Journal of Theoretical Biology 357 (2014) 26–34

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Multispecies population dynamics of prebiotic compositional assemblies

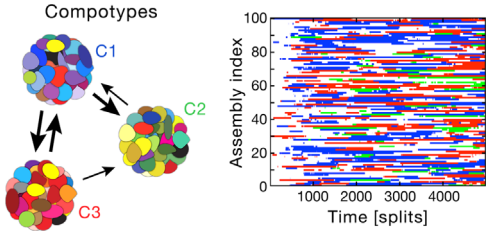
Omer Markovitch, Doron Lancet*

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

HIGHLIGHTS

- Population dynamics of compositional assemblies studied.
- Species dynamics resemble that of natural populations.
- Dynamics analyzed by Lotka–Volterra model.
- Species carrying capacity correlates with molecular repertoire size.
- Species growth rate inversely correlates with molecular repertoire size.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 October 2013
Received in revised form
14 April 2014
Accepted 1 May 2014
Available online 14 May 2014

Keywords:
Metabolism first
Lipid world
Composomes
Origin of life

ABSTRACT

Present life portrays a two-tier phenomenology: molecules compose supramolecular structures, such as cells or organisms, which in turn portray population behaviors, including selection, evolution and ecological dynamics. Prebiotic models have often focused on evolution in populations of self-replicating molecules, without explicitly invoking the intermediate molecular-to-supramolecular transition. Here, we explore a prebiotic model that allows one to relate parameters of chemical interaction networks within molecular assemblies to emergent population dynamics. We use the graded autocatalysis replication domain (GARD) model, which simulates the network dynamics within amphiphile-containing molecular assemblies, and exhibits quasi-stationary compositional states termed compotype species. These grow by catalyzed accretion, divide and propagate their compositional information to progeny in a replication-like manner. The model allows us to ask how molecular network parameters influence assembly evolution and population dynamics parameters. In 1000 computer simulations, each embodying different parameter set of the global chemical interaction network parameters, we observed a wide range of behaviors. These were analyzed by a multi species logistic model often used for analyzing population ecology (r–K or Lotka–Volterra competition model). We found that compotypes with a larger intrinsic molecular repertoire show a higher intrinsic growth (r) and lower carrying capacity (K), as well as lower replication fidelity. This supports a prebiotic scenario initiated by fast-replicating assemblies with a high molecular diversity, evolving into more faithful replicators with narrower molecular repertoires.

1. Introduction

The origin of life is perhaps the most exhaustive systems chemistry experiment (Kiedrowski et al., 2010). The path from organic mixtures (i.e., the primeval "soup") to reproducing lifelike protocells has traditionally been dominated by the genetic, or replicator-first, approach, and the metabolism-first approach, both requiring the emergence of replicating entities capable of undergoing Darwinian evolution (Anet, 2004; Bernhardt, 2012; Orgel, 2004; Rasmussen et al., 2009). Models often focus on the emergence of self-replicating polymers (e.g., RNA), or on the evolution of populations of such entities. Yet, less attention is devoted to

* Corresponding author.
E-mail addresses: omermar@weizmann.ac.il (O. Markovitch), doron.lancet@weizmann.ac.il (D. Lancet).

http://dx.doi.org/10.1016/j.jtbi.2014.05.005
0022-5193/

studying the intermediate stage of molecules-to-population dynamics. Of note, a similar transition has been studied in a model of RNA-like replicators, in which supramolecular entities (traveling waves) were found to play a role in the ecology and evolution of replicators (Takeuchi et al., 2008). Present life portrays a two-tier phenomenology: molecules compose self-replicating supramolecular structures such as cells or organisms, which in turn portray population behavior, including selection, evolution. Thus, understanding how molecular mixtures gave rise to evolving entities will greatly contribute to our understanding of the origin of life and to a degree akin to the on-going pursuit to understand and predict the dynamics of ecological populations from the, often complex, metabolic or genetic networks of the underlying species (Bailey et al., 2009).

The GARD model of the lipid world scenario for the origin of life is an embodiment of the metabolism-first approach (Segre et al., 2001a). GARD is a systems-chemistry model which entails supramolecular assembly of amphiphiles with intrinsic network of mutually catalytic interactions, which has been shown to portray a capacity to undergo replication mediated by homeostatic growth, as well as selection and evolution (Markovitch and Lancet, 2012; Segre et al., 2000). Its species-like quasi-stationary states in compositional space are called composomes. GARD's forte is in invoking a supramolecular structure that is on the one hand replicable and evolvable and on the other hand simple enough so that the molecular parameters may be directly and quantitatively related to resulting dynamics.

An admitted shortcoming of GARD is the paucity of experimental verification of many of its predictions. A proof-of-principle experiment should address the question of whether vesicles are capable of homeostatic growth and even rudimentary transfer of compositional information to fission-generated progeny. Such experiments would require complex setups, accurate compositional monitoring of individual amphiphile assemblies, not yet fully elaborated. A promising lead would be the recent experimental exploration of multi-component vesicles (Maurer et al., 2009; Vequi-Suplicy et al., 2010). Another critique of GARD asserts that it simulates abstract molecules without specified chemical properties. This point has been recently addressed in an extension of the simulated model to incorporate realistic physicochemical properties of amphiphilic molecules, showing that a measure of compositional heredity may be observed (Armstrong et al., 2011).

GARD simulations are used here to quantitatively follow population dynamics of composmal species. In the foregoing analyses a multivariate logistic equation is used to relate systems chemical parameters of GARD assemblies, including chemical diversity, replication fidelity and compositional similarity to specific ecology-like measures such as the carrying capacity, the intrinsic growth rate and the competition parameters.

2. Methods

GARD is a kinetic model which describes the growth and fission of a non-covalent molecular assembly, typically assumed to consist of amphiphilic molecules (Markovitch and Lancet, 2012; Segre et al., 2000). Molecules from a buffered environment form and join an aggregate and molecules within it can leave, and once the assembly reaches a pre-defined size a random fission action is applied to produce two progenies of same size which can grow again and again in growth-fission cycles. Assembly dynamics is described by the set of ordinary differential equations

$$\frac{dn_i}{dt} = (k_f \rho_i N - k_b n_i)\left(1 + \sum_{j=1}^{N_G} \beta_{ij} \frac{n_j}{N}\right) \quad (1)$$

where for a molecular type i, n_i is its count within an assembly, ρ_i is its environmental concentration and β_{ij} is the rate-enhancement exerted by an assembly molecule of type j on incoming or outgoing molecule of type i. k_f and k_b are the basal forward and backward rate constants (joining and leaving, respectively), N_G is the size of environmental repertoire and N is the current assembly size. The simulations are done by the stochastic Gillespie algorithm (Gillespie, 1976), as described (Segre et al., 2000).

Addressing population dynamics, the chemical dynamics of GARD compositional assemblies in a reactor under constant population conditions are simulated in a buffered environment with a repertoire of $N_G = 100$ amphiphilic molecular types (Markovitch and Lancet, 2012). Assembly growth is controlled by its molecular composition and by β network of mutually-catalytic parameters (with N_G nodes and ${N_G}^2$ edges; Eq. (1)). Simulated fission occurs when the number of molecules in the assembly reaches a predetermined maximal value $N_{max} = 100$ (Markovitch and Lancet, 2012; Segre et al., 2000). This is a simplified representation of a physical process whereby larger assemblies are more prone to fission due to surface tension and shearing effects.

A group of similar composomes, gleaned by K-means clustering algorithm (Shenhav et al., 2007), is termed a compotype, and may be regarded as species in the framework of lipid world and GARD. The pre-fission composition of each assembly is assessed as belonging to one of the N_C compotype species characterizing the specific β or to "drift", where the latter represents random compositions rather than faithfully replicating one. Compotypes are found by subjecting a simpler mode of the simulation to the clustering algorithm (Markovitch and Lancet, 2012). Each chemostat reactor is seeded with $L_{pop} = 1000$ random compositions. Assemblies are then allowed to simultaneously grow based on their idiosyncratic kinetic parameters (Eq. (1)), which effectively means of degree of competition as a result of different growth rates due to different compositions. When one of the assemblies reaches the predefined maximal size, it undergoes random fission and a randomly chosen assembly from among the remaining 999 assemblies is removed, thus keeping the population size constant. This protocol is based on the classical Moran process (Moran, 1958). Each simulation is performed for 50,000 split events in the reactor, typically sufficient to reach steady state, and data is saved every 10 split events. For statistical rigor, 1000 such simulations performed, each with a different β randomly drawn from a lognormal distribution (Segre et al., 2001b). The simulations give rise to a total of 1859 compotypes. The GARD10 MATLAB code is employed for all simulations, using parameter values identical to those previously employed (Markovitch and Lancet, 2012).

The size of a compotype's intrinsic molecular repertoire (N_{mol}) is measured as the number of molecule types (out of the total N_G types) whose fractional counts in a compotype are bigger than 1.0. Compotype replication fidelity (F_{rep}) and speed (t_{rep}) are measured using the following method: an assembly with exactly the same composition as in the compotype's center of mass (rounded to nearest integer) is used as parent. This parent than undergoes split and each of the two progeny is grown according to its idiosyncratic composition (Eq. (1)) until it reaches the maximal size. This is repeated for 4000 times, each time beginning with the same parent, giving a total of 8000 fully grown progeny of this compotype. F_{rep} of a compotype is than defined as the average compositional similarity (measured as the dot product between two vectors representing the compositions of two assemblies; Segre et al., 2000) between the fully grown progeny to the parent. This is an extension to a previous analysis, where the fidelity was assessed based only on the split action (Segre et al., 2001b). Each event of a molecule joining (or leaving) the assembly has a rate, and the total growth time of each progeny is the sum of 1 over each of these rates. t_{rep} of a compotype is than defined as the

average growth time of all grown progeny who are highly similar to this compotype.

3. Results, discussion and conclusions

Different simulations with different underlying β networks result in widely different dynamic behaviors, such as delayed growth, different plateaus and "takeover" of a fast-rising compotype by a slower one (Figs. 1 and A.1). Such dynamics are typical of natural ecosystems that harbor multiple species with competition or predator–prey relationships. The resulted dynamics are analyzed by a multi species logistic model for population ecology (r–K or Lotka–Volterra competition model) (Gabriel et al., 2005; Strobeck, 1973; Vandermeer, 1975), as per the equation

$$\frac{dC_i}{dt} = r_i C_i \left(\frac{K_i - C_i - \sum_{i \neq j}^{N_C} \alpha_{ij} C_j}{K_i} \right) \quad (2)$$

where for compotype species i, C_i is its population fraction (out of 1000 assemblies) at time t (time measured in split events), r_i is the intrinsic growth rate, K_i is the carrying capacity (compotype maximal fraction in the absence of competition) and α_{ij} is the extent of competition exerted by compotype j on compotype i. This equation has a steady state $C_i^{ss} = K_i - \sum a_{ij} C_j^{ss}$ and can be solved analytically only for the case of a single species ($N_C = 1$). For each simulation, the $N_C{}^2 + 2N_C$ logistic parameters for all N_C compotypes are obtained by least square fitting and numerical integration, with the following constraints: $r_i > 0$, $0 < K_i \leq 1.0$, $0 \leq \alpha_{ij} \leq 10.0$ and $0 < C_i(0) < 1.0$ (see Appendix B). Notably, an adequate fit to such equation was observe for practically all GARD simulations performed, with average root mean square difference $= 0.019 \pm 0.011$. In contrast, several other models with similar overall characteristics gave an inferior fit (Fig. B.1). We note that the focus of the present manuscript is the full-fledged dynamics of GARD populations, often with more than one species. This is unlike the quasispecies model, typically aimed towards a population of sequences stemming from a single master sequence (Biebricher and Eigen, 2006; Eigen and Schuster, 1977), which we are currently applying to GARD in a separate study.

We subsequently perform analyses aimed at relating the chemistry-base molecular parameters of GARD to the parameters of the logistic equation. Each GARD compotype contains a subset (N_{mol}) of the total N_G molecular types present in the environment. Such repertoire restriction occurs as a result of the intermolecular catalytic interactions in β and in the present simulations an average of $N_{mol} = 16 \pm 5$ is obtained (Fig. 2). The effect of this chemical diversity parameter, N_{mol}, on the K_i and r_i values of individual compotypes is examined (Fig. 3). It is found that K values are inversely correlated with N_{mol}, and in contrast, r values show a weak positive correlation. Thus, compotypes with a large N_{mol} will tend to have a larger growth rate and a smaller carrying capacity. For cases with negligible competition parameters this will amount to a steeper ascent and a relatively low plateau in cases of large N_{mol}.

The dependence of K and r on N_{mol} may be explained considering the random nature of the processes involved and the fact that external concentrations of all molecular types are equal. A higher value of N_{mol} increases the probability that a randomly impinging molecule will be part of the compotype's intrinsic molecule repertoire, enhancing homeostatic growth rate. Conversely, low N_{mol} means that on average every molecular type exists inside the compotype in higher counts, so when split occurs there is a better chance that a progeny will contain the same composition as the parent.

One may ask, whether there could be a parallelism for any of these results in present day life. An interesting analogous trend was observed in experimental data for 113 bacteria, whereby a negative correlation was seen between measured doubling time and metabolic network size (Freilich et al., 2009). However, direct comparison between compotype dynamics and present-day metabolism might not be possible, as the latter is controlled by a genome, centralized informational entity acting via a complex hierarchy of interactions (Heinemann and Sauer, 2010; Sauer, 2006).

Next, the relation between K and the replication fidelity (F_{rep}) is analyzed. F_{rep} measures the average degree of compositional similarity between an assembly representing a compotype and its progeny, both in fully-grown state. When analyzing ecology, K,

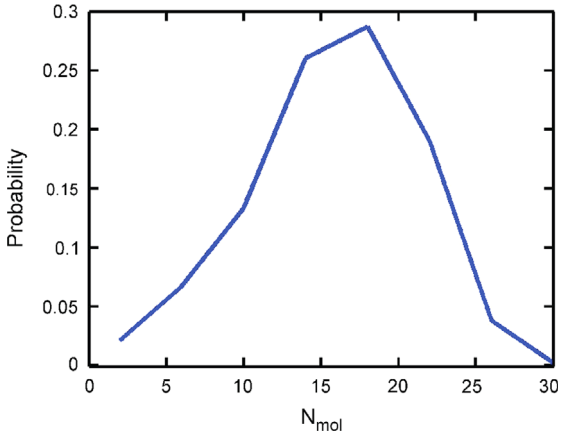


Fig. 2. Distribution of intrinsic molecular repertoire size (N_{mol}) values for all compotypes.

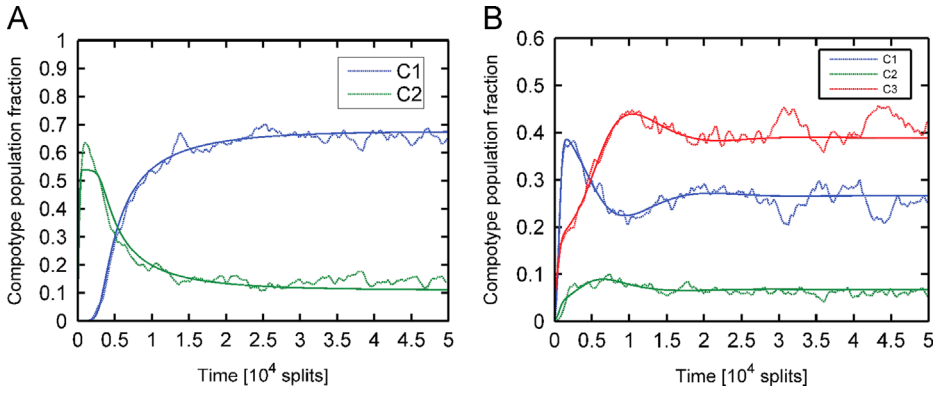


Fig. 1. Examples of GARD simulations and fit to logistic growth. Simulation data is broken line and fit is solid line. The two simulations respectively exhibit 2 and 3 compotype species (A and B). Panel (A) shows an example of a takeover. Fitted parameters are collected in Table A.1.

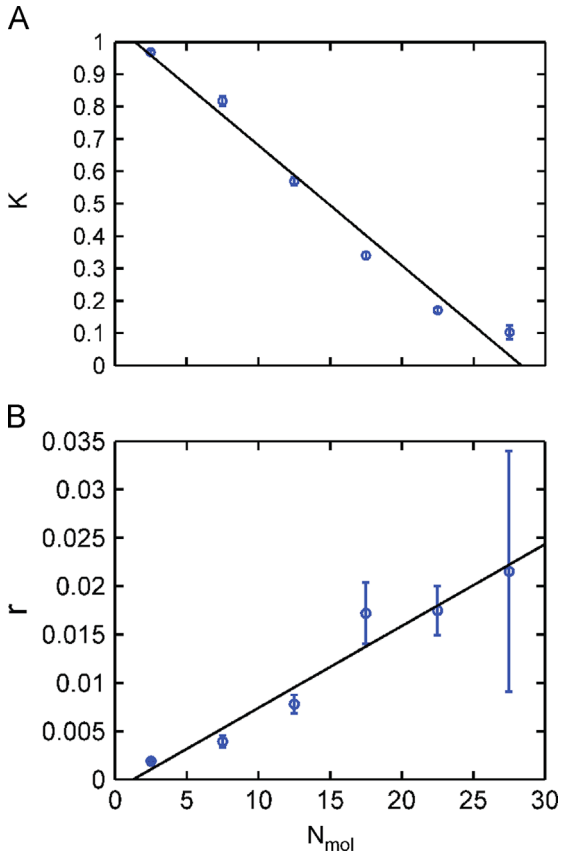


Fig. 3. The dependence on N_{mol}. Data are binned according to N_{mol} values and vertical lines represent standard error of the mean. Black solid line is a linear fit. (A) K vs. N_{mol}. Linear fit: $K=-0.0371N_{mol}+1.052$, $R^2=0.978$ and (B) r vs. N_{mol}. Linear fit: $r=8.46\times10^{-4}N_{mol}-1.06\times10^{-3}$, $R^2=0.946$.

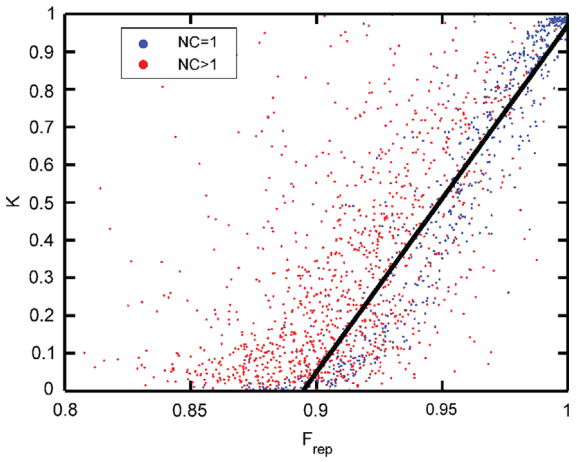


Fig. 4. K vs. F_{rep}. Blue and red dots are compotypes taken from simulations exhibiting $N_C=1$ and $N_C>1$, respectively. Black solid line is linear fit to $N_C=1$ data: $K=9.23F_{rep}-8.23$, $R^2=0.876$.

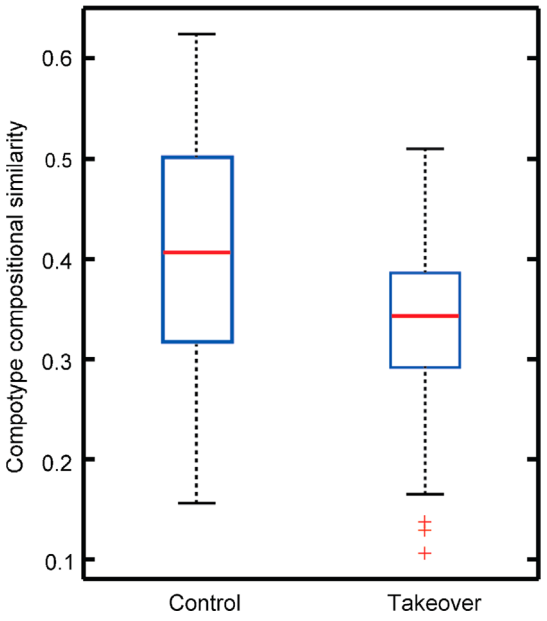


Fig. 5. Compotype similarity in the takeover and control groups. Average cross-compotype similarity (H; Segre et al., 2000) for the takeover group is lower than for the control group (0.33 ± 0.09 vs. 0.40 ± 0.12; supported by a two-sample Kolmogorov–Smirnov test with p-value$=1.3\times10^{-3}$).

the carrying capacity, represents the maximal number of individuals of a given species that may be sustained in an ecological niche. In the original Verhulst formalism, death was introduced to counter the Malthusian exponential growth. Later, the r–K logistic formalism defined K=birth/death (Gabriel et al., 2005). In GARD, a positive correlation between K and F_{rep} is observed (Fig. 4). Unfaithful replication (low F_{rep}) means that the progeny has lost its compotype state, either to another compotype species or to drift, somewhat comparable to death of the species in question. This may rationalize the somewhat unexpected positive correlation between an emergent molecular parameter such as F_{rep} and the logistic equation's carrying capacity. Other relationships explored, between r and F_{rep} and between K and r and the replication-time (t_{rep}) showed no appreciable correlations (Figs. C.1 and C.2).

Lastly, the molecular mechanism behind the takeover was addressed. This phenomenon is exemplified in Fig. 1A by the observation that compotype C_2 shows a much faster ascent, reaching a 538 fold excess over C_1 at time 990. Subsequently, C_1 increases substantially, becoming 5.82 fold more abundant than C_2 at steady state. This was examined by analyzing an extended set of 316 β networks that exhibit $N_C=2$. Two parameters, MP and PP, were defined to quantify takeover behavior: MP=Max(C_{low})/Plateau(C_{low}), PP=Plateau(C_{high})/Plateau(C_{low}), where C_{low} is the compotype with the lower plateau (Fig. D.1). Two subgroups were examined: one showing clear takeover, with MP > 2 and PP > 5, and another in which no takeover occurs, with MP < 1.5 and PP < 4 (control). The inter-compotype compositional similarity for pairs that exhibit takeover is found to be significantly lower than for control pairs (Fig. 5). These two behavior types are also seen to be partially segregated in a principle component analysis of the 6 fitted logistic parameters (r_1, r_2, K_1, K_2, α_{12} and α_{21}) (Fig. D.2). Intriguingly, the majority of the variance in this plot is contributed by the competition parameter α_{21} (Fig. D.3). Work is underway to study the molecular mechanism that governs across-compotype competition.

In a majority of the cases the logistic (or Lotka–Volterra) formalism is used for cases such as predator–prey or inter-species competition for resources. Should one use such formalism in the case of GARD populations? These populations are characterized by a different dynamics, whereby species interconvert into each other. In GARD analyses, some of the logistic equation parameters are thus interpreted differently than in classical ecology. The carrying capacity (K) is related to the chance that a compotype will produce progeny belonging to the parent's

compotype (and not to drift, or another compotype). Hence, K is related to the replication fidelity of a given species, independent of environmental parameters. Specifically, in the simulations presented here there is no competition for resources as the environment is buffered. GARD's α_{ij} parameter measures the extent of species inter-conversion, made possible by the fact that every compotype is a sub-network of the global β network. Thus, the forgoing results could seed a better understanding of early evolution, whereby protocellular entities were sufficiently malleable so as to reveal aspects of ecology-like population dynamics.

We utilize here the logistic equation to fit the dynamic behavior of GARD compotypes. This equation can show oscillations for certain parameter ranges (May, 1974; Rescigno, 1967, 1968). Notably, in 1000 different sets of fitted logistic parameters here no oscillatory behavior observed. It is important, though, that such parameters are derived from chemical rate-enhancement values embodied in beta. Future analyses could give insights into the conditions for the existence of stationary states vs. oscillations.

Acknowledgments

We thank Raphael Zidovetzki for fruitful discussions, and Paul Higgs and Avi Mayo for insights on the analyses. We also thank the reviewers for their comments and suggestions. This work is partly supported by EU-FP7 project MATCHIT, the Crown Human Genome Center and the Minerva Center for Life Under Extreme Planetary Conditions, both at the Weizmann Institute of Science.

Appendix A

See Fig. A.1 and Table A.1.

Appendix B

Fit procedure is as follows:

(1) $C_i(t)$ data of each compotype is smoothed 100 times by a 5-point moving average.
(2) In order to avoid over sampling of the long times over the short times, the fit time range is until twice the time the variance of the data (for each time point, the variance is calculated until that point) drops below half its maximal size, plus 100 points along the tail. This is calculated for each compotype individually and then the longest time of all compotypes in each simulation is picked.
(3) Compotypes with $\langle C_i \rangle < 0.01$ are ignored and their assemblies classified as drift.
(4) For simulation with $N_C=1$, if the time curve exhibits a plateau lower than the maximum by more than 20%, then this simulation is ignored.
(5) MATLAB *lsqcurvefit* is used to perform least-squares curve fitting with the following function parameters (the rest are at their default values): TolFun$=1e-10$; TolX$=1e-10$; MaxFunEval$=200N_C(N_C+1)$; MaxIter$=1000$.
(6) *ode45* and *ode15s* ordinary differential equation solvers are used to numerically solve the logistic equation (see main text), and the fit with the lowest residuals is considered. The following function parameters are used and the rest are at their default values: AbsTol$=1e-10$; RelTol$=1e-10$.
(7) Initial parameter guesses are: $K_i=\max(C_i)$; $a_{ij}=0.1$; $r_i=\sum_{t=1}^{100} dC_i/100$ (dC_i is approximated by 5th order numerical differentiation); $C_i(0)=\text{mean}[C_i(1...100)]$.
(8) Constraints are $r_i>0$, $0\le K_i\le 1.0$, $0\le a_{ij}\le 10.0$, $0\le C_i(0)\le \max(C_i)$.

The root-mean-squared-difference (RMSD) is defined as

$$\text{RMSD}=\sqrt{\left\langle\left(\sum_{i=1}^{N_C}[f_i(t)-C_i(t)]\right)^2\right\rangle} \tag{B.1}$$

where f_i is the fitted equation for compotype i in a simulation, and $<\ldots>$ marks an average over all time-points in that simulation (Fig. B.1).

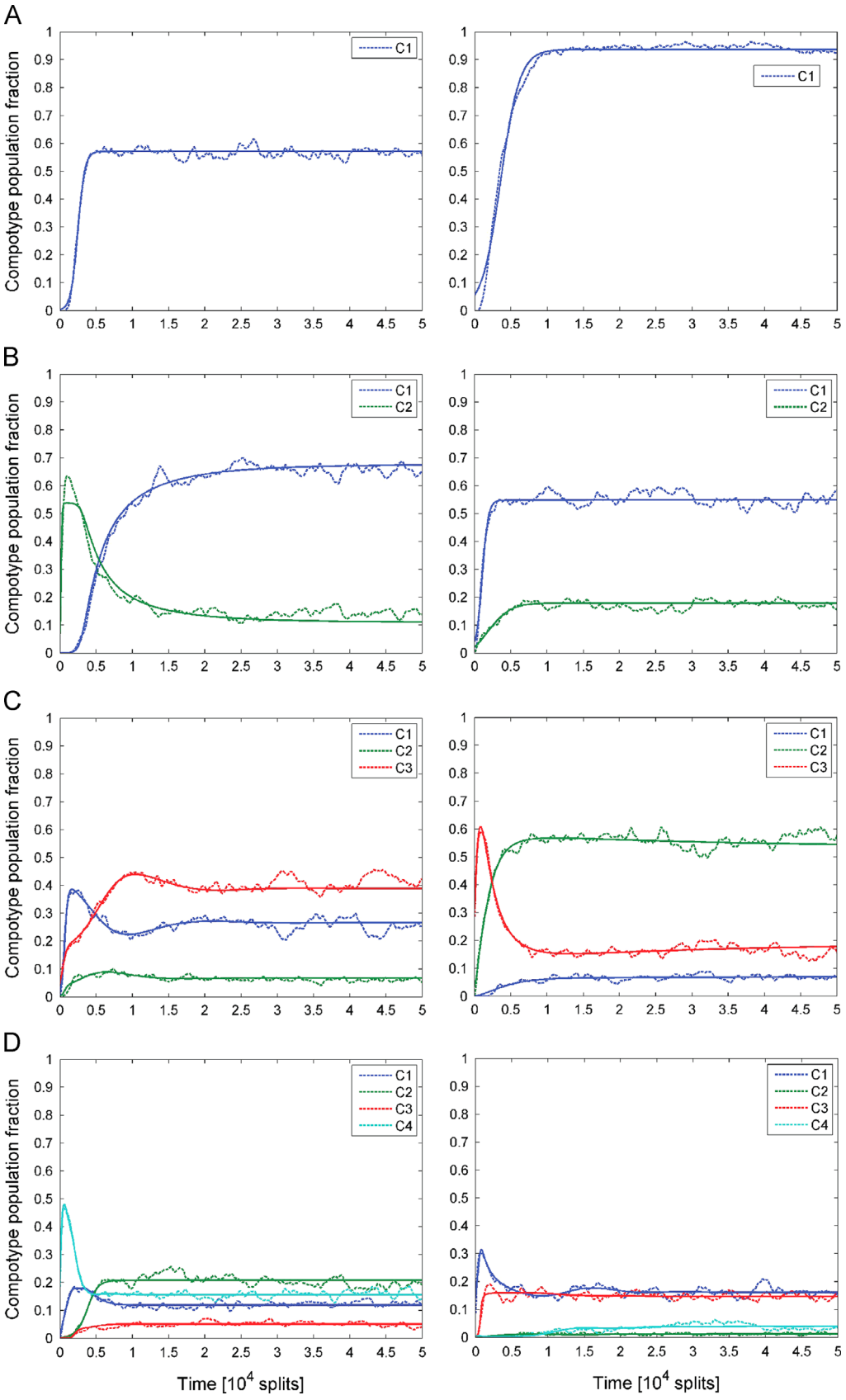


Fig. A.1. Additional examples of GARD population dynamics, showing different number of compotypes (N_C=1..4). Broken lines are the data and solid lines are the fit to the logistic equation. Fitted parameters are given in Table A.1. Panel B left and right sides respectively show takeover and control (see Section 3), with compositional similarity=0.30 and 0.55, respectively.

Table A.1
Logistic equation fitted parameters for all the examples given in Fig. A.1. β Seed is used for the generation of this β network using MATLAB pseudo random generator.

Panel	β Seed		K	r [splits^{-1}]	α_{ij}				$C(0)$
					$j=1$	$j=2$	$j=3$	$j=4$	
A, left	18	$i=1$	0.572	2.06E−03					3.58E−03
A, right	30	$i=1$	0.937	7.70E−04					5.68E−02
B, left	170	$i=1$	0.843	8.48E−02		1.518			7.86E−11
		$i=2$	0.538	1.28E−02	0.633				6.89E−02
B, right	261	$i=1$	0.548	2.40E−03		0.000			4.15E−02
		$i=2$	0.273	1.00E−03	0.170				2.25E−02
C, left	45	$i=1$	0.555	4.90E−03		1.487	0.486		1.94E−02
		$i=2$	0.494	4.30E−03	0.788		0.557		3.10E−03
		$i=3$	0.724	2.00E−03	1.260	0.000			5.46E−02
C, right	7	$i=1$	0.462	4.80E−03		0.601	0.359		1.00E−03
		$i=2$	0.734	5.30E−03	1.018		0.663		3.50E−02
		$i=3$	0.828	3.60E−03	0.000	1.919			3.13E−01
D, left	149	$i=1$	0.348	7.10E−03		0.274	2.094	0.414	1.57E−02
		$i=2$	0.448	3.20E−03	0.585		1.914	0.456	1.40E−03
		$i=3$	0.113	4.51E−02	0.281	0.038		0.130	1.59E−11
		$i=4$	0.581	6.20E−03	1.292	0.000	5.277		2.15E−01
D, right	133	$i=1$	0.518	2.40E−03		7.198	1.814	0.000	1.46E−01
		$i=2$	0.342	1.06E−02	0.002		2.034	0.791	1.91E−06
		$i=3$	0.187	8.50E−03	0.101	0.885		0.334	1.20E−03
		$i=4$	0.341	2.70E−03	1.266	4.141	0.304		4.10E−03

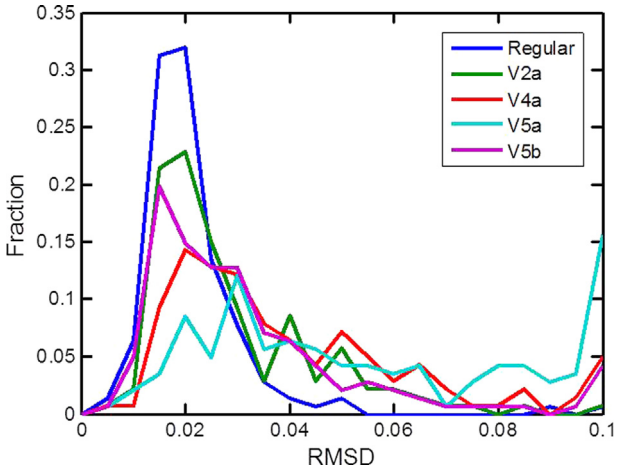


Fig. B.1. Distribution of RMSD values (Eq. (B.1)) simulations, for several logistic equation types. Regular (Eq. (2)) is based on the dataset of 1000 simulations. V2a–V5b (see Table B.1) are based on the extended dataset of 316 $N_C=2$ simulations.

Table B.1
Four additional type of logistic equation tested with $N_C=2$.

V2a	$\frac{dC_1}{dt}=r_1C_1\left(\frac{K_1-C_1+\alpha_{1D}C_D}{K_1}\right)$
	$\frac{dC_2}{dt}=r_2C_2\left(\frac{K_2-C_2+\alpha_{2D}C_D}{K_2}\right)C_D=L_{pop}-C_1-C_2$
V4a	$\frac{dC_1}{dt}=r_1C_1\left(\frac{K-C_1-\alpha_{12}C_2}{K}\right)$
	$\frac{dC_2}{dt}=r_2C_2\left(\frac{K-C_2-\alpha_{21}C_1}{K}\right)$
V5a (taken from [Kuno, E., 1992, Researches on Population Ecology, 34, 275])	$\frac{dC_1}{dt}=\left\{b_1\left(\frac{C_1}{C_1+\alpha_{12}C_2}\right)-d_1\right\}C_1-h_1C_1^2$
	$\frac{dC_2}{dt}=\left\{b_2\left(\frac{C_2}{C_2+\alpha_{21}C_1}\right)-d_2\right\}C_2-h_2C_2^2$
V5b	$\frac{dC_1}{dt}=\left\{b_1-d_1\left(\frac{C_1+\alpha_{12}C_2}{C_1}\right)\right\}C_1-h_1C_1^2$
	$\frac{dC_2}{dt}=\left\{b_2-d_2\left(\frac{C_2+\alpha_{21}C_1}{C_2}\right)\right\}C_2-h_2C_2^2$

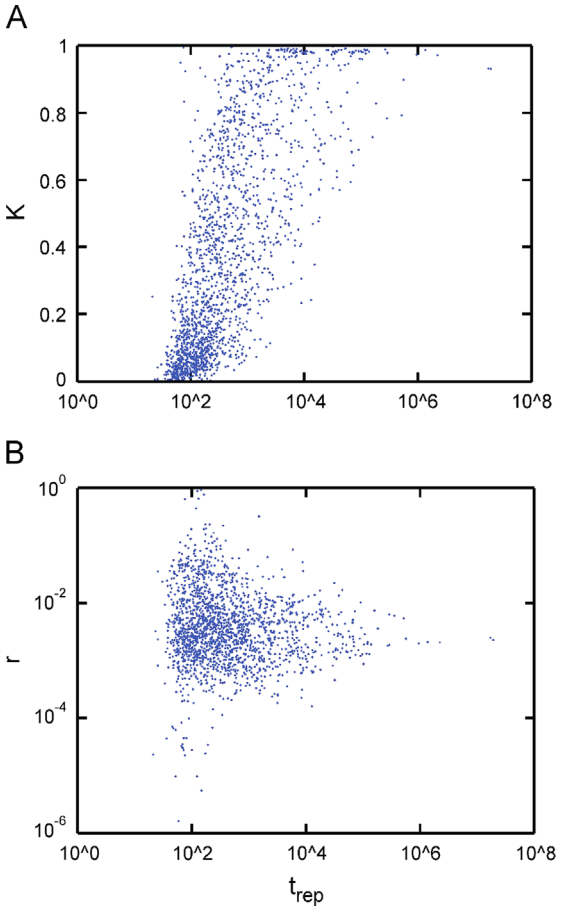


Fig. C.1. K and r vs. t_{rep} of all compotypes (A and B, respectively). Linear fit: K vs. $\log_{10}(t_{rep})$ gives $R^2=0.48$; $\log_{10}(r)$ vs. $\log_{10}(t_{rep})$ gives $R^2=0.007$.

Appendix C

See Figs. C.1 and C.2.

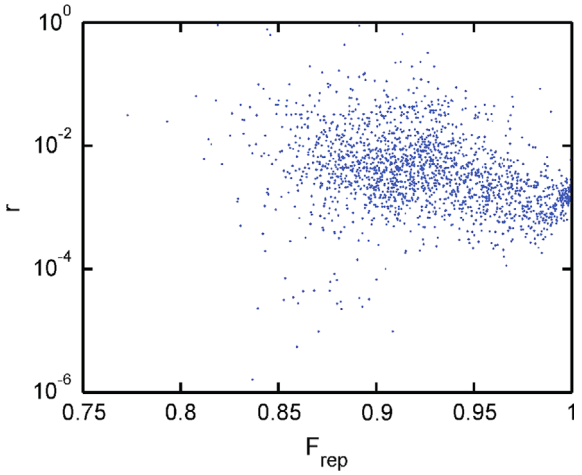


Fig. C.2. r vs. F_{rep} of all compotypes. Linear fit: $\log_{10}(r)$ vs. F_{rep} gives $R^2=0.10$.

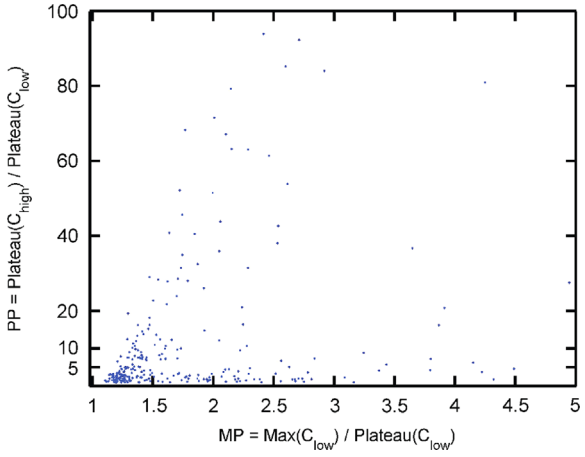


Fig. D.1. Takeover in simulations exhibiting $N_C=2$. Additional simulations performed for this part, giving a total of 316 such simulations. Takeover is represented by MP > 2 and PP > 5 and control is MP < 1.5 and PP < 4 (each group consists of 40 and 87 simulations, respectively). Max(C) is the value at the highest point and Plateau(C) is the average value along the plateau.

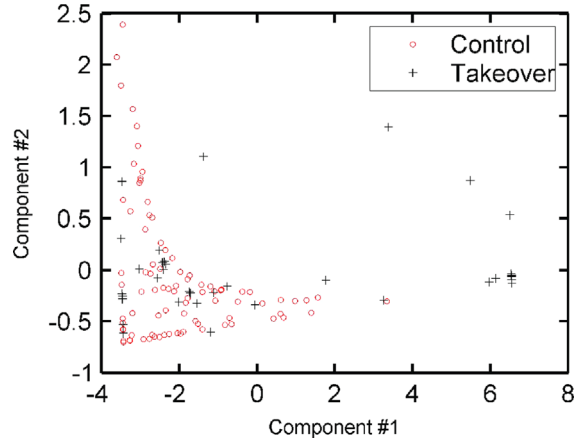


Fig. D.2. Principle component analysis (PCA) of the 6 fitted parameters (r_1, r_2, K_1, K_2, α_{12} and α_{21}), performed using MATLAB *princomp* routine. The first two components are responsible for 96% and 3% of the variance in the data, respectively.

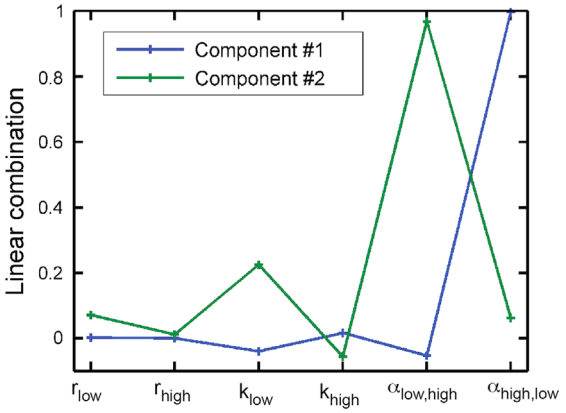


Fig. D.3. Linear combination of the first two components found by the PCA. A two sample Kolmogov–Smirlov test found that the takeover group have higher $\alpha_{high,low}$ values than the takeover (p-value$=3.2\times10^{-4}$).

Appendix D

See Figs. D.1–D3.

References

Anet, F.A., 2004. The place of metabolism in the origin of life. Curr. Opin. Chem. Biol. 8, 654–659.

Armstrong, D.L., Markovitch, O., Zidovetzki, R., Lancet, D., 2011. Replication of simulated prebiotic amphiphile vesicles controlled by experimental lipid physicochemical properties. Phys. Biol. 8, 066001, http://dx.doi.org/10.1088/1478-3975/8/6/066001.

Bailey, J.K., Hendry, A.P., Kinnison, M.T., Post, D.M., Palkovacs, E.P., Pelletier, F., Harmon, L.J., Schweitzer, J.A., 2009. From genes to ecosystems: an emerging synthesis of eco-evolutionary dynamics. N. Phytol. 184, 746–749.

Bernhardt, H.S., 2012. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). Biol. Direct 7, 23.

Biebricher, C.K., Eigen, M., 2006. What is a quasispecies? Curr. Top. Microbiol. Immunol. 299, 1–31.

Eigen, M., Schuster, P., 1977. Hypercycle – principle of natural self-organization. A. Emergence of hypercycle. Naturwissenschaften 64, 541–565.

Freilich, S., Kreimer, A., Borenstein, E., Yosef, N., Sharan, R., Gophna, U., Ruppin, E., 2009. Metabolic-network-driven analysis of bacterial ecological strategies. Genome Biol. 10, R61.

Gabriel, J.R., Saucy, F., Bersier, L.F., 2005. Paradoxes in the logistic equation? Ecol. Model. 185, 147–151.

Gillespie, D.T., 1976. General method for numerically simulating stochastic time evolution of coupled chemical-reactions. J. Comput. Phys. 22, 403–434.

Heinemann, M., Sauer, U., 2010. Systems biology of microbial metabolism. Curr. Opin. Microbiol. 13, 337–343.

Kiedrowski, G., Otto, S., Herdewijn, P., 2010. Welcome home, systems chemists. J. Syst. Chem. 1, 1.

Markovitch, O., Lancet, D., 2012. Excess mutual catalysis is required for effective evolvability. Artif. Life 18, 243–266.

Maurer, S.E., Deamer, D.W., Boncella, J.M., Monnard, P.A., 2009. Chemical evolution of amphiphiles: glycerol monoacyl derivatives stabilize plausible prebiotic membranes. Astrobiology 9, 979–987.

May, R.M., 1974. Biological populations with nonoverlapping generations – stable points, stable cycles, and chaos. Science 186, 645–647.

Moran, P.A.P., 1958. Random processes in genetics. Math. Proc. Camb. Philos. Soc. 54, 60–71.

Orgel, L.E., 2004. Prebiotic chemistry and the origin of the RNA world. Crit. Rev. Biochem. Mol. Biol. 39, 99–123.

Rasmussen, S., Bedau, M., Chen, L., Deamer, D.W., Krakauer, D.C., Norman, P.H., Peter, S.F., 2009. Protocells Bridging Nonliving and Living Matter. MIT Press, London.

Rescigno, A., 1967. Struggle for life. I. 2 Species. Bull. Math. Biophys. 29377.

Rescigno, A., 1968. Struggle for life. II. 3 Competitors. Bull. Math. Biophys. 30, 291.

Sauer, U., 2006. Metabolic networks in motion: 13C-based flux analysis. Mol. Syst. Biol. 2, 62.

Segre, D., Ben-Eli, D., Lancet, D., 2000. Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. Proc. Natl. Acad. Sci. USA 97, 4112–4117.

Segre, D., Ben-Eli, D., Deamer, D.W., Lancet, D., 2001a. The lipid world. Orig. Life Evol. Biosph. 31, 119–145.

Segre, D., Shenhav, B., Kafri, R., Lancet, D., 2001b. The molecular roots of compositional inheritance. J. Theor. Biol. 213, 481–491.

Shenhav, B., Oz, A., Lancet, D., 2007. Coevolution of compositional protocells and their environment. Philos. Trans. R. Soc. B – Biol. Sci. 362, 1813–1819.

Strobeck, C., 1973. *N* species competition. Ecology 54, 650–654.
Takeuchi, N., Salazar, L., Poole, A.M., Hogeweg, P., 2008. The evolution of strand preference in simulated RNA replicators with strand displacement: implications for the origin of transcription. Biol. Direct 3, 33.
Vandermeer, J.H., 1975. Interspecific competition – new approach to classical theory. Science 188, 253–255.
Vequi-Suplicy, C.C., Riske, K.A., Knorr, R.L., Dimova, R., 2010. Vesicles with charged domains. Biochim. Biophys. Acta – Biomembr. 1798, 1338–1347.