopping condition is.

$$S_0 - \frac{n_1 - n_1(0)}{Y_1} - \frac{n_2 - n_2(0)}{Y_2} = 0 \tag{8}$$

which can be written as:  $n_2 + rn_1 = c$ , the ‘‘stopping line’’ (using the yield ratio  $r$  and defining the constant  $c \equiv S_0 Y_2 + n_2(0) + rn_1(0)$ ). The form of the equation describing the stopping line shows that if  $r = 1$  all trajectory endpoints have the same final number of cells  $n_1 + n_2$ , while for  $r \neq 1$  the final number of cells varies along the stopping line. Figure 2 illustrates the space and the dynamics by showing some individual trajectories of micro-population growth: starting from a given initial condition  $(n_1(0), n_2(0))$ , each trajectory shows the history of a single population which, depending on the order of divisions of each of the two types, proceeds randomly in the phenotypic plane until it hits the stopping line where all trajectories are absorbed. In Figure 2A the two cell types are uniform in their properties and the random walk is therefore symmetric in the two directions. The final number of cells is deterministic and only the final population composition - what fraction of each type - is variable, depending on the details of the random trajectory, with a symmetric distribution. Figure 2B shows that non-uniform division rates of the two types cause a bias in the

trajectories and they become curves in the plane. Still, the final number of cells is deterministically fixed because both types have the same yield. In contrast Figure 2C shows that non-uniform yields modify the slope of the stopping line, and therefore each trajectory on that line has a different final number of cells. In this graphical representation, it is clear that the two metabolic properties - division rate and yield - play separate and different roles in the problem: the first determines the bias of trajectories while the second determines the slope of the stopping line. Any assumed relation between them can be inserted as additional input to the model.

**Probability distributions at the stopping line.** The probability distribution at the stopping line is determined by the probabilities of trajectories arriving at each point; this is what allows us to leave out of the discussion the temporal structure of trajectories. As we have seen, the composition of the final population, namely the number of  $n_1$  cells, is always a random variable even if the cell types are identical in their phenotypic properties. Our goal is to calculate  $f_{r,c}(n_1)$ , the probability density function of  $n_1$  on the stopping line  $n_2 + rn_1 = c$ , given an initial condition  $\overleftarrow{n}(0)$ , or at least to estimate its moments.

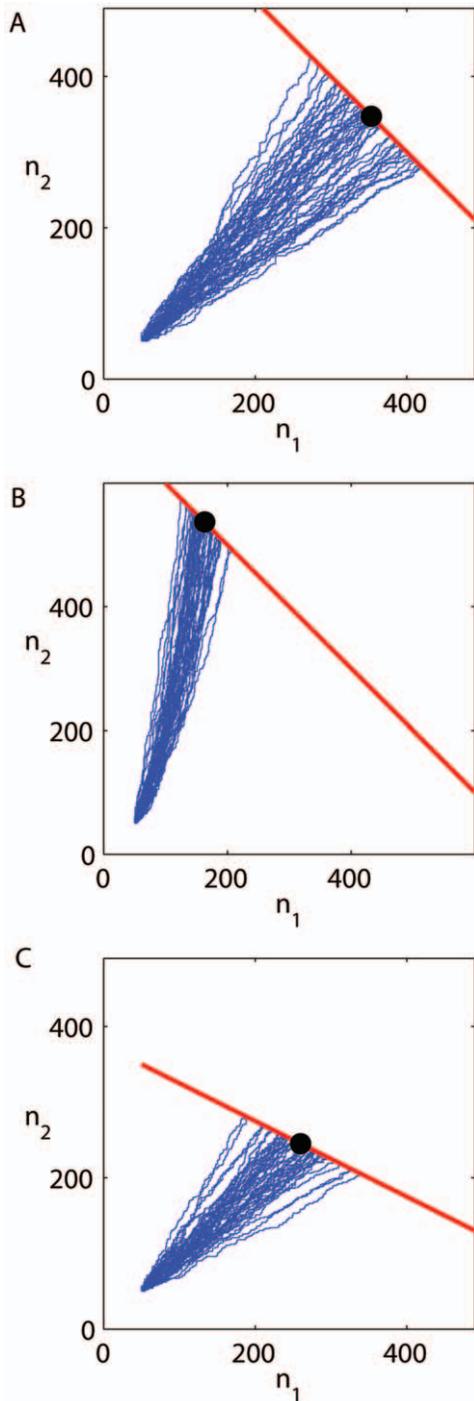
We first consider the simpler case where the yield in the population is uniform ( $r = 1$  or  $Y_1 = Y_2 = Y$ ) and the division rate is uniform ( $\rho = 1$  or  $\mu_1 = \mu_2 = \mu$ ), corresponding to Figure 2A. The total number of cell divisions in all trajectories is fixed at  $N(t^*) - N_0 = YS_0$ , and we have two types of events - division of type 1 or 2 - occurring with the same probability out of a total given number of events. This can be mapped to a classic problem in probability theory known as Polya’s Urn [38,39]. In Polya’s original problem, white and black balls are drawn at random from an urn; each time one ball is chosen it is returned back to the urn with an additional ball of the same color. In our analogy the probability that out of  $k$  draws from the urn  $n_1$  will be white is equivalent to the probability that out of  $k$  divisions  $n_1$  will be of type 1. This probability is the well known Beta-Binomial distribution:

$$P_k(n_1; n_1(0), n_2(0)) = \binom{k}{n_1 - n_1(0)} \frac{\mathbf{B}(n_1, k + n_1(0) + n_2(0) - n_1)}{\mathbf{B}(n_1(0), n_2(0))} \tag{9}$$

where  $\mathbf{B}(x, y)$  is the Beta function. Using the moments of this distribution, the ratio between its standard-deviation and mean is

$$\eta \equiv \frac{\sqrt{\text{Var}(n_1(t^*))}}{\langle n_1(t^*) \rangle} = \sqrt{\frac{N(t^*)}{N(t^*) - N_0}} \sqrt{\frac{n_2(0)}{n_1(0)}} \frac{1}{\sqrt{N_0 + 1}}. \tag{10}$$

This ratio, representing the relative width of the final distribution, does not decay with an increasing number of divisions; in the limit  $N(t^*) = N_0 + k \gg N_0$  it converges to a quantity proportional to the inverse square root of the inoculum size. An additional factor represents the initial composition, namely how many cells of each type were in the inoculum. Assuming that in a typical experiment this initial composition is averaged over some set of possible preparations, there remains still a strong dependence on the inoculum size:  $\eta \propto 1/\sqrt{(N_0 + 1)}$ . At the extreme, starting many copies of an initial population with a single cell of each type, the final ensemble of populations has all possible values of  $n_1$  between 0 and  $N$  with equal probability (a uniform distribution), no matter how many generations have passed. In contrast to our intuition



**Figure 2. Trajectories in phenotypic space for heterogeneous micro-populations grown to saturation of resource.** Each point in the plane  $(n_1, n_2)$  represents the number of cells of each type in the population. Division of type 1 increases  $n_1$  by one and corresponds to the trajectory advancing along the  $x$ -direction, and similarly for type 2. All three panels show trajectories that start from an initial population of  $n_1 = n_2 = 50$ . (A) symmetric types ( $r = 1, \rho = 1$ ). Trajectories are equally likely to proceed along  $x$  or  $y$ , and the stopping condition is of slope  $(-1)$ . (B) Types differ by their division rates ( $r = 1, \rho = 0.5$ ), causing the trajectories to be biased towards the faster growing type. (C) Types differ in their yield with respect to the finite resource ( $r = 0.5, \rho = 1$ ), causing the stopping line to be of slope different from  $(-1)$ ; growth will stop after a variable total number of divisions depending on trajectory, since each type consumes a different amount of resource at division. doi:10.1371/journal.pone.0052105.g002

from many repetitions of a coin-flipping, there is no convergence after a long time to a distribution that is strongly peaked around  $1/2$ . This result illustrates an important property of all models studied here: for a large number of cell divisions, the distribution tends to a limit whose shape is independent of the number of divisions but still maintains a fingerprint of the initial population size [40].

A generalization of Polya's urn can be constructed for the more general case of variable division rates, (illustrated in Figure 2B). Since the ratio  $\rho = \mu_1/\mu_2$  is the only relevant parameter for the phenotype-plane trajectories, we may approximate it by the ratio of two large integers  $M_1, M_2$  and map the problem to an urn in which each cell of type  $i$  is represented by  $M_i$  balls of color  $i$ . If each ball chosen at random is returned to the urn with an additional  $M_i$  balls of the same color, the ratio of probabilities for choosing each color remains faithful to the ratio of division rates, and the number of cells of each color in the urn at any given time is proportional to the number of cells of the corresponding type. Formulation of this construction can be found in Text S1.

The limiting distribution shape is revealed in term of scaling variables. Figure 3 shows several such limiting distributions for the case of uniform yield as a function the scaling variable  $X_N = n/N^\rho$ . For uniform division rate ( $\rho = 1$ ) this variable reduces to  $X_N = n/N$ , the fraction of type-1 divisions, and the limit distribution is just the bounded Beta distribution  $X_N \rightarrow X \tilde{\text{Beta}}(n_1(0), n_2(0))$ . For non-uniform division rate the scaling variable is a more abstract mathematical quantity, and its distribution converges as  $X_N \rightarrow X \tilde{U}^{-\rho} V$  where  $U$  and  $V$  are two Gamma distributed random variables -  $U \tilde{\Gamma}(n_1(0), 1)$ ,  $V \tilde{\Gamma}(n_2(0), 1)$ . This distribution has no closed form density but has known moments (see Text S1 for details). The relative width of these distributions in both cases, equal and variable division rates, decreases as the inoculum size increases.

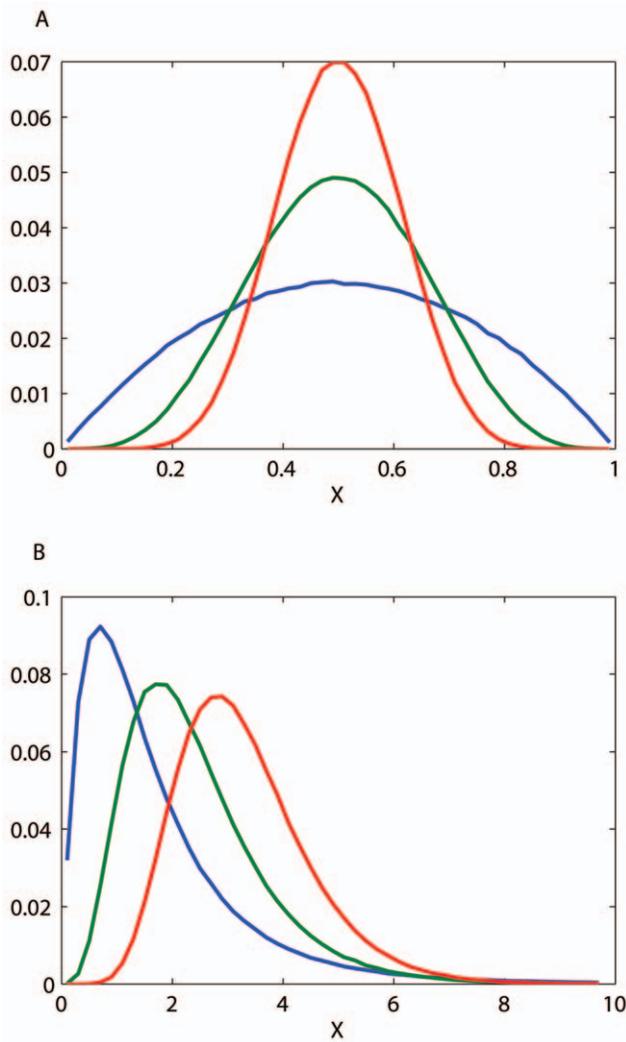
In the example above, Eq. (10) and the distributions of Figure 3 manifestly show how the size of the initial preparation affects in the variation in population composition when yield is uniform. This is somewhat analogous to the effect of population bottlenecks on variability through genetic drift. Next we consider the more interesting case of a population with variable yield, in which not only the composition but the total final number of cells is itself a random variable.

**Yield variability and distribution of final population size.** In terms of the graphical phenotypic plane representation, we are interested in trajectories that end on a line with slope  $\neq -1$ , since  $r$  is now different from one (see Eq. (8)). In our model trajectories in phase plane can advance only in one of two directions: increasing  $n_1$  or  $n_2$  (this property is lost if transitions between phenotypes are accounted for). Using the stopping condition and this monotone property of the trajectories, one can find a relation between the cumulative probability function of  $n_1$  with non-uniform yields  $F_{r,c}(n_1)$  and the uniform case discussed in the previous section, but with an effective value of  $c$ :

$$F_{r,c}(n_1) = F_{r=1, c+n_1(1-r)}(n_1). \quad (11)$$

This relation provides the key to computing the distributions in the variable-yield population from the uniform-yield results presented above. It is illustrated graphically in Figure 4.

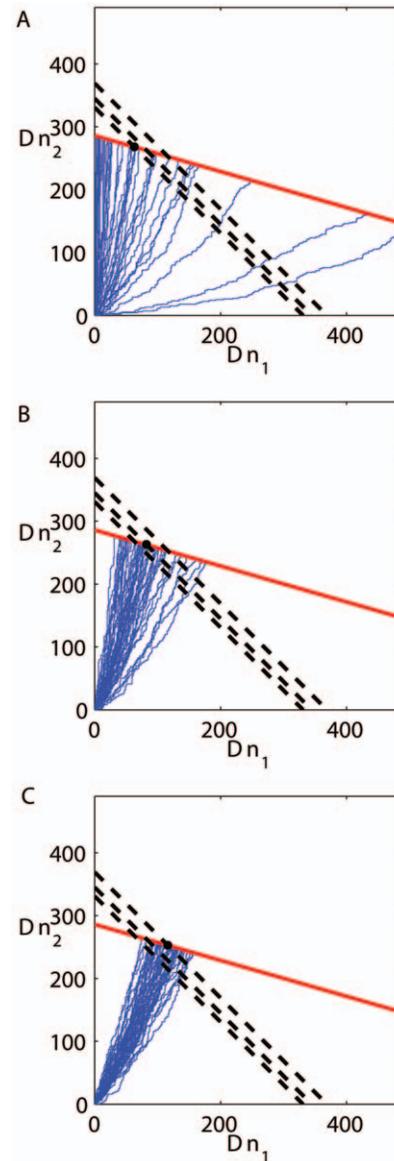
When yields are variable, each point on the stopping line corresponds to a different final population size. It is determined by the  $n_1$  coordinate of that point and by the initial conditions:



**Figure 3. Scaled distribution of the number of cells of metabolic type 1 in the final population.** All distributions are for symmetric initial composition, equal yields and a large number of divisions. Different distributions in a plot are for different initial populations  $N_0$  (Blue -  $N_0=4$ , Green -  $N_0=10$ , Red -  $N_0=20$ ). These distributions are plotted as a function of the scaling variable  $x$  (see text for details), and their shape does not depend on the number of divisions but does depend on the initial number of cells. (A) The two types have the same growth rate and are therefore equal in all their properties. Population composition varies only because individual trajectories are composed of different sequences of divisions of the two types. Because of the symmetry between types, all distributions are symmetric around 0.5. (B) The two types have different growth rates, and the distribution of final composition becomes skewed. doi:10.1371/journal.pone.0052105.g003

$$N(t^*) = n_1(t^*) + n_2(t^*) = c + n_1(t^*)(1-r). \quad (12)$$

This relation is used to define the scaling variable  $x(n_1)$ , which reveals the limiting distribution shape for a large number of divisions (see Text S1 for details). The mean and variance of this distribution are found by inverting this scaling variable as  $n_1 = n_1(x)$ , as follows:

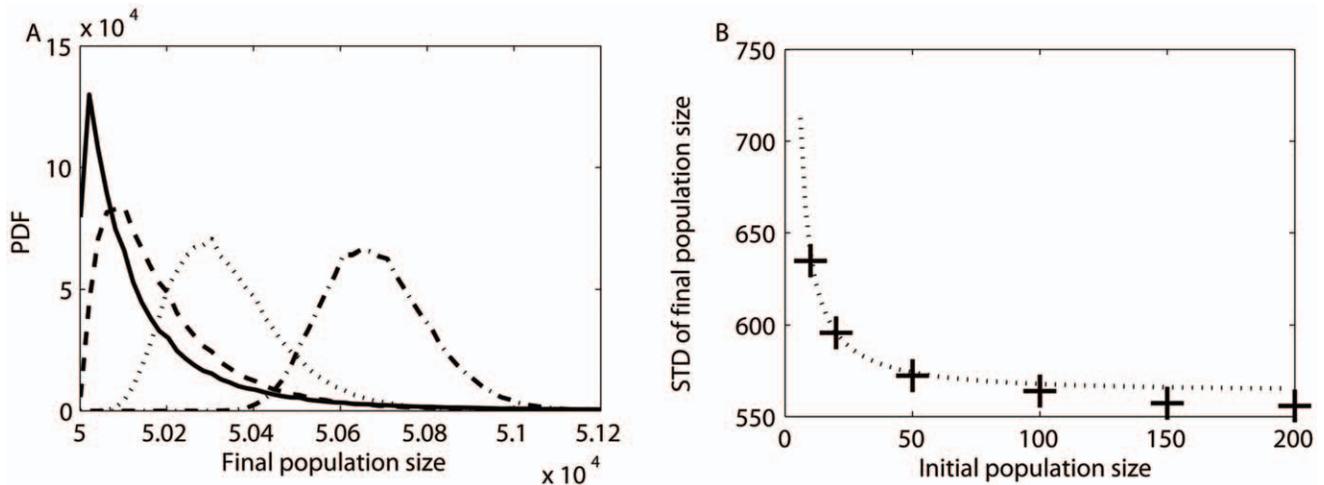


**Figure 4. Relating trajectories of uniform and variable yield populations.** All gray trajectories, ending on the solid red line (variable-yield stopping line) at points corresponding to  $n_1 < 200$ , build up the cumulative probability for the final population to have less than 200 cells of type 1. Due to the monotone property of trajectories, they all cross also the dashed line (uniform-yield stopping line) that passes through the point  $n_1 = 200$  obeying the same constraint and thus the cumulative probability is the same. The parameters of the two lines are simply related through  $n_1$  (see Eq. (11)). doi:10.1371/journal.pone.0052105.g004

$$\langle N \rangle = N_0 + Y_2 S_0 + (1-r) \langle n_1(x) \rangle \quad (13)$$

$$\text{Var}(N) = (1-r)^2 \text{Var}(n_1(x)). \quad (14)$$

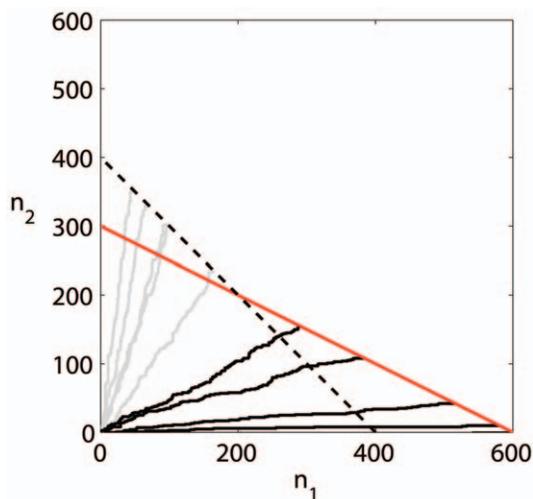
These expressions show that the mean final population size obtains an additional term from yield variability ( $r \neq 1$ ) which can be negative or positive depending on the correlation between



**Figure 5. Trajectories in phenotypic space illustrating dependence on inoculum size.** Results of Monte Carlo simulations are shown for initial population sizes of 2,20,50 equally divided between the two types. Axes are number of divisions. Parameters:  $\rho = r = 2/3$ . All trajectories end on the stopping line (red line) which has an asymmetric slope; trajectories ending at different points on the line have a different final population size (dashed lines). The mean of all trajectory endpoints on the line is marked with a black circle. It can be seen that as the inoculum size increases, going from (A) through (B) to (C), the spread of trajectories decreases and their mean crosses a dashed line corresponding to a higher number of cells. doi:10.1371/journal.pone.0052105.g005

division rate and yield (see also Figure 1). The variance in final population size, of course, becomes positive from yield variability with a variance that increases with the extent of this variability, represented by  $(1-r)$  in this two-state model. A detailed derivation of these results is presented in Text S1.

The dependence on the inoculum size is incorporated in the moments of the random variable  $n_1(x)$ . To illustrate this dependence graphically in phenotype space, we show in Figure 5 sets of trajectories starting at different inoculum sizes with all other parameters held fixed. In this figure, for ease of comparison, the axes are the number of divisions during growth ( $\Delta n_1, \Delta n_2$ ) rather



**Figure 6. Final population size for a heterogeneous micro-population with metabolic tradeoff.** (A) Distributions of the final populations size from simulations with division rate ratio  $\rho = 0.5$  and yield ratio  $r = 0.5$ , for different initial population sizes - 2 cells (solid line), 4 cells (dashed line), 8 cells (dotted line), 64 cells (dash-dot line). In (B) we can see the Standard deviation of the final population size as function of initial population size, in good agreement with the analytic approximation. doi:10.1371/journal.pone.0052105.g006

than the absolute number of cells of each type. The difference in growth rates causes a bias of trajectories towards  $\Delta n_2$ , the faster growing type. For small inoculum sizes (Figure 5A), an initial step in the  $n_2$  direction can rapidly amplify itself causing a heavy bias of the mean towards this type; because this rapidly growing type is less efficient, the mean number of cells at the stopping line is smaller, namely the stopping line intersects a dashed line  $\Delta n_1 + \Delta n_2 = const$  with a smaller value of the constant. At the same time, this amplified sensitivity causes a wide spread of the trajectories. As inoculum size increases, the relative effect of this bias at the stopping line becomes smaller (Figures 5B, 5C); the trajectories are less spread-out and the mean final number of cells increases.

In several special cases of interest the moments of the distribution of final population size can be explicitly calculated or approximated, for example the case of equal growth rates, or growth rate ratio of exactly 2. Analytic calculation is harder in the general case of variable growth rates and variable yields. Details of the calculations for special cases which allow solution are presented in the Text S1. In any case the moments can be calculated directly from averages over the limiting distribution by numerically inverting the relation defining the scaling variable.

In Figure 6 we show several examples of the distribution of final population size (A), and approximations for the standard deviation which are in good agreement with the simulation (B) and decreases as a power law of the inoculum size.

The stochastic analysis of the model allows the computation of variance and higher moments, analytically in some cases and numerically in others. The average is consistent with the deterministic approximation solution presented earlier, however it is expressed by moments of another random variable which is not generally available in closed form. In the deterministic analysis it was possible to develop a controlled approximation for the limit of large number of divisions, which is not easily done in the general stochastic equations.

## Discussion

Recent advances in microdroplet and microfluidic technology for growing cell populations have the potential to provide basis for

many practical applications. At the same time, they open up an interesting regime for cell population behavior that has previously been unexplored. The intermediate sizes of populations, the control over initial conditions and the availability of an ensemble of populations equally prepared, offer to think about these micro-populations as repeated experiments of mesoscopic sized systems and to consider the stochastic properties of this ensemble of finite populations.

Cell populations have a structured and dynamic phenotypic composition in terms of their metabolic properties, and in particular their growth rate and their yield with respect to growth-limiting resources. This structure can strongly influence the dynamics of the population in the intermediate-size regime. We have formulated and solved a theoretical model that relates metabolic variability of cells to measurable statistical properties illustrating the emergence of novel behavior in this regime of micro-populations. In the absence of asymptotic takeover that results from infinite-time competition, the presence of both efficient and fast-growing phenotypes creates a dependence between the dynamics of the population composition and the effective consumption rate of the resource, resulting in a variability of the final population composition and size. Moreover, even the mean final population size exhibits an inoculum dependency which does not disappear after a large number of divisions. We have demonstrated this effect for a two state model with large ratios of yields and divisions rates. With smaller, more realistic ratios of the metabolic parameters, the two state model produce a smaller but qualitatively identical effects. However, modeling a more realistic situation would also require including a broader range of phenotypes. These more realistic models remain subject for future work.

As a theoretic approach to analyzing this regime of population growth we have constructed a mapping to a probabilistic problem, a generalization of Polya's Urns [41], which allowed us to compute stochastic properties of the population. In this class of problems it has been shown that limiting distributions appear which are independent of the number of trials but do depend on the initial conditions; in the case of a dividing population, it means that statistical properties of the ensemble of populations converge to a limit after a large number of divisions, which still reflects the initial condition of the inoculum.

All of these predictions obtained in our work can be tested in experiments on microdroplets. For example, the dependence of the final distribution on initial condition (number of cells in

preparation) but its independence on number of divisions (amount of resource in each droplet) can be tested directly. At the present state of the technology both these parameters can be controlled to good accuracy. The dependence of variance on inoculum size can be easily measured [27]. The extent of yield variability, represented in our model by  $(1-r)$ , is expected to be sensitive to the type of growth medium and other details of growth conditions [42]. A systematic investigation of these statistical properties as a function of growth parameters can indirectly provide much information about actual yield variability in a microorganism population and its dependence on environmental parameters.

The limiting environment and the finite growth time in microdroplets bring into light the importance of yield variability, which has been largely overlooked in previous studies. In unconstrained populations the fastest dividing cells take over and thus division rate is often equated with the Darwinian concept of "fitness". In micro-population, on the other hand, the yield becomes a selective parameter – the yield composition of the population changes by the time the resource is depleted, with the higher-yield cells having a larger representation in the population regardless of their growth rate. This effect can be used to select high-yield cells via cycles of growth to saturation followed by mixing (B. Teusink, unpublished results, 2012). Such a controlled growth protocol can have implications to biotechnology by providing an empirical way to maximize biomass yield in microorganisms building on existing yield variability in the population. Developing such protocols is subject for future theoretical and experimental research.

## Supporting Information

### Text S1

(PDF)

## Acknowledgments

We are grateful to J. Bibette for suggesting this problem and for sharing with us his unpublished data, to E. Braun for reading and commenting on the manuscript and to N. Shnerb for illuminating remarks.

## Author Contributions

Conceived and designed the experiments: NB YE. Performed the experiments: NB YE. Wrote the paper: YE NB.

## References

- Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135: 216–226.
- Brenner N, Farkash K, Braun E (2006) Dynamics of protein distributions in cell populations. *Phys Biol* 3: 172–182.
- Tsuru S, Ichinoise J, Kashiwagi A, Ying BW, Kaneko K, et al. (2009) Noisy cell growth rate leads to fluctuating protein concentration in bacteria. *Phys Biol* 6: 036015.
- Huh D, Paulsson J (2011) Random partitioning of molecules at cell division. *Proc Natl Acad Sci USA* 108: 15004–15009.
- Salman H, Brenner N, Tung Ck, Elyahu N, Stolovicki E, et al. (2012) Universal protein fluctuations in populations of microorganisms. *Phys Rev Lett* 108: 238105.
- Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427: 415–418.
- Friedlander T, Brenner N (2008) Cellular properties and population asymptotics in the population balance equation. *Phys Rev Lett* 101: 18104.
- Shahrezaei V, Ollivier JF, Swain PS (2008) Colored extrinsic fluctuations and stochastic gene expression. *Mol Sys Biol* 4: 196.
- Booth IR (2002) Stress and the single cell: intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. *Int J Food Microbiol* 78: 19–30.
- Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity. *Nat Rev Microbiol* 4: 577–87.
- Dhar N, McKinney JD (2007) Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr Opin Microbiol* 10: 30–38.
- Veening J, Smits WK, Kuipers OP (2008) Bistability, Epigenetics, and Bet-Hedging in Bacteria. *Ann Rev Microbiol* 62: 193–210.
- Fraser D, Kaern M (2009) A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol Microbiol* 71: 1333–1340.
- Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormance. *Ann Rev Microbiol* 9: 119–130.
- Lachmann M, Jablonka E (1996) The inheritance of phenotypes: an adaptation to fluctuating environments. *J Theoret Biol* 181: 1–9.
- Thattai M, van Oudenaarden A (2004) Stochastic gene expression in fluctuating environments. *Genetics* 167: 523–530.
- Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S (2004) Bacterial persistence as a phenotypic switch. *Science* 305: 1622–1625.
- Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. *Science* 309: 2075–2078.
- Filiba E, Lewin D, Brenner N (2012) Transients and tradeoffs of phenotypic switching in a fluctuating limited environment. *Theoret Pop Biol* 82: 187–199.
- Stolovicki E, Braun E (2011) Collective Dynamics of Gene Expression in Cell Populations. *PLoS One* : e20530.
- Elhanati Y, Schuster S, Brenner N (2011) Dynamic modeling of cooperative protein secretion in microorganism populations. *Theoret Pop Biol* 80: 49–63.

22. Smith HL, Waltman P (1995) The theory of the chemostat: dynamics of microbial competition. Cambridge University Press.
23. Pfeiffer T, Schuster S, Bonhoefer S (2001) Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292: 504–507.
24. MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press.
25. Boyce MS (1984) Restitution of r- and K-Selection as a model of density-dependent natural selection. *Ann Rev Ecol Syst* 15: 427–447.
26. Theberge A, Courtois F, Schaerli Y, Fischlechner M, Abell C, et al. (2010) Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology. *Angew Chem Int Ed* 49: 5846–5868.
27. Um E, Lee S, Park J (2010) Random breakup of microdroplets for single-cell encapsulation. *Appl Phys Lett* 97: 153703.
28. Marcoux P, Dupoy M, Mathey R, Novelli-Rousseau A, Heran V, et al. (2011) Micro-confinement of bacteria in w/o emulsion droplets for rapid detection and enumeration. *Coll Surf A: Physicochem Eng* 377: 54–62.
29. Baraban L, Bertholle F, Salverda MLM, Bremond N, Panizza P, et al. (2011) Microfluidic droplet analyser for microbiology. *Lab on a Chip* 11: 4057.
30. Dewan A, Kim J, McLean RH, Vanapalli SA, Karim MN (2012) Growth kinetics of microalgae in microfluidic static droplet arrays. *Biotech Bioeng* 109: 2987–2996.
31. Lu T, Shen T, Bennett M, Wolynes P, Hasty J (2007) Phenotypic variability of growing cellular populations. *Proc Natl Acad Sci USA* 104: 18982–18987.
32. Novak M, Pfeiffer T, Lenki RE, Sauer U, Bonhoeffer S (2006) Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. *The American Naturalist* 168: 242–251.
33. MacLean RC (2008) The tragedy of the commons in microbial populations: insights from theoretical, comparative and experimental studies. *Heredity* 100: 471–477.
34. Beardmore RE, Gudelj I, Lipson DA, Hurst LD (2011) Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature* 472: 342–346.
35. Trinh CT, Wlaschin A, Sreenc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotech* 81: 813–826.
36. Schauble S, Schuster S, Kaleta C (2011) Hands-on metabolism analysis of complex biochemical networks using elementary flux modes. *Meth Enzymol* 500: 437–456.
37. Maharjan RP, Seeto S, Ferenci T (2007) Divergence and Redundancy of Transport and Metabolic Rate-Yield Strategies in a Single *Escherichia coli* Population. *J Bact* 189: 2350–2358.
38. Johnson NL, Kotz S (1977) Urn models and their application: an approach to modern discrete probability theory. Wiley New York.
39. Feller W (1971) An introduction to probability theory and its applications. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley. URL <http://books.google.fr/books?id=L RNAAAAMAAJ>.
40. Janson S (2005) Limit theorems for triangular urn schemes. *Prob Theor Rel Fields* 134: 417–452.
41. Mahmoud H (2008) Pólya Urn Models, volume 76 of *Texts in Statistical Science*. Taylor & Francis Group. URL <http://books.google.fr/books?id=7Bizo28c2LQC>.
42. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420: 190–193.