Supplemental materials for Selection dynamics in transient compartmentalization

Alex Blokhuis,[1,2] David Lacoste,[1] Philippe Nghe,[2] and Luca Peliti[3,*]

[1]*Gulliver Laboratory, UMR CNRS 7083, PSL Research University,*
ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 05, France

[2]*Laboratory of Biochemistry, PSL Research University,*
ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 05, France

[3]*SMRI, 00058 Santa Marinella (RM), Italy*

(Dated: March 12, 2018)

[*] luca@peliti.org

1

Here we provide more details on (i) the determination of the parameter Λ from experimental data, (ii) the impact of mutations in the limit $\Lambda \to \infty$, (iii) details on the derivation of the asymptotes, for $\lambda \to \infty$ and $\Lambda \simeq 1$, (iv) further aspects of the phase diagram, and (v) a comparison of the phase diagram for linear and non-linear selection functions.

I.   EXPONENTIAL GROWTH AND THE VALUE OF Λ

The following table contains experimental values measured in Ref. [1] for the ribozyme and three different parasites. The nucleotide length, doubling time (T_d), and relative replication rate (r), are reported, from which we infer Λ in the final column. The doubling time T_d for the ribozyme is related to the growth rate α introduced in the main text by $T_d = \ln(2)/\alpha$, and similarly the doubling times of the parasites is related to the γ by $T_d = \ln(2)/\gamma$.

Type	Length (nt) 2	$T_d(s)$	Relative r	Λ
Ribozyme	362	25.0	1.00	1
Parasite 1	245	20.7	1.21	13
Parasite 2	223	17.1	1.46	107
Parasite 3	129	14.6	1.71	473

In the experiment, a typical compartment contains λ RNA molecules that can be ribozyme or parasite, $2.6 \cdot 10^6$ molecules of Qβ replicase, and $1.0 \cdot 10^{10}$ molecules of each NTP. Replication takes place by complexation of RNA with Qβ replicase, which uses NTPs to make a complementary copy. This copy is then itself replicated to reproduce the original. There is a large amount of nucleotides, so that exponential growth of the target RNA proceeds until $N \approx n_{Q\beta}$. Starting from a single molecule, it takes $n_D = \log_2 n_{Q\beta} = 21.4$ doubling times to reach this regime. In a parasite-ribozyme mixture, we can estimate Λ using the relative r:

$$\Lambda = \frac{2^{n_D}}{2^{n_D/r}} = 2^{n_D(1-\frac{1}{r})}. \tag{1}$$

Another important assumption of our model, is that we neglect a possible dependence of Λ on n and m. In order to test this assumption we have estimated the fluctuations of Λ in the following way. We recall that the total number of RNA at the end of the exponential

phase is the constant $n_{Q\beta}$ given above, thus $N = (n - m)2^{n_D} + m2^{n_D/r} = n_{Q\beta}$. We first solve for n_D in this equation for a given n and m and then we use this result into Eq. (1) to obtain Λ for a given n and m. We show in figure 1 a typical plot of the values taken by this function $\Lambda(n, m)$ for a particular choice of n and m, together with the probability distribution $P_\lambda(n, x, m)$ defined in Eq. (4) of the main text. In general $\Lambda(n, m)$ is close to a constant for soft parasites ($r = 1.2$), and is less constant for hard parasites ($r = 1.7$). Even in the later case however, Λ hardly varies in the range of n, m values where the probability distribution takes significant weight. We conclude that the assumption of neglecting a possible dependence of Λ on n and m has only a minor effect on our results.
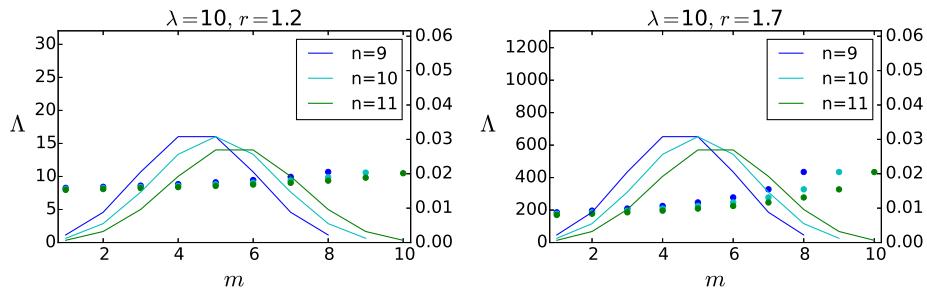
FIG. 1. Plots for Λ (colored dots) as function of the parameters (n, m) characterizing the initial composition of the compartments relative growth rates r and $\lambda = 10$, together with the probability distribution $P_\lambda(n, x, m)$ (solid lines) for $x = 0.5$.

II. GROWTH RATES FLUCTUATIONS

The model presented in the main text is purely deterministic, but fluctuations are still present due to the initial condition. Since growth rates typically depend on the initial condition, they will fluctuate when sampling the initial condition. While these fluctuations are already present in a deterministic model, effects associated to other fluctuations in the growth rates, which may occur at time scales smaller than the duration of the growth phase require a stochastic approach. Replication is intrinsically stochastic and therefore this additional source of fluctuations in grow rates could be present. In order to estimate such an effect, we have implemented below a stochastic version of our model.

A stochastic component in the growth phase of the ribozymes and parasites can be

3

included using a discrete Langevin approach. In such an approach, Eq. (1) is modified to become:

$$\ln \frac{\bar{m}}{m} = \alpha T + \xi_1, \tag{2}$$

where the first term on the right hand side accounts for the deterministic contribution we had before and ξ_1 is a Gaussian random variable of mean zero and variance σ_1. Similarly for the parasites:

$$\ln \frac{\bar{y}}{y} = \gamma T + \xi_2, \tag{3}$$

where ξ_2 is another similar noise controlling the growth of the parasites.

The noise which has been introduced here could describe either fluctuations of growth rate α or of the duration of replication. Note also that we still define Λ with respect to the mean time and growth rates as $\Lambda = \exp((\gamma - \alpha)T)$. Since N is fundamentally fixed by the number of enzymes $n_{Q\beta}$ in this problem, we still assume that N is fixed in this stochastic model. Then this condition $N = \bar{m} + \bar{y}$ leads to a constraint between the noise ξ_1 and ξ_2, which means that these two noises must be correlated. Then, Eq (2) is modified as

$$\bar{x}(n, m, \eta) = \frac{\bar{m}}{N} = \frac{m}{m + (n - m)\Lambda \exp(\eta)}, \tag{4}$$

where we have introduced the random variable $\eta = \xi_2 - \xi_1$. Let us introduce the variance of η which we call σ^2. This is the main parameter controlling the growth noise.

Eq. (5) is now modified as follows

$$x'(\lambda, x) = \frac{\int d\eta g(\eta) \sum_{n,m} \bar{x}(n, m, \eta) f(\bar{x}(n, m, \eta)) P_\lambda(n, x, m)}{\int d\eta g(\eta) \sum_{n,m} f(\bar{x}(n, m, \eta)) P_\lambda(n, x, m)}, \tag{5}$$

where $g(\eta)$ is the Gaussian distribution of mean zero and variance σ^2. In order to evaluate the correction due to the noise η, we expand the integrand in the numerator and denominator with respect to η and we perform the Gaussian integrals. The result is a modified recursion relation which contains a correction term proportional to σ^2. The explicit expression of this correction is lengthy and not given here, it was evaluated numerically.

This Taylor expansion is justified if the amplitude of the noise σ^2 is sufficiently small. In order to assess this, we investigate the various sources of noises in this problem. The noise could be due to the arrival times of $Q\beta$ or from the replication process itself. For the first source of noise, the time scale to form an $Q\beta$ - RNA complex due to diffusion can be evaluated as $t_D \approx (D_{RNA} \, c_{Q\beta}^{\frac{2}{3}})^{-1}$, where D_{RNA} is the diffusion constant for an

4

RNA strand (length ≈ 300) and $c_{Q\beta}$ the concentration of $Q\beta$ replicase. We found this timescale to be over $2 \cdot 10^4$ times smaller than replication times (15-25s, see SM 1). Due to this large difference in timescales, the noise in this problem should primarily be caused by the replication rather than by the binding of a $Q\beta$ to an RNA strand. Let us now look at the noise due to replication. Once a $Q\beta$ enzyme is bound to a single RNA molecule, the total replication time τ can be written as a sum of the dwell times of all the nucleotides to be added to the template and which are themselves exponentially distributed. When the binding rates are identical, the resulting distribution of τ is a gamma distribution with a coefficient of variation $1/\sqrt{n}$, in terms of n the number of nucleotides as shown in D. Floyd et al. (2010). Since our replicating molecules are long, typically between $n \simeq 150$ to 300, this coefficient of variation is quite small. This coefficient of variation is expected to increase due to the number of doubling times in the replicating phase, which is typically of the order of 20. In the end although we can not provide an accurate estimate of the noise of η, all these factors suggest a small noise amplitude for σ^2.

In the worst possible case, we would have $\sigma^2 = 1$ which is the case shown in the two figures below. The red contour plots of Δx corresponding to that of Fig 1 of the manuscript, the blue ones correspond to the prediction of a stochastic version of the same model including the correction due to σ^2. We only show the plots for $\Lambda = 4$ (left) and $\Lambda = 2.5$ (right), because we find that for high values of Λ the noise has only a negligible effect even in this worst case scenario, which is reasonable. While we see that the noise η affects significantly the contour plots in this worst case scenario, when $\sigma^2 = 1$, the effect is quite small with a more realistic estimate of the noise namely $\sigma^2 = 0.1$ as shown in figure 3:

All these results support our view that while the growth is intrinsically stochastic the deterministic model we have developed with a constant N and fluctuations only in the initial conditions indeed capture the main features of the experiment we are interested in. As shown in figures 2 and 3, the stability of the ribozyme phase in the new stochastic model is enhanced with respect to the deterministic model. This confirms that fluctuations are essential to stabilize the ribozyme phase, and favor it whether they come from the initial condition as in the deterministic model or from other sources as in the stochastic developed here.

5
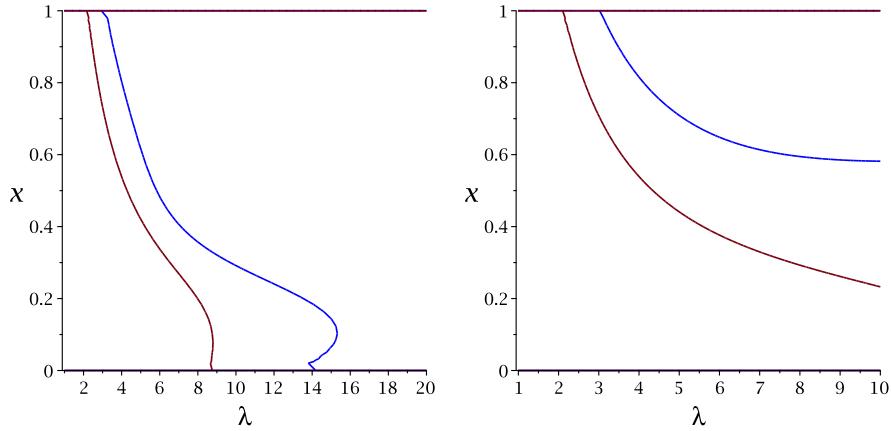
FIG. 2. Contour plots of Δx similar to that of Fig 1 of the manuscript, with no correction due to noise (red solid line) and including corrections due to noise with $\sigma^2 = 1$ (blue solid line). The figures correspond to $\Lambda = 4$ (left) and $\Lambda = 2.5$ (right).
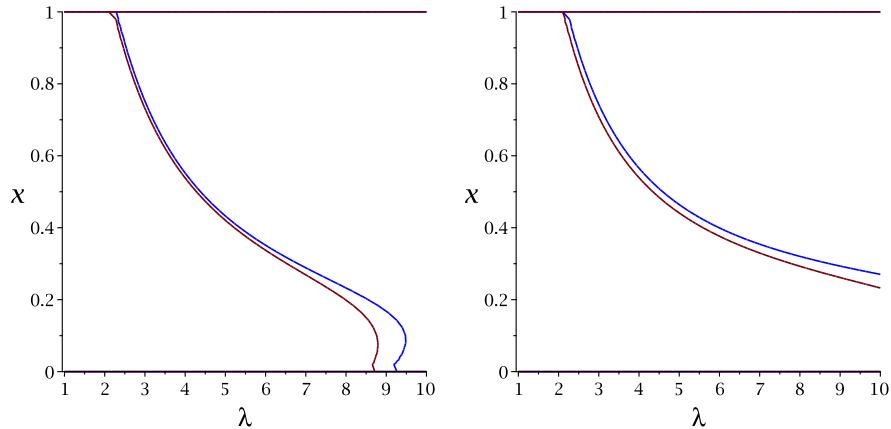
FIG. 3. Contour plots of Δx similar to that of Fig 1 of the manuscript, with no correction due to noise (red solid line) and including corrections due to noise with $\sigma^2 = 0.1$ (blue solid line). The figures correspond to $\Lambda = 4$ (left) and $\Lambda = 2.5$ (right).

III.   IMPACT OF MUTATIONS IN THE LIMIT $\Lambda \gg 1$

In this section, we explain how the approach described in the main text needs to be amended in the presence of mutations. We focus here only on the case that $\Lambda \gg 1$, because one can expect that the effect of mutations will be more dramatic in this limit corresponding to *hard* parasites. If one of such parasites is present in a compartment, it

6

invades the population very quickly, provided it appears during the exponential growth phase. As explained in the main text, to describe the limit $\Lambda \gg 1$, we can introduce three inoculation probabilities, p_{ribo} for compartments containing only ribozymes, p_{para} for compartments containing parasites and p_{zero} for empty compartments:

$$p_{ribo} = \sum_{n=1}^{\infty} \frac{x^n \lambda^n}{n!} e^{-\lambda} = (e^{\lambda x} - 1)e^{-\lambda},$$

$$p_{zero} = e^{-\lambda}, \tag{6}$$

$$p_{para} = 1 - p_{ribo} - p_{zero} = 1 - e^{\lambda(x-1)}.$$

Now, let us also introduce p_{mut} as the probability that a ribozyme is turned into a parasite as a result of a mutation during one replication event of the molecule. Then the probability that there is no mutation occurring during n^* replication events is

$$\zeta = (1 - p_{mut})^{n*}. \tag{7}$$

A typical value for this n^* corresponds to what is denoted n_D in the previous section, namely the number of replications until the end of the exponential growth regime.

With p_{zero} unchanged, the new probabilities for compartments containing ribozyme (resp. parasites) p'_{ribo} (resp. p'_{para}) are simply

$$p'_{ribo} = p_{ribo}\zeta, \tag{8}$$

$$p'_{para} = p_{para} + p_{ribo}(1 - \zeta). \tag{9}$$

Using these expressions in the recurrence relation, we obtain

$$x' = \frac{p'_{ribo}f(1)}{p'_{ribo}f(1) + p'_{para}f(0)} \tag{10}$$

Evaluating the fixed point stability at $x = 0$ then yields the equation for the asymptote

$$\zeta \lambda f(1) = f(0)(e^{\lambda} - 1). \tag{11}$$

Similarly, we can evaluate the fixed point stability at $x = 1$, to obtain

$$(e^{\lambda} - 1)\left(1 + \zeta\left(\frac{f(1)}{f(0)} - 1\right)\right)^2 = \lambda \frac{f(1)}{f(0)} e^{\lambda}. \tag{12}$$

For $p_{mut} \to 0$, $\zeta \to 1$ and we obtain the asymptotes mentioned in the text. For $\zeta < 1$, the asymptotic values of λ for both $x = 0$ and $x = 1$ become smaller. As a result, both the ribozyme and the bistable regions shrink as one would expect. In the extreme case where $\zeta \to 0$, both regions disappear completely since then the only solution to Eqs. (11-12) corresponds to $\lambda = 0$.

7

IV. ASYMPTOTIC BEHAVIOR FOR $\lambda \to \infty$

For large λ, for Λ close to 1 and x close to 1 (resp. 0), the most abundant compartments verify $m = n$ or $m = n - 1$ (resp. $m = 0$ or $m = 1$). As λ is large, we can neglect fluctuations in n and we can take $n = \lambda$. We therefore only look at the recursion for a typical compartment with $n = \lambda$, with a simplified notation $P_\lambda(n = \lambda, x, m) = P_\lambda(x, m)$, where

$$P_\lambda(x, m) = B_m(\lambda, x), \tag{13}$$

obtaining

$$x' = \frac{f(1)P_\lambda(x, \lambda) + \bar{x}P_\lambda(x, \lambda - 1)f(\bar{x})}{f(1)P_\lambda(x, \lambda) + P_\lambda(x, \lambda - 1)f(\bar{x})}, \tag{14}$$

where

$$\bar{x} = \bar{x}(\lambda, \lambda - 1) = \frac{\lambda - 1}{\lambda + \Lambda - 1} \simeq 1 - \frac{\Lambda}{\lambda}. \tag{15}$$

We have therefore

$$x' = \frac{x^\lambda f(1) + \lambda x^{\lambda-1}(1-x)\bar{x}f(\bar{x})}{x^\lambda f(1) + \lambda x^{\lambda-1}f(\bar{x})} = \frac{xf(1) + \lambda(1-x)\bar{x}f(\bar{x})}{xf(1) + \lambda(1-x)f(\bar{x})}. \tag{16}$$

Taking the derivative with respect to x we obtain

$$\frac{dx'}{dx} = \frac{\lambda(1-\bar{x})f(1)f(\bar{x})}{(xf(1) + \lambda(1-x)f(\bar{x}))^2}, \tag{17}$$

which for $x = 1$ yields

$$\left.\frac{dx'}{dx}\right|_{x=1} = \frac{\lambda(1-\bar{x})f(\bar{x})}{f(1)}. \tag{18}$$

Thus the boundary defined by the equation

$$\left.\frac{dx'}{dx}\right|_{x=1} = 1, \tag{19}$$

is given by

$$\Lambda \simeq 1 + \frac{f'(1)}{f(1)\lambda} = 1 + 6.12\,10^{-6}/\lambda. \tag{20}$$

Evaluating the stability around the fixed point $x = 0$ we obtain likewise

$$x' = \frac{\lambda x(1-x)^{\lambda-1}\bar{x}f(\bar{x})}{(1-x)^\lambda f(0) + \lambda x(1-x)^{\lambda-1}f(\bar{x})} = \frac{\lambda x \bar{x} f(\bar{x})}{(1-x)f(0) + \lambda x f(\bar{x})}, \tag{21}$$

where now $\bar{x}$ is given by

$$\bar{x} = \bar{x}(\lambda, 1) = \frac{1}{(\lambda - 1)\Lambda + 1} \simeq \frac{1}{\Lambda\lambda}. \tag{22}$$

Evaluating the derivative of $x'(x)$ at $x = 0$ we obtain

$$\left. \frac{dx'}{dx} \right|_{x=0} = \frac{\lambda \bar{x} f(\bar{x})}{f(0)}. \tag{23}$$

This gives the boundary as

$$\Lambda = 1 + \frac{f'(0)}{f(0)\lambda} = 1 + 19.8661/\lambda. \tag{24}$$

V. ADDITIONAL FEATURES OF THE PHASE DIAGRAM
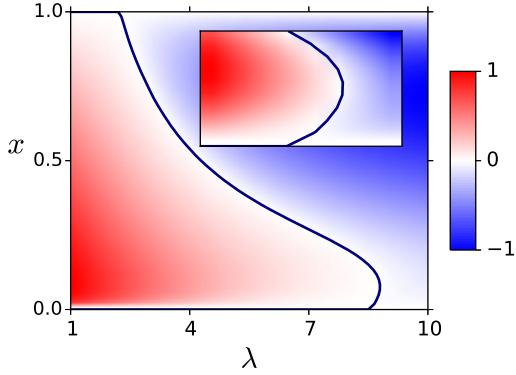
FIG. 4. Contour plots of Δx vs. x for $\Lambda = 4$ in the plane (x, λ). Inset shows a blow-up of the region near $\lambda = 8$, which exhibits features of both the bistable and coexistence regions.

The construction of the phase diagram of the main text is based on the condition of stability of the two fixed points $x = 0$ and $x = 1$. This treatment only gives a complete picture of the phase behavior if there are at most three fixed points. While this is true for most pairs (λ, Λ), notice that in the special case of Fig. 4a for $\Lambda = 4$ and near $\lambda = 8$, the curve turns back. In this region, there are four fixed points, with $x = 0$ and $0 < x^* < 1$ being stable. The novel aspect of the region near $\lambda = 8$ and x below 0.1 (shown in the inset as a blow-up), is that there is a bistability between points $x = 0$ and $x = x^*$ whereas in the phase diagram of the main text, the bistability only concerned points $x = 0$ and $x = 1$. For $x > 0.1$ and λ between approximately 2 and 8, we have a standard coexistence phase.

9

VI. COMPARISON BETWEEN LINEAR AND NON-LINEAR SELECTION FUNCTION

In the main text, we have introduced the following selection function

$$f(\bar{x}) = 0.5 \left(1 + \tanh \left(\frac{\bar{x} - x_{th}}{x_w} \right) \right), \tag{25}$$

with $x_{th} = 0.25$ and $x_w = 0.1$, which is now represented as the blue solid line in fig. 5. Note that this selection function takes a small but non-zero value for $x = 0$, namely $0.5(1 - \tanh(x_{th}/x_w)) = 0.0067$, which represents the fraction of false positives in the selection process.
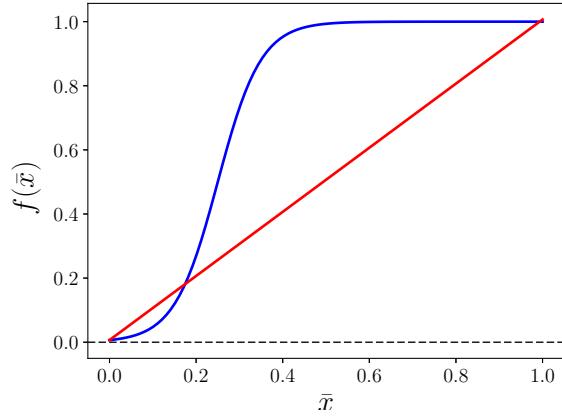
FIG. 5. Representation of the two selection functions used in this section, namely the one defined in the main text (red solid line) and a linear one (blue solid line), which approximately take the same values at $x = 0$ and $x = 1$. The horizontal dashed line represents the level of false positives given by $f(0)$.

It is interesting to compare the phase diagram given in the main text with that obtained for a linear selection function, f_{lin} shown as the red solid line in fig. 5. We choose $f_{lin}(\bar{x}) = 0.0067 + \bar{x}$, such that we have the same approximate values for $f(0)$ and $f(1)$ as with the previous function defined in Eq. (25). Consequently, we expect to find the same vertical asymptotes at $\lambda \simeq 6.95$ and $\lambda \simeq 149$, as confirmed in fig. 6. The equations of these vertical asymptotes are indeed only a function of $f(0)$ and $f(1)$. They are

$$\lambda f(0)e^{\lambda} = (e^{\lambda} - 1)f(1), \tag{26}$$

10

for the boundary between the bistable and parasite regions, and

$$\lambda f(1) = (e^\lambda - 1)f(0), \tag{27}$$

for the boundary between the bistable and ribozyme regions.

When comparing the two phase diagrams obtained with the non-linear and linear selection functions shown in fig. 6, we observe a similar general structure except for the center of the diagram and for the two asymptotes for Λ close to 1. This is to be expected for the center region where none of the simple approximations hold. Concerning the asymptotes near $\Lambda = 1$, as shown in the main text they represent boundaries between the coexistence and parasite regions (resp. coexistence and ribozyme regions) and they depend on the logarithmic derivative of the selection function near $x = 0$ (resp. $x = 1$).
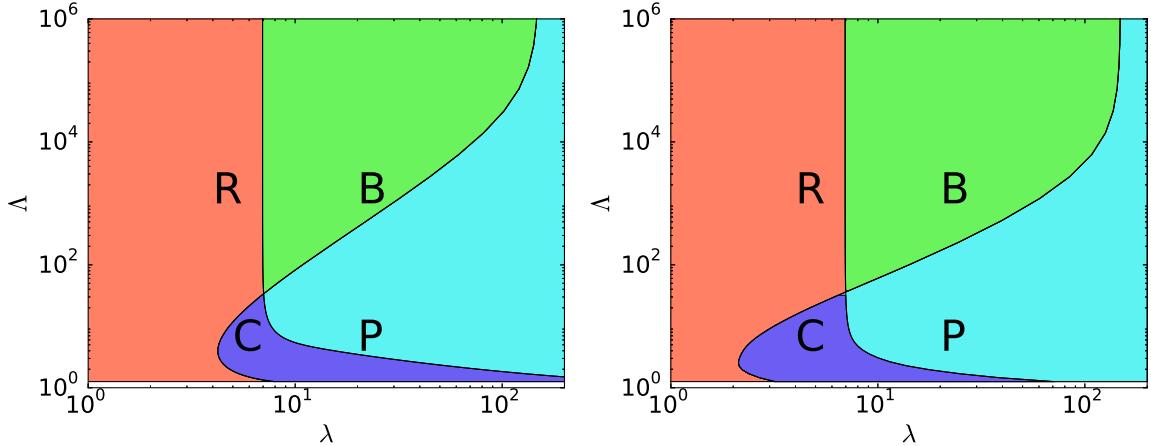
FIG. 6. Left: Phase diagram of the transient compartmentalization dynamics with the linear selection function $f(\bar{x}) = 0.0067 + \bar{x}$ in the (λ, Λ) plane. Right: idem with the non-linear selection function represented in Fig. 5. Phases are R: pure Ribozyme, B: Bistable, C: Coexistence and P: pure Parasite.

[1] S. Matsumura, A. Kun, M. Ryckelynck, F. Coldren, A. Szilágyi, F. Jossinet, C. Rick, P. Nghe, E. Szathmáry, and A. D. Griffiths, Science **354**, 1293 (2016).