ARTICLE

doi:10.1038/nature11549

Spontaneous network formation among cooperative RNA replicators

Nilesh Vaidya[1], Michael L. Manapat[2], Irene A. Chen[3]†, Ramon Xulvi-Brunet[3], Eric J. Hayden[4] & Niles Lehman[1]

The origins of life on Earth required the establishment of self-replicating chemical systems capable of maintaining and evolving biological information. In an RNA world, single self-replicating RNAs would have faced the extreme challenge of possessing a mutation rate low enough both to sustain their own information and to compete successfully against molecular parasites with limited evolvability. Thus theoretical analyses suggest that networks of interacting molecules were more likely to develop and sustain life-like behaviour. Here we show that mixtures of RNA fragments that self-assemble into self-replicating ribozymes spontaneously form cooperative catalytic cycles and networks. We find that a specific three-membered network has highly cooperative growth dynamics. When such cooperative networks are competed directly against selfish autocatalytic cycles, the former grow faster, indicating an intrinsic ability of RNA populations to evolve greater complexity through cooperation. We can observe the evolvability of networks through *in vitro* selection. Our experiments highlight the advantages of cooperative behaviour even at the molecular stages of nascent life.

The 'RNA world' is a plausible stage in the development of life because RNA simultaneously possesses evolvability and catalytic function[1]. An RNA organism that could evolve in such a fashion is likely to have been one of the Earth's first life forms. A search is underway[2,3] for an RNA autoreplicase that relies on its individual genotype to compete for survival and reproduction by Darwinian-type evolution in a fitness landscape. Yet the transition from a prebiotic chemistry to this stage of life is not understood. Several authors have proposed that the most primitive life thrived less on discrete genotypes and instead on collections of molecular types more subject to systems chemistry than to straightforward selection dynamics[4–9]. In particular, it was suggested that webs of functionally linked, genetically related replicators were required in the earliest phases of life's appearance to prevent informational decay (the so-called error catastrophe)[4,10–12].

An empirical demonstration of RNA replicator networks could illuminate critical features of this early stage of life. Ribozymes are good candidates for this because they can evolve outside of an organismal context, construct other RNAs, exhibit self-sustained reproduction, and explore sequence space in efficient ways[13–15]. However, their ability to form catalytic networks capable of expanding as predicted from theory has not yet been shown, despite the observation that collections of nucleic acids have the potential to manifest complexity[6,16]. Simulations show that molecular networks should arise, evolve and provide a population with resistance against parasitic sequences[8]. These results are robust within structured environments such as cells or on grids, but are less so in a solution phase. Recent experimental work *in vitro* has been very successful at demonstrating simple ecologies[17–19], reciprocity between two species[6,16,20], and sustained exponential growth via cross catalysis[15]. Empirical efforts to date have been limited by an inability to expand past reciprocal interactions between two species to prebiotically relevant systems that have the capacity to increase their complexity by expanding to three, and then more, members[18,21]. Specifically, the use of systems in which the recognition domain in the catalyst and the target domain in the substrate are co-located in each replicator has prevented networks of more than two members from forming. If this molecular feature could be circumvented, larger networks could be realized within RNA populations in the test tube and help demonstrate a potential escape from the error catastrophe problem that tends to plague selfish systems.

The *Azoarcus* ribozyme system

The ~200-nucleotide (nt) *Azoarcus* group I intron ribozyme[22] can be broken into fragments that can covalently self-assemble by catalysing recombination reactions in an autocatalytic fashion[23,24] (Supplementary Fig. 1). By allowing variation in the sequence recognition mechanism by which this assembly occurs, which is provided by the 3-nt internal guide sequence (IGS) at the 5′ end of the ribozyme, many such autonomously self-assembling ribozymes become possible. We sought to determine if these ribozymes could display cooperative behaviour if their IGS sequences target the assembly of other ribozymes, but not themselves.

To create a cooperative network, we fragmented the *Azoarcus* ribozyme into two pieces in three different ways with the intent of observing how they could spontaneously reassemble via intermolecular cooperation (Fig. 1a, b). We manipulated the IGS (canonically GUG) and its target triplet to generate both matched and mismatched partners. We mixed various IGS and target pairs in two-piece constructs to test the ability of mismatched pairs to promote self-assembly (Supplementary Fig. 2). From these data, we chose three mismatched pairs that exhibit relatively little autocatalysis: GUG/CGU, GAG/CAU, and GCG/CUU. These crippled pairs are denoted I_1, I_2 and I_3, respectively, meaning that they are informational subsystems, albeit weakly autocatalytic.

We chose the triplet pairs so that when the three subsystems are mixed together, they should constitute a cyclical cooperative network in which the output of one subsystem can catalyse the replication of the next one in the cycle (Fig. 1b). This occurs because the IGS of one subsystem is matched to the target in the next subsystem, and the physical separation of the IGS and its target allows for cycles of more than two members. When the six RNAs (**W**, ***h*•X•Y•Z**, **W•X**, ***h*•Y•Z**, **W•X•Y** and ***h*•Z**; • indicates covalent bonding) are allowed to fold together and be co-incubated in equimolar ratios, we expect the subsystems first to

[1]Department of Chemistry, Portland State University, PO Box 751, Portland, Oregon 97207, USA. [2]School of Engineering and Applied Sciences and Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts 02138, USA. [3]FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Department of Bioengineering, Stanford University, Stanford, California 94305, USA. †Present address: Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA.

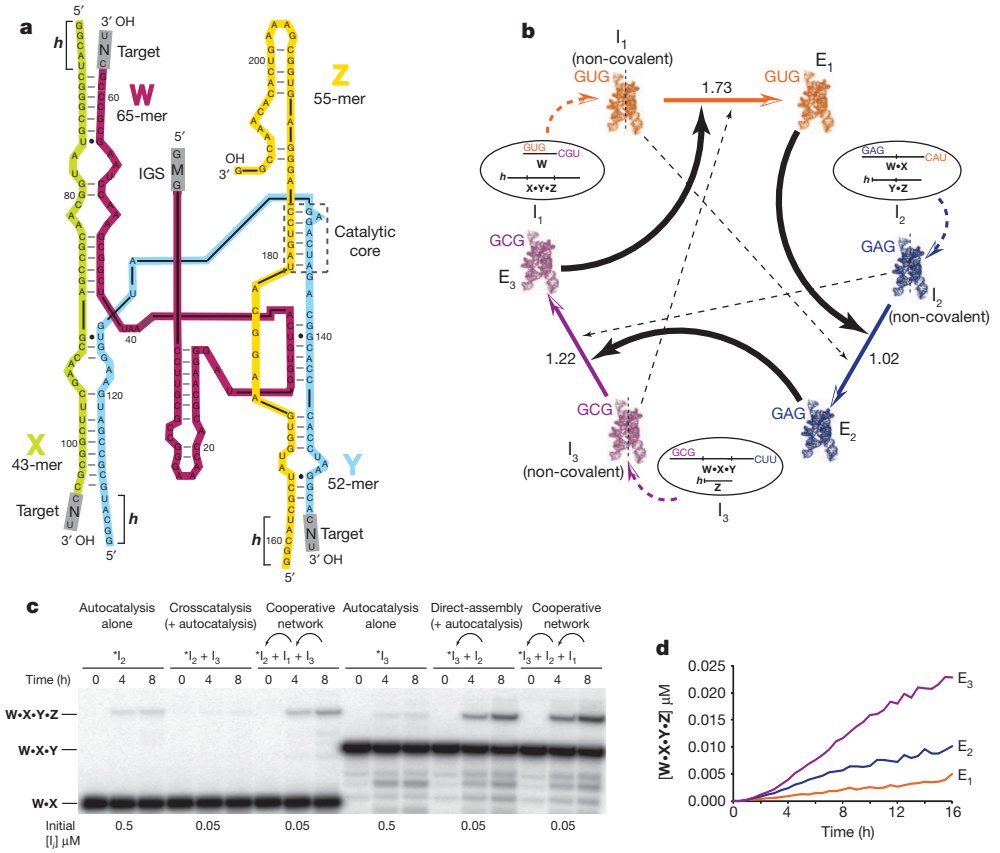

Figure 1 | Cooperative covalent assembly of recombinase ribozymes. **a**, Design of recombinase ribozymes capable of spontaneous cooperative covalent assembly from fragments. The *Azoarcus* ribozyme[25] can be broken at three loop regions to obtain four oligomers capable of self-assembling into a full-length molecule[26,27]. The grey box in **W** (magenta) is the internal guide sequence (IGS), whereas those at the 3′ ends of the **W**, **X** (lime) and **Y** (blue) fragments are recombination targets (tags) recognized by the IGS, which guides the catalysis of a covalent closure (•) of the loops. **b**, A cooperative system comprised of three subsystems, each created from partitioning the molecule into two pieces at different junctions: I_1 (**W** + ***h***•**X**•**Y**•**Z**), I_2 (**W**•**X** + ***h***•**Y**•**Z**) and I_3 (**W**•**X**•**Y** + ***h***•**Z**). Numbers over arrows estimate the cooperative advantage for each step (see text). **c**, Electrophoretic observation of assemblies of E_2 and E_3. The 5′ fragments of I_2 or I_3 were independently 5′-radiolabelled with ^{32}P (that is, $*I_2$ or $*I_3$). The reactions were performed by incubating 0.5 μM (for autocatalysis) or 0.05 μM (for direct assembly, cross catalysis and cooperation) of each fragment for 8 h. Where appropriate, the arrows identify the subsystems being assembled by the previous subsystems in the network, where the IGS and recombination tags match. **d**, Yields of individual E_i ribozymes over time, measured every 30 min for 16 h when all six I_i RNA fragments are co-incubated at 0.05 μM.

form non-covalent versions of ribozymes, and then catalyse the formation of covalent versions of the next ribozyme in the cycle.

To test whether cooperation between enzymes occurred, we took several approaches. First, for the cycle to exhibit positive feedback[4], there should be a distinct advantage to being a covalently contiguous ribozyme (E_i), as opposed to remaining fragmented (I_i). Once covalent ribozymes are formed, they should further promote synthesis of their target ribozymes, at faster rates than the non-covalent versions would. When we tested each in isolation, we found that the E_i ribozymes recombined their respective target substrates into products 1.3–6.3-fold more than the I_i versions when assayed separately (Supplementary Fig. 3). Second, by examining each subsystem in isolation or in pairs, we could compare the relative strengths of autocatalysis (E_i synthesizing E_i), cross-catalysis (E_{i+1} synthesizing E_i), and what should be the most efficient, direct catalysis (E_i synthesizing E_{i+1}). When we incubated just the two RNAs from any one subsystem, such as I_2, alone, there is minimal synthesis of the corresponding ribozyme E_2; after a few hours roughly 0.1% of **W**•**X** is converted into **W**•**X**•**Y**•**Z**. This low background level of autocatalytic synthesis reflects residual catalytic activity available to a mismatched IGS and IGS target, for example GAG with CAU[25], showing that each I_i subsystem has severely limited information-replication potential in isolation. Likewise, when the four RNAs of two subsystems were co-incubated, the cross-catalytic synthesis of the ribozyme corresponding to the preceding subsystem in the cycle is similarly poor, again hindered by an IGS–IGS-target mismatch (Fig. 1c). After only 1 h of incubation, the yield of E_3 from 0.5 μM I_3 is 0.10 ± 0.02% (autocatalysis), and the yield of E_3 from 0.5 μM I_3 and 0.5 μM E_1 is 0.7 ± 0.06% (cross-catalysis), but the yield of E_3 from 0.5 μM I_3 and 0.5 μM I_2 is 13 ± 0.5% (direct catalysis) (data not shown; errors given as s.e.m.). These differences are all statistically significant as measured by *t*-tests several planned comparisons ($P < 0.001$). From these data we determined that direct catalysis is significantly more efficient than catalysis resulting from mismatched IGS sequences and their targets.

When all six RNAs of all three subsystems are co-incubated, cooperation causes the synthesis of **W**•**X**•**Y**•**Z** to rapidly escalate, as expected. The composite yield of full-length RNA after 16 h when I_1, I_2 and I_3 are mixed is 125-fold higher than the sum of the yields of the three subsystems in isolation (Supplementary Fig. 4). This enhancement can be readily visualized after shorter periods of time (Fig. 1c). Each subsystem grows at a different rate (Fig. 1d). The synthesis of E_3 by E_2 is more rapid than that of the other two ribozymes, presumably because the non-covalent version of the enzyme (I_2) is nearly as efficient as the covalent version (E_2); it could also be because certain IGS–IGS target pairs are more efficient[25]. Importantly, we can detect two-step (relayed) cooperativity by comparing the yields with and without the intervening enzyme. In the case of E_1 for example, after 4 h the increase in yield of E_1 upon addition of I_2 to I_1 with I_3 present is 2.5%, whereas the increase in yield of adding I_2 to I_1 without I_3 present is only 0.02%, showing the operation of E_2 through E_3 onto E_1 (Supplementary Table 1); this is supported by doping experiments (Supplementary Fig. 5).

To observe the advantage of cooperation in another way, we constructed a control system in which the I_i molecules could act as catalysts, but could not be covalently assembled themselves because their target sequences were not a match for any catalyst in the system (Supplementary Fig. 5). Cooperation would be manifest when enzymes synthesize other enzymes, and there is some benefit to being covalent. Thus we measured the yields of **W•X•Y•Z** molecules at 8 h in this control system and in our normal system (that is, Fig. 1b). The yields in the control system were consistently worse, and we calculated the ratio (E_i catalysis + I_i catalysis) to (I_i catalysis only) as the advantage of being covalent in each leg of the cycle. These ratios, indicated above the coloured arrows in Fig. 1b, are 1.73, 1.02 and 1.22 for $i = 1, 2$ and 3, respectively. Assuming these values are multiplicative, the cooperative benefit is about 2.2 for the entire cycle.

An impediment to truly hyperbolic growth for such a system[4] is the occasional formation of non-productive complexes (for example, **W**–**Y•Z**) through partially complementary base pairing (Fig. 1a). We can detect such complexes (Supplementary Fig. 6), but when they are minimized by pre-folding each RNA separately, the yield after 2 h increases by 25–50% (Supplementary Fig. 7). As shown by heat–cool regimes, reverse reactions that have the net effect of breaking down covalent ribozymes into fragments may also have a small role in preventing hyperbolic growth (Supplementary Fig. 8).

Cooperation versus selfishness

Next we tested whether a three-membered cooperative system has the potential to have higher fitness than purely autocatalytic systems when placed in direct competition (Fig. 2). To construct 'selfish' autocatalytic subsystems (S_i), we reverted the IGS–IGS target pairs within each subsystem so that they would match. To create S_1 we used $_{GUG}$**W**$_{CAU}$ and ***h*•X•Y•Z**, to create S_2 we used $_{GAG}$**W•X**$_{CUU}$ and ***h*•Y•Z**, and to create S_3 we used $_{GCG}$**W•X•Y**$_{CGU}$ and ***h*•Z**. Each of these subsystems replicates well in isolation. Upon mixing of RNAs, we tracked selfish and cooperative ribozymes by the composition (matched or mismatched, respectively) of the **W**-containing fragments because these contain the IGS and hence the most crucial genetic element (Fig. 2a). When we compared the total yield of $S_1 + S_2 + S_3$ to that of $I_1 + I_2 + I_3$, the former out-performed the latter at all time points (that is, selfishness wins in isolation). One reason for this result is that there would be less time delay in initiating covalent synthesis in the all-selfish system. However, when we placed all six subsystems (12 RNAs: $I_1 + I_2 + I_3 + S_1 + S_2 + S_3$) in the same reaction, the relative yields at later times are reversed, and the growth of the enzymes resulting from the cooperative network now exceeds those from the selfish subsystems (that is, cooperation wins in competition). These results are independent of the exact RNA fragments we chose, as the same result can be seen in other systems with different IGS and IGS targets (see Supplementary Fig. 9). The yield reversal upon mixing happens because the selfish enzymes now participate in—and effectively expand—the cooperative network (Supplementary Fig. 10). This would be a mechanism for a network connectivity increase when the subsystems involved are competing for at least one shared resource, in this case the catalytic core (**Y**–**Z**), because all **W**-containing fragments can use the same 3′ fragments. Whereas selfish enzymes can also benefit from the network, the asymmetry in the proficiencies of the various IGS–IGS-target pairings creates potential for an asymmetry in the relative benefits of the various enzymes in the mixed environment. This feature would have been common in primordial genetic systems, allowing us to posit that cooperation could have been predisposed even in homogeneously mixed environments.

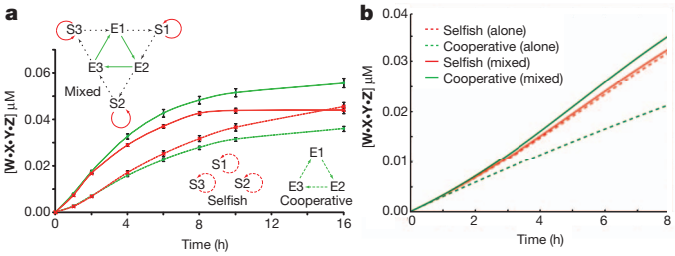

Figure 2 | Cooperative chemistry out-competes selfish chemistry when directly competed. **a**, Empirical results using cooperative (I_1, I_2 and I_3, that is, Fig. 1b) and selfish subsystems (S_1, S_2 and S_3, where IGS and IGS targets were changed to be matching in each subsystem). Yields of total **W•X•Y•Z** RNA tracked the concentrations of cooperative (mismatched) or selfish (matched) **W**-containing RNAs (0.05 μM initial concentrations) over time either when the cooperative (green) and selfish (red) sets of subsystems were incubated separately (dashed lines) or together in the same reaction mixture (solid lines; upper left inset). Data points are averages of three independent trials. Error bars show the standard error of the mean (s.e.m.), and the yields of the cooperative trials in the mixed experiment are significantly greater than those of the selfish trials at the 10- and 16-h time points ($P < 0.05$ by *t*-tests using Sidák's correction for multiple *a posteriori* comparisons). **b**, Simulation of growth dynamics using a toy model of the network of cooperation and selfish interactions (see Supplementary Information). Cooperative enzymes fare better in competition than do selfish enzymes, as demonstrated empirically in panel **a**.

Modelling

Empirical systems such as the one described above are subject to the particularities of chemical and methodological idiosyncrasies, so we sought to generalize these results by constructing mathematical models that show that under a certain set of parameters, the laboratory results should indeed be possible. First we constructed an ordinary differential equation (ODE) model for the three-membered network shown in Fig. 1b. We tracked the yield of each of the three E_i ribozymes separately—using three identical replicates from the same initial reaction mixture—by taking aliquots every 30 min for 16 h (Fig. 1d). We used standard optimization techniques to find the rate constants of all the possible reactions in Fig. 1b that produced trajectories in the ODE system closest to the observed data (Supplementary Information). We used these estimated rate constants to construct a second ODE model that would mimic the cooperative growth of the three subsystems. In general, the non-covalent versions of the ribozymes form relatively tight complexes, with K_d values in the low nanomolar range. When we built cooperative behaviour into the model by relying on differential equations of type $dE_j/dt = k_{ij}[I_j][E_i]$, the experimental data were fit very well in all three subsystems (Supplementary Fig. 11). When we removed direct catalysis from the model and inserted only autocatalysis instead, the quality of the fit decayed substantially such that the root mean squared error was 2.4-fold greater (Supplementary Fig. 12), confirming these results. These data support the contention that replication of the subsystems is indeed cooperative.

Next we constructed a toy model comparing the cooperative and selfish behaviours seen in Fig. 2a using the dynamical relationships that can exist among all enzymes (Fig. 2b). The 'selfish' enzymes perform some altruistic catalysis when alternative substrates become available. The empirical data display more striking yield differences than the model, perhaps because the time delays in bringing the results of the selfish catalytic events back to the selfish subsystems are exacerbated by physical processes such as diffusion. Again this result is general, at least within this network topology, and does not depend on the particular IGS–IGS-target pairings chosen. In essence, although the selfish replicators can parasitize the cooperators, the cooperative network benefits more by incorporating the selfish RNAs. Interestingly, the opposite is generally true in evolutionary dynamics: groups of cooperative individuals grow more quickly than groups of selfish individuals, but a group consisting of both types will eventually be dominated by the selfish[26]. One limitation to the experiment shown in Fig. 2a is that there is only a single iteration of selection. The RNAs used to seed the experiment limit its evolutionary potential; Supplementary Fig. 13 depicts joint genotype frequency changes over time. Experiments in a serial transfer format are needed to show the selection of one strategy over the other (see below), but we

can use both our data and modelling to predict that cooperation would have been advantageous in simpler chemical systems that preceded organismal biology.

Randomization experiment

The system described in Fig. 1 is only one of a very large number of possibilities. To test the notion that cooperative networks of RNA could form spontaneously, we randomized the middle nucleotide of both the IGS (M) and its target triplet (N) in fragments of the ribozyme, generating both matched and mismatched partners within a population. We created three pools of randomized fragments containing the IGS on the 5′ end of the ribozyme: $_{GMG}\mathbf{W}_{CNU}$, $_{GMG}\mathbf{W}\bullet\mathbf{X}_{CNU}$ and $_{GMG}\mathbf{W}\bullet\mathbf{X}\bullet\mathbf{Y}_{CNU}$, plus three fragments containing the catalytic core and the 3′ end of the ribozyme: **X•Y•Z**, **Y•Z** and **Z** (Fig. 3a). Fourfold variation in M and in N, combined with threefold variation in the junction (*j*) where recombination occurs (before **X**, **Y** or **Z**) leads to 48 genotypic possibilities (Fig. 3a). These assembled ribozymes can be distinguished by three variables: (1) the middle nucleotide of the IGS (M), (2) the location of the junction (x, y or z) and (3) the middle nucleotide of the target (N). We therefore denote each ribozyme with the three-letter code M*j*N, where *j* = x, y or z. Each of these ribozymes can be covalently assembled by any other ribozyme, itself covalently contiguous or not, provided that M in the catalyst is complementary to N in the substrate.

When we incubated equimolar amounts of these six RNA sets, all 48 possible full-length **W•X•Y•Z** *Azoarcus* ribozymes arose. The relative frequencies of the 48 possible full-length ribozymes recovered at each time point over an 8 h time course (Supplementary Table 2) show that, in accordance with the above and published data[25], recombination at the **Y**–**Z** junction is favoured, but no single genotype ever exceeded 13% of the total. Growth in the randomization experiment showed markedly greater yields (2–12-fold) than in our engineered three-membered system (Fig. 3b), indicating that far more productive interactions among RNA species are occurring in the former.

From approximately three million **W•X•Y•Z** genotypes sampled at each time point, distinct trends portray indirect evidence of a rapid succession from smaller to increasingly larger networks of cooperators (Fig. 3c, d). Genotypes that could easily propagate by selfish autocatalytic replication peak at or before the first time point at 30 min (Fig. 3c, dotted line with crosses). These are S_i genotypes (for example, those in Fig. 2) where M and N are complementary. A prime example is CyG, which could increase in number from the association of $_{GCG}\mathbf{W}\bullet\mathbf{X}_{CGU}$ and **Y•Z** molecules, and this genotype rose in frequency from 4.8% to 7.2% between 30 min and 2 h. Out of the 48 possible product genotypes, twelve (25%) are of this type.

After peaking early, the frequencies of autocatalysts dropped below random expectation and then slowly climbed. Because of extremely large sample sizes, these deviations are highly significant (two-tailed *G*-tests of independence; $P \ll 0.001$). However, this later frequency increase may not be a consequence of autocatalysis per se, but of the incorporation of autocatalysts into higher-ordered networks, akin to the mechanism by which cooperative networks assimilate selfish replicators (Fig. 2 and Supplementary Fig. 10). Analyses of the frequencies of the product genotypes cannot reveal the identities of the catalysts that made them, and thus do not provide direct evidence of replicator cycles. Nevertheless, we examined whether networks of two or more distinct members could be increasing over time. Some pairs of genotypes can cooperate with each other to form two-membered cycles (for example, AxC + GzU), whereas others cannot (for

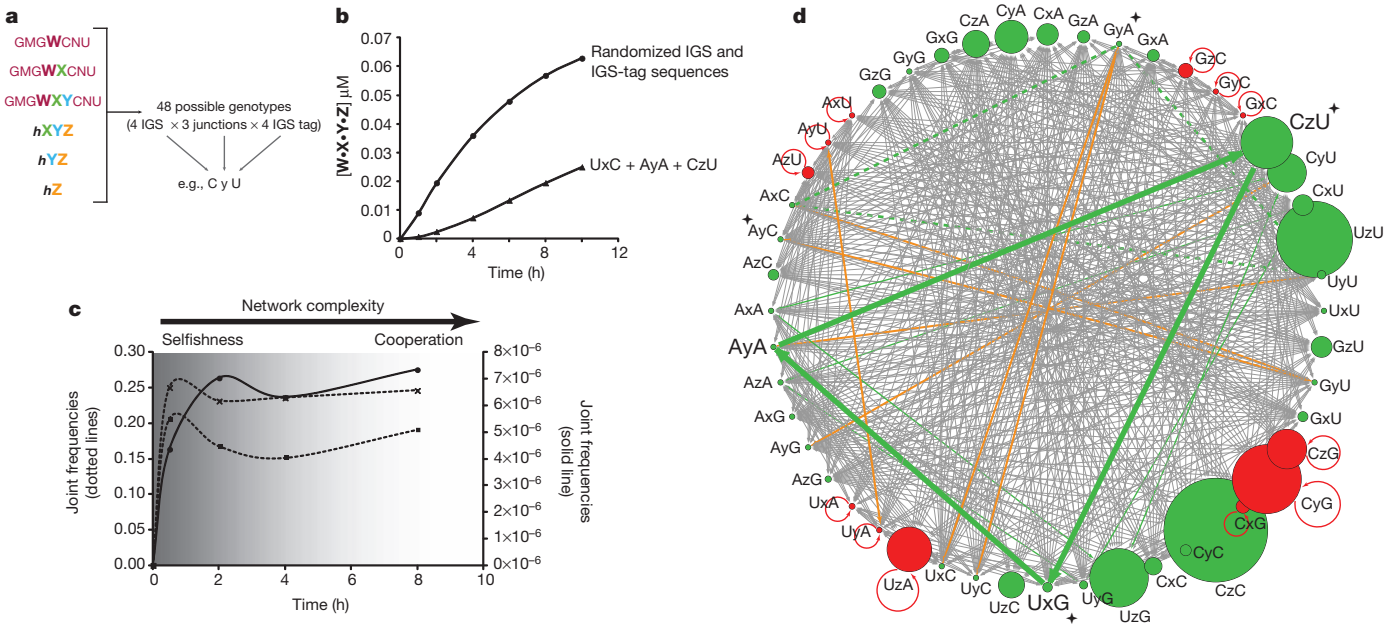

Figure 3 | The randomization experiment. **a**, Experimental design. The middle nucleotides of the IGS and the tags were randomized to create diverse RNA pools. A reaction of 300 pmol each (0.5 μM) of $_{GMG}\mathbf{W}_{CNU}$, $_{GMG}\mathbf{W}\bullet\mathbf{X}_{CNU}$, $_{GMG}\mathbf{W}\bullet\mathbf{X}\bullet\mathbf{Y}_{CNU}$, **X•Y•Z**, **Y•Z** and **Z** was sampled at 0.5, 2, 4 and 8 h, and millions of recombined full-length **W•X•Y•Z** ribozymes were genotyped by nucleotide sequence analysis (Supplementary Table 2). **b**, Comparison of growth curves from fixed and randomized RNAs. Yields over time were compared for the simple three-membered cycle (filled triangles, UxG + AyA + CzU; the sum of the three curves in Fig. 1d) to that in the randomized format (filled circles, panel **a**) when both were performed at the same RNA pool concentrations (0.05 μM). **c**, Proposed succession from simple to complex networks using genotype frequency data from experiment in panel **a**. Simple autocatalytic cycles where M and N are complementary were directly tracked by the sum of such **W•X•Y•Z** molecules (dashed line with crosses; for example, AzU). Reciprocal two-membered cycles were tracked by the sum (×10, for ease of presentation) of the joint frequencies of all genotypes that can potentially participate in such cycles (dashed line with squares; for example, AxA + UxU). The rise of three-membered cycles can be seen from the sum (×10,000 for ease of presentation) of joint frequencies of three sets of genotypes: Fig. 1b and its two permutations by junction (solid line; UxG + AyA + CzU; UyG + AzA + CxU; UzG + AxA + CyU). See Supplementary Information for calculation of the joint frequencies. **d**, The potential network of RNA genotypes. Each node is one of the 48 possible M*j*N genotypes; size scales with relative frequency in the 8 h pool. Nodes are autocatalysts (red) or those that must replicate cooperatively (green). Grey arrows show all possible direct catalytic events; orange arrows show reciprocal two-membered cycles in which the frequencies of both members at least double between 30 min and 2 h; green arrows show key three-membered networks: thick green is the system studied in depth (Fig. 1b), thin green are permutations by junction, dotted green is AxC + GyA + UyU. Starred genotypes can participate in a four-membered network.

example, AxC + UzG). We noticed that the global joint frequencies of the members comprising all possible two-membered cycles peaked at 30 min, declined and recovered, although delayed with respect to the autocatalysts (Fig. 3c). Support for the succession from autocatalysts to these two-membered cycles is found in the frequencies of two possible partners for the autocatalysts GjC, which are CxG and CzG (autocatalysts themselves); the sum of these rose monotonically between 2–8 h (3.7% to 6.1%).

At roughly 2 h, a succession to three-membered cycles may have occurred. Although there are hundreds of such possible assemblages, the joint frequencies of the members of diverse ones requiring synthesis at all three junctions (such as UxG + AyA + CzU) jump at the 2 h mark (Fig. 3c, solid line). Many others peak then as well; the joint frequency of the AxC + GyA + UyU trio increases nearly 20-fold after the 30 min point. At 4 h and later the possibility of succession to even higher-ordered networks that subsume all simpler ones obfuscates individual trends. Visualization of all possible connections among genotypes underscores these conclusions (Fig. 3d). By 8 h the network is dominated by genotypes that can only be replicated via cooperation (green circles). In fact, the variance in the genotype frequencies drops monotonically over the course of the experiment, indicating that all genotypes increasingly participate in the network over time.

Serial transfer of the randomized population

The experiments depicted in Fig. 3 portray the dynamic changes that occur on a kinetic time scale as a batch of RNAs approaches equilibrium. In an actual prebiotic scenario, however, this effect would be iterated and perhaps magnified over several generations, as opposed to being an asymptotic value that results from mixing several RNAs in a single reaction vessel. To bring a stronger evolutionary flavour, we repeated the randomization experiment but in a serial transfer format. Starting with another aliquot of the exact same set of RNAs (that is, products from the same *in vitro* transcription), we carried a population through eight serial transfers, taking 10% of the population each hour into a fresh tube of fragments. In this manner the **W•X•Y•Z** molecules that spontaneously assemble are continually being fed with new fragments, such that selection will favour those molecules and networks that grow faster and persist over iterations. Given that the assembly that occurs each round can be strongly influenced by the actions of naive RNAs from the 90% fresh material, we opted to assay genotypic change by sampling only the most high-frequency genotypes: those present in an abundance greater than random chance (1/48). By manually sequencing the same number of genotypes (75) from transfers number 1 and 8 and enumerating those genotypes present more frequently than random expectation ($2/75 > 1/48$), we were able to observe the amalgamation of an RNA network over time (Fig. 4). At the 1 h time point, no closed network was possible and autocatalysts were relatively frequent (33%), but by 8 h a reflexively autocatalytic set was present in which every reaction is catalysed by at least one molecule involved in any of the reactions of the set[27]. This set included nine genotypes and fewer autocatalysts (25%), although the latter drop is not quite statistically significant (one-tailed *G*-test of independence; $P = 0.14$). Such expansion of the network to add additional genotypes is a more general case than the direct competition that we described in Fig. 2. As another indicator of the effect of serial transfer, the outcome of this experiment differed markedly from the batch assembly experiment (Fig. 3). After 8 h in the batch experiment the genotypes were dominated by pyrimidine-containing IGSs and targets (YzY; Fig. 3d). By contrast, the serial transfer experiment, although also reiterating the bias for the **Y–Z** junction, distinctly favoured IGS and target sequences containing purines (RzR; Fig. 4).

Fragmentation into four pieces

Lastly, we tested whether increased fragmentation of the RNA could provide additional complexity, and enhance the pre-biological relevance. We did this by breaking the molecule up into four pieces instead of two, creating four-piece versions of I_1, I_2 and I_3 analogously

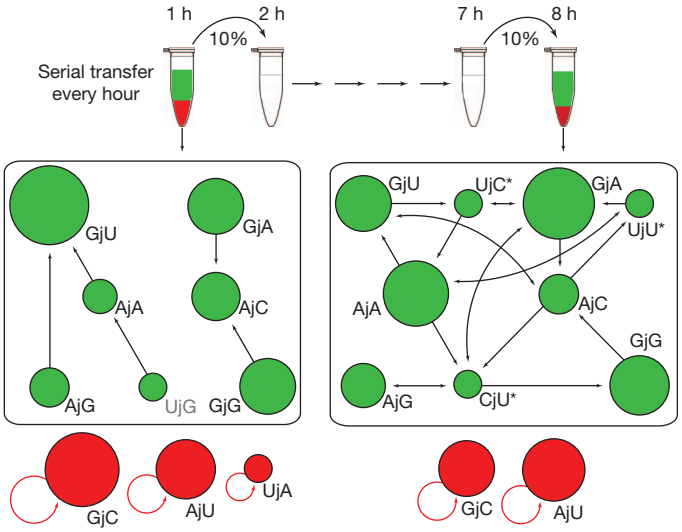

Figure 4 | The serial transfer experiment. The same RNA used to seed the randomization experiment (Fig. 3) was also subjected to a serial transfer protocol. For the first iteration, 50 pmol each of $_{GMG}\mathbf{W}_{CNU}$, $_{GMG}\mathbf{W•X}_{CNU}$, $_{GMG}\mathbf{W•X•Y}_{CNU}$, **X•Y•Z**, **Y•Z** and **Z** were incubated in a 100 µl volume. After 1-h time points, 10% of the reaction mixture was transferred to a new tube containing 90% fresh RNA with a total volume of 100 µl. The population was sampled via 5′ RACE and RT–PCR to capture variation in all positions of any **W•X•Y•Z** molecules present in the population. The 1 and 8 h populations were cloned, and genotype frequencies were obtained by manual sequence analysis of 75 clones each (Supplementary Table 3). Any genotype present twice or more was included on this diagram (see text); size of the circles scales to relative frequencies within their respective populations. All possible catalytic interactions are shown with arrows among non-autocatalytic genotypes (green), with autocatalytic genotypes (red) not participating in the network. Grey genotype in the 1st iteration disappears. Genotypes with asterisks appear by the 8th iteration.

to Fig. 1a, b. When we mixed the resulting 12 RNAs together, we observed two interesting phenomena (Fig. 5). First, the growth curve was distinctly sigmoidal, indicating that when more fragments are involved, the cooperativity of the system becomes even more apparent. In the four-piece fragmentation, **W•X•Y•Z** ribozymes can be made via many pathways, including those in which more than one enzyme cooperates to construct the product: for example, an E_1 ribozyme could recombine the **W–X** junction, an E_2 ribozyme could recombine the **X–Y** junction, and an E_3 ribozyme could recombine the **Y–Z** junction. Second, analysis of the sequences of the product **W•X•Y•Z** ribozymes showed that such cooperation was common (Supplementary Fig. 14). In fact 85% of all ribozymes required help from enzymes from at least two subsystems (Fig. 5).

Discussion

Our results illustrate a scenario in which simple autocatalytic cycles form easily but are later supplanted by more complex cooperative networks that take advantage of the autocatalysts. Our system describes the short-term kinetic phenomena that provide the foundation for evolutionary behaviour[10] in the presence of sequence variation throughout the ribozymes analogous to those described as "prelife"[9]. Features of the system described here that would make it relevant to early evolution are that it is comprised solely of RNA (although other polymers could display cooperative behaviour[17,18]) and that the 3-nt IGS or IGS targets are essentially the tag sequences[28] that have been suggested as a means to form molecular coalitions that can partition genetic information in a homogeneous milieu. Closure of autocatalytic sets would have been facilitated by the cooperative aggregation of oligomers with related tags[21]. Subsequent expansion of cooperative networks as shown here is possible by invasion of the network by a new set with a distinct tag sequence, for example, moving from the

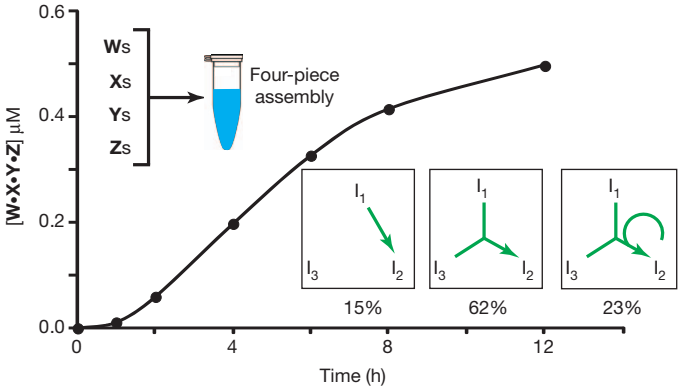

Figure 5 | Growth curve of a four-piece system. A more highly fragmented system based on that shown in Fig. 1b was created by breaking the ribozyme into four fragments for each I_i subsystem. The resulting 12 RNAs were co-incubated at 0.5 μM each, and samples were removed over time for both yield analysis (plot) and nucleotide sequence analysis (frequencies). The **W•X•Y•Z** RNAs can be assembled from a minimum of one, two or three IGS-bearing enzymes (examples shown with diagrams); the high frequencies of the latter two classes demonstrates the system's cooperativity.

three-membered cycle to a four-membered cycle such as by inclusion of a new IGS–IGS-target pair (Fig. 3d, starred genotypes), and then well beyond four members (Fig. 4). Longer-term evolutionary optimization would have required spatial heterogeneity[29] or compartmentalization[8,30] to provide lasting immunity against parasitic species or short autocatalytic cycles. Over time, a transition back to purely selfish replicators based on polymerization chemistry could proceed[7].

In our system, we show how RNA networks have the potential to arise spontaneously and to buffer informational decay. A key to the latter is the use of recombination for replication. Although allowing for some genotypic variability, it does not lead to the accumulation of deleterious mutations as does template-directed polymerization[31]. Highly interdependent networks of genetically related replicators as a means to circumvent the error catastrophe in nascent life have been proposed[11]. The three-membered cycle shown here resembles a hypercycle as envisioned previously[4,21,32], but without hyperbolic growth. We prefer to focus on the observation that the cycle can be derived from simpler cycles and has the potential to expand to more complex ones as evidence that RNA molecular coalitions can show spontaneous order-producing dynamics, which already has theoretical support[27]. Molecular ecological succession is a plausible model for a bridge between selfish replicators and cooperative systems.

METHODS SUMMARY

Experimental. Ribozyme assays or covalent self-assembly from oligomers were performed as described previously[24,25]. Briefly, RNA oligomers were incubated together in 100 mM $MgCl_2$ and 30 mM EPPS buffer (pH 7.5) for 5 min–16 h at 48 °C at a final concentration of 0.01–2.0 μM each. Visualization and quantification was possible via phosphorimaging when **W**-containing fragments were 5′-end-labelled with $\gamma[^{32}P]$ATP before use. For genotyping, ~200-nt RNA was excised from a gel and subject to PCR with reverse transcription (RT–PCR) using **W**- and **Z**-specific primers. High-throughput sequence analysis on the Illumina platform was possible after 5′ RACE to capture the sequence variability in the IGS of assembled ribozymes. For manual sequence analysis, the PCR products were cloned into *Escherichia coli* and individual colonies were picked for colony PCR reactions. Resulting amplicons were either subjected to nucleotide-sequence analysis or restriction fragment length polymorphism (RFLP) analysis.

Modelling. The cooperative system was modelled as a set of six differential equations describing the concentrations over time of the six principal species (see Supplementary Information). These equations are derived from the detectable catalysis reactions (encompassing six direct-catalysis reactions and three cross-catalysis reactions). The experimental time series data from the full three-component system and from the two-component subsystems yielding detectable product were fit simultaneously to the model by standard optimization techniques.

Received 11 March 2011; accepted 6 September 2012.
Published online 17 October; corrected online 31 October 2012.

1. Joyce, G. F. RNA evolution and the origins of life. *Nature* **338,** 217–224 (1989).
2. Zaher, H. S. & Unrau, P. J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13,** 1017–1026 (2007).
3. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332,** 209–212 (2011).
4. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* **64,** 541–565 (1977).
5. Kauffman, S. A. *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ. Press, 1993).
6. Sievers, D. & von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* **369,** 221–224 (1994).
7. Levy, M. & Ellington, A. D. The descent of polymerization. *Nature Struct. Biol.* **8,** 580–582 (2001).
8. Szathmáry, E. The origin of replicators and reproducers. *Phil. Trans. Royal Soc. B* **361,** 1761–1776 (2006).
9. Nowak, M. A. & Ohtsuki, H. Preevolutionary dynamics and the origin of evolution. *Proc. Natl Acad. Sci. USA* **105,** 14924–14927 (2008).
10. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58,** 465–523 (1971).
11. Maynard Smith, J. Hypercycles and the origin of life. *Nature* **280,** 445–446 (1979).
12. Kun, Á., Santos, M. & Szathmáry, E. Real ribozymes suggest a relaxed error threshold. *Nature Genet.* **37,** 1008–1011 (2005).
13. Doudna, J. A. & Szostak, J. W. RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339,** 519–522 (1989).
14. Joyce, G. F. Forty years of *in vitro* evolution. *Angew. Chem. Int. Ed.* **46,** 6420–6436 (2007).
15. Lincoln, T. A. & Joyce, G. F. Self-sustained replication of an RNA enzyme. *Science* **323,** 1229–1232 (2009).
16. Levy, M. & Ellington, A. D. Exponential growth by cross-catalytic cleavage of deoxyribozymogens. *Proc. Natl Acad. Sci. USA* **100,** 6416–6421 (2003).
17. Lee, D. H., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* **390,** 591–594 (1997).
18. Lee, D. H., Severin, K. & Ghadiri, M. R. Autocatalytic networks: the transition from molecular self-replication to molecular ecosystems. *Curr. Opin. Chem. Biol.* **1,** 491–496 (1997).
19. Voytek, S. B. & Joyce, G. F. Niche partitioning in the coevolution of 2 distinct RNA enzymes. *Proc. Natl Acad. Sci. USA* **106,** 7780–7785 (2009).
20. Kim, D.-E. & Joyce, G. F. Cross-catalytic replication of an RNA ligase ribozyme. *Chem. Biol.* **11,** 1505–1512 (2004).
21. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part C: the realistic hypercycle. *Naturwissenschaften* **65,** 341–369 (1978).
22. Reinhold-Hurek, B. & Shub, D. A. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357,** 173–176 (1992).
23. Hayden, E. J. & Lehman, N. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem. Biol.* **13,** 909–918 (2006).
24. Hayden, E. J., von Kiedrowski, G. & Lehman, N. Systems chemistry on ribozyme self-construction: evidence for anabolic autocatalysis in a recombination network. *Angew. Chem. Int. Ed.* **47,** 8424–8428 (2008).
25. Draper, W. E., Hayden, E. J. & Lehman, N. Mechanisms of covalent self-assembly of the *Azoarcus* ribozyme from four fragment oligonucleotides. *Nucleic Acids Res.* **36,** 520–531 (2008).
26. Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard Univ. Press, 2006).
27. Hordijk, W. & Steel, M. Detecting autocatalytic, self-containing sets in chemical reaction systems. *J. Theor. Biol.* **227,** 451–461 (2004).
28. Weiner, A. M. & Maizels, N. 3′ terminal tRNA-like structures tag genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl Acad. Sci. USA* **84,** 7383–7387 (1987).
29. Boerlijst, M. C. & Hogeweg, P. Spiral wave structures in prebiotic evolution: hypercycles stable against parasites. *Physica D* **48,** 17–28 (1991).
30. Szathmáry, E. & Demeter, L. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128,** 463–486 (1987).
31. Lynch, M., Burger, R., Butcher, D. & Gabriel, W. The mutational meltdown in asexual populations. *J. Hered.* **84,** 339–344 (1993).
32. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part B: the abstract hypercycle. *Naturwissenschaften* **65,** 7–41 (1978).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank A. Burton, R. Ghadiri, P. Higgs, B. Larson, K. Chacón and A. López García de Lomana for help during preparation of this manuscript. This work was supported by NASA grant NNX10AR15G to N.L., the Center for Life in Extreme Environments at Portland State University, and a fellowship from the Human Frontier Science Program Organization to R.X.-B.

Author Contributions N.L. and N.V. conceived the basic idea and conducted the experiments; E.J.H. and I.A.C. contributed to the evaluation of the results; I.A.C., M.L.M. and R.X.-B. constructed the mathematical models; N.L. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.L. (niles@pdx.edu).