Birth of immunity in early life: stepwise emergence of resistance and immunity to complex molecular parasites via progressively degenerating hyperparasites

Short title:

parasite immunity from hyperparasitism

Authors:

Magnus Pirovino[1], Christian Iseli[2], Joseph A. Curran[3], Bernard Conrad[4*]

Affiliations: [1]OPIRO Consulting Ltd., FL-9495 Triesen, Principality of Liechtenstein, [2]Bioinformatics Competence Center, EPFL and Unil, CH-1015 Lausanne, [3]Department of Microbiology and Molecular Medicine, and Institute of Genetics and Genomics of Geneva (iGE3), Geneva University, CH-1211 Genève 4, [4]Genesupport, Avenue de Sévelin 18, CH-1004 Lausanne

*Correspondence to: bernard.conrad@genesupport.ch

Autocatalytic networks likely played a central role in the shift from inanimate systems to protocells. Using an established hyperparasite framework, we model the emergence of antiparasite immunity using the Lotka Volterra equations. It builds on autocatalytic, increasingly complexed RNA networks reflecting the historical exposure to ubiquitous molecular parasites. First, upon parasite encounter, specific catalysts providing parasite resistance appear in primordial autocatalytic cycles, supplanted by promiscuous, efficiently self-replicating, but parasite sensitive ribozymes. This is supported by catalyst tradeoff analysis. Substrate promiscuity confers high catalytic activity, but entails parasite exposure, higher substrate specificity implies relatively lower activity, yet it offers parasite resistance. Second, sustained parasite influx generates hyperparasite cycles, reactivating the original, parasite-specific catalyst-subunit that now once more provides parasite resistance, in a homeostatically stabilized habitat, empirically consistent with phage satellites and defective RNA viruses induced across a range of parasite loads. Third, under continuous exposure to progressively elaborate parasite populations, hyperparasites degenerate and embody antiparasite immunity. Again, the observed triggering of antiviral immunity by phage satellite encoded antiphage systems and defective RNA virus genomes are in agreement with this model. As such, it offers an attractive and unifying theory for the spontaneous birth of immunity in early life.

Introduction

The multistep transition process from inanimate to living systems has been dated to >3.9 Ga [1]. A non-exhaustive list of thresholds that needed to be progressively surmounted during this process has been drawn, comprising, among others, chirality symmetry breaking, spontaneous polymerization, self-assembly of compartments, and particularly the catalytic closure threshold [2]. One central pillar in the theory of how early life could have materialized pertains to the “RNA world”. This stipulates that the initial informational polymer was RNA, and that the same molecule was also the fundamental catalyst [3,4]. The underlying concept of autocatalytic sets was laid out more than half a century ago [5,6]. It was subsequently refined [7] and validated in a proof of principle experiment showing that RNA fragments could build self-replicating ribozymes from individual RNA fragments via cooperative catalytic cycles [8]. In parallel, an RNA polymerase ribozyme was generated by *in vitro* evolution that exhibited a high level of processivity [9]. Collectively, both theoretical considerations and experimental evidence support the view of a gradual evolution towards RNA self-catalysis. This started with self-assembly, followed by the formation of autocatalytic networks using self-assembled substrates. From this emerged different ribozymes (recombinases, ligases) and template based ligation, leading to a template based, self-replicase catalyzed assembly of monomers [10].

Prebiotic heredity in chemical networks is theorized to comprise variable autocatalytic cycles that form a core (genotype) and a non-catalytic periphery (phenotype), allowing limited heredity and selectability (11). Complexed mixtures of cross-replicating RNA enzymes were shown to undergo self-sustained Darwinian evolution with the appearance of both, highly active enzymes that maintained the population, and poorly active enzyme passengers [12]. In conclusion, the coexistence of prebiotic heredity with an emerging, template-based canonical Darwinian heredity appears plausible [10].

The ribozyme catalyzed RNA-polymerization process is inherently error prone [9], generating variant sequence space mutants called a quasispecies [13]. Mutants vary in frequency, length, sequence and structural-complexity, as well as in ribozyme activity, thus providing selectable traits in the dimensional fitness landscape, epitomizing the mutation-selection balance [14]. This variant sequence-space intrinsically implies the obligate presence and evolutionary persistence of the molecular parasites associated with all life forms, as argued essentially for thermodynamic reasons [15]. In precellular life, the most abundant class of genetic parasites were most likely capsidless. These subsequently gave rise to capsidless selfish genetic elements and to capsid-encoding viruses that diverged from cellular, ribosome-encoding organisms [16]. Unsurprisingly, nucleic acid polymerase evolution also most likely bifurcated along the same paths based on distinct RNA recognition motifs [17].

It has been noted within the realm of the virosphere - in all Baltimore classes [18] - that genetic parasites tend to be parasitized, typically forming a tripartite relationship. This is exemplified by bacteriophages (phages), notorious parasites of bacteria that are themselves parasitized by phage satellites [19,20]. Other typical threesome associations are observed in RNA viruses and Nucleocytoplasmic Large DNA viruses (NCLDVs): for RNA viruses, it comprises the host, the RNA virus and defective viral genomes [21], for NCLDLVs it is the eukaryotic host, the NCLDV and its dedicated virophage from the same group of Bamfordviruses [22].

We previously developed a hyperparasite framework based on the Lotka Volterra equations [23], demonstrating that superior homeostasis arose from the tripartite configuration in a given habitat, namely, the host, the parasite and the hyperparasite, i.e. the parasite of the parasite. The aforementioned empirical evidence lends substantial credit to this. This reasoning has now led us to extend the model, building on complexed autocatalytic RNA networks with prebiotic heredity reflecting the historical parasite exposure, and showing, how parasite resistance and antiparasite immunity can be selected for and gradually emerge from progressively degenerating hyperparasites in an intrinsic fashion in early life, in hosts exposed to increasingly more complexed parasite populations.

Methods

Definitions and Terms

In what follows we assume the following: the Malthusian fitness r of a replicating (biochemical) system (or species) is defined by $\frac{d}{dt}R = \dot{R} = r\,R$, where R denotes the population size of the replication sys-tem or species. Evolutionary dynamics between two or more interacting species or systems, e.g. R, S, P, … are described by the following type of replicator differential equation e.g. for R:

$\dot{R}$ = "(A) terms of dynamics using habitat resources" x "(B) term of habitat resource restriction" + "(C) terms of dynamics not using habitat resources"

Terms A include both "terms of first order kinetics" and "terms of second order kinetics"

- "Terms of first order kinetics" are typically natural growth or decay rates, e.g. $\dot{R} = r\,R$. If $r > 0$, growth process described by rR eats into resources, whereas if $r < 0$, r is a decay rate, and de-cay dynamics does not.

  Terms of second order kinetics are derived from encounters between different species. For instance, the term *"bRS"* in *"$\dot{R} = -aR + bRS$"* has the following meaning: $bRS =$
  $(probabiliy\ of\ an\ encounter\ between\ members\ of\ R\ and\ S)$ x $(impact\ of\ encounter\ on\ the\ fitness\ of\ R)$.

  Here we assume, according to the Lotka Volterra calculus that the probability of an encounter between members of R and S is proportional to the product RS.

  The term B of resource restriction is used in a habitat of limited resources. For instance, in the classic logistic equation $\dot{R} = r\,R\left(1 - \frac{R}{K}\right)$ the term $\left(1 - \frac{R}{K}\right)$ indicates that the habitat has limited resources supporting a maximum population K for R.

  Since every encounter involves resource consuming, we assume here, that all terms of second order kinetics are subject to resource restrictions and therefore need to be multiplied with the term B.

Example of a host-parasite interaction (R=host, P=parasite):

$$\dot{R} = (a\,R - bPR)\left(1 - \frac{R+P}{K}\right)$$

$$\dot{P} = -c\,P + dRP\left(1 - \frac{R+P}{K}\right)$$

where $a > 0, b > 0, c > 0, d > 0$. The first order dynamics growth process aR and all the second order processes eat into the habitat's resources, whereas the first order decay process $-c\,P$ does not. The habitat resources restriction is represented by the term $\left(1 - \frac{R+P}{K}\right)$, indicating that a maximum population size of $K = R + P$ can be nourished by the habitat and that both, S and P, compete for the same resources.

Definition of catalysts, ribozymes and enzymes

We use the term *catalyst*, sometimes also *perfect catalyst*, for any substance that accelerates the rate of a reaction without undergoing any change during the process. We call an *imperfect catalyst* a substance that accelerates the rate of a reaction, but that may undergo a change during the process; antibodies in the transition from hyperparasitism to antiparasitism are examples of imperfect catalysts (see process $(3')$, Suppl. Inf.). C*atalysts* displaying relatively higher substrate specificity generally have relatively lower catalytic activity, for instance the protoreplicase. A subgroup of catalysts is called here *enzymes*, if they are generated based on RNA templates. An enzyme may be either a *ribozyme*, if it is also used as an RNA template to generate other enzymes, or a *pure enzyme* if it cannot be used as an RNA template for other polymerization processes. The term *ribozyme* encompasses different enzymatic activities, such as recombinase, ligase and polymerase. Recombinases and ligases predominate in early protocycles, while polymerases appear only in advanced networks.

Model $\mathbf{1}^{(1)}$: habitat with one host autocatalytic cycle exposed to parasite invasion

$$\frac{dR_2}{dt} = a\ R_1 R_1 \Pi - d_{R_2} R_2$$
$$\frac{dR_1}{dt} = (b\ R_2 R_1 - \delta R_2 P)\Pi - d_{R_1} R_1$$
$$\frac{dP}{dt} = c R_2 P \Pi + \alpha - d_P P,$$

where a, b and c are positive constants, d_{R_i}, d_P are the decay rates of populations $R_i, i = 1, 2$, and of parasite P absent its host R_2; δ is parasite's P rate of exploitation of the host replicase R_2; and the habitat restriction term is defined as $\Pi = \left(1 - \frac{R_1 + R_2 + P/q_P}{K}\right)$, where $q_P > 1$, since typically parasites P are more efficient using the resources of the habitat than their hosts R. α is the parasite invasion parameter, i.e., per unit of time the habitat is invaded by α population units of parasites P.

Model $\mathbf{1}^{(0)}$habitat with primordial autocatalytic cycle, exposed to template invasion $\boldsymbol{P}$ (to become parasite later in evolution) and protoenzyme $\boldsymbol{E_1}$:

$$\frac{dP}{dt} = \alpha - d_P P$$
$$\frac{dE_1}{dt} = \beta - d_{E_1} E_1$$
$$\frac{dE_2}{dt} = a E_1 F_1 \Pi - d_{E_2} E_2$$

$$\frac{dF_1}{dt} = bE_2P\Pi - d_{F_1}F_1,$$

where

E_2 is the protoreplicase produced by this primitive autocatalytic cycle,

F_1 is the prototemplate produced by this primitive autocatalytic cycle,

α is the inflow rate of RNA-template P into the habitat,

β is the inflow rate of protoenzyme/catalyst E_1 into the habitat,

a and b are positive constants,

d_{E_i}, d_{F_1}, d_P are the decay rates of enzyme populations $E_i, i = 1, 2$, of prototemplate d_{F_1}, and of RNA-template P,

$\Pi = \left(1 - \frac{E_2 + {}^{F_1}/_{q_{F_1}}}{K}\right)$ is the habitat restriction term where $q_{F_1} > 1$, since we assume that proto-templates F_1 are more efficient using the resources of the habitat than the more complex protoreplicase E_2.

Model* $\mathbf{1}^{(1+)}$: *stabilized host-parasite-hyperparasite cycle

$$\frac{dR_2}{dt} = (a\ R_1R_1 - \delta_2R_1F_1)\Pi - d_{R_2}R_2$$

$$\frac{dR_1}{dt} = (b\ R_2R_1 - \delta_1R_2P)\Pi - d_{R_1}R_1$$

$$\frac{dP}{dt} = (cR_2P - \delta_3F_2P)\Pi + \alpha - d_PP$$

$$\frac{dF_1}{dt} = (d\ F_2P + \beta R_2P)\Pi - d_{F_1}F_1$$

$$\frac{dF_2}{dt} = (e\ R_1F_1)\Pi - d_{F_2}F_2,$$

where

a, b, c, d, e are positive (reproduction) constants,

R_1 is a ribozyme of the autocatalytic host cycle,

R_2 is the replicase, i.e. another ribozyme (or may be also pure enzyme since it is not used as template) of the autocatalytic host cycle,

P is the parasite, exploiting replicase R_2 of the host cycle (exploitation parameter δ_1),

F_1 is the prototemplate exploiting the host R_1 (exploitation parameter δ_2) for protoreplicase F_2 production, triggered upon encounter of R_2 with P (trigger parameter β) and further produced by protoreplicase F_2 upon encounter with P,

F_2 is the protoreplicase (pure enzyme or ribozyme) producing F_1 upon encounter with parasite templates P ($F_2 = E_2$ from the primitive autocatalytic cycle of Model $1^{(0)}$), thus, exploiting template of parasite P (exploitation parameter δ_3),

d_{R_i}, d_{F_i}, d_P are the decay rates of populations $R_i, F_i, i = 1, 2$, and of P, absent new inflow or reproduction in the habitat,

α is the inflow rate of RNA template P into the habitat,

β is trigger parameter of replicase R_2 producing F_1 by encounter with parasite P,

$\Pi = \left(1 - \frac{R_1 + R_2 + {}^{P}/_{q_P} + {}^{F_1}/_{q_{F_1}} + {}^{F_2}/_{q_{F_2}}}{K}\right)$ is the habitat restriction term where $q_P, q_{F_1}, q_{F_2} > 1$, since we assume that parasitizing cycles of P and F_1, F_2 are more efficient using the resources of the habitat than the more complex host cycle R_1, R_2.

***Model* $\mathbf{1}^{(2)}$: integrated host-parasite-hyperparasite cycle without new parasite inflow**

The simulation results of Model $1^{(2)}$ are realized as follows: Model $1^{(1+)}$ provides the starting basis, with arbitrary initial populations to establish a stabilized situation, given a moderate parasite inflow with $\alpha > 0$ (Fig. 3lefthand part). Subsequently, Model $1^{(1+)}$ is adapted to Model $1^{(2)}$. All the parameters of Model $1^{(2)}$ are the same as those of Model $1^{(1+)}$, with the exception of the parasite inflow, now $\alpha = 0$. As initial conditions for the populations of Model $1^{(2)}$, we assume the terminal populations of the simulations of Model $1^{(2)}$, as calculated for Fig. 3 at the beginning of the righthand part. The larger fluctuations of the population at the beginning of the righthand part of Fig. 3 stem from the transition from Model $1^{(1+)}$ to Model $1^{(2)}$, where the parasite inflow abruptly decreases from $\alpha > 0$ to $\alpha = 0$. The fact that parasites in Model $1^{(2)}$ are now reproduced without new inflow shows that they can be viewed as fully integrated into the autocatalytic network, realized in $1^{(2)}$ (see Suppl. Inf.).

Model* $\mathbf{1}^{(1-)}$: *transition from hyperparasitism to antiparasitism

$$\frac{dR_2}{dt} = (a\ R_1R_1 - \delta_2 R_1F_1)\Pi - d_{R_2}R_2$$

$$\frac{dR_1}{dt} = (b\ R_2R_1 - \delta_1 R_2P)\Pi - d_{R_1}R_1$$

$$\frac{dP}{dt} = (cR_2P - \delta_3 F_2P)\Pi + \alpha - \theta_P F_2P - d_P P$$

$$\frac{dF_1}{dt} = (d\ F_2P + \beta R_2P)\Pi - d_{F_1}F_1$$

$$\frac{dF_2}{dt} = (e\ R_1F_1)\Pi - \theta_{F_2}F_2P - d_{F_2}F_2,$$

Where all the parameters are the same as in Model $1^{(1+)}$, with the exception that now two additional parameters are added, namely θ_P and θ_{F_2}. The term $\theta_P F_2 P$ denotes the number of population units of parasites P made unviable per unit of time during the encounter of protoreplicase F_2 with P, now also functioning as antibody. In addition, the term $\theta_{F_2} F_2 P$ denotes the number of population units of protoreplicase/antibody F_2 made unviable per unit of time during the encounter of protoreplicase/antibody F_2 with P. For instance, if $\theta_p = \theta_{F_2}$, then the same number of antibodies and parasites are made unviable during their encounter. In the evolutionary transition from hyperparasitism to antiparasitism F_2 can still act both ways, as imperfect protoreplicase and as antibody directed against the parasite P. We call it *imperfect* protoreplicase or imperfect catalyst, because it still produces/catalyzes prototemplate F_1 upon encounter with template P. But now F_2 may be damaged during this process, therefore cannot be called a *perfect* catalyst anymore. The replication differential equations of Model $1^{(1-)}$ realizing a pure antiparasitism situation are the same as in Model $1^{(1-)}$ of Fig. 4A, except that the reproduction parameters hyperparasite d and δ_3 are now set to zero: $\delta_3 = d = 0$.

Results

In a stepwise and gradual process outlined in the Supplementary Information (proofs, Suppl. Inf.) and Methods sections, we lay out using the Lotka Volterra equations, how simple autocatalytic cycles are initially built and how over time they become increasingly more elaborated as a result of molecular parasite exposure. The coexistence of prebiotic heredity with an emerging canonical, template-based Darwinian heredity [10-12] allows for the complexification of the quasispecies that in turn reflects its historical parasite exposure. The previously established tripartite habitat framework featuring a host, a parasite that is itself parasitized (hyperparasite; [23]), is instrumental for homeostasis and the emergence of resistance, and finally immunity to the increasingly more exhaustive parasite populations.

Autocatalytic cycles in transitional early RNA-settings

The seminal duality of the RNA world, namely the fact that the same RNA molecule provides both the template and the self-replicating catalyst [3,4] constitutes the foundation of the initial autocatalytic cycles (replication equations (1.1), (1.2), forming a two-step autocatalytic cycle; proof outlined in Suppl. Inf.). Since the template and the enzyme/ribozyme remain in principle unchanged during this process, both can be viewed as catalysts, as long as the molecules do not undergo deleterious loss of function mutations.

235

236 ***Parasitism in autocatalytic cycles***

237 Molecular parasites are an inherent feature of the replication process, they are inseparably
238 associated with all life forms [24], for a number of reasons, the most important being the accuracy-
239 rate tradeoff [25-26]. Average enzymes operate orders of magnitude (k_{cat}/K_M of $\sim 10^5$ $s^{-1}M^{-1}$) below
240 the diffusion rate (10^8-10^9 $s^{-1}M^{-1}$), which implies promiscuous substrate binding, high catalytic activity,
241 and as a secondary tradeoff a high mutation rate with a large variance in the quasispecies (speed
242 being more important than enzyme accuracy; [26,27]). In an initial autocatalytic cycle, self-replicating
243 ribozymes emerge (ribozyme R_2 acting on a replicating template R_1; proof outline in Suppl. Inf.;
244 Model $1^{(1)}$ in Methods). The catalyst tradeoff opposes specific, but catalytically relatively less active
245 molecules, and promiscuous, yet catalytically highly efficient molecules [26]. Therefore, the
246 catalytically very efficient and promiscuous R_2 will inevitably be parasitized by RNA templates P.
247 Figure 1 illustrates the population sizes of the habitat constituents under high (Fig. 1A), and low (Fig.
248 1B) parasite exposure. As expected, high parasite loads overwhelm the habitat with P templates, an
249 observation that begs a solution.

250

251 ***A stable, primordial autocatalytic cycle in the presence of parasites***

252 In a primordial habitat with abundant inflow of molecular RNA species including parasites, a
253 quasispecies will eventually emerge with the central feature that it also comprises molecules with
254 superior parasite-specificity, which implies relatively lower catalytic activity, and confers parasite
255 resistance (protoreplicase $E_{2,}$ prototemplate F_1; equations and proofs outlined in the Suppl. Inf.).
256 This composite feature progressively arises in successive rounds of autocatalytic cycles, initially
257 comprising the encounter of the catalyst $E_{1,}$ provided by a neighboring habitat, with F_1 evolved
258 autocatalytically in the habitat, producing the primordial catalyst $E_{2,}$ that is capable of amplifying F_1
259 but not P. It is the twofold consequence of the primordial parasite-specificity of $E_{2,}$, and of the
260 complexity of P that exceeds the processivity of the protoreplicase. As a corollary and as displayed in
261 Fig. 2, selection of such molecules confers considerable resistance to parasite invasion, and restores
262 host template and catalyst abundance (Model $1^{(0)}$, Methods). There is both, strong theoretical and
263 experimental evidence to support these findings. Indeed, while it is generally assumed that catalysts
264 arising in early autocatalytic cycles have low activity and high promiscuity [25], an in depth physico-
265 chemical analysis of how enzymes meet demands of selectivity and accuracy points towards
266 accuracy-rate tradeoffs [26]. Selectivity (accuracy, specificity) implies the ability to discriminate
267 between two templates, provided both are present [26], and rate denotes speed [25,26]. Ground
268 state discrimination means that specificity is achieved mainly through substrate binding, which

269 imposes strong accuracy-rate tradeoffs [26]. Improvements in selectivity mediated by tighter cognate
270 substrate binding invariably lead to lower catalytic efficiency (parallel decreases in the constants for
271 the cognate substrates K_M^{cog} and k_{cat}^{cog}; [26]). This is precisely what the model stipulates, the
272 protoreplicase has a primal parasite-selectivity, which comes at the cost of a relatively reduced
273 catalytic activity, yet it allows for adequate host molecule-, but not parasite-amplification. In
274 conclusion, if catalysts with relatively reduced function in the quasispecies confer parasite resistance,
275 yet restore host catalyst and template abundance, they will be positively selected [14].
276
277 ***Evolution of the primordial autocatalytic cycle into a full-fledged autocatalytic cycle***
278 Natural selection in the mutant quasispecies space will ultimately produce, starting from E_2, a fully
279 efficient, self-replicating ribozyme (replicase R_2) that is bestowed with high catalytic activity [9],
280 capable of generating all enzymes and templates in the habitat (equations and proofs outlined in the
281 Suppl. Inf.). However, this has a cost. R_2 is promiscuous, catalytically highly efficient and therefore
282 exposed to parasite invasion, which was not the case for its primordial catalyst precursor with
283 relatively lower activity (protoreplicase E_2; Model $1^{(1+)}$ stabilized host-parasite-hyperparasite cycles,
284 Methods).
285 At this point, the protoreplicase E_2-subdomain of R_2 confers P selectivity, restoring parasite
286 resistance and rescuing the abundance of the partial catalyst and its templates, even with a moderate
287 parasite inflow ($\alpha > 0$, Fig. 3 left panel). This process has as its central feature a hyperparasite cycle
288 (equations and proofs outlined in the Suppl. Inf.). It is further noteworthy that F_2 (formerly E_2) is still
289 produced (Fig. 3), and is acting additively to R_2, which is important to generate antiparasitism later.
290 Under permanent and moderate, thus manageable parasite inflow ($P^{special}$, Suppl. Inf.), an
291 increasingly ramified network of autocatalytic cycles arrives in the habitat that persists even when
292 the parasite exposure ceases, constituting a historical record of past parasite exposures ($\alpha = 0$, Fig. 3
293 right panel).
294 There is considerable evidence in favor of the essential components of this model, in
295 particular the catalyst accuracy-rate tradeoff and the hyperparasite cycle. Firstly, a solid body of
296 theoretical and experimental evidence supports catalyst accuracy-rate tradeoffs [25,26]. Secondly,
297 metagenomic studies demonstrate the virtually universal presence of tripartite constituents across
298 molecular parasite habitat classes, most persuasively for bacteria-phages-phage satellites [19,20], for
299 RNA viruses partnering the host- virus-defective interfering RNAs [21], and for NCLDVs, with
300 eukaryotic host-NCLDV-virophage [22]. Thirdly, additional support comes from the observation that
301 hyperparasites are induced across the range of parasite concentrations, and especially at high
302 multiplicity infection for RNA viruses [28,29].
303

304 ***Emergence of antiparasitism in autocatalytic cycles***

305 Under conditions, where the sheer number and complexity of incoming parasite populations would
306 overwhelm the carrying capacity of the habitat, a degenerative process is initiated that erodes
307 catalysis. Formally, the hyperparasite cycle degenerates to an antiparasite cycle (equations and proofs
308 outlined in the Suppl. Inf.; Model $1^{(1-)}$, Methods). In Figures 4 and 5, the transition from
309 hyperparasitism to antiparasitism is illustrated, in the sense that both coexist at equal activity, partial
310 catalyst and non-catalytic antiparasitic activities (Fig. 4A, see Methods for parameters). In Figure 4B,
311 antiparasitism is fully implemented; hyperparasitism has been fully replaced by the non-catalytic
312 antiparasitism (reproduction parameters hyperparasite d and δ_3 are set to zero: $\delta_3 = d = 0$), making
313 the difference between pure antiparasitism, and the transition from hyperparasitism to
314 antiparasitism minor. In Figure 4B, the population fluctuations are more volatile than in Figure 4A, but
315 pure antiparasitism does effectively stabilize the autocatalytic cycle, without a hyperparasite
316 component. Figure 5 shows the gradual transition from pure hyperparasitism (top panel) to
317 antiparastism (botton panel), in four successive steps (0, 0.25, 05. 0.75 and 1), progressively replacing
318 hyperparasitisim with antiparasitism. Therefore, we conclude that antiparasitism can replace
319 hyperparasitism in situations with a stronger and highly diverse parasite inflow that would overrun
320 the carrying capacity of the autocatalytic cycle in the habitat.

321

322

323 **Discussion**

324 We have outlined in a stepwise fashion, how increasingly elaborate autocatalytic cycles gradually
325 emerge in an early life RNA-habitat. Initially, catalysts with relatively lower catalytic activity are
326 selected, because they are parasite-specific and confer resistance (primordial autocatalytic cycles).
327 Soon, complexed mixtures evolve, featuring self-replicating catalysts of high catalytic activity that
328 become stabilized through the selection and integration of hyperparasite cycles, including relatively
329 less active catalyst-subunits that confer powerful parasite resistance (stabilized host-parasite-
330 hyperparasite cycles). One quintessential building block materializes via prebiotic heredity,
331 encompassing a core with variable autocatalytic cycles (genotype), and a non-catalytic periphery
332 (phenotype). Conjointly with the mutation-selection balance, natural selection can operate on the
333 genotype rendering possible complexed networks that reflect the history of parasite exposure. Under
334 conditions in which there is the influx of overwhelmingly broad and complexed parasite populations
335 into the habitat, the carrying capacity of the autocatalytic networks are exceeded, and a non-
336 catalytic, degenerative process of hyperparasite cycles ensues that leads to the appearance of anti-
337 parasitism, and the birth of immunity in early life.

338 One of the strong features of this model is the solid empirical evidence supporting it. First of
339 all, the catalyst tradeoff analyses [26] applied to autocatalytic RNA networks [25] predict the
340 emergence of increasingly active ribozymes, binding promiscuity and high catalytic activity being
341 balanced with specificity and relatively lower activity, which is precisely what the model shows.
342 Moreover, tripartite microbial populations composed of a host, a parasite and a hyperparasite are
343 virtually ubiquitous, as evidenced for bacteria-phages-phage satellites [19,20], host-RNA viruses-
344 defective interfering RNA viruses [21], and for eukaryotic host-NCLDVs-virophages [22,30]. We
345 previously developed a hyperparasite framework based on the Lotka Volterra equations that
346 corroborates the superior homeostatic viability of tripartite habitats [23]. It thus becomes credibly
347 likely, that early life networks of habitats with RNA populations were constructed following this
348 architectural logic. Second, prebiotic heredity has a theoretical founding [8] and experimental
349 support [9,10,31]. Perhaps concepts based on information theory [32] could more appropriately
350 capture the meaning attached to the genotypic core composed of autocatalytic cycles, and its
351 exchange with the world of selfish genetic elements, as we describe it here in our model. Third, it is
352 striking that the appearance of microbial immunity is associated with hyperparasites. Indeed, several
353 recent studies show that phages and their satellites encode a set of diverse antiphage systems
354 [19,33], and defective interfering RNAs activate innate immunity [21,28]. Phage satellite-encoded
355 immune systems protect bacteria from phage predation and were proposed to be an integral innate
356 immune component [34]. This notion could easily be extended to omnipresent, bona fide
357 hyperparasites, such as endogenous retroviruses [35,36]. In conclusion, our model provides an
358 overarching theoretical foundation for the notorious association of immunity with the parasite-
359 hyperparasite sphere that is rooted in early life.
360 Furthermore, it should be noted that extant RNA viral quasispecies most closely resemble the
361 primordial RNA populations described here. Large populations exhibiting fast replication recapitulate
362 the fundamental tradeoff, favoring speed over polymerase accuracy [27]. High mutation rates enable
363 the emergence and maintenance of substantial genetic variation, which leaves marks of the past
364 ecological history in viral genomes via selective sweeps [37]. Moreover, in all contemporary tripartite
365 models the host diverged to cellular life forms carrying polymerases phylogenetically distinct from
366 those prevalent in viruses and selfish genetic elements [17], thus severely limiting genetic exchange.
367 This represented a marked change from the primordial tripartite populations that freely swapped
368 catalysts.
369 This model has by design intrinsic limitations, notably the fact that the conclusions have not
370 yet been experimentally verified. Some components of the model may even not be amenable to
371 realistic experimentation, especially highly efficient self-replicating RNA catalysts [9,10]. Along the
372 same line, it is impossible to infer causation from statistical associations, particularly the pervasive

horizontal gene transfer associated with selfish genetic elements blurs the lines. These aspects will be addressed in future studies.

References

1 Betts H. C., Puttick M. N., Clark J. W., Williams T. A., Donoghue P. C. J., et al. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556-1562 (2018).

2 Jeancolas C, Malaterre C, Nghe P. Thresholds in Origin of Life Scenarios. *iScience* **23**, 101756; 10.1016/j.isci.2020.101756 (2020).

3 Robertson M. P., Joyce G. F. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608; 10.1101/cshperspect.a003608 (2012).

4 Joyce G. F, Szostak J. W. Protocells and RNA Self-Replication. *Cold Spring Harb. Perspect. Biol.* **10**, a034801; 10.1101/cshperspect.a034801 (2018).

5 Kauffman, S. A. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *J. Cybernet.* **1**, 71-96 (1971).

6 Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465-523 (1971).

7 Kauffman S. A. Autocatalytic sets of proteins. *J. Theor. Biol.* **119**, 1-24 (1986).

8 Vaidya N., Manapat M. L., Chen I.A., Xulvi-Brunet R., Hayden E. J., et al. Spontaneous network formation among cooperative RNA replicators. *Nature*. **491**, 72-77 (2012).

9 Tjhung K. F., Shokhirev M. N., Horning D. P., Joyce G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. U S A.* **117**, 2906-2913 (2020).

10 Pavlinova P., Lambert C. N., Malaterre C., Nghe P. Abiogenesis through gradual evolution of autocatalysis into template-based replication. *FEBS Lett*. **597**, 344-379 (2023).

11 Vasas V., Fernando C., Santos M., Kauffman S., Szathmáry E. Evolution before genes. *Biol. Direct*. **7**, 10.1186/1745-6150-7-1 (2012).

12 Sczepanski J. T., Joyce G. F. Synthetic evolving systems that implement a user-specified genetic code of arbitrary design. *Chem. Biol.* **19**, 1324-32 (2012).

13 Domingo E., García-Crespo C., Perales C. Historical Perspective on the Discovery of the Quasispecies Concept. *Annu. Rev. Virol*. **8**, 51-72 (2021).

14 Bull J. J., Meyers L. A., Lachmann M. Quasispecies made simple. *PLoS Comput Biol*. **1**, e61; 10.1371/journal.pcbi.0010061 (2005).

406 15 Koonin E. V., Wolf Y. I., Katsnelson M. I. Inevitability of the emergence and persistence of
407 genetic parasites caused by evolutionary instability of parasite-free states. *Biol. Direct.* **12**,
408 31; 10.1186/s13062-017-0202-5 (2017).
409 16 Koonin E. V., Dolja V. V. Virus world as an evolutionary network of viruses and capsidless
410 selfish elements. *Microbiol. Mol. Biol. Rev*. **78**, 278-303 (2014).
411 17 Koonin E. V., Krupovic M., Ishino S., Ishino Y. The replication machinery of LUCA: common
412 origin of DNA replication and transcription. *BMC Biol*. **18**, 61; 10.1186/s12915-020-00800-
413 9 (2020).
414 18 International Committee on Taxonomy of Viruses Executive Committee. The new scope of
415 virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat. Microbiol*. **5**,
416 668-674 (2020).
417 19 de Sousa J. A. M., Fillol-Salom A., Penadés J. R., Rocha E. P. C. Identification and
418 characterization of thousands of bacteriophage satellites across bacteria. *Nucleic Acids*
419 *Res.* **51**, 2759-2777 (2023).
420 20 Eppley J. M., Biller S. J., Luo E., Burger A., DeLong E. F. Marine viral particles reveal an
421 expansive repertoire of phage-parasitizing mobile elements. *Proc. Natl. Acad. Sci. U S A.*
422 **119**, e2212722119; 10.1073/pnas.2212722119 (2022).
423 21 Vignuzzi M., López C. B. Defective viral genomes are key drivers of the virus-host
424 interaction. *Nat. Microbiol*. **4**, 1075-1087 (2019).
425 22 Katzourakis A., Aswad A. The origins of giant viruses, virophages and their relatives in
426 host genomes. *BMC Biol*. **12**, 51; 10.1186/s12915-014-0051-y (2014).
427 23 Conrad B., Iseli C., Pirovino M. Energy-harnessing problem solving of primordial life:
428 Modeling the emergence of catalytic host-nested parasite life cycles. *PLoS One*. **18**,
429 e0281661; 10.1371/journal.pone.0281661 (2023).
430 24 Iranzo J., Puigbò P., Lobkovsky A. E., Wolf Y. I., Koonin E. V. Inevitability of Genetic
431 Parasites. *Genome Biol. Evol*. **8**, 2856-2869 (2016).
432 25 Janzen E., Blanco C., Peng H., Kenchel J., Chen I. A. Promiscuous Ribozymes and Their
433 Proposed Role in Prebiotic Evolution. *Chem. Rev*. **120**, 4879-4897 (2020).
434 26 Tawfik D. S. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and
435 catalytic efficiency? *Curr. Opin. Chem. Biol*. **21**, 73-80 (2014).
436 27 Fitzsimmons W. J., Woods R. J., McCrone J. T., Woodman A., Arnold J. J., et al. A speed-
437 fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol*.
438 **16**, e2006459; 10.1371/journal.pbio.2006459 (2018).
439 28 Genoyer E., López C. B. The Impact of Defective Viruses on Infection and Immunity. *Annu.*
440 *Rev. Virol*. **6**, 547-566 (2019).

441 29 Ibarra-Chávez R., Brady A., Chen J., Penadés J. R., Haag A. F. Phage-inducible
442 chromosomal islands promote genetic variability by blocking phage reproduction and
443 protecting transductants from phage lysis. *PLoS Genet*. **18**, e1010146;
444 10.1371/journal.pgen.1010146 (2022).
445 30 Meng L., Endo H., Blanc-Mathieu R., Chaffron S., Hernández-Velázquez R., et al.
446 Quantitative Assessment of Nucleocytoplasmic Large DNA Virus and Host Interactions
447 Predicted by Co-occurrence Analyses. *mSphere*. **6**, e01298-20; 10.1128/mSphere.01298-
448 20 (2021).
449 31 Yeates J. A. M., Nghe P., Lehman N. Topological and thermodynamic factors that influence
450 the evolution of small networks of catalytic RNA species. *RNA*. **23**, 1088-1096 (2017).
451 32 Koonin E. V. The meaning of biological information. *Philos. Trans. A Math. Phys. Eng. Sci.*
452 **374**, 20150065; 10.1098/rsta.2015.0065 (2016).
453 33 Rousset F., Depardieu F., Miele S., Dowding J., Laval A. L., et al. Phages and their satellites
454 encode hotspots of antiviral systems. *Cell Host Microbe*. **30**, 740-753 (2022).
455 34 Fillol-Salom A., Miguel-Romero L., Marina A., Chen J., Penadés J. R. Beyond the CRISPR-
456 Cas safeguard: PICI-encoded innate immune systems protect bacteria from bacteriophage
457 predation. *Curr. Opin. Microbiol*. **56**, 52-58 (2020).
458 35 Stauffer Y., Marguerat S., Meylan F., Ucla C., Sutkowski N., et al. Interferon-alpha-induced
459 endogenous superantigen. a model linking environment and autoimmunity. *Immunity*.
460 **15**, 591-601 (2001).
461 36 Malfavon-Borja R., Feschotte C. Fighting fire with fire: endogenous retrovirus envelopes
462 as restriction factors. *J. Virol*. 89, 4047-4050 (2015).
463 37 Zhao L., Seth-Pasricha M., Stemate D., Crespo-Bellido A., Gagnon J., et al. Existing Host
464 Range Mutations Constrain Further Emergence of RNA Viruses. *J. Virol.* 93, e01385-18;
465 10.1128/JVI.01385-18 (2019).
466
467
468 **Acknowledgements**
469 We thank Thomas Curran for discussions.
470
471
472 **Author contributions**

Conception and design of the work: MP and BC. Data acquisition MP; data analysis MP, CI; interpretation of data MP, CI, JAC, BC. Creation of new software used in the work: MP, CI. Drafted the work and revised it: BC, MP and JAC.

Data availability

The data, including scripts, is fully made available at the following sites (https://github.com/BICC-UNIL-EPFL/parasite-immunity-from-hyperparasitism [github.com]; https://doi.org/10.5281/zenodo.7970788 [doi.org]).

Competing interests

The authors declare no competing interests.

Figures and Figure Legends

Fig. 1

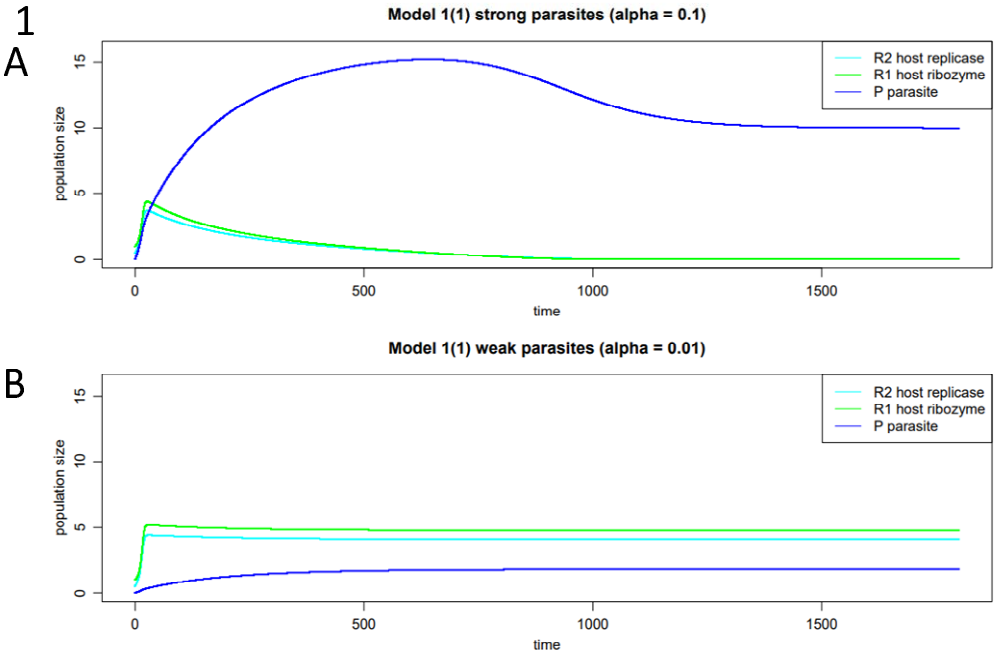

Figure 1. Habitat with one autocatalytic host cycle, variably parasite exposed

Fig. 1A. The catalysts in an autocatalytic cycle (ribozyme R_2, template R_1) exposed to incoming parasites P are promiscuous and hence also amplify P. Under high parasite exposure, parasites will therefore outcompete host molecules.

Fig. 1B. Under weak parasite exposure, the host molecules R_1 and R_2 prevail.

Fig. 2

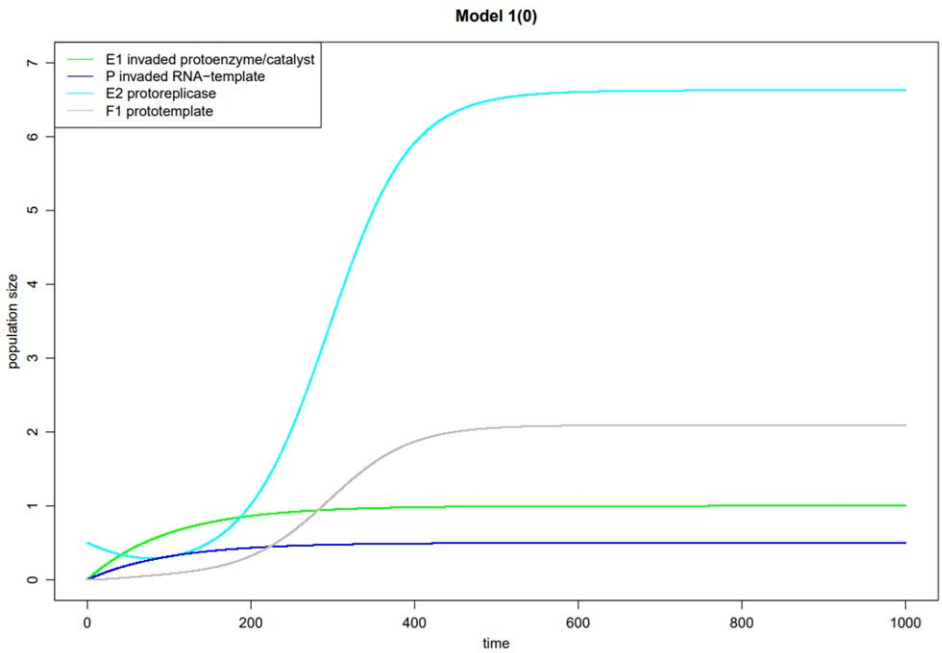

Figure 2. Habitat with primordial autocatalytic cycle

Parasite-specific catalysts with relatively lower catalytic activity in the habitat (E_2 protoreplicase) will confer parasite resistance and thus ensure abundance of habitat molecules in the presence of parasite inflow. Host molecules are expectedly less prevalent compared to Fig. 1B, at least prototemplate F_1.

Fig. 3

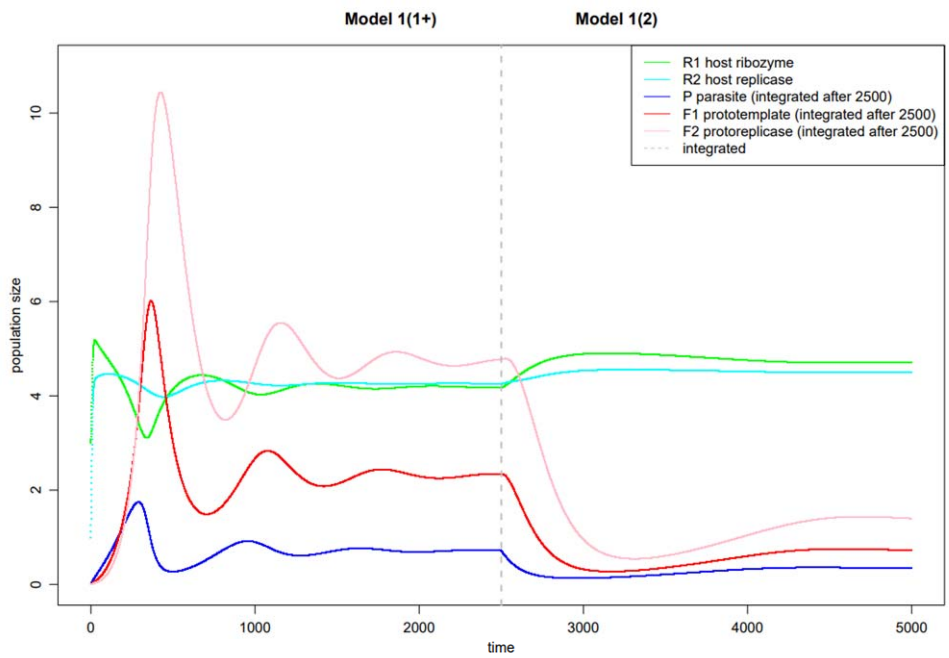

Figure 3. Stabilized host-parasite-hyperparasite cycle with variable parasite inflow

Fig. 3 left panel (Model 1(1+)). A habitat with highly efficient catalysts (self-replicase R_2, template R_1) becomes stabilized with moderate parasite inflow via its relatively less active catalyst-subunit E_2 that confers parasite resistance. F_2 (formerly E_2 in Fig. 2) is generated as well, it acts additively with *R2* and will be important for antiparasitism.

Fig. 3 right panel (Model 1(2)). Under constant, moderate and thus manageable parasite inflow ($P^{special}$, Suppl. Inf.), the hyperparasite cycle mutates and becomes integrated into the habitat, even when parasite inflow ceases (prebiotic heredity and mutation-selection balance). Host molecule prevalence is comparable to a habitat without parasites (Fig. 1B).

Fig. 4A

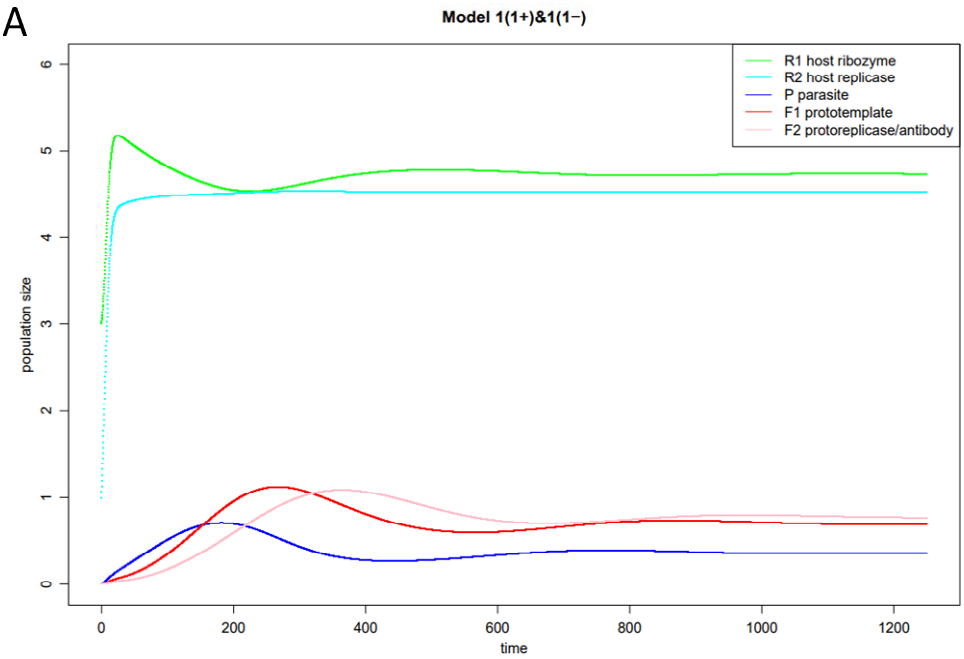

Figure 4. Transition from hyperparasitism to antiparasitism under strong parasite inflow

Fig. 4A Under strong, permanent exposure to complex parasite populations the hyperparasite cycle partially loses its catalytic activity and becomes antiparasitic. The protoreplicase F_2 now continues to function as a partially active catalyst, but also exerts a non-catalytic antiparasitic function (antibody). The parameters in Fig. 4A are set to have these two activities equally represented.

Fig. 4B

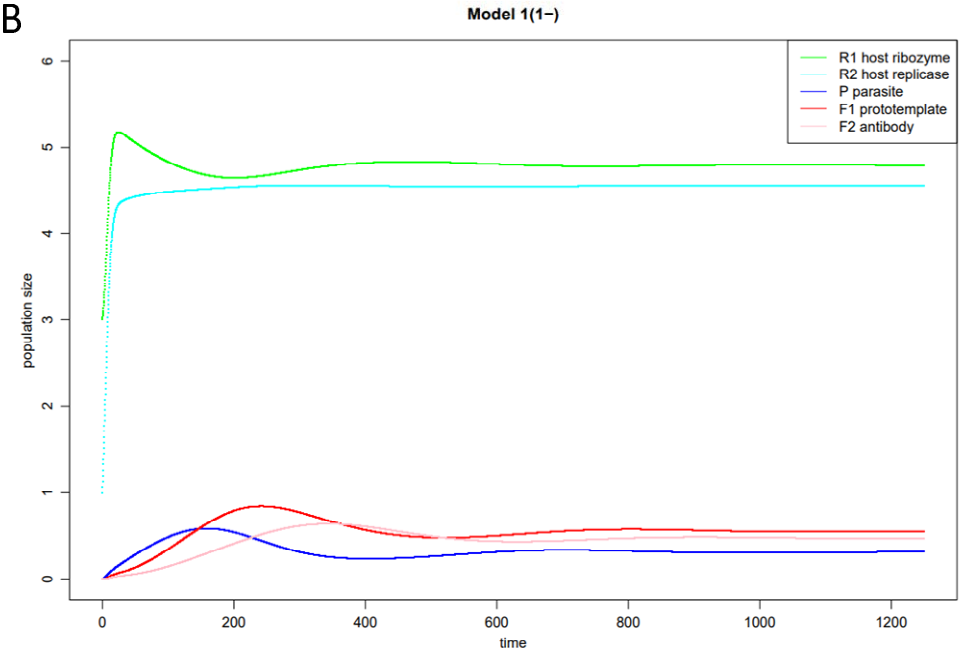

Fig. 4B The parameters are now set to completely shift to antiparasitism (fully degenerated hyperparasite cycle, completely non-catalytic). Antiparasitism has fully replaced hyperparasitism, and the habitat is still stabilized, yet with stronger fluctuations.

Fig. 5

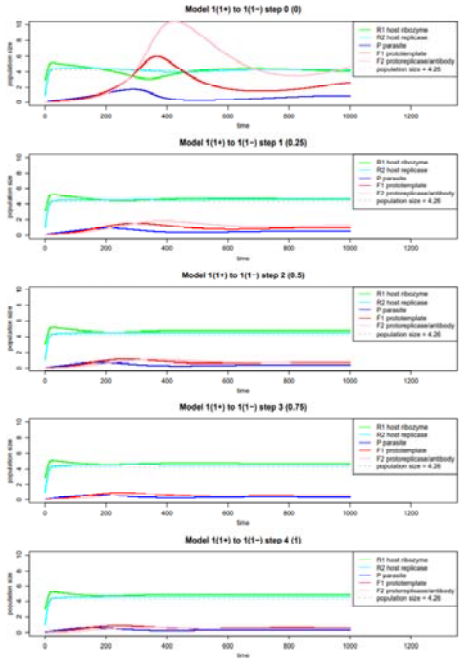

Figure 5. Gradual transition from hyperparasitism to antiparasitism in four steps

The stepwise transition from pure hyperparasitism (top panel) to pure antiparasitism (botton panel) in four steps is shown (0, 0.25. 0.5, 0.75 and 1).