The origin of the central dogma through conflicting multilevel selection

Nobuto Takeuchi*[1,2] and Kunihiko Kaneko[1,3]

[1]Research Center for Complex Systems Biology, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan

[2]School of Biological Sciences, Faculty of Science, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

[3]Department of Basic Science, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan

Abstract

The central dogma of molecular biology rests on two kinds of asymmetry between genomes and enzymes: informatic asymmetry, where information flows from genomes to enzymes but not from enzymes to genomes; and catalytic asymmetry, where enzymes provide chemical catalysis but genomes do not. How did these asymmetries originate? Here we show that these asymmetries can spontaneously arise from conflict between selection at the molecular level and selection at the cellular level. We developed a model consisting of a population of protocells, each containing a population of replicating catalytic molecules. The molecules are assumed to face a trade-off between serving as catalysts and serving as templates. This trade-off causes conflicting multilevel selection: serving as catalysts is favoured by selection between protocells, whereas serving as templates is favoured by selection between molecules within protocells. This conflict induces informatic and catalytic symmetry breaking, whereby the molecules differentiate into genomes and enzymes, establishing the central dogma. We show mathematically that the symmetry breaking is caused by a positive feedback between Fisher's reproductive values and the relative impact of selection at different levels. This feedback induces a division of

*Corresponding author. E-mail: nobuto.takeuchi@auckland.ac.nz

labour between genomes and enzymes, provided variation at the molecular level is sufficiently large relative to variation at the cellular level, a condition that is expected to hinder the evolution of altruism. Taken together, our results suggest that the central dogma is a logical consequence of conflicting multilevel selection.

Keywords: reproductive division of labour | origin of genetic information | RNA world hypothesis | prebiotic evolution | Price equation

1 Introduction

At the heart of living systems lies a distinction between genomes and enzymes—a division of labour between the transmission of genetic information and the provision of chemical catalysis. This distinction rests on two types of asymmetry between genomes and enzymes: informatic asymmetry, where information flows from genomes to enzymes but not from enzymes to genomes; and catalytic asymmetry, where enzymes provide chemical catalysis but genomes do not. These two asymmetries constitute the essence of the central dogma in functional terms [1].

However, current hypotheses about the origin of life posit that genomes and enzymes were initially undistinguished, both embodied in a single type of molecule, RNA or its analogues [2]. While these hypotheses resolve the chicken-and-egg paradox of whether genomes or enzymes came first, they raise an obvious question: How did the genome-enzyme distinction originate?

To address this question, we explore the possibility that the genome-enzyme distinction arose during the evolutionary transition from replicating molecules to protocells [3–6]. During this transition, competition occurred both between protocells and between molecules within protocells. Consequently, selection operated at both cellular and molecular levels, and selection at one level was potentially in conflict with selection at the other [7, 8]. Previous studies have demonstrated that such conflicting multilevel selection can induce a partial and primitive distinction between genomes and enzymes in replicating molecules [9, 10]. Specifically, the molecules undergo catalytic symmetry breaking between their complementary strands, whereby one strand becomes catalytic and the other

becomes non-catalytic. However, the molecules do not undergo informatic symmetry breaking—i.e., one-way flow of information from non-catalytic to catalytic molecules—because complementary replication necessitates both strands to be replicated. Therefore, the previous studies have left the most essential aspect of the central dogma unexplained.

Here we investigate whether conflicting multilevel selection can induce both informatic and catalytic symmetry breaking in replicating molecules. To this end, we extend the previous model by considering two types of replicating molecules, denoted by P and Q. Although P and Q could be interpreted as RNA and DNA, their chemical identity is unspecified for simplicity and generality. To examine the possibility of spontaneous symmetry breaking, we assume that P and Q initially do not distinguish each other. We then ask whether evolution creates a distinction between P and Q such that information flows irreversibly from one type (either P or Q) that is non-catalytic to the other that is catalytic.

2 Model

Our model is an agent-based model with two types of replicators, P and Q. We assume that both P and Q are initially capable of catalysing four reactions at an equal rate: the replication of P, replication of Q, transcription of P to Q, and transcription of Q to P, where complementarity is ignored (Fig. 1a; note that this figure does not depict a two-member hypercycle because in our model replicators undergo transcription [11]; see Discussion for more on comparison with hypercycles).

Replicators compete for a finite supply of substrate denoted by S (hereafter, P, Q, and S are collectively called particles). S is consumed through the replication and transcription of P and Q, and recycled through the decay of P and Q (Fig. 1b). Thus, the total number of particles, i.e., the sum of the total numbers of P, Q, and S is kept constant (the relative frequencies of P, Q, and S are variable).

All particles are compartmentalised into protocells, across which P and Q do not diffuse at all, but S diffuses rapidly (Fig. 1c; Methods). This difference in diffusion induces the passive transport of S from protocells in which S is converted into P and Q slowly, to

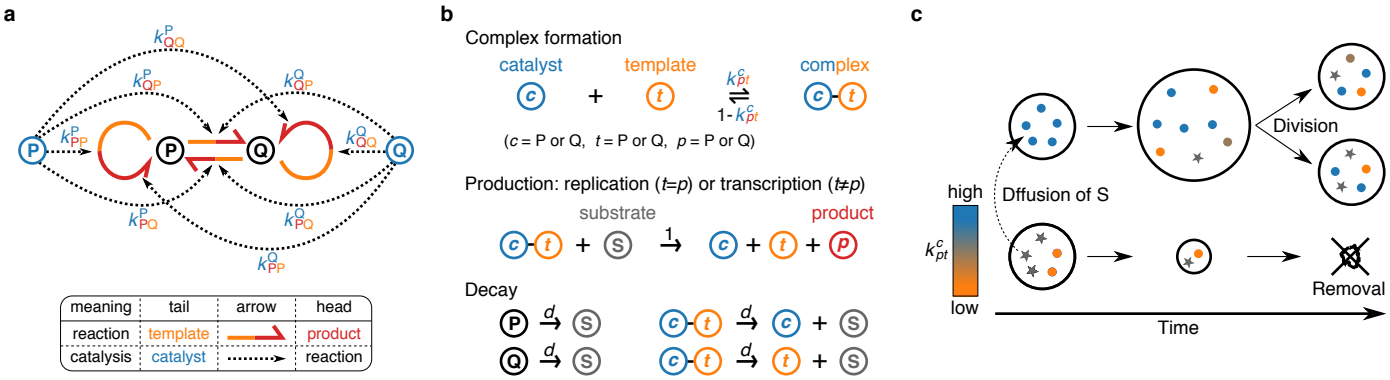

Figure 1: The agent-based model (see Methods for the details). **a**, Two types of replicators, P and Q, can serve as templates and catalysts for producing either type. Circular harpoons indicate replication; straight harpoons, transcription (heads indicate products; tails, templates). Dotted arrows indicate catalysis (heads indicate reaction catalysed; tails, replicators providing catalysis). **b**, Replicators undergo complex formation, replication, transcription, and decay. Rate constants of complex formation are given by the k^c_{pt} values of a replicator serving as a catalyst (whose type, P or Q, is denoted by c). The catalyst can form two distinct complexes with another replicator serving as a template (whose type is denoted by t) depending on whether it replicates ($p = t$) or transcribes ($p \neq t$) the template. **c**, Protocells exchange substrate (represented by stars) through rapid diffusion. Protocells divide when the number of internal particles exceeds V. Protocells are removed when they lose all particles.

protocells in which this conversion is rapid. Consequently, the latter grow at the expense of the former [12]. If the number of particles in a protocell exceeds threshold V, the protocell is divided with its particles randomly distributed between the two daughter cells; conversely, if this number decreases to zero, the protocell is discarded.

Crucial in our modelling is the incorporation of a trade-off between a replicator's catalytic activities and templating opportunities. This trade-off arises from the constraint that providing catalysis and serving as a template impose structurally-incompatible requirements on replicators [13, 14]. Because replication or transcription takes a finite amount of time, serving as a catalyst comes at the cost of spending less time serving as a template, thereby inhibiting replication of itself. To incorporate this trade-off, the model assumes that replication and transcription entail complex formation between a catalyst and template (Fig. 1b) [15]. The rate constants of complex formation are given by the catalytic activities (denoted by k^c_{pt}) of replicators, as described below.

Each replicator is individually assigned eight catalytic values denoted by $k^c_{pt} \in [0, 1]$, where the indices (c, p, and t) are P or Q (Fig. 1a). Four of these k^c_{pt} values denote the

catalytic activities of the replicator itself; the other four, those of its transcripts. For example, if a replicator is of type P, its catalytic activities are given by its k_{pt}^{P} values, whereas those of its transcripts, which are of type Q, are given by its k_{pt}^{Q} values. The indices p and t denote the specific type of reaction catalysed, as depicted in Fig. 1a. When a new replicator is produced, its k_{pt}^{c} values are inherited from its template with potential mutation of probability m (Methods).

The k_{pt}^{c} values of a replicator determine the rates at which this replicator forms a complex with another replicator and catalyses replication or transcription of the latter (Fig. 1b; Methods). The greater the catalytic activities (k_{pt}^{c}) of a replicator, the greater the chance that the replicator is sequestered in a complex as a catalyst and thus unable to serve as a template—hence a trade-off. Note that the trade-off is relative: if all replicators in a protocell have identical k_{pt}^{c} values, their multiplication rate increases monotonically with their k_{pt}^{c} values, assuming all else is held constant.

The above trade-off creates a dilemma: providing catalysis brings benefit at the cellular level because it accelerates a protocell's uptake of substrate; however, providing catalysis brings cost at the molecular level because it decreases the relative opportunity of a replicator to be replicated within a protocell [9]. Therefore, selection between protocells tends to maximise the k_{pt}^{c} values of replicators (i.e., cellular-level selection), whereas selection within protocells tends to minimise the k_{pt}^{c} values of replicators (i.e., molecular-level selection).

3 Results

3.1 Computational analysis

Using the agent-based model described above, we examined how k_{pt}^{c} values evolve as a result of conflicting multilevel selection. To this end, we set the initial k_{pt}^{c} values of all replicators to 1, so that P and Q are initially identical in their catalytic activities (the initial frequencies of P or Q are also set to be equal). We then simulated the model for various values of V (the threshold at which protocells divide) and m (mutation rate).

Our main result is that for sufficiently large values of V and m, replicators undergo spontaneous symmetry breaking in three aspects (Figs. 2a-d and S1). First, one type of replicator (either P or Q) evolves high catalytic activity, whereas the other completely loses it (i.e., $k_{pt}^{c} \gg k_{pt}^{c'} \approx 0$ for $c \neq c'$): catalytic symmetry breaking (Fig. 2bc). Second, templates are transcribed into catalysts, but catalysts are not reverse-transcribed into templates (i.e., $k_{ct}^{c} \gg k_{tc}^{c} \approx 0$): informatic symmetry breaking (Fig. 2bc). Finally, the copy number of templates becomes smaller than that of catalysts: numerical symmetry breaking: (Fig. 2d). This three-fold symmetry breaking is robust to various changes in model details (see SI Text 1.1 and 1.2; Figs. S2, S3, and S4).

A significant consequence of the catalytic and informatic symmetry breaking is the resolution of the dilemma between providing catalysis and getting replicated. Once symmetry is broken, tracking lineages reveals that the common ancestors of all replicators are almost always templates (Fig. 2ef; Methods). That is, information is transmitted almost exclusively through templates, whereas information in catalysts is eventually lost (i.e., catalysts have zero reproductive value). Consequently, evolution operates almost exclusively through competition between templates, rather than between catalysts. How the catalytic activity of catalysts evolves, therefore, depends solely on the cost and benefit to templates. On one hand, this catalytic activity brings benefit to templates for competition across protocells. On the other hand, this activity brings no cost to templates for competition within a protocell (neither does it bring benefit because catalysis is equally shared among templates). Therefore, the catalytic activity of catalysts is maximised by cellular-level selection operating on templates, but not minimised by molecular-level selection operating on templates, hence the resolution of the dilemma between catalysing and templating. Because of this resolution, symmetry breaking leads to the maintenance of high catalytic activities (Figs. S6 and S7).

3.2 Mathematical analysis

To understand the mechanism of the catalytic and informatic symmetry breaking, we simplified the agent-based model into mathematical equations. These equations allow us

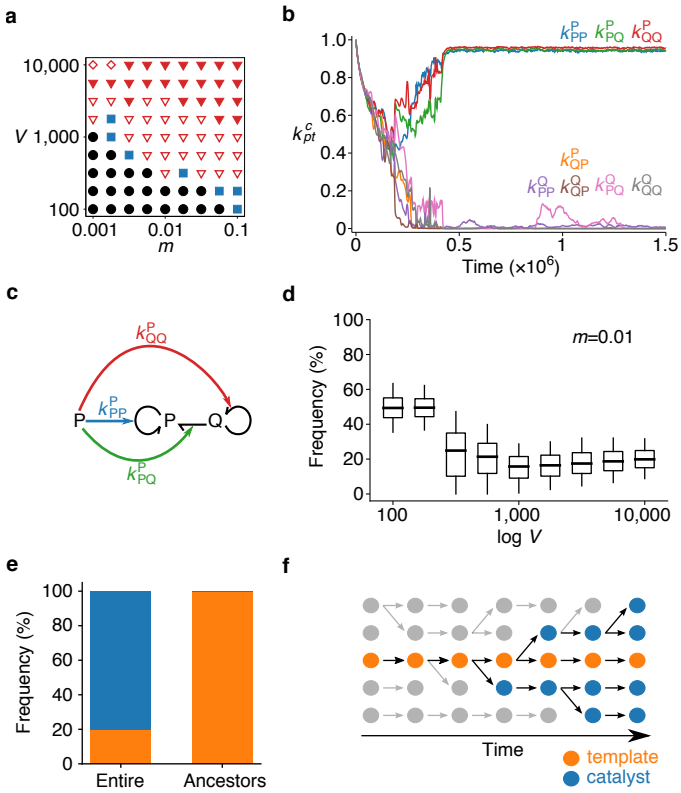

Figure 2: The evolution of the central dogma. **a**, Phase diagram: circles indicate no symmetry breaking (Fig. S1ab); squares, uncategorised (Fig. S1cd); open triangles, incomplete symmetry breaking (Fig. S1e-h); filled triangles, three-fold symmetry breaking as depicted in b, c, and d; diamonds, catalytic and informatic symmetry breaking without numerical symmetry breaking (Fig. S5a). The initial condition was $k^c_{pt} = 1$ for all replicators. **b**, Dynamics of k^c_{pt} averaged over all replicators. $V = 10000$ and $m = 0.01$. **c**, Catalytic activities evolved in b. **d**, Per-cell frequency of minority replicator types (P or Q) at equilibrium as a function of V: boxes, quartiles; whiskers, 5th and 95th percentiles. Only protocells containing at least $V/2$ particles were considered. **e**, Frequencies of templates (orange) and catalysts (blue) in the entire population or in the common ancestors. $V = 3162$ and $m = 0.01$. **f**, Illustration of e. Circles represent replicators; arrows, genealogy. Extinct lineages are grey. Common ancestors are always templates, whereas the majority of replicators are catalysts.

to consider all the costs and benefits involved in the provision of catalysis by $c \in \{\mathrm{P}, \mathrm{Q}\}$: molecular-level cost to c (denoted by γ^c_c) and cellular-level benefit to $t \in \{\mathrm{P}, \mathrm{Q}\}$ (denoted by β^t_c). The equations calculate the joint effects of all these costs and benefits on the evolution of the average catalytic activities of c (denoted by $\bar{k}^c$). The equations are derived with the help of Price's theorem [7, 8, 16] and displayed below (see Methods and

SI Text 1.3 for the derivation):

$$
\begin{aligned}
\Delta\bar{k}^{\mathrm{P}} &\approx \bar{\omega}^{\mathrm{P}}\left(\beta_{\mathrm{P}}^{\mathrm{P}}\sigma_{\mathrm{cel}}^{2} - \gamma_{\mathrm{P}}^{\mathrm{P}}\sigma_{\mathrm{mol}}^{2}\right) + \bar{\omega}^{\mathrm{Q}}\beta_{\mathrm{P}}^{\mathrm{Q}}\sigma_{\mathrm{cel}}^{2} \\
\Delta\bar{k}^{\mathrm{Q}} &\approx \bar{\omega}^{\mathrm{P}}\beta_{\mathrm{Q}}^{\mathrm{P}}\sigma_{\mathrm{cel}}^{2} + \bar{\omega}^{\mathrm{Q}}\left(\beta_{\mathrm{Q}}^{\mathrm{Q}}\sigma_{\mathrm{cel}}^{2} - \gamma_{\mathrm{Q}}^{\mathrm{Q}}\sigma_{\mathrm{mol}}^{2}\right),
\end{aligned}
\tag{1}
$$

where Δ denotes evolutionary change per generation, $\bar{\omega}^{c}$ is the average normalised reproductive value of c, $\sigma_{\mathrm{cel}}^{2}$ is the variance of catalytic activities among protocells (cellular-level variance), and $\sigma_{\mathrm{mol}}^{2}$ is the variance of catalytic activities within a protocell (molecular-level variance).

The first and second terms on the right-hand side of equations (1) represent evolution arising through the replication of P and Q, respectively, weighted by the reproductive values, $\bar{\omega}^{\mathrm{P}}$ and $\bar{\omega}^{\mathrm{Q}}$. The terms multiplied by $\beta_{c}^{t}\sigma_{\mathrm{cel}}^{2}$ represent evolution driven by cellular-level selection; those by $-\gamma_{c}^{c}\sigma_{\mathrm{mol}}^{2}$, evolution driven by molecular-level selection.

The derivation of equations (1) involves various simplifications that are not made in the agent-based model, among which the three most important are noted below (see Methods and SI Text 1.3 for details). First, equations (1) simplify evolutionary dynamics by restricting the number of evolvable parameters to a minimum required for catalytic and informatic symmetry breaking. More specifically, equations (1) assume that k_{pt}^{c} is independent of p and t (denoted by k^{c}), i.e., catalysts do not distinguish the replicator types of templates and products. Despite this simplification, catalytic symmetry breaking can still occur (e.g., $k^{\mathrm{P}} > k^{\mathrm{Q}}$), as can informatic symmetry breaking: the trade-off between catalysing and templating causes information to flow preferentially from less catalytic to more catalytic replicator types. However, numerical symmetry breaking is excluded as it requires k_{pt}^{c} to depend on p; consequently, the frequencies of P or Q are fixed and even in equations (1) (this is not the case in the agent-based model described in the previous section). Therefore, while equations (1) are useful for identifying the mechanism of catalytic and informatic symmetry breaking, they are not useful for identifying the mechanism of numerical symmetry breaking. In a supplementary material, we use different equations to identify the mechanism of numerical symmetry breaking (see SI Text 1.4 and Fig. S5).

The second simplification involved in equations (1) is that variances σ^2_{mol} and σ^2_{cel} are treated as parameters although they are actually dynamic variables dependent on m and V in the agent-based model (in supplementary material, we examine this assumption; see SI Text 1.5 and Fig. S8). In addition, these variances are assumed to be identical between $\bar{k}^{\mathrm{P}}$ and $\bar{k}^{\mathrm{Q}}$ because no difference is a priori assumed between P and Q.

The third simplification involved in equations (1) is that the terms of order greater than σ^2_{cel} and σ^2_{mol} are ignored under the assumption of weak selection [16].

Using equations (1), we can now elucidate the mechanism of the symmetry breaking. Consider a symmetric situation where P and Q are equally catalytic: $\bar{k}^{\mathrm{P}} = \bar{k}^{\mathrm{Q}}$. Since P and Q are identical, the catalytic activities of P and Q evolve identically: $\Delta\bar{k}^{\mathrm{P}} = \Delta\bar{k}^{\mathrm{Q}}$. Next, suppose that P becomes slightly more catalytic than Q for whatever reason, e.g., by genetic drift: $\bar{k}^{\mathrm{P}} > \bar{k}^{\mathrm{Q}}$ (catalytic asymmetry). The trade-off between catalysing and templating then causes P to be replicated less frequently than Q, so that $\bar{\omega}^{\mathrm{P}} < \bar{\omega}^{\mathrm{Q}}$ (informatic asymmetry). Consequently, the second terms of equations (1) increase relative to the first terms. That is, for catalysis provided by P (i.e., $\bar{k}^{\mathrm{P}}$), the impact of cellular-level selection through Q (i.e., $\bar{\omega}^{\mathrm{Q}}\beta^{\mathrm{Q}}_{\mathrm{P}}\sigma^2_{\mathrm{cel}}$) increases relative to those of molecular-level and cellular-level selection through P (i.e., $-\bar{\omega}^{\mathrm{P}}\gamma^{\mathrm{P}}_{\mathrm{P}}\sigma^2_{\mathrm{mol}}$ and $\bar{\omega}^{\mathrm{P}}\beta^{\mathrm{P}}_{\mathrm{P}}\sigma^2_{\mathrm{cel}}$, respectively), resulting in the relative strengthening of cellular-level selection. By contrast, for catalysis provided by Q (i.e., $\bar{k}^{\mathrm{Q}}$), the impacts of molecular-level and cellular-level selection through Q (i.e., $-\bar{\omega}^{\mathrm{Q}}\gamma^{\mathrm{Q}}_{\mathrm{Q}}\sigma^2_{\mathrm{mol}}$ and $\bar{\omega}^{\mathrm{Q}}\beta^{\mathrm{Q}}_{\mathrm{Q}}\sigma^2_{\mathrm{cel}}$, respectively) increase relative to cellular-level selection through P (i.e., $\bar{\omega}^{\mathrm{P}}\beta^{\mathrm{P}}_{\mathrm{Q}}\sigma^2_{\mathrm{cel}}$), resulting in the relative strengthening of molecular-level selection. Consequently, a small difference between $\bar{k}^{\mathrm{P}}$ and $\bar{k}^{\mathrm{Q}}$ leads to $\Delta\bar{k}^{\mathrm{P}} > \Delta\bar{k}^{\mathrm{Q}}$, the amplification of the initial difference—hence, symmetry breaking. The above mechanism can be summarised as a positive feedback between reproductive values and the relative impact of selection at different levels.

We next asked whether, and under what conditions, the above feedback leads to symmetry breaking such that either P or Q completely loses catalytic activity. To address this question, we performed a phase-plane analysis of equations (1) as described in Fig. 3 (see Methods and SI Text 1.6 for details). Figure 3 shows that $\bar{k}^{\mathrm{P}}$ and $\bar{k}^{\mathrm{Q}}$ diverge from

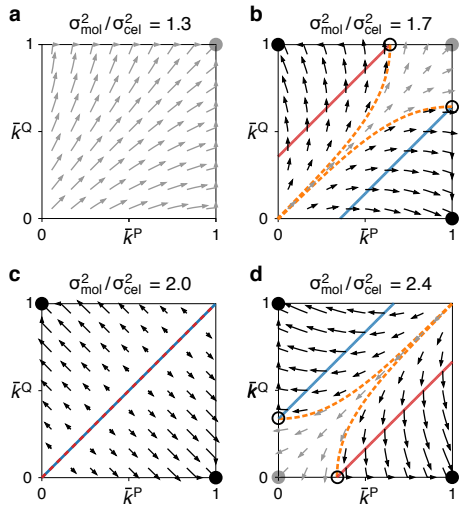

Figure 3: Phase-plane analysis. For this analysis, equations (1) were adapted as follows: β_c^t and γ_c^c were set to 1; $\bar{\omega}^c$ was calculated as $e^{-\bar{k}_c}/(e^{-\bar{k}_P}+e^{-\bar{k}_Q})$; Δ was replaced with time derivative ($\frac{d}{d\tau}$); and $\frac{d}{d\tau}\bar{k}^c$ was set to 0 if $\bar{k}^c = 0$ or $\bar{k}^c = 1$ to ensure that $\bar{k}^c$ is bounded within $[0, 1]$ as in the agent-based model. Solid lines indicate nullclines: $\frac{d}{d\tau}\bar{k}^P = 0$ (red) and $\frac{d}{d\tau}\bar{k}^Q = 0$ (blue). The nullclines at $\bar{k}^c = 0$ and $\bar{k}^c = 1$ are not depicted for visibility. Filled circles indicate symmetric (grey) and asymmetric (black) stable equilibria; open circles, unstable equilibria; arrows, short-duration flows ($\Delta\tau = 0.15$) leading to symmetric (grey) or asymmetric (black) equilibria. Dashed lines (orange) demarcate basins of attraction. $\sigma^2_{cel} = 1$. **a**, Molecular-level variance is so small that cellular-level selection completely dominates; consequently, $\bar{k}^c$ is always maximised. **b**, Molecular-level variance is large enough to create asymmetric equilibria; however, cellular-level variance is still large enough to make $\bar{k}^P = \bar{k}^Q = 1$ stable. **c**, A tipping point; the nullclines overlap. **d**, Molecular-level variance is so large that $\bar{k}^P = \bar{k}^Q = 1$ is unstable; the asymmetric equilibria can be reached if $\bar{k}^P \approx \bar{k}^Q \approx 1$.

symmetric states (i.e., $\Delta\bar{k}^P \neq \Delta\bar{k}^Q$), confirming the positive feedback described above. However, symmetry breaking occurs only if molecular-level variance σ^2_{mol} is sufficiently large relative to cellular-level variance σ^2_{cel} [i.e., if genetic relatedness between replicators, $\sigma^2_{cel}/(\sigma^2_{mol} + \sigma^2_{cel})$, is sufficiently low; see Methods]. Large $\sigma^2_{mol}/\sigma^2_{cel}$ is required because if $\sigma^2_{mol}/\sigma^2_{cel}$ is too small, cellular-level selection completely dominates over molecular-level selection, maximising both $\bar{k}^P$ and $\bar{k}^Q$ (Fig. 3a). The requirement of large $\sigma^2_{mol}/\sigma^2_{cel}$ is consistent with the fact that the agent-based model displays symmetry breaking for sufficiently large V: the law of large numbers implies that $\sigma^2_{mol}/\sigma^2_{cel}$ increases with V [9, 17]. This consistency with the agent-model suggests that equations (1) correctly describe the mechanism of symmetry breaking in the agent-based model (see SI Text 1.5 and Fig. S8 for an additional consistency check in terms of both m and V).

4 Discussion

Our results show that conflicting multilevel selection can induce informatic and catalytic symmetry breaking in replicating molecules. The symmetry breaking is induced because molecular-level selection minimises the catalytic activity of one type of molecule (either P or Q), whereas cellular-level selection maximises that of the other. The significance of the symmetry breaking is that it results in one-way flow of information from non-catalytic to catalytic molecules—the central dogma. The symmetry breaking thereby establishes a division of labour between the transmission of genetic information and the provision of chemical catalysis. This division of labour resolves a dilemma between templating and catalysing, the very source of conflict between levels of selection. Below, we discuss our results in relation to four subjects, namely, chemistry, hypercycle theory, kin selection theory, and reproductive division of labour.

Our theory does not specify the chemical details of replicating molecules, and this abstraction carries two implications. First, our theory suggests that the central dogma, if formulated in functional terms, is a general feature of living systems that is independent of protein chemistry. When the central dogma was originally proposed, it was formulated in chemical terms as the irreversible flow of information from nucleic acids to proteins [1]. Accordingly, the chemical properties of proteins have been considered integral to the central dogma [18]. By contrast, the present study formulates the central dogma in functional terms, as the irreversible flow of information from non-catalytic to catalytic molecules. Our theory shows that the central dogma, formulated as such, is a logical consequence of conflicting multilevel selection. Therefore, the central dogma might be a general feature of life that is independent of the chemical specifics of material in which life is embodied.

The second implication of the chemical abstraction is that our theory could be tested by experiments with existing materials. Our theory assumes that a replicator faces a trade-off between providing 'catalysis' and getting replicated. However, it does not restrict catalysis to being replicase activity: although our agent-based model assumes that catalysts are replicases, our mathematical analysis does not. Therefore, existing RNA and

DNA molecules could be used to test our theory [19]. For example, one could compare two systems, one where RNA serves as both templates and catalysts, and one where RNA serves as catalysts and DNA serves as templates. According to our theory, the latter is expected to maintain higher catalytic activity through evolution, provided the mutation rate and the number of molecules per cell are sufficiently large (see also [20]). In addition, using RNA and DNA is potentially relevant to the historical origin of the central dogma, given the possibility that DNA might have emerged before the advent of protein translation [21–24].

While our theory is similar to hypercycle theory in that both are concerned with the evolution of complexity in replicator systems, our theory proposes a distinct mechanism for evolving such complexity. Whereas hypercycle theory proposes symbiosis between multiple lineages of replicators [11], our theory proposes symmetry breaking (i.e., differentiation) in a single lineage of replicators—a fundamental distinction that is drawn between 'egalitarian' and 'fraternal' major evolutionary transitions as defined by Queller [25] (egalitarianism implies equality, which is involved in the evolution of complexity through symbiosis, whereas fraternalism implies kinship, which is involved in the evolution of complexity through differentiation; these terms are taken from the French Revolutionary slogan, *'Liberté, Egalité, Fraternité'*).

Moreover, our theory differs from hypercycle theory in terms of the roles played by non-catalytic templates. In hypercycle theory, the evolution of non-catalytic templates jeopardises hypercycles because such templates (called parasites) can replicate faster than catalytic templates constituting the hypercycles [15, 26]. In our theory, the evolution of non-catalytic templates is one of the essential factors leading to the division of labour between genomes and enzymes.

While our theory differs from hypercycle theory in the above aspects, it does not contradict the latter. In fact, there is a potential synergy between the evolution of complexity through symmetry breaking and that through symbiosis. Our theory posits that a distinction between genomes and enzymes resolves the dilemma between templating and catalysing, thereby increasing the evolutionary stability of catalytic activities in repli-

cators. Likewise, this distinction might also contribute to the evolutionary stability of symbiosis between replicators, hence the potential synergy (however, we should add that the specific mechanism of symbiosis proposed by hypercycle theory is not unique [27–33]).

While our theory is consistent with kin selection theory, it makes a novel prediction for evolution under a condition of low genetic relatedness. Kin selection theory posits that altruism can evolve if genetic relatedness is sufficiently high [34]. Consistent with this, our theory posits that for sufficiently high genetic relatedness (i.e., for sufficiently high $\sigma^2_{\text{cell}}/(\sigma^2_{\text{cel}} + \sigma^2_{\text{mol}})$, or sufficiently small m and V), cellular-level selection maximises the provision of catalysis by all molecules, establishing full altruism (providing catalysis can be viewed as altruism [35]: providing catalysis brings no direct benefit to a catalyst because a catalyst cannot catalyse the replication of itself in our model). However, the two theories diverge for sufficiently low genetic relatedness. In this case, kin selection theory predicts that evolution cannot lead to altruism; by contrast, our theory predicts that evolution can lead to a division of labour between the transmission of genetic information and the provision of chemical catalysis. Whether this reproductive division of labour should be called altruism is up for debate.

In relation to reproductive division of labour, our theory suggests a novel mechanism for its evolution in terms of a distinction between genomes and enzymes. In previous theories, reproductive division of labour has been regarded as an adaptation caused by natural selection [4–6]. For example, Michod has shown that reproductive division of labour can evolve because it maximises the group-level fitness of replicating entities if a trade-off curve between the replicating capacity and other functional capacities of the entities is convex [36] (see [37] for a historical reference). In our theory, division of labour between genomes and enzymes evolves, not because it maximises the fitness of a protocell (i.e., group), but because it is a stable equilibrium between evolution driven by molecular-level selection and evolution driven by cellular-level selection, an emergent outcome of conflicting multilevel selection (note that the fitness of a protocell is maximal if all replicators in the protocell are maximally catalytic and hence display no division of labour, a state that evolves for sufficiently small V and m). Parallel results have

Table 1: **Division of labour between information transmission and other functions transcends the levels of biological hierarchy.**

hierarchy		differentiation	
whole	parts	information	other
cell	molecules	genome	enzyme
symbiont population*	prokaryotic cells	transmitted	non-transmitted
ciliate	organelles	micronucleus	macronucleus
multicellular organism	eukaryotic cells	germ	soma
eusocial colony	animals	queen	worker

*Bacterial endosymbionts of ungulate lice (*Haematopinus*) and planthoppers (*Fulgoroidea*) [39].

been obtained from previous studies, where conflicting multilevel selection is shown to evolve various states that are not directly selected for at any single level [9, 20, 38]. Taken together, these results suggest the possibility that biological complexity evolves as emergent outcomes of conflicting multilevel selection.

Finally, we note that the division of labour between the transmission of genetic information and other functions is a recurrent pattern throughout biological hierarchy. For example, multicellular organisms display differentiation between germline and soma, as do eusocial animal colonies between queens and workers (Table 1) [3–6]. Given that all these systems potentially involve conflicting multilevel selection and tend to display reproductive division of labour as their sizes increase [6], our theory might provide a basis on which to pursue a universal principle of life that transcends the levels of biological hierarchy.

5 Methods

5.1 Details of the model

The model treats each molecule as a distinct individual with uniquely-assigned k_{pt}^{c} values. One time step of the model consists of three sub-steps: reaction, diffusion, and cell division.

In the reaction step, the reactions depicted in Fig. 1b are simulated with the algorithm

described previously [9]. The rate constants of complex formation are given by the k^c_{pt} values of a replicator serving as a catalyst. For example, if two replicators, denoted by X and Y, serve as a catalyst and template, respectively, the rate constant of complex formation is the k^x_{py} value of X, where x, y, and p are the replicator types (i.e., P or Q) of X, Y, and product, respectively. If X and Y switch the roles (i.e., X serves as a template, and Y serves as a catalyst), the rate constant of complex formation is the k^y_{px} value of Y. Complexes are distinguished not only by the roles of X and Y, but also by the replicator type of product p. Therefore, X and Y can form four distinct complexes depending on which replicator serves as a catalyst and which type of replicator is being produced.

The above rule about complex formation implies that whether a template is replicated ($p = t$) or transcribed ($p \neq t$) depends entirely on the k^c_{pt} values of a catalyst. In other words, a template cannot control how its information is used by a catalyst. This rule excludes the possibility that a template maximises its fitness by biasing catalysts towards replication rather than transcription. Excluding this possibility is legitimate if the backbone of a template does not directly determine the backbone of a product as in nucleic acid polymerisation.

In addition, the above rule about complex formation implies that replicators multiply fastest if their k^c_{pt} values are maximised for all combinations of c, p, and t (this is because X and Y form a complex at a rate proportional to $\sum_p k^x_{py} + k^y_{px}$ if all possible complexes are considered). Consequently, cellular-level selection tends to maximize k^c_{pt} values for all combinations of c, p, and t (because cellular-level selection tends to maximise the multiplication rate of replicators within protocells). If k^c_{pt} values are maximised for all combinations of c, p, and t, P and Q coexist. Therefore, coexistence between P and Q is favoured by cellular-level selection, a situation that might not always be the case in reality. We ascertained that the above specific rule about complex formation does not critically affect results by examining an alternative model in which cellular-level selection does not necessarily favour coexistence between P and Q (see SI Text 1.1).

In the diffusion step, all substrate molecules are randomly re-distributed among protocells with probabilities proportional to the number of replicators in protocells. In other

words, the model assumes that substrate diffuses extremely rapidly.

In the cell-division step, every protocell containing more than V particles (i.e. P, Q, and S together) is divided as described in Model.

The mutation of k^c_{pt} is modelled as unbiased random walks. With a probability m per replication or transcription, each k^c_{pt} value of a replicator is mutated by adding a number randomly drawn from a uniform distribution on the interval $(-\delta_{\mathrm{mut}}, \delta_{\mathrm{mut}})$ ($\delta_{\mathrm{mut}} = 0.05$ unless otherwise stated). The values of k^c_{pt} are bounded above by $k_{\max}$ with a reflecting boundary ($k_{\max} = 1$ unless otherwise stated), but are not bounded below to remove the boundary effect at $k^c_{pt} = 0$. However, if $k^c_{pt} < 0$, the respective rate constant of complex formation is regarded as zero.

We ascertained that the above specific model of mutation does not critically affect results by testing two alternative models of mutation. One model is nearly the same as the above, except that the boundary condition at $k^c_{pt} = 0$ was set to reflecting. The other model implements mutation as unbiased random walks on a logarithmic scale. The details are described in SI Text 1.2.

Each simulation was run for at least 5×10^7 time steps (denoted by $t_{\min}$) unless otherwise stated, where the unit of time is defined as that in which one replicator decays with probability d (thus, the average lifetime of replicators is $1/d$ time steps). The value of d was set to 0.02. The total number of particles in the model N_{tot} was set to $50V$ so that the number of protocells was approximately 100 irrespective of the value of V. At the beginning of each simulation, 50 protocells of equal size were generated. The initial values of k^c_{pt} were set to $k_{\max}$ for every replicator unless otherwise stated. The initial frequencies of P and Q were equal, and that of S was zero.

5.2 Ancestor tracking

Common ancestors of replicators were obtained in two steps. First, ancestor tracking was done at the cellular level to obtain the common ancestors of all surviving protocells. Second, ancestor tracking was done at the molecular level for the replicators contained by the common ancestors of protocells obtained in the first step. The results shown in

Fig. 2e were obtained from the data between 2.1×10^7 and 2.17×10^7 time steps, so that the ancestor distribution was from after the completion of symmetry breaking.

5.3 Outline of the derivation of equations (1)

To derive equations (1), we simplified the agent-based model in two ways. First, we assumed that k^c_{pt} is independent of p and t. Under this assumption, a catalyst does not distinguish the replicator types of templates (i.e., $k^c_{pt} = k^c_{pt'}$ for $t \neq t'$) and products (i.e., $k^c_{pt} = k^c_{p't}$ for $p \neq p'$). This assumption excludes the possibility of numerical symmetry breaking, but still allows catalytic and informatic symmetry breaking as described in Results.

Second, we abstracted away chemical reactions by defining ω^t_{ij} as the probability that replicator j of type t in protocell i is replicated or transcribed per unit time. Let $n^t_{ij}(\tau)$ be the population size of this replicator at time τ. Then, $n^t_{ij}(\tau)$ is expected to satisfy

$$\begin{bmatrix} n^{\mathrm{P}}_{ij}(\tau+1) \\ n^{\mathrm{Q}}_{ij}(\tau+1) \end{bmatrix} = \begin{bmatrix} \omega^{\mathrm{P}}_{ij} & \omega^{\mathrm{Q}}_{ij} \\ \omega^{\mathrm{P}}_{ij} & \omega^{\mathrm{Q}}_{ij} \end{bmatrix} \begin{bmatrix} n^{\mathrm{P}}_{ij}(\tau) \\ n^{\mathrm{Q}}_{ij}(\tau) \end{bmatrix}. \tag{2}$$

The fitness of the replicator can be defined as the dominant eigenvalue λ_{ij} of the 2×2 matrix on the right-hand side of equation (2): $\lambda_{ij} = \omega^{\mathrm{P}}_{ij} + \omega^{\mathrm{Q}}_{ij}$. Fisher's reproductive values of P and Q are given by the corresponding left eigenvector $\boldsymbol{u}_{ij} = [\omega^{\mathrm{P}}_{ij}, \omega^{\mathrm{Q}}_{ij}]$.

The evolutionary dynamics of the average catalytic activity of replicators can be described with Price's equation [7, 8]. Let κ^c_{ij} be the catalytic activity of replicator j of type c in protocell i (we use κ instead of k to distinguish κ^c_{ij} from k^c_{pt}). Price's equation states that

$$\langle \lambda_{\tilde{i}\tilde{j}} \rangle \Delta \langle \kappa^c_{\tilde{i}\tilde{j}} \rangle = \sigma^2_{\tilde{i}}[\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa^c_{i\tilde{j}} \rangle] + \mathbb{E}_{\tilde{i}}\big[\sigma^2_{i\tilde{j}}[\lambda_{ij}, \kappa^c_{ij}]\big], \tag{3}$$

where $\langle x_{i\tilde{j}} \rangle$, $\langle x_{\tilde{i}\tilde{j}} \rangle$, and $\mathbb{E}_{\tilde{i}}[x]$ are x averaged over the indices marked with tildes, $\sigma^2_{\tilde{i}}[x, y]$ is the covariance between x and y over protocells, and $\sigma^2_{i\tilde{j}}[x, y]$ is the covariance between x and y over the replicators in protocell i. One replicator is always counted as one sample in calculating all moments.

To approximate equation (3), we assumed that covariances between κ^{P}_{ij} and κ^{Q}_{ij} and between $\langle\kappa^{\mathrm{P}}_{i\tilde{j}}\rangle$ and $\langle\kappa^{\mathrm{Q}}_{i\tilde{j}}\rangle$ are negligible because the mutation of κ^{P}_{ij} and that of κ^{Q}_{ij} are uncorrelated in the agent-based model (see SI Text 1.6 for an alternative justification of this assumption). Under this assumption, equation (3) is approximated by equations (1) up to the second central moments of κ^c_{ij} and $\langle\kappa^c_{i\tilde{j}}\rangle$, with the following notation (see SI Text 1.3 for the derivation):

$$\bar{\omega}^t = \frac{\langle\omega^t_{\tilde{i}\tilde{j}}\rangle}{\langle\lambda_{\tilde{i}\tilde{j}}\rangle}, \qquad \sigma^2_{\mathrm{cel}} = \sigma^2_{\tilde{i}}[\langle\kappa^c_{i\tilde{j}}\rangle, \langle\kappa^c_{i\tilde{j}}\rangle], \qquad \sigma^2_{\mathrm{mol}} = \mathbb{E}_{\tilde{i}}[\sigma^2_{i\tilde{j}}[\kappa^c_{ij}, \kappa^c_{ij}]],$$

$$\bar{k}^c = \langle\kappa^c_{\tilde{i}\tilde{j}}\rangle, \qquad \gamma^c_c = -\mathbb{E}_{\tilde{i}}\left[\frac{\partial \ln \omega^c_{ij}}{\partial \kappa^c_{ij}}\right], \qquad \beta^t_c = \frac{\partial \ln\langle\omega^t_{i\tilde{j}}\rangle}{\partial\langle\kappa^c_{i\tilde{j}}\rangle},$$

where $\bar{\omega}^t$ is the normalised average reproductive value of type-t replicators, σ^2_{cel}, σ^2_{mol}, and $\bar{k}^c$ are the simplification of the notation, γ^c_c is an average decrease in the replication rate of a type-c replicator due to an increase in its own catalytic activity, and β^t_c is an increase in the average replication rate of type-t replicators in a protocell due to an increase in the average catalytic activity of type-c replicators in that protocell. We assumed that σ^2_{cel} and σ^2_{mol} do not depend on c because no difference is a priori assumed between P and Q.

The values of γ^c_c and β^t_c can be interpreted as the cost and benefit of providing catalysis. Let us assume that V is so large that $\langle\kappa^c_{i\tilde{j}}\rangle$ and κ^c_{ij} can be regarded as mathematically independent of each other if i and j are fixed (if i and j are varied, $\langle\kappa^c_{i\tilde{j}}\rangle$ and κ^c_{ij} may be statistically correlated). Under this assumption, increasing κ^c_{ij} does not increase $\langle\kappa^c_{i\tilde{j}}\rangle$, so that γ^c_c reflects only the cost of providing catalysis at the molecular level. Likewise, increasing $\langle\kappa^c_{i\tilde{j}}\rangle$ does not increase κ^c_{ij}, so that β^t_c reflects only the benefit of receiving catalysis at the cellular level. Moreover, the independence of $\langle\kappa^c_{i\tilde{j}}\rangle$ from κ^c_{ij} implies that $\partial\omega^{c'}_{ij}/\partial\kappa^c_{ij} = 0$ for $c \neq c'$, which permits the following interpretation: if a replicator of type c provides more catalysis, its transcripts, which is of type c', pay no extra cost (i.e., $\gamma^{c'}_c = 0$).

5.4 Outline of the phase-plane analysis

To perform the phase-plane analysis depicted in Fig. 3, we defined ω_{ij}^t as a specific function of κ_{ij}^t (see above for the meaning of ω_{ij}^t and κ_{ij}^t):

$$\omega_{ij}^t = e^{\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle + \langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle} e^{-s\kappa_{ij}^t} \left[\langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{P}}} \rangle + \langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{Q}}} \rangle \right]^{-1}, \tag{4}$$

where the first factor $e^{\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle + \langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle}$ represents the cellular-level benefit of catalysis provided by the replicators in protocell i, the second factor $e^{-s\kappa_{ij}^t}$ represents the molecular-level cost of catalysis provided by the focal replicator, the last factor normalises the cost, and s is the cost-benefit ratio. The above definition of ω_{ij}^t was chosen to satisfy the requirement that a replicator faces the trade-off between providing catalysis and serving as a template, i.e., γ_t^t and β_c^t are positive. Apart from this requirement, the definition was arbitrarily chosen for simplicity.

Under the definition in equation (4), we again approximated equation (3) up to the second central moments of κ_{ij}^c and $\langle \kappa_{i\tilde{j}}^c \rangle$, obtaining the following (see SI Text 1.6 for the derivation):

$$\bar{\omega}^t = e^{-s\bar{k}^t} / (e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}), \qquad \gamma_c^c = s, \qquad \beta_c^t = 1. \tag{5}$$

Equations (1) and (5) can be expressed in a compact form as

$$\begin{bmatrix} \Delta\bar{k}^{\mathrm{P}} \\ \Delta\bar{k}^{\mathrm{Q}} \end{bmatrix} \approx \sigma_{\mathrm{tot}}^2 \nabla \big[RB - (1-R)C \big],$$

where $\nabla = [\partial/\partial\bar{k}^{\mathrm{P}}, \partial/\partial\bar{k}^{\mathrm{Q}}]^{\mathrm{T}}$ (^{T} denotes transpose), $\sigma_{\mathrm{tot}}^2 = \sigma_{\mathrm{mol}}^2 + \sigma_{\mathrm{cel}}^2$, $R = \sigma_{\mathrm{cel}}^2/\sigma_{\mathrm{tot}}^2$, $B = \bar{k}^{\mathrm{P}} + \bar{k}^{\mathrm{Q}}$, and $C = -\ln(e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}})$. R can be interpreted as the regression coefficient of $\langle \kappa_{i\tilde{j}}^c \rangle$ on κ_{ij}^c [40] and, therefore, the coefficient of genetic relatedness [41]. The potential $RB - (1-R)C$ can be interpreted as inclusive fitness.

Competing interests

We have no competing financial interests.

Author Contributions

N.T. conceived the study, designed, implemented and analysed the models, and wrote the paper. K.K. discussed the design, results and implications of the study, and commented on the manuscript at all stages.

Acknowledgements

The authors thank Stuart A. West and his group, and Ulrich F Müller for discussion, Daniel J. van der Post, Austen R. D. Ganley, and Anthony M. Poole for help with the manuscript, and Paulien Hogeweg for inspiration. The authors wish to acknowledge the contribution of NeSI to the results of this research. New Zealand's national compute and analytics services and team are supported by the New Zealand eScience Infrastructure (NeSI) and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation and Employment. URL http://www.nesi.org.nz

Funding

The authors have been supported by JSPS KAKENHI (grant number JP17K17657 and JP17H06386). NT has been supported by grants from the University of Tokyo and the School of Biological Sciences, the University of Auckland.

References

[1] Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561–563.

[2] Gesteland RF, Cech T, Atkins JF, editors. The RNA World. 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006.

[3] Buss LW. The Evolution of Individuality. Princeton, NJ: Princeton University Press; 1987.

[4] Maynard Smith J, Szathmáry E. The Major Transitions in Evolution. Oxford: W. H. Freeman/Spektrum; 1995.

[5] Michod RE. Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality. Princeton, NJ: Princeton University Press; 1999.

[6] Bourke AFG. Principles of Social Evolution. Oxford, UK: Oxford University Press; 2011.

[7] Price GR. Extension of covariance selection mathematics. Annals of Human Genetics. 1972;35(4):485–490.

[8] Hamilton WD. Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox R, editor. Biosocial Anthoroplogy. London: Malaby Press; 1975. p. 133–153.

[9] Takeuchi N, Hogeweg P, Kaneko K. The origin of a primordial genome through spontaneous symmetry breaking. Nature Communications. 2017;8(1):250.

[10] von der Dunk S, Colizzi E, Hogeweg P. Evolutionary conflict leads to innovation: symmetry breaking in a spatial model of RNA-like replicators. Life. 2017;7(4):43.

[11] Eigen M, Schuster P. The Hypercycle: A Principle of Natural Self Organization. Berlin: Springer-Verlag; 1979.

[12] Chen IA, Roberts RW, Szostak JW. The emergence of competition between model protocells. Science. 2004;305:1474–1476.

[13] Durand PM, Michod RE. Genomics in the light of evolutionary transitions. Evolution. 2010;64(6):1533–1540.

[14] Ivica NA, Obermayer B, Campbell GW, Rajamani S, Gerland U, Chen IA. The paradox of dual roles in the RNA world: resolving the conflict between stable folding and templating ability. Journal of Molecular Evolution. 2013;77(3):55–63.

[15] Takeuchi N, Hogeweg P. The role of complex formation and deleterious mutations for the stability of RNA-like replicator systems. Journal of Molecular Evolution. 2007;65(6):668–686.

[16] Iwasa Y, Pomiankowski A, Nee S. The evolution of costly mate preferences II. the 'handicap' principle. Evolution. 1991;45(6):1431–1442.

[17] Takeuchi N, Kaneko K, Hogeweg P. Evolutionarily stable disequilibrium: Endless dynamics of evolution in a stationary population. Proceedings of the Royal Society B. 2016;283(1830):20153109.

[18] Koonin EV. Why the central dogma: on the nature of the great biological exclusion principle. Biology Direct. 2015;10(1):52.

[19] Müller S, Appel B, Balke D, Hieronymus R, Nübel C. Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? F1000Research. 2016;5:1511.

[20] Takeuchi N, Hogeweg P, Koonin EV. On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. PLoS Computational Biology. 2011;7(3):e1002024.

[21] Powner MW, Zheng SL, Szostak JW. Multicomponent assembly of proposed DNA precursors in water. Journal of the American Chemical Society. 2012;134(33):13889–13895.

[22] Poole AM, Horinouchi N, Catchpole RJ, Si D, Hibi M, Tanaka K, et al. The case for an early biological origin of DNA. Journal of Molecular Evolution. 2014;79(5-6):204–212.

[23] Ritson DJ, Sutherland JD. Conversion of biosynthetic precursors of RNA to those of DNA by photoredox chemistry. Journal of Molecular Evolution. 2014;78(5):245–250.

[24] Gavette JV, Stoop M, Hud NV, Krishnamurthy R. RNA-DNA chimeras in the

context of an RNA world transition to an RNA/DNA world. Angewandte Chemie International Edition. 2016;55(42):13204–13209.

[25] Queller DC. Cooperators since life began. The Quarterly Review of Biology. 1997;72(2):184–188.

[26] Maynard Smith J. Hypercycles and the origin of life. Nature. 1979;280:445–446.

[27] Szathmáry E, Demeter L. Group selection of early replicators and the origin of life. Journal of Theoretical Biology. 1987;128(4):463–486.

[28] Maynard Smith J, Szathmáry E. The origin of chromosomes I. Selection for linkage. Journal of Theoretical Biology. 1993;164(4):437–446.

[29] Czárán T, Szathmáry E. Coexistence of replicators in prebiotic evolution. In: Dieckman U, Law R, Metz JAJ, editors. The Geometry of Ecological Interactions: Simplifying Spatial Complexity. Cambridge, UK: Cambridge University Press; 2000. p. 116–134.

[30] Hogeweg P, Takeuchi N. Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. Origins of Life and Evolution of the Biosphere. 2003;33(4/5):375–403.

[31] Takeuchi N, Hogeweg P. Evolution of complexity in RNA-like replicator systems. Biology Direct. 2008;3(1):11.

[32] Takeuchi N, Hogeweg P. Evolutionary dynamics of RNA-like replicator systems: a bioinformatic approach to the origin of life. Physics of Life Reviews. 2012;9(3):219–263.

[33] Kim YE, Higgs PG. Co-operation between polymerases and nucleotide synthetases in the RNA world. PLOS Computational Biology. 2016;12(11):e1005161.

[34] Hamilton WD. The genetical evolution of social behaviour. I. Journal of Theoretical Biology. 1964;7(1):1–16.

[35] West SA, Griffin AS, Gardner A. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. Journal of Evolutionary Biology. 2007;20(2):415–432.

[36] Michod RE. The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. Proceedings of the National Academy of Sciences. 2006;103(24):9113–9117.

[37] Michod RE. Population biology of the first replicators: On the origin of the genotype, phenotype and organism. American Zoologist. 1983;23(1):5–14.

[38] Takeuchi N, Hogeweg P. Multilevel selection in models of prebiotic evolution II: A direct comparison of compartmentalization and spatial self-organization. PLoS Computational Biolology. 2009;5:e1000542.

[39] Frank SA. Host control of symbiont transmission: the separation of symbionts into germ and soma. The American Naturalist. 1996;148(6):1113–1124.

[40] Rice SH. Evolutionary Theory: Mathematical and Conceptual Foundations. Sunderland, MA, USA: Sinauer Associates; 2004.

[41] Hamilton WD. Selfish and spiteful behaviour in an evolutionary model. Nature. 1970;228(5277):1218–1220.

Supplementary material

The supplementary material file containing Supplementary Texts and Fig. S1 to S8 is appended to the main manuscript file.

Supplementary Material to "The origin of the central dogma through conflicting multilevel selection"

Nobuto Takeuchi*[1,2] and Kunihiko Kaneko[1,3]

[1]Research Center for Complex Systems Biology, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan
[2]School of Biological Sciences, Faculty of Science, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[3]Department of Basic Science, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan

1 Supporting Texts

1.1 An alternative agent-based model in which coexistence between P and Q is selectively neutral

In this section, we describe an alternative agent-based model in which coexistence between P and Q is neutral with respect to cellular-level selection. In the agent-based model described in the main text, coexistence between P and Q is favoured by cellular-level selection. This is due to a specific rule about complex formation, which implies that replicators multiply fastest if both P and Q provide and receive catalysis (see Methods for details). To ascertain that this specific rule about complex formation does not critically affect results, we additionally examined an alternative model in which replicators multiply fastest even if only either P or Q provides and receives catalysis. In this model, cellular-level selection does not favour coexistence between P and Q while it still tends to maximise the multiplication rate of replicators within protocells.

In the alternative model, the reaction rate constants of complex formation are defined as a function of the k^c_{pt} values of a replicator serving as a catalyst as follows:

$$\max(k^c_{\mathrm{P}t}, k^c_{\mathrm{Q}t})\frac{k^c_{pt}}{k^c_{\mathrm{P}t} + k^c_{\mathrm{Q}t}}.$$

Under this definition, two replicators, denoted by X and Y, form a complex at a rate proportional to $\max(k^x_{\mathrm{P}y}, k^x_{\mathrm{Q}y}) + \max(k^y_{\mathrm{P}x}, k^y_{\mathrm{Q}x}) \leqslant 2k_{\max}$ if all possible complexes are considered, where x and y are the replicator types of X and Y, respectively (in the original

*Corresponding author. E-mail: nobuto.takeuchi@auckland.ac.nz

model, this rate is proportional to $\sum_p k^x_{py} + k^y_{px} \leqslant 4k_{\max}$). Accordingly, replicators multiply fastest not only if $k^c_{pt} = k_{\max}$ for all combinations of c, p, and t, but also if $k^c_{cc} = k_{\max}$ for either $c = \mathrm{P}$ or $c = \mathrm{Q}$ and $k^c_{pt} = 0$ for all the other combinations of c, p, and t. In other words, replicators multiply fastest even if only either P or Q provides and receives catalysis (this is in contrast to the model described in the main text). While cellular-level selection always tends to maximise the multiplication rate of replicators within protocells, it is indifferent to how this maximisation is achieved. Therefore, cellular-level selection does not necessarily tend to maximise k^c_{pt} values for all combinations of c, p, and t; i.e., it does not necessarily favour coexistence between P and Q.

To examine the effect of coexistence between P and Q on symmetry breaking, we simulated the alternative model described above with two initial conditions, symmetric and asymmetric. In the symmetric initial condition, both P and Q were present—this is the same initial condition as used for the original agent-based model. In the asymmetric initial condition, only Q was present (see Fig. S2 for details)—this condition might be closer to what is typically imagined in the RNA world hypothesis. For both initial conditions, the model displays the same three-fold symmetry breaking as displayed by the original model (Fig. S2), indicating that the results do not depend on whether coexistence between P and Q is favoured by cellular-level selection.

1.2 Alternative agent-based models in which the mutation of k^c_{pt} is modelled differently

In this section, we describe alternative models for the mutation of k^c_{pt}. In the agent-based model described in the main text, the mutation of k^c_{pt} is modelled as unbiased random walks in a half-open interval $(-\infty, k_{\max})$ with a reflecting boundary at $k^c_{pt} = k_{\max}$. To ascertain that this specific model of mutation does not critically affect results, we additionally examined two alternative models of mutation. The first alternative model is nearly the same as the above, except that the reflecting boundary condition is set at $k^c_{pt} = 0$. In the second alternative model, each k^c_{pt} value is mutated by multiplying $\exp(\epsilon)$, where ϵ is a number randomly drawn from a uniform distribution on the interval $(-\delta_{\mathrm{mut}}, \delta_{\mathrm{mut}})$, with a reflecting boundary at $k^c_{pt} = k_{\max}$. Both models of mutation produce essentially the same result as described in the main text (Figs. S3 and S4), indicating that the results do not depend on the specific models of mutation.

1.3 The derivation of equation (1)

In this section, we describe the derivation of equations (1) that is outlined in Methods.

To derive equations (1), we simplified the agent-based model in two ways. First, we assumed that k^c_{pt} is independent of p and t. Under this assumption, a catalyst does not distinguish the replicator types of templates (i.e., $k^c_{pt} = k^c_{pt'}$ for $t \neq t'$) and products (i.e., $k^c_{pt} = k^c_{p't}$ for $p \neq p'$). This assumption excludes the possibility of numerical symmetry breaking, but still allows catalytic and informatic symmetry breaking as described in the main text (see Results).

Second, we abstracted away chemical reactions by defining ω^t_{ij} as the probability that replicator j of type t in protocell i is replicated or transcribed per unit time. Let $n^t_{ij}(\tau)$ be the population size of this replicator at time τ. Then, the dynamics of $n^t_{ij}(\tau)$ can be

mathematically described as

$$\begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau+1) \\ n_{ij}^{\mathrm{Q}}(\tau+1) \end{bmatrix} = \begin{bmatrix} \omega_{ij}^{\mathrm{P}} & \omega_{ij}^{\mathrm{Q}} \\ \omega_{ij}^{\mathrm{P}} & \omega_{ij}^{\mathrm{Q}} \end{bmatrix} \begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau) \\ n_{ij}^{\mathrm{Q}}(\tau) \end{bmatrix}. \tag{S1}$$

The fitness of the replicator can be defined as the dominant eigenvalue λ_{ij} of the 2×2 matrix on the right-hand side of equation (S1). The equilibrium frequencies of P and Q are given by the right eigenvector $\boldsymbol{v}_{ij}$ associated with λ_{ij}. Fisher's reproductive values of P and Q are given by the corresponding left eigenvector $\boldsymbol{u}_{ij}$. These eigenvalue and eigenvectors are calculated as follows:

$$\lambda_{ij} = \omega_{ij}^{\mathrm{P}} + \omega_{ij}^{\mathrm{Q}}, \qquad \boldsymbol{v}_{ij} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{u}_{ij} = \begin{bmatrix} \omega_{ij}^{\mathrm{P}} & \omega_{ij}^{\mathrm{Q}} \end{bmatrix}. \tag{S2}$$

Based on the above simplification, we now derive equations (1). For concreteness, we focus on the evolution of the average catalytic activity of P (denoted by $\bar{k}^{\mathrm{P}}$ in the main text). However, the same method of derivation is applicable to that of Q if P and Q are swapped.

Let κ_{ij}^{P} be the catalytic activity of replicator j of type P in protocell i (we use κ instead of k to distinguish κ_{ij}^{P} from k_{pt}^{P}). Price's equation [1, 2] states that

$$\langle \lambda_{\tilde{i}\tilde{j}} \rangle \Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle = \sigma_{\tilde{i}}^2[\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle] + \mathbb{E}_{\tilde{i}}\big[\sigma_{i\tilde{j}}^2[\lambda_{ij}, \kappa_{ij}^{\mathrm{P}}]\big] \tag{S3}$$

where $\langle x_{i\tilde{j}} \rangle$, $\langle x_{\tilde{i}\tilde{j}} \rangle$, and $\mathbb{E}_{\tilde{i}}[x]$ are x averaged over the indices marked with tildes, $\sigma_{\tilde{i}}^2[x, y]$ is the covariance between x and y over protocells, and $\sigma_{i\tilde{j}}^2[x, y]$ is the covariance between x and y over the replicators in protocell i (one replicator is always counted as one sample in calculating all moments). Below, we show that equation (S3) is approximated by equations (1) up to the second moments of $\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle$ and κ_{ij}^{P}, namely, $\sigma_{\tilde{i}}^2[\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle]$ and $\mathbb{E}_{\tilde{i}}[\sigma_{i\tilde{j}}^2[\kappa_{ij}^{\mathrm{P}}, \kappa_{ij}^{\mathrm{P}}]]$.

To approximate the first term on the right-hand side of equation (S3), we assume that $\langle \lambda_{i\tilde{j}} \rangle$ is a function of $\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle$ and $\langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle$ that can be expanded as a Taylor series around $\langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle$ and $\langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}} \rangle$. Substituting this series into $\sigma_{\tilde{i}}^2[\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle]$, we obtain

$$\sigma_{\tilde{i}}^2[\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle] = \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \frac{\partial \langle \lambda_{i\tilde{j}} \rangle}{\partial \langle \kappa_{i\tilde{j}}^{c} \rangle} \sigma_{\tilde{i}}^2[\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle, \langle \kappa_{i\tilde{j}}^{c} \rangle] + O(\sigma_{\tilde{i}}^3), \tag{S4}$$

where $O(\sigma_{\tilde{i}}^3)$ consists of terms involving the third or higher (mixed) central moments of $\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle$ and $\langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle$ over protocells [3].

To approximate the second term on the right-hand side of equation (S3), we likewise assume that λ_{ij} is a function of κ_{ij}^{P} and κ_{ij}^{Q} that can be expanded as a Taylor series around $\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle$ and $\langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle$. Substituting this series into $\sigma_{i\tilde{j}}^2[\lambda_{ij}, \kappa_{ij}^{\mathrm{P}}]$, we obtain

$$\sigma_{i\tilde{j}}^2[\lambda_{ij}, \kappa_{ij}^{\mathrm{P}}] = \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \frac{\partial \lambda_{ij}}{\partial \kappa_{ij}^{c}} \sigma_{i\tilde{j}}^2[\kappa_{ij}^{\mathrm{P}}, \kappa_{ij}^{c}] + O(\sigma_{i\tilde{j}}^3),$$

where $O(\sigma_{i\tilde{j}}^3)$ consists of terms involving the third or higher (mixed) central moments of κ_{ij}^{P} and κ_{ij}^{Q} over the replicators in protocell i [3]. Applying $\mathbb{E}_{\tilde{i}}$ to both sides of the above

equation and assuming that $\partial\lambda_{ij}/\partial\kappa^c_{ij}$ is independent of $\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^c_{ij}]$, we obtain

$$\mathbb{E}_{\tilde{i}}\big[\sigma^2_{\tilde{i}\tilde{j}}[\lambda_{ij}, \kappa^{\mathrm{P}}_{ij}]\big] = \sum_{c\in\{\mathrm{P},\mathrm{Q}\}} \mathbb{E}_{\tilde{i}}\left[\frac{\partial\lambda_{ij}}{\partial\kappa^c_{ij}}\right]\mathbb{E}_{\tilde{i}}\big[\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^c_{ij}]\big] + \mathbb{E}_{\tilde{i}}\big[O(\sigma^3_{\tilde{i}\tilde{j}})\big]. \tag{S5}$$

Substituting equations (S4) and (S5) into equation (S3), we obtain

$$\Delta\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle = \frac{1}{\langle\lambda_{\tilde{i}\tilde{j}}\rangle}\sum_{c\in\{\mathrm{P},\mathrm{Q}\}}\left(\frac{\partial\langle\lambda_{\tilde{i}\tilde{j}}\rangle}{\partial\langle\kappa^c_{\tilde{i}\tilde{j}}\rangle}\sigma^2_{\tilde{i}}[\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^c_{\tilde{i}\tilde{j}}\rangle] + \mathbb{E}_{\tilde{i}}\left[\frac{\partial\lambda_{ij}}{\partial\kappa^c_{ij}}\right]\mathbb{E}_{\tilde{i}}\big[\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^c_{ij}]\big]\right) + O', \tag{S6}$$

where $O' = O(\sigma^3_{\tilde{i}}) + E_{\tilde{i}}[O(\sigma^3_{\tilde{i}\tilde{j}})]$.

Next, we assume that covariances $\sigma^2_{\tilde{i}}[\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle]$ and $\mathbb{E}_{\tilde{i}}\big[\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{Q}}_{ij}]\big]$ are negligible because the mutation of κ^{P}_{ij} and that of κ^{Q}_{ij} are uncorrelated in the simulation model (this assumption is alternatively justified in SI Text 1.6). Under this assumption, equation (S6) is transformed into

$$\Delta\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle = \frac{1}{\langle\lambda_{\tilde{i}\tilde{j}}\rangle}\left(\frac{\partial\langle\lambda_{\tilde{i}\tilde{j}}\rangle}{\partial\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle}\sigma^2_{\tilde{i}}[\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle] + \mathbb{E}_{\tilde{i}}\left[\frac{\partial\lambda_{ij}}{\partial\kappa^{\mathrm{P}}_{ij}}\right]\mathbb{E}_{\tilde{i}}\big[\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]\big]\right) + O'. \tag{S7}$$

Using equation (S2) (i.e., $\lambda_{ij} = \omega^{\mathrm{P}}_{ij} + \omega^{\mathrm{Q}}_{ij}$), we can transform equation (S7) into

$$\Delta\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle = \frac{1}{\langle\lambda_{\tilde{i}\tilde{j}}\rangle}\sum_{t\in\{\mathrm{P},\mathrm{Q}\}}\left(\frac{\partial\langle\omega^t_{\tilde{i}\tilde{j}}\rangle}{\partial\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle}\sigma^2_{\tilde{i}}[\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle] + \mathbb{E}_{\tilde{i}}\left[\frac{\partial\omega^t_{ij}}{\partial\kappa^{\mathrm{P}}_{ij}}\right]\mathbb{E}_{\tilde{i}}\big[\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]\big]\right) + O'. \tag{S8}$$

Moreover, it can be shown that

$$\begin{aligned}
\mathbb{E}_{\tilde{i}}\left[\frac{\partial\omega^t_{ij}}{\partial\kappa^c_{ij}}\bigg|_{\substack{\kappa^{\mathrm{P}}_{ij}=\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle\\ \kappa^{\mathrm{Q}}_{ij}=\langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle}}\right] &= \mathbb{E}_{\tilde{i}}\left[\omega^t_{ij}(\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle)\frac{\partial\ln\omega^t_{ij}}{\partial\kappa^c_{ij}}\bigg|_{\substack{\kappa^{\mathrm{P}}_{ij}=\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle\\ \kappa^{\mathrm{Q}}_{ij}=\langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle}}\right]\\
&= \mathbb{E}_{\tilde{i}}\big[\omega^t_{ij}(\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle)\big]\mathbb{E}_{\tilde{i}}\left[\frac{\partial\ln\omega^t_{ij}}{\partial\kappa^c_{ij}}\bigg|_{\substack{\kappa^{\mathrm{P}}_{ij}=\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle\\ \kappa^{\mathrm{Q}}_{ij}=\langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle}}\right] + O(\sigma^2_i)\\
&= \langle\omega^t_{\tilde{i}\tilde{j}}\rangle\mathbb{E}_{\tilde{i}}\left[\frac{\partial\ln\omega^t_{ij}}{\partial\kappa^c_{ij}}\bigg|_{\substack{\kappa^{\mathrm{P}}_{ij}=\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle\\ \kappa^{\mathrm{Q}}_{ij}=\langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle}}\right] + \mathbb{E}_{\tilde{i}}\big[O(\sigma^2_{\tilde{i}\tilde{j}})\big] + O(\sigma^2_i).
\end{aligned}$$

Using the above equation, we can transform equation (S8) into

$$\Delta\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle = \sum_{t\in\{\mathrm{P},\mathrm{Q}\}}\frac{\langle\omega^t_{\tilde{i}\tilde{j}}\rangle}{\langle\lambda_{\tilde{i}\tilde{j}}\rangle}\left(\frac{\partial\ln\langle\omega^t_{\tilde{i}\tilde{j}}\rangle}{\partial\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle}\sigma^2_{\tilde{i}}[\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle] + \mathbb{E}_{\tilde{i}}\left[\frac{\partial\ln\omega^t_{ij}}{\partial\kappa^{\mathrm{P}}_{ij}}\right]\mathbb{E}_{\tilde{i}}\big[\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]\big]\right) + O'', \tag{S9}$$

where $O'' = O' + O(\sigma^2_{\tilde{i}})\ \mathbb{E}_{\tilde{i}}[O(\sigma^2_{\tilde{i}\tilde{j}})] + \mathbb{E}_{\tilde{i}}[O(\sigma^2_{\tilde{i}\tilde{j}})]\mathbb{E}_{\tilde{i}}[O(\sigma^2_{\tilde{i}\tilde{j}})]$.

We adopt the following notation:

$$\bar{\omega}^t = \frac{\langle\omega^t_{\tilde{i}\tilde{j}}\rangle}{\langle\lambda_{\tilde{i}\tilde{j}}\rangle}, \qquad \sigma^2_{\mathrm{cel}} = \sigma^2_{\tilde{i}}[\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle], \qquad \sigma^2_{\mathrm{mol}} = \mathbb{E}_{\tilde{i}}[\sigma^2_{\tilde{i}\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]],$$

$$\bar{k}^{\mathrm{P}} = \langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \qquad \gamma^{\mathrm{P}}_{\mathrm{P}} = -\mathbb{E}_{\tilde{i}}\left[\frac{\partial\ln\omega^{\mathrm{P}}_{ij}}{\partial\kappa^{\mathrm{P}}_{ij}}\right], \qquad \beta^t_{\mathrm{P}} = \frac{\partial\ln\langle\omega^t_{\tilde{i}\tilde{j}}\rangle}{\partial\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle},$$

where $\bar{\omega}^t$ is the normalised average reproductive value of type-t replicators, σ^2_{cel}, σ^2_{mol}, and $\bar{k}^{\mathrm{P}}$ are the simplification of the notation, $\gamma^{\mathrm{P}}_{\mathrm{P}}$ is an average decrease in the replication rate of a type-P replicator due to an increase in its own catalytic activity, and β^t_{P} is an increase in the average replication rate of type-t replicators in a protocell due to an increase in the average catalytic activity of type-P replicators in that protocell.

We assume that V is so large that $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$ and κ^{P}_{ij} can be regarded as mathematically independent of each other, provided i and j are fixed (if i and j are varied, $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$ and κ^{P}_{ij} may be statistically correlated). Under this assumption, increasing κ^{P}_{ij} does not increase $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$, so that $\gamma^{\mathrm{P}}_{\mathrm{P}}$ reflects only the cost of providing catalysis at the molecular level. Likewise, increasing $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$ does not increase κ^{P}_{ij}, so that β^t_{P} reflects only the benefit of receiving catalysis at the cellular level. Moreover, the independence of $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$ from κ^{P}_{ij} implies that $\partial \omega^{\mathrm{Q}}_{ij} / \partial \kappa^{\mathrm{P}}_{ij} = 0$, which permits the following interpretation: if a replicator of type P provides more catalysis, its transcripts, which is of type Q, pay no extra cost (i.e., $\gamma^{\mathrm{Q}}_{\mathrm{P}} = 0$).

Using the above notation and the fact that $\partial \omega^{\mathrm{Q}}_{ij} / \partial \kappa^{\mathrm{P}}_{ij} = 0$, we can transform equation (S9) into

$$\Delta \bar{k}^{\mathrm{P}} \approx \bar{\omega}^{\mathrm{P}} (b^{\mathrm{P}}_{\mathrm{P}} \sigma^2_{\mathrm{cel}} - \gamma^{\mathrm{P}}_{\mathrm{P}} \sigma^2_{\mathrm{mol}}) + \bar{\omega}^{\mathrm{Q}} b^{\mathrm{Q}}_{\mathrm{P}} \sigma^2_{\mathrm{cel}}, \tag{S10}$$

where O'' is omitted. Equation (S10) is identical to equations (1).

Finally, to derive the equation for $\Delta \bar{k}^{\mathrm{Q}}$ (i.e., $\Delta \langle \kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}} \rangle$), we swap P and Q in the above derivation. Moreover, we assume that $\sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle] = \sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle]$ and $\mathbb{E}_{\tilde{i}}[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{Q}}_{ij}, \kappa^{\mathrm{Q}}_{ij}]] = \mathbb{E}_{\tilde{i}}[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]]$ because no difference is a priori assumed between P and Q.

1.4 The mathematical analysis of numerical symmetry breaking

In this section, we show that numerical symmetry breaking occurs because while it is neither favoured nor disfavoured by molecular-level selection, it is favoured by cellular-level selection if catalytic and informatic symmetry breaking has occurred. To this end, we will again simplify the agent-based model into mathematical equations in a mannar analogous to that used to derive equations (1).

Before describing the mathematical analysis, we first need to note that the proximate—as opposed to ultimate—cause of numerical symmetry breaking is the self-replication of catalysts (i.e., $k^c_{cc} > 0$, where c is the replicator type of catalysts) in the absence of the reverse transcription of catalysts (i.e., $k^c_{tc} = 0$, where t is the replicator type of templates). This fact can be inferred from the following two results. First, when catalytic, informatic, and numerical symmetry breaking occurs, the replication and transcription of templates are catalysed at about the same rate, i.e., $k^c_{tt} \approx k^c_{ct}$ (Fig. 2b). Therefore, the replication and transcription of templates cannot cause numerical asymmetry. Second, when catalytic and informatic symmetry breaking occurs without numerical symmetry breaking, the self-replication of catalysts is absent (Fig. S5). Taken together, these results indicate that the proximate cause of numerical symmetry breaking is the self-replication of catalysts in the absence of the reverse transcription of catalysts. Therefore, to understand why numerical symmetry breaking occurs, we need to understand why the self-replication of catalysts evolves if catalytic and informatic symmetry breaking has occurred.

To address the above question, we assume that replicators have already undergone catalytic and informatic symmetry breaking and consider how the fitness of those replicators depends on the self-replication of catalysts. The population dynamics of replicators

with catalytic and informatic asymmetry can be described as follows. Let $n_{ij}^{t}(\tau)$ be the population size of replicator j of type t in protocell i at time τ. Let catalysts and templates be P and Q, respectively. Then, the dynamics of $n_{ij}^{t}(\tau)$ is mathematically described as follows:

$$\begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau+1) \\ n_{ij}^{\mathrm{Q}}(\tau+1) \end{bmatrix} = \begin{bmatrix} w_{ij}^{\mathrm{PP}} & \omega_{ij}^{\mathrm{Q}} \\ 0 & \omega_{ij}^{\mathrm{Q}} \end{bmatrix} \begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau) \\ n_{ij}^{\mathrm{Q}}(\tau) \end{bmatrix}, \tag{S11}$$

where w_{ij}^{PP} is the self-replication probability of catalysts, and ω_{ij}^{Q} is the replication and transcription probabilities of templates, which are assumed to be identical to each other. The fitness of replicators can be defined as the dominant eigenvalue (denoted by λ_{ij}) of the 2×2 matrix on the right-hand side of equation (S11):

$$\lambda_{ij} = \begin{cases} \omega_{ij}^{\mathrm{Q}} & \text{if } \omega_{ij}^{\mathrm{Q}} > w_{ij}^{\mathrm{PP}} \\ w_{ij}^{\mathrm{PP}} & \text{otherwise.} \end{cases} \tag{S12}$$

The associated right eigenvector, which determines the stationary frequencies of P and Q, is

$$\boldsymbol{v}_{ij} = \begin{cases} \frac{1}{2-w_{ij}^{\mathrm{PP}}/\omega_{ij}^{\mathrm{Q}}} \begin{bmatrix} 1 \\ 1 - w_{ij}^{\mathrm{PP}}/\omega_{ij}^{\mathrm{Q}} \end{bmatrix} & \text{if } \omega_{ij}^{\mathrm{Q}} > w_{ij}^{\mathrm{PP}} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{otherwise.} \end{cases} \tag{S13}$$

Equation (S13) shows that we must assume $\omega_{ij}^{\mathrm{Q}} > w_{ij}^{\mathrm{PP}}$ in order for P and Q to coexist. Equation (S13) also shows that the frequency of catalysts (i.e., $1/(2-w_{ij}^{\mathrm{PP}}/\omega_{ij}^{\mathrm{Q}})$) increases with the self-replication of catalysts (i.e., w_{ij}^{PP}), as stated in the beginning of this section.

We first examine whether the self-replication of catalysts is favoured by molecular-level selection. To this end, we consider how the fitness of replicators (i.e., λ_{ij}) depends on the self-replication of catalysts (i.e., w_{ij}^{PP}). According to equation (S12), λ_{ij} does not directly depend on w_{ij}^{PP}. However, λ_{ij} can indirectly depend on w_{ij}^{PP} because λ_{ij} increases with the frequency of catalysts in a protocell (i.e., $\mathbb{E}_{i\tilde{j}}[1/(2-w_{i\tilde{j}}^{\mathrm{PP}}/\omega_{i\tilde{j}}^{\mathrm{Q}})]$). This frequency increases with w_{ij}^{PP} if V is so small that a particular replicator can influence the frequency of catalysts in the protocell. However, if λ_{ij} increases with w_{ij}^{PP}, the average fitness of replicators in the protocell (i.e., $\langle \lambda_{i\tilde{j}} \rangle$) must also increase. Therefore, we need to consider the relative fitness (i.e., $\lambda_{ij}/\langle \lambda_{i\tilde{j}} \rangle$). The relative fitness is independent of w_{ij}^{PP} because catalysis is equally shared among templates within a protocell. Therefore, the self-replication of catalysts is neither favoured not disfavoured by molecular-level selection.

We next examine whether the self-replication of catalysts is favoured by cellular-level selection. To this end, we consider how the fitness of a protocell depends on the average self-replication of catalysts in that protocell (i.e., $\langle w_{i\tilde{j}}^{\mathrm{PP}} \rangle$). The fitness of a protocell can be defined as the average fitness of the replicators in that protocell (i.e., $\langle \lambda_{i\tilde{j}} \rangle$). According to equation (S12), $\langle \lambda_{i\tilde{j}} \rangle$ does not directly depend $\langle w_{i\tilde{j}}^{\mathrm{PP}} \rangle$. However, $\langle \lambda_{i\tilde{j}} \rangle$ indirectly depends on $\langle w_{i\tilde{j}}^{\mathrm{PP}} \rangle$ because $\langle \lambda_{i\tilde{j}} \rangle$ increases with the frequency of catalysts in a protocell (i.e., $\mathbb{E}_{i\tilde{j}}[1/(2-w_{i\tilde{j}}^{\mathrm{PP}}/\omega_{i\tilde{j}}^{\mathrm{Q}})]$). This frequency increases with $\langle w_{i\tilde{j}}^{\mathrm{PP}} \rangle$, so that $\langle \lambda_{i\tilde{j}} \rangle$ must also increase with $\langle w_{i\tilde{j}}^{\mathrm{PP}} \rangle$. Therefore, the self-replication of catalysts is favoured by cellular-level selection.

Taken together, the above considerations indicate that the self-replication of catalysts is neutral with respect to molecular-level selection, but advantageous with respect to

cellular-level selection. Therefore, numerical symmetry breaking results from the maximisation of fitness at the cellular level in the presence of catalytic and informatic asymmetry.

Finally, we mention an important consequence of numerical symmetry breaking. Numerical symmetry breaking causes a bottleneck effect on the population of replicators within a protocell. This bottleneck effect increases among-cell variance relative to within-cell variance (i.e., $\sigma^2_{\mathrm{cel}}/\sigma^2_{\mathrm{mol}}$); therefore, it has a stabilising effect on protocells [4, 5]. In this regard, numerical symmetry breaking can be compared to life-cycle bottlenecks displayed by multicellular organisms and eusocial colonies (i.e., an organism or colony develops from only one or a few propagules), which are considered to reduce within-group conflict [6–8].

1.5 The hierarchical Wright-Fisher model

In this section, we describe a model that stochastically simulates the population dynamics described by equations (1), in which σ^2_{mol} and σ^2_{cel} are treated as dynamic variables dependent on m and V.

The simplifications involved in the derivation of equations (1), while illuminating, make the comparison between equations (1) and the agent-based model indirect. Specifically, equations (1) cannot be compared with the agent-based model in terms of the same parameters, because the equations treat σ^2_{mol} and σ^2_{cel} as parameters, which are actually dynamic variables dependent on m and V in the agent-based model. To fill this gap, we constructed a model that stochastically simulates the population dynamics described by equations (1) and treats σ^2_{mol} and σ^2_{cel} as dynamic variables dependent on m and V.

This model is formulated as a hierarchical Wright-Fisher process. Replicators are partitioned into a number of groups (hereafter, protocells). Each replicator is individually assigned replicator type $c \in \{\mathrm{P}, \mathrm{Q}\}$ and two k^c values. The fitness of a replicator is calculated according to equation (S14). In each generation, replicators are replicated or transcribed with probabilities proportional to ω^c_{ij}, so that the population dynamics matches equation (S1) on average. After the replication-transcription step, the protocells containing greater than V replicators are divided with their replicators randomly distributed between the two daughter cells. The protocells containing no replicators are discarded.

The mutation of k^c is modelled as unbiased random walks with reflecting boundaries. That is, each k^c value of a replicator is mutated with a probability m per replication or transcription by adding a number randomly drawn from a uniform distribution on the interval $(-\delta_{\mathrm{mut}}, \delta_{\mathrm{mut}})$ ($\delta_{\mathrm{mut}} = 0.1$). The values of k^c are bounded in $[0, 1]$ with reflecting boundaries at both bounds.

To determine the condition for symmetry breaking, we simulated the above Wright-Fisher model for various values of V and m. The simulations show that symmetry breaking occurs only if V and m are sufficiently large (Fig. S8), a result that is consistent with the outcomes of the original agent-based model (Fig. 2). Given that the Wright-Fisher model involves many of the simplifications involved in equations (1), the above consistency supports the validity of the symmetry breaking mechanism described by equations (1).

1.6 The phase-plane analysis

In this section, we describe the phase-plane analysis outlined in Methods.

To perform the phase-plane analysis depicted in Fig. 3, we adapted equations (1) by defining ω^t_{ij} as a specific function of κ^t_{ij} (see the previous section for the meaning of ω^t_{ij}

and κ_{ij}^t). The following definition was employed:

$$\omega_{ij}^t = e^{\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle + \langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle} \frac{e^{-s\kappa_{ij}^t}}{\langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{P}}} \rangle + \langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{Q}}} \rangle}. \tag{S14}$$

where the factor $e^{\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle + \langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle}$ represents the cellular-level benefit of catalysis provided by the replicators in protocell i, the numerator $e^{-s\kappa_{ij}^t}$ represents the molecular-level cost of catalysis provided by the focal replicator, the denominator $1/(\langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{P}}} \rangle + \langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{Q}}} \rangle)$ normalises the cost, and s is the cost-benefit ratio. The above definition of ω_{ij}^t was chosen to satisfy the requirement that a replicator faces the trade-off between providing catalysis and serving as a template, so that γ_t^t and β_c^t are positive; for example, if the cost γ_t^t were negative, it would actually be a benefit, so that there would be no trade-off. This requirement is satisfied if $\partial\omega_{ij}^t/\partial\kappa_{ij}^t < 0$ and $\partial\langle\omega_{i\tilde{j}}^t\rangle/\partial\langle\kappa_{i\tilde{j}}^c\rangle > 0$ for $c = t$ and $c \neq t$. Apart from this requirement, the definition was arbitrarily chosen for simplicity.

Under the definition of ω_{ij}^t in equation (S14), we obtain equations describing the evolution of $\langle \kappa_{\tilde{i}\tilde{j}}^c \rangle$ (denoted as k^c in the main text) as follows. Since the evolution of $\langle \kappa_{\tilde{i}\tilde{j}}^c \rangle$ is described by equation (S6), we substitute equation (S14) into equation (S6). For this substitution, we need to calculate the derivatives of fitness. According to equation (S2), the fitness of a replicator is $\lambda_{ij} = \omega_{ij}^{\mathrm{P}} + \omega_{ij}^{\mathrm{Q}}$. Therefore,

$$\begin{aligned}
\mathbb{E}_{\tilde{i}}\left[\left.\frac{\partial\lambda_{ij}}{\partial\kappa_{ij}^c}\right|_{\substack{\kappa_{ij}^{\mathrm{P}}=\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle \\ \kappa_{ij}^{\mathrm{Q}}=\langle\kappa_{i\tilde{j}}^{\mathrm{Q}}\rangle}}\right] &= \mathbb{E}_{\tilde{i}}\left[-ce^{\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle+\langle\kappa_{i\tilde{j}}^{\mathrm{Q}}\rangle}\frac{e^{-s\langle\kappa_{i\tilde{j}}^c\rangle}}{\langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{P}}}\rangle + \langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{Q}}}\rangle}\right] \\
&= -ce^{\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}\rangle+\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}\rangle}\frac{e^{-s\langle\kappa_{\tilde{i}\tilde{j}}^c\rangle}}{e^{-s\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}\rangle} + e^{-s\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}\rangle}} + \mathbb{E}_{\tilde{i}}\big[O(\sigma_{i\tilde{j}}^2)\big] + O(\sigma_{\tilde{i}}^2).
\end{aligned}$$

Moreover, the average fitness of replicators in a protocell is $\langle\lambda_{i\tilde{j}}\rangle = e^{\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle+\langle\kappa_{i\tilde{j}}^{\mathrm{Q}}\rangle}$, so

$$\left.\frac{\partial\langle\lambda_{i\tilde{j}}\rangle}{\partial\langle\kappa_{i\tilde{j}}^c\rangle}\right|_{\substack{\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle=\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}\rangle \\ \langle\kappa_{i\tilde{j}}^{\mathrm{Q}}\rangle=\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}\rangle}} = e^{\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}\rangle+\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}\rangle}.$$

We substitute these derivatives into equation (S6) and use the fact that

$$\langle\lambda_{\tilde{i}\tilde{j}}\rangle = e^{\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}\rangle+\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}\rangle} + O(\sigma_{\tilde{i}}^2)$$

to obtain

$$\Delta\langle\kappa_{\tilde{i}\tilde{j}}^c\rangle = (1+\rho_{\mathrm{cel}})\sigma_{\mathrm{cel}}^2 - s\frac{e^{-s\langle\kappa_{\tilde{i}\tilde{j}}^c\rangle} + \rho_{\mathrm{mol}}e^{-s\langle\kappa_{\tilde{i}\tilde{j}}^{c'}\rangle}}{e^{-s\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}\rangle} + e^{-s\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}\rangle}}\sigma_{\mathrm{mol}}^2 + O'', \tag{S15}$$

where $c' \neq c$, ρ_{cel} is the correlation coefficient between $\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle$ and $\langle\kappa_{i\tilde{j}}^{\mathrm{Q}}\rangle$ (i.e., $\rho_{\mathrm{cel}} = \sigma_{\tilde{i}}^2[\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle, \langle\kappa_{i\tilde{j}}^{\mathrm{Q}}\rangle]/\sigma_{\mathrm{cel}}^2$), and ρ_{mol} is the average correlation coefficient between κ_{ij}^{P} and κ_{ij}^{Q} (i.e., $\rho_{\mathrm{mol}} = \mathbb{E}_{\tilde{i}}[\sigma_{i\tilde{j}}^2[\kappa_{ij}^{\mathrm{P}}, \kappa_{ij}^{\mathrm{Q}}]]/\sigma_{\mathrm{mol}}^2$). To derive equation (S15), we have assumed that the variances of $\langle\kappa_{i\tilde{j}}^c\rangle$ and κ_{ij}^c are independent of c; i.e., $\sigma_{\mathrm{cel}}^2 = \sigma_{\tilde{i}}^2[\langle\kappa_{i\tilde{j}}^c\rangle, \langle\kappa_{i\tilde{j}}^c\rangle]$ and $\sigma_{\mathrm{mol}}^2 = \mathbb{E}_{\tilde{i}}[\sigma_{i\tilde{j}}^2[\kappa_{ij}^c, \kappa_{ij}^c]]$ for $c = \mathrm{P}$ and $c = \mathrm{Q}$.

Equation (S15) can be expressed in a compact form as follows:

$$\begin{bmatrix}\Delta\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}\rangle \\ \Delta\langle\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}\rangle\end{bmatrix} = \sigma_{\mathrm{tot}}^2\nabla\big[RB - (1-R)C\big] + O'',$$

where ∇ is a nabla operator (i.e., $\nabla = [\partial/\partial\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle, \partial/\partial\langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle]^{\mathrm{T}}$, where ^{T} denotes transpose), $\sigma^2_{\mathrm{tot}} = \sigma^2_{\mathrm{mol}} + \sigma^2_{\mathrm{cel}}$, $R = \sigma^2_{\mathrm{cel}}/(\sigma^2_{\mathrm{cel}} + \sigma^2_{\mathrm{mol}})$, $B = (1 + \rho_{\mathrm{cel}})(\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}} + \kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}})$, and $C = (\rho_{\mathrm{mol}} - 1)\ln(e^{-s\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}} + e^{-s\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}}) + \rho_{\mathrm{mol}} s(\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}} + \kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}})$. R can be interpreted as the regression coefficient of $\langle\kappa^c_{\tilde{i}\tilde{j}}\rangle$ on κ^c_{ij} [9] and, therefore, the coefficient of genetic relatedness [10]. The potential function $RB - (1 - R)C$ can then be interpreted as inclusive fitness.

Next, we set $\rho_{\mathrm{mol}} = 0$ and $\rho_{\mathrm{cel}} = 0$ in equations (S15) and let $\langle\kappa^c_{\tilde{i}\tilde{j}}\rangle$ be denoted by $\bar{k}^c$, obtaining

$$\begin{aligned}\Delta\bar{k}^c &= \sigma^2_{\mathrm{cel}} - s\frac{e^{-s\bar{k}^c}}{e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}}\sigma^2_{\mathrm{mol}} + O'' \\ &= \frac{e^{-s\bar{k}^c}}{e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}}(\sigma^2_{\mathrm{cel}} - s\sigma^2_{\mathrm{mol}}) + \frac{e^{-s\bar{k}^{c'}}}{e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}}\sigma^2_{\mathrm{cel}} + O'',\end{aligned} \tag{S16}$$

where $c' \neq c$. Comparing equations (S16) and (S10), we infer that

$$\begin{aligned}\bar{\omega}^c &= \frac{e^{-s\bar{k}^c}}{e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}}, \\ \gamma^c_c &= s, \\ \beta^t_c &= 1,\end{aligned}$$

which are identical to equations (5).

Next, we omit O'' in equation (S16) and replace Δ with time derivative $d/d\tau$, obtaining

$$\frac{d}{d\tau}\bar{k}^c = \sigma^2_{\mathrm{cel}} - s\frac{e^{-s\bar{k}^c}}{e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}}\sigma^2_{\mathrm{mol}}. \tag{S17}$$

Finally, to allow for the restriction on the range of $\bar{k}^c$ (i.e., $\bar{k}^c \in [0, k_{\max}]$), we multiply the right-hand side of equation (S17) with a function, denoted by $\Theta(\bar{k}^c)$, that is 1 if $0 < \bar{k}^c < k_{\max}$ and 0 if $\bar{k}^c = 0$ or $\bar{k}^c = k_{\max}$. Multiplying $\Theta(\bar{k}^c)$ with the right-hand side of equation (S17), we obtain

$$\frac{d}{d\tau}\bar{k}^c = \Theta(\bar{k}^c)\left[\sigma^2_{\mathrm{cel}} - s\frac{e^{-s\bar{k}^c}}{e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}}\sigma^2_{\mathrm{mol}}\right].$$

The above equation was numerically integrated for $s = 1$ to obtain the phase-plane portrait depicted in Fig. 3.

Equation (S15) allows for statistical correlations between κ^{P}_{ij} and κ^{Q}_{ij} at the molecular and cellular levels, i.e., ρ_{mol} and ρ_{cel}. Therefore, it can be used to examine the consequence of ignoring these correlations, which is one of the simplifications made in the derivation of equations (1) described in SI Text 1.3. For this sake, we calculate the nullcline of $\Delta\langle\kappa^c_{\tilde{i}\tilde{j}}\rangle$. Setting $\Delta\langle\kappa^c_{\tilde{i}\tilde{j}}\rangle = 0$ in equation (S15) and omitting O'', we obtain

$$\langle\kappa^{c'}_{\tilde{i}\tilde{j}}\rangle \approx \langle\kappa^c_{\tilde{i}\tilde{j}}\rangle + s^{-1}\ln\frac{\rho_{\mathrm{mol}} s\sigma^2_{\mathrm{mol}} - (1 + \rho_{\mathrm{cel}})\sigma^2_{\mathrm{cel}}}{(1 + \rho_{\mathrm{cel}})\sigma^2_{\mathrm{cel}} - s\sigma^2_{\mathrm{mol}}}.$$

This equation shows that all parameters only appear in the intercept of the nullcline with the $\langle\kappa^{c'}_{\tilde{i}\tilde{j}}\rangle$-axis. Let us denote this intercept as $s^{-1}\ln I$. The way I qualitatively depends on σ^2_{cel} and $s\sigma^2_{\mathrm{mol}}$ is independent of ρ_{cel} because $-1 < \rho_{\mathrm{cel}} < 1$. Therefore, we

can assume that $\rho_{\mathrm{cel}} = 0$ without loss of generality. Next, to see how ρ_{mol} influences I, we focus on the singularity of I by setting $(1 + \rho_{\mathrm{cel}})\sigma^2_{\mathrm{cel}} = s\sigma^2_{\mathrm{mol}} + \epsilon$, where $\epsilon > 0$. Then, $I = (1 - \rho_{\mathrm{mol}})s\sigma^2_{\mathrm{mol}}/\epsilon - \rho_{\mathrm{mol}}$. The way I qualitatively depends on $s\sigma^2_{\mathrm{mol}}/\epsilon$ is independent of ρ_{mol} because $-1 < \rho_{\mathrm{mol}} < 1$. Therefore, we can assume that $\rho_{\mathrm{mol}} = 0$ without loss of generality. Taken together, these calculations show that ignoring correlations between κ^{P}_{ij} and κ^{Q}_{ij} does not qualitatively affect the results, supporting the validity of equations (1).

2 Supporting Figures

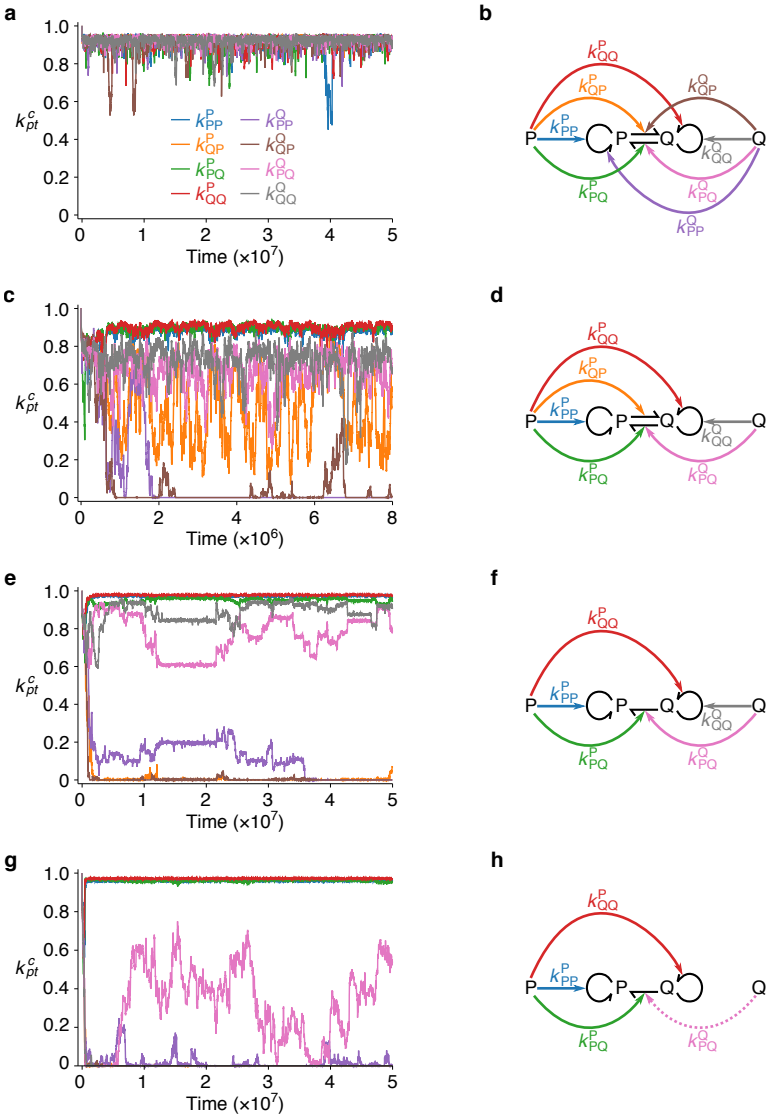

Figure S1: The evolutionary dynamics of the agent-based model. **a**, The dynamics of k^c_{pt} averaged over all replicators for parameters corresponding to 'no symmetry breaking' in Fig. 2a: $V = 178$ and $m = 0.01$. **b**, Catalytic activities evolved in a. **c**, **d**, Parameters corresponding to 'uncategorised' in Fig. 2a: $V = 178$ and $m = 0.1$. **e**, **f**, Parameters corresponding to 'incomplete symmetry breaking' in Fig. 2a: $V = 562$ and $m = 0.01$. **g**, **h**, Parameters corresponding to 'incomplete symmetry breaking' in Fig. 2a: $V = 1778$ and $m = 0.01$.

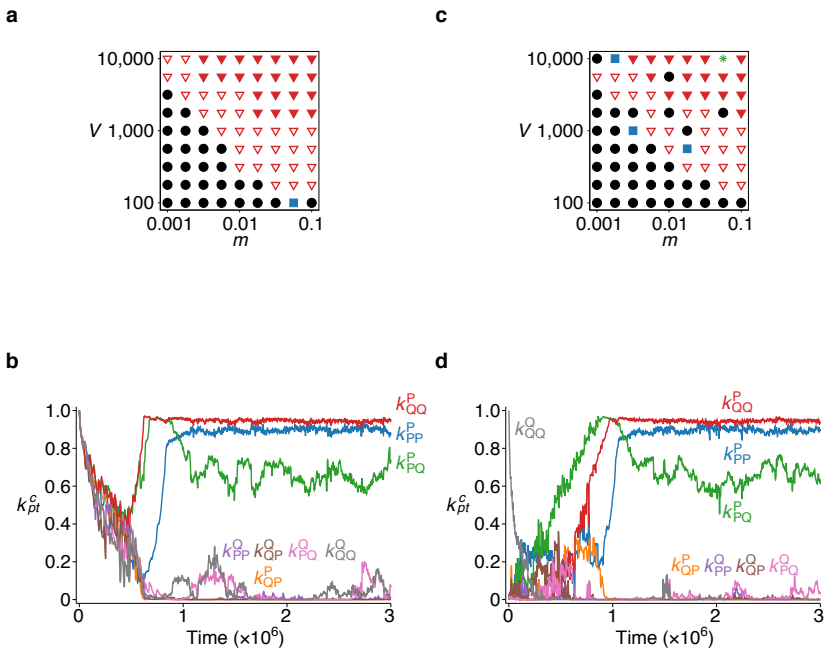

Figure S2: Symmetry breaking with an alternative definition of complex formation rates (see SI Text 1.1). The rate constants of complex formation were defined in such a way that coexistence between P and Q is neither favoured nor disfavoured by cellular-level selection. **a**, Phase diagram with a symmetric initial condition: $k^c_{pt} = 1$ for all combinations of c, p, and t, with both P and Q present at the beginning of each simulation. The symbols are the same as in Fig. 2a, except that the circles include cases in which one replicator type goes extinct. **b**, Dynamics of k^c_{pt} averaged over all replicators for $m = 0.01$ and $V = 10000$ in a. **c**, Phase diagram with an asymmetric initial condition: $k^{\mathrm{Q}}_{\mathrm{QQ}} = 1$ and $k^c_{pt} = 0$ for all the other combinations of c, p, and t, with only Q present at the beginning of each simulation. The symbols are the same as in a, except that stars indicate the extinction of replicators. **d** Dynamics of k^c_{pt} averaged over all replicators for $m = 0.01$ and $V = 10000$ in b.

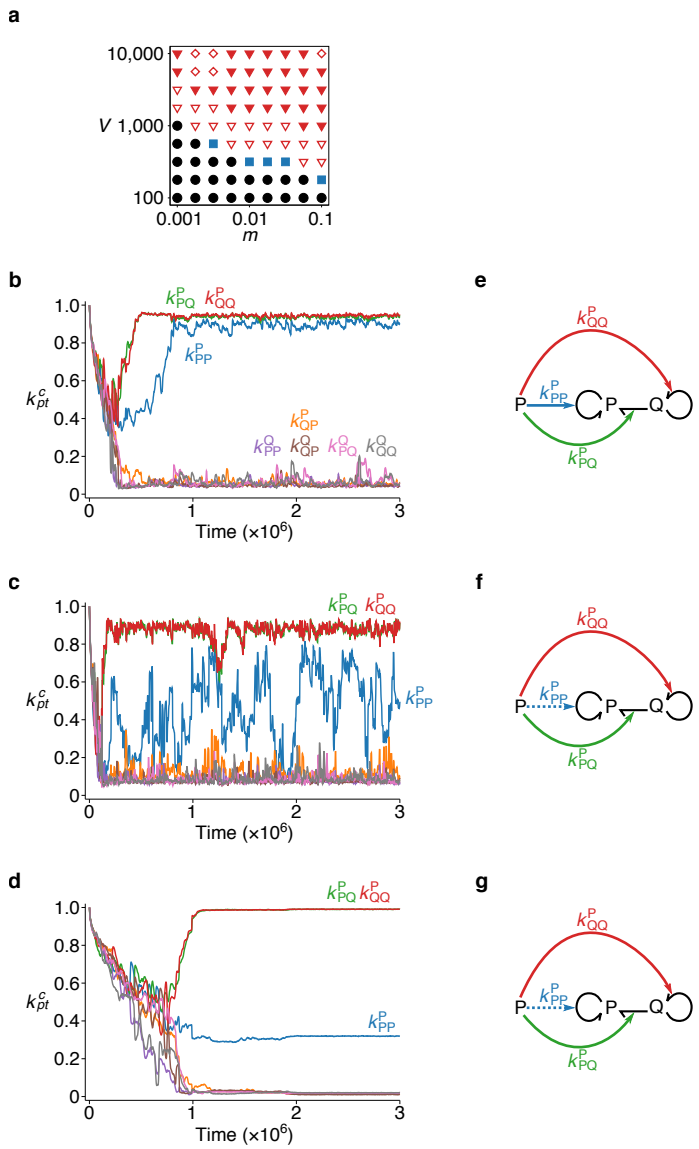

Figure S3: Symmetry breaking with reflecting mutation (see SI Text 1.2). The mutation of k^c_{pt} is modelled as unbiased random walk with reflecting boundaries at 0 and 1. **a**, Phase diagram. The symbols are the same as in Fig. 2a ($t_{\min} > 3.9 \times 10^7$ for $m = 0.1$ and $V = 10000$). **b** Dynamics of k^c_{pt} averaged over all replicators. $m = 0.01$ and $V = 10000$. Three-fold symmetry breaking occurs. **c**, $m = 0.0562$ and $V = 10000$. Numerical symmetry breaking is slight. **d**, $m = 0.00178$ and $V = 10000$. Numerical symmetry breaking is slight. **e**, **f**, **g**, Catalytic activities evolved in b, c, d, respectively.

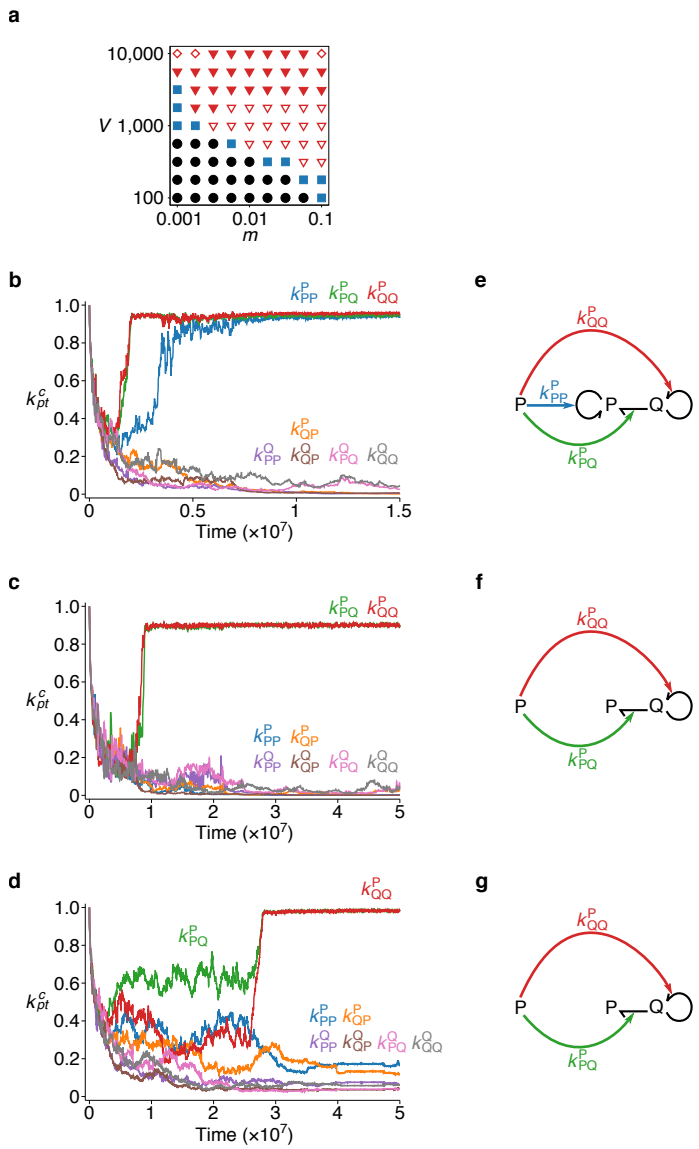

Figure S4: Symmetry breaking with log-space mutation (see SI Text 1.2). The mutation of k^c_{pt} is modelled as unbiased random walks on a logarithmic scale. **a**, Phase diagram. The symbols are the same as in Fig. 2a ($t_{\min} > 3.9 \times 10^7$ only for $m = 0.1$ and $V = 10000$). **b**, Dynamics of k^c_{pt} averaged over all replicators. $m = 0.01$ and $V = 10000$. Three-fold symmetry breaking occurs. **c**, $m = 0.1$ and $V = 10000$. No numerical symmetry breaking occurs. **d**, $m = 0.00178$ and $V = 10000$. No numerical symmetry breaking occurs. **e**, **f**, **g**, Catalytic activities evolved in b, c, d, respectively.

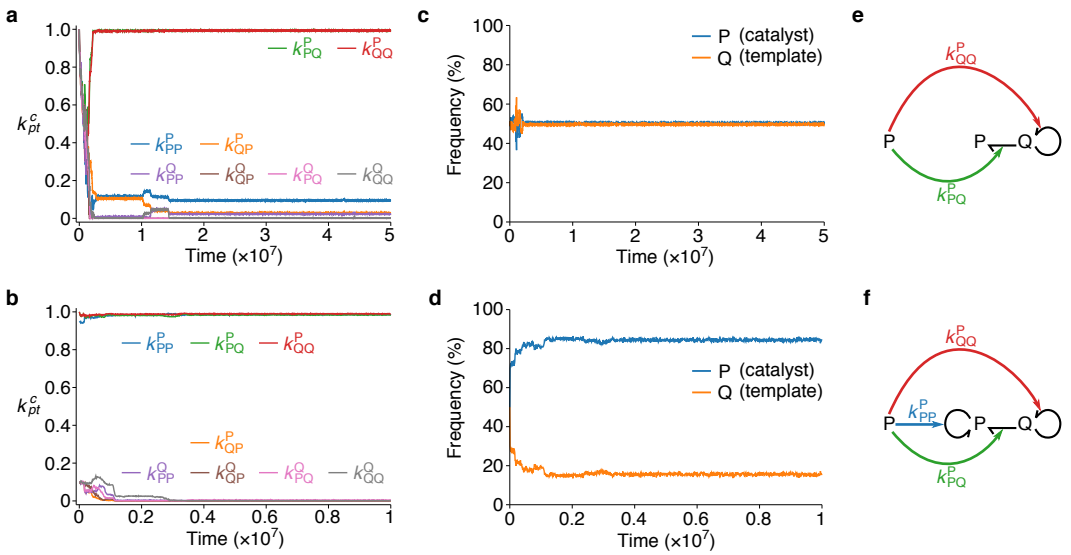

Figure S5: The absence of numerical symmetry breaking for small m and large V (see SI Text 1.4). **a**, **b**, The dynamics of k^c_{pt} averaged over all replicators is shown for $V = 10000$ and $m = 0.001$ with two different initial conditions: a symmetric initial condition, where $k^c_{pt} = 1$ (a); an asymmetric initial condition, where $k^{\mathrm{P}}_{\mathrm{PP}} = 0.95$, $k^{\mathrm{P}}_{\mathrm{PQ}} = 0.1$, $k^{\mathrm{P}}_{\mathrm{QP}} = 1$, $k^{\mathrm{P}}_{\mathrm{QQ}} = 1$, and $k^{\mathrm{Q}}_{pt} = 0.1$ (b). The self-replication of catalysts does not evolve for the symmetric initial condition, whereas it is maintained for the asymmetric initial condition ($t_{\min} > 1.2 \times 10^7$). The dependence of the results on the initial conditions suggests the presence of bistability for $V = 10000$ and $m = 0.001$. **c**, **d**, The frequencies of P (catalysts) and Q (templates) are plotted as the functions of time. Numerical symmetry breaking does not occur for the symmetric initial condition, whereas it occurs for the asymmetric initial condition. The results indicate that numerical asymmetry depends on the self-replication of catalysts. **e**, **f**, Catalytic activities evolved for the symmetric initial condition (e) and for the asymmetric initial condition (f).

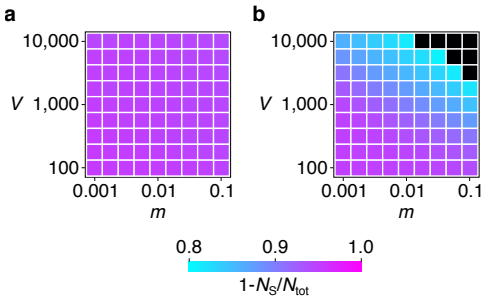

Figure S6: The effect of symmetry breaking on catalytic activities. The fraction of replicators $1 - N_{\mathrm{S}}/N_{\mathrm{tot}}$, which is a proxy for the overall catalytic activity of replicators, is shown as a function of m and V, where N_{S} is the total number of S molecules in the system, and $N_{\mathrm{tot}} = N_{\mathrm{P}} + N_{\mathrm{Q}} + N_{\mathrm{S}}$. **a**, The original model, which allows symmetry breaking (i.e., Fig. 1). **b**, The model that excludes the possibility of symmetry breaking; specifically, it allows only one type of replicator (either P or Q). Black squares indicate extinction (i.e. $N_{\mathrm{tot}} = N_{\mathrm{S}}$). $t_{\mathrm{min}} > 1.5 \times 10^7$.

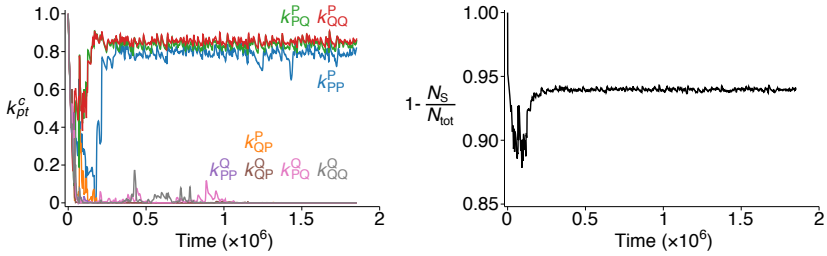

Figure S7: Result for large m and V values. The dynamics of the agent-based model is shown for $m = 0.1$ and $V = 10^5$, parameters outside the range examined in Fig. 2a and Fig. S6a. **a**, The dynamics of k^c_{pt} averaged over all replicators. **b**, The dynamics of the fraction of replicators $1 - N_S/N_{tot}$, where N_{tot} and N_S are the total numbers of particles and S molecules in the system, respectively. $t_{min} > 1.8 \times 10^6$.

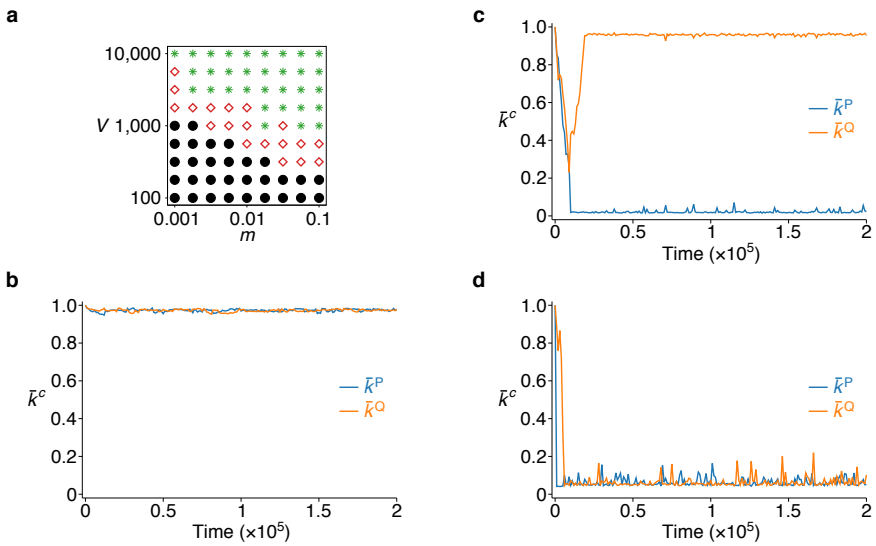

Figure S8: Symmetry breaking in a hierarchical Wright-Fisher model (see SI Text 1.5). The model stochastically simulates the population dynamics described by equations (1), treating σ^2_{mol} and σ^2_{cel} as variables dependent on m and V (see SI Text 1.5). **a**, Phase diagram. Circles indicate no symmetry breaking (i.e., $\bar{k}^{\mathrm{P}} \approx \bar{k}^{\mathrm{Q}} \approx 1$); diamonds, symmetry breaking (i.e., $\bar{k}^{c} \approx 0$ and $\bar{k}^{c'} \approx 1$ for $c \neq c'$); stars, extinction (i.e., $\bar{k}^{\mathrm{P}} \approx \bar{k}^{\mathrm{Q}} \approx 0$). $s = 1$ (cost-benefit ratio). The total number of replicators was $50V$ (approximately 130 protocells throughout simulations). The initial condition was $k^{\mathrm{P}} = k^{\mathrm{Q}} = 1$ for all replicators. Each simulation was run for 4×10^5 generations. The extinction (i.e., $\bar{k}^{\mathrm{P}} \approx \bar{k}^{\mathrm{Q}} \approx 0$) for large m and V is consistent with the phase-plane analysis of equations (1), which also shows extinction (i.e., $\bar{k}^{\mathrm{P}} \approx \bar{k}^{\mathrm{Q}} \approx 0$) for sufficiently large $\sigma^2_{\mathrm{mol}}/\sigma^2_{\mathrm{cel}}$ (parameters outside the range examined in Fig. 3). The discrepancy between Fig. S8a and Fig. 2a is due the simplifying assumption made in equations (1) that k^c_{pt} is independent of p and t. If k^c_{pt} is allowed to depend on p and t, the flow of information from templates to catalysts can become completely unidirectional. Such unidirectional flow of information can resolve the dilemma between catalysing and templating and leads to the maintenance of high catalytic activities as described in Results. **b**, The dynamics of $\bar{k}^c$ for $m = 0.001$ and $V = 1000$ (no symmetry breaking). **c**, $m = 0.01$ and $V = 1000$ (symmetry breaking). **d**, $m = 0.1$ and $V = 1000$ (extinction).

References

[1] Price GR. Extension of covariance selection mathematics. Annals of Human Genetics. 1972;35(4):485–490.

[2] Hamilton WD. Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox R, editor. Biosocial Anthoroplogy. London: Malaby Press; 1975. p. 133–153.

[3] Iwasa Y, Pomiankowski A, Nee S. The evolution of costly mate preferences II. the 'handicap' principle. Evolution. 1991;45(6):1431–1442.

[4] Kaneko K, Yomo T. On a kinetic origin of heredity: minority control in a replicating system with mutually catalytic molecules. Journal of Theoretical Biology. 2002;214(4):563–576.

[5] Takeuchi N, Hogeweg P, Kaneko K. The origin of a primordial genome through spontaneous symmetry breaking. Nature Communications. 2017;8(1):250.

[6] Maynard Smith J, Szathmáry E. The Major Transitions in Evolution. Oxford: W. H. Freeman/Spektrum; 1995.

[7] Michod RE. Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality. Princeton, NJ: Princeton University Press; 1999.

[8] Bourke AFG. Principles of Social Evolution. Oxford, UK: Oxford University Press; 2011.

[9] Rice SH. Evolutionary Theory: Mathematical and Conceptual Foundations. Sunderland, MA, USA: Sinauer Associates; 2004.

[10] Hamilton WD. Selfish and spiteful behaviour in an evolutionary model. Nature. 1970;228(5277):1218–1220.