The origin of the central dogma through conflicting multi-level 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

The central dogma of molecular biology rests on two kinds of asymmetry between genomes and enzymes[1]. Information flows from genomes to enzymes, but not from enzymes to genomes: informatic asymmetry. Enzymes provide catalysis, whereas genomes do not: catalytic asymmetry. How did these asymmetries originate? Here we demonstrate that these asymmetries can spontaneously arise from conflict between selection at the molecular level and selection at the cellular level. Our model consists 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 multi-level selection: serving as catalysts is favoured by cellular-level selection, whereas serving as templates is favoured by molecular-level selection. This conflict induces informatic and catalytic symmetry breaking, whereby the molecules differentiate into genomes and enzymes, hence establishing the central dogma. We show mathematically that the symmetry breaking is caused by positive feedback between

*nobuto.takeuchi@auckland.ac.nz

Fisher's reproductive values and the relative impact of selection at different levels. Our work proposes that the central dogma is a logical consequence of conflicting multi-level selection, hence making it no longer a 'dogma.'

At the heart of living systems lies the distinction between genomes and enzymes—the division of labour between the transmission of genetic information and the provision of chemical catalysis. 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[2] or its analogues[3]). How then did this 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[4–7]. During this transition, selection operated at both molecular and cellular levels, and selection at one level was potentially in conflict with selection at the other. Previously, we demonstrated that such conflicting multi-level selection can induce catalytic symmetry breaking in replicating molecules[8]. We thus hypothesised that conflicting multi-level selection could also induce the evolution of the genome-enzyme distinction and, hence, the origin of the central dogma.

To examine this hypothesis, we consider a model with two types of replicators, denoted by P and Q. The chemical identity of P and Q is unspecified for simplicity and generality. For simplicity, we separate the origin of the genome-enzyme distinction from the origin of protein translation. For generality, we formulate our model to be independent of chemical specifics (see also Supplementary Discussion 1). To examine the possibility of spontaneous symmetry breaking, we assume no a priori difference between P and Q. We assume that both P and Q can serve as templates for replication ($\mathrm{P} \to 2\mathrm{P}$ and $\mathrm{Q} \to 2\mathrm{Q}$) and transcription ($\mathrm{P} \to \mathrm{P} + \mathrm{Q}$ and $\mathrm{Q} \to \mathrm{Q} + \mathrm{P}$), where complementarity is ignored (Fig. 1a). Moreover, both P and Q can serve as catalysts for replication and transcription. Each replicator is individually assigned eight catalytic values denoted by $k^c_{pt} \in [0, 1]$, where c, p, and t are the replicator types of catalyst, product, and template, respectively. Four of these k^c_{pt} values denote the catalytic activities of the replicator itself, and the other four denote those of its transcripts; e.g., if a replicator is of type P, its catalytic activities are given by k^{P}_{pt}, and those of its transcripts, which are of type Q, by k^{Q}_{pt}. A replicator inherits k^c_{pt} values from its template with potential mutation. Mutation randomly changes each k^c_{pt} value with probability m per replication or transcription (see Methods). For simplicity, catalysts are assumed not to distinguish between different templates of the same replicator type (either because catalysts are unspecific or because templates are sufficiently similar to each other).

Replicators compete for a finite supply of substrate denoted by S (the abstraction

of monomers). The substrate is recycled through the decay of P and Q to keep the total number of P, Q, and S (hereafter, collectively called particles) constant (Fig. 1b).

All particles are compartmentalised into protocells, across which P and Q do not diffuse at all, but S diffuses rapidly (Fig. 1c; see Methods). This difference in diffusion induces the passive transport of S from protocells in which S is converted into replicators slowly, to protocells in which this conversion is rapid. Consequently, the latter grow at the expense of the former[9]. 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 is considered to arise from a constraint that providing catalysis and serving as a template impose structurally-incompatible requirements on replicators[10,11]. 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 self-replication. To incorporate this trade-off, the model assumes that replication and transcription entail complex formation between a catalyst and template (Fig. 1b)[12]. The rate constants of complex formation are given by the k^c_{pt} values of a replicator serving as a catalyst. Thus, the greater the values of k^c_{pt}, the greater the chance that a replicator, or its transcript, is sequestered in a complex as a catalyst and thus unable to serve as a template.

The above trade-off creates a dilemma: providing catalysis brings benefit at the cellular level because it accelerates a protocell's uptake of S, but brings cost at the molecular level because it inhibits a replicator's self-replication[8]. Therefore, selection between protocells tends to maximise the k^c_{pt} values of replicators (i.e., cellular-level selection), whereas selection within protocells tends to minimise the k^c_{pt} values of replicators (i.e., molecular-level selection).

To determine the outcome of this conflicting multi-level selection, we simulated our 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 symmetry breaking in three aspects (Fig. 2a). First, one type of replicator (either P or Q) evolves high catalytic activity, whereas the other completely loses it (i.e., $k^c_{pt} \gg k^{c'}_{pt} \approx 0$ 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., $k^c_{ct} \gg k^c_{tc} \approx 0$): 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 Supplementary Discussion 2 and 3; Extended Data Figs. 3, 4, and 5). Below, we focus on catalytic and informatic symmetry breaking because they are directly related to the central dogma (see Supplementary Discussion 4 for numerical symmetry breaking).

The significant consequence of 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; see Methods for ancestor tracking). 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, independent of competition between catalysts. How the catalytic activity of catalysts evolves, therefore, depends solely on the cost and benefit to templates. On the 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, but not minimised by molecular-level selection, hence the resolution of the dilemma between catalysing and templating. Because of this resolution, symmetry breaking leads to the maintenance of high catalytic activities (Extended Data Figs. 6 and 7).

To understand the mechanism of the symmetry breaking, we simplified the model into mathematical equations. These equations allow us to consider all the costs and benefits involved in providing catalysis: for catalysis provided by $c \in \{\mathrm{P}, \mathrm{Q}\}$, its molecular-level cost to c (denoted by γ_c^c), and its cellular-level benefits to $t \in \{\mathrm{P}, \mathrm{Q}\}$ (denoted by β_c^t). 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[13–17] and displayed below (see Methods 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, σ_{cel}^2 is the variance of catalytic activities among protocells (cellular-level variance), and σ_{mol}^2 is the variance of catalytic activities within a protocell (molecular-level variance).

The derivation of equations (1) involves various simplifications, among which the three most important are noted below (see Methods for details). First, equations (1) assume that catalysts do not distinguish the replicator types of templates and products (i.e., k^c_{pt} is independent of p and t, hence denoted by k^c). Such distinction is required for numerical symmetry breaking, which is thus excluded under this assumption. However, 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. Second, equations (1) treat σ^2_{mol} and σ^2_{cel} as parameters although they are actually variables dependent on m and V in the simulation model. In addition, these variances are assumed to be identical between $\bar{k}^{\mathrm{P}}$ and $\bar{k}^{\mathrm{Q}}$. Third, equations (1) ignore the terms of order greater than σ^2_{cel} and σ^2_{mol}[16].

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 $\beta^t_c\sigma^2_{\mathrm{cel}}$ represent evolution driven by cellular-level selection; those by $-\gamma^c_c\sigma^2_{\mathrm{mol}}$, evolution driven by molecular-level selection.

Using equations (1), we can now elucidate the mechanism of 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 positive feedback between reproductive values and the relative impacts of selection at different levels.

To link the above analysis to the simulation model, we need to allow for the restriction on the range of $\bar{k}^c$ (i.e., $\bar{k}^c \in [0,1]$). This restriction can be taken into

account through a phase-plane analysis of equations (1), which we have performed numerically (see Methods). The results shown in Fig. 3 indicate that symmetry breaking occurs only when $\sigma^2_{\mathrm{mol}}/\sigma^2_{\mathrm{cel}}$ is sufficiently large (i.e., when genetic relatedness R is sufficiently small, where $R = \sigma^2_{\mathrm{cel}}/(\sigma^2_{\mathrm{mol}} + \sigma^2_{\mathrm{cel}})$[17–19]; see Methods). This result is consistent with the simulation model (Fig. 2a) because by the law of large numbers, cellular-level variance (σ^2_{cel}) decreases relative to molecular-level variance (σ^2_{mol}) as V increases[8,20] (see Supplementary Discussion 5 and Extended Data Fig. 8 for an additional confirmation in terms of V and m instead of $\sigma^2_{\mathrm{mol}}/\sigma^2_{\mathrm{cel}}$). This result indicates that equations (1) correctly describe the mechanism of symmetry breaking in the simulation model.

In summary, our results show that a positive feedback between conflicting multi-level selection and reproductive values causes symmetry breaking of replicators that establishes a division of labour between the transmission of genetic information and the provision of chemical catalysis. Such division of labour between information transmission and the other functions is a recurrent pattern throughout biological hierarchy; e.g., multicellular organisms display differentiation between germline and soma; eusocial animal colonies, queens and workers (Extended Data Table 1)[4–7]. Given that all these systems potentially involve conflicting multi-level selection and tend to display the respective division of labour as their sizes increase, our theory provides a basis on which to pursue a universal principle of life.

Figures

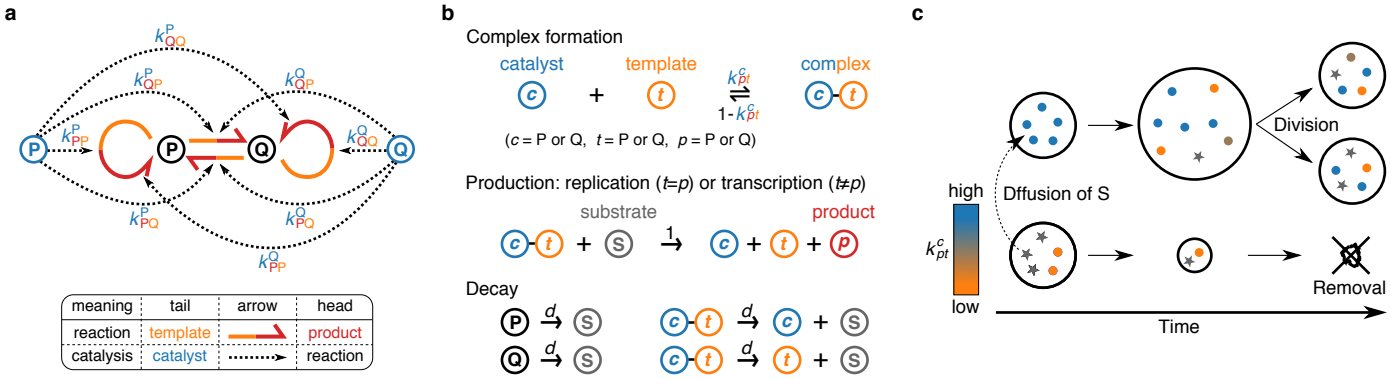


Figure 1: **The model.** **a**, Two types of replicators, P and Q, can serve as templates and catalysts for producing either type. Circular harpoons indicate replication; straight harpoons, transcription (heads indicate products; tails, templates). Dotted arrows indicate catalysis (heads indicate reaction 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 (denoted by c). **c**, Protocells exchange substrate S (represented by stars) through rapid diffusion. They divide when the number of internal particles exceeds V. They are removed when they lose all particles. See Methods for the details of the model.

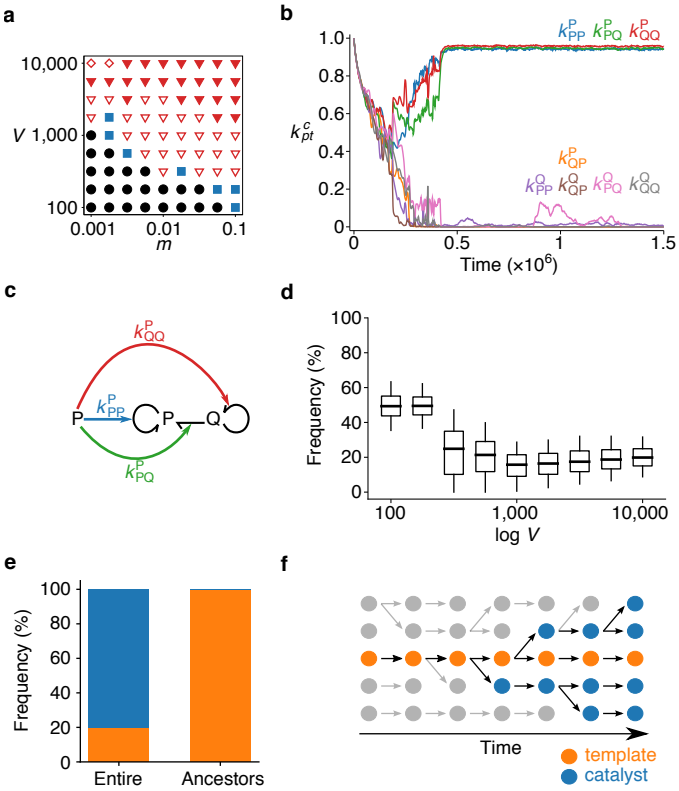


Figure 2: **The evolution of the central dogma. a**, Phase diagram: circles indicate no symmetry breaking (Extended Data Fig. 1ab); squares, uncategorised (Extended Data Fig. 1cd); open triangles, incomplete symmetry breaking (Extended Data Fig. 1e-h); filled triangles, three-fold symmetry breaking as depicted in b and c; diamonds, catalytic and informatic symmetry breaking without numerical symmetry breaking (Extended Data Fig. 2a). 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**, Replicator evolving 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 $V/2$ molecules 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 are catalysts.

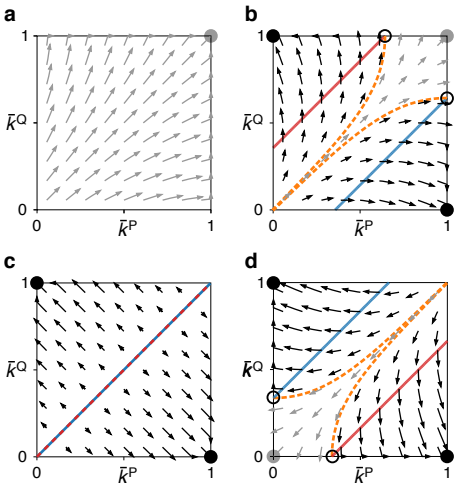


Figure 3: **Phase-plane portrait.** For this figure, equations (1) were adapted as described in Methods, and Δ was replaced with time derivative ($\frac{d}{d\tau}$). Solid lines indicate nullclines: $\frac{d}{d\tau}\bar{k}^{\mathrm{P}} = 0$ (red) and $\frac{d}{d\tau}\bar{k}^{\mathrm{Q}} = 0$ (blue). 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 demarcate basins of attraction. To ensure that $0 \leq \bar{k}^{c} \leq 1$, $\frac{d}{d\tau}\bar{k}^{c}$ is set to 0 if $\bar{k}^{c} = 0$ or $\bar{k}^{c} = 1$. The nullclines at $\bar{k}^{c} = 0$ and $\bar{k}^{c} = 1$ are not depicted for visibility. Parameters: $\sigma^2_{\mathrm{cel}} = 1$, $s = 1$, $\rho_{\mathrm{cel}} = 0$, $\rho_{\mathrm{mol}} = 0$ (see Methods). **a**, $\sigma^2_{\mathrm{mol}}/\sigma^2_{\mathrm{cel}} = 1.3$. Cellular-level variance is so large relative to molecular-level variance that $\bar{k}^{c}$ is always maximised. **b**, $\sigma^2_{\mathrm{mol}}/\sigma^2_{\mathrm{cel}} = 1.7$. Asymmetric equilibria emerge, but cellular-level variance is large enough to make $\bar{k}^{\mathrm{P}} = \bar{k}^{\mathrm{Q}} = 1$ stable. **c**, $\sigma^2_{\mathrm{mol}}/\sigma^2_{\mathrm{cel}} = 2.0$. The tipping point. **d**, $\sigma^2_{\mathrm{mol}}/\sigma^2_{\mathrm{cel}} = 2.4$. Cellular-level variance is small enough to make $\bar{k}^{\mathrm{P}} = \bar{k}^{\mathrm{Q}} = 1$ unstable. The asymmetric equilibria can be reached if $\bar{k}^{\mathrm{P}} \approx \bar{k}^{\mathrm{Q}} \approx 1$.

Methods

The model.

The model treats each molecule as a distinct individual with uniquely-assigned k^c_{pt} variables. 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[8]. 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 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. Therefore, X and Y can form four distinct complexes depending on which replicator serves as a catalyst (X or Y) and which type of replicator is being produced (p = P or p = Q).

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. Thus, the 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 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). Therefore, all k^c_{pt} values of replicators tend to be maximised by cellular-level selection. If all k^c_{pt} values are maximised, P and Q coexist. Thus, 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 this specific aspect of the model does not critically affect results by examining an alternative model in which coexistence between P and Q is neutral with respect to cellular-level selection (see Supplementary Discussion 2).

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 the main text.

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 $(-\delta_{\rm mut}, \delta_{\rm mut})$ ($\delta_{\rm mut}$ = 0.05 unless otherwise stated). The values of k^c_{pt} are bounded above by $k_{\rm max}$ with a reflecting boundary ($k_{\rm max}$ = 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 Supplementary Discussion 3.

Each simulation was run for at least 5×10^7 time steps (denoted by $t_{\rm min}$) 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 $N_{\rm tot}$ 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 $k_{\rm max}$ for every replicator unless otherwise stated. The initial frequencies of P and Q were equal, and that of S was zero.

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.

The derivation of equations (1).

To derive equations (1), we simplified the simulation 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 $p \neq p'$). As described in the main text, this assumption excludes the possibility of numerical symmetry breaking, but still allows catalytic and informatic symmetry breaking.

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

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

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). 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:

$$\lambda_{ij} = \omega^{\mathrm{P}}_{ij} + \omega^{\mathrm{Q}}_{ij}, \qquad \boldsymbol{v}_{ij} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{u}_{ij} = \begin{bmatrix} \omega^{\mathrm{P}}_{ij} & \omega^{\mathrm{Q}}_{ij} \end{bmatrix}. \tag{3}$$

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^{P}_{ij} be the catalytic activity of replicator j of type P in protocell i (we use κ instead of k to distinguish κ^{P}_{ij} from k^{P}_{pt}). Price's equation[14,15] states that

$$\langle \lambda_{\tilde{i}\tilde{j}} \rangle \Delta \langle \kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}} \rangle = \sigma^2_{\tilde{i}} \left[\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle \right] + \mathbb{E}_{\tilde{i}} \left[\sigma^2_{i\tilde{j}} \left[\lambda_{ij}, \kappa^{\mathrm{P}}_{ij} \right] \right] \tag{4}$$

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). Below, we show that equations (1) approximate equation (4) up to the second moments of κ^{P}, viz., $\sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle]$ and $\mathbb{E}_{\tilde{i}}[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]]$.

To approximate the first term on the right-hand side of equation (4), we assume that $\langle \lambda_{i\tilde{j}} \rangle$ is a function of $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$ and $\langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle$ that can be expanded as a Taylor series

around $\langle \kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}} \rangle$ and $\langle \kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}} \rangle$. Substituting this series into $\sigma^2_{\tilde{i}}[\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle]$, we obtain

$$\sigma^2_{\tilde{i}}[\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle] = \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \frac{\partial \langle \lambda_{i\tilde{j}} \rangle}{\partial \langle \kappa^{c}_{i\tilde{j}} \rangle} \sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{c}_{i\tilde{j}} \rangle] + O(\sigma^3_{\tilde{i}}), \tag{5}$$

where $O(\sigma^3_{\tilde{i}})$ consists of terms involving the third or higher (mixed) central moments of $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$ and $\langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle$ over protocells[16].

To approximate the second term on the right-hand side of equation (4), we likewise assume that λ_{ij} is a function of κ^{P}_{ij} and κ^{Q}_{ij} that can be expanded as a Taylor series around $\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle$ and $\langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle$. Substituting this series into $\sigma^2_{i\tilde{j}}[\lambda_{ij}, \kappa^{\mathrm{P}}_{ij}]$, we obtain

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

where $O(\sigma^3_{i\tilde{j}})$ consists of terms involving the third or higher (mixed) central moments of κ^{P}_{ij} and κ^{Q}_{ij} over the replicators in protocell i[16]. 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 $\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{c}_{ij}]$, we obtain

$$\mathbb{E}_{\tilde{i}}\big[\sigma^2_{i\tilde{j}}[\lambda_{ij}, \kappa^{\mathrm{P}}_{ij}]\big] = \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \mathbb{E}_{\tilde{i}}\left[\frac{\partial \lambda_{ij}}{\partial \kappa^{c}_{ij}}\right] \mathbb{E}_{\tilde{i}}\big[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{c}_{ij}]\big] + \mathbb{E}_{\tilde{i}}\big[O(\sigma^3_{i\tilde{j}})\big]. \tag{6}$$

Substituting equations (5) and (6) into equation (4), we obtain

$$\Delta \langle \kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}} \rangle = \frac{1}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle} \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \left(\frac{\partial \langle \lambda_{i\tilde{j}} \rangle}{\partial \langle \kappa^{c}_{i\tilde{j}} \rangle} \sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{c}_{i\tilde{j}} \rangle] + \mathbb{E}_{\tilde{i}}\left[\frac{\partial \lambda_{ij}}{\partial \kappa^{c}_{ij}}\right] \mathbb{E}_{\tilde{i}}\big[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{c}_{ij}]\big] \right) + O', \tag{7}$$

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

Next, we assume that covariances between κ^{P} and κ^{Q}, namely, $\sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle]$ and $\mathbb{E}_{\tilde{i}}\big[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{Q}}_{ij}]\big]$, 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 the next section). Under this assumption, equation (7) is transformed into

$$\Delta \langle \kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}} \rangle = \frac{1}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle} \left(\frac{\partial \langle \lambda_{i\tilde{j}} \rangle}{\partial \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle} \sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle] + \mathbb{E}_{\tilde{i}}\left[\frac{\partial \lambda_{ij}}{\partial \kappa^{\mathrm{P}}_{ij}}\right] \mathbb{E}_{\tilde{i}}\big[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]\big] \right) + O'. \tag{8}$$

Using equation (3) (i.e., $\lambda_{ij} = \omega^{\mathrm{P}}_{ij} + \omega^{\mathrm{Q}}_{ij}$), we can transform equation (8) into

$$\Delta \langle \kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}} \rangle = \frac{1}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle} \sum_{t \in \{\mathrm{P},\mathrm{Q}\}} \left(\frac{\partial \langle \omega^{t}_{i\tilde{j}} \rangle}{\partial \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle} \sigma^2_{\tilde{i}}[\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle] + \mathbb{E}_{\tilde{i}}\left[\frac{\partial \omega^{t}_{ij}}{\partial \kappa^{\mathrm{P}}_{ij}}\right] \mathbb{E}_{\tilde{i}}\big[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{P}}_{ij}]\big] \right) + O'. \tag{9}$$

Moreover, it can be shown that

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

Using the above equation, we can transform equation (9) into

$$\Delta\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle=\sum_{t\in\{\mathrm{P},\mathrm{Q}\}}\frac{\langle\omega_{i\tilde{j}}^{t}\rangle}{\langle\lambda_{i\tilde{j}}\rangle}\left(\frac{\partial\ln\langle\omega_{i\tilde{j}}^{t}\rangle}{\partial\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle}\sigma_{\tilde{i}}^{2}[\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle,\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle]+\mathbb{E}_{\tilde{i}}\left[\frac{\partial\ln\omega_{ij}^{t}}{\partial\kappa_{ij}^{\mathrm{P}}}\right]\mathbb{E}_{\tilde{i}}\big[\sigma_{i\tilde{j}}^{2}[\kappa_{ij}^{\mathrm{P}},\kappa_{ij}^{\mathrm{P}}]\big]\right)+O'',\tag{10}$$

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

We adopt the following notation:

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

$$\bar{k}^{\mathrm{P}}=\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle,\qquad \gamma_{\mathrm{P}}^{\mathrm{P}}=-\mathbb{E}_{\tilde{i}}\left[\frac{\partial\ln\omega_{ij}^{\mathrm{P}}}{\partial\kappa_{ij}^{\mathrm{P}}}\right],\qquad \beta_{\mathrm{P}}^{t}=\frac{\partial\ln\langle\omega_{i\tilde{j}}^{t}\rangle}{\partial\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle},$$

where $\bar{\omega}^{t}$ is the normalised average reproductive value of type-t replicators, $\sigma_{\mathrm{cel}}^{2}$, $\sigma_{\mathrm{mol}}^{2}$, 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_{P}^{t} 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 $\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle$ and κ_{ij}^{P} can be regarded as mathematically independent of each other, provided i and j are fixed (if i and j are varied, $\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle$ and κ_{ij}^{P} may be statistically dependent). Under this assumption, increasing κ_{ij}^{P} does not increase $\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle$, so that $\gamma_{\mathrm{P}}^{\mathrm{P}}$ reflects only the cost of providing catalysis at the molecular level. Likewise, increasing $\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle$ does not increase κ_{ij}^{P}, so that β_{P}^{t} 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 (10) into

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

where O'' is omitted.

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 $\sigma^2_{\tilde{i}} [\langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{Q}}_{\tilde{i}j} \rangle] = \sigma^2_{\tilde{i}} [\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle, \langle \kappa^{\mathrm{P}}_{\tilde{i}j} \rangle]$ 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.

The phase-plane analysis.

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^t_{ij}). The following definition was employed:

$$\omega^t_{ij} = e^{\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle + \langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle} \frac{e^{-s\kappa^t_{ij}}}{\left\langle e^{-s\kappa^{\mathrm{P}}_{i\tilde{j}}} \right\rangle + \left\langle e^{-s\kappa^{\mathrm{Q}}_{i\tilde{j}}} \right\rangle}. \tag{11}$$

where the factor $e^{\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle + \langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle}$ represents the cellular-level benefit of catalysis provided by the replicators in protocell i, the numerator $e^{-s\kappa^t_{ij}}$ represents the molecular-level cost of catalysis provided by the focal replicator, the denominator $1/(\langle e^{-s\kappa^{\mathrm{P}}_{i\tilde{j}}} \rangle + \langle e^{-s\kappa^{\mathrm{Q}}_{i\tilde{j}}} \rangle)$ normalises the cost, and s is the cost-benefit ratio. The above definition of ω^t_{ij} 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^t_c are positive (e.g., 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 $\partial \omega^t_{ij} / \partial \kappa^t_{ij} < 0$ and $\partial \langle \omega^t_{i\tilde{j}} \rangle / \partial \langle \kappa^c_{i\tilde{j}} \rangle > 0$ for $c = t$ and $c \neq t$. Apart from this requirement, the definition was arbitrarily chosen for simplicity.

Under the definition of ω^t_{ij} in equation (11), we obtain equations describing the evolution of $\langle \kappa^c_{\tilde{i}\tilde{j}} \rangle$ (denoted as $\bar{k}^c$ in the main text) as follows. Since the evolution of $\langle \kappa^c_{\tilde{i}\tilde{j}} \rangle$ is described by equation (7), we substitute equation (11) into equation (7). For this substitution, we need to calculate the derivatives of fitness. According to

equation (3), the fitness of a replicator is $\lambda_{ij} = \omega^{\mathrm{P}}_{ij} + \omega^{\mathrm{Q}}_{ij}$. Therefore,

$$\mathbb{E}_{\tilde{i}}\left[\frac{\partial \lambda_{ij}}{\partial \kappa^{c}_{ij}}\bigg|_{\substack{\kappa^{\mathrm{P}}_{ij}=\langle\kappa^{\mathrm{P}}_{i\tilde{j}}\rangle \\ \kappa^{\mathrm{Q}}_{ij}=\langle\kappa^{\mathrm{Q}}_{i\tilde{j}}\rangle}}\right] = \mathbb{E}_{\tilde{i}}\left[-ce^{\langle\kappa^{\mathrm{P}}_{ij}\rangle+\langle\kappa^{\mathrm{Q}}_{ij}\rangle}\frac{e^{-s\langle\kappa^{c}_{i\tilde{j}}\rangle}}{\langle e^{-s\kappa^{\mathrm{P}}_{i\tilde{j}}}\rangle + \langle e^{-s\kappa^{\mathrm{Q}}_{i\tilde{j}}}\rangle}\right]$$

$$= -ce^{\langle\kappa^{\mathrm{P}}_{\tilde{i}j}\rangle+\langle\kappa^{\mathrm{Q}}_{\tilde{i}j}\rangle}\frac{e^{-s\langle\kappa^{c}_{\tilde{i}j}\rangle}}{e^{-s\langle\kappa^{\mathrm{P}}_{\tilde{i}j}\rangle} + e^{-s\langle\kappa^{\mathrm{Q}}_{\tilde{i}j}\rangle}} + \mathbb{E}_{\tilde{i}}\left[O(\sigma^2_{i\tilde{j}})\right] + O(\sigma^2_{\tilde{i}}).$$

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

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

We substitute these derivatives into equation (7) and use the fact that

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

to obtain

$$\Delta\langle\kappa^{c}_{\tilde{i}\tilde{j}}\rangle = (1+\rho_{\mathrm{cel}})\sigma^2_{\mathrm{cel}} - s\frac{e^{-s\langle\kappa^{c}_{\tilde{i}\tilde{j}}\rangle} + \rho_{\mathrm{mol}}e^{-s\langle\kappa^{c'}_{\tilde{i}\tilde{j}}\rangle}}{e^{-s\langle\kappa^{\mathrm{P}}_{\tilde{i}\tilde{j}}\rangle} + e^{-s\langle\kappa^{\mathrm{Q}}_{\tilde{i}\tilde{j}}\rangle}}\sigma^2_{\mathrm{mol}} + O'', \tag{12}$$

where $c' \neq c$, ρ_{cel} is the correlation coefficient between $\langle\kappa^{\mathrm{P}}_{i\tilde{j}}\rangle$ and $\langle\kappa^{\mathrm{Q}}_{i\tilde{j}}\rangle$ (i.e., $\rho_{\mathrm{cel}} = \sigma^2_{\tilde{i}}[\langle\kappa^{\mathrm{P}}_{i\tilde{j}}\rangle, \langle\kappa^{\mathrm{Q}}_{i\tilde{j}}\rangle]/\sigma^2_{\mathrm{cel}}$), and ρ_{mol} is the average correlation coefficient between κ^{P}_{ij} and κ^{Q}_{ij} (i.e., $\rho_{\mathrm{mol}} = \mathbb{E}_{\tilde{i}}[\sigma^2_{i\tilde{j}}[\kappa^{\mathrm{P}}_{ij}, \kappa^{\mathrm{Q}}_{ij}]]/\sigma^2_{\mathrm{mol}}$). To derive equation (12), we have assumed that the variances of $\langle\kappa^{c}_{i\tilde{j}}\rangle$ and κ^{c}_{ij} are independent of c; i.e., $\sigma^2_{\mathrm{cel}} = \sigma^2_{\tilde{i}}[\langle\kappa^{c}_{i\tilde{j}}\rangle, \langle\kappa^{c}_{i\tilde{j}}\rangle]$ and $\sigma^2_{\mathrm{mol}} = \mathbb{E}_{\tilde{i}}[\sigma^2_{i\tilde{j}}[\kappa^{c}_{ij}, \kappa^{c}_{ij}]]$ for $c = \mathrm{P}$ and $c = \mathrm{Q}$.

Equation (12) can be expressed in a compact form as follows:

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

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 $^{\mathrm{T}}$ 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 $\langle\kappa^{c}_{i\tilde{j}}\rangle$ on κ^{c}_{ij}[17] and, therefore, the coefficient of genetic relatedness[18,19]. The potential function $RB - (1-R)C$ can then be interpreted as inclusive fitness.

Next, we omit O'' from equations (12), replace Δ with time derivative $d/d\tau$, and let $\langle \kappa^c_{\tilde{i}\tilde{j}} \rangle$ be denoted by $\bar{k}^c$, obtaining

$$\frac{d}{d\tau}\bar{k}^c = (1+\rho_{\rm cel})\sigma^2_{\rm cel} - s\frac{e^{-s\bar{k}^c} + \rho_{\rm mol}e^{-s\bar{k}^{c'}}}{e^{-s\bar{k}^{\rm P}} + e^{-s\bar{k}^{\rm Q}}}\sigma^2_{\rm mol}. \tag{13}$$

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 (13) 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 (13), we obtain

$$\frac{d}{d\tau}\bar{k}^c = \Theta(\bar{k}^c)\left[(1+\rho_{\rm cel})\sigma^2_{\rm cel} - s\frac{e^{-s\bar{k}^c} + \rho_{\rm mol}e^{-s\bar{k}^{c'}}}{e^{-s\bar{k}^{\rm P}} + e^{-s\bar{k}^{\rm Q}}}\sigma^2_{\rm mol}\right].$$

The above equation was numerically integrated to obtain the phase-plane portrait depicted in Fig. 3.

Equation (13) allows for statistical correlations between $\kappa^{\rm P}_{ij}$ and $\kappa^{\rm Q}_{ij}$ at the molecular and cellular levels, i.e., $\rho_{\rm mol}$ and $\rho_{\rm cel}$. 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). For this sake, we calculate the nullcline of $\frac{d}{d\tau}\bar{k}^c$. From equation (13), we obtain

$$\bar{k}^{c'} = \bar{k}^c + s^{-1}\ln\frac{\rho_{\rm mol}s\sigma^2_{\rm mol} - (1+\rho_{\rm cel})\sigma^2_{\rm cel}}{(1+\rho_{\rm cel})\sigma^2_{\rm cel} - s\sigma^2_{\rm mol}}.$$

This equation shows that all parameters only appear in the intercept of the nullcline with the $\bar{k}^{c'}$-axis. Let us denote this intercept as $s^{-1}\ln I$. The way I qualitatively depends on $\sigma^2_{\rm cel}$ and $s\sigma^2_{\rm mol}$ is independent of $\rho_{\rm cel}$ because $-1 < \rho_{\rm cel} < 1$. Therefore, we can assume that $\rho_{\rm cel} = 0$ without loss of generality. Next, to see how $\rho_{\rm mol}$ influences I, we focus on the singularity of I by setting $(1+\rho_{\rm cel})\sigma^2_{\rm cel} = s\sigma^2_{\rm mol} + \epsilon$, where $\epsilon > 0$. Then, $I = (1-\rho_{\rm mol})s\sigma^2_{\rm mol}/\epsilon - \rho_{\rm mol}$. The way I qualitatively depends on $s\sigma^2_{\rm mol}/\epsilon$ is independent of $\rho_{\rm mol}$ because $-1 < \rho_{\rm mol} < 1$. Therefore, we can assume that $\rho_{\rm mol} = 0$ without loss of generality. Taken together, these calculations show that ignoring correlations between $\kappa^{\rm P}_{ij}$ and $\kappa^{\rm Q}_{ij}$ does not qualitatively affect the results, supporting the validity of equations (1).

References

1. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–563 (1970).

454 2. *The RNA World* 3rd (eds Gesteland, R. F., Cech, T. & Atkins, J. F.) (Cold
455 Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2006).

456 3. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. The case for an
457 ancestral genetic system involving simple analogues of the nucleotides. *Proceed-*
458 *ings of the National Academy of Sciences of the United States of America* **84,**
459 4398–4402 (1987).

460 4. Buss, L. W. *The Evolution of Individuality* (Princeton University Press, Prin-
461 ceton, NJ, 1987).

462 5. Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (W.
463 H. Freeman/Spektrum, Oxford, 1995).

464 6. Michod, R. E. *Darwinian Dynamics: Evolutionary Transitions in Fitness and*
465 *Individuality* (Princeton University Press, Princeton, NJ, 1999).

466 7. Bourke, A. F. G. *Principles of Social Evolution* (Oxford University Press, Ox-
467 ford, UK, 2011).

468 8. Takeuchi, N., Hogeweg, P. & Kaneko, K. The origin of a primordial genome
469 through spontaneous symmetry breaking. *Nature Communications* **8,** 250 (2017).

470 9. Chen, I. A., Roberts, R. W. & Szostak, J. W. The emergence of competition
471 between model protocells. *Science* **305,** 1474–1476 (2004).

472 10. Durand, P. M. & Michod, R. E. Genomics in the light of evolutionary trans-
473 itions. *Evolution* **64,** 1533–1540 (2010).

474 11. Ivica, N. A. *et al.* The paradox of dual roles in the RNA world: resolving the
475 conflict between stable folding and templating ability. *Journal of Molecular*
476 *Evolution* **77,** 55–63 (2013).

477 12. Takeuchi, N. & Hogeweg, P. The role of complex formation and deleterious
478 mutations for the stability of RNA-like replicator systems. *Journal of Molecular*
479 *Evolution* **65,** 668–686 (2007).

480 13. Price, G. R. Selection and covariance. *Nature* **227,** 520–521 (1970).

481 14. Price, G. R. Extension of covariance selection mathematics. *Annals of Human*
482 *Genetics* **35,** 485–490 (1972).

483 15. Hamilton, W. D. Innate social aptitudes of man: an approach from evolutionary
484 genetics. in *Biosocial Anthoroplogy* (ed Fox, R) 133–153 (Malaby Press, London,
485 1975).

486 16. Iwasa, Y., Pomiankowski, A. & Nee, S. The evolution of costly mate preferences
487 II. the 'handicap' principle. *Evolution* **45,** 1431–1442 (1991).

17. Rice, S. H. *Evolutionary Theory: Mathematical and Conceptual Foundations* (Sinauer Associates, Sunderland, MA, USA, 2004).

18. Taylor, P. D. & Frank, S. A. How to make a kin selection model. *Journal of Theoretical Biology* **180,** 27–37 (1996).

19. Levin, S. & West, S. Kin selection in the RNA world. *Life* **7,** 53 (2017).

20. Takeuchi, N., Kaneko, K. & Hogeweg, P. Evolutionarily stable disequilibrium: Endless dynamics of evolution in a stationary population. *Proceedings of the Royal Society B* **283,** 20153109 (2016).

Supplementary Information

Extended Data Figures and Supplementary Discussion are included in this manuscript.

Acknowledgements

The authors have been supported by JSPS KAKENHI (grant number JP17K17657 and JP17H06386). NT has been supported by the University of Tokyo through the national university reform and strengthening project.

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.

Author Information

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to nobuto.takeuchi@auckland.ac.nz.

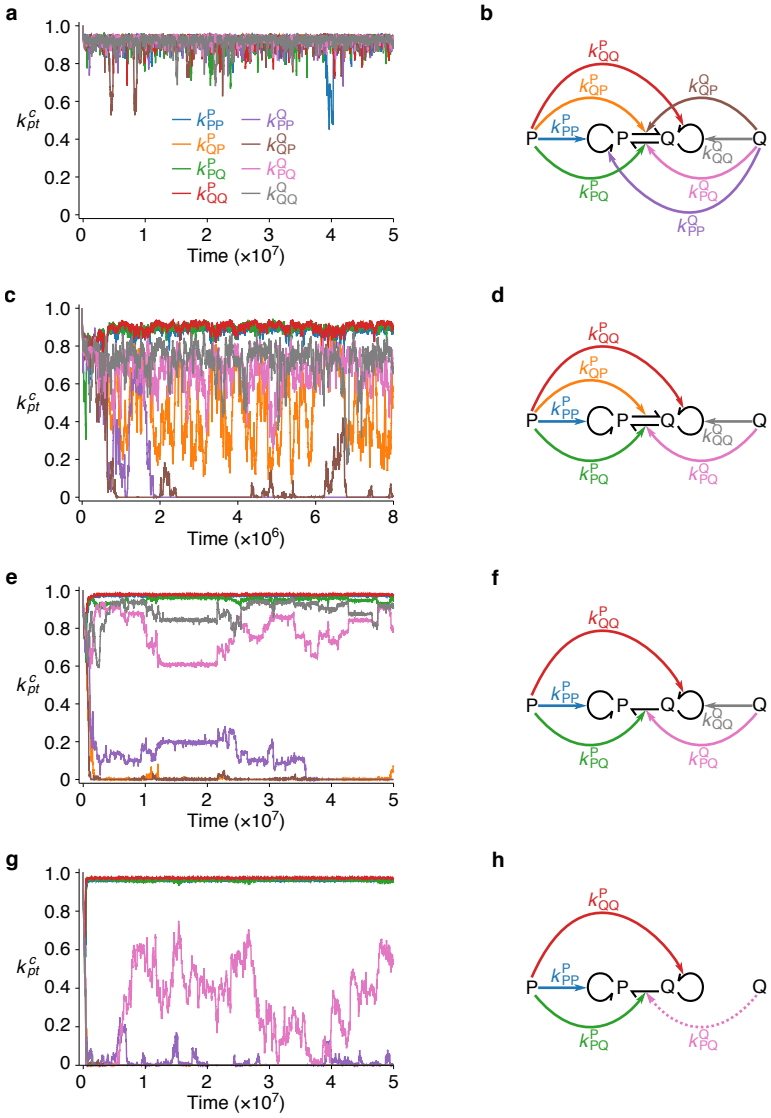
Code availability

The C++ source codes implementing the models are available from the corresponding author upon request.

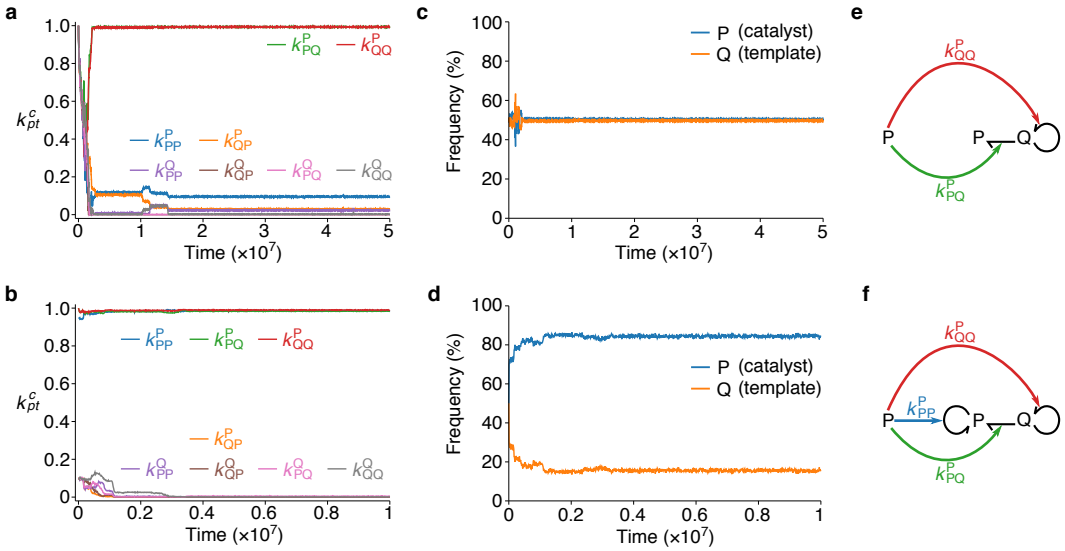
Data availability

The authors declare that the data supporting the findings of this study are available within the paper and Extended Data Figures.

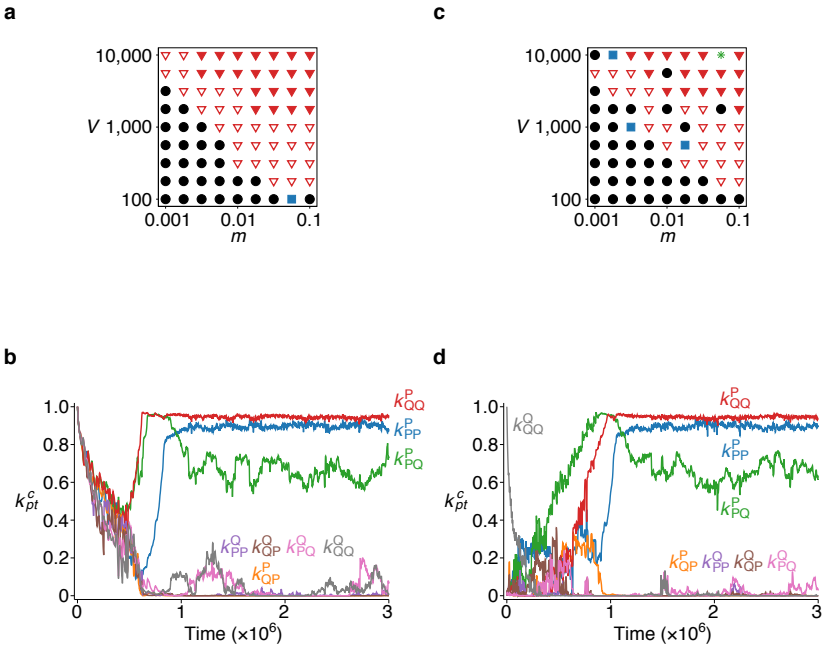
Extended Data



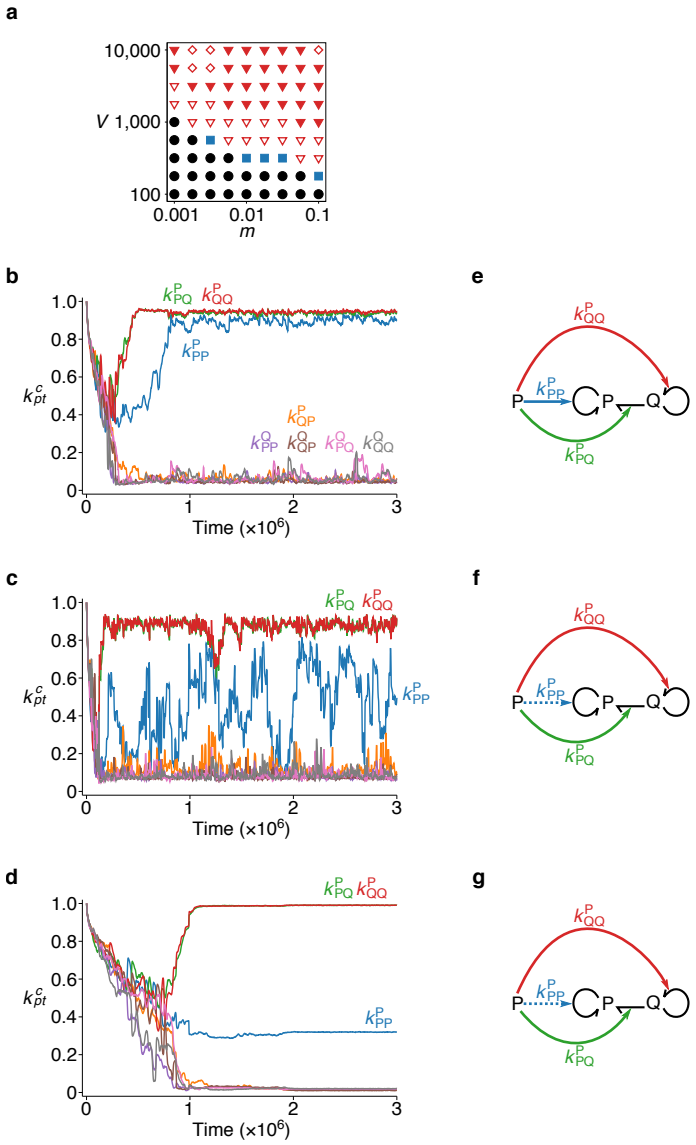
Extended Data Figure 1: **The evolutionary dynamics of the model.** **a**, The dynamics of k_{pt}^{c} averaged over all replicators for parameters corresponding to 'no symmetry breaking' in Fig. 2a: $V = 178$ and $m = 0.01$. **b**, Replicators evolving in a. **c**, **d**, Parameters corresponding to '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 $m = 0.01$.



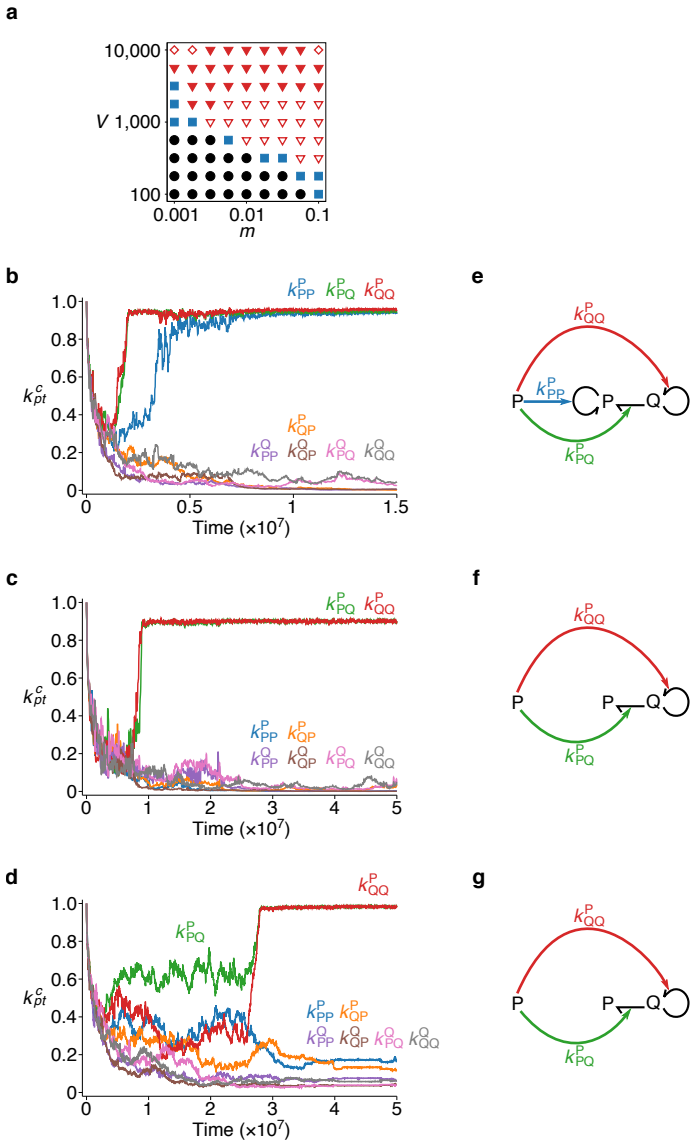
Extended Data Figure 2: **The absence of numerical symmetry breaking for small m and large V. a**, **b**, The dynamics of k_{pt}^{c} averaged over all replicators is shown for $V = 10000$ and $m = 0.001$ with two different initial conditions: a symmetric initial condition, where $k_{pt}^{c} = 1$ (a); an asymmetric initial condition, where $k_{\mathrm{PP}}^{\mathrm{P}} = 0.95$, $k_{\mathrm{PQ}}^{\mathrm{P}} = 0.1$, $k_{\mathrm{QP}}^{\mathrm{P}} = 1$, $k_{\mathrm{QQ}}^{\mathrm{P}} = 1$, and $k_{pt}^{\mathrm{Q}} = 0.1$ (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**, Replicators evolving for the symmetric initial condition (e) and for the asymmetric initial condition (f).



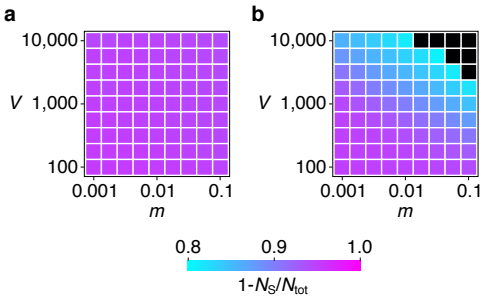
Extended Data Figure 3: **Symmetry breaking with an alternative definition of complex formation rates.** The rate constants of complex formation were defined in such a way that coexistence between P and Q is neither favoured nor disfavoured by cellular-level selection (see Supplementary Discussion 2). **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.



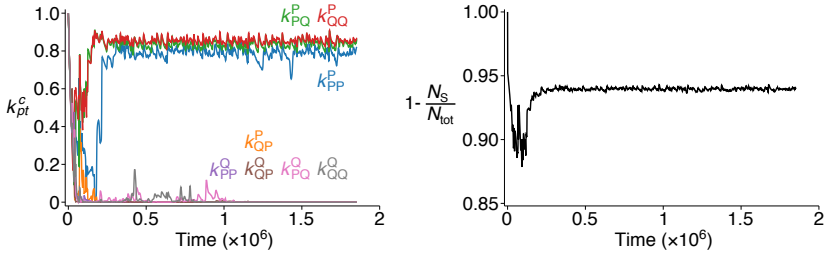
Extended Data Figure 4: **Symmetry breaking with reflecting mutation.** The mutation of k^c_{pt} is modelled as unbiased random walk with reflecting boundaries at 0 and 1 (see Supplementary Discussion 3). **a**, Phase diagram. The symbols are the same as in Fig. 2a ($t_{\min} > 3.9 \times 10^7$ for $m = 0.1$ 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**, Replicators evolving in b, c, d, respectively.



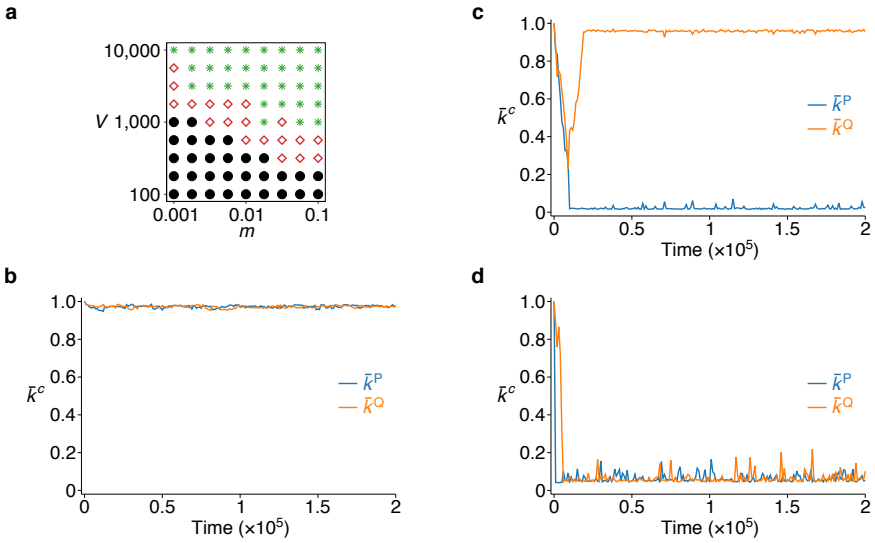
559 Extended Data Figure 5: **Symmetry breaking with log-space mutation.** The
560 mutation of k^c_{pt} is modelled as unbiased random walks on a logarithmic scale (see
561 Supplementary Discussion 3). **a**, Phase diagram. The symbols are the same as in
562 Fig. 2a ($t_{\min} > 3.9 \times 10^7$ only for $m = 0.1$ and $V = 10000$). **b**, Dynamics of k^c_{pt} averaged
563 over all replicators. $m = 0.01$ and $V = 10000$. Three-fold symmetry breaking occurs.
564 **c**, $m = 0.1$ and $V = 10000$. No numerical symmetry breaking occurs. **d**, $m = 0.00178$
565 and $V = 10000$. No numerical symmetry breaking occurs. **e**, **f**, **g**, Replicators evolving
566 in b, c, d, respectively.



Extended Data Figure 6: **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 (Fig. 1). **b**, The model which excludes the possibility of symmetry breaking; specifically, it allows only one type of replicator (either P or Q). Black squares indicate extinction (i.e. $N_{\mathrm{tot}} = N_{\mathrm{S}}$). $t_{\mathrm{min}} > 1.5 \times 10^7$.



Extended Data Figure 7: **Result for large m and V values.** The dynamics of the simulation model is shown for $m = 0.1$ and $V = 10^5$, parameters outside the range examined in Fig. 2a and Extended Data Fig. 6. **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. $t_{min} > 1.8 \times 10^6$.



Extended Data Figure 8: **Symmetry breaking in a hierarchical Wright-Fisher model.** 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 Supplementary Discussion 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 **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).

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

Extended Data Table 1: **Differentiation between reproductive and non-reproductive elements is a universal property of life.** *Bacterial symbionts of ungulate lice (*Haematopinus*) and planthoppers (*Fulgoroidea*)[21].

Supplementary Discussion

1. On the the chemical identity of P and Q

The present study formulates the central dogma in functional (as opposed to chemical) terms as the one-way flow of information from non-catalytic molecules to catalytic molecules. This formulation is advantageous for simplicity and generally as mentioned in the main text. In particular, it makes our theory independent of the chemical details of replicating molecules. For example, our theory assumes that a molecule faces a trade-off between catalysing and templating, but it does not restrict catalysis to being replicase activity (although our simulation model explicitly assumes that catalysts are replicases, our mathematical theory based on equation (1) does not make this assumption). Therefore, our theory offers a great degree of freedom for experimental testing. One possibility for such experiments might be to use RNA and DNA to embody P and Q of our theory, given the availability of various catalytic RNA and DNA molecules[22–24]. 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 proteins[25–28].

2. Model in which coexistence between P and Q is selectively neutral

In the simulation model described in the main text, the reaction rate constants of complex formation are defined as the k^c_{pt} values of a replicator serving as a catalyst. Under this definition, coexistence between P and Q is favoured by cellular-level selection because replicators multiply fastest if their k^c_{pt} values are maximised for all combinations of c, p, and t, as described in Methods. To ascertain that this specific aspect of the model does not critically affect results, we additionally examined an alternative model in which cellular-level selection neither favours nor disfavours coexistence between P and Q.

In this 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 $\max(k^x_{\mathrm{P}y}, k^x_{\mathrm{Q}y}) + \max(k^y_{\mathrm{P}x}, k^y_{\mathrm{Q}x}) \leq 2k_{\max}$ if all possible complexes are considered, where x and y are the replicator types of X and Y, respectively (note

that in the original simulation model, this rate is proportional to $\sum_p k^x_{py} + k^y_{px} \leq 4k_{\max}$). 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. Therefore, coexistence between P and Q is not necessarily favoured by cellular-level selection.

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. In the asymmetric initial condition, only Q was present. For both initial conditions, the model displays the same symmetry breaking as displayed by the original model (Extended Data Fig. 3).

3. Alternative models for the mutation of k^c_{pt}

In the simulation 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 model described in the main text, except that the boundary condition at $k^c_{pt} = 0$ is set to reflecting. 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 (Extended Data Figs. 4 and 5)

4. Numerical symmetry breaking

In this section, we show that numerical symmetry breaking occurs because it is favoured by cellular-level selection in the presence of catalytic and informatic asymmetry and neither favoured nor disfavoured by molecular-level selection. To this end, we will use a similar mathematical framework as used to derive equations (1) (see Methods).

The proximate—as opposed to ultimate—cause of numerical symmetry breaking is the self-replication of catalysts (i.e., $k^c_{cc} > 0$). 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 (Extended Data Fig. 2). Taken together, these results indicate that the proximate cause of numerical symmetry breaking is the self-replication of catalysts. Therefore, to understand why numerical symmetry breaking occurs, we need to understand why the self-replication of catalysts evolves.

To address this 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:

$$\begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau+1) \\ n_{ij}^{\mathrm{Q}}(\tau+1) \end{bmatrix} = \begin{bmatrix} w_{ij}^{\mathrm{PP}} & \omega_{ij}^{\mathrm{Q}} \\ 0 & \omega_{ij}^{\mathrm{Q}} \end{bmatrix} \begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau) \\ n_{ij}^{\mathrm{Q}}(\tau) \end{bmatrix}, \tag{14}$$

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 (14):

$$\lambda_{ij} = \begin{cases} \omega_{ij}^{\mathrm{Q}} & \text{if } \omega_{ij}^{\mathrm{Q}} > w_{ij}^{\mathrm{PP}} \\ w_{ij}^{\mathrm{PP}} & \text{otherwise.} \end{cases} \tag{15}$$

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

$$\boldsymbol{v}_{ij} = \begin{cases} \frac{1}{2-w_{ij}^{\mathrm{PP}}/\omega_{ij}^{\mathrm{Q}}} \begin{bmatrix} 1 \\ 1-w_{ij}^{\mathrm{PP}}/\omega_{ij}^{\mathrm{Q}} \end{bmatrix} & \text{if } \omega_{ij}^{\mathrm{Q}} > w_{ij}^{\mathrm{PP}} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{otherwise.} \end{cases} \tag{16}$$

Equation (16) shows that we must assume $\omega_{ij}^{\mathrm{Q}} > w_{ij}^{\mathrm{PP}}$ in order for P and Q to coexist. Equation (16) also shows that the frequency of catalysts (i.e., $(2 - w_{ij}^{\mathrm{PP}}/\omega_{ij}^{\mathrm{Q}})^{-1}$) 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 (15), λ_{ij} is not directly dependent 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., $\mathbb{E}_{i\tilde{j}}[(2-w_{i\tilde{j}}^{\mathrm{PP}}/\omega_{i\tilde{j}}^{\mathrm{Q}})^{-1}]$). This frequency can increase 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., $\langle\lambda_{i\tilde{j}}\rangle$). Thus, the fitness of a protocell increases with the frequency of catalysts in that protocell (i.e., $\mathbb{E}_{i\tilde{j}}[(2-w_{i\tilde{j}}^{\mathrm{PP}}/\omega_{i\tilde{j}}^{\mathrm{Q}})^{-1}]$), which in turn increases 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 genome-enzyme differentiation.

Finally, we add two general remarks about numerical symmetry breaking. First, numerical symmetry breaking is always observed in the systems displaying the division of labour between the transmission of genetic information and the other functions (Extended Data Table 1); e.g., the number of germ-line cells is smaller than that of somatic cells per organism, and the number of queens is smaller than that of workers per colony[4–7]. Numerical symmetry breaking can therefore be considered as an integral aspect of the reproductive division of labour although it is not considered as such in the central dogma.

Second, the important consequence of numerical symmetry breaking is that it 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., $\sigma_{\mathrm{cel}}^2/\sigma_{\mathrm{mol}}^2$); therefore, it has a stabilising effect on protocells[8,29]. 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[5–7].

5. The hierarchical Wright-Fisher model

Although the simplifications involved in the derivation of equations (1) allow us to elucidate the mechanism of symmetry breaking, they also make the comparison between equations (1) and the simulation model indirect. Specifically, equations (1) cannot be compared with the simulation model in terms of the same parameters, because the former treat σ^2_{mol} and σ^2_{cel} as parameters, which are actually variables dependent on m and V in the latter. To fill this gap, we constructed a model that stochastically simulates the population dynamics described by equations (1), but nevertheless treats σ^2_{mol} and σ^2_{cel} as 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 (11). In each generation, replicators are replicated or transcribed with probabilities proportional to ω^c_{ij}, so that the population dynamics matches equation (2) 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. With a probability m per replication or transcription, each κ^c value of a replicator is mutated by adding a number randomly drawn from a uniform distribution on the interval $(-\delta_{\mathrm{mut}}, \delta_{\mathrm{mut}})$ ($\delta_{\mathrm{mut}} = 0.1$). The values of κ^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 (Extended Data Fig. 8), a result that is consistent with the outcomes of the original simulation 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).

References

21. Frank, S. A. Host control of symbiont transmission: the separation of symbionts into germ and soma. *The American Naturalist* **148,** 1113–1124 (1996).

22. Silverman, S. K. Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends in Biochemical Sciences* **41,** 595–609 (2016).

23. 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* **5,** 1511 (2016).

24. Samanta, B. & Joyce, G. F. A reverse transcriptase ribozyme. *eLife* **6,** e31153 (2017).

25. Powner, M. W., Zheng, S.-L. & Szostak, J. W. Multicomponent assembly of proposed DNA precursors in water. *Journal of the American Chemical Society* **134,** 13889–13895 (2012).

26. Poole, A. M. *et al.* The case for an early biological origin of DNA. *Journal of Molecular Evolution* **79,** 204–212 (2014).

27. Ritson, D. J. & Sutherland, J. D. Conversion of biosynthetic precursors of RNA to those of DNA by photoredox chemistry. *Journal of Molecular Evolution* **78,** 245–250 (2014).

28. Gavette, J. V., Stoop, M., Hud, N. V. & Krishnamurthy, R. RNA-DNA chimeras in the context of an RNA world transition to an RNA/DNA world. *Angewandte Chemie International Edition,* 13204–13209.

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