

# Species Abundance Patterns in Complex Evolutionary Dynamics

Kei Tokita\*

*Program for Evolutionary Dynamics, Harvard University, One Brattle Square, Cambridge, Massachusetts 02138, USA*  
(Received 3 May 2004; published 20 October 2004)

An analytic theory of species abundance patterns (SAPs) in biological networks is presented. The theory is based on multispecies replicator dynamics equivalent to the Lotka-Volterra equation, with diverse interspecies interactions. Various SAPs observed in nature are derived from a single parameter. The abundance distribution is formed like a widely observed left-skewed lognormal distribution. As the model has a general form, the result can be applied to similar patterns in other complex biological networks, e.g., gene expression.

DOI: 10.1103/PhysRevLett.93.178102

PACS numbers: 87.23.-n, 75.10.Nr, 87.10.+e, 87.90.+y

If we investigate the number and populations of species in an ecosystem, we can observe universal characteristic patterns in that ecosystem. How to clarify the mechanisms underlying those species abundance patterns (SAPs) has been one of the “unanswered questions in ecology in the last century” [1] even though the knowledge obtained from it would affect vast areas of nature conservation. Various models have been applied to ecosystem communities where species compete for niches on a trophic level [2–16], but these models have left the more complex systems a mystery. Such systems occur on multiple trophic levels and include various types of interspecies interactions, such as prey-predator relationships, mutualism, competition, and detritus food chains. Although SAPs are observed universally in nature, their essential parameters have not been fully clarified.

I consider, then, a widely adopted model of biological networks represented by the so-called  $N$ -species replicator equation (RE) [17],

$$\frac{dx_i}{dt} = x_i \left( \sum_j J_{ij} x_j - \frac{1}{N} \sum_{j,k} J_{jk} x_j x_k \right) \quad (1)$$

to calculate the abundance  $x_i(t) (\in [0, N])$  of species  $i (= 1, 2, \dots, N)$ . Here we assume that  $(J_{ij})$  is a time-independent random symmetric ( $J_{ij} = J_{ji}$ ) matrix whose elements have a normal distribution with mean  $m(>0)$  and variance  $\tilde{J}^2/N$  as

$$P(J_{ij}) = \sqrt{\frac{N}{2\pi\tilde{J}^2}} \exp \left[ -\left( \frac{N}{2\tilde{J}^2} \right) (J_{ij} - m)^2 \right]. \quad (2)$$

Self-interactions are all set to a negative constant as  $J_{ii} = -u(<0)$ . Note that the essential parameter is unique as  $p \equiv (u + m)/\tilde{J}$  because the transformation of the interaction  $K_{ij} \equiv (J_{ij} - m)/\tilde{J}$  does not change the trajectory of the dynamics (1). Although ecologists do not generally believe in the randomness of interspecies interactions in nature, the discipline has been affected by the random interaction model [18] as a prototype of complex systems.

The RE appears in various fields [17]. In sociobiology, it is a game dynamical equation for the evolution of behavioral phenotypes; in macromolecular evolution, it is the basis of autocatalytic reaction networks (*hypercycles*); and in population genetics it is the continuous-time selection equation in the symmetric ( $J_{ij} = J_{ji}$ ) case. The symmetric RE also corresponds to a classical model of competitive community for resources [19].

Particularly in the context of ecology, the  $N - 1$  species Lotka-Volterra (LV) equation

$$\frac{dy_i}{dt} = y_i \left( r_i - \sum_j^{N-1} b_{ij} y_j \right) \quad (3)$$

is equivalent to the  $N$  species RE [17]. That is, the abundance  $y_i$  and the parameters in the corresponding LV equation are described by those in the present RE model as

$$y_i = x_i/x_M \quad (i = 1, 2, \dots, N), \quad (4)$$

$$r_i = J_{iM} - J_{MM} = J_{iM} + u, \quad (5)$$

$$b_{ij} = J_{ij} - J_{Mj}, \quad (6)$$

where the “resource” species  $M(y_M = 1)$  can be arbitrarily chosen from  $N$  species in the RE. The ecological interspecies interactions ( $b_{ij}$ ) ( $i \neq j$ ) have a normal distribution with mean 0 and variance  $2\tilde{J}^2/N$  from Eq. (6), and they are no longer symmetric ( $b_{ij} \neq b_{ji}$ ). The present model therefore describes an ecological community with complex prey-predator interactions [ $(b_{ij}, b_{ji}) \rightarrow (+, -)$  or  $(-, +)$ ], mutualism  $(+, +)$ , and competition  $(-, -)$ . Moreover, a community can have a “loop” (detritus) food chain [ $(b_{ij}, b_{ji}) \rightarrow (+, -)$ ,  $(b_{jk}, b_{kj}) \rightarrow (+, -)$ ,  $(b_{ki}, b_{ik}) \rightarrow (+, -)$ ]. The intraspecific interaction  $b_{ii}$  turns out to be related to the intrinsic growth rate  $r_i$  as  $b_{ii} = J_{ii} - J_{Mi} = -u - J_{Mi} = -r_i$  and is therefore competitive ( $b_{ii} < 0$ ) for producers ( $r_i > 0$ ) or mutualistic ( $b_{ii} > 0$ ) for consumers ( $r_i < 0$ ).

By Eq. (5), the intrinsic growth rates also have a normal distribution with mean  $u + m$  and variance  $\bar{J}^2/N$ . The probability at which  $r_i$  is positive—that is, that the  $i$ th species is a producer—is therefore given by the error function

$$\text{Prob}(r_i > 0) = \int_{-p\sqrt{N/2}}^{\infty} \frac{dt}{\sqrt{\pi}} \exp(-t^2). \quad (7)$$

Consequently, the parameter  $p$  can be termed as the “productivity” of a community because the larger the  $p$ , the greater the number of producers. The parameter  $p$  is also connected to the maturity of an ecosystem because  $m$  increases in time in an evolutionary model [20].

The symmetry ( $J_{ij} = J_{ji}$ ) makes the average fitness  $\bar{f} \equiv \sum_{j,k}^N J_{jk} x_j x_k$  [the second term of the right-hand side of Eq. (1)] a Lyapunov function [17], which is a non-decreasing function of time in dynamics (1). Therefore, every initial state converges to a local maximum of  $\bar{f}$  as  $t \rightarrow \infty$ . Interpreting  $\mathcal{H} \equiv -\frac{1}{2}\bar{f}$  as an energy function, we can study macroscopic functions such as free energy at such a maximum by using the technique of statistical mechanics of random systems [21–26].

Information on equilibrium states of dynamics (1) at  $t \rightarrow \infty$  is derived from the zero-temperature limit of free energy density,

$$f \equiv -\lim_{\beta \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{1}{N\beta} [\ln Z]_J, \quad (8)$$

$$Z \equiv \int_0^\infty \left( \prod_i^N dx_i \right) \delta\left(N - \sum_k^N x_k\right) e^{-\beta \mathcal{H}}, \quad (9)$$

where  $[\cdots]_J \equiv \int_{-\infty}^\infty dJ_{ij} P(J_{ij}) (\cdots)$  denotes the “sample average” [21] over random interactions. The Dirac delta function in Eq. (9) reflects the conservation of total abundance  $\sum_i^N x_i(t) = N$  satisfied at any  $t$  in Eq. (1). The calculation of Eq. (8) is similar to calculations in the previous works [22–26] and yields mean-field equations for the order parameters  $q$  and  $v$  as

$$p - v = \sqrt{q} \int_{-\Delta}^\infty (z + \Delta) D z, \quad (10)$$

$$(p - v)^2 = \int_{-\Delta}^\infty (z + \Delta)^2 D z, \quad (11)$$

where  $\Delta \equiv \sqrt{q}(p - 2v)$  and  $D z \equiv dz \exp(-z^2/2)/\sqrt{2\pi}$ . The resulting equations turn out to be formally the same as the case where  $m = 0$  and  $\bar{J} = 1$  [22]. For each value of  $p$ , Eqs. (10) and (11) are solved numerically.

Among macroscopic functions calculated in the present framework, the most significant for a theory of SAPs is the *survival function*  $\alpha_p(x) \equiv \frac{1}{N} \sum_i^N \theta(x_i - x)$ , the proportion of species whose abundance is larger than  $x$ , where  $\theta(z) \equiv [1(z > 0); 0(z \leq 0)]$  is the step function. Similar to the fashion in which free energy was calculated, the survival function  $\alpha_p(x)$  is analytically calcu-

lated and represented by the order parameters as

$$\begin{aligned} \alpha_p(x) &\equiv \lim_{\beta \rightarrow \infty} \lim_{N \rightarrow \infty} \left[ \int_0^\infty \left( \prod_j^N dx_j \right) \theta(x_i - x) \right. \\ &\quad \times \delta\left(N - \sum_k^N x_k\right) \frac{\exp(-\beta \mathcal{H})}{Z} \Big]_J \\ &= \alpha_p(0) \tilde{\theta}(x) - \int_{-\Delta}^\infty \tilde{\theta}\left(x - \frac{\sqrt{q}(z + \Delta)}{p - v}\right) D z, \end{aligned} \quad (12)$$

where the definition of the step function above is given by  $\tilde{\theta}(z) \equiv 1(z \geq 0); 0(z < 0)$ .

The resulting function  $\alpha_p(0) \equiv v(p - v)$  and  $\alpha_p(1)$  of  $p$  can be termed “diversity,” i.e., the proportion of non-extinct species and that of the species with abundance larger than unity, respectively, as depicted in Fig. 1. This demonstrates a typical positive correlation between productivity and diversity [27]. Numerical results for  $\alpha_p(1)$  are also depicted in Fig. 1 for comparison. We see good agreement between the analytical and the numerical results for  $p \gtrsim 1$ , while some deviations appear for small values of  $p$ . This small-value deviation is attributable to the occurrence of replica symmetry breaking (RSB) [21,23] for  $p < \sqrt{2}$ , which yields a number of metastable states of Eq. (8), and the replicator dynamics (1) essentially converges to not only a ground state of (8) but also to the metastable states. Since the energy  $\mathcal{H}$  and the diversity are both nonincreasing functions of time in dynamics (1), the mean-field results here give a lower minimum of diversity. Interestingly, the metastable states enhance the diversity. The analysis of RSB is expected to improve the quantitative agreement [23].

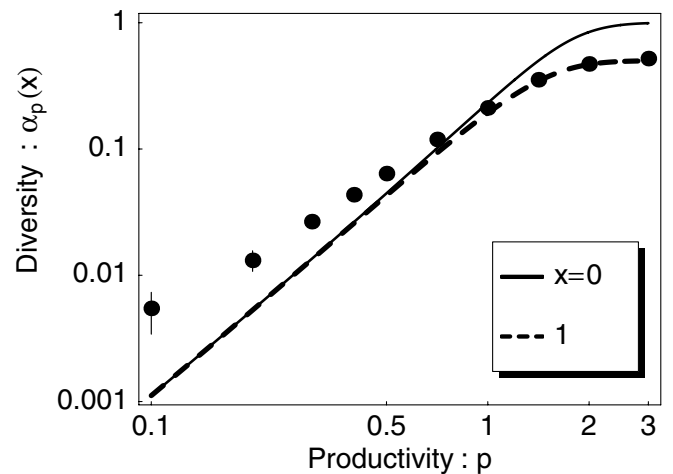


FIG. 1. Diversity  $\alpha_p(x=0, 1)$  as a function of  $p$  of log-log scales. Black circles show numerical solutions of  $\alpha_p(1)$  averaged over 50 samples of  $(J_{ij})$  for Eq. (1) with  $N = 2048$  and  $p = 0.1, 0.2, 0.3, 0.4, 0.5, \sqrt{2}/2, 1, \sqrt{2}, 2, 3$ . Error bars indicate the maximum and minimum values found in the samples.

Note that  $\alpha_p(x)$  is also represented as a function of species rank  $n$ :

$$\alpha_p(x) = \frac{n}{N} \quad \text{for } x \in [x^{(n+1)}, x^{(n)}] \quad (13)$$

( $n = 1, 2, \dots, S \leq N$ ) if the species abundance is ranked in descending order, as in  $x^{(1)} \geq x^{(2)} \geq \dots \geq x^{(n)} \geq \dots \geq x^{(S)} > 0$ . As the function  $\alpha_p(x)$  is a nonincreasing monotonic function, the species abundance relation, i.e., the abundance  $x^{(n)}$  as a function of a rank  $n$ , is given by the inverse function of  $\alpha_p(x)$  as  $x^{(n)} = x_p(n/N) = \alpha_p^{(-1)}(x)$ , depicted in Fig. 2 for some values of  $p$ . We observe two typical SAPs in different regions [12] and with different species compositions [6]: one is a straight line like the geometric series [2] for a small value of  $p$ , and the other consists of sigmoid curves on a logarithmic vertical axis for some range of  $p$ . This latter SAP denotes a lognormal-like abundance distribution. Remarkably, the transition of the SAPs from low  $p$  to high is identical to the observed transition from low- to high-productivity areas, that is, from a species-poor area such as an alpine or polar region to a species-rich tropical rain forest [12]. The transition also corresponds to the secular variation of SAPs observed in abandoned cultivated land [7]. This supports the contention that  $p$  (or  $m$ ) is a maturity parameter, as is suggested by an evolutionary model [20].

The abundance distribution is also derived from the survival function. As  $C_p(x) \equiv 1 - \alpha_p(x)$  is a cumulative distribution function of abundance, the abundance distribution is given by the derivative  $F_p(x) \equiv dC_p(x)/dx$  and

$$F_p(x) = \frac{p-v}{\sqrt{2\pi}q} \exp\left\{-\frac{(p-v)^2}{2q}\left(x - \frac{q(p-2v)}{p-v}\right)^2\right\} + C_p(0)\delta(x), \quad (14)$$

where the second term denotes the rate of species extinction. The first term is a normal distribution but not a lognormal distribution. Nevertheless, the curves in Fig. 2 demonstrate a typical sigmoid pattern on a logarithmic vertical axis. This pattern indicates the coexistence of very abundant species with rare ones. This multiscale of abundance is intuitively understood by a divergent behavior of the variance  $\sigma^2 \equiv q/(p-v)^2$  of  $F_p(x)$  for small  $p$  because  $q \rightarrow \infty$  and  $v \rightarrow 0$  for  $p \rightarrow 0$ . Moreover, the mode of  $F_p(x)$  per “natural” octave [5]  $\ln(x)$  is always a positive value (as shown in Fig. 3) at  $x^* = \frac{\sigma}{2}(\Delta + \sqrt{\Delta^2 + 4}) > 0$ , which denotes a unimodal distribution. Indeed, the mode diverges as

$$x^* \rightarrow \frac{\sigma}{|\Delta|} = \frac{1}{(p-v)|p-2v|} \rightarrow \infty \quad (15)$$

for  $p \rightarrow 0$ . As a result, the abundance distribution is a truncated normal distribution with a large variance  $\sigma^2 \rightarrow \infty$  and a negatively divergent mean  $\mu \equiv \frac{q(p-2v)}{p-v} \rightarrow -\infty$  satisfying  $\frac{\sigma}{\mu} = \frac{1}{\Delta} \rightarrow 0$  for  $p \rightarrow 0$ . This is why the abundance distribution per octave looks like a left-skewed lognormal distribution [10] in Fig. 3.

Moreover, we derive an analytical expression for the population  $x_{\max}$  of the most abundant species, the position of the individual curve mode  $x_N$ , and therefore the ratio of their logarithm defined as  $\gamma \equiv \log(x_N)/\log(x_{\max})$ . According to the *canonical hypothesis* [5,8], the parameter  $\gamma$  takes a value near unity in various real communities. To check the validity of the canonical hypothesis, we first need to evaluate an expected value of the most abundant species  $x_{\max}$ . From the definition of  $x_{\max}$ , that is  $NF_p(x_{\max}) = 1$ , and the conservation of the total abundance  $\int_0^\infty F_p(x)xdx = N$ , which is equivalent to

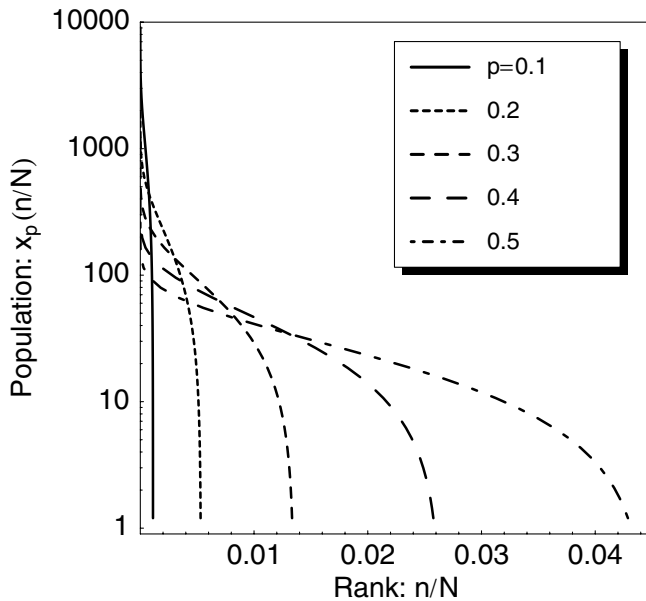


FIG. 2. Rank-abundance relations as a function of productivity  $p$  on normal-log scales.

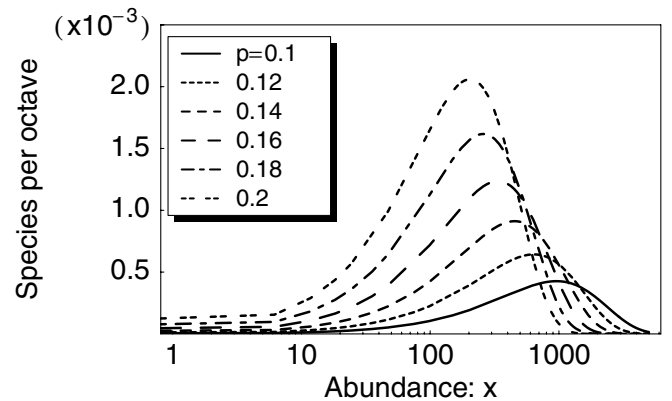


FIG. 3. Abundance distribution per natural octave  $\ln(x)$ . Functions  $F_p(x)x\{d[\ln(x)] = F_p(x)dx\}$  for some values of  $p$  are depicted, whereas Preston originally defined octaves as logarithms to base 2 [5].

$\sum_i^N x_i = N$ , we obtain

$$x_{\max} = \frac{q(p - v) + \sigma \sqrt{2 \ln \left( \frac{\sigma[1 - \alpha_p(0)] + \Delta \alpha_p(0)}{\sqrt{2\pi}} \right)}}{p - v}. \quad (16)$$

On the other hand, the mode of the individual curve  $F_p(x)x$  per octave is given by  $x_N = \frac{\sigma}{2}(\Delta + \sqrt{\Delta^2 + 8})$ , and finally, the parameter  $\gamma \equiv \log(x_N)/\log(x_{\max})$  is evaluated by substituting the values of the order parameters  $q$  and  $v$  for each value of  $p$ . In the present model,  $\gamma$  is a monotonically increasing function of  $p$  and  $0.96 < \gamma < 1.04$  for  $0.1 < p < 0.6$ , denoting that the canonical hypothesis is supported in the range of  $p$  giving the typical SAPs in Fig. 3. Although the canonical hypothesis was demonstrated to be merely a mathematical consequence of lognormal distribution [8] rather than anything biological, it is noteworthy that the lognormal-like abundance distribution with  $\gamma \approx 1$  derives from basic ecological dynamics. This still suggests a biological foundation for the hypothesis in a large complex ecosystem, in the same way that a biological foundation was indicated for the theory of a local competitive community [9].

In the present model, all species coexist only in the limit  $p \rightarrow \infty$ , that is, in the trivial cases in which interspecies interactions are negligible ( $\tilde{J} \ll u$ ) or homogeneous ( $\tilde{J} \rightarrow 0$ ), thereby giving  $\alpha_\infty(x) = \theta(1 - x)$ ,  $x^{(n)} = x_\infty(n/N) = 1$  for all  $n$  and  $F_\infty(x) = \delta(x - 1)$ .

The present theory seeks to capture the influence of productivity on the SAPs under the assumption that all species interact randomly; nevertheless, this assumption itself is never justified because it ignores a biological correlation between interactions produced by evolution. However, note that the randomness is assumed only for an initial state with  $N$  species in Eq. (1). Actually, the simulation reveals the resulting interactions of nonextinct species to be nonrandom: every sample for  $p \in [0.1, 3]$  in Fig. 1 evolves to only flora,  $\forall i, r_i > 0$ . Moreover, by ordering species as  $r_i > r_j$  for any  $i < j$ , we observe a hierarchy: there are only three types of interactions, that is, mutualism ( $b_{ij}, b_{ji} = (+, +)$ ), competition ( $-, -$ ), and exploitation of  $i$  on  $j$  as  $(+, -)$ , but no reverse  $(-, +)$ . This suggests the applicability of the present model to a plant community.

It has been demonstrated that empirically supported patterns are derived from a single parameter of general population dynamics. This not only suggests the importance of globally coupled biological interactions in a large assemblage but also provides a unified viewpoint on mechanisms of similar patterns observed in other biological networks with complex interactions, for example, a lognormal abundance distribution of a protein in cells [28–30], which is revealed by gene expression networks.

The author thanks R. Frankham, Y. Iwasa, E. Matsen, R. May, M. Nowak, and J. Plotkin for their helpful comments. This work was supported by Grants-in-Aid from

MEXT, Japan.

\*Permanent address: Large-Scale Computational Science Division, Cybermedia Center, Osaka University, 1-32 Machikaneyama-cho, Toyonaka, Osaka 560-0043, Japan.

Electronic addresses: tokita@cmc.osaka-u.ac.jp

http://www.cp.cmc.osaka-u.ac.jp/~tokita

- [1] R. M. May, *Philos. Trans. R. Soc. London B* **264**, 1951 (1999).
- [2] I. Motomura, *Zool. Mag. (Tokyo)* **44**, 379 (1932).
- [3] A. S. Corbet, R. A. Fisher, and C. B. Williams, *J. Anim. Ecol.* **12**, 42 (1943).
- [4] R. H. MacArthur, *Am. Nat.* **94**, 25 (1960).
- [5] F. W. Preston, *Ecology* **43**, 410 (1962).
- [6] R. H. Whittaker, *Communities and Ecosystems* (Macmillan, New York, 1970).
- [7] F. A. Bazzaz, *Ecology* **56**, 485 (1975).
- [8] R. M. May, in *Ecology and Evolution of Communities*, edited by M. L. Cody and J. M. Diamond (Belknap, Cambridge, 1975), pp. 81–120.
- [9] G. Sugihara, *Am. Nat.* **116**, 770 (1980).
- [10] S. Nee, P. H. Harvey, and R. M. May, *Proc. R. Soc. London B* **243**, 161 (1991).
- [11] M. Tokeshi, *Species Coexistence* (Blackwell, Oxford, 1999).
- [12] S. P. Hubbel, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, 2001).
- [13] M. Hall, K. Christensen, S. A. di Collabiano, and H. J. Jensen, *Phys. Rev. E* **66**, 011904 (2002).
- [14] B. J. McGill, *Nature (London)* **422**, 881 (2003).
- [15] I. Volkov, J. R. Banavar, S. P. Hubbel, and A. Maritan, *Nature (London)* **424**, 1035 (2003).
- [16] S. Pigolotti, A. Flammini, and A. Maritan, *Phys. Rev. E* **70**, 011916 (2004).
- [17] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, 1998).
- [18] R. M. May, *Nature (London)* **238**, 413 (1972).
- [19] R. MacArthur and R. Levins, *Am. Nat.* **101**, 377 (1967).
- [20] K. Tokita and A. Yasutomi, *Theor. Pop. Biol.* **63**, 131 (2003).
- [21] M. Mezard, G. Parisi, and A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [22] S. Diederich and M. Oppen, *Phys. Rev. A* **39**, 4333 (1989).
- [23] P. Biscari and G. Parisi, *J. Phys. A* **28**, 4697 (1995).
- [24] V. M. de Oliveira and J. F. Fontanari, *Phys. Rev. Lett.* **85**, 4984 (2000).
- [25] V. M. de Oliveira and J. F. Fontanari, *Phys. Rev. E* **64**, 051911 (2001).
- [26] V. M. de Oliveira and J. F. Fontanari, *Phys. Rev. Lett.* **89**, 148101 (2002).
- [27] R. B. Waide *et al.*, *Annu. Rev. Ecol. Syst.* **30**, 257 (1999).
- [28] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, *Nature (London)* **422**, 633 (2003).
- [29] K. Kaneko, *Phys. Rev. E* **68**, 031909 (2003).
- [30] K. Sato, Y. Ito, T. Yomo, and K. Kaneko, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14086 (2003).