Multilevel Evolution
and the Emergence of Function

Enrico Sandro Colizzi

MULTILEVEL EVOLUTION
AND THE EMERGENCE OF FUNCTION

Multilevel Evolutie
en het Ontstaan van Functies
(met een samenvatting in het Nederlands)

Evoluzione Multilivello
e l'Emergere di Funzioni
(con un compendio in Italiano)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 21 december 2016 des middags te 4.15 uur

door Enrico Sandro Colizzi

geboren op 23 mei 1986 te Monselice, Italië

Promotor: Prof. dr. P. Hogeweg

Preface

Evolved organisms can be characterized by their traits, whose synergy and organisation make up the whole living being. For instance, an elephant has a trunk, tusks, eyes, a tail... We can understand the functional significance of these traits - e.g. what are the eyes of the elephant good for. We can also grasp their adaptive significance, in the sense that elephants with no eyes are worse off than elephants with eyes, and conclude that vision confers a reproductive advantage over blindness. We can, finally, reconstruct the evolutionary history of these traits by tracing how they came about event by event. Nevertheless, we can ask functional, adaptive and historical questions about these traits only after they emerged. In other words, these questions assume the traits, either *a priori*, as is the case for functional and adaptive questions, or *a posteriori* in the case of the historical one. Then, the answers to these questions can only be, respectively, *ad hoc* or *post hoc*.

These questions apply to virtually every evolved trait, given that evolution has been an inventive process for the past four billion years, and many of the answers have been of great value to understand evolution. Yet, a deeper question is: how do novel functions come about?

Let us re-state the problem in more general terms. Individual entities at one level of biological organisation (e.g. the cells) self-organise to generate a higher level of organisation (the organism). The information processed locally (that which orchestrates interactions) is integrated at the higher level. This information is functional at a higher organisational level because novel behaviours (patterns) can be observed and ascribed at that level. As a consequence, the evolution of lower levels cannot be fully understood without considerations of the higher ones (individual cells are neither functional nor evolutionary independent from the whole organism).

The questions of this thesis are: When does information integrated at one level of organisation acquire novel functional significance due to the feedback from emergent, higher-order organisations? And, what are the evolutionary consequences of this?

In this thesis we show that evolutionary inventivity readily arises in simple model sys-

tems.

Because we are studying the evolution of novel functions in an emergent context, it is impossible to preconceive what the novel functions are. We make several computational and mathematical models and endow evolution with sufficiently many degrees of freedom, but we can only recognise results after they appear. However, we keep our models within the boundaries of biological feasibility, so that the principles on which our results stand can be comprehended in depth, and can provide a useful frame of reference for understanding the evolution of living systems.

Contents

1

Introduction

1.1 General considerations

1.1.1 Evolution as a multilevel process

Spatio-temporal complexity of biological systems From single electron manipulations to trans-continental migrations, biological processes span a breathtakingly vast and continuous range of spatial scales. In order to grasp the enormity of this range, we commonly describe living systems as a hierarchical set of levels of organisation. For instance, the collectivity of macromolecular processes and their interactions define the behaviour of a cell, many cells organise into tissues and organs (in multicellular organisms), and so on progressing to ever larger spatial scales, all the way to populations of different species interacting with each other in ecosystems.

This is a simplification: it is far from being universally true (one may find a complete ecosystem under a pebble, which is smaller than the eye that is observing it), and boundaries between levels of organisation are often blurred (e.g. the gut - an organ, deals with the microbiome - an ecosystem).

Because each single level of organisation is itself of staggering complexity (think of gene regulatory networks, ant nests, ecosystems), it is common modelling practice to focus on few elements within a level and study their interplay at the level in which they belong.

This interplay may generate higher levels of organisation, which were not explicitly incorporated in the model, and are therefore emergent. For instance, a group of identical bird-like objects display several features of the dynamics of a flock by introducing simple rules about the interactions of each individual with its neighbours (Hildenbrandt *et al.*, 2010, Reynolds, 1987). Moreover, emergent patterns can be functional because they integrate information at a larger scale than the individual: the slug stage of Dictyostelium discoideum can migrate coherently and with precise directionality by integrating signals that are not observable at the level of the individual cells (Maree *et al.*, 1999). Similarly, a school of fish-like individuals swimming in turbulent waters can trace a food source when a fish alone cannot (Torney *et al.*, 2009). Thus, emergent levels of organisation influence the dynamics of the composing individuals. In "Chance and necessity", Monod wrote (Monod, 1970):

> The elementary interactions upon which everything hinges, these, thanks to their "mechanical" character, are relatively easy to grasp. Much less readily come by is an intuitive global picture of living systems whose phenomenal complexity defies assimilation.

Considerations of multilevel information processing seem necessary to understand the complexity of biological systems.

Evolution The very fact that we can appreciate the complexity of biological systems means that these systems are dynamically stable at some spatio-temporal scale: they are able to preserve and re-generate themselves. A natural question is how does this complexity come about? In other words, how do biological systems behave in time? Palaeontological record (and, recently, experiments) show that changes accumulate along the line of descent of all biological entities. By studying the chronological progression of these changes, e.g. by comparing the genomes of different species, we may recognise patterns and trajectories, and we may wish to construct an explanation about the order of these changes.

Darwin's theory states that biological systems evolve because some heritable changes make individuals more adapted to the environment where they live, which allows them to have more offspring than those who have not changed (or have changed in a way that makes them less adapted). Because the offspring also carries the adaptive change which allows a competitive advantage, the original population will eventually be replaced by the progeny of the original mutated individual (Darwin, 1872).

Darwin did not know as much as we do about biological complexity and, by his own admission, spoke of "chance variation" in the lack of a better explanation on the cause of the differences between parents and offspring. Nevertheless, Darwin's theory works because the directionality in the changes (a particular way in which mutations accumulate) is not a property of the changes themselves, but a property of the effect of these changes

- mediated by the environment - on the number of offspring the mutant has relative to the rest of the population (it is a teleonomy).

There is something atypical in the definition of this theory. First, it says nothing about how or which mutations make an individual fitter. Second, it says nothing about the level at which the mutation has an effect, and what the effect is. Third, despite the two points above, evolution at many levels of organisation can be understood in the light of this theory (whether a mutation speeds up the rate of an enzyme, or it makes the neck of the giraffe longer).

Unlike Darwin, we can glimpse this complexity. We know that most of heritable variability is caused by genetic mutations. Yet, to a large extent we do not know *how* genetic mutations generate variability in a certain trait. This is because we know little about the map between the genotype - the level at which information is stored and mutations occur, and the phenotype - where the effect of these mutations are visible.

Thus, attempting to understand the multilevel complexity of this map can bring about a clearer understanding of evolution.

Multilevel evolution and long-term information integration   After appreciating both the multilevel nature of biological organisation, and the evolution of biological systems by mutation and selection, the next step is to understand the interplay between them. Mutations occur in a genome and alter its information content, but understanding at which scale they produce an effect, if any at all, is far from obvious. Moreover, because different levels of organisation impose different selection pressures, it is often difficult to understand the origin of mutational trends in genomes. The general picture that emerges from the study of multilevel evolutionary dynamics, is that multilevel systems integrate information in the long-term evolution (Hogeweg, 2012). A striking consequence of this is, for instance, that random mutations may have non-random effects, and can be beneficial more frequently than expected (Crombach and Hogeweg, 2008). This happens because the map between genotype and phenotype is itself a product of evolution, and information about the environment can be integrated in the genome during the evolutionary process to shape such map. Another consequence is that ecosystems can process information beyond the capabilities of individuals, as a side effect of self-organised spatial patterns (Hogeweg, 2007). Finally, the methodological advantage of studying evolution as a multilevel process is that much of what happens is not preconceived, nor is the explicit target of evolution, but simply emerges.

We will show in this thesis that multilevel evolution allows for novel functions to arise as well.

1.1.2 The RNA world

Many evolutionary insights described in this Introduction, as well as part of the results in this thesis, hinge on the concept of an RNA world (Gilbert, 1986). The RNA world is a hypothetical stage in evolution that occurred before cellular life and the division of labour between information and function in DNA and proteins.

Although information mostly flows from DNA to RNA to proteins in present-day cells (Crick, 1958), the central tenet of the RNA world hypothesis is that information storage and catalytic functions can co-occur on RNA molecules. RNA stores information in the succession of four nucleotides, much like DNA does, and catalyses chemical reactions by bringing in close proximity different functional groups, i.e. by folding, thus performing protein-like functions (Ward *et al.*, 2014). It is worth of notice that the catalytic core of the ribosome is a ribozyme (an RNA enzyme), and that steady progress has been made to synthesise a general template-dependent polymerase ribozyme (Attwater *et al.*, 2013, Horning and Joyce, 2016). Moreover, RNAs are commonly employed for regulation (Serganov and Nudler, 2013, Sherwood and Henkin, 2016, Wilson and Doudna, 2013) in present day cellular context, and transmembrane RNAs can be synthesised (Janas *et al.*, 2006). Thus RNA is capable of information storage, functional activity and regulation. Finally, because RNAs can be mutated and selected for different functions, they can undergo Darwinian evolution (Joyce, 2007). All this provides strong evidence for the evolutionary independence of RNA from DNA and proteins, and weak evidence for an RNA world.

From a theoretical viewpoint, however, the fact that RNA molecules are capable of evolution makes the RNA world hypothesis an excellent framework in which fundamental evolutionary dynamics can be studied. This is because the RNA world is the period in Earth history in which evolution - as we understand it today - was at its simplest stage. As we will see below, "simplest" does not quite mean "simple" at all, as complex evolutionary dynamics can be observed considering RNAs alone. And yet, compared to the complexity of present-day organisms, RNA world models can only be toy models (The RNA world is the playground of theoretical biologists). But as it comes often with playing, we learnt a great deal about what we were playing with, and certainly more than if we studied the whole "phenomenal complexity" of living organisms.

As a side note, assuming an RNA world sets aside the question of the "chemical" (or kinetic) origin of evolution itself, which is instead being explored in (theoretical) models where adaptation can occur without an explicit information storage medium (Hordijk *et al.*, 2010, Vasas *et al.*, 2012). Moreover, severe problems arise at all stages of RNA monomers and polymers synthesis in (experimental) prebiotic model conditions, even though both novel pathways (Powner *et al.*, 2009) and novel environmental conditions (Mast *et al.*, 2013) that ameliorate these difficulties are being discovered (Higgs and Lehman, 2014). Because of these difficulties, it has been proposed that RNA may not be the original molecule dedicated to information storage and could be a later invention, product of evolution itself (Hud *et al.*, 2013, Robertson and Joyce, 2012).

In conclusion, whether an RNA world actually existed is an important question. Nevertheless, the amount of "evolutionary understanding" the RNA world has brought to us is profound, and therefore it is worthwhile studying it even if it never existed.

1.1.3   This introduction

The rest of this introduction describes previous work relevant for the other chapters. We first review the evolutionary dynamics of information maintenance and integration in genomes when some global target must be reached by the evolutionary process. Then we consider information processing in ecosystems, where different trophic levels co-evolve by being selected on the basis of local information. Next, we focus on the evolutionary effects of self-organisation, i.e. how they integrate the information of their components in models where selection is context dependent. Finally, we introduce and discuss the results of the following chapters, focusing on 1) their contribution to the topics mentioned in this Introduction and 2) their common feature, i.e. the evolution of novel functions as a result of the feedback between levels of organisation.

Notice that although we mention the word "information" repeatedly, we never define it explicitly. Previous attempts at defining information (or complexity) (Adami *et al.*, 2000, Hazen *et al.*, 2007, Lindgren *et al.*, 2004) do not measure ecosystem-based information processing properly, or do not measure genome complexity properly. Despite the lack of a formal definition, we try to keep the discourse pragmatic and bound to the results (without indulging in speculations). After all, "biological information" may well be another one of those biological concepts that escapes definition despite being self-evident, such as gene or species.

1.2   The problem of information integration in evolution

1.2.1   Quasispecies and mutation rates

As mentioned above, Darwin's theory of evolution rests on two notions: 1) mutations generate variability in a population, and 2) the better mutants are adapted to the environment in which they live, the more offspring they make. Therefore, we should expect that fitter individuals arise from less fit populations by mutations and wins the competition in that population, i.e. it is selected, if the environment does not change.

It turns out that this expectations is not unconditionally true, despite its seemingly tautological nature. A very simple model of Darwinian evolution (arguably, the simplest) is sufficient to show this.

Consider a population of RNA-like molecular species that self-replicate and mutate. For instance, we can imagine sequences of nucleotides that autonomously undergo template-dependent polymerisation with a certain degree of accuracy (strictly speaking, RNAs do not easily undergo enzyme-free template-dependent polymerisation, but certain modified nucleotides do (Zhang *et al.*, 2013)) but the mathematical formulation does not depend on these particular assumptions. Each molecular species is characterised by a genotype, in this case a bit string of length ν (one could imagine Purines and Pyrimidines). The genotype determines the rate A at which the species replicate, i.e. there is one-to-one relation in the map between genotypes and phenotypes (the phenotype being replication rate, in this case). Replication involves the possibly inaccurate copying of a genotype: given a per-base accuracy q, perfect replication occurs with rate $Q = q^\nu$. Inaccurate copying of the genotype leads to point mutations - bit flips of the genotype, so that the replication of a species i can result into the generation of another species j. The system is constrained by introducing a term for chemostat ϕ, which keeps the total concentration of molecules constant.

Let us consider the "best case" scenario of a very large and well-mixed population, so that stochastic fluctuations are avoided and global competition ensues. We can write a system of Ordinary Differential Equations (ODE) in which the variable x_i represents the concentration of species i as follows:

$$\dot{x}_i = A_i Q x_i + \sum_j A_j Q_{ij} x_j - \phi x_i \qquad (1.1)$$

where $\phi = \sum_i A_i x_i$ is the mean excess production (obtained by setting the total concentration to one and its derivative to zero). This is the quasispecies equation (Eigen, 1971).

Effectively, the term ϕ introduces selection; this is clearer if we look at the system in the absence of mutations, where Eq. 1.1 becomes

$$\dot{x}_i = (A_i - \phi)x_i \qquad (1.2)$$

and we can see that the chemostat washes out molecules that at any moment replicate slower than the population average (i.e where $A_i < \phi$), to the advantage of those that replicate faster than average (results do not differ if we considered the molar fraction of the species instead of introducing the chemostat (Schuster and Stadler, 2008, Tejero *et al.*, 2010): the fastest growing species "dilutes" all the others even if their growth is unconstrained, due to exponential expansion). Thus, if the system is initialised with a diverse set of species only the fastest replicating ones survive in the absence of mutations when $Q = 1$. Hence, survival of the fittest.

Eq. 1.1 can be rewritten in matrix form as $\dot{X} = \boldsymbol{AQ}X - \phi X$, where X is the vector of concentration $(x_0, x_1, ...)$, $\boldsymbol{A}$ is the vector of replication rates $(A_0, A_1, ...)$ and $\boldsymbol{Q}$

is the mutational matrix that encapsulates the frequency of mutational interconversion among sequences. AQ is a so called fitness landscape, and gives a representation of the interplay of fitness and mutations. Because of the mutational process, the system does not converge to a single species. If all genotypes can be reached by the mutational process, however, evolution is guaranteed to converge to the distribution of species of the dominant eigenvector of AQ (i.e. the one with the largest eigenvalue). Such distribution is a quasispecies.

Given the formalism introduced above, we can now study the dynamics of the fittest replicator. To do this, we simplify Eq. 1.1 by collapsing all but the fittest species of Eq. 1.1 into a single class, effectively leaving two species: x_0 the fittest (or master sequence), and "the rest" x_r. Genotype class x_r contains the vast majority of possible sequences. Because most of these will be mutationally far from the master sequence, we can at first neglect mutations from x_r to x_0 (and drop the subscripts for the quality of replication Q_{ij}). The ODE system becomes:

$$\dot{x}_0 = A_0 Q x_0 - \phi x_0$$
$$\dot{x}_r = A_r x_r + A_0(1 - Q)x_0 - \phi x_r$$

(1.3)

Asking whether the fittest species is selected and survives in an evolutionary context, amounts to find if a rare x_0 mutant can invade a steady state population of x_r (i.e. $\dot{x}_0 > 0$ with x_0, x_r evaluated at $0, \bar{x}_r$). This yields the condition:

$$A_0 Q > A_r$$

(1.4)

which states that fitness (a larger replication rate) is not the sole requirement for survival and replication accuracy must be taken into account as well. This condition is the so called Error Threshold. Crossing this bifurcation point, the system undergoes a phase transition where the fittest genotype is lost due to mutational meltdown. Unpacking the fitness classes from x_r shows that crossing the Error Threshold leads to a steady state distribution where each genotype is present proportionally to its replication rate A_i. Because the number of less fit genotypes is much larger than fit ones (in our case in particular there is only one master sequence class), the concentration of master sequences reduces to an insignificant fraction (and in a discrete population system goes extinct) (Eigen, 1971).

The same condition can be viewed in terms of evolutionary information integration by calculating the longest "informative" genome that can be maintained by selection in the face of the mutational process (assuming that selective advantage and information are directly related). Recalling that $Q = q^\nu$ and defining the relative selective advantage of the master sequence as $\sigma = A_0/A_r$, the Information Threshold is

$$\nu < -\frac{\ln \sigma}{\ln q} \approx \frac{\ln \sigma}{1 - q}.$$

(1.5)

The Information Threshold states that the maximum "amount of information" that can be maintained in a genome depends on the relative fitness advantage it confers as well as, more importantly, on mutation rates. When the mutation rate is beyond the Error Threshold, information is diluted and eventually lost in the system, despite its selective superiority.

The system presented so far allows only for mutations that turn the fit replicator into a less fit one, and we have resorted to an invasion experiment to introduce the fittest genotype class. Including a small back-flow of mutations ϵ, so that $1 - \epsilon \gg Q$:

$$
\begin{aligned}
\dot{x}_0 &= A_0 Q x_0 + A_r \epsilon x_r - \phi x_0 \\
\dot{x}_r &= A_r (1 - \epsilon) x_r + A_0 (1 - Q) x_0 - \phi x_r
\end{aligned}
\tag{1.6}
$$

does not change the biological significance of these results. Even though the bifurcation disappears, x_0 becomes vanishingly small for approximately the same values of Q (the state space does not change dramatically as long as ϵ is not too large).

From a different perspective, the characteristic time for the fittest sequence to appear and reach steady state increases both with Q approaching A_r/A_0 and with $1 - \epsilon$ approaching 1 (the characteristic time is the average time to steady state along an orbit). When Q approaches A_r/A_0 mutations occur frequently, which allows for a faster exploration of the fitness landscape and consequently a faster generation of fitter mutants; however, $Q = A_r/A_0$ is also the value at which fitter mutants have no competitive advantage over less fit sequence classes. In this simplified fitness landscape there is no optimal Q to balance these two effects (Marín *et al.*, 2012). Such optimum exists instead when the general model (with multiple sequence classes) is considered instead (Marín *et al.*, 2012, Stich and Manrubia, 2011), typically at the value where the sequence classes with highest fitness are marginally stable (Van Nimwegen and Crutchfield, 2001).

Much of the theory on optimal exploration overlooks the important factor that changes in sequence length (which depends on mutation rate via the Information Threshold) change the volume of genotype space that can be explored. We will see how this conflict is resolved below. Here, we add two more points. First, the Error Threshold does not appear as a bifurcation but as a first-order (continuous) phase transition, when all sequence classes are considered and a sharp fitness landscape is introduced, i.e. $A_0 > A_i$ and A_i constant and independent of distance from the master sequence; in fact, it may be that no Error Threshold (no phase transition) exists at all, e.g. under assumption of a multiplicative fitness landscape ($A_i = s^{-d}$, where $i > 0$, $s > 1$ and d the mutational distance from the master sequence). Second, both Error Threshold (or lack thereof) and characteristic time should be weighed against finite population sizes. Small population sizes translate to extinction for all fitness landscapes, and long characteristic times must be compared with the time scale of genetic drift: high mutation rates lead to the recurrent extinction of the fittest sequence class as the population delocalises from the fitness peak. Before the Error Threshold, instead, evolution optimises the genotypes in a population, as shown by

the fact that after sufficiently long time the ancestor's distribution is independent of initial conditions and can be almost always traced back to the fittest individual (Hermisson *et al.*, 2002).

Having established that the size of an informative genome is limited by mutation rates, we are left with a dilemma. If a replicator has evolved to the maximum complexity given a certain mutation rate, how can it further increase in complexity? In other words, how can it decrease its mutation rate? To do so, it should evolve a more complex machinery (e.g. proof reading activity), i.e. it should encode more complexity in its genome, which is impossible given that the system is at its maximal complexity. Because replication accuracy was likely weakest at the origin of life, it is quite paradoxical that evolution has achieved the degree of complexity we see today. This is Eigen's paradox (Eigen, 1971), and typifies the problem of integrating information through evolution.

Later in this Introduction we show some classical attempts to solve the problem (and the problems these attempts generate in turn). Moreover, we show how evolution can exploit the very existence of Eigen's Paradox to maintain, integrate and generate (new) information (Chapter 4). A few more concepts must be introduced before delving into it.

The effect of neutrality 1 - Survival of the flattest   So far, we have considered a *steep* fitness landscape, where fitness drops in the close mutational neighbourhood of the master sequence. However, while some mutations drastically decrease fitness, others may have no effect at all on a phenotype, i.e. they are neutral. As a first step to incorporate the effect of neutral mutations, we extend the fitness landscape by including a *flat* region, i.e. a mutational network of selectively indistinguishable species, all with the same growth rate A_1. As we have done for Eq. 1.3, we can now cluster all species into three classes: 1) the fittest sequence, with growth rate A_0, 2) all sequences with growth A_1, which form a single phenotype ξ_1, 3) all others belonging to a lower fitness class ξ_r, with growth rate A_r. The effect of neutrality can be modelled by assuming that a fraction Λ of the mutations occurring to sequences in phenotype class ξ_1 retain the same phenotype. Disregarding back-mutations, we can write:

$$\dot{x}_0 = A_0 Q x_0 - \phi x_0$$
$$\dot{\xi}_1 = A_1 Q \xi_1 + A_1 \Lambda (1 - Q) \xi_1 - \phi \xi_1 \qquad (1.7)$$
$$\dot{\xi}_r = A_r \xi_r + A_0 (1 - Q) x_0 + A_1 (1 - \Lambda)(1 - Q)\xi_1 - \phi \xi_r.$$

The similarity with Eq. 1.3 allows to write immediately the condition for the survival of ξ_1 in the absence of x_0 as

$$A_1 (Q + \Lambda(1 - Q)) > A_r. \qquad (1.8)$$

Comparing the survival conditions for x_0 and ξ_1 (Eq. 1.4 and Eq. 1.8) reveals that a population evolving on this landscape at high mutation rates may distribute on the flat region

even if growth rate there is lower than in the steep part of the landscape, i.e. if $A_1 < A_0$. This is because a single genotype x_0 must be maintained against mutations in a steep region of the fitness landscape, whereas a distribution of mutationally interconverting equireplicative genotypes ξ_0 are maintained in a flat part of the fitness landscape. Therefore, the beneficial effect of neutral mutations can overcompensate for a lower growth rate. This highlights that the target of selection is the quasispecies with the largest average growth rate, and the fitness of a single genotype is important insofar it is reachable by the mutational process (Schuster and Swetina, 1988). This effect was dubbed "Survival of the flattest" (Wilke *et al.*, 2001).

The effect of neutrality 2 - Phenotypic Error Threshold   We have seen that a quasispecies evolving on a flat fitness landscape tolerates better the deleterious effect of mutations (this is the reason for outcompeting a quasispecies on a steep landscape). Should we then expect a beneficial effect on the Error Threshold? The answer turns out to be (partially) negative: the effect of neutral mutations is limited close to the Error Threshold for all reasonable degrees of neutrality.

This can be shown directly from Eq. 1.7 by disregarding the steep part of the fitness landscape. We remain with only two fitness classes (as in Eq. 1.3): ξ_0 is the master phenotype and ξ_r the collection of all other less fit ones. As above, a proportion Λ of the mutations occurring to genotypes in the class ξ_0 generate another member of the same class and we exclude back-mutations that turn genotype of phenotype class ξ_r into ξ_0. With these assumptions, we can write (Takeuchi *et al.*, 2005, Van Nimwegen *et al.*, 1999)

$$\dot{\xi}_0 = A_0 Q_0 \xi_0 + A_0 \Lambda (1 - Q_0) \xi_0 - \phi \xi_0$$
$$\dot{\xi}_r = A_r \xi_r + A_0 (1 - \Lambda)(1 - Q_0) \xi_0 - \phi \xi_r, \tag{1.9}$$

and the condition for survival of ξ_0 is obviously Eq. 1.8.

The assumption of a simple, low-dimensional fitness landscape (with peaks, plateaus and valleys) used so far can be misleading in the context of large neutrality and high mutation rates. This is because it does not provide any intuition about how to model multiple mutations occurring in a single replication event, with many of them leading to viable sequences. To characterise Phenotypic Error and Information Threshold we must therefore model a neutral network of mutationally connected equireplicative sequences in a little more details. To do so, we can assume an additive effect of mutations and that the per-base probability of neutral mutation is λ, uniform over a genome (so that neutral mutations are binomially distributed). Then, the per base effective replication accuracy is $Q_e = (q + (1 - q)\lambda)^\nu$, with ν the size of the genome. The Phenotypic Error Threshold can be derived from $A_0 Q_e > A_r$, and the relative Information Threshold is

$$\nu < -\frac{\ln \sigma}{\ln(q + (1 - q)\lambda)} \approx \frac{\ln \sigma}{(1 - q)(1 - \lambda)}. \tag{1.10}$$

Eq. 1.10 states that although a population evolving on a neutral network can maintain

a larger genome than one evolving on a steep region of a fitness landscape (the extra term $1 - \lambda$, comparing Eq. 1.5 and Eq. 1.10), the gain in genome size is still severely limited by mutations. Importantly, maintaining a larger genome size despite frequent mutations does not entail a larger information content. On the contrary, we can only expect genomes to be able to accommodate some randomness. Moreover, the benefit of neutrality is limited by the actual value of λ (which rarely exceeds 0.5, as obtained by counting the number of neutral substitutions RNA molecules tolerate, see next section).

An experimental "enzymatic activity landscape" can be constructed by mutagenic analysis of present-day ribozymes (Kun *et al.*, 2005). Because many close mutants of these ribozymes show little to none activity, the resulting Error Threshold can be much larger than what is expected from Eq. 1.4. Surprisingly, combining this observation with the (experimentally determined) value of λ (via Eq. 1.10) indicates that a hypothetical ribo-organism could sustain a functionally rich and diverse genome.

The effect of neutrality 3 - Neutral evolution of robustness   In order to show that mutational robustness can evolve to the detriment of replication rate (hence information content), we considered the evolutionary dynamics on a fitness landscape with one flat region and one peak, separated by a "deleterious" valley. We can take this one step further and include a neutral network. Essentially, we consider that several somewhat flat regions exist in the fitness landscape. Importantly, if we consider only two fitness classes (as above), the probability of deleterious mutations also varies locally in the fitness landscape.

Much like before, we expect evolution to maximise the average growth rate of a population. Here, however, larger average growth rate will be determined entirely by the shape of the neutral network (because replication rate is constant), and we expect to find an evolving population on the region of the neutral network that has larger than average connectivity (i.e. a larger Λ).

That is to say, evolution neutrally maximises mutational robustness (Van Nimwegen *et al.*, 1999). This argument, derived with infinitely large populations, applies to finite populations as well provided that mutation rate and population sizes are both large enough to sustain sufficient variability, so to avoid that an entire population converges on a single genotype (in which case evolution would approximate a diffusive process, and the population distribution would not reflect the heterogeneity in the neutral network) (Van Nimwegen *et al.*, 1999).

At the beginning of the section we have asked whether evolutionary optimisation is able to generate the fittest individual. In hindsight, a more precise (a better nuanced) question could have been whether evolution can integrate an arbitrary amount of information. Despite the answer turns out to be negative, we have not done this for nothing. Allowing one to better formulate questions (to see a problem in a clearer light) seems to

us an important purpose of theoretical models. Furthermore, we have seen that populations, and not single genomes, are the target of evolution, and that the particular fitness landscape on which a population lives is of great importance in determining the evolutionary outcome of the system. Moreover, by considering neutrality (λ) rather than steepness or flatness we have began to move away from simple fitness landscapes. In the following paragraph we caution against other shortcomings of this concept.

Some remarks on fitness landscapes In this section, we have extensively used the metaphor of fitness landscapes (Wright, 1932), because it is intuitive (almost pictorial) and simplifies much of the description of evolutionary processes. There is danger in the abuse of this notion, however.

First, we tend to think in few dimensions, whereas biological landscapes have enormous dimensionality: if a genome of length ν is a point in genotype space, the dimension of the space is ν, and the combination of all possible nucleotides that generate genomes of such length is $4^\nu \approx 10^{0.6\nu}$. Thus a fitness landscape would assign fitness to each of these points, as well as connecting them by mutations. Neglecting the high dimensionality of a fitness landscape can lead to results that are difficult to defend on biological ground (e.g. (Wagner and Krall, 1993), and (Takeuchi and Hogeweg, 2007a) for its resolution).

A second problem is that we considered static fitness landscapes. This implies that genomes are of constant length, and that mutations preserve genome size (we have considered substitutions). Clearly, biological genomes are not subject to either of these constraints, which alter the dimensionality of the landscape and affect its connectivity, thereby altering the effect mutations themselves have on genomes. In Chapter 5 we show that larger (and biased) mutation rates can have a beneficial effect on mutational load (hence adaptation) in a cellular context, when genome size and content evolve in response to a co-evolving transcriptional load.

Finally, the evolutionary dynamics on fitness landscapes considered so far amount to a mere competition between independent individuals with different growth rates. Although this improves mathematical tractability and ease of interpretation, it may be an excessive simplification. For instance, population-dependent selection pressures and environmental variations (including the "internal environment" of an organism) can change the height of the fitness peaks and the depths of fitness valleys during the process of adaptation. In Chapter 4 we show that a population evolves to a steep region of (a biologically more sound) fitness landscape as a side effect of the interactions between replicators.

One may conclude that a fitness sea-scape of variable dimensionality would make a better image for understanding evolutionary dynamics. Except that, instead, the image has not become clearer at all, suggesting that we drop the metaphor altogether in favour of greater biological realism.

This notwithstanding, the insights of quasispecies models are useful in the research on origin of life (Kun *et al.*, 2005, Takeuchi and Hogeweg, 2012) and virus evolution,

(Domingo *et al.*, 1985, Más *et al.*, 2010) and can be extended by incorporating molecular details of polymerisation (Rajamani *et al.*, 2010) or other biological processes (Boerlijst *et al.*, 1996, de Aguiar *et al.*, 2015, Takeuchi and Hogeweg, 2007b). Building on the isomorphism between the original quasispecies model (Eigen, 1971) and a spin glass model in Statistical Mechanics (Leuthäusser, 1986), an interesting connection can be made between quasispecies dynamics and quantum mechanical phenomena (Baake *et al.*, 1997, Bianconi and Rahmede, 2011). This connection may make possible that tools used in a highly mathematical field such as quantum mechanics be used in understanding more complex evolutionary dynamics (or the other way around).

1.2.2 Individual based solutions to evolutionary information integration

Coding and Genotype to phenotype maps In its simplest interpretation, the Information Threshold poses a limit on the amount of complexity a single genotype can evolve. In the quasispecies models we presented, however, there was no complexity at all to be incorporated in genomes: replicators did not encode anything because mutations affected replication rate directly. This one-to-one map between genomes and their fitness is different from that of biological systems, where the order of nucleotides in a genome (a genotype) is informative only in the light of the translation machinery, which makes this information functional (a phenotype). Conceptually, we seek for a map that unambiguously determines a phenotype given a genotype, and in which the genotype is informative only in the context of the map itself, the rules for which are also in the genome and are therefore evolvable (a symbolic information processing system (Maynard Smith, 2000)). As we have mentioned before, the functional similarity of RNA with common cellular information processing systems, as well as the theoretical simplicity of conceiving an RNA world, makes RNA a perfect candidate to study biological evolutionary dynamics in these terms.

The most common approach is to consider the minimum free energy secondary structure as the phenotype of an RNA sequence (Fig. 1.1). The *a priori* reason for this choice is that fast folding algorithms exist, e.g. (Lorenz *et al.*, 2011) (unlike for proteins), and that the map is both redundant and frustrated.[1] Although the tertiary structure is ultimately necessary for the functional properties of RNAs, it is often disregarded because its folding is directed by secondary structure elements (and because no fast 3D folding algorithms exist). RNA sequences fold into a stable secondary structure by forming internal hydrogen bonds with both Watson-Crick base pairs (A-U and C-G) and wobble pairs (most importantly U-G). Two subsequences with complementary nucleotides form stem, with free energies depending on the length of the matching strands and on the strength of the matching pairs of nucleotides (C-G bonds are stronger than A-U ones, which in turn are stronger than G-U). However, it is often that case that a subsequence can form a match

---

[1]"Frustration" occurs because several competing interactions generate a rugged energy landscape with many globally different low-energy structures.

with multiple others. Frustration in RNA folding arises because not all possible inter-actions can be realised. The folding algorithm minimises free energy on a smaller scale first by assigning energies to all the possible stacks of nucleotides, and then globally by selecting the collection of non-overlapping stacks that minimise the total energy (exclud-ing pseudo-knots) (Hofacker *et al.*, 1994). On the one hand, far apart subsequences can be brought close to each other in the folding process, but few mutations can disrupt the fold (due to frustration) and lead to a very different one. Hence, the landscape is rugged. On the other hand, given a stack of matching nucleotides, several pairs in the stack can be changed without altering the fold. In other words, several sequences fold in the same secondary structure. Thus RNA folding is also redundant.
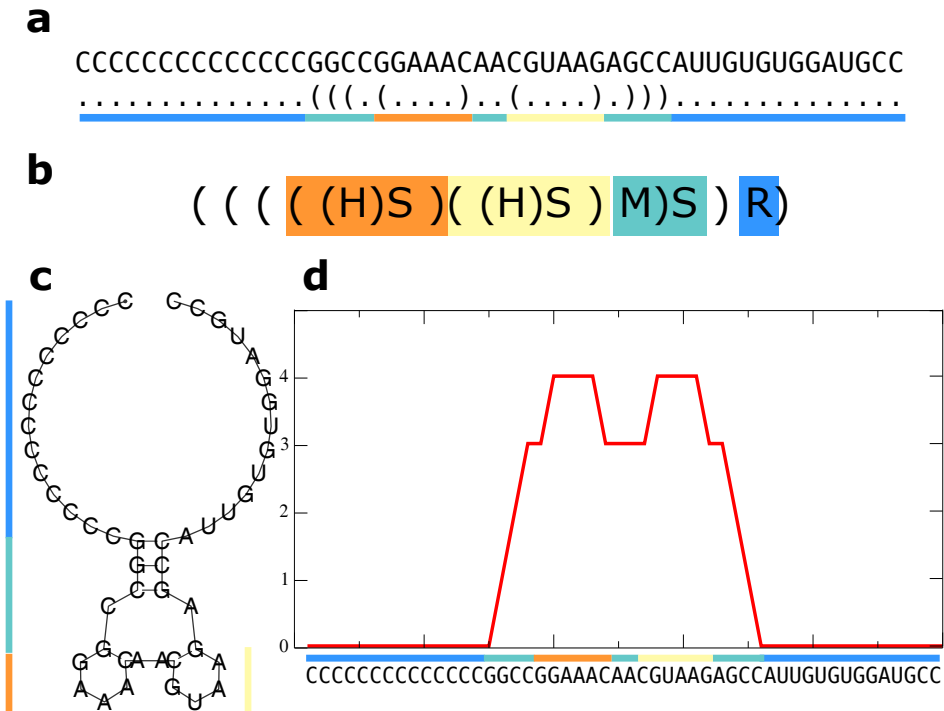
Figure 1.1. Different representations of the minimum free energy folding of an RNA se-quence. **a**: Sequence and "dots and brackets" representation, **b**: coarse grained "Shapiro" representation (H: hairpin loop; S: stem; M: multi-loop connector; R: unfolded remainder of the sequence), **c**: graphical representation, **d**: mountain plot representation. Corresponding substructures in different representations are indicated with the same colour. Sequence folded with Vienna Package (Hofacker *et al.*, 1994).

Ruggedness and redundancy interplay in a peculiar way, and lead to several surprising properties of this map:

- Extensive random sampling of the genotype space reveals that most sequences fold

> into one of few common secondary structures (Grüner *et al.*, 1996). The longer the sequences, however, the more structures become available (although the rate at which the number of structures increases with sequence length depends on the exact level at which secondary structure is coarse grained).

- Given a sequence that folds into a common structure, few mutations (much fewer than sequence length) are sufficient to find most other common structures (Schuster *et al.*, 1994).

- Yet, it is possible to traverse the entire genotype space by mutating one or two nucleotides at a time (Schuster *et al.*, 1994).

- Novel structures can be found at constant rate in the mutational neighbourhood along a path in the neutral network of a structure, no matter the distance travelled. On the same path, structures similar to the one of interest are repeatedly rediscovered (phenotypes bring their own shadow along in a neutral walk) (Huynen, 1996).

Evolutionary dynamics with complex genotype-to-phenotype maps   The introduction of genotype-to-phenotype maps shifts the problem of evolutionary optimisation from attaining a target genotype to a target phenotype. To this end, we can follow the evolutionary dynamics of a population of RNA-like replicators, selected on the bases of their similarity with a predefined target secondary structure (Huynen *et al.*, 1996). Starting from a random sequence with low fitness, the evolution of phenotypes proceeds by long periods of stasis where no phenotypic differences are visible, interrupted by sudden fitness "jumps", where one or more phenotypic mutations make the population fitter. In contrast, mutations are accumulated in sequences somewhat more continuously.

In the long periods of phenotypic stasis, sequences explore the part of the genotype space connected by neutral mutations. As seen above, neutral exploration does not sample the network homogeneously, but favours the regions that are more connected. Fitness jumps, i.e. beneficial mutations, typically occur as transitions between different neutral networks rather than through crossing a fitness valley (Van Nimwegen and Crutchfield, 2000) (the latter process is called stochastic tunnelling in simplified fitness landscapes (Iwasa *et al.*, 2004, Komarova *et al.*, 2003)).

After a beneficial mutation occurred, selection concentrates the population on the fitter genotype, which is likely not located on a highly connected part of the new neutral network. Then, neutral exploration can restart until the target structure is found.

A different outcome of the evolution of RNA-like replicators can be obtained by considering all the stable structures within a certain energy range, rather than only the minimum free energy fold. Assuming that a single RNA lives long enough to switch between these conformations (and energy barriers between them do not affect the switching), the

proportion of time spent in each fold is equal to the probability of the configuration (i.e. its energy E_i, via Boltzmann relation $\exp(-E_i/kT)/Z$, where Z is the partition function, i.e. the normalising sum of the probability of each configuration). Then, overall fitness can be calculated as an average of the fitness contributions of different structures (Ancel and Fontana, 2000).

Despite the fact that a larger portion of the phenotype space can be "seen" at any moment, a population of replicators evolving with this criterion does not converge to the target phenotype and gets trapped in a local optimum. This happens because at any point selection minimises the time spent in less favourable conformations and increases the time spent in the most favourable ones. Because of plastogenetic congruence (the similarity of phenotypes between the energetic neighbourhood and the mutational neighbourhood), selection effectively increases mutational robustness to the point that adaptive steps (which are accompanied by a large reduction of robustness) become selected against.

Altogether, evolution reaches a dead end when selection acts on the energy neighbourhood, because of implicit selection for robustness (due to plastogenetic congruence). In contrast, selection for minimum free energy structure leads to mutational robustness as an evolutionary side effect. We hypothesise that this leads to a more stable energy neighbourhood due to plastogenetic congruence.

Multiple coding and the evolution of coding structure   In the previous paragraphs we showed that replicators heavily exploit the complexity of RNA folding as a genotype-to-phenotype map to integrate information. This map is similar to that of complex biological systems because it allows for a large freedom of coding the information for a phenotype. However, it differs from biological systems in one important feature, namely the ability to evolve itself. For example, the genetic code in present day organisms is arbitrary (in the sense that there is no *a priori* reason for a certain codon to translate to a certain aminoacid), has undergone a certain degree of evolutionary optimisation (Freeland *et al.*, 2000, Woese, 1965), while being still evolvable (Isaacs *et al.*, 2011). Moreover multiple open reading frames can overlap (Huvet and Stumpf, 2014) (and references therein, also see (Hogeweg and Hesper, 1992) for an evolutionary model) and some genes can be spliced in different ways to obtain functionally distinct proteins (Barash *et al.*, 2010). Finally, proteins can fold in multiple functionally different states depending on environmental signals (Tokuriki and Tawfik, 2009).

These examples point out that biological genotype-to-phenotype maps are versatile: they change the "meaning" of the information stored in the genome depending on context (as emphasised in (Trifonov, 2011)), as well as modify the map itself (a form of evolvability).

Examples of RNAs that are active in different conformations abound: (Marek *et al.*, 2011, Schultes and Bartel, 2000, Serganov and Nudler, 2013). Hence, the results showed in the previous section do not exhaust the evolutionary potential of RNAs.

In the previous paragraph we have included the observation that RNAs can fold into different structures, all contributing to the same fitness target. This needs not be so: different phenotypes can be achieved by one sequence as suboptimal folds (within a certain energy range from the minimum free energy structure). We consider that a single replicator consists of an ensemble of sequences evolving towards a set of target phenotypes (de Boer and Hogeweg, 2014). A population of such replicators -we can imagine RNAs enclosed in membrane, and call them protocells- could use different sequences for different target structures.

A general observation is that several sequences code for more than one phenotype at evolutionary steady state. Moreover, increasing mutation rates increases the frequency of multiple coding. This is because higher mutation rates decrease the total genome size that can be maintained against mutations, which forces the evolution of "smarter" alternatives, in this case multiple coding. As a result high fitness is maintained at high mutation rates.

The occurrence of multiple coding under high mutation rates reveals a conceptual difference between the models discussed before and the last one. In the former cases, large mutation rates improves the rate of evolutionary exploration but threaten the maintenance of the information contained in the fittest sequence. Neutral networks can buffer against the deleterious effect of mutations to some extent, by coding for more robust solutions. In this last case, however, mutation rates should be though of as having a novel role: they tune which "coding regime" the system adopts. Clearly, this is possible because the genotype-to-phenotype map itself can evolve.

Evolution of symbolic information   The previous experiment shows that the ability of RNA to store information is limited by what the genotype-to-phenotype map can do with it, rather than by mutation rates. This result can be taken one step further by a simple modification of the same model. We assume that RNAs can fold into different structures by binding adapters (short RNA sequences) that mask the complementary residues (de Boer and Hogeweg, 2012). Because sequences and adapters co-evolve, protocells evolve at high mutation rates to code for multiple functions by having few long sequences whose folding is guided by several short adapters.

This system shares several similarities with the genetic code:

- it is arbitrary and can modify itself, i.e. it is evolvable,

- it makes information stored in a genotype explicit in a context-dependent manner,

- the machinery that "translates" this code is also encoded in the genome.

Hence the evolution of a symbolic information processing system.

In conclusion, we have transitioned from Quasispecies models evolving on simple fitness landscapes, passing by the introduction of RNA as a genotype-to-phenotype map, to models where replicators have larger freedom to chose the coding structure for solving the evolutionary problem of information integration. At the same time, our focus has shifted from the idea of an Error Threshold, where mutations pose a limit to information, to evolvable maps in which mutation rate is a degree of freedom (a "tunable" parameter) for code structuring.

In Chapter 4 we show that multiple coding arises naturally in an evolutionary system even though it is not selected for, as an ecological strategy against competing replicators and parasites.

Alternatives to RNA folding as genotype-to-phenotype map In the previous section we made large use of RNA folding as a genotype to phenotype map. We chose RNA folding because it is chemically well-grounded, fast, (fairly) accurate and of clear Biological significance, and we concluded that evolution exploits heavily its properties (e.g. its peculiar combination of ruggedness and redundancy). Despite its apparent specificity, many of the concepts developed by adopting such specific genotype-to-phenotype map generalise to biological systems. For instance, evidence of the synergy between mutational robustness and evolvability has been found in Eukaryotic gene regulatory networks (Payne and Wagner, 2014) (the network of regulators, such as transcription factors, that controls gene expression) as well as metabolic networks in bacteria (Samal *et al.*, 2011, Wagner *et al.*, 2014). In the latter case, the large modularity in metabolic network can be understood as a form of neutrality in metabolic space, and seems to be a by-product of the selection pressure for metabolising multiple carbon sources (Samal *et al.*, 2011). Furthermore, experimental addition of edges to gene networks in bacteria are often neutral, and occasionally cause mutants to outgrow the wild-type (Isalan *et al.*, 2008), indicating that bacteria are both robust and evolvable.

From a theoretical view point, RNA has proved useful for clarifying the mechanisms that evolution adopts to increase (or maintain) fitness. However, while more complex models may not allow such deep characterisation of the processes studied, they can show novel concepts and novel evolutionary mechanisms. [2] Cell-based models, for instance, have been used to clarify the interplay among genome length, evolvability and selection pressures on coding and non-coding genome (Knibbe *et al.*, 2007), to explain the long term effects on fitness of genome expansion (Cuypers and Hogeweg, 2012, 2014), to show the relation between adaptation to target metabolites and exaptation to metabolise other compounds (Barve and Wagner, 2013), as well as a proof of principle that an evolutionary process can increase its own capacity to evolve, the so called evolution of evolvability (Crombach and Hogeweg, 2008, Draghi and Wagner, 2008).

---

[2]Somewhat humorously, we could say that evolution has not departed from the RNA world for nothing!

In the next section, we introduce a particular genotype-to-phenotype map to explicitly address information integration and processing in evolving systems.

1.3 Ecosystem-based information processing and spatial pattern formation

In the previous section we considered the evolution of a single population at a time, and we analysed how information was integrated in genomes. However, we neglected the functional properties of the interactions between individuals because of the global fitness criterion we introduced. In biological systems, evolving entities are never confronted with a full problem (e.g. there is no selection for a full, functional eye before its simpler components are evolved). Models can reflect this when fitness is based on interactions and information becomes context dependent. However, once we have this in mind, the question "how can a system integrate information over evolutionary time scales" becomes less clear, because now we may not always have an explicit search image any more, and we have to recognise what emerges from the system. This is by no means a weakness of the following models. On the contrary, this allows to move from the "artificial" dynamics of a system as imposed by an arbitrary fitness function, to its "natural", i.e. the non-predetermined, properties (Hogeweg, 1993), while still using the concepts developed in the previous section for a comparative search image.

1.3.1 Eco-evolutionary dynamics of simple predator-prey models

A possible way to introduce a "local" fitness criterion consists of considering the fitness of an individual as dependent on both itself and those it is interacting with. A clear biological example of this are trophic interactions, where the fitness of a predator depends on the specific preys it eats.

A simple implementation of this idea consists of modelling the co-evolving interactions between two trophic levels, predators and preys.

Predators and prey are both characterised by a (one dimensional) trait value. As predators feed on preys with similar values, predators' fitness increases when a one dimensional continuous trait-value approximate the corresponding value of a prey and, vice-versa, the prey gains by not being matched by the predator (Van Der Laan and Hogeweg, 1995). A simple expectation for the evolutionary dynamics of this system is that predators' trait chases that of preys forever, making evolutionary stability impossible (a form of Red Queen dynamics (Van Valen, 1973)). In contrast, we observe that the two species stabilise (and Red Queen dynamics are avoided) because mutations dampen the oscillatory ecological dynamics of the species, thus stabilising them. This model shows that the

19

interplay between ecological interactions and population diversity is important for the evolutionary stability of an ecosystem. But what about information?

Analogously to the previous section, we can introduce a genotype-to-phenotype map, rather than dealing with phenotypic parameters. Let predators and prey have genotype and phenotype, with the map between them determined by RNA folding. Predators' fitness depends on the similarity of their secondary structure with the prey, whereas the prey survives if its secondary structure is sufficiently different from that of the predator (Huynen, 1993). Unsurprisingly, Red Queen Dynamics ensues, whereby preys endlessly mutate away from predators and predators mutate closer to preys. The long-term evolutionary effect of the short-term selection for diversity is that the prey is driven to a more rugged region of the phenotype space, where mutations have the largest effects. This effect is stronger when predator's mutation rate is larger (notice that prey's mutation rates remain constant in these experiments).

While in the previous paragraph we have shown that introducing mutations affects the dynamics of an ecosystem, here we have shown that introducing ecological dynamics change the way organisms evolve.

1.3.2 Spatial pattern formation and ecosystem-based evolutionary info integration

Having seen that interactions between populations does affect the evolutionary capacity of a species, we next ask how an ecosystem integrates information. This question can be tackled explicitly by evolving a population in order to solve a certain problem by exposing individuals only to parts of the problem itself. Such *sparse* fitness evaluation also suggests that problems and problem-solvers should co-localise. Thus, we can study both the well-mixed system and the spatially extended one (embedded in a two dimensional grid). Here we focus on the spatial system (see (de Boer, 2012) for the complete exposition).

With an ecosystem in mind we consider three co-evolving trophic levels: preys, predators scavengers (de Boer and Hogeweg, 2010). The problem itself consists of approximating a target function through evolution (within a predefined domain and up to some precision). Preys are numerical instances of the problem, while predators and scavengers are evolvable solving functions (represented in LISP notation as a tree of operations). Predators and scavengers' fitness depends on producing the correct value of the target function for the problem instance provided by the prey (competition between species is excluded by fixing eating order). Preys are more successful the worse predators perform.

Predators evolve full solutions at low mutation rates by generalising based on the small

subset of problems encountered by the lineage. After the first solution is found, long-term evolution shortens the coding length so that individuals are less affected by mutations. The initial expansion of the genome is functional because it makes individuals more evolvable (cfr. (Cuypers and Hogeweg, 2012) where initial genome inflation makes virtual cells more evolvable in the long evolutionary term). Increasing mutation rates limits the genome size that can be maintained by evolution, thereby limiting the total volume of genome space that can be explored.

Individual-based solutions are never found above a threshold mutation rate. While an individual predator alone cannot reach a global solution, sub-populations can specialise on different parts of the problem, speciation occurs, and two lineages co-exist. Because a large part of the problem is left unsolved at any moment, scavengers can specialise as well, speciating to complement predators' solutions. The alternation of the two pairs of predators and scavengers in the spatially embedded system generates a pattern of mutually invading waves (de Boer and Hogeweg, 2010). Due to spatial pattern formation, cooperation between predators and scavengers the problem can be fully solved.

In the previous section, we showed a system evolving towards a global fitness target selects the most suitable coding structure given the constraints imposed by mutation rates, e.g. multiple coding evolved at high mutation rates. Here we see that an ecosystem evolves a suitable coding structure to solve a problem depending on mutation rates. Division of labour within and between species arises at high mutation rates. In other words, ecosystems process and integrate information when individuals cannot. In Chapter 4 we present a different coding structure that allows large information integration and processing at high mutation rates.

1.3.3 The problem of cooperation at the origin of life

In the previous section we studied the evolution of information integration by confronting replicators with local sub-problems. This localised (and sparse) fitness evaluation allowed for co-evolution of different species and the emergence of coding structure. However, the overall target fitness was still pre-defined.

In this section we present models where no fitness target is predefined (there is no evolutionary problem), and fitness depends solely on interactions with other individuals. A common theme will be that interactions allow for self-organisation in spatial models, which often result in recognisable spatial patterns. As we will see, spatial pattern formation constitute a novel level of selection that can modify the outcome of evolutionary dynamics in a counterintuitive way.

Hypercycles and the evolution of parasites   The earliest attempt to solve Eigen's Paradox in this terms (in fact, the earliest proposed solution) (Eigen and Schuster, 1978) stems

from the idea that an ecosystem of several short sequences may be able to maintain information at high mutation rates where a single sequence cannot. Because every master sequence must win the competition within its own quasispecies, competitive exclusion applies also between quasispecies. In other words, because different quasispecies cannot coexist independently, an interaction structure must be introduced for all sequences to survive. Assuming that different sequences are mutationally distant and do not interconvert by few mutations, and that mutation rate is sufficiently low to guarantee the survival of each sequence, we can neglect mutations entirely and write (in analogy with Eq. 1.1) the kinetic equations for the concentrations of the species:

$$\dot{x}_i = \sum_j A_{ji} x_i x_j - \phi x_i \tag{1.11}$$

where the quadratic term $A_{ji} x_i x_j$ introduces an interaction network between species pairs (and $\phi = \sum_j A_{ji} x_i x_j$). To visualise this, we can imagine that a molecular species i specifically catalyses the replication of the species j, even though the mathematical formulation glosses over the details of enzymatic polymerisation (and hypercyclic structures may be common in present day ecosystems (Maynard Smith and Szathmary, 1995)). A hypercycle is a cyclic interaction structure where replicator i catalyses replicator $i + 1$ (the last replicator n replicates the first 0). This is the simplest arrangement that maintains all species, whereas any strictly non-cyclic arrangement quickly faces extinction (for an extensive discussion which includes first order terms as well see (Stadler and Stadler, 2003, Stadler, 1991)). Due to second order interactions between different species, an unbounded hypercycle grows hyperbolically (whereas quasispecies grows exponentially). Although infinite growth (in finite time!) is not reasonable, its consequence is that the competition between different hypercycles always favours the most abundant, i.e. the one that established first (this effect is called "once-forever" selection (Eigen and Schuster, 1977), see also (Szathmáry, 1991)). Steady state dynamics are stable for small hypercycles (less than 4 members), and oscillatory (on a limit cycle) for larger ones.

Although conceptually appealing, the hypercycle is both ecologically and evolutionary unstable. Ecological instability is due to large oscillations in the concentrations of species when hypercycles have many members, which take each species close to extinction every cycle. Evolutionary instability is due to mutations that make replicators better target for replication. These selfish mutants are selected because of their faster growth and, symmetrically, altruistic mutations that increase the replicating efficiency to the next member in the cycle are counterselected (Bresch et al., 1980). At the extreme end of the spectrum of selfish mutations are those that maximise the probability of being a target of replication and minimise those of replicating others. These parasitic mutants interrupt the cycle altogether, and lead to the extinction of the system.

Spatial pattern formation and its evolutionary consequences Both ecological and evolutionary instability are a consequence of assuming a stirred system without popu-

lation structure. We can construct the same model as Eq. 1.11 in a discrete, spatially extended setting. To do so, replication of a sequence of type i occurs if in its neighbourhood there is 1) a sequence of type $i-1$, i.e. its replicator and 2) empty space in which the product of replication is placed. A chemostat assumption is unnecessary in this case, as the finiteness of space prohibits unbounded growth. A probability of death is introduced for empty space to be regenerated, so that the system does not to get stuck in its initial configuration.

In this system, the cyclic succession of species in a hypercycle with more than four members organise into a concentric succession of rotating spirals (Boerlijst and Hogeweg, 1991b). The overall ecological stability is greatly ameliorated because several spirals can establish at the same time (each with its dynamics), and the total population sizes do not oscillate widely. The spatial system display several novel eco-evolutionary properties:

- It is resistant to parasites. Parasite invasion is spatially limited to one arm of the spiral (their target), but cannot spread to all target replicators in the system because successive spiral arms of the same type are spatially disconnected. As the spirals are generated from the core, the local invasion of parasites is of little consequence for the rest of the spiral. If parasites happen to reach the core of a spiral, instead, they are likely to destroy the whole spiral. Yet, as long as other spirals are present, the overall stability of the system is not compromised.

- Competition between hypercycles, i.e. selection, is automatically re-introduced. If some variability is present in the system, self-organised spirals can be formed from different original sets of replicators. Competition occurs at the level of spirals: the one that rotates faster erodes the space occupied by the other, until the core is reached and the slower spiral goes extinct (Boerlijst and Hogeweg, 1991b). A counterintuitive consequence of the competition between spirals is, for instance, selection for early death. Larger death rates increase the rate at which a type of replicators is replaced by the next, thus making rotation faster. In this sense, spirals "enslave" the evolution of its composing elements.

- Automatic reproductive differentiation between core and periphery. Due to the sequential nature of replication, replicators in the arms of a spiral move outward from the core generation after generation (independently from diffusion). The replicators that remain in the core, instead, re-generate the arms of the spiral. In fact, tracing the ancestor of all replicators shows that all individuals descend from the core of the spirals. That is, once spirals have formed, the long term fitness of replicators depends only on their location.

Despite its surprising properties, the spatially self-organised hypercycle does not solve Eigen's paradox. Shortcutting mutants that reduce the number of species needed to complete one cycle are selected because they make spirals rotate faster, i.e. a cycle is completed in shorter time. Hence, hypercycle cannot integrate information in evolution, and

is actually prone to losing it. Moreover, hypercycles do not multiply because the centre of the spiral cannot escape, thus variability does not accumulate.

Nevertheless, it does highlight the importance of self-organisation for the evolutionary dynamics of the system. In particular, the feedback of spatial patterns on the evolutionary dynamics of its composing elements is not a peculiarity of the hypercycle, and is instead much more general (see below and e.g. (Savill *et al.*, 1997, van Ballegooijen and Boerlijst, 2004)).

1.3.4 Spatial cooperation and travelling waves

The hypercycle model shows that cooperation - and its relation to parasites - is an evolutionary problem since the RNA world. In this context, the simplest view on the problem is to consider a single species of replicators which catalyses its own replication, i.e. it acts both as replicase and as template. For instance, assume that a replicator x_1 replicates others with rate k_1, and that template behaviour is controlled by a parameter β_1. We can ask whether a mutant x_2 (with replication and template behaviour controlled by, respectively, k_2 and β_2) can invade when its template behaviour is different, i.e. $\beta_1 \neq \beta_2$. The dynamics of a well-mixed system read:

$$
\begin{aligned}
\dot{x}_1 &= k_1\beta_1 x_1^2 + k_2\beta_1 x_1 x_2 - \phi x_1 \\
\dot{x}_2 &= k_2\beta_2 x_2^2 + k_1\beta_2 x_1 x_2 - \phi x_2,
\end{aligned}
\tag{1.12}
$$

where ϕ is the total growth rate, as before. The invasion condition for x_2 is $\beta_2 > \beta_1$, that is: a mutant can invade when it is a better template than the original replicator.[3] This results mirrors the well-known problem of cooperation in sociobiology: the lack of population structure in a cooperative system leads to selection for "selfish" replicators.

Clearly, parasites that do not give replication and are better templates than replicators, should be able to invade too. To see the effect of parasites we must first abandon the chemostat, as it forces the total population size to 1 even when almost no replication occurs in the system. This awkward effect can be avoided by assuming that replication rates are scaled by the availability of some resources $\theta = 1 - x_1 - x_2$. Because parasites do not replicate other species, we can set $k_2 = 0$ and re-name variables as $\kappa_1 = k_1\beta_1$ and $\kappa_2 = k_1\beta_2$ (in the absence of a chemostat that washes away excess production, we must also introduce a decay term d for all species). The new system reads (Takeuchi and Hogeweg, 2012):

$$
\begin{aligned}
\dot{x}_1 &= \kappa_1\theta x_1^2 - dx_1 \\
\dot{x}_2 &= \kappa_2\theta x_1 x_2 - dx_2.
\end{aligned}
\tag{1.13}
$$

Parasites outcompete replicators when $\kappa_1 < \kappa_2$, i.e. when the former are better template than the latter. As a consequence of this, replicators go extinct and parasites quickly
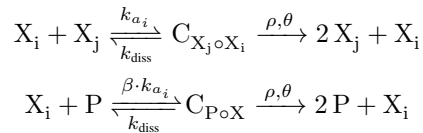
---

[3]Results do not change assuming that replicators spend a fraction of time τ in active (catalytic) state, and the rest of the time in a state of template, so that $k = \tau$ and $\beta = 1 - \tau$.

follow, as they are unable of self-maintenance.

Spatial selection for cooperation   As in the case of the spatially extended version of the hypercycle, a discrete, spatially extended model of the dynamics between cooperative replicators can reverse the selective trend to selfishness. When replicators are embedded on a surface, local replication clusters similar replicators, which together can grow faster than (cluster of) mutants (Szabó *et al.*, 2002). Moreover, limited dispersion does not permit selfish molecules to spread efficiently and outcompete the cooperative replicators. The larger the rate at which molecules diffuse (approaching a well mixed system), the more likely it becomes that selfish mutants end up in the centre of a cluster of cooperative replicators, destroying it.

However, there is a caveat for these results to hold, namely that cooperation cannot be too costly. In this system, this is (presumably) achieved by 1) the possibility of an increase in replicase activity along with with template quality and 2) assuming that replicators do not spend much time replicating others.

Complex formation strengthens selection for parasites   With trans-acting replicators, template dependent polymerisation takes a certain amount of time, in which the replicator itself is not being replicated. Parasites do not suffer from this problem instead, as they spend their entire life-time being replicated. As a result, replication is more costly for replicators than we have assumed so far. This can be incorporated in the replicator-parasite system presented above by introducing complex formation as follows:

$$\text{X}_\text{i} + \text{X}_\text{j} \xrightleftharpoons[k_\text{diss}]{k_{a_i}} \text{C}_{\text{X}_\text{j} \circ \text{X}_\text{i}} \xrightarrow{\rho, \theta} 2\,\text{X}_\text{j} + \text{X}_\text{i}$$

$$\text{X}_\text{i} + \text{P} \xrightleftharpoons[k_\text{diss}]{\beta \cdot k_{a_i}} \text{C}_{\text{P} \circ \text{X}} \xrightarrow{\rho, \theta} 2\,\text{P} + \text{X}_\text{i}$$

where X_i and X_i are replicators that form complexes with rate k_a, P is parasite and has an advantage $\beta > 1$ at complex formation, C stands for complex (with subscripts indicating the composing members; complexes dissociate with rate k_diss), which generates a novel replicator or parasite with rate ρ if resources θ are present.

In a well mixed system, this strengthens the selection for selfishness and, consequently, for parasites (Takeuchi and Hogeweg, 2007b). In a spatial system as well, increased costs of replication overturn the emergent selection for active replicators. The spatial system does not go extinct, however. Replicators survive at the minimum rate of complex formation even though mutations that further decrease replicators' activity are still selected. Survival is possible because of the heterogeneity in the system: local extinction ensues when replicators cannot sustain themselves, but the remaining empty space can be re-colonised by neighbouring replicators (see Chapter 2 and (Takeuchi *et al.*, 2011)).

Despite this, in Chapter 3 we will see that cooperation in a replicator system can evolve also when costs are large. The evolutionary mechanism for this depends on the emergent evolutionary properties of self-organised spatial patterns, which we present in the next section.

Travelling waves dynamics and evolution   Parasites do not lead to global extinction in a spatially extended system, contrary to the results obtained from Eq. 1.13. Instead, replicators and parasites coexist for a large range of parameters (provided that $\kappa_2 > \kappa_1$). Coexistence is mediated by the self-organised arrangement of the two populations into travelling waves, as shown in Fig. 1.2. The origin of these spatial patterns is quite simple. A small population of replicators inoculated on a surface grows by expanding in all directions. Introducing parasites next to replicators prohibits the expansion of the latter in that direction. Parasites, in turn, do not grow autonomously and must be in close contact with replicators. Limited dispersal hinders parasites because they locally cause the extinction of replicators (due to superior template quality). When replicators are locally extinct, only those parasites that are adjacent to other replicators can replicate further. Hence a travelling wave, where replicators - the wave front - expand in the direction opposite to parasites, and parasites - the back of the wave - replicate in the direction of replicators and leave empty space behind.
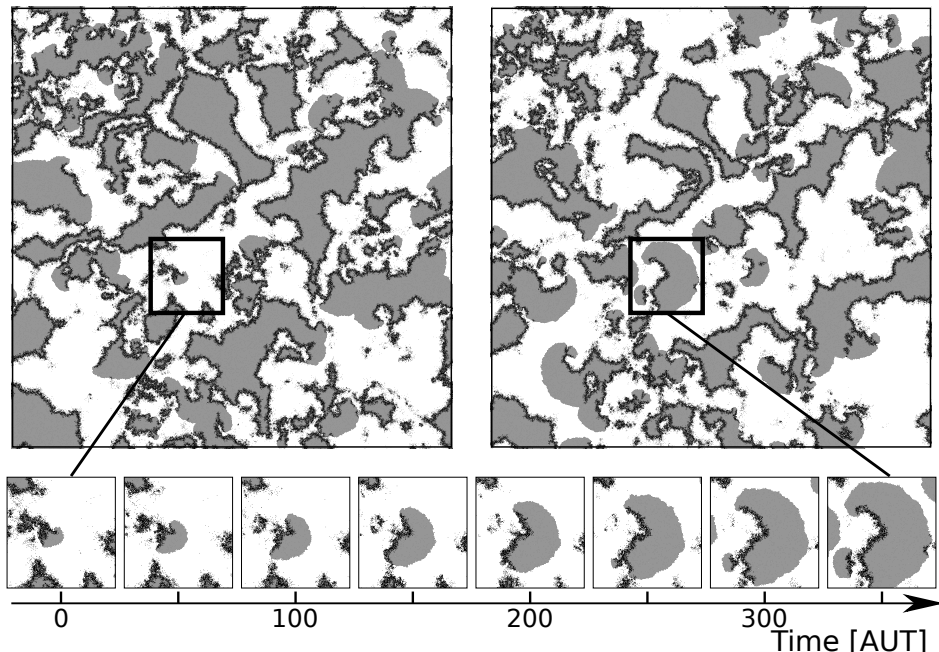

Figure 1.2. Travelling waves formed by replicators (grey) and parasites (black). The two snapshots are taken from the simulations described in Chapter 2 at 300 time steps (Arbitrary Units of Time) from each other. Below, the "birth" and "development" of a wave.

Multiple waves occur simultaneously on a sufficiently large surface. This leads to frequent collisions between waves, which in turn can result in the partial or complete annihilation of the two waves, as well to the generation of new waves. Additionally, new waves can establish when replicators creep through the back of a pre-existing wave if parasites are sufficiently weak. Thus waves are capable of multiplication (unlike hypercycles).

In a well-mixed system parasites are selected to become stronger when their affinity for replicators is free to evolve (Takeuchi and Hogeweg, 2009). In the spatial system, instead, weaker parasites permit replicators to escape from the back of the waves more frequently, thereby establishing new waves. Waves with weaker parasites tend to be more turbulent and cannot be invaded by waves with stronger parasites, which are more stable. Thus wave-level selection acts on the evolution of parasites to make waves more "fertile". Hence, selection at the level of waves reverses the individual-based selection pressure for stronger parasites (waves enslave their components, similarly to hypercycles). Moreover, new waves inherit the parasites of the waves they come from. Thus, waves display compositional inheritance.

Altogether, waves display variability and inheritance, and can multiply as well as respond to selection pressures. In other words, we can describe the evolution of travelling waves in terms of mutation and selection *independently* from the evolution of its composing elements. In this sense, waves are units of evolution (Takeuchi and Hogeweg, 2009).

In Chapter 2 we extend these results by studying the effect of wave-level evolution on replicators, and on the the co-evolution of replicators and parasites.

Replicators with sequence, structure and interactions   The replicator-parasite model discussed so far is clearly a simplified description of an RNA world. This simplicity allows to pin-point the effects of the interplay between evolution and self-organisation. However, we have also seen that complex genotype-to-phenotype maps deeply affect the evolutionary dynamics of replicators.

The two approaches can be combined straightforwardly by considering trans-acting replicators with genotype-to-phenotype map determined by RNA folding, in a spatially extended system (Takeuchi and Hogeweg, 2008). Replicators form complexes with a probability that depends on the degree of complementarity between the unfolded 5' and 3' end subsequences. Replication occurs if a replicator folds into a predefined secondary structure and binds with its 5' dangling end another sequence (by the 3' end). The complementary sequence is generated upon replication.

The long-term evolutionary stability of the system depends on mutation rates in a counterintuitive way. A single quasispecies evolves and persists at the highest viable mutation rates, the sequence variability in the system is minimal and a clear master sequence can be found. The quasispecies maximises the probability of complex formation by increasing the frequency of C and G nucleotides in their 5' and 3' end (exploiting the larger

free energy of C-G bonds over A-U bonds). The reason for the stability of a compact quasispecies hinges on the interplay between coding structure and spatial process, and is the object of Chapter 4 (see below).

Decreasing mutation rates, parasites readily evolve and organise with replicators in travelling waves. Parasites exploit sequence homogeneity and large association rates of the replicators' quasispecies. Even though parasites can evolve by a few mutations, it takes them several more to optimise their affinity to the replicators. Thus parasites form a separate lineage which co-exists with replicators.

Further decreasing mutation rates, a new species of replicator evolve which specialises on A and U nucleotides to mediate complex formation. The new quasispecies is selected because it is not affect by the parasites, due to its different sequence specificity. A new parasite can evolve which exploits the second quasispecies, if mutation rate is decreased once more.

Altogether, the model shows how the interplay between genotype, phenotype and spatial pattern formation generates an eco-evolutionary process of niche-creation and speciation.

Replicators in compartments   Spatial pattern formation imposes a new level of selection that stabilise the dynamics of replicators and parasites. For the sake of completeness (although outside the scope of this thesis), we briefly mention that parasites can also be stabilised by explicit compartmentalisation. If compartments host replicators and parasites in small numbers, and assuming that both are needed for optimal growth rate, an overall stable steady state can be achieved because of the stochastic repartition of the two species upon division (Szathmáry and Demeter, 1987). Moreover, protocell-level dynamics impose a higher level of selection, which affects the evolution of parasites (Takeuchi and Hogeweg, 2009). Finally, complex evolutionary dynamics can be observed also when protocells contain only replicators, as a result of the conflicts between levels of selection. For instance, the conflict between individual-based selection for selfishness (which lowers the rate of complex formation), and selection for protocell growth (which relies on replicators with larger growth rates) can cause evolutionary cycling in different lineages of replicators, even though the overall population seems to have reached an evolutionary stable steady state (Takeuchi *et al.*, 2016).

1.4   This thesis: Multilevel evolution of novel functions

In this thesis we extend and connect the concepts presented above by studying the behaviour of several evolutionary models. In the next paragraphs we introduce the following chapters by focusing on the emergence of functions in a multilevel evolutionary context.

1.4.1 Replicator-parasite systems

In Chapter 2 we extend the analysis on the spatial evolutionary dynamics of replicator-parasite systems by 1) fixing parasite strength and letting replicators evolve and 2) allowing the two species to co-evolve. We complement the results presented in the previous section [REF TH09] from the "perspective" of replicators' evolution, and we characterise the phase space of the system in terms of wave-level properties. We also find a novel evolutionary regime in which wave-level dynamics allow replicators to become more cooperative when parasites are stronger. Finally, we explore the co-evolutionary dynamics of replicators and parasites by tuning the duration of replication (i.e. the amount of time a replicator spends before completing replication). We show that spatial pattern formation directs the co-evolutionary trajectories of replicators and parasites, which in turn can reverse the wave-level evolutionary dynamics.

A novel function for parasites   The increase of association rates in replicators depends (once again) on wave-level dynamics. Selection pressure for chaotic waves (fecundity) is reduced when strong parasites do not allow replicators to escape frequently from the back of a wave. Because of this, empty space is more abundant behind parasites, and waves are more stable and live longer. The abundance of uninterrupted empty space sets the conditions for replicators to increase association rate: on the wave front, the replicators the invade space first (by increasing association rate) are those whose progeny survives, the others remain behind and are outcompeted by parasites. Stronger replicators further strengthen parasites, establishing a feedback mechanism that selects for the increase of cooperation. Hence, strong parasite are beneficial to the evolution of replicators, because they set the spatial conditions in which replicators become more cooperative.

1.4.2 The evolution of cooperation at high costs

Increasing the timespan of replication increases its cost further for a replicator, because it steepens the trade-off between replicating others and being replicated. As a side effect, even weak parasites are likely to kill all replicators in a wave, which in the long run selects for parasites that are worse templates than replicators and may go extinct (Chapter 2). Treating replicators and parasites as different species is reasonable if one assumes that a long evolutionary history is necessary for parasites to be optimised templates (evidence for this is presented in the next section). The principle on which replicators increase their cooperative traits, however, does not depend strictly on this separation. Because replicators act *in trans*, longer time-spans for replication strengthens selection for selfishness, and parasite-like replicators evolve. Selfish and cooperative replicators then self-organise in travelling waves, which re-inforce both selfish and cooperative replicators, with the two populations forming separate lineages. Hence, selfish and cooperative behaviour coexist, co-evolve and enhance each other by means of travelling waves dynamics.

To show this point more clearly, in Chapter 3 we reformulate the system in the classical terms of sociobiology, i.e. in terms of costs and benefits of cooperation. Drawing from microbiology, we model a population in which individuals can evolve the rate at which they produce a diffusible substance, i.e. a public good. An individual benefits from the total public good it comes in contacts with, but pays a cost when producing it. Thus public good production is a cooperative trait. In a well-mixed system, selfish mutants that produce less public good can always invade a resident population of cooperative individuals as long as benefits are larger then costs (the dynamics analogous to Eq. 1.12 and the calculation is straightforward). This is according to intuition: when the benefits of producing costly public good are shared, any selfish mutant the produces less than average will experience a direct competitive advantage, because a slight decrease in the shared benefits is overcompensated by a large decrease in the individuals' production cost.

When a population of cooperators is embedded in a grid (or a network), instead, results can be different because groups of cooperators have collectively an advantage over selfish individuals (we have seen an example of this above). This form of spatial selection in "viscous populations" holds as long as costs are sufficiently low (or benefits sufficiently high) (Van Baalen and Rand, 1998, Wilson $et\ al.$, 1992). As a simple heuristic rule, if b is the benefit per cooperator in a neighbourhood, and c is the cost, individuals can invade in a grid (or a network) of connectivity k as long as $b > ck$ (Ohtsuki $et\ al.$, 2006). Indeed, we observe that public good production evolves to large values in the entire population for low costs, but decreases to the minimum viable when costs approach b/k.

What happens when costs are increased further? Strikingly, public good production evolves to larger values again. The initial selection for lower public good production leads to the evolution of selfish individuals which rely on cooperators for survival, thereby forming travelling waves. The invasion dynamics of travelling waves increases public good production on the wave front (Van Dyken $et\ al.$, 2013), and provides support for selfish individuals, which are further selected to decrease their production rate. Thus, cooperation evolves at high costs due to the mutual feedback between evolution and self-organisation.

This result enriches the point about emergence of novel functions due to multi-level evolution in two ways: 1) The results obtained with the RP system assumed replicators and parasites as separate lineages; here we have shown that parasites are not a necessary assumption, but automatically evolve from cooperators; 2) The fact that the same principle can be applied in both the replicator-parasite system and the public good model highlights its generality (the public good model is an extremely simple and general model for the evolution of cooperation).

1.4.3 Mutational division of labour in evolved quasispecies

In Chapter 4 we return to the RNA-like replicator system, and we study the effect RNA folding as a genotype-to-phenotype map on the eco-evolutionary dynamics of the system at high mutation rate. We have mentioned before that a single quasispecies evolves in this parameter regime and that the system displays little variability both sequence-wise and spatially (for instance, no parasitic lineage evolves). We show that a master sequence evolves to a steep region of the fitness landscape (very few of its mutants are neutral) with a stable distribution of closely related mutants descending from it. The master sequence exploits an evolved coding structure that prevents both a survival of the flattest effect and exploitation by selfish replicators and parasites while being the genotype that replicates the fastest. The analysis of the mutational neighbourhood of the master sequence allows the identification of novel functions as phenotypes of non-viable sequences (that do not replicate). Because mutations occur frequently, the unusual mutational neighbourhood of the master sequence translates into a spatial distribution of these novel phenotypes. This stabilises the master sequence and prevents the invasion of competitors (with "worse" mutational neighbourhoods) as well as parasites.

In the previous section we have seen that a coding structure can evolve at high mutation rates so that a single sequence codes for multiple phenotypes. These individual-based solutions to the problem of evolutionary information integration were obtained by selecting for a global fitness target. However, we explicitly included the capacity to evolve the genotype-to-phenotype map itself in the system (by suboptimal foldings or adapters), because the properties of single-sequence, minimum free energy folding are immutable. Here this effect is purely emergent, and comes about because evolution exploits the fact that mutations occur "reliably". Moreover, we have seen that population-based (or ecosystem-based) solutions can integrate and process information at high mutation rates (when individuals cannot), provided that the fitness function is local. Here, there is no explicit fitness function (fitness is based on interactions), and yet we observe that information processing is mediated by the properties of the emergent functional ecosystem.

In conclusion, there needs not be a rigid distinction between individual-based and population-based solutions to evolutionary problems, in the sense that evolution is not bound to one of them if it is given enough degrees of freedom.

1.4.4 Evolution of mutational biases

The approach we have taken so far allowed us to identify the emergence of innovation in (relatively) simple multilevel model system. In Chapter 5 we take a different approach: we focus on an experimental observation that exemplifies such emergence in a cellular context and we build a model to identify the evolutionary and functional significance of this behaviour.

It was found that S. cerevisiae can increase its ribosomal DNA copy number in response to nutrient availability (Jack *et al.*, 2015) (ribosomal DNA is organised in a cluster of about 150 to 180 copies of the ribosomal RNA gene on chromosome XII). A larger number of rDNA genes does not lead to growth differences (French *et al.*, 2003), thus the increase in copy number is not directly under selection. It is known, however, that damage accumulates quickly in yeast's DNA if there are less than about 30 copies of the gene (see e.g. (Ide *et al.*, 2010)).

We hypothesise that the evolution of these "regulated mutations" is a form of evolution of evolution (Hindré *et al.*, 2012). A typical hallmark of the evolution of evolvability is the long-term evolutionary integration of environmental information to modify the genotype-to-phenotype map, which can be "bent" so that random mutations have non-random effects (Hogeweg, 2012). In the case of regulated mutations in the rDNA gene, the evolved genotype-to-phenotype map biases mutations by means of an intricate inter-play among the signalling pathway, transcription regulation and DNA replication. The outcome is that non-homologous recombination leads to frequent duplications of rDNA genes when rDNA copy number is low.

In Chapter 5 we begin the exploration of this system by studying the evolution of a popu-lation of cells with (a phenomenological) metabolism, regulation and a beads-on-a-string model of genes in a genome. Mutations can be caused by large transcriptional load (see (Helmrich *et al.*, 2013) for review) as well as random mistakes during replication, and cause gene duplications, deletions or inactivations. We find that the the size of the cod-ing genome is severely limited by the ratio between different mutations. At evolutionary steady state, the amount of non coding genes is large, and cells have low overall fitness when the three types of mutations are equiprobable. In contrast, when we increase the overall mutation rate but we bias mutations towards duplications and deletions, evolu-tion reaches a much fitter steady state, characterised by a large coding part, and very few inactive genes. Furthermore, the relative frequency of transcription-induced muta-tions evolves to increase duplications and deletions, and reduce inactivating mutations. We also find that cells' growth rate is hardly diminished when we artificially remove a considerable fraction of rDNA gene copies from evolved cells, as is the case in yeast.

Altogether, our results suggest that there exists a selection pressure in yeast to bias tran-scriptional mutations towards duplications and deletions, which results, as a side effect, in a large copy number of ribosomal genes.

2

Parasites Sustain and Enhance RNA-like Replicators Through Spatial Self-Organisation

ENRICO SANDRO COLIZZI, PAULIEN HOGEWEG (2016)

Abstract

In a prebiotic RNA world, parasitic behaviour may be favoured because template dependent replication happens *in trans*, thus being altruistic. Spatially extended systems are known to reduce harmful effects of parasites.

Here we present a spatial system to show that evolution of replication is (indirectly) enhanced by strong parasites, and we characterise the phase transition that leads to this mode of evolution. Building on the insights of this analysis, we identify two scenarios, namely periodic disruptions and longer replication time-span, in which speciation occurs and an evolved parasite-like lineage enables the evolutionary increase of replication rates in replicators. Finally, we show that parasites co-evolving with replicators are selected to become weaker, i.e. worse templates for replication when the duration of replication is increased.

We conclude that parasites may not be considered a problem for evolution in a prebiotic system, but a degree of freedom that can be exploited by evolution to enhance the evolvability of replicators, by means of emergent levels of selection.

2.1 Author Summary

The RNA world is a stage of evolution that preceded cellular life. In this world, RNA molecules would both replicate other RNAs and behave as templates for replication. A known evolutionary problem of this world is that selection should favour parasitic templates that do not replicate others, because they would be replicated the most. A possible solution to this problem comes from spatial self-organisation: local accumulation of parasites lead to their own local extinction, which leaves empty space for replicators to invade. We show that the spatial organisation generated by interacting replicators and parasites sets the (spatial) conditions that enhance replicase activity when parasites are stronger. Moreover, we find that the co-evolution of replicators and parasites is severely constrained by the type of spatial patterns they form, and we explore this feedback between evolution and self-organisation. We conclude that spatial self-organisation may have played a prominent role in prebiotic evolution.

2.2 Introduction

According to the RNA world hypothesis, RNA self-replication preceded the current DNA-RNA-protein replication mechanism (Gilbert, 1986). RNA molecules can store information much like DNA as well as catalyse reactions (Guerrier-Takada *et al.*, 1983, Kruger *et al.*, 1982, Nissen *et al.*, 2000), including self-replication (Draper *et al.*, 2008, Paul and Joyce, 2002), and are capable of undergoing Darwinian evolution (Lincoln and Joyce, 2009, Martin *et al.*, 2015, Mills *et al.*, 1967). From a theoretical view point, one of the simplest evolutionary systems consists of ribozymes that perform template dependent polymerization, even though such ribozymes are not fully functional yet experimentally (Attwater *et al.*, 2013, Wochner *et al.*, 2011).

Replication *in trans* requires a catalytic molecule to bind and copy a template, and is thus prone to exploitation by molecules that behave more often as templates than as catalysts. At the extreme end of the spectrum lie parasites: RNAs that never replicate others and may be better templates than replicators (Bresch *et al.*, 1980, Maynard Smith, 1979). A well-mixed pre-biotic soup is indeed evolutionarily unstable because selection for better templates progresses until replicators cannot sustain themselves. Clearly, the evolutionary instability of this system is aggravated by parasitism. Some form of higher-level organisation is therefore necessary for the persistence of self-replicating RNAs.
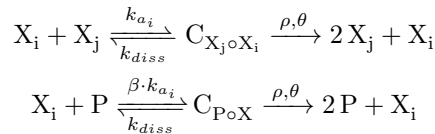
We focus on the emergent levels of selection introduced by spatial self-organisation (a viable alternative is explicit compartmentalisation, such as vesicles (Hogeweg and Takeuchi, 2003, Niesert *et al.*, 1981, Szathmáry and Demeter, 1987)). The problem of parasites is alleviated in spatially extended systems due to spatial pattern formation (Boerlijst and

Hogeweg, 1991c, Hogeweg and Takeuchi, 2003, Könnyű *et al.*, 2008, Szabó *et al.*, 2002, Takeuchi *et al.*, 2011), because parasites automatically segregate from replicators due to limited diffusion, and their accumulation leads to local, but not global, extinction (see also Chapter 3). Furthermore, spatial patterns can constrain the evolutionary dynamics of parasites and select for weaker ones (Könnyű *et al.*, 2008, Takeuchi and Hogeweg, 2009, van Ballegooijen and Boerlijst, 2004).

Here we show that parasites may not be a "problem" for the evolution of RNA-like self-replication in spatially extended system, but are actually beneficial, in that they sustain and enhance replication through higher levels of selection. This effect was briefly mentioned in (Takeuchi *et al.*, 2011). Here we characterise it in terms of how the strength of parasites affects the selection pressures generated by spatial patterns. Building on this, we identify a phase transition-like behaviour, where the selection regime changes abruptly and we analyse the selection pressures at the replicators' levels. Then, we explore the co-evolutionary dynamics of replicators and parasites.

2.3 Methods

With an RNA-world in mind, we model the dynamics of a population of replicators and parasites in an individual based, spatially extended, stochastic simulation system. Individuals are located on a two-dimensional square lattice with eight neighbours and wrapped boundaries (based on CASH, (de Boer and Staritsky, 2000)). Each node of the lattice can be empty or occupied by one individual. Replicators form complexes with other replicators at rate k_a, and with parasites at rate $k_a \cdot \beta$ in order to replicate them (i.e. the replicator behaves as replicase). Assuming that replicators behave as templates with rate set to one, $\beta \geq 1$ represents the relative advantage a parasite experiences as a template over replicators. Complexes occupy always two adjacent nodes. Complex dissociation happens with constant rate k_{diss}. Upon successful complex formation between two adjacent individuals, and in the presence of empty space in the neighbourhood, the template is copied with rate ρ. After replication the complex breaks and the molecules return to an unbound state. Assume X_i is a replicator attempting to form a complex either with a replicator X_j or with a parasite P, the reaction scheme reads:

$$X_i + X_j \xrightleftharpoons[k_{diss}]{k_{a_i}} C_{X_j \circ X_i} \xrightarrow{\rho, \theta} 2\,X_j + X_i$$

$$X_i + P \xrightleftharpoons[k_{diss}]{\beta \cdot k_{a_i}} C_{P \circ X} \xrightarrow{\rho, \theta} 2\,P + X_i$$

where C is a complex and θ represents empty space. Mutations happen with probability μ and affect the complex formation rate k_a of the newly generated individual by adding a small random number (drawn from a uniform distribution $[-\frac{\delta_\mu}{2}, \frac{\delta_\mu}{2}]$). Individuals de-

grade with rate d, leaving empty space (also when in complex, in which case the other molecule survives and returns to an unbound state). Diffusion is modelled by swapping node contents between neighbouring nodes, it happens with rate D and may involve single individuals as well as complexes. See Supporting Information, Section 2.6.1 for more details.

Several important assumptions are made in order to simplify our model. First, we do not take into account that replication yields the complementary strand of a template. Second, we assume that replication rate ρ is the same for all replicators and parasites, and does not evolve in this study. Third, we assume that replicators behave as templates with rate set to one, which does not evolve, and parasites have are relatively more available for complex formation, i.e. $\beta > 1$. Notice that we let parasites' relative advantage β evolve (where specified). Fourth, parasites are modelled as a different class of molecules, in line with the results of a previous study (Takeuchi and Hogeweg, 2008). There, RNA-like replicators were modelled with sequences and secondary structures, and it was found that although few mutations could turn a replicator into a parasite, several mutations were necessary for parasites to optimise and establish themselves, thereby forming a new lineage. Furthermore, replicators evolved so that no close mutants of theirs was strongly parasitic (Chapter 4). Fifth, parasites do not form complexes with each other. In fact, complex formation was determined by sequence-complementarity in the above-mentioned study (Takeuchi and Hogeweg, 2008), and parasites evolved to minimise interactions with each other.

2.4   Results

2.4.1   Individual-based selection decreases replication

When we let the association rate (k_a) mutate in a replicators' system without parasites, k_a decreases (Fig. 2.1a) because replicators with lower-than-average k_a behave more as templates than as replicators, thus being replicated the most. Eventually, the minimum association rate needed for survival is reached ($k_a \approx 0.05$, just above death rate, $d = 0.03$). There, the system persists indefinitely because mutations that further decrease the probability of complex formation lead to local extinction, followed by the invasion of neighbouring replicators. Mixing the system, as well as a large increase in the diffusion rate, leads to global extinction as k_a becomes too low to sustain replication in the system (Supporting Information, Section 2.6.2).
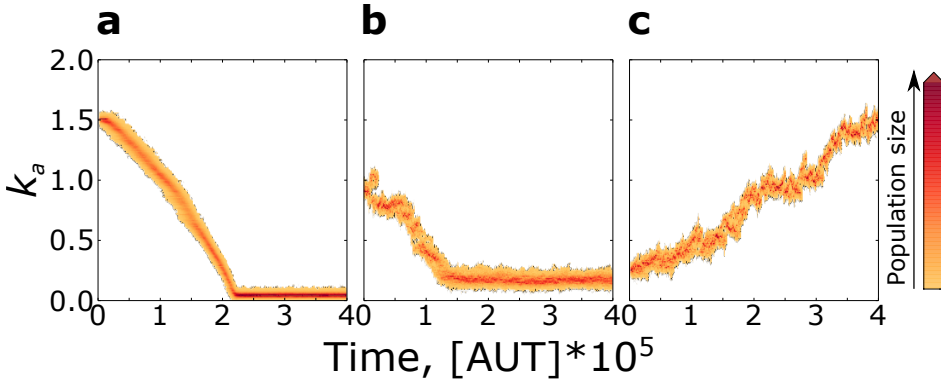
Figure 2.1. Evolutionary dynamics of replicators' association rate without and with parasites: k_a decreases in the absence of parasites, and evolves to higher values the stronger the parasites. **a**: Only replicators are present in the system; **b**: Parasite advantage $\beta = 1.40$; **c**: Parasite advantage $\beta = 1.70$. Other parameters: $k_a \in [0, 2]$, $k_{diss} = 0.25$, $\rho = 1$, $\Delta t_r = 0$, $\mu = 0.005$, $\delta_\mu = 0.05$, $d = 0.03$, $D = 0.1$. Results are robust to moderate changes in diffusion rate (Supporting Information, Section 2.6.2.

2.4.2 A phase transition to high association rates for larger parasite strength

We now introduce parasites and let the evolutionary dynamics of the association rate k_a unfold in response to a large constant parasite advantage ($\beta \in [1.0, 1.99]$).

Importantly, parasites are at advantage over replicators because 1) complex formation automatically shifts replication towards parasites by introducing a replicase/template implicit trade-off for replicators (Takeuchi and Hogeweg, 2007b), 2) they form complexes more frequently than replicators ($\beta > 1$) and 3) when in complex they do not replicate others, but are exclusively replicated. Yet, the deleterious effect of parasites is limited in spatial systems because local replication prevents them from becoming a global stability threat. The processes of spatial pattern formation in discrete replicators-parasites systems often results in travelling waves (Hogeweg and Takeuchi, 2003, Takeuchi and Hogeweg, 2009), where replicators expanding into empty space constitute the wave-front, and parasites outcompeting those replicators make up the back. Parasites leave empty space behind themselves due to local extinction. Notice that, because β is a multiplicative term, larger k_a and larger β contribute synergistically to parasite replication. One could expect that lower rates of complex formation are selected in the face of stronger parasites until the system collapses and goes extinct.

Contrary to this expectation, replicators evolve to larger association rates in response to stronger parasites (Fig. 2.1b and c). The increase in k_a is limited for lower parasite advantage ($1.1 \leq \beta \leq 1.4$), while for higher parasite advantage ($\beta \geq 1.7$), k_a is maximised (Fig. 2.2a). The system is bistable for intermediate values of β ($1.5 \leq \beta \leq 1.6$).
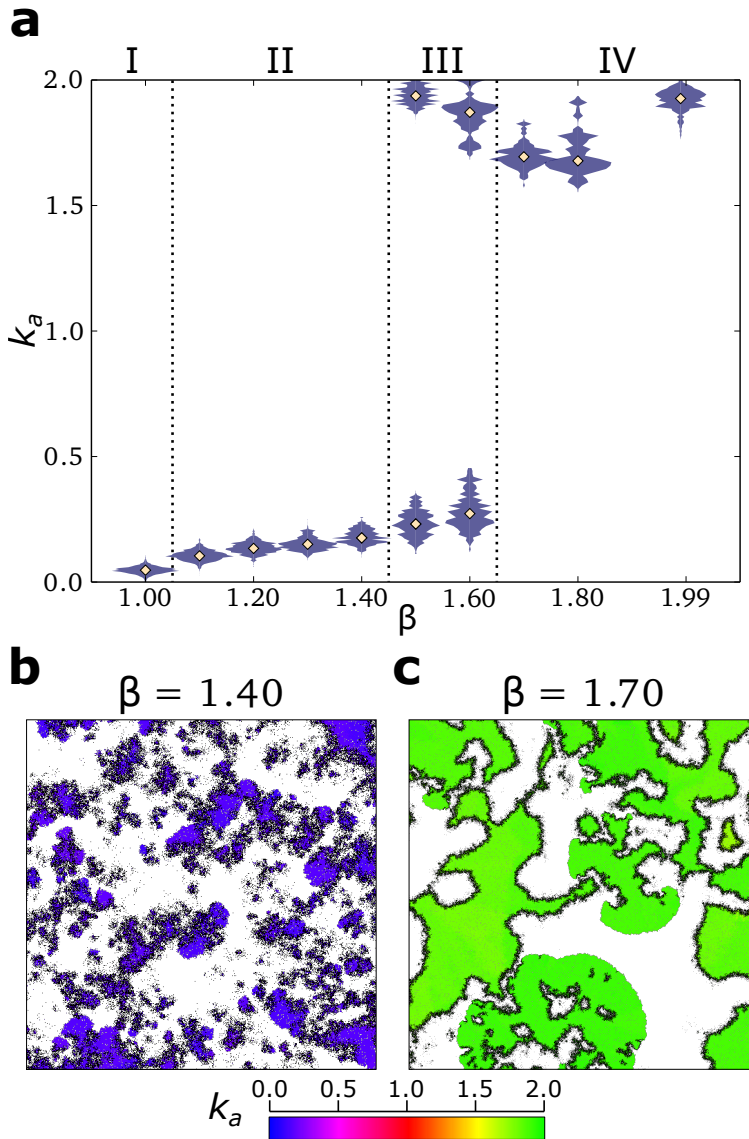
Figure 2.2. Stronger parasites lead to a phase transition in the eco-evolutionary dynamics. **a**: distribution of the association rate k_a at evolutionary steady state in a population of replicators, for different values of parasite advantage β. **b**: chaotic waves at evolutionary steady state when $\beta = 1.40$ (lattice size 512^2, for clarity 1/4 of the lattice is displayed). **c**: stable waves at evolutionary steady state when $\beta = 1.70$ (lattice size 2048^2, 1/4 of the lattice is displayed). All other parameters are as Fig. 2.1.

Parasites cannot be sustained in the system when they are too weak ($\beta \leq 1.0$, Supporting Information Section 2.6.3). For low values of both β and k_a, the replication of parasites is comparable to that of replicators, while local accumulation of parasites leads to their extinction. After parasites disappear, k_a decreases to the minimum needed for survival (Fig. 2.2aI).

For increasing β ($1.1 \leq \beta \leq 1.4$, Fig. 2.2aII), parasites persist and replicators reach an evolutionary steady state. Replicators and parasites organise in small and chaotic travelling waves, while the lattice looks overall patchy (Fig. 2.2b). Observations of the spatial dynamics suggest that new waves are often established by replicators escaping from the back of an existing wave (Supplementary video S1)[1]. Starting from large values of k_a, replicators evolve to decrease it because those with smaller k_a generate new waves more frequently (Supporting Information Section 2.6.4). k_a does not decrease without bound. Selection is stabilising because too low replication rates cannot support the persistence of both replicators and parasites. Surprisingly, the population size of replicators increases with parasite advantage (provided that $\beta \geq 1.3$, Fig. 2.3a). Notice that although parasites outnumber replicators at lower β, increasing parasite advantage steadily decreases the population of parasites (Fig. 2.3b).

For $1.5 \leq \beta \leq 1.6$ the system exhibits bi-stability, reaching either the lower or the higher (see below) final value of k_a depending on initial conditions (Fig. 2.2aIII).

When parasite advantage is large, $\beta \geq 1.7$, k_a is maximised over evolutionary timescales (Fig. 2.2aIV) and travelling waves become larger and stable (Fig. 2.2c, Supplementary video S2), provided that the lattice is large enough to contain them (Supporting Information Section 2.6.5). As the value of k_a increases, parasites become stronger and it becomes progressively less likely that new waves establish by escaping from the back of an existing one. When waves with different replicators compete side by side, those where replicators have higher k_a win because they invade space faster than the others (Supporting Information Section 2.6.6). Fig. 2.3 shows that the population size of replicators is much larger than that of parasites, and that population sizes do not change much with increasing parasite strength.

In summary, weaker parasites lead to small and chaotic waves. Replicators evolve low association rates and establish new waves by escaping from the back of older waves (cf. (Takeuchi and Hogeweg, 2009)). In contrast, limited escape is possible from stronger parasites. Replicators respond by evolving higher association rates, and organise with parasites into larger and more stable travelling waves. The transition from one behaviour to the other is akin to a first order phase transition, even though the system is far from equilibrium.

---

[1]The videos are available from the publisher's web site (dx.doi.org/10.1371/journal.pcbi.1004902)
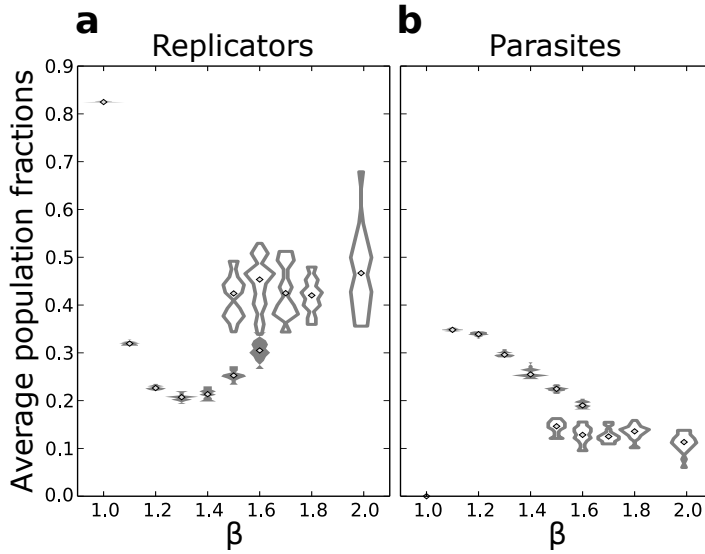
Figure 2.3. Fractions of **a** replicators and **b** parasites at evolutionary steady state. Full/empty distributions for replicators and parasites with the same β correspond in the two figures.

2.4.3 Strong parasites induce high k_a by generating empty space

Since the two regimes differ in the way replicators and parasites organise in space (respectively chaotic vs. stable waves), we characterise the phase transition in terms of the amount of uninterrupted empty space experienced by the expanding fronts of travelling waves. To this end, we run several ecological experiment (i.e. without mutations) with different values of k_a and β, and (qualitatively) measure the amount of empty space experienced by replicators on the front of travelling waves (Supporting Information Section 2.6.7). In Fig. 2.4, the results from the ecological experiments and those from the evolutionary dynamics (Fig. 2.2) are overlaid (the measure of the available empty space from the ecological system is represented as a heat-map for the different values of k_a and β).

For lower parasite advantage ($\beta \leq 1.60$), the evolutionary outcome matches the values of k_a obtained from the ecological experiments where the amount of uninterrupted empty space is minimal (but not the total empty space, see Fig. 2.4a), i.e. where wave birth is maximal. Thus, both wave-level selection (for higher wave birth) and replicator-level selection (for replication) lead to decreased k_a when it is large. When k_a is small, wave-level selection opposes individual-level selection, and k_a increases. A side effect of smaller steady state value of k_a is that replicators' population size at evolutionary steady state is minimal (Fig. 2.4b).

In contrast, for stronger parasites ($\beta \geq 1.70$), replicators evolution maximises k_a, and parasite population is smaller (Fig. 2.4c). Fig. 2.4 shows that uninterrupted empty space

is abundant in this parameter region regardless of the values of k_a. Because parasites are strong, replicators do not escape frequently from the back of the waves and do not form new waves, hence the large amount of uninterrupted empty space behind waves.

The availability of uninterrupted empty space drives the evolution of higher association rate, because sub-populations with higher k_a invade empty space faster and eventually dominate the expansion front.

We confirmed that k_a increases when empty space is available in two ways: by letting a population of replicators expand into unlimited empty space (Fig. 2.5a), and by periodically disrupting a resident population of replicators with large scale ablations (Fig. 2.5b, see Supporting Information Section 2.6.8 for more details).

Altogether, we find that a positive feedback loop establishes: as populations of replicators evolve to higher k_a to colonise space faster, parasites benefit from increased complex formation, becoming stronger. As parasite strength increases, invadable space is increasingly perceived as limitless by replicators, which enables the further increase of k_a. Thus, this process reverses the selection for becoming better templates.

In conclusion, strong parasitism leads to the formation of stable travelling waves, which in turn generate the selection pressure for increasing association rates in replicators.

2.4.4   Speciation as a consequence of spatial dynamics

So far we have shown that depending on the strength of the parasites, replicators evolve to moderate or high k_a through spatial self-organisation. Here we study replicator-only systems and show that in fact parasites arise automatically either as a consequence of disruptions or due to an increased cost of replication.

Periodic disruptions can select for higher association rates and lead to the speciation of parasites   Replicators evolve to large association rates with large scale disruptions (Fig. 2.5b). In contrast, the evolutionary dynamics become richer due to speciation when a large number of patches with intermediate size are ablated: k_a increases in one lineage, while it decreases in the other, which behaves as a parasite (Fig. 2.6a). Empty space allows for increasing k_a at the population's expansion front, while selection for becoming parasitic occurs behind it (i.e. k_a decreases). Because the two selection pressures are spatially segregated, they can both be maintained and do not balance, hence speciation occurs. Parasite-like replicators survive when ablations are of intermediate size because they can expand to an area larger than the size of the ablation itself between two disruptive events (whereas they could not with larger ablations).

The persistence of the two lineages is dependent on the occurrence of disruptions, and
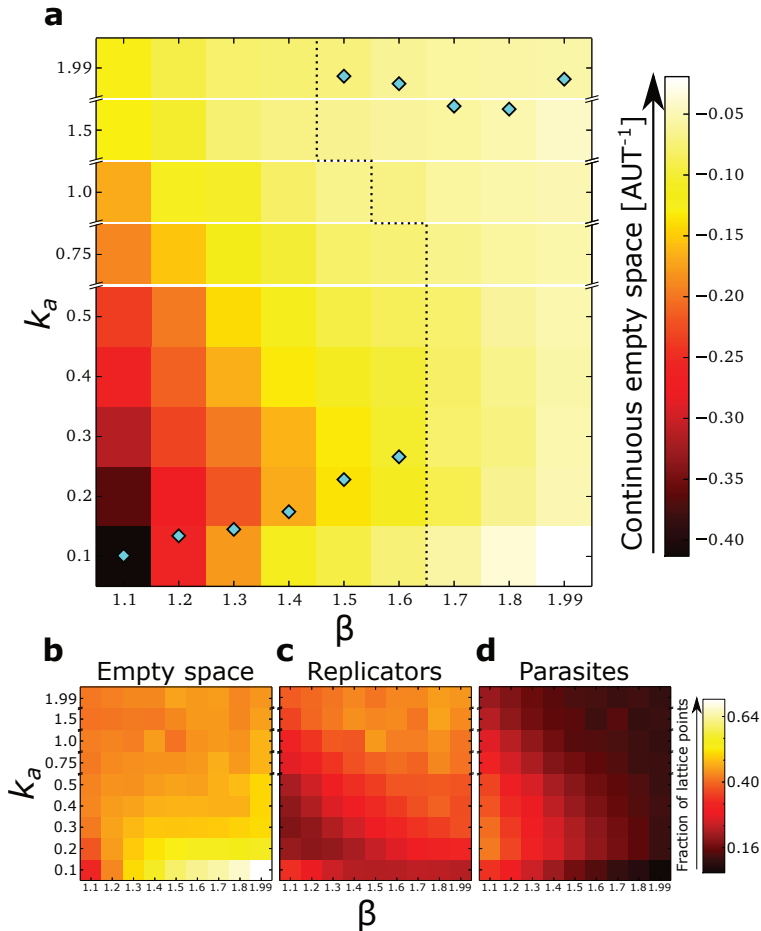
Figure 2.4. The amount of available empty space for waves predicts the phase transition in eco-evolutionary regimes better than other population statistics. **a**: Bifurcation diagram: Overlay of the median k_a (at the end point of the evolutionary dynamics) and the amount of uninterrupted empty space (from the ecological simulations). Cyan diamonds: median of the distributions of k_a after evolution, as taken from Fig. 2.2; dotted line: tentative sketch of the separatrix between the two regimes; tiled β vs. k_a surface: colours are according to an index (see colour bar) that measures the amount of empty space in front of a wave, calculated from ecological simulations for a combination of β and k_a (see Supporting Information Section 2.6.7 for details). **b**: Average fraction of lattice sites that are empty, **c**: occupied by replicators, **d** occupied by parasites, in the same ecological simulations used in **a**. All parameters as in Fig. 2.1. For the ecological simulations $\mu = 0$.

a minimum ablation size exists for k_a to increase: microscopic perturbations, achieved by increasing decay rates, do not introduce sufficient continuous space for speciation to occur (Supporting Information Section 2.6.8).

Longer replication times induce speciation Besides externally imposed disruptions, larger scale heterogeneities can be autonomously generated in the system if replicators spend more time replicating another individual. The reaction scheme presented in Methods is modified so that $C_{X_j \circ X_i} \xrightarrow{\rho, \theta, \Delta t_{repl}} 2\,X_j + X_i$ and similarly for parasites.

So far, we have assumed that after a replicator and a template form a complex, replication is only limited by finding empty space before the complex breaks apart. With template-dependent polymerization in mind (e.g RNA replication), it is reasonable to assume that a complex spends time before replication is complete, e.g. undergoing conformational changes to activate the replication machinery, as well as spending time actually copying the template. Notice that longer replication times make replication more costly, because replicators spend a larger fraction of their life-time replicating others. This strengthens the selection pressure to decrease k_a (we assume that if a complex breaks before the replication time has passed, no product is formed).

When replication time is sufficiently long ($\Delta t_{repl} \geq 3.5$ AUT, Supporting Information Section 2.6.9), a population of replicators undergoes speciation and two lineages form: in one k_a increases, in the other it decreases and these replicators behave as parasites (Fig. 2.6c). The two species organise in travelling waves (Fig. 2.6d) and establish an evolutionary feedback: parasite-like replicators cannot persist autonomously and exploit replicators with higher k_a for replication; the empty space they leave behind can be re-colonised by other replicators, which leads to increased k_a.

Thus, longer replication times intensify the selection pressure to decrease k_a. While in the previous paragraph empty space was imposed on the system, here large patches of empty space are generated when k_a becomes too low and replicators go extinct. The resulting invasion dynamics trigger the selection pressure for increasing k_a on the expanding front, which triggers the individual-based pressure to decrease k_a behind it and leads to parasite-like replicators. We conclude that the long term evolutionary consequence of longer replication time is that an emergent feedback establishes between evolution and spatial organisation. This feedback destabilises the evolutionary steady state presented in Fig. 2.1a and two evolutionarily codependent species arise (cf. Chapter 3).

Notice however that parasite-like replicators experience no relative advantage as templates (β is fixed to 1 in their case). The next question is, then, whether the co-evolution of replicators and parasites under a regime of more costly replication leads to stronger or weaker parasites, i.e. better or worse templates.
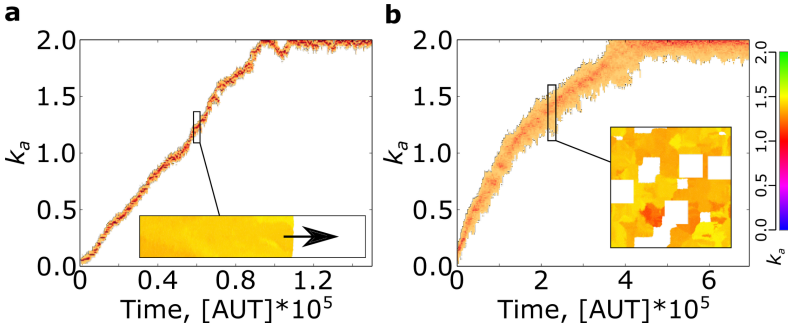
Figure 2.5. Association rate increases on the front of an expanding population, as well as when large scale disruptions occur. **a**: A population expands into an infinite space (Only the data about the front is plotted). Inset: a snapshot of the expansion dynamics (direction indicated by the arrow) **b**: evolutionary dynamics of k_a with large scale disruptions. Disruptions are as follows: every 50 AUT, 4 square patches (200^2) of the lattice are turned to empty at random positions. Lattice size 1024^2. Inset: a snapshot of the spatial dynamics.
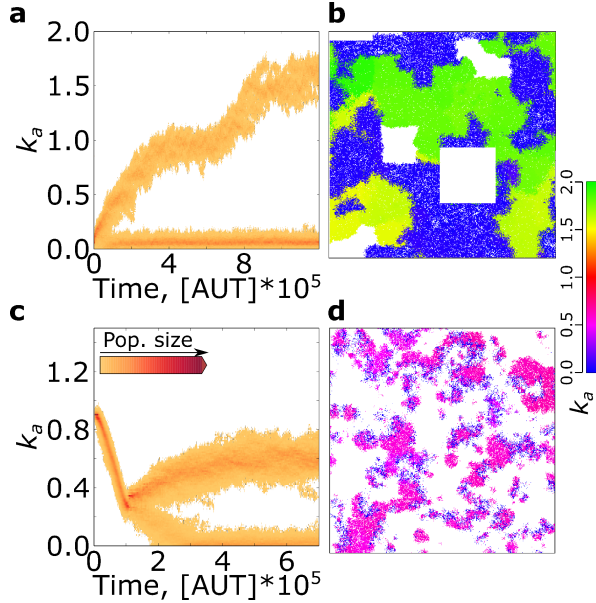
Figure 2.6. Smaller disruptions and increased cost of replication lead to speciation **a**: Time plot of the distribution of k_a in a simulation with smaller disruptions. Disruptions are as follows: every 50 time steps, 16 square (100^2) patches of the lattice are turned to empty at random positions. Lattice size 1024^2, other parameters as in Fig. 2.1. **b**: A snapshot of the simulation with disruptions (an area of 400^2 is displayed). **c**: Time plot of the distribution of k_a in a simulation where duration of replication is $\Delta t_{repl} = 4$ AUT. All other parameters are the same as above, except for $k_{diss} = 0.1$. **d**: A snapshot of the simulation with longer replication time (lattice size 1024^2, displayed: 400^2).

2.4.5 Longer replication time selects for weaker parasites in replicator-parasite co-evolution

We extend the replicator-parasite system presented above by letting the parasite advantage β co-evolve with the replicators' association rate (k_a), and we vary the duration of a replication event (Δt_{repl}).

Co-evolutionary dynamics In Fig. 2.7 the co-evolutionary steady state values of β and k_a are plotted for different replication times.

The replication rates of both replicators and parasites is maximised in the long term co-evolutionary dynamics when replication time Δt_{repl} is set to zero (Fig. 2.7a), i.e. the setting used for the replicators-parasites system analysed above. While parasites are selected to constantly increase β, the evolutionary trajectory of k_a depends on initial conditions (Fig. 2.8): when k_a and β are initialised at lower values, k_a first remains stationary around a (meta)stable equilibrium line (cf. Fig. 2.4), and reaches larger values only when β is large enough. Taken together, the evolutionary trajectory of k_a and observation of the spatial dynamics show that the system autonomously (dynamically) undergoes the phase transition between chaotic and stable waves described in Fig. 2.2.

The evolutionary maximisation of both β and k_a occurs for replication times up to $\Delta t_{repl} = 2.0$.

However, a second steady state emerges for $\Delta t_{repl} > 0$ for which β and k_a are not maximised, but rather they stabilise at lower values (Fig. 2.7a). Following this steady state to larger Δt_{repl}, we observe that β decreases and k_a increases, i.e. weaker parasites and stronger replicators evolve. When $\Delta t_{repl} \geq 4.0$ the median value of β is less than one, i.e. parasites become worse templates than replicators. For $\Delta t_{repl} = 4.5$ parasites evolve to extinction, and replicators may speciate into two lineages, one of which becomes parasite-like (k_a approaches zero, see previous paragraph).

Spatial population dynamics The spatial co-evolutionary dynamics that lead to these steady states are straightforward.

For β and k_a sufficiently large, and lower Δt_{repl}, replicators and parasites organise into stable travelling waves (Fig. 2.7b, $\Delta t_{repl} = 0$ and the higher steady state of $\Delta t_{repl} = 1.5$). There, parasites with larger β outcompete those with lower β because the former form more complexes with replicators, and therefore invade faster. The empty space generated behind parasites is then occupied faster by replicators with larger k_a. Meanwhile, parasites with larger β profit the most from replicators with larger values of k_a. This leads to maximise both replicators and parasites' replication rates (Fig. 2.8).

In contrast, for longer replication times (Fig. 2.8, $\Delta t_{repl} > 1.5$) or with smaller initial

Figure 2.7. Bistability in the co-evolutionary steady states of replicators (k_a) and parasites (β), in response to longer replication times. **a**: A full/empty distribution in the parasite pane corresponds to the full/empty distribution for replicators at the same Δt_{repl}. All parameters as above, $\beta \in [0, 2]$. **b**: Snapshots of the lattice from simulations with different Δt_{repl}. Left snapshot: spatial distribution of replicators (parasites in grey); right snapshot: parasites (replicators in grey). Lattice size from top to bottom: 2048^2, 1024^2, 1024^2, 512^2, 512^2 (1/4 of the lattice is displayed for clarity).
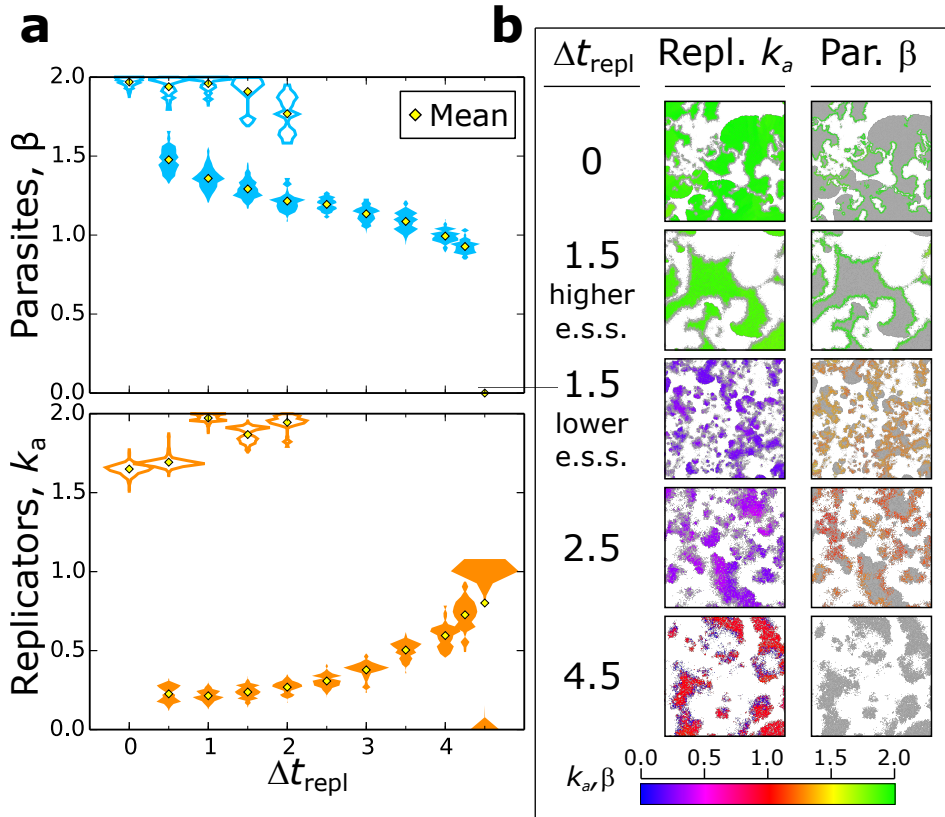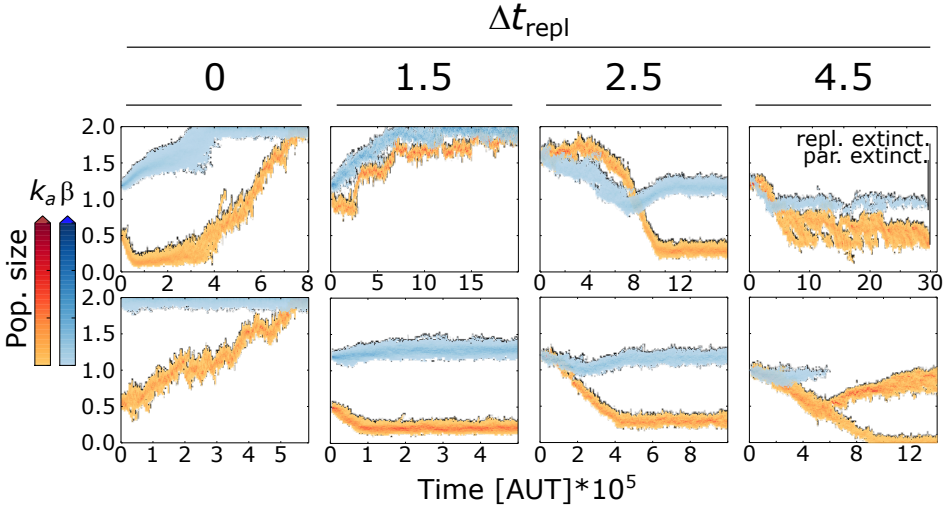
Figure 2.8. Bistability in the co-evolutionary steady states of replicators and parasites, in response to longer replication time. Pictures show the co-evolutionary trajectories of β and k_a for different values of Δt_{repl} and different initial conditions.

values of k_a relative to β (Fig. 2.8 $\Delta t_{repl} = 2.5$, lower pane), the spatial dynamics of this steady state are characterised by chaotic travelling waves. New waves are established by replicators creeping through the back of older waves (Fig. 2.7b, $\Delta t_{repl} = 1.5$-lower steady state, $\Delta t_{repl} = 2.5$ and 4.5). Chaotic waves cause the co-evolutionary stabilisation of the replication rates of both parasites (cf. (Takeuchi and Hogeweg, 2009)) and replicators (see above). This effect is striking because a larger Δt_{repl} leads to both a stronger selection pressure on replicators for decreasing k_a, and a larger intrinsic advantage to parasites (because they do not pay any cost for replication). Yet, increasing Δt_{repl} leads to co-evolutionary steady states where replicators' association rate is larger and parasites' advantage as templates is smaller.

Spatial pattern formation causes directional changes in the evolutionary trajectories of replicators The selection pressure on replicators can change direction during the co-evolutionary dynamics, making the evolutionary trajectory of k_a non monotonic (Fig. 2.8 upper row). This happens because spatial patterns themselves change over the course of evolution (cf. (Boerlijst and Van Ballegooijen, 2010)).

Although the evolutionary steady state with low k_a is not stable for $\Delta t_{repl} = 0$, its "ghost" (cf. (Sardanyés and Solé, 2006)) can be observed in the trajectory of the co-evolutionary dynamics (Fig. 2.8). It can take repeated "evolutionary attempts" for the system to transition completely to the phase with stable waves and replicators with higher k_a (Supporting Information Section 2.6.10), because smaller waves locally outcompete larger ones (cf. (Takeuchi and Hogeweg, 2009)).

For $\Delta t_{repl} = 1.5$ a similar effect can be seen, where replicators initially decrease their k_a in response to weak parasites, but then increase k_a when parasites have evolved to sufficiently high β (Fig. 2.8). Interestingly, parasites are initialised at the same value in the two panes of Fig. 2.8 for $\Delta t_{repl} = 1.5$. Their evolution is dependent on the kinds of replicators they form waves with. Because in both cases replicators face "weak" parasites, they respond by decreasing k_a. However, in one case replicators start with high k_a so parasites increase β, in the other replicators have low initial k_a and parasites do not maximise β.

The opposite situation may also happen for longer replication times (Fig. 2.8, $\Delta t_{repl} > 1.5$), replicators initially respond to strong parasites by increasing k_a. Later, when β is sufficiently small, replicators are selected to decrease k_a (notice that although β's trajectories are non monotonic, parasites evolve a monotonically decreasing complex formation rate $\beta * k_a$). When replication time is long enough and the initial value of k_a is low relative to β, travelling waves destabilise because parasites invade replicators faster than replicators expand into empty space (the expansion front becomes narrower). In these limit conditions (if parasites were any stronger the system would go extinct) replicators cannot escape from the back of older waves and initially evolve to larger k_a, while the parasitic erosion of the invasion front often isolates small groups of replicators from the rest of the wave (Supporting Information Section 2.6.11), which survive longer if the associated parasites are weaker.

Finally, the limit viable replication time is $\Delta t_{repl} = 4.5$, for which parasites drive themselves to extinction (global extinction quickly ensued for $\Delta t_{repl} = 4.7$ and larger). When parasites are extinct, replicators cannot benefit from the selection pressure that sustained higher k_a (deriving from spatial pattern formation) and may themselves face extinction (Fig. 2.8, $\Delta t_{repl} = 4.5$ upper pane). If parasites are not allowed to evolve β below 1, instead, both replicators and parasites persist indefinitely (Supporting Information Section 2.6.12). Replicators can persist in the absence of parasites if they succeed at evolving a parasite-like lineage and re-organise in spatial pattern (Fig. 2.8, $\Delta t_{repl} = 4.5$ lower pane).

In conclusion, the emergent selection pressures originating from self-organised spatial patterns locks the evolution of replicators to the evolution of parasites. This either leads to stable travelling waves that allow for the evolutionary increase in replication rates of both replicators and parasites, or to chaotic travelling waves that bring the system to an evolutionary pressure that selects for weaker parasites and stronger replicators, when the cost of replicating others becomes higher. We stress that results are due to the co-evolution of replicators and parasites, which allows a larger degree of complexity to unfold. In fact, these results are lost when only replicators can evolve at longer replication times, and instead we obtain results qualitatively similar to Fig. 2.2, namely that stronger parasites lead to higher association rates in replicators (Supporting Information Section 2.6.13).

2.5   Discussion

In this study we analysed the eco-evolutionary dynamics of minimal replicator-parasite systems. Replicators copy templates after forming complexes with them, and parasites may be better templates than replicators.

An earlier study on a similar system (Takeuchi and Hogeweg, 2009) showed that when only parasites could mutate, selection for wave-level fertility (i.e. small and chaotic waves) resulted in the evolution of weaker parasites. Here we have paralleled that result when replicators mutate: replicators evolve to smaller association rates when parasites advantage is weak by escaping more frequently from the back of their waves, which forms more (smaller and chaotic) waves. Whether parasites or replicators mutate, waves evolve so that new waves are generated more easily. Hence there is selection for wave-fertility.

However, we have also identified a novel mode of evolution when parasites are stronger, which produces stable and long-lived travelling waves composed of replicators with larger association rates. Therefore, waves can also experience selection for longevity.

A phase transition separates these two spatial patterns and their evolutionary regimes. This phase transition can be studied by measuring the amount of empty space generated by parasites, and invaded by replicators at the front of a travelling wave (cf. Cwynar and MacDonald (1987), Korolev (2013), Shine *et al.* (2011), Travis and Dytham (2002) and Chapter 3).

We conclude that although replicators are prone to decrease association rates (to spend more time as templates), introducing parasites allows replicators to sustain a higher association rate.

Notice that 1) we recover both evolutionary regimes when we set k_a to constant and let only parasites mutate (Supporting Information Section 2.6.14) and 2) longer replication time-spans $\Delta t_{repl} > 0$ do not qualitatively change these results.

Finally, we analysed the co-evolution of replicators and parasites when the time-span needed for replication is prolonged. For smaller Δt_{repl} both evolutionary strategies are attainable, and the system shows evolutionary bi-stability. For larger Δt_{repl}, only the co-evolutionary steady states with relatively lower k_a and β is reachable.

We introduced the term Δt_{repl} to study the evolutionary dynamics of the system when replication rates do not depend solely on the availability of empty space. However, we did not let Δt_{repl} evolve because we did not specify any molecular detail of RNA replication (for instance, the evolution of larger or smaller Δt_{repl} should be functionally related to that of β and k_a), which instead could be better targeted by sequence-based models (Szabó *et al.* (2002), Takeuchi and Hogeweg (2008) and Chapter 4).

Nevertheless, let us assume that Δt_{repl} scaled with template length, i.e. $\Delta t_{repl} > 0$, and that k_a and β evolved independently from it. Because Δt_{repl} would likely be selected to decrease, following the lower equilibrium line of Fig. 2.7, evolution would reach a steady state in which neither k_a nor β are maximised. Thus, our results would hold and we would recover the conclusions of (Takeuchi and Hogeweg, 2009).

Pre-biotic evolution   In the context of prebiotic evolution, mutually replicating molecules (among which e.g. hypercycles (Eigen and Schuster, 1978)) are known to be evolutionarily unstable: selection at the individual level causes the evolution of better templates to the detriment of replication (Bresch *et al.*, 1980, Maynard Smith, 1979). Higher-order organisation, such as spatial extension ((McCaskill *et al.*, 2001, Szabó *et al.*, 2002)), spatial pattern formation ((Boerlijst and Hogeweg, 1991c, Hogeweg and Takeuchi, 2003, Scheuring *et al.*, 2003)) or vesicles ((Niesert *et al.*, 1981, Szathmáry and Demeter, 1987)), is often invoked to solve this problem.

Here we have shown that individuals that behave only as templates may actually aid the evolution of higher replication rates. Parasitic behaviour is, in fact, "functional" because it contributes to the spatial structure that selects for higher levels of replication.

An earlier study on metabolism-based replicator models (Könnyű *et al.*, 2008) showed that metabolic parasites could evolve into replicases, providing a group-level (albeit relatively costless) evolutionary benefit to the system. In our system instead parasites are beneficial *as parasites*, since we do not pre-conceive extra functional possibilities for them. Parasites induce more replicase activity in pre-existing replicases despite the cost of being a stronger replicase.

We conclude that parasites may be considered a functional degree of freedom that can be exploited by evolution through higher-order organisation.

The evolution of multilevel evolution   In general, spatial pattern formation can deeply affect the evolution of its components (Boerlijst and Hogeweg, 1991a,c, van Ballegooijen and Boerlijst, 2004), and can lead to selection that reinforces the spatial patterns even at the expenses of its composing individuals (Savill *et al.*, 1997), or to self-organised evolutionary switching between different spatial patterns (Boerlijst and Van Ballegooijen, 2010). This higher-level selection can be recognised in travelling waves as well (Hogeweg and Takeuchi, 2003, Takeuchi and Hogeweg, 2009, Takeuchi *et al.*, 2011). Travelling waves, however, also display compositional (and spatial) inheritance and variation, and therefore are units of evolution ((Takeuchi and Hogeweg, 2009)).

Here we have shown that the system can autonomously undergo the phase transition between chaotic and stable waves, as a result of a feedback between evolution and self-organisation. This means that the self-reinforcing selection pressure can change directionality. Thus we have observed the evolution of wave-level evolution, transitioning

from selection for fertility to selection for longevity.

It has been recently asked: "How far can the RNA World go without being encapsulated in a cell?" (Higgs and Lehman, 2014). We have not yet seen its limit, it seems.

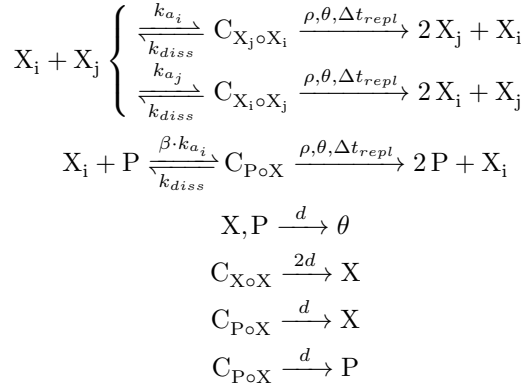## 2.6 Supporting Information

2.6.1 Details of the model

The model is a spatially extended, individual oriented Monte Carlo simulation system. Space is modelled as a two-dimensional square lattice with toroidal boundaries (based on CASH, de Boer and Staritsky (2000)). Each node has eight neighbours (the Moore neighbourhood) and can be empty or occupied by one individual. All dynamics implemented is local, involving only nodes in the Moore neighbourhood, and all nodes are updated in random order in one iteration of algorithm.

Replicators and parasites' complex formation and replication There can be two types of individuals, replicators and parasites. They differ in two ways: 1) replicators can both *form* and *accept* complexes, whereas parasites can only *accept* complexes; it follows that replicators can form complex with each other and with parasites, but the latter can never form complexes with other parasites. 2) Parasites can be better than replicators at forming complexes. This is modelled by assuming that replicators accept complexes with rate scaled to 1, while parasites accept complexes at rate $\beta \geq 1$ (unless it evolves otherwise), which represents the advantage parasites experience at accepting complexes relative to replicators. The rate at which replicators form complexes, called the "association rate" throughout the text, is k_a. Complex molecules always occupy two adjacent nodes of the lattice (and each molecule in complex keeps a "flag" that signals where the other molecule is located).

Complex dissociation happens with constant rate k_{diss}. When a complex is formed between two adjacent individuals, Δt_{repl} AUT (units of time) must pass without the complex dissociating; then if an empty node is available in the neighbourhood, the template is copied with rate ρ. After replication, the complex breaks and molecules return to an unbound state. Notice that the time passing, i.e. the counter that starts from Δt_{repl} and reaches zero, is modelled as a first order reaction. Individuals decay with rate d, leaving empty space. Molecules in complexes can also degrade.

Assume X_i and X_j are replicators, P is a parasite, C is a complex (the suffixes indicate the individuals in complex, as well as which one is template and which one is replicase)

and θ represents an empty node. The full reaction scheme reads:

$$X_i + X_j \begin{cases} \xrightarrow{\ k_{a_i}\ } C_{X_j \circ X_i} \xrightarrow{\rho, \theta, \Delta t_{repl}} 2\,X_j + X_i \\[2mm] \xleftarrow[k_{diss}]{} \\[-4mm] \xrightarrow{\ k_{a_j}\ } C_{X_i \circ X_j} \xrightarrow{\rho, \theta, \Delta t_{repl}} 2\,X_i + X_j \\[1mm] \xleftarrow[k_{diss}]{} \end{cases}$$

$$X_i + P \xrightarrow[k_{diss}]{\beta \cdot k_{a_i}} C_{P \circ X} \xrightarrow{\rho, \theta, \Delta t_{repl}} 2\,P + X_i$$

$$X, P \xrightarrow{\ d\ } \theta$$

$$C_{X \circ X} \xrightarrow{\ 2d\ } X$$

$$C_{P \circ X} \xrightarrow{\ d\ } X$$

$$C_{P \circ X} \xrightarrow{\ d\ } P$$

Mutations happen after replication with probability μ and may affect the association rate k_a of the newly generated replicator, or parasite advantage β, by adding a small random number drawn from a uniform distribution $[-\frac{\delta_\mu}{2}, \frac{\delta_\mu}{2}]$. No other parameter is evolved.

Diffusion   Diffusion is modelled by swapping the contents of neighbouring nodes, it happens with rate D and may involve single individuals as well as complexes. The algorithm for diffusion follows Takeuchi and Hogeweg (2009): diffusion between two molecules or a molecule and empty space is modelled as a second order reaction, diffusion between a complex (which occupies two lattice nodes) and a free molecule or an empty node is considered a third order reaction, diffusion between two complexes is a fourth order reaction. Suppose that the focal molecule X_1 is in complex with X_2, and a randomly chosen neighbour Y is a single molecule. Diffusion swaps their locations in order to maintain adjacency between molecules in complex as follows: X_1 takes the place of Y, X_2 takes the place where X_1 was, and Y is moved to the node where X_2 was. Two complexes, $X_1 \circ X_2$ and $Y_1 \circ Y_2$, swap places as follows: assuming X_1 is the focal molecule, and Y_1 the random neighbour, X_1 moves to the node of Y_2 and vice versa, X_2 moves to Y_1's location and vice versa.

Algorithm update   One iteration of the alogrithm runs as follows.

- One node is drawn randomly, and one if its neighbouring nodes is selected randomly.

- Based on the content of the nodes, the possible events are determined, e.g. a replicator next to a parasite can form a complex, swap places or decay.

- One event is drawn. Each event happens with probability proportional to its rate, scaled by a constant which is larger than the sum of all the rates of all possible

reactions (because one event is drawn per iteration, whether it happens or not, the constant also define the number of iterations of the algorithm which determines the unit of time).

- The rate of a first order reaction (such as decay or decreasing the counter of Δt_{repl}) is used "as such", whereas the rate of second order reaction (such as replication or diffusion of two individuals not in complex) is halved. The rate of third order reaction (e.g. complex diffusion) is divided by 6 and so on.

2.6.2   The effect of mixing and large diffusion rates

Mixing and large diffusion lead to extinction in the system with only replicators

Well-mixed system   Mixing is approximated by greatly increasing the rate of diffusion. Unsurprisingly, selection for becoming a better template dominates when spatial patterns cannot form and global extinction happens when k_a reaches values too low to sustain replication (Fig. 2.9a).

Large diffusion rates   In the main text, we have shown that low diffusion allows the persistence of replicators at the minimum viable association rate. This happens because local accumulation of replicators with too low values of k_a leads to their extinction, followed by re-invasion from neighbouring replicators. Increasing diffusion favours replicators with low k_a because it prevents their local accumulation, and allows them to exploit replicators further away from them, which leads to a lower steady state distribution of k_a (Fig. 2.9b). When diffusion is too large the system approaches a well-mixed conditions, k_a becomes too small to sustain replication, and extinction ensues. Interestingly, the limit value for diffusion rate before extinction ensues ($D = 1$) also displays the most variability. This is presumably due to the fact that local extinction of replicators happens on a somewhat larger spatial scale, so that the subsequent re-invasion endures for long enough to trigger some selection for larger k_a (see main text).

A moderate increase in diffusion rate does not qualitatively change results in replicators-parasites system

Increasing diffusion rate 5 folds does not lead to qualitatively different results (Fig. 2.10). We did not test the full range of β values due to computational load. However, replicators behave much like the default case ($D = 0.1$) in that 1) they reach the minimal viable k_a when no parasites are introduced, 2) they reach an evolutionary stable steady state for weaker parasites ($\beta = 1.30$), and maximise k_a when parasites

are stronger ($\beta = 1.80$). Notice that, in general, an increase in diffusion makes spatial patterns larger. In order to contain such patterns for e.g. $D = 1$, the lattice should be much larger than what is computationally feasible.

2.6.3   Too weak parasites are not maintained in the system

When $\beta \leq 1$, parasites cannot prevent replicators from reaching low values of k_a and go extinct, as shown in Fig. 2.11.

2.6.4   Chaotic waves competition and selection for lower k_a

In order to illustrate how wave-level selection causes the decrease of k_a in replicators, we modified the set-up of the simulation by shaping the field as a long narrow strip. We initialised replicators and parasites in two waves, one at each end of the strip. In one wave replicators have larger k_a than in the other. We set mutation rates to zero to focus only on the selection process (the ecological competition). We let the two waves collide, as shown in Fig. 2.12. Although replicators with smaller k_a expand slower, they escape from the back of the original wave more frequently. The space behind the wave with weaker replicators is thus much fuller and works as a reservoir of replicators when the two initial waves collide and annihilate each other. The remaining empty space is then invaded by waves in which replicators have lower k_a.

2.6.5   Global extinction due to small lattice size

The size of spatial patterns can become comparable to that of the lattice when β and k_a are large. In this condition, the coexistence of multiple spatial patterns becomes impossible, and extinction can happen, as shown in Fig. 2.13. Larger lattices are therefore required to allow the evolutionary dynamics unfold properly (at the expenses of a larger computational load).

2.6.6   Stable waves competition and selection for larger k_a

When parasites are strong $\beta \leq 1.70$, travelling waves become larger and more stable, because limited (or no) escape is possible from their back. This means that, given the same parasite (i.e. the same $beta$) competition is determined by invasion into empty space, as shown in Fig. 2.14 (mutation rate is set to zero). Replicators with larger k_a are selected because they invade faster, and take over the expansion front.
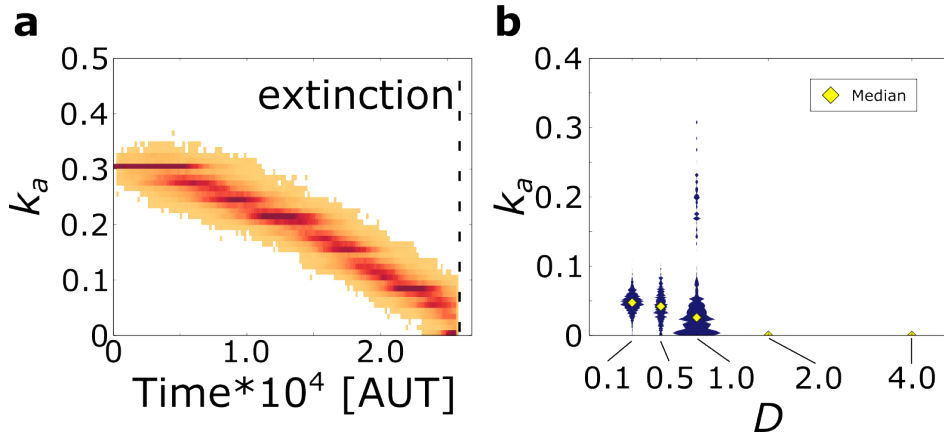
a

b

Figure 2.9. a: In a mixed system, extinction due to selection for becoming a template cannot be prevented. $D = 479.04$, meaning that out of 100 events, 96 on average are diffusion, each happening with probability $p_D = 0.998$. Lattice size 128^2. All other parameters are the same as in the main text. **b**: Evolutionary steady state of a replicators-only system with different diffusion coefficients D. Lattice size 512^2.
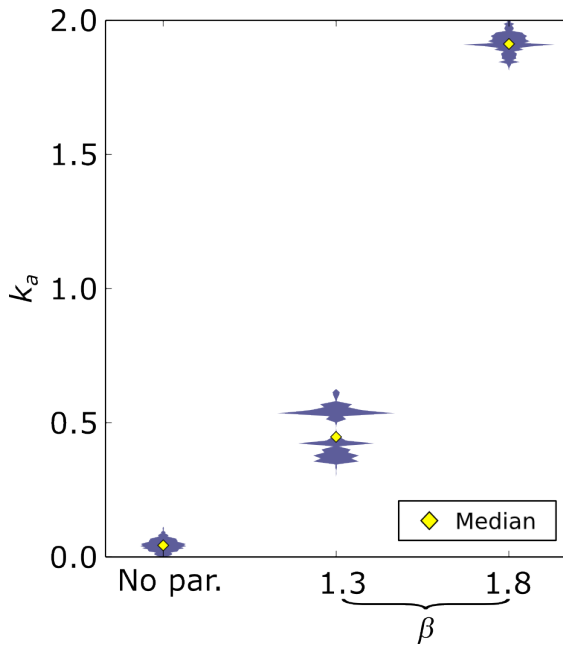

Figure 2.10. No qualitative difference observed in evolutionary steady state with a larger rate of diffusion. $D = 0.5$. Other parameters as in main text.
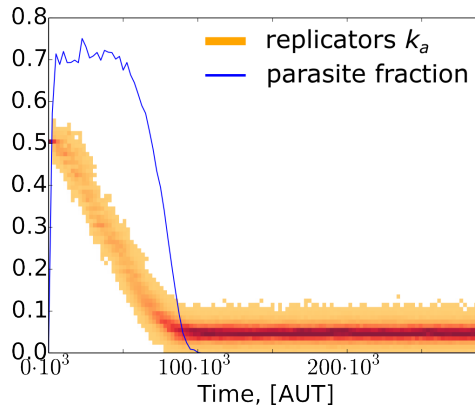
56

Figure 2.11. Evolutionary dynamics of replicators in the presence of very weak parasites, parasites go extinct when k_a becomes too small. $\beta = 1.0$, other parameters as in main text.
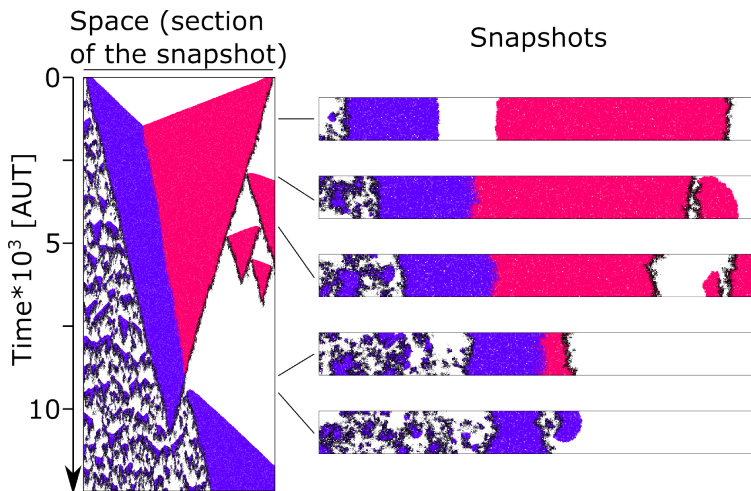

Figure 2.12. Wave level selection for lower k_a. Left: space-time plot of a replicator-parasite system (i.e. a stack of the middle row of the lattice at successive time points). Right: snapshots of the lattice at selected time points. Parasites (black) $\beta = 1.3$, replicators (indigo on the left) $k_a = 0.20$ and (magenta on the right) $k_a = 0.80$.
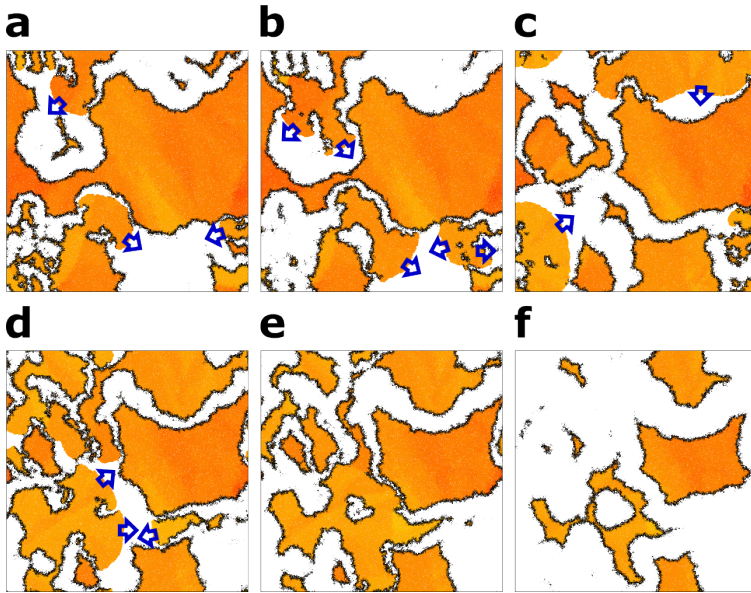
Figure 2.13. Small lattices lead to extinction when travelling waves become larger. **a**, **b**, **c**, **d**, **e**, **f**: Subsequent snapshots of the lattice. The expansion front of replicators is indicated by blue arrows. Parasites advantage $\beta = 1.7$. Notice that boundaries are wrapped.

Figure 2.14. Replicators with larger values of k_a are selected in the presence of stronger parasites. Snapshots are at subsequent time steps. Parasites (black) $\beta = 1.80$, replicators above (red) $k_a = 1.0$, below (yellow) $k_a = 1.2$.

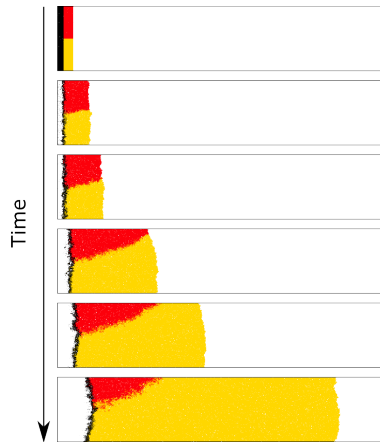### 2.6.7 Method for calculating the index of continuous empty space (Main Text Fig. 2.4)

The index in Main Text Fig. 2.4 is calculated as follows. For every combination of β and k_a shown in the figure, an ecological simulation (i.e. $\mu = 0$) is run with replicators and parasites. First a transient (5000 AUT) is let pass where the system self-organises and forms spatial patterns, then, 3x3 equally distant nodes (to minimise data correlations) in the field are screened for presence/absence of individuals. Every consecutive time step a node is empty, a counter is incremented. If the node becomes occupied, the number of consecutive time steps it spent as empty is recorded and the counter is re-set to zero. When the node turns empty again, the process can restart.

All the values of the counters are accumulated and a histogram is generated which represents the distribution of time spans (t_{empty}) nodes spent as empty. We used this as a proxy for the amount of continuous empty space a wave experiences. The log-transformed data can be well fitted by a linear function of the form $\alpha - \gamma t_{empty}$ (meaning that the original distribution is exponential), as shown in Fig. 2.15. $-\gamma$ is the index of continuous empty space of Main Text Fig. 2.4.

We chose $-\gamma$ because it increases when spatial patters generate larger amounts of empty space behind them.

2.6.8 Ablations

Methods   Periodical ablations are introduced in the system by deleting square patches every 50 AUT. In order to introduce n ablations of size η^2, n random coordinates are generated, which represent the centres of the ablations. The surface of size η^2 is then turned to empty, and the dynamics proceeds by normally (if an individual is in complex with one that is ablated, the complex breaks).

Ablations sustain the two species system   Ablations of an intermediate size lead to the establishment of two distinct lineage in a system with only replicators. If ablations are stopped, however, the selection for increasing k_a disappears, and the species with the lowest k_a outcompetes the other, as shown in Fig. 2.16

Higher death rate does not lead to speciation   In order to generate enough empty space to trigger selection for higher k_a, ablations must be of a minimum size. Point-sized ablations, for instance, are too small to achieve this. We modelled point size ablations by increasing the decay rate of replicators from the default $k_{death} = 0.03$ to 0.2 (i.e. 6.6 folds).
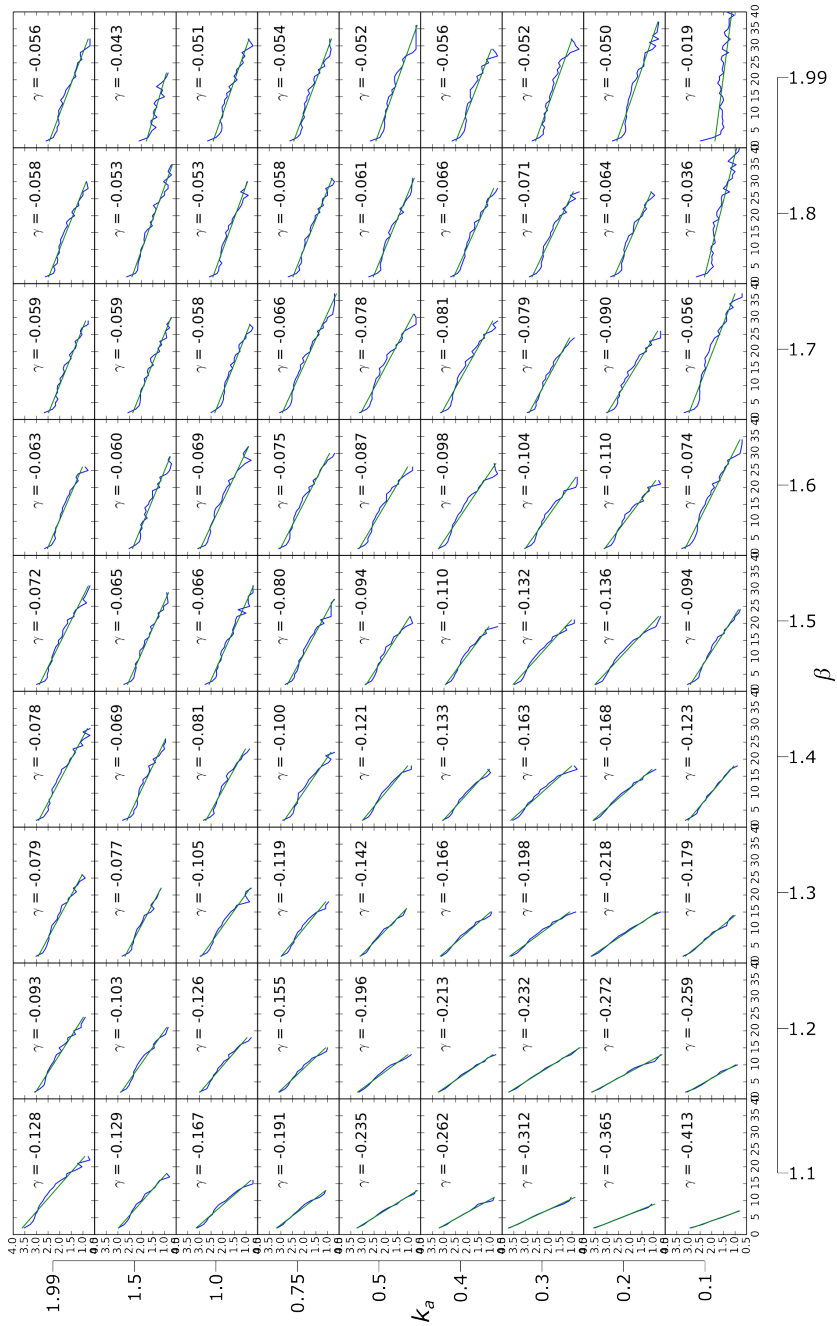
Figure 2.15. Fitting the time spans nodes spend as empty space with exponential distributions. Each plot is a combination of k_a and β ($\mu = 0$). Distribution is log-transformed (base 10), units on x-axis is AUT*500. Blue: original data, green: line-fit.

Figure 2.16. Interrupting ablations lead to the extinction of the species with higher k_a (green). Lattice size 1024^2.

Figure 2.17. Increasing death rate does not lead to speciation. $k_{death} = 0.2$, other parameters as in main text.

Figure 2.18. Evolutionary steady state values of k_a for replicators-only systems with longer replication times. $kdiss = 0.1$, other parameters as in main text.

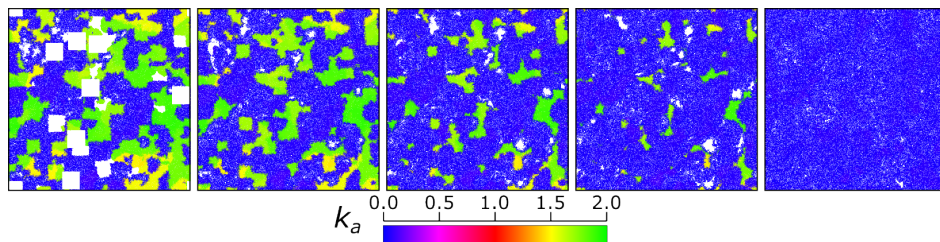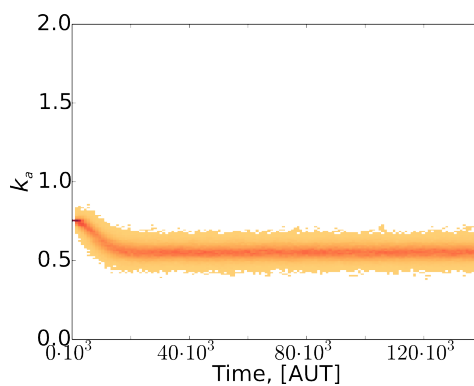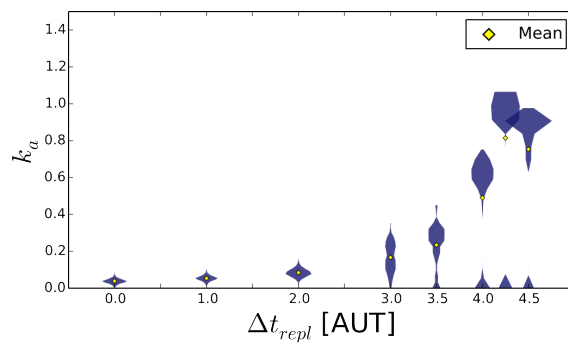The evolutionary steady state value of k_a in Fig. 2.17 is larger than in Main Text Fig. 2.1a because individuals die much more frequently. However, we do not observe qualitative differences in evolutionary steady state behaviour, i.e. no speciation occurs.

2.6.9 Longer replication time leads to speciation

The evolutionary separation of two lineages (one of which behaves as parasite) is observed for sufficiently duration of replication ($\Delta t_{repl} \leq 3.5$ AUT), as shown in Fig. 2.18.

2.6.10 Evolutionary transients during wave phase transition

When both β and k_a can mutate, we observe that evolution maximises both of them for $\Delta t_{repl} = 0$. Starting from smaller values of β and k_a, the system dynamically undergoes the phase transition between chaotic and stable waves. As lattice size must be large enough to contain the stable waves, spatial heterogeneities may lead to undergoing such phase transition locally, so that stable waves form in one part of the lattice (k_a and β increase) but not in another. Stable waves, however, are outcompeted by chaotic waves, because replicators in the latter can behave as parasites of the replicators in the former. Therefore, several "evolutionary attempts" may occur for a successful phase transition to actually happen at $\Delta t_{repl} = 0$, as shown in Fig. 2.19.

2.6.11 Limit behaviour for the generation of new waves

When β is very large, the longer-term stability of waves is threatened by the parasitic erosion of the invasion front. However, as the invasion front becomes narrower, small groups of replicators can be isolated from the side of the older wave, and establish a new one. Fig. 2.20 shows that this process selects for lower parasite strength because more waves are born from the sides of older waves when parasites are weaker, and constitutes the limit behaviour of the selection pressure that leads to chaotic travelling waves.

2.6.12 Indefinite persistence for $\Delta t_{repl} = 4.5$ if $\beta > 1$

When $\Delta t_{repl} = 4.5$, parasites may drive themselves to extinction because they are selected to become worse templates than replicators (β evolves to lower than 1, see main text). The consequent loss of spatial patterns may lead to the extinction of replicators as well. Instead, both replicators and parasites persist indefinitely if parasites are not allowed to decrease β below 1, because travelling wave dynamics sustain the two species, as shown in Fig. 2.21.
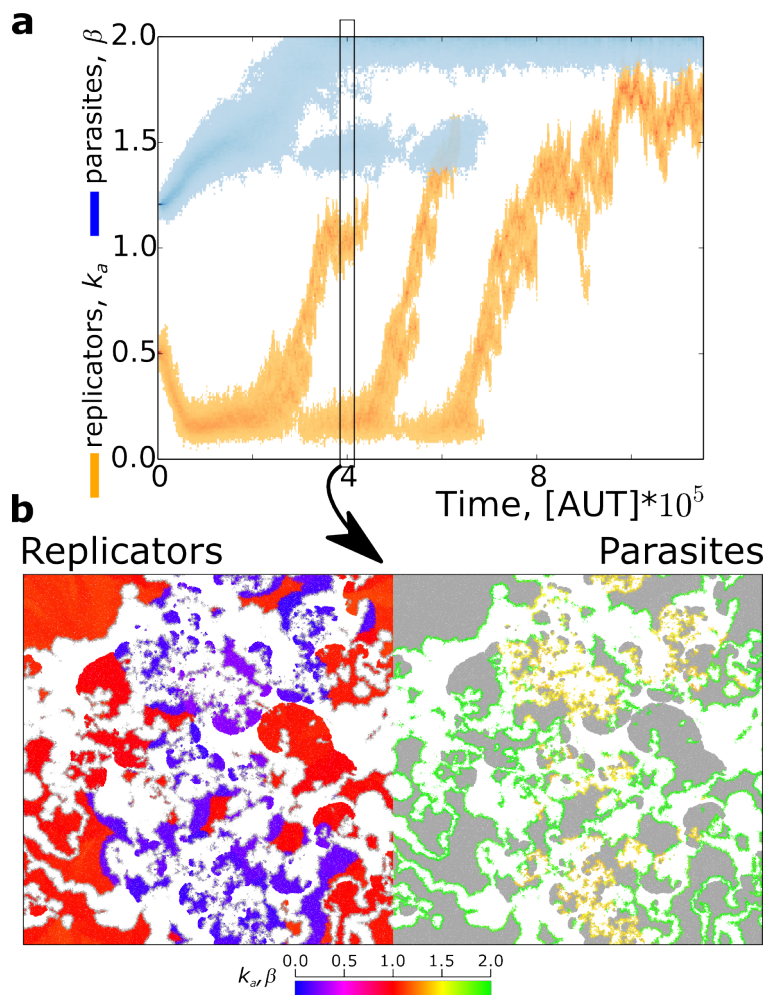
Figure 2.19. Repeated "evolutionary attempts" may be necessary to undergo the wave phase transition due to wave-level competition. **a**:Co-evolutionary dynamics of k_a and β for $\Delta t_{repl} = 0$. **b**: spatial distribution of replicators and parasites.
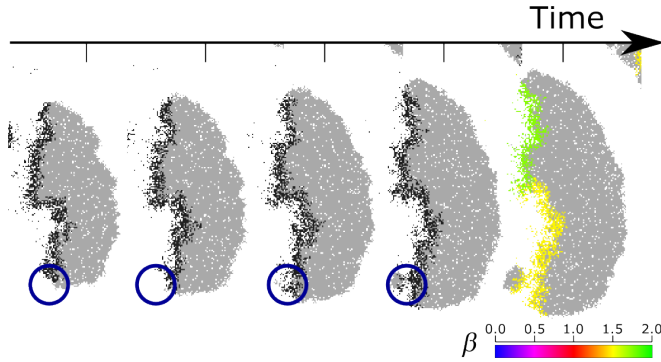
Figure 2.20. New waves are born from the side of older ones, when parasites are weaker. The same portion of the lattice is displayed at subsequent time steps (each 50 AUT). The circles highlight where the new wave is formed relative to the older one (all circles are at the same coordinates). Parameters as in main text.
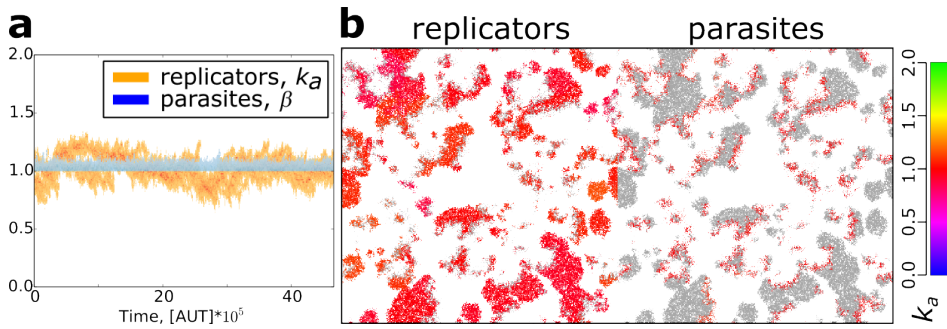
Figure 2.21. Spatial patterns allows indefinite persistence of replicators and parasites for $\Delta t_{repl} = 4.5$ when β cannot evolve below 1. Other parameters as in main text.
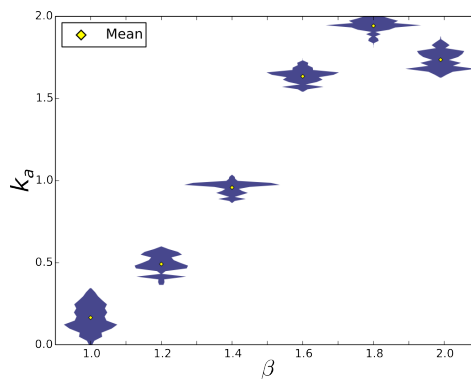
Figure 2.22. When only replicators mutate with longer replication times ($\Delta t_{repl} = 3$), they evolve to larger k_a in response to larger parasite advantage β.

2.6.13 The effect of longer replication time is lost when only replicators evolve

Fig. 2.22 shows that when parasites are not allowed to co-evolve with replicators under conditions of longer replication time ($\Delta t_{repl} = 3$), we obtain qualitatively the same results of Main Text Fig. 2.2, i.e. that evolutionary steady state k_a increases with parasite strength. Notice that, even though we did not extensively test for the presence of a phase transition, the evolutionary maximisation of k_a is not observed in the co-evolutionary model.

2.6.14 The evolution of parasites and previous results

In earlier work, it has been shown that parasites can evolve to lower probabilities of complex formation with a model similar to ours Takeuchi and Hogeweg (2009). In the main text we have seen that replicators can evolve to both higher or lower probabilities of complex formation (k_a) depending on parasite strength. Moreover, we have shown that replicator-parasites co-evolution can evolve in two different ways. Here we re-examine the results of Takeuchi and Hogeweg (2009) in the light of our results, and show that parasites also evolve to higher or lower complex formation rates as response to, respectively, weaker or lower replicators.

We set k_a constant and let β evolve. First we present results for $\Delta t_{repl} = 0$ and later we show how this compares to the case of $\Delta t_{repl} > 0$.

The evolution of β depends on k_a and on spatial pattern formation   Fig. 2.23 shows that an evolutionary stable steady state exists for β when replicators' k_a is set at lower values (for $\Delta t_{repl} = 0$). This is the same result as in Takeuchi and Hogeweg (2009). However, we find that parasites evolve to increase β when association rate of replicators is larger.

Importantly, to recover the results from Takeuchi and Hogeweg (2009), we had to set the lattice boundaries to fixed (i.e. individuals disappear when they cross the edge of the lattice). In fact, stronger parasites organise in larger waves, which need more space to unfold properly. Therefore, fixed boundary conditions may confound results because larger waves are more likely to "fall out" of a lattice than smaller ones.

Because implementing larger lattices is computationally prohibitive, we turned boundary conditions to wrapped, so that larger waves did not disappear when they reached the boundaries, but instead re-entered from the opposite side of the lattice. We find that the parasites speciate and organise in both chaotic and stable travelling waves. While stronger parasites are more efficient at invading replicators (with which they form larger patterns), new waves are created from older ones where parasites are weaker (Fig. 2.24a).

Thus, we observe co-existence Fig. 2.24b. Notice that the lineage with large β goes extinct because the associated waves become too large, and wrapped boundary conditions are no longer sufficient to counteract the effects of a small lattice. After extinction, evolution of a lineage which increases β occurs again.

Parasite evolution for longer replication time span For the sake of completeness, we repeated the analysis above for larger values of Δt_r (we tested $\Delta t_r = 1.5$ and 2.5). Results are in general agreement with above, i.e. parasites evolution leads to larger β when replicators' k_a is larger (Fig. 2.25). There are three points to make. First, before the phase transition, increasing k_a leads to a lower steady state value of β, in contrast to the case where replicators evolve with fixed parasites (compare with Main Text Fig. 2.2 and Supplementary Fig. 2.22). Second, the phase transition occurs to larger value for larger value of k_a when Δt_{repl} is increased. Third, we observe long term co-existence of two parasite species for longer replication time span and large k_a (Fig. 2.25b, $k_a = 1.99$).

Source code The source code used to run the simulations can be found at the publisher's website (`dx.doi.org/10.1371/journal.pcbi.1004902`).
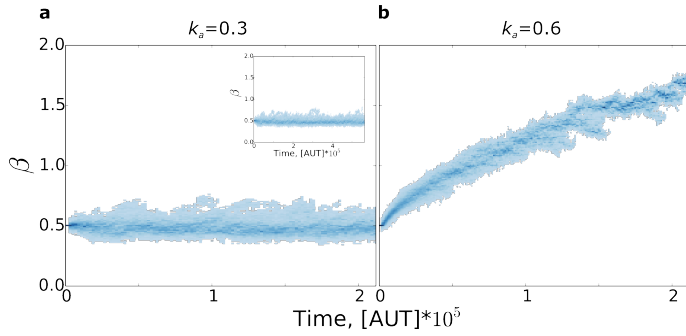
Figure 2.23. Parasite evolve to two different evolutionary steady state when replicators are weaker or stronger, and boundary conditions are fixed. **a**: Evolution of β while replicators have constant $k_a = 0.3$, inset: long term evolution; **b**: same, but $k_a = 0.6$. Other parameters as in main text.
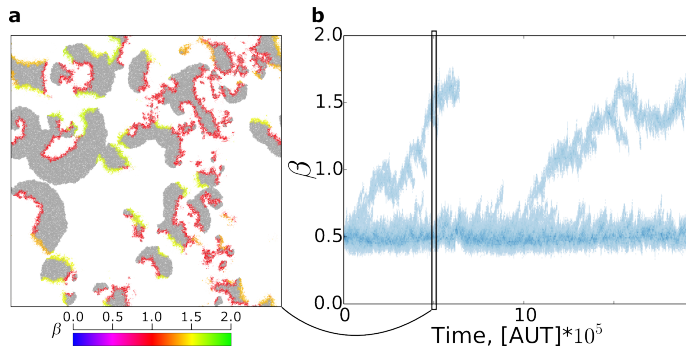
Figure 2.24. Co-existence of two lineages of parasites, mediated by the dynamics of different spatial patterns, for wrapped boundary conditions. **a**: A snapshot of the lattice. **b**: Evolution of β while replicators have constant $k_a = 0.3$. Other parameters as in main text.
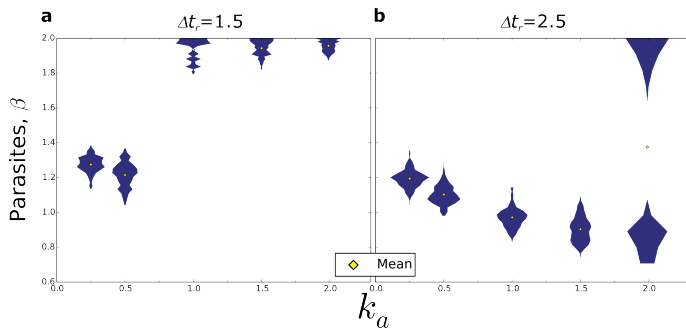
Figure 2.25. Parasite reach alternative evolutionary steady states depending on replicators' strength also larger replication time spans. **a**: $\Delta t_{repl} = 1.5$. **b**: $\Delta t_{repl} = 2.5$. Other parameters as in main text

3

High cost enhances cooperation through the interplay between evolution and self-organisation

Enrico Sandro Colizzi, Paulien Hogeweg (2016)

Abstract

Background   Cooperation is ubiquitous in biological systems, yet its evolution is a long lasting evolutionary problem. A general and intuitive result from theoretical models of cooperative behaviour is that cooperation decreases when its costs are higher, because selfish individuals gain selective advantage.

Results   Contrary to this intuition, we show that cooperation can increase with higher costs. We analyse a minimal model where individuals live on a lattice and evolve the degree of cooperation. We find that a feedback establishes between the evolutionary dynamics of public good production and the spatial self-organisation of the population. The evolutionary dynamics lead to the speciation of a cooperative and a selfish lineage. The ensuing spatial self-organisation automatically diversifies the selection pressure on the two lineages. This enables selfish individuals to successfully invade cooperators at the expenses of their autonomous replication, and cooperators to increase public good production while expanding in the empty space left behind by cheaters. We show that this emergent feedback leads to higher degrees of cooperation when costs are higher.

Conclusions   An emergent feedback between evolution and self-organisation leads to high degrees of cooperation at high costs, under simple and general conditions. We propose this as a general explanation for the evolution of cooperative behaviours under seemingly prohibitive conditions.

3.1   Background

When cooperation is costly to the individual but its benefits are equally shared in a group, one would expect progressively more selfish behaviour to be selected. This indeed happens when interactions between cooperative and selfish individuals are random. Instead, cooperation can be selected in a population when the interactions among cooperative and selfish individuals are structured, be it genetically, spatially or socially (Hamilton, 1964, Mitteldorf and Wilson, 2000, Nowak and May, 1992, Santos and Pacheco, 2005, Van Baalen and Rand, 1998, Wilson, 1975). Population structure favours cooperation when it allows for cooperators to be in contact with each other more frequently than with selfish individuals (Fletcher and Doebeli, 2009).

The result is a (locally or globally) stable equilibrium configuration in which cooperators persist indefinitely, and selfish individuals may co-exist. For instance, this can be due to spatial clustering of cooperators (Killingback *et al.*, 1999), or to an inherent structure of the interaction network (Ohtsuki *et al.*, 2006). Then, the conditions for selfish individuals invading and overtaking a group of cooperators represent the limit to the stability of the solutions found. In general, these conditions state that higher costs of cooperation increase the selective advantage of selfish individuals.

While spatial structure alone can favour cooperators due to population viscosity, a growing body of experimental and theoretical work indicates that self-organised spatial patterns may have profound and complex effects on cooperative interactions, due to emergent heterogeneities in the local distributions and densities of cooperators and selfish individuals (Datta *et al.*, 2013, Korolev, 2013, Momeni *et al.*, 2013, Müller *et al.*, 2014, Nahum *et al.*, 2011, Smaldino *et al.*, 2013, Suweis *et al.*, 2013, Szolnoki *et al.*, 2014, Van Dyken *et al.*, 2013, Wakano *et al.*, 2009). For instance, as a population of cooperators invade empty space, its expansion front can be enriched in altruistic individuals, while selfish individuals lag behind (Datta *et al.*, 2013, Korolev, 2013, Van Dyken *et al.*, 2013). Alternatively, mutualistic interactions that are favoured in a resident population can be automatically broken on such expansion front (Müller *et al.*, 2014). Furthermore, spatial self-organisation can sort cooperative strains from selfish ones (Boerlijst and Hogeweg, 1991c, Hogeweg, 1994, Hogeweg and Takeuchi, 2003, Momeni *et al.*, 2013, Szabó *et al.*, 2002, Takeuchi and Hogeweg, 2009), thus limiting the spread of the latter.

The customary approach to study the stability of cooperation (under a specific set of assumptions) consists of fixing the strategies of the interacting cooperative and selfish individuals, and analyse the population dynamics of the system (e.g. whether cooperators and defectors coexist, or if one lineage outcompetes the other). Because such pre-determined strategies do not mutate over time, their evolutionary stability remains unexplored.

Although exceptions to this approach exist (e.g. Doebeli *et al.* (2004), Killingback *et al.* (1999, 2006), Szabó *et al.* (2002)), little is known about what strategies evolve (by mutation and selection) and how they feedback on the spatial self-organisation of a population, even though it is clear that spatial self-organisation affects the population dynamics of

cooperative traits (see examples above). Here, we seek to study this feed-back between evolution and self-organisation with a minimal model where individuals can evolve the degree of cooperation in a spatially extended system.

We model the cooperative trait in terms of public good production (inspired by social dynamics in microbes (Crespi, 2001, West *et al.*, 2007)), and we let the amount of public good produced mutate in a continuous fashion. Thus, we can study the long term evolutionary dynamics of cooperation without preconceiving the extent to which individuals cooperate or defect. As we will show, selfish individuals that produce zero public good evolve readily at higher costs, and quickly invade cooperators. Rather than leading to global extinction, this enables cooperators to thrive and selects for a higher degree of public good production over evolutionary time scales.

3.2 Results

The model is a straightforward implementation of a population in which individuals replicate depending on the amount of public good produced in their close neighbourhood.

Individuals are embedded on a lattice. They may reproduce, die or move (locally). Competition for reproduction into neighbouring empty nodes is based on fitness, calculated as the difference between benefits and costs (Fig. 3.1). An individual benefits from the public good produced in its neighbourhood, but pays a cost for producing it. Thus, public good production is a cooperative trait. We assume that reproductive success is solely based on cooperation, so that individuals do not reproduce if public good in their neighbourhood is insufficient. Mutations slightly change the offspring's production rate (see Methods for details).

High cost leads to the evolution of larger public good production We set the benefits per unit of public good b $= 10$, and we let the spatial self-organisation and the evolutionary dynamics unfold under different costs c.

When costs are much smaller than benefits (c ≤ 1.5, Fig. 3.2), the public good production steadily increases because an individuals' own production increases its fitness, rather than decreasing it. Moreover, because replication is a local process, mutants with higher than average production rates benefit from each other due to limited dispersal (Fletcher and Doebeli, 2009, Killingback *et al.*, 1999, Ohtsuki *et al.*, 2006), outcompeting more selfish lineages. Thus, cooperation is maximised in the long run. For increasing costs, public good production suddenly drops ($2 \leq$ c ≤ 3, Fig. 3.2). In this regime, the clustering advantage of cooperators is insufficient and more selfish individuals replicate the most because, by producing less, they pay a lower cost. Eventually, public good production stabilises at the minimum value needed for survival.
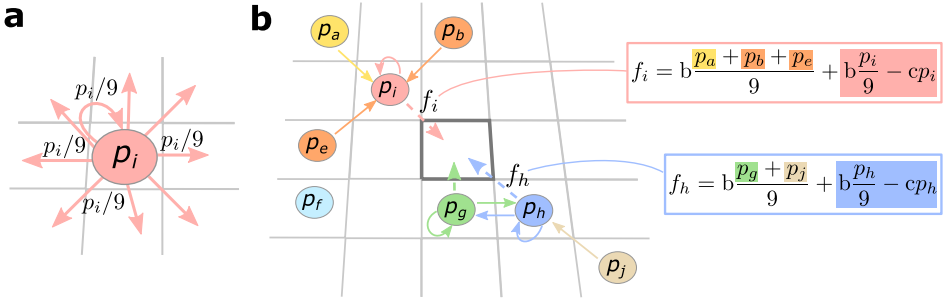
Figure 3.1. The model. **a** The world is a square lattice with connectivity $k = 8$ (every node has 8 neighbours) and wrapped boundary conditions. Individuals produce public good at rate p, shared in equal parts ($p/9$) among all neighbouring nodes and self. **b** Individuals compete for reproduction into an adjacent empty spot. Probability of reproduction depends on fitness f_i, which is the difference between benefits and costs. The sum of the public good an individual collects from itself and the neighbours, if any, is $p_{sum} = (p_a + p_b + ... + p_i)/9$, and confers a benefit b·p_{sum}. Individuals pay a fitness cost proportional to the public good they produce: $c \cdot p$. Successful reproduction yields a copy of the selected individual. Mutations occur with probability μ and change the public good production rate by a small random number chosen uniformly in the interval $[-\delta/2, \delta/2]$. Individuals have a small probability k_{move} to move to a random adjacent node, and can die with probability k_{death}, leaving the node empty (See Methods for the details of the models).

Strikingly, further increasing costs leads to an increase in cooperation (c > 3, Fig. 3.2, see Supplementary Section 3.6.1 for the full snapshots of the system). The distribution of public good production is bimodal at evolutionary steady state, with most of the population having higher rates of public good production and a minority producing almost no public good at all.

The long term evolutionary dynamics of cooperation at high cost   Following the evolutionary and the spatial dynamics of a single case elucidates why cooperation increases and persists for higher costs (c = 4.5, Fig. 3.3, video at (Colizzi and Hogeweg, 2015)). When we initialise the system with highly producing individuals, public good production decreases rapidly due to strong selection for selfishness (compare Fig. 3.3b and c). Where public good production drops below the minimum for survival, large patches of individuals go extinct (Fig. 3.3b). The surviving individuals can expand into the empty space (Fig. 3.3c). As the expansion progresses, a selfish and cooperative lineage separate from each other (Fig. 3.3d). The selfish strain evolves to zero public good production, becoming incapable of autonomous persistence and relying on the public good produced by cooperators for survival.

While the two strains differentiate from each other, they organise spatially to form travelling waves (Fig. 3.3e and 3.3f, similar to (Hogeweg and Takeuchi, 2003, Takeuchi and Hogeweg, 2009)). Cooperators constitute the front of a wave, and expand into empty
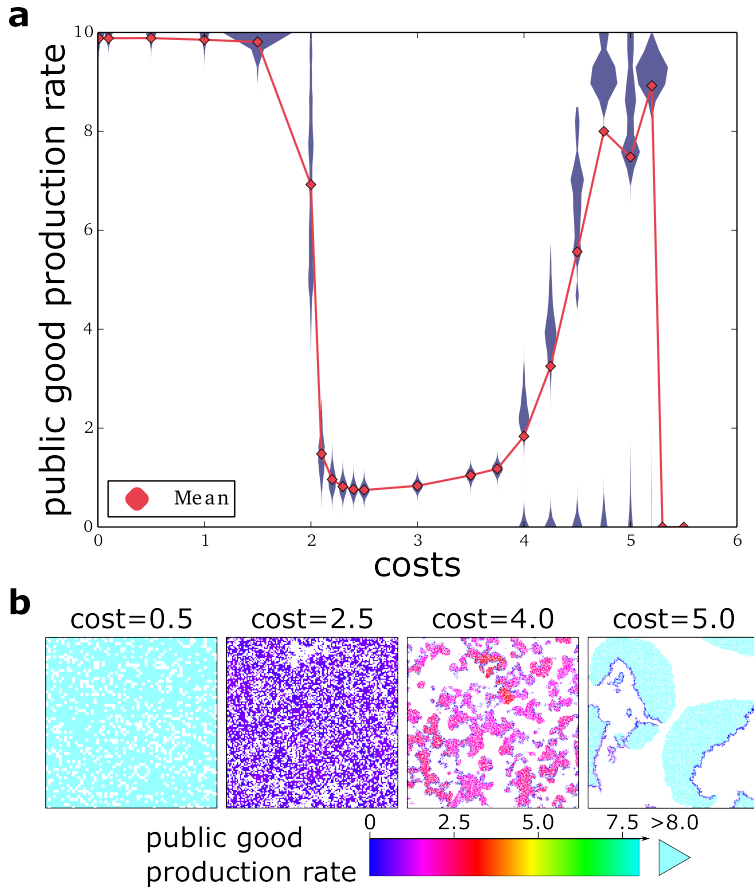
Figure 3.2. An increase in costs results in an evolutionary increase in cooperation. a Evolutionary steady state distribution (in blue) and mean (red diamond) of public good production are plotted for different values of costs (benefits are kept constant). Parameters: benefit (per unit of public good produced) $b = 10$, $k_{death} = 0.2$, $k_{move} = 0.02$, $\mu = 0.05$, $\delta = 0.1$. The maximum public good production is set to $p_{max} = 10$. **b** Snapshots of the lattice at evolutionary steady state. Colour coding depends on public good production rate. White is background. Lattice dimensions used for the simulations from left to right: 256^2, 512^2, 2048^2, 2048^2 (1/16 of the lattice is displayed for clarity, see Supplementary Section 3.6.1 for the full snapshots).
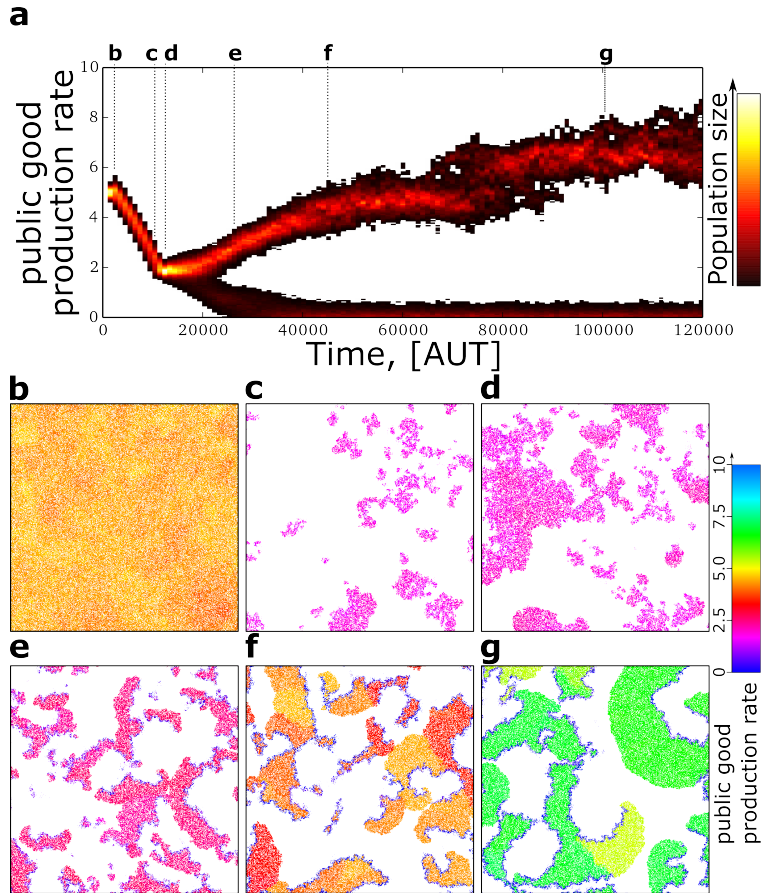
Figure 3.3. Evolutionary dynamics of public good production a At each time point the distribution of public good production in the lattice is plotted as a heat map. **b-g** Snapshots of the lattice at subsequent time steps (letters correspond between time plot and snapshots). Costs (per unit of public good) c = 4.5, other parameters and colour coding in the snapshot as in Fig. 3.2. Lattice size = 2048^2 (1/16 of the lattice is displayed for clarity). Time units are Monte Carlo steps. See also Supplementary Section 3.6.2 for the full snapshots, and video at (Colizzi and Hogeweg, 2015).

space by replicating into it (generation after generation); selfish individuals invade those cooperators, and constitute the back of the wave. Selfish individuals leave empty space behind a wave after they die, causing the semblance of movement (See Fig. 3.4 and (Colizzi and Hogeweg, 2015) for videos). The progression of a wave, however, happens on a time scale that is much longer than the life time of an individual, which in turn experiences a fairly constant environment throughout its life. Because waves persist longer than individuals, they can integrate information over several generations.

New waves are "born" from the collisions of older waves. As cooperators on the front evolve to larger public good production, waves become larger (compare Fig. 3.3e,f and g, and Supplementary Section 3.6.2 for the full snapshots). The formation of spatial patterns allows the populations in the system to persist indefinitely despite selfish individuals continuously invading cooperators (Fig. 3.3g and Supplementary Section 3.6.6), provided that the lattice is much larger than the spatial patterns (see Supplementary Section 3.6.3).

Spatial population dynamics of cooperators and selfish individuals   Spatial self-organisation drives the evolution of cooperation in the system. When spatial patterns are destroyed, e.g. by mixing, only selfish individuals are selected and public good production decreases, leading to global extinction (Supplementary Section 3.6.4), in accordance with the result that random interactions favour selfish behaviour.

To unravel the interplay between the two lineages and their spatial organisation, we analysed the spatial population dynamics for cooperative and selfish individuals separately. To this end, we shaped the lattice into a long, narrow strip, and set the mutation rate to zero (see Material and Methods for details). Cooperators expand faster into empty space when they produce more public good, and slower when costs are higher (Fig 3.5a, red). When two clustered populations compete at the expansion front (Fig 3.5c), the one with the largest public good production wins because, by replicating faster, it occupies space before the competing one and eventually overtakes the entire wave front.

The replication rate of selfish individuals invading a population of cooperators is higher when the cost of public good production is higher, and it is insensitive to how much public good is produced (Fig 3.5a, blue). Clearly, when two strains compete in the back of a wave, the winner is the more selfish one (Fig 3.5b).

The picture emerging from these experiments is that different selection pressures operate depending on the spatial context: a population expanding into empty space (the wave front) is selected for higher degrees of cooperation (in agreement with Datta *et al.* (2013), Korolev (2013), Momeni *et al.* (2013), Van Dyken *et al.* (2013)), competition in the back (behind the wave front) selects for more selfishness. Importantly, even though costs, benefits and fitness function are the same, spatial pattern formation automatically segregates these two opposing evolutionary pressures to spatially different contexts so that they do not balance each other: hence the evolution of a cooperative and a selfish lineage.
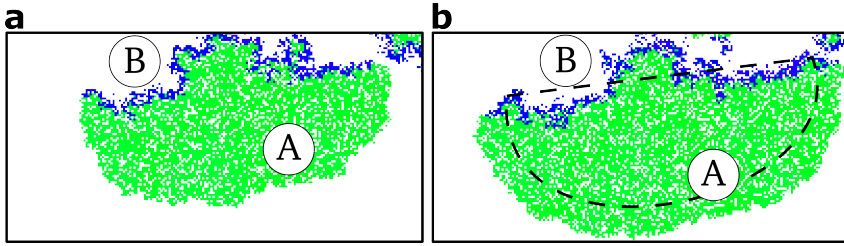
Figure 3.4. The movement of a travelling wave. Magnification (200x100) of the same portion of the lattice at 10 time steps distance. A) the wave-front, composed of cooperative individuals (in green), B) the back, composed of selfish individuals (in blue). The snapshots are from the same simulation run as in Fig 3.3, at time steps 400000 (left) and 400010 (right). The dashed line in the right pane marks approximately the position of the wave in the left pane. Reference (Colizzi and Hogeweg, 2015) shows one such travelling wave in a video.
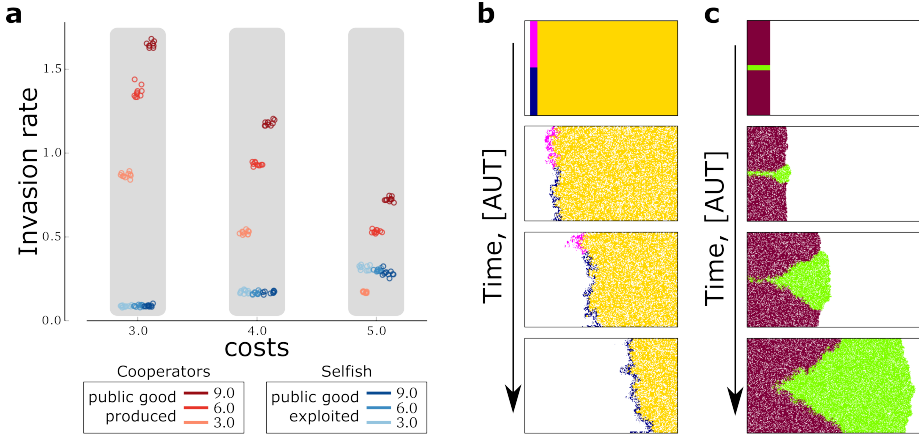
Figure 3.5. Invasion rates and spatial dynamics of competition for cooperative and selfish individuals. **a** Cooperators' invasion rate increases with larger public good production and lower costs; selfish invasion rate increases with increasing costs and it is insensitive to the production rates of the cooperators that support them. The invasion rates for cooperators invading empty space (red circles) and selfish individuals invading a population of cooperators (blue circles) was measured for each combination of cost and public good production rate (benefit is constant, and set to b = 10) in 10. **b** A population with lower public good production rate out-competes one with larger production in the back of a wave. Parameters: cooperators (yellow wave-front) $p_{yellow} = 6.0$, selfish individuals (magenta and blue wave-back) $p_{blue} = 0.2$ and $p_{magenta} = 1.0$. **c** The population with a larger production rate out-competes the one with lower production at the front of a wave. Parameters: $p_{purple} = 5.0$, $p_{green} = 6.0$. For both **a** and **b**, $\mu = 0$, b = 10, c = 4.0, background in white. Other parameters as in Fig. 3.2.

In the full system, cooperators and selfish individuals are ecologically and evolutionarily interdependent (Supplementary Section 3.6.5). The two lineages establish an evolutionary feedback mediated by their spatial organisation. Empty space is generated by selfish individuals after invading a population of cooperators. Therefore, the condition for increasing public good production, i.e. the availability of continuous empty space, is mediated by the invasion dynamics of the selfish lineage. With higher costs, selfish individuals propagate faster, and space is left empty at a higher rate. The larger the empty space, the more cooperators can increase public good production. This evolutionary feedback reaches an evolutionary steady state because highly producing cooperators reduce the empty space faster upon faster expansions (Fig 3.5a). We checked the long term stability of the steady state (Supplementary Section 3.6.6).

Altogether, cooperation evolves to a higher degree for higher costs due to an emergent feedback between self-organised interaction structures (the spatial patterns) and the evolution of the individuals composing them.

Robustness to parameter change   Our results are robust when death rates, movement rates and benefits are changed, provided that benefits-to-costs ratio is maintained (Supplementary Section 3.6.7). At lower costs, individuals directly benefiting themselves with their own public good are sometimes dubbed *weak* altruists (Fletcher and Doebeli, 2009, Wilson, 1979) (in our case, for $c < b/(k+1)$, with $k+1 = 9$ the connectivity of the lattice including self), whereas they are considered *strong* altruists when their public good is only shared among others (in the context of game theory, these situations are called, respectively, snowdrift game and prisoner's dilemma (Doebeli *et al.*, 2004)). In our spatial model, we observed no qualitative difference in the evolutionary dynamics when individuals did or did not benefit from their own public good (Supplementary Section 3.6.8). This could be expected because individuals' own payoffs at high costs are negative in both models. Indeed, in both cases the evolving populations underwent speciation of a selfish and a cooperative lineage.

Weak or strong altruism do make a difference in the corresponding well-mixed systems, where strong altruists go extinct at lower costs, while weak altruists maximise public good production (Supplementary Section 3.6.9).

3.3   Discussion

It is known that during population range expansion, cooperation can be promoted on the front of the expansion range (Datta *et al.*, 2013, Korolev, 2013, Momeni *et al.*, 2013, Van Dyken *et al.*, 2013). One could argue that in these models cooperation could evolve only as long as empty space is available, and should eventually be out-competed globally by selfish strategies when the invasion dynamics reach an end. Here we have shown that selfish individuals provide the empty space to allow continuous expansion within a

limited area.

Generally, stable solutions to the problem of cooperation are based on the condition that at equilibrium selfish individuals do not locally invade a population of cooperators. For instance, it is well known that cooperation can be stabilised in spatially extended systems, as cooperators cluster and segregate selfish individuals to the edges of those clusters (Doebeli and Hauert, 2005, Killingback *et al.*, 1999, Smaldino *et al.*, 2013). A side effect of these solutions is that higher costs undermines the stability of such clusters. In the parameter region where travelling waves do not form (Fig. 3.2, for c ≤ 3) our work is in agreement with those results in that there exists an inverse relationship between costs and cooperation (e.g. Hamilton (1964), Van Baalen and Rand (1998), Wilson (1975)), and in particular with the heuristics (Ohtsuki *et al.*, 2006) that cooperation evolves when the benefit-to-cost ratio is larger than the connectivity of the lattice ($b/c > k$, while it becomes progressively more unlikely when b/c approaches k).

However, we have shown that a novel class of solutions exists at high costs, where large degrees of cooperation are maintained in a locally out-of-equilibrium fashion, with selfish individuals always successfully invading cooperators and setting the stage for the evolutionary increase and the global stability of cooperation. We conclude that spatial self-organisation can reverse the relationship between costs and cooperation, thus extending the evolutionary viability of cooperation to higher costs.

Our results rest on two assumptions: population size can vary and some degree of cooperation is necessary for reproduction. Variable population size is obviously realistic, even though it is not often included in evolutionary models of cooperation. Although the assumption of necessary cooperation is not always met, it is reasonable in several cases. Examples in microbiology include, cooperative protection or cooperative virulence in bacterial infections (Diggle *et al.*, 2007, Stewart and Costerton, 2001); invertase production in yeasts while growing on sucrose (Carlson and Botstein, 1982); siderophore production in iron-limited environments (Luján *et al.*, 2015, Miethke and Marahiel, 2007, Schrettl *et al.*, 2004), cooperative secretion of digestive enzyme in microbial hunting (Crespi, 2001, Konovalova *et al.*, 2010). Outside the microbial world, situations where our model may apply are e.g. dangerous behaviours in cooperative nest defence (Olendorf *et al.*, 2004), and replication *in trans* in prebiotic evolution (Higgs and Lehman, 2014, Takeuchi and Hogeweg, 2012).

Two recent studies have come to conclusions that at first sight are similar to ours (Smaldino *et al.*, 2013, Szolnoki *et al.*, 2014). By making the assumption, as we do, that the lack of cooperation leads to death, they observe (quasi) static spatial patterns in which cooperation is maintained because despite relatively high costs, clusters of cooperators cannot be invaded by selfish individuals.

In contrast, cheaters can always invade in our system, and do so faster when costs are higher. This shows that costs are qualitatively higher in our model. Cooperation is maintained despite, and due to the evolution of true cheaters (in the sense of (Jones *et al.*, 2015)). Furthermore, we show that the amount of public good produced by an individual increases in evolutionary times, whereas the evolutionary stability of the solutions in (Smaldino *et al.*, 2013, Szolnoki *et al.*, 2014) is left unexplored, and only the long population dynamical transient is analysed.

More in general, the importance of spatial self-organisation for understanding the population dynamics of cooperators and defectors has been highlighted both from a theoretical (Wakano *et al.*, 2009) and from an experimental point of view (Penn *et al.*, 2012). Here we make a similar point, but with an evolutionary twist: in our case selfish individuals are not merely a burden to cooperators; instead, their emergence as a separate lineage is necessary for the evolution of high degrees of public good production because they generate the spatial conditions in which cooperators thrive and evolve.

Allowing mutations to change public good production in a continuous range resulted in the evolution of two separate strains, a selfish and a cooperative one. The evolution of stable heterogeneity in a population has been observed before in models of cooperation (Doebeli *et al.*, 2004, Koella, 2000, Szabó *et al.*, 2002). Here, besides stressing that the evolution of two lineages from a single ancestral one might be a rather general feature of models with variable investments (as very simple assumptions were needed, in contrast to (Doebeli *et al.*, 2004)), we make the case that true cheating behaviour (*sensu* Jones *et al.* (2015)) is actually functional and beneficial to the long term evolution of cooperation.

3.4   Conclusions

In conclusion, besides extending the theoretical limits of cooperation, our results broaden the search image of cooperative behaviour in nature by suggesting that there need not be a strict trade-off between costs and benefits; rather, a wider view of the self-organised eco-evolutionary processes must be taken into account to understand the occurrence of costly cooperation.

3.5   Material and Methods

General system   Our system is an individual-based, Monte Carlo simulation run on a square lattice with connectivity $k = 8$ and toroidal boundary conditions. The nodes of the lattice can be empty or occupied by at most one individual. Individuals produce public good with rate p per time step (alternatively, p can be considered the degree of

altruism of an individual). An individual i produces p_i public good per time step ($0 \leq p_i \leq 10$), which is divided equally among neighbouring nodes and self, each receiving $p_i/9$ public good. All n neighbours, in turn, share a fraction of the public good they produce $p_{\{1,2,\dots,n\}}/9$ with individual i. The benefit from the public good received from each neighbour and from self is $B_i = b \left(\frac{p_i}{9} + \frac{1}{9} \sum_{j=1}^{n} p_j \right)$, where b is the benefit per unit of public good. Individuals pay a cost proportional to the public good they produce $C_i = cp_i$. Public good is not accumulated over multiple time steps. The fitness of an individual is the difference between benefits and costs: $f_i = B_i - C_i$ (set to zero if costs exceed benefits).

Each Monte Carlo step, all nodes are updated in random order (although synchronous updating rules do not affect results). If a node is empty, the individuals in its neighbourhood (if any) compete for replication. Assume an individual i is competing with m other individuals, and let us name f_{tot} the sum of the fitness of all individuals competing for the same empty node, then $f_{\text{tot}} = f_i + \sum_{k=1}^{m} f_k$. Individual i is chosen for reproduction over its competitors with probability $P(i \text{ replicates}) = \frac{f_i}{f_{\text{tot}}} \left(1 - e^{-f_{\text{tot}}} \right)$. The term in parenthesis is the probability that at least an individual replicates, which models the assumption replication should be more frequent in a neighbourhood where there is more public good, and conversely it should be rare if little public good is produced. Notice that this term does not affect death.

Upon successful replication, mutations may happen with probability μ and affect p by adding a small random number drawn with uniform probability from the interval $[-\delta/2, \delta/2]$. If a node is not empty, with probability k_{move} its content is swapped with that of a randomly chosen adjacent node. Moreover, every non empty node can turn to empty with probability k_{death}. See Fig. 3.1 for a cartoon of the model and the caption of Fig. 3.2 for the actual values of the parameters. The algorithm is implemented using the CASH libraries (de Boer and Staritsky, 2000).

Invasion dynamics of cooperators and selfish individuals (Fig. 3.5)   We modified the system described above as follows: 1) we shaped the lattice into a narrow strip of arbitrary length; 2) we changed the boundary conditions to no-flux, and in particular we removed individuals when they moved or replicated into a boundary node of the lattice; 3) we set mutation rates to zero to better focus on spatial population dynamics. The rules for the local dynamics remained the same as above. We initialised all populations on one side of the lattice and waited until they reached the other side. For cooperators this meant that they invaded empty space, whereas selfish individuals invaded a resident homogeneous population of cooperators. In all cases, the number of Monte Carlo time steps it took for the first individuals to arrive to the other side of the lattice (generation by generation) was recorded. The invasion rate plotted in Fig. 3.5 was calculated as the length of the space invaded divided by the time it took for the population to invade it.

3.6 Supplementary Material

3.6.1 Full snapshots of the system at evolutionary steady state for different costs

The snaphsots of the lattice at evolutionary steady state for different costs (main text, Fig. 3.2) are magnified to highlight the spatial patterns. The full snapshots are presented here in Fig. 3.6 and the correct proportions between different lattices are maintained. In the parameter conditions where spatial patterns do not form, i.e. $c \leq 2.5$ a smaller lattice was used (because simulations are computationally expensive). A larger lattice was used for larger costs (see Supplementary Section 3.6.3 for the consequences of smaller lattice with large spatial patterns).

3.6.2 Full snapshots of the evolutionary dynamics of public good production

The snaphsots of the lattice at evolutionary steady state for different costs (main text, Fig. 3.3) are magnified to highlight spatial pattern formation. The full snaphsots are presented here in Fig. 3.7, 3.8, 3.9 (black squares mark displayed portion in main text).

3.6.3 Evolutionary consequences of small vs. large lattice

Because spatial pattern dynamics play an important role, the size of the lattice must be large enough to accommodate them. When costs are high relative to benefits, the spatial scale of pattern formation may become comparable to the lattice size. When this is the case, extinction is often observed as a consequence of wave-wave collision (see snapshots in Fig. 3.10), or because the front of a wave touches its own back (Fig. 3.11, in both cases $c = 5.3$, recall that lattice has wrapped boundary conditions). Hence, results in the main text may be extended to higher costs if lattice size were larger (at the expense of computational load). However, the evolutionary mechanism by which cooperation increases in the system remains the same as that reported in the main text.

3.6.4 In a well-mixed system at high costs, only selfish behaviour is selected

In the main text, we have shown that cooperation increases for larger costs. The spatial organisation of the cooperative and the selfish lineage play a fundamental role in this. In Fig 3.12 we show that for high costs in well mixed conditions (where spatial patterns are
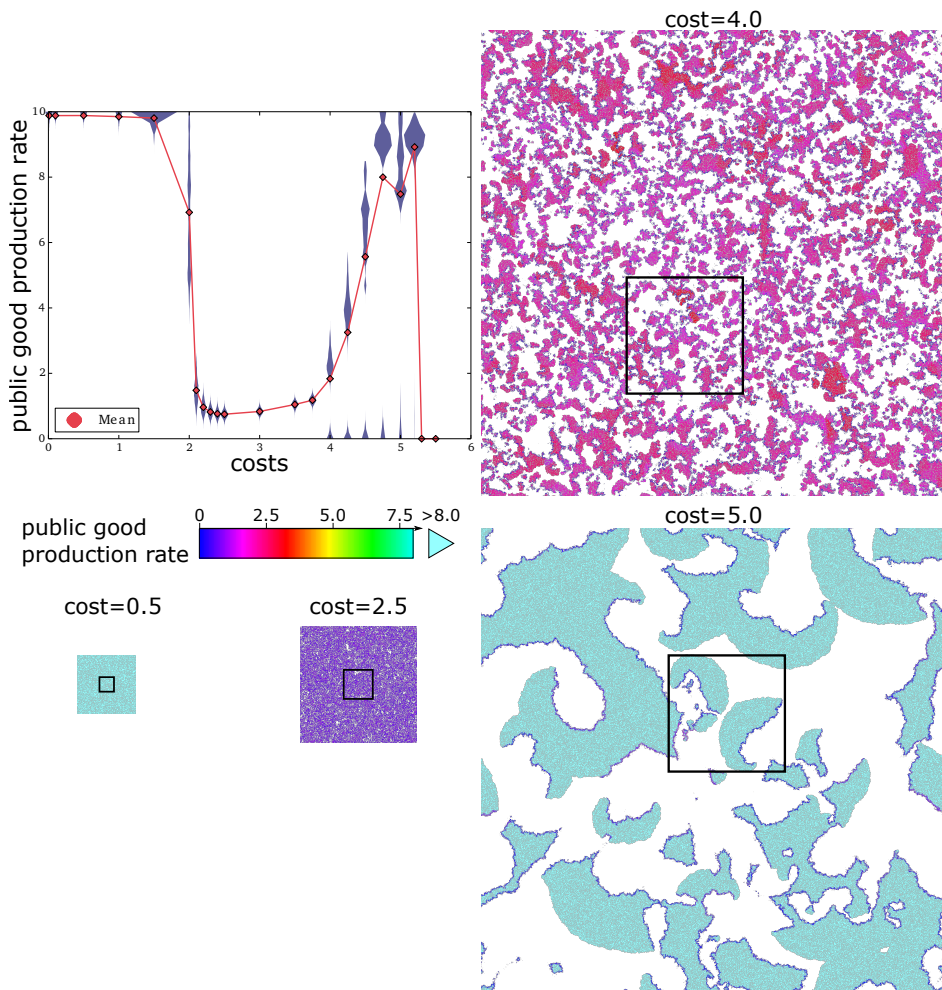
Figure 3.6. Evolutionary steady state of public good production for different costs (same figure as Fig.3.2). The snapshots depict the full lattice for $c = 0.5$, $c = 2.5$, $c = 4.0$ and $c = 5.0$. The black square marks the part of the snapshot that is displayed in Fig. 3.2.
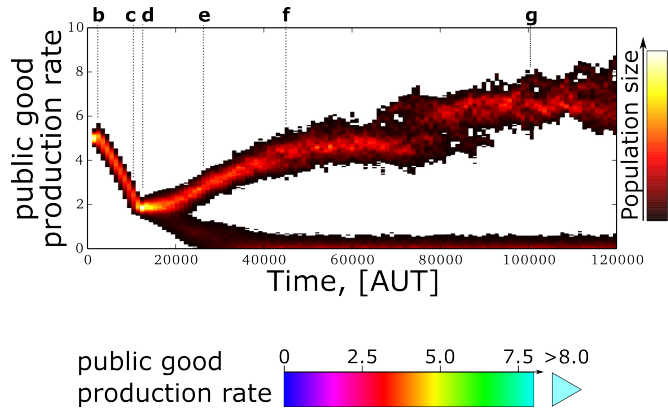
Figure 3.7. Evolutionary dynamics of public good production for $c = 4.5$ (same figure as Fig.3.3). This figure is meant for reference to the following.
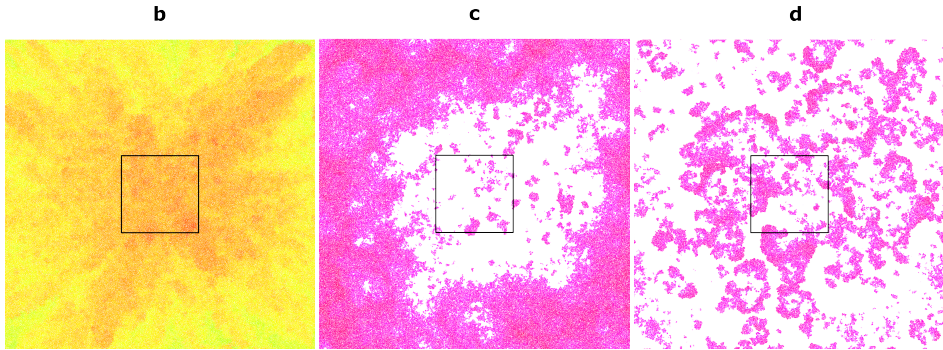
Figure 3.8. Snapshot of the full lattice at time point **b**, **c** and **d** of Fig. 3.7.
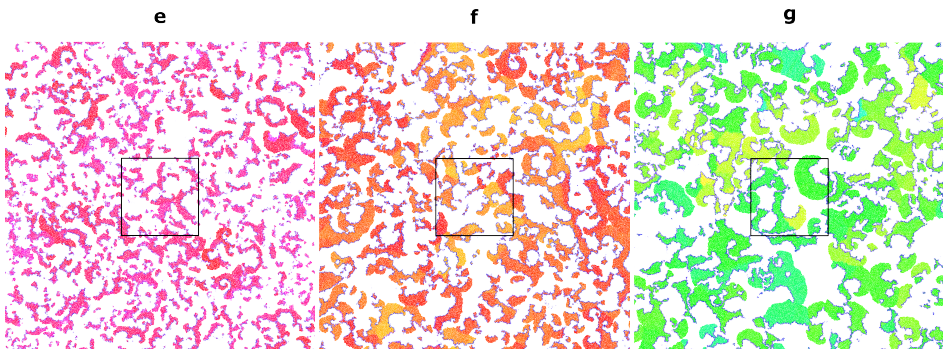
Figure 3.9. Snapshot of the full lattice at time point **e**, **f** and **g** of Fig. 3.7.
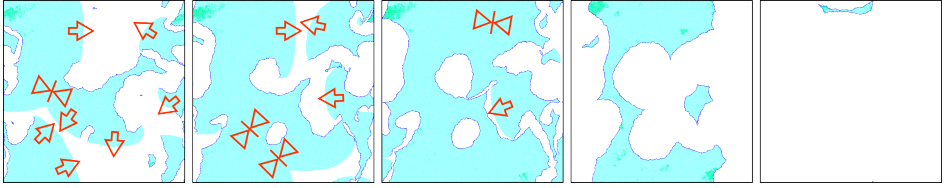
Figure 3.10. Giant wave-wave collisions may lead to extinction with high costs. Arrows indicate the direction of expansion of the waves, the ribbon-like symbols indicate where two wave-fronts have collided. Costs = 5.3
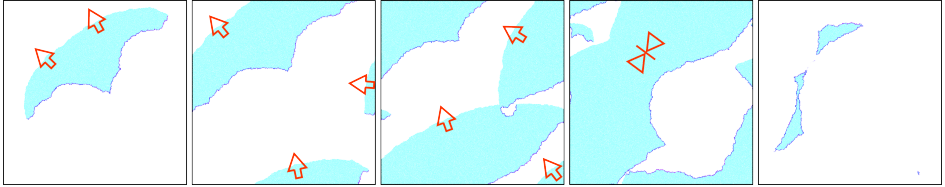
Figure 3.11. Giant single waves may touch their own back, letting selfish individuals invade the cooperators in the front. Under high costs this may lead to extinction. Arrows indicate the direction of expansion of the wave, the ribbon-like symbols indicate where two sides of the same wave-front collided on each other. Costs=5.3

destroyed), evolution favours selfishness. Public good production rapidly decreases in the system until it reaches values that are too small for survival. There, global extinction ensues.
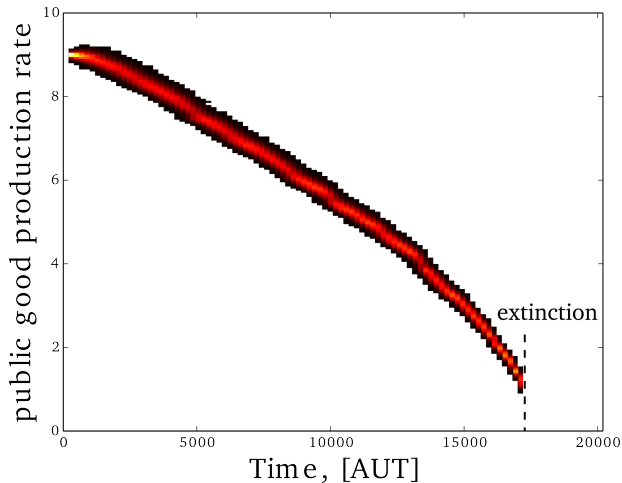
Figure 3.12. In a well mixed system with costs $c = 4.0$ cooperation is always minimised until extinction. Other parameters as in main text Fig. 3.3.

85

3.6.5   Consequences of removing cooperative or selfish individuals

We tested directly the interdependence between cooperative and selfish species by re-moving one species or the other from the system. In practice, we continued a simulation with high costs ($c = 4.5$) and after 10000 AUT we removed all individuals with public good production rate either larger or smaller than $p = 1.5$.

Removing all cooperators results in an almost instantaneous extinction of the selfish in-dividuals (data not shown).

Removing all selfish individuals at one time point (Fig. 3.13a) results in a quick filling of the lattice with cooperators. With no selfish individuals to generate empty space, only individual-based selection for selfishness is present, and we observed a global decrease in public good production which leads to the same evolutionary dynamics observed in Fig. 3.3 of the main text. Eventually both the cooperative and the selfish species are restored and the system reaches the evolutionary state it had before removing all selfish individuals.

As a complementary approach, we ran a simulation and continuously removed selfish individuals, i.e. any individual with production rate $p < 1.5$ (costs are set to $c = 4.5$, as before). Removing selfish individuals does not allow for waves to form properly, which inhibits the evolutionary feedback between evolution and self-organisation described in the main text. Hence, we do not expect the evolution of large degrees of cooperation. This is confirmed by the results in Fig. 3.13b.

3.6.6   Long term stability of the evolutionary steady state

The evolutionary dynamics of cooperators can be very noisy in the short run (Fig 3.14, upper pane). Mutliple lineages separate and persist long enough to evolve to different degrees of cooperation (public good production rates between 4 and 10). Nevertheless, we confirmed the long term stability of the steady state reached by the evolutionary dy-namics by letting the system (shown in Main text Fig. 3.3) run much longer than the time scale needed to reach such steady state (Fig 3.14, lower pane).

3.6.7   Moderate parameter changes do not affect results

Death rate   Individuals have an average life span of 1/[death rate] $= 1/0.2 = 5$ AUT. Decreasing the death rate to 0.1 does not change results (Fig. 3.15, top pane).

Figure 3.13. a: Removing all selfish individuals (at time 10000 AUT) results in a new evolutionary cycle which ultimately restores cooperative and selfish individuals. **b**: Continuously removing selfish individuals inhibits the evolution of cooperation.

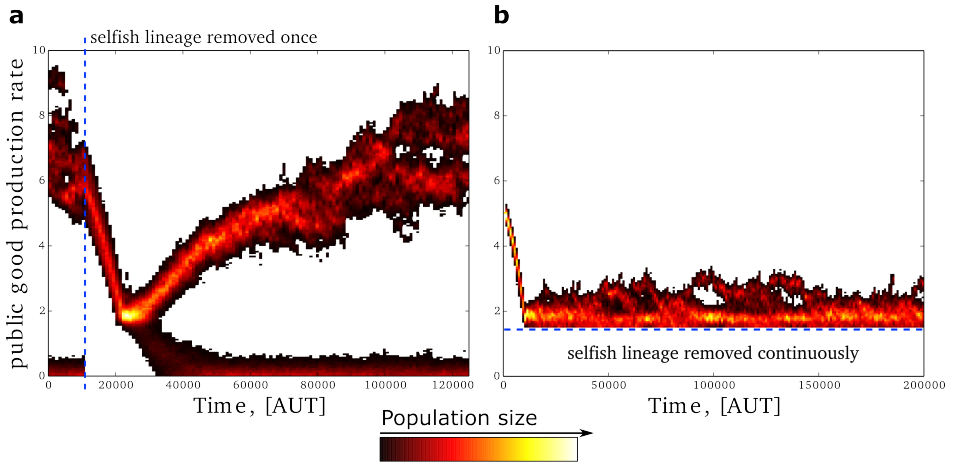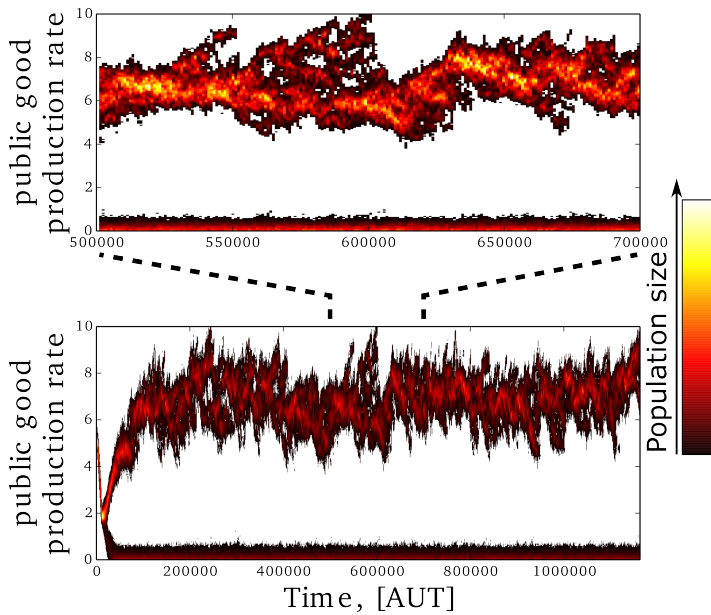**Figure 3.14.** Short vs. long term evolutionary dynamics after reaching evolutionary steady state. The simulation is the same as in the Main text Fig. 3.3.

Movement Movement is implemented as a random diffusive process (see Material and Methods in the main text). The probability that individuals move is kept to very low values ($p_{\text{move}} = 0.02$ per individual, per time step) but not zero, so that spatial patterns do not get stuck in frozen configurations. This enhances the robustness of the observed spatial patterns. The stability of our results was checked by starting from backups and setting the rate of diffusion to 10 times higher (i.e. to $p_{\text{move}} = 0.2$). We analysed three data points, corresponding to the three behaviours observed in the system for low, intermediate and high costs. Results are shown in Fig. 3.15 (central pane), and are not qualitatively different. However, at higher costs ($c = 4.5$), we observe a much larger increase of public good production. A moderate increase in movement also increases the frequency with which other individuals are met. This constitutes an evolutionary advantage for selfish individuals, which translates in the generation of more empty space, thus giving cooperators a greater opportunity to expand into it.

The limit of very large diffusion is the well-mixed system. There, the advantage of selfish individuals becomes large, because spatial patterns cannot form. In this limit, cooperation breaks down at high costs, and for low costs with strong altruism (see Supplementary Section 3.6.9).

Different values for benefits and costs In the main text, the benefits are kept always constant ($b = 10$) and simulations are run for different costs. Results do not change qualitatively if benefits are increased, as long as the benefit-to-cost ratio is kept approximately as in the main text. We report results for $b = 20$ in Fig. 3.15 (bottom). For $c = 10$, the system shows large scale evolutionary fluctuations in public good production, with the selfish lineage occasionally producing more public good. However, the mechanism by which public good production increases remains the same as explained in the main text.

3.6.8 Strong or weak altruism do not differ qualitatively

Whether an individual contributes positively to its own fitness or not is sometimes referred to as, respectively, weak and strong altruism (Fletcher and Doebeli, 2009, Wilson, 1979). In our model, a fraction of public good remains by the individual that produced it. We can calculate the costs for which an individual has a positive fitness gain from its own public good production. The contribution to fitness an individual i gains from its own public good production is $f_{\text{self}} = bp_i/9 - cp_i$. For the benefits used in the main text, $b = 10$, f_{self} is positive only for $c \lesssim 1.1$. Thus we do not expect qualitative differences in the evolutionary dynamics for higher costs if we implemented strong altruism (where $f_{\text{self}} = -cp_i$, which can never be positive).

Fig.3.16 confirms our prediction. Results are qualitatively the same to those described in the main text (Fig. 3.2) if public good is shared among neighbours, but no fraction is given to self. The maximum cost for which the system is viable is presumably higher than $c = 4.5$. However, spatial patterns become very large at higher costs, and an adequate lattice
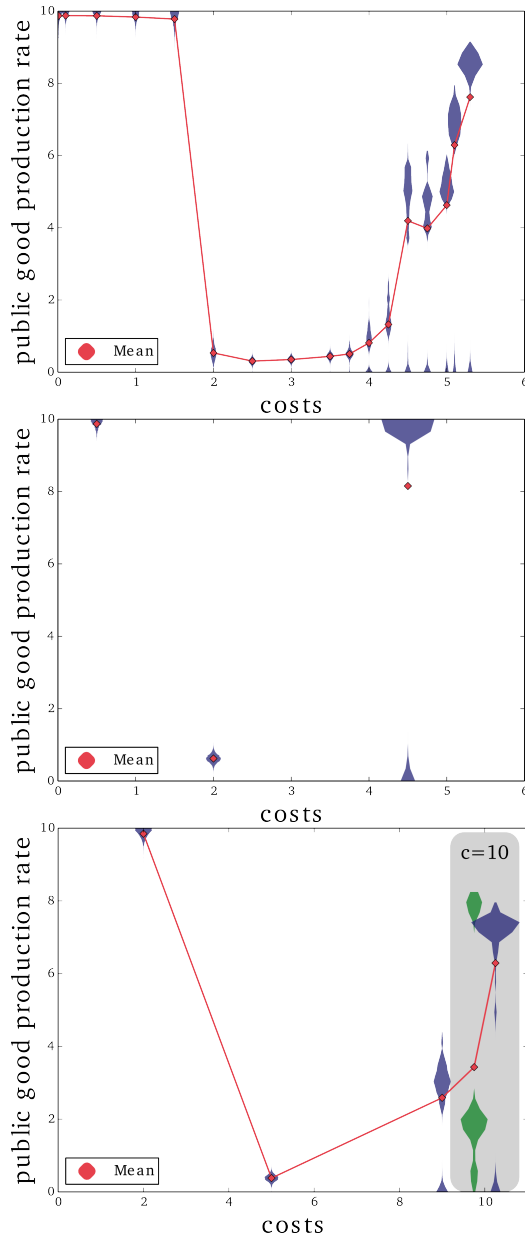
Figure 3.15. Top: Final distribution for $k_{\text{death}} = 0.1$. **Centre:** Final distribution for larger diffusion, $p_{\text{move}} = 0.2$. **Bottom:** Distribution for benefits $b = 20$, twice as much as those used in the main text after reaching evolutionary steady state. Both distributions in the shaded area correspond to c = 10 and are taken at different time points ($t_{\text{green}} = 450000$, $t_{\text{blue}} = 550000$ time steps). In all cases other parameters are identical to main text Fig. 3.2.

size render the simulation computationally infeasible (see Section 3.6.3 for a discussion on small lattice size and its consequences). More importantly, the evolutionary dynamics also lead to larger public good production with higher costs.

3.6.9 Under well-mixed conditions, strong or weak altruism differ at low costs

As shown in Section 3.6.8, the spatial system is largely indifferent to whether a fraction of produced public good is retained by the producer or not. Instead, if the spatial patterns are destroyed by well-mixing the system, we can observe that at lower costs, weak altruism leads to maximising public good production, while under strong altruism the system goes extinct ($b = 10$, $c = 0.5$ in Fig 3.17, left pane). When public good production costs are higher, the system goes extinct for both strong and weak altruism ($b = 10$, $c = 4.0$ in Fig 3.17, right pane).
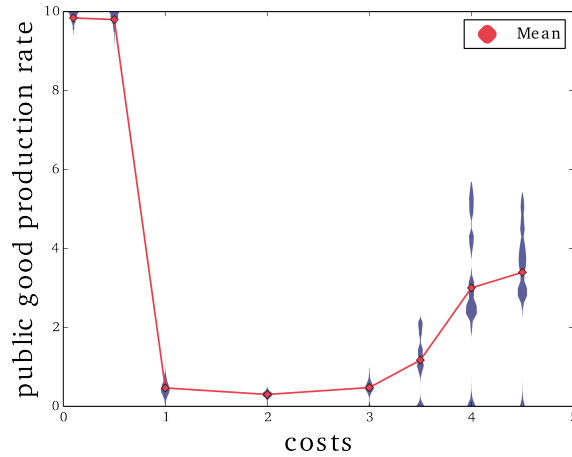
Figure 3.16. Increasing costs lead to increased cooperation when individuals do not retain the public good they produce. Parameters: b $= 10$, $k_{death} = 0.1$, $k_{move} = 0.02$, $\mu = 0.05$, $\delta = 0.1$
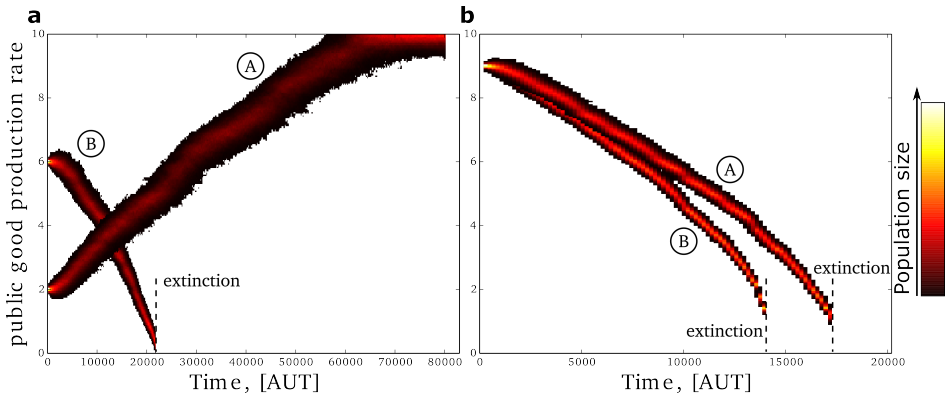
Figure 3.17. a With smaller costs $c = 0.5$, when individuals retain a fraction of public good they produce (A), public good production is maximised in the long run. In contrast, when producers do not retain a fraction of their own public good cooperation is minimised until extinction (B). **b** Whether individuals retain (A), or not (B) a fraction of public good they produce, in well mixed system with high costs $c = 4.0$ cooperation is always minimised until extinction. Notice that (A) is identical to Fig. 3.12.

4

Evolution of functional diversification within quasispecies

ENRICO SANDRO COLIZZI, PAULIEN HOGEWEG (2014)

Abstract

According to Quasispecies Theory, high mutation rates limit the amount of information genomes can store (Eigen's Paradox), while genomes with higher degrees of neutrality may be selected even at the expenses of higher replication rates (the "Survival of the Flattest" effect). Introducing a complex genotype to phenotype map, such as RNA folding, epitomises such effect because of the existence of neutral networks and their exploitation by evolution, affecting both population structure and genome composition.

We re-examine these classical results in the light of an RNA-based system that can evolve its own ecology. Contrary to expectations, we find that quasispecies evolving at high mutation rates are steep and characterized by one master sequence. Importantly, the analysis of the system and the characterization of the evolved quasispecies reveal the emergence of functionalities as phenotypes of non replicating genotypes, whose presence is crucial for the overall viability and stability of the system. In other words, the master sequence codes for the information of the entire ecosystem, while the decoding happens, stochastically, via mutations. We show that this solution quickly outcompetes strategies based on genomes with a high degree of neutrality.

In conclusion, individually coded but ecosystem based diversity evolves and persists indefinitely close to the Information Threshold.

4.1   Introduction

   In the classical formulation of quasispecies theory, populations are modelled as collections of mutationally interconnected genotypes with different growth rates. The outcome of the mutation-selection dynamics is a stable distribution of closely related genotypes: a quasispecies.

The structure of the quasispecies depends on the particular choice of the mutational scheme and the relative growth rates of the mutants, features often summarized together in the concept of fitness landscape. In a *steep* landscape the overall growth rate drops drastically in the close mutational neighbourhood, whereas in a *flat* one many genotypes have similar fitness. Since selection targets the population with the highest average growth rate (Schuster and Swetina, 1988), in a fitness landscape with both steep and flat regions, the quasispecies can be distributed in the flatter parts at high mutation rates, even if the fittest genotypes there have lower growth rates. This effect is called "survival of the flattest".

In these models, genotypes are defined solely in terms of growth rate, and fitness landscapes are static and pre-determined. Alternatively, replicators can be characterised by an explicit genotype, and a genotype-to-phenotype map. To this end, a biologically grounded instance consists of "coding" genotypes as RNA-like sequences, and phenotypes by RNA folding to secondary structure, thus co-locating information and functionality on a single molecule. RNA folding as a genotype-to-phenotype map is known to 1) be very rugged ((Huynen *et al.*, 1993), (Fontana *et al.*, 1993)) and 2) have intertwined neutral networks which percolate throughout the entire genotype space (Schuster *et al.*, 1994),(Huynen, 1996). A population of replicators that evolves on a neutral network eventually spends most of the time on its highly connected regions, provided that mutation rate and population size are large enough (Van Nimwegen *et al.*, 1999). Hence, neutrality increases automatically in the long term evolution of replicators with explicitly defined genotypes and phenotypes.

A quasispecies can be maintained in the system only if mutation rate is below a threshold value, the Error Threshold, above which the effect of too frequently arising mutants cannot be counteracted by selection, Darwinian optimization breaks down and the quasispecies delocalizes. A consequence of the Error Threshold is that, if the per-base mutation rate is constant, longer sequences suffer from mutations more than shorter ones, and there exists a maximum length a sequence can sustain above which the accumulation of deleterious mutations cannot be prevented: the Information Threshold (Eigen, 1971).

   The Information Threshold poses a serious limit on the evolutionary accumulation of information in a genome. Earlier approaches to overcome such limitation consisted of modelling different types of replicators interacting with each others, which, collectively, would integrate more information than each individual species (e.g. the hypercycle (Eigen and Schuster, 1978)). By introducing interacting replicators, the problem of inform-

ation integration is taken from the quasispecies level to the ecosystem level. However, in well-mixed conditions, these systems are evolutionarily unstable to species that benefit from being replicated without giving replication, i.e. parasites. Not so if some form of compartmentalization is taken into account, be it explicit (as in (Szathmáry and Demeter, 1987)), or emergent as a consequence of the dynamics in discrete spatially extended systems, e.g. in the form of spiral waves ((Boerlijst and Hogeweg, 1991c)), or travelling waves of replicators/parasites ((Hogeweg and Takeuchi, 2003)). More generally, spatial pattern formation has been shown to have important consequences for the eco-evolutionary dynamics of a system (e.g. (Takeuchi and Hogeweg, 2009)).

(Eigen and Schuster, 1978) (part B) considered functionally cooperating partners belonging to different lineages to be the only possible solution to the problem of integrating more information in a system (i.e. the hypercycle). Quasispecies-based solutions were excluded on the rationale that genotypic kinship relations cannot confer functional phenotypic coupling.

Here, in stark contrast, we show that a single quasispecies can integrate a large amount of information at high mutation rates, and that it behaves functionally like an ecosystem. In particular, we extend the analysis of a recently developed model (Takeuchi and Hogeweg, 2008) by focusing on quasispecies dynamics and the survival mechanisms of interacting replicators: we characterize the evolved quasispecies not only in terms of replication rates, but also by the emergent functional roles of mutants.

The Results Section is structured as follows: 1) a quasispecies which survives at high mutation rates is evolved, 2) the master sequence is determined and it is shown that replication rate is neither maximized, nor can it alone explain the survival of such sequence, 3) it is established that, counter to expectations, the evolved quasispecies is exceptionally steep, i.e. neutrality is low and most mutants are not viable, 4) we propose a functional classification of non viable phenotypes and the evolved quasispecies is explored in depth, 5) the role of the functional classes is analyzed in simplified models as well as by their spatial distribution. Finally, results are put together and the unified picture of a "functional ecosystem", populated mostly by non viable mutants, is presented. Since results are presented by means of one case study, the generality of the results and the (rare) qualitatively different outcomes close the Results section.

4.2 Material and Methods

Model The system is a spatially extended, individual oriented, Monte Carlo simulation model. Individuals consist of RNA-like strings of constant length (50 nucleotides) which are folded (to minimum free energy secondary structure, with Vienna Package version 1.7, (Hofacker *et al.*, 1994)). They are located on a two-dimensional square grid

of 512x512 cells with toroidal boundaries (based on CASH libraries, (de Boer and Staritsky, 2000)). Each cell of the grid can be occupied by at most one individual. Fig. 4.1 gives a representation of the model. One Monte Carlo step is as follows: all cells are chosen in random order and, if not empty 1) complex formation or complex dissociation may happen; 2) if in complex with a catalytically active molecule (see below), and in presence of empty space, a sequence can be replicated and the complementary sequence is generated (after which the complex breaks apart), mutations may occur 3) diffusion takes place as one step of random walk and 4) sequences can decay, i.e. the cell they occupy turns to empty. Suppose X is a catalytically active molecule and Y is not, then, schematically:

$$X + Y \underset{k_2(X,Y)}{\overset{k_1(X,Y)}{\rightleftharpoons}} \begin{cases} C_{X\circ Y} \\ C_{Y\circ X} \end{cases}$$

$$C_{Y\circ X} \begin{cases} \xrightarrow{\kappa,\theta,(1-\mu)} X + Y + Y^{-1} \\ \xrightarrow{\kappa,\theta,\mu} X + Y + Y^{-1}_\mu \end{cases}$$

$$X \xrightarrow{d} \theta$$
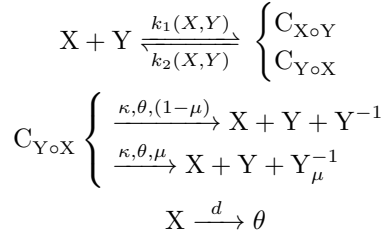
where C is a complex molecule, θ represent empty space (which constitutes the resource for replication), Y^{-1} is the complementary sequence (the subscript μ refers to the mutated sequence), κ is the replication rate and d is the decay rate. $k_1(X,Y)$ and $k_2(X,Y)$ are the probabilities that, respectively, complex formation and complex dissociation happen. Complex formation happens by binding the 5' dangling end and the 3' dangling end of two molecules adjacent on the grid. The probability of binding depends on the complementarity of the two dangling ends: the two stretches are aligned by sliding one strand on the other, to find the minimum energy score ($\mathcal{G}_{min}(X \circ Y)$), calculated as the minimum sum of the contributions of base pair matches in a continuous stretch (i.e. gaps are not allowed). $G - C$ contribution is -0.15, $A - U$ is -0.1 and $G - U$ is -0.05, the contribution of all other base pairs is zero. The probability of binding is then $p_{X\circ Y} = 1 - \exp(\mathcal{G}_{min}(X \circ Y))$. Since two molecules can form complexes in two ways (because either can be at the 5' end), the probability of binding is calculated for both configurations (so, in the same way: $p_{Y\circ X} = 1 - \exp(\mathcal{G}_{min}(Y \circ X))$). If $p_{X\circ Y} + p_{Y\circ X} > 1$ the two probabilities are normalized, i.e. $k_1(X \circ Y) = p_{X\circ Y}/(p_{X\circ Y} + p_{Y\circ X})$, otherwise $k_1(X,Y) = p_{X\circ Y}$. The probability of complex dissociation is $k_2(X,Y) = 1 - k_1(X,Y)$.

Quantitatively, if two dangling ends match only for a $G - C$ pair, the probability of complex formation is negligible (≈ 0.015), while complex dissociation is very likely ($1 - 0.015 = 0.985$); moreover, it takes at least five $G - C$ pairs (or seven $A - U$ pairs) for the probability of complex formation to be larger than that of complex dissociation ($p_{X\circ Y} > 0.5 \iff \mathcal{G}_{min}(X \circ Y) < -0.694$, which, for $G - C$ only pairs is true when $\mathcal{G}(5 \times (C \circ G)) = -0.75$, and for $A - U$ pairs when $\mathcal{G}(7 \times (A \circ U)) = -0.7$).
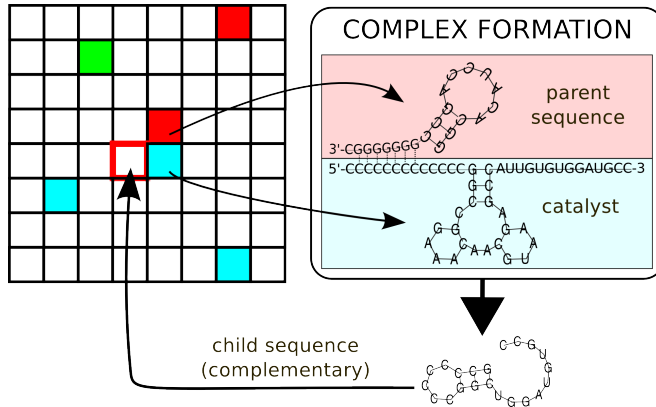
Figure 4.1. Schematic representation of complex formation and replication: two adjacent sequences form a complex based on sequence similarity; if the molecule binding with the 5' dangling end is catalytic, in presence of empty space, replication takes place and the complementary of the other sequence is produced. Mutations may happen at this step. In the model, all sequences are 50 nt long, here they are depicted shorter for clarity.

Figure 4.2. Classification of the mutational classes. Phenotypes are determined for pairs of complementary sequences (black: + strand, red: - strand), and are coarse grained to presence/absence of 5' dangling end, catalytic structure, 3' dangling end. Since different phenotypes may fall in the same functional class, some of all the possibilities are depicted (the larger ones are the most frequent configurations). "Helpers" are defined by having a 5' dangling end and the catalytic structure on at least one (and the same) strand, and not having a 3' dangling end in both strands. "Stallers" are defined by having a 5' dangling end but no catalytic structure on at least one (and the same) strand, and not having a 3' dangling end in both strands. "Junk" are sequences that do not have any 5' dangling end on either strands, and no 3' dangling end on at least one. "Hybrid" sequences display a helper-like phenotype on one strand and staller-like on the other.

Catalytic activity consists of being able to replicate other molecules. A sequence is defined as catalytic if it folds into a predetermined, coarse grained secondary structure, arbitrarily chosen to consist of a multiloop which connects a stem to two hairpins, ((((H)S)((H)S)M)S), in Shapiro notation (as implemented in the Vienna Package). Two more conditions need to be satisfied for replication to take place: the catalytic molecule must be engaged in complex with its 5' dangling end and empty cell must be present in the neighbourhood of the complex. The sequence complementary to that at the 3' end of the complex is generated, folded and located on the empty cell. The only mutations implemented are substitutions, which happen with a per base probability μ. The replication rate, κ is set to 1 (i.e. replication depends only on complex formation and the availability of empty space), the decay probability d is set to 0.03. A small probability of not moving ($p = 0.1$) is introduced for complex molecules, to take into account their slower diffusion.

4.2.1 Phenotype recognition and classification

For the analysis of the simulation output we have developed a classification of individual molecules which is based on presence/abscence of catalytic domain, as well as on 5' and 3' dangling ends. Considering both a sequence and its complement, which together define a genotype, any phenotype can be coarse-grained to 6 bits of information, meaning that there exist $2^6 = 64$ possibilities. If a dangling end is short, a molecule would have a small probability of forming a complex. In order to classify a molecule as having/not having a dangling end, the complementary stretch is generated, and the probability of complex formation is calculated as above. We set a threshold for the energy score of the complex: if $G \leq -0.75$ (corresponding to a probability ≈ 0.5) then the molecule is classified as having that dangling end, otherwise, as not having it.

Phenotypes with minor differences can be grouped into phenotype classes, which can be further joined by integrating functional considerations. In Fig. 4.2 the proposed classification is presented. The dependence on a particular threshold value to recognize a dangling end as such ($G \leq -0.75$) is minimal in the evolved system (within a reasonable range; see Supplementary material Fig. 4.12, left pane). In contrast, the distribution of the functional classes in random sequences does depend on it (right pane).

A necessary (intrinsic) condition for genotypes to be replicated is to have a 3' dangling end on both strands. However, in order to *give* replication, a sequence must fold into the catalytic structure and have a 5' dangling end. A "Unit of replication", is a viable pair of complementary sequences, of which at least one is able to give replication. It is trivial to observe that, for a minimal system to persist, units of replication have to be present. Nonetheless, the minimum viable phenotype consists of a pair of sequences that do not fold into the catalytic structure, have 3' dangling end, but no 5' ones. Such phenotype could exploit the units of replication for catalysis, and as such would be a parasite.

Units of replication and parasites are the only two viable classes. Among the non viable ones, we define "Helpers" as those sequences which can replicate other molecules but cannot be replicated, "Stallers" can engage molecules in complex, but can neither replicate them nor be replicated, "Junk" cannot form complexes (these phenotypes are mostly inert) and "Hybrids" sequences display a helper-like phenotype on one strand and staller-like on the other.

4.3 Results

4.3.1 Evolving persistence to high mutation rates

Sequences have to be evolved in order to withstand high mutation rates because initializing the system already at high mutation rates ($\mu \geq 0.013$) with randomly generated units of replication leads to a quick extinction (see below). Starting from low mutation rates, each time a sequence is mutated, a small positive random number is added to its mutation rate (if the distribution is unbiased and negative numbers can be drawn, mutation rate decreases). During the evolutionary run, the system displays various dynamic regimes which can be characterized by the structure and the stability of the evolving ecosystem (see Supplementary Material and Fig. 4.13). The dependence of the number of species and ecosystem structure on mutation rate is in line with (Takeuchi and Hogeweg, 2008): while for lower mutation rates multiple lineages coexist in the field, when mutation rate is sufficiently high ($\mu \geq 0.014$) only one quasispecies is present. At this point the simulation is continued by setting the value of μ to constant. For the case we focus on, $\mu = 0.015$, which corresponds to a probability of at least one mutation happening per replication event $1 - (1 - \mu)^\nu = 0.53$ (with ν -length of a sequence- $= 50$ nucleotides).

4.3.2 The Master Sequence of the Quasispecies

Clustering the sequences reveals that only one quasispecies is present in the field, with a high degree of sequence similarity (Fig. 4.3, top). The consensus sequence is also the most abundant genotype occurring, thus the master sequence, and the center of the quasispecies. The inspection of the ancestors tree (Fig. 4.3, bottom) confirms it, as such sequence is the most frequently observed ancestor of every other molecule (cf. (Hermisson *et al.*, 2002)). However, at any time point, several *sub-lineages* coexist and compete even for long time periods before going extinct. Along the line of descent (in yellow), as well as in the other lineages (in red), the presence of the master sequence is intermittent, and the ancestor of the lineage becomes a close mutant. Nonetheless, back mutations restore the original sequence in the long run. This is an indication that the system is close to the error threshold (cf. metastable states close to the Error Threshold in (Takeuchi and Hogeweg, 2007a)).
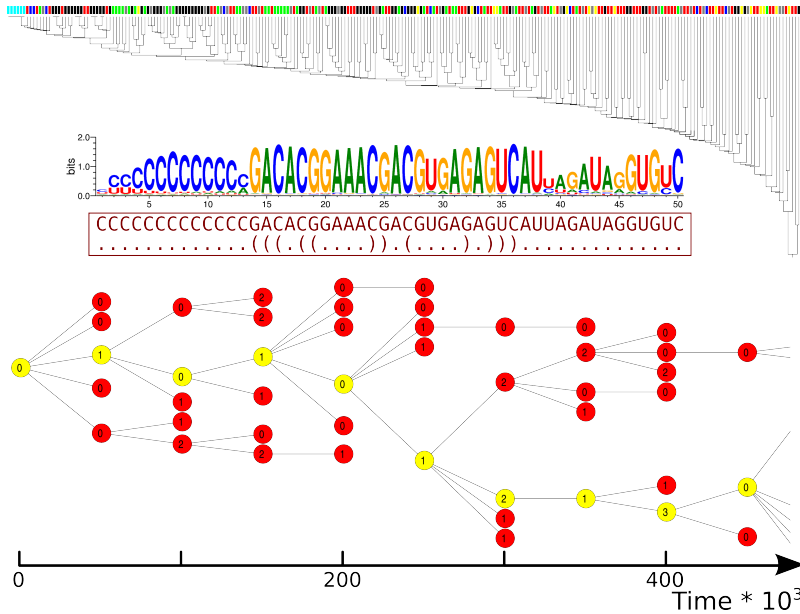
Figure 4.3. Evolved quasispecies: clustering, sequence logo, consensus sequence and ancestor tree. **Top**: Clustering, sequence Logo and most abundant sequence with its secondary structure (in dots-and-brackets notation) of the quasispecies at $\mu = 0.015$. 1000 sequences are randomly sampled from the population at one time point, clustered (500 hundred are displayed for clarity) and the sequence logo is generated (sampling at different time points, or larger samples do not produce qualitatively different results. Colours on the leaves correspond to the functional class in which sequences fold (see Models and Methods). Cyan: most abundant sequence, black: units of replication, green: helpers, red: stallers, grey: junk, blue: hybrids. The tree is visualised with iTOL ((Letunic and Bork, 2011)) **Bottom**: Ancestors tree over the first $450 * 10^3$ time steps of a simulation run (the simulation is initialised with a homogenous population of master sequences, it consisted of $\approx 2.5 * 10^6$ time steps, after which it was interrupted). The rest of the simulation as well as different runs show qualitatively the same pattern. The tree is built so that the nodes at a given time step are every individual's ancestor $50 * 10^3$ time steps later, edges connect lines of descent; yellow nodes are those in the line of descent that persists until the end of the simulation, in red the others. Numbers mark the Hamming Distance of the ancestor from the master sequence. The tree is visualised in Cytoscape ((Smoot *et al.*, 2011)).

4.3.3 Replication rate

Since in Quasispecies Theory fitness is defined (solely) by replication rate, we calculate the replication rate of the master sequence and analyze to which extent it is maximized in the evolved quasispecies.

The phenotype associated to the master sequence is catalytic and has two dangling ends. The 5' one is composed exclusively of C's, while the 3' end lacks a clear pattern in sequence composition. The complementary strand is not catalytic, with a closed 5' end, and a 3' dangling end with only G's. Taking into account both strands, the overall phenotype class is that of a unit of replication. In order to replicate, the catalytic strand has to be able to form a complex with both itself and the complementary sequence. Clearly, the 5' end of the catalytic strand matches perfectly with the 3'end of the opposite strand (the probability of complex formation is ≈ 0.86). However, forming a complex with another catalytic strand poses a problem: having too many G's on the 3' end, which would, in principle, ensure a high complex formation probability, could cause a molecule to fold on itself (forming a stem), thus leaving no dangling ends. This explains the intricate mixture of nucleotides at the 3' dangling end of the catalytic strand. The probability that two catalytic strands of the master sequence form a complex is 0.90.

The self-replication rate of the master sequence is rather high, due to its long danglind ends and the $C - G$-based strategy for complex formation. To test whether self-replication is maximized we implemented an evolutionary optimization algorithm that selects sequences for replication rate, without considering any interaction between genetically different individuals. Catalytically perfect units of replication evolve quickly by using an evolutionary optimization algorithm that selects sequences for replication rate. Genotypes consist of a pairs of complementary sequences, both catalytic, for which the total probability of binding both the same strand and the complementary are equal to 1 (i.e. when $\mu = 0$, the only limit to growth is diffusion; see Material and Methods). Sequence composition on the dangling ends of units of replication evolved with this method is far from being dominated by C's, rather, A's and U's (interspersed within each tail) are the most frequently observed nucleotides (see caption of Fig. 4.4).

Since the optimized sequences achieve the highest absolute degree of optimization for replication rate, while those in the full model do not, the former do replicate faster also in the full model, when $\mu = 0$. However, all the sequences tested quickly go extinct for $\mu \geq 0.013$, where the master sequence easily survives.

A unit of replication which has both strands catalytic is never observed in the model, possibly because both its strands can be exploited by parasites as well as other (non symmetric) units of replication. Moreover, from a pool of randomly generated units of replication, we found that those catalytic on both strands are the minority (37%), and for only 4% of the sequences each strand replicates the other better than self. Units of
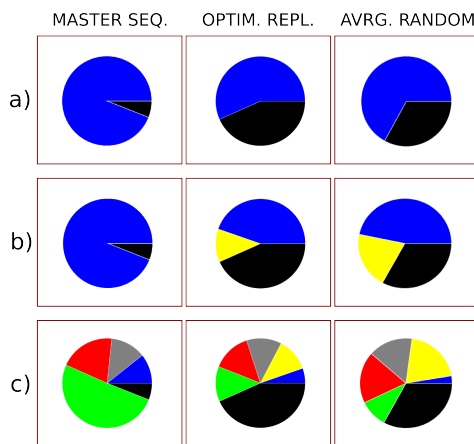
Figure 4.4. Hamming Distance = 1 mutational neighbourhood of the evolved master sequence (left), one typical maximally fit sequence after evolutionary optimization for replication rate (centre), the average random unit of replication (right). The piecharts show the fraction of mutants that fold into the various functional classes. **a)**: units of replication (in black),i.e. neutrality vs. the other foldings (in blue). **b)**: as above, with, in yellow, the fraction of mutants that fold into parasite; **c)**: the full mutational neighbourhood. Colours of functional classes are the same as for the cluster tree. The sequence of the optimised unit of replication is AAAACGUGUAAAGGAGCGAAUCGCAGGCAGAGCCACCAUAAAAGUUAUUA. Random units of replication are obtained by generating 10^6 random sequences and screening their function: 228 units of replication are found. All HD=1 mutants of the latter are generated and the fractions of mutants folding into each functional class are determined.

replication optimized for replication rate so that only one strand can be catalytic achieve slightly higher replication rates than those in the full model. Moreover they are not biased in nucleotide composition and they invariably go extinct at high mutation rates (data not shown).

In conclusion, the selection of the master sequence cannot be explained by considering exclusively its (partially optimized) replication rate, or the structure of its dangling ends.

4.3.4   The mutational neighbourhood of the master sequence

Having excluded that optimization at the level of replication rate is the sole outcome of evolution, we turn to study the effect of interactions within the quasispecies, i.e. the interactions with the mutants of the master sequence. We first consider the viable mutants, i.e. neutral ones and parasites, then we characterize the rest of the mutational neighbourhood.

Neutrality   The fraction of all the Hamming Distance (HD) = 1 mutants of the master sequence that fold into units of replication is $\lambda_{ms} = 0.06$ (note that a broad definition of neutrality is adopted, as folding into a unit of replication suffices, and replication rate is not taken into account). In comparison, the average degree of neutrality for random units of replication is $\lambda_r = 0.33$ and for a typical unit of replication obtained by optimizing replication rate is even larger ($\lambda_{opt} > 0.40$, Fig. 4.4, top row). Interestingly, the master sequence seems to belong to a *steep* (as opposed to *flat*) quasispecies. The master sequence is also non-modular: the fraction of units of replication obtained when mutating only the dangling ends of the catalytic strand is only slightly higher than the neutrality of the whole sequence ($\lambda_{tails} = 0.11$). Moreover, the master sequence is a far outlier with respect to the neutrality distribution for random units of replication (Fig. 4.5, first pane).

Parasites   In this model, parasitic lineages readily evolve at lower mutation rates and, because of spatial structuring, they do not destroy the ecosystem (Supplementary material, and (Takeuchi and Hogeweg, 2008)). Strikingly, the chances of generating a parasite as a $HD = 1$ mutant of the master sequence are exactly zero (0.0005 at $HD = 2$), despite the fact that a single mutation disrupting the catalytic structure of a unit of replication could suffice. In contrast, the fraction of parasites in the $HD = 1$ mutational neighbourhood of both random units of replication and the ones optimized for replication rate is much higher (Fig. 4.4, center row). When these sequences are inoculated in the full model at high mutation rates, parasites are going to be generated often and, being very similar to the unit of replication they originate from ($HD = 1$), they become strong competitors for getting replicated. As close mutants of units of replication, they are not spatially separated (as it happens in, e.g., travelling waves) and lead the system
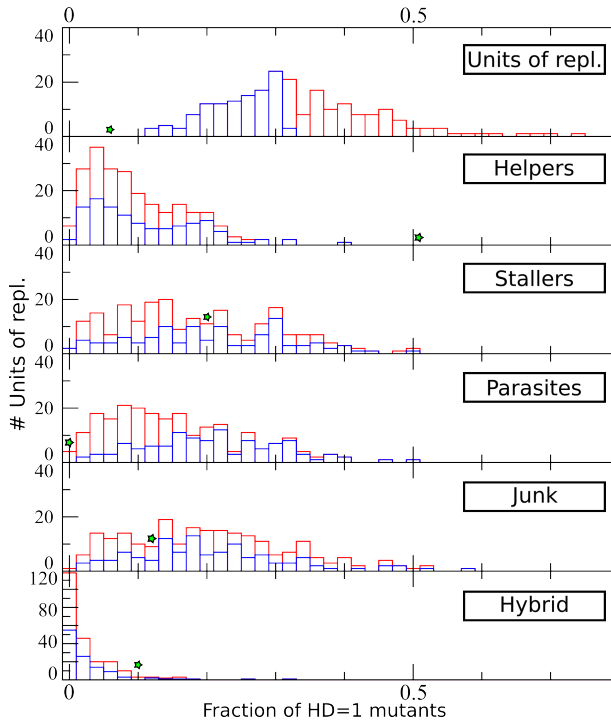
Figure 4.5. Distribution of HD=1 mutants for random sequences and for the master sequence. Each histogram displays the distribution of number of units of replications for their fraction of mutants that fold into the functional class indicated (bin size = 0.02). In blue, the distribution of the 50% least neutral sequences among different classes, in red, the 50% most neutral. Green stars: fraction of mutants of the master sequence belongiong to the class indicated.

to extinction, similarly to the situation in a well mixed system (cf. e.g (Takeuchi and Hogeweg, 2007b)). This explains why these sequences fail to survive despite having a high replication rate. In conclusion, long term evolution at high mutation rates minimizes the chances of producing parasites in the quasispecies.

So far, only 6% of the $HD = 1$ mutational neighbourhood has been characterized, considering the low degree of neutrality of the master sequence ($\lambda_m s = 0.06$) and the fraction of its mutants that turn into parasites ($\lambda_P = 0$). This completes the description of the viable mutants. In the next section, the remaining 94% of non-viable mutants are better characterized.

The $HD = 1$ mutational neighbourhood   In Section Material and Methods the classification of all the possible phenotypes has been introduced. Here we analyze the distri-

bution of such classes in the mutational neighbourhood of the master sequence, to conclude that the evolved quasispecies is selected for the peculiar distribution of its mutants.

In Fig. 4.4, bottom row, the relative fractions of functional classes arising as mutants of the master sequence are compared to those of optimized sequences and random ones (as before). Helpers make up to about 50% of the mutational neighbourhood of the master sequence, in contrast to the case of random sequences ($\approx 10\%$) and those optimized for replication. The master sequence does not seem to differ much from random sequences for what regards junk, stallers and hybrids.

In Fig. 4.5, the mutants of the master sequence are compared with the full distribution of mutants from random units of replication, rather than just with the average values. The green star in each subplot represents the frequency of each functional class in the $HD = 1$ mutational neighbourhood of the master sequence. The degree of neutrality of the master sequence ($\lambda_{ms} = 0.06$), the fraction of its mutants that fold into parasites ($\lambda_P = 0$) or helpers ($\lambda_H = 0.51$) are far outliers of their respective "null" distributions: the first two are under-represented, while helpers are over-represented. Stallers, junk and hybrid sequences in the neighbourhood of the master sequence seem not to be significantly different from random.

It could be the case that, since the total number of possible mutants produced is constant (i.e. 3 possible substitutions $*$ sequence length $= 150$ mutants), the lower degree of neutrality in the master sequence would allow for more different kinds of functional classes as a side effect. To address this caveat, the population of random units of replication is split into two groups: the 50% least neutral sequences (blue in Fig. 4.5) are separated from the 50% most neutral (red), and the two groups are compared for the other functional classes. We find no evidence that such is the case, as the two groups distribute roughly in the same way for all functional classes. Altogether, this indicates that the properties of the $HD = 1$ mutational neighbourhood of the master sequence are evolved and that the master sequence is selected for its mutational neighbourhood wich minimizes neutrality and frequency of parasites, as well as maximizes the amount of helpers.

Comparing the mutational neighbourhood of random units of replication with the global occurrence of functions in genotype space (Supplementary material, Fig. 4.12, right pane), we observe that the global occurrence of helpers and hybrids is limited, which explains why they are so infrequent as mutants of random units of replication. Instead, it is remarkable that some units of replication have only few stallers, junk or parasites, given their abundance in the genotype space. However, as mentioned above, most of the sequences in these functional classes might be unreachable with few mutational steps from a unit of replication. From which we conclude that the genotype space from the "mutational point of view" of a particular phenotype looks biased from the global picture. Of course, selection can act on the former, and not on the latter.

The mutational neighbourhood at larger distance   Units of replication arise as mutants of the evolved master sequence, and, in turn, make more mutants. Analyzing what the mutational neighbourhood is like at higher Hamming Distances allows us to understand to which extent of the mutational neighbourhood selection pressure reaches.

The first search to higher Hamming Distances is performed by selecting units of replication that have a replication rate as high as (or higher than) that of the master sequence. Starting from the master sequence, the average mutational neighbourhood of these units of replication, at progressively higher Hamming Distances is plotted in Fig. 4.6 (core neutral mutants). The frequency of units of replication mutants of neutral sequences increases slightly and seems to saturates for $HD > 4$, whereas the fraction of core neutral mutants among them remains extremely low. The fraction of helpers drops quickly, whereas, symmetrically, the fraction of stallers increases. Very few parasites are present, and there seems to be little variation for Junk and Hybrid sequences. Taking into account that mutants in the close mutational neighbourhood are also close in space in the full system (because replication is a local process), the picture that emerges is that the master sequence is able to outcompete neutral mutants by being replicated more often than anyone else by helpers, while being hindered less than every other unit of replication by stallers.

A second search is performed by selecting any unit of replication (pseudo neutral in Fig. 4.6, second row). The main differences with the results above are the marked increase in units of replication at higher Hamming Distances and the larger amount of parasites, while stallers do not increase as much as in the previous case. Although units of replication are increasingly neutral at higher Hamming Distances, they do not outcompete the master sequence.

Finally, a third procedure is implemented, by sampling the mutants of the master sequence at progressively higher Hamming Distances. This procedure is implemented for the sake of completeness: some units of replication may be generated only with multiple mutations (hence, they would not be taken into account by the previous procedures) and may contribute to the overall shape of the landscape. However, the mutational neighbourhood emerging from this method is very similar to the one described above (Fig. 4.6, third row).

Altogether, the analysis of the mutational neighbourhood suggests that selection acts to shape and finely tune the mutational neighbourhood in such a way that small genotypic variations (single substitutions) have large effects on the functional role of the phenotypes. Moreover, interactions with strong competitors are minimized, as they are mediated by either non competitive units of replication (which dilute the latter) or non viable sequences. The distribution of the non viable mutants of the master sequence is shaped in such a way to contribute to the replication of the master sequence itself (the helpers at low Hamming Distance) or hinder competitors (stallers at high Hamming Distance). The mutational neighbourhood of units of replication selected with the second
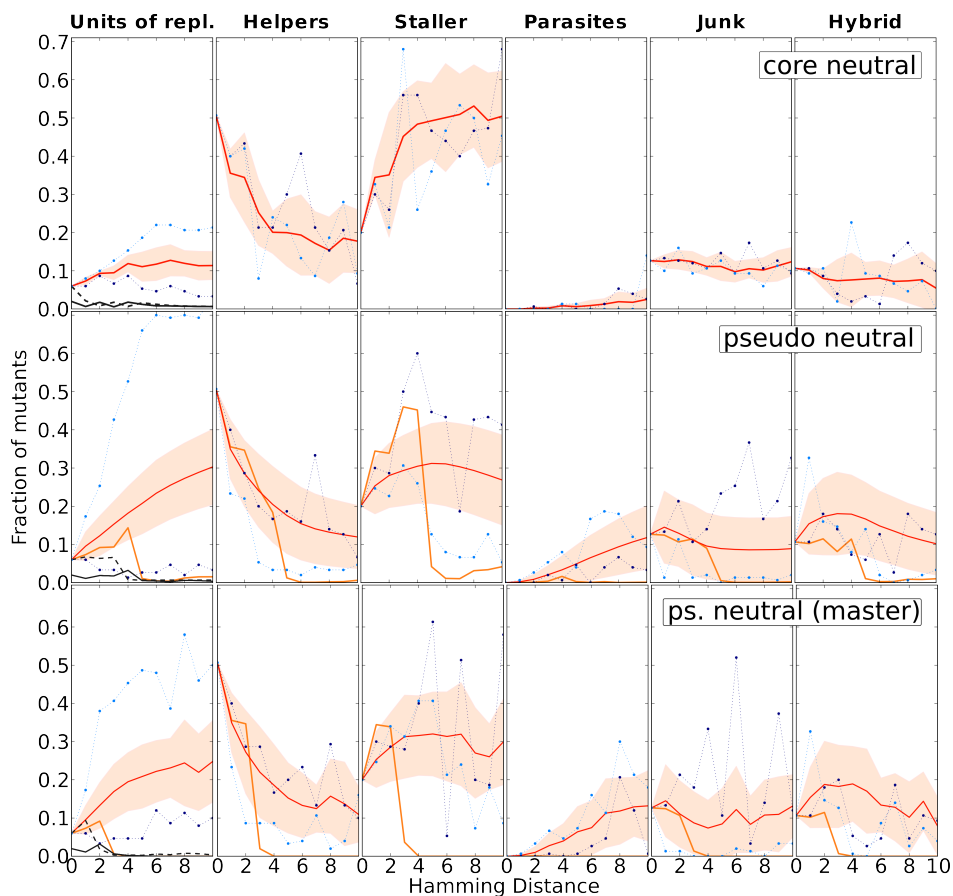
Figure 4.6. Mutational neighbourhood of the master sequence at higher Hamming Distances. **First row**: fully neutral mutants (units of repl. with replication rate as high or higher than that of the master sequence), **Second row**: pseudo neutral mutants (units of repl. with repl. rate lower than that of the master sequence). Random sampling for units of replication if more than 10^4 are found in both cases), **Third row**: pseudo neutral mutants sampled from the master sequence at Hamming Distances from 1 to 10 (if the total number of units found exceeds 10^4 or if, after 10^5 tries, no more units are found, the algorithm moves on to the next Hamming Distance value). In the first row. Continuous red line: average fraction of phenotypes in each class. In the second and third row: Thick red line: average fraction of phenotypes in each class for pseudo neutral mutants; orange line: average fraction of mutants for core neutral units of replication. For all rows. Shaded area: between +/ - standard deviation. Continuous black line: fraction of total mutants that are core mutants. Dashed black line: fraction of total mutants that are pseudo neutral mutants. Blue dotted line: mutational neighbourhood of the sequence with lowest neutrality at a certain Hamming Distance. Light blue dotted line: as before, except for highest neutrality.

method becomes more and more similar to that of random units of replication the higher the Hamming Distance. Above, we have seen that random units of replication are not viable at high mutation rates, hence, they cannot be competitors of the master sequence. The lack of a structured mutational neighbourhood (especially for the abundance of parasites) explains their quick extinction (see above).

4.3.5   Spatial population dynamics

So far, we have presented the functional classes and analyzed the quasispecies from a "static" perspective, i.e. by exploring the mutational neighbourhood of the master sequence. We now turn to the field to assess which effects the functions defined above have, and at which scale.

Limited diffusion is essential for survival, as increasing the number of random walk-steps per reaction step (to a ratio 3:1), as well as mixing the system, leads to extinction quickly. Nevertheless, at a first glance, the field looks patchy and disorganized. Copies of the master sequence are more or less clustered and separated from parasites, which (as expected) are closer to the regions with empty space. The other units of replications, as well as helpers, stallers and junk, are widespread, with limited apparent spatial clustering (Supplementary Material, Fig. 4.14. Notice that the system is far from being mixed, as increasing the frequency of random walk-steps per reaction step leads to extinction quickly, and so does mixing the system every time step).

The total number of individuals oscillates in time (occupying in total about half of the field), however, since the relative ratio of functional classes remains almost constant (Fig. 4.8), we can study the distribution in the field from a single time point. The individuals at HD=1 from the master sequence (Fig. 4.8) distribute similarly to the HD=1 mutants (see Fig. 4.7), the exception being the great abundance of units of replication. The relative fractions of functional classes, however, distribute in a way roughly similar to random (Fig. 4.7).

Units of replication and parasites are selected in the short run (their number is large) because they are viable. Nontheless, most sequences are non viable at any given moment: helpers, and to a lower magnitude, junk, are more frequent at lower Hamming Distances, while stallers reach their peak at higher ones. The distribution of parasites has a peak even further in Hamming Distance. Notice that in the field it is possible to discriminate the strand of a hybrid genotype, which can be assigned to helpers or stallers. Fig. 4.8 and Fig. 4.6 share some degree of similarity (until $HD = 10$) in that helpers decrease with higher Hamming Distances, stallers increase, parasites increase only for $HD > 3$ and the fraction of junk does not show much variation. Analogously, Fig. 4.3 (top) shows that helpers and units of replication are more frequently found close (genotypically) to the master sequence, while stallers and parasites arise far from it.
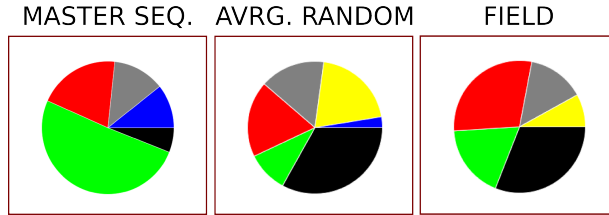
Figure 4.7. The piecharts represent **left** the mutational neighbourhood of the master sequence, **centre** the mutational neighbourhood of random units of replication, **right** frequency of functional classes present in the system at one time point (Fractions : units of replication 0.31, helpers 0.18, stallers 0.29, junk 0.14, parasites 0.08). The colours in the piecharts are as in Fig. 4.4.
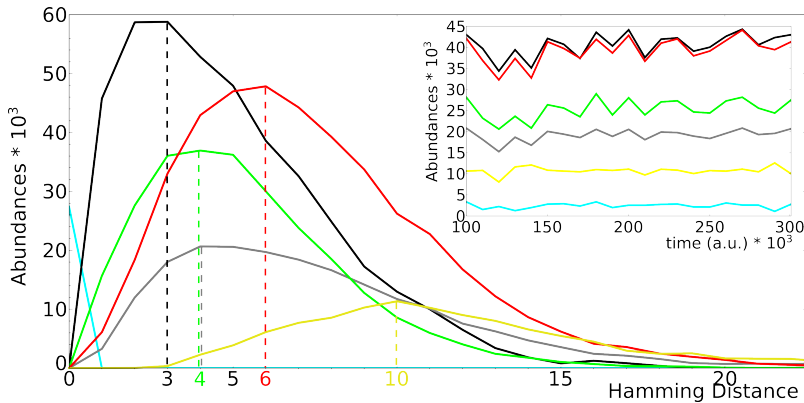
Figure 4.8. Distribution of the abundances of the functional classes in the field (at one time point) as function of the Hamming Distance of their sequence from the master sequence. The Hamming Distance value is calculated as the minimum between the Hamming Distance of the sequence and the master sequence, and the Hamming Distance of the complementary sequence with the master sequence. Inset: Time plot of the abundances of the various functional classes (the total population is between $100 * 10^3$ and $150 * 10^3$ individuals). Colour coding as in Fig. 4.4; in cyan, the master sequence

Altoghether, there seems to be a correlation between the distributions of mutants of the master sequence and the distribution of functional classess present in the field. To better explain the particular distribution of the functional classes, more information about the behaviour of the non viable phenotypes is needed, which is the concern of the next section.

4.3.6   The role of non-viable mutants for the stability of the system

So far, the non-viable phenotypes have been classified, assigned a function and their presence in the field has been shown. It still remains to determine what role they have in terms of ecological stability of the system, especially for what regards helpers and stallers, which is the concern of this section.

Helpers   To assess the role of helpers we exclude them from the field in two ways: either by removing them, leaving empty space (in which case they still benefit the system in that the empty space becomes a resource), or by turning them into junk (i.e. inert material). In both cases the system goes extinct quickly. Helpers are crucial for the viability of the whole system.

Because helpers are widespread in the field (Supplementary Material, Fig. 4.14), parasites benefit from them as well. To investigate the interplay of helpers and parasites for the stability of the system (especially at high mutation rates), we study a simple system of ordinary differential equation (ODE) model. We assume that units of replication (X, we ignore the +/- strand difference), can form complexes with other units, helpers (H) and parasites (P), and the last two can also form complexes. Upon complex formation, new molecules are generated. The mutational products of units of replication are helpers, parasites and junk (which is inert). We assume that parasites do not mutate, which is justified by their (generally) high neutrality and the lack of helpers in their mutational neighbourhood; for simplicity, we do not include junk produced by them, thereby mod-

elling *strong* parasites. The reaction scheme goes as follows:

$$
2\,\mathrm{X} \underset{b_{xx}}{\overset{a_{xx}}{\rightleftharpoons}} \mathrm{C}_{xx}
\begin{cases}
\xrightarrow{\kappa,\theta,(1-\mu)} 3\,\mathrm{X} \\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_H} 2\,\mathrm{X} + \mathrm{H} \\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_P} 2\,\mathrm{X} + \mathrm{P} \\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} 2\,\mathrm{X} + \mathrm{J}
\end{cases}
$$

$$
\mathrm{X} + \mathrm{H} \underset{b_{xh}}{\overset{a_{xh}}{\rightleftharpoons}} \mathrm{C}_{xh}
\begin{cases}
\xrightarrow{\kappa,\theta,(1-\mu)} 2\,\mathrm{X} + \mathrm{H} \\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_H} \mathrm{X} + 2\,\mathrm{H} \\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_P} \mathrm{X} + \mathrm{H} + \mathrm{P} \\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} \mathrm{X} + \mathrm{H} + \mathrm{J}
\end{cases}
\tag{4.1}
$$

$$
\mathrm{X} + \mathrm{P} \underset{b_{px}}{\overset{a_{px}}{\rightleftharpoons}} \mathrm{C}_{px} \xrightarrow{\kappa,\theta} \mathrm{X} + 2\,\mathrm{P}
$$

$$
\mathrm{H} + \mathrm{P} \underset{b_{ph}}{\overset{a_{ph}}{\rightleftharpoons}} \mathrm{C}_{ph} \xrightarrow{\kappa,\theta} \mathrm{H} + 2\,\mathrm{P}
$$

$$
\mathrm{X},\mathrm{H},\mathrm{P},\mathrm{J},\mathrm{C}_{xx},\mathrm{C}_{xh},\mathrm{C}_{px},\mathrm{C}_{ph} \xrightarrow{d} \theta,
$$
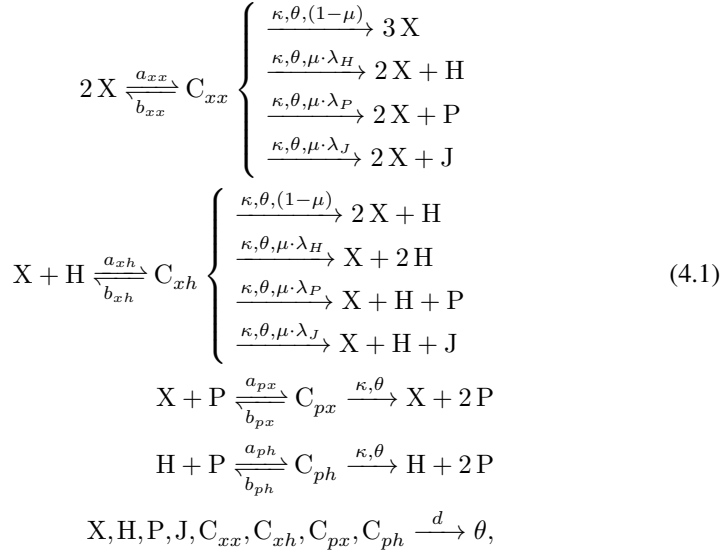
where a is the rate of complex formation for the molecules indicated by the subscripts, b is the complex dissociation rate, C is a complex between two molecules indicated in the subscript. κ is the replication rate (it can be thought of as a polymerization rate), θ is a phenomenological term for competition (e.g. resources, such as empty space), μ is mutation rate, λ is the fraction of mutants that turn into the class in the subscript. The ODE system from this reaction scheme and a summary of the bifurcation analysis can be found in Supplementary Material.

In Fig. 4.9 (left pane), the steady state values of the units of replication are plotted against mutation rate (μ) for different values of the fraction of mutants that turn into helpers, λ_H (bifurcation plot). Interestingly, the system is destabilized at lower mutation rates for higher values of λ_H. This effect is entirely due to the presence of parasites and stays the same if the fraction of mutants that turn into parasites (λ_p) is set to zero. To show this, the parasite equation is removed from the ODE system, and a similar bifurcation plot is built (Fig. 4.9, right pane). In contrast with the previous case, increasing the fraction of mutants that turn into helpers (λ_h) makes the system more resistant to mutations (see inset of Fig. 4.9, right pane). The comparison of the two systems shows that in the absence of parasites, helpers stabilize units of replication to a large extent, but it still remains that the system is fragile to parasites: at high mutation rates, the latter can invade and cause extinction (not shown). Conversely, the production of parasites as mutants of units of replication destabilizes the system and leads to extinction at much lower mutation rates, no matter the value of λ_H. We conclude that in the full system helpers are only favorable for the master sequence (and, possibly, for the units of replication very close to it), because of its lack of parasites in both the mutational and the physical neighbourhood.

In the next section we show how the latter is mediated by stallers.

Stallers   To investigate the role of stallers (similarly to the case with helpers) we exclude them from the field either by removing them or by replacing them with junk. In both cases the density of the units of replication in the field increases abruptly, and two new lineages evolve: one catalytic and one parasitic, with the system self-organizing into travelling waves. The strategy of units of replication in the new lineage is still 'C'-based, whereas parasites are similar in structure to those evolving at lower mutation rates (see (Takeuchi and Hogeweg, 2008)). In the case where stallers are removed from the field, the new catalytic lineage lacks a persisting master sequence, has high neutrality (≈ 0.3) and few helpers in the mutational neighbourhood. An increase in neutrality of units of replication is both a way to cope with high mutation rates and a mechanism to partially weaken the parasitic exploitation, because it increases the variability in the system. In contrast, for the case where stallers are turned to junk, the units of replication are characterised by low neutrality, a moderate amount of helpers and a large fraction of stallers (tuned to junk). However, the new master sequence evolves so that what is classified as junk behaves partially as stallers. This quasispecies outcompetes the parasitic one, which goes extinct.

That stallers hinder the growth of units of replication (in general) can be seen from the initial, sudden increase of density in the field, when stallers have just been removed. Since the quasispecies is highly homogeneous across the field (see Fig. 4.3, top), the increased number of units of replication makes it easier for a parasite to invade (to some extent, further facilitated by the fact that the stallers in their mutational neighbourhood are also excluded). In this sense, stallers affect parasites mostly by removing the substrate for their replication (i.e. units of replication), and only to a lesser extent by direct interaction with them. In such conditions, parasites optimize the exploitation of the units of replication, by loosing both 5' dangling ends while increasing the density of 'G' on both 3' ends. Given the evolution of faster replicating parasites, units of replication that rely on helpers for survival, such as the master sequence, are counter-selected and go extinct (as explained above).

Altogether, we conclude that stallers are an intrinsic problem for the system, and that the master sequence of the full model has evolved some *mutational control* over them.

Similarly to the case with helpers, an ODE model is studied to understand the interplay of units of replication, stallers and parasites. When units of replication (X) replicate erroneously, mutants can be junk (J), parasites (P) or stallers (S); the erroneous replication of parasites produce mutants as well, namely junk or stallers. Stallers engage in complex with both parasites and units of replications, but no replication happens. The reaction
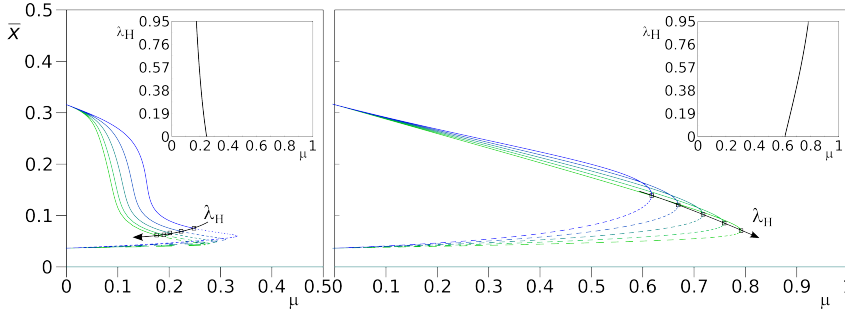
Figure 4.9. Bifurcation diagrams for the ODE systems with helpers (Reaction scheme 4.1). **Left**: with parasites. Parameters: $a_{hp} = 0.7, a_{xh} = 0.7, a_{xp} = 0.7, a_{xx} = 0.7, d = 0.03, \lambda_P = 0.03, \kappa = 1, \theta = 1$. Continuous line: stable equilibrium; dashed line, unstable equilibrium; black squares: bifurcation points. The arrow highlights the direction of change for the bifurcation point with progressively increasing λ_H values ($\lambda_H \in \{0, 0.25, 0.5, 0.75, 0.95\}$). For simplicity, the Hopf bifurcation is marked rather than the homoclinic bifurcation. However the two bifurcations are very close. For all parameters combinations, the extinction state $(\overline{0})$ is a stable equilibrium. Inset: two parameter bifurcation plot (μ vs. λ_H). **Right**: Same as above, except parasites are not included in the ODE. The arrow highlights the direction of change of the limit point. Numerical integration and bifurcation analysis performed with GRIND and CONTENT ((de Boer and Pagie, 2005), (Kuznetsov, 1999)).
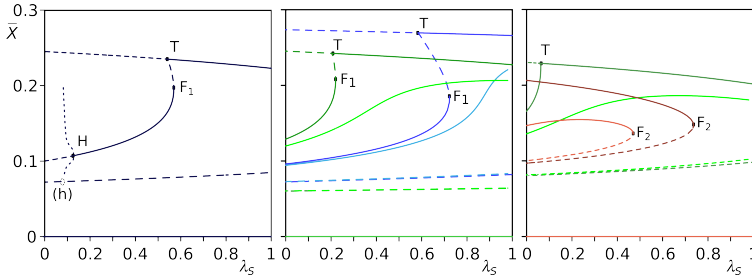
Figure 4.10. Bifurcation diagrams for the ODE system with stallers (Reaction scheme 4.2). Solid lines: stable equilibria, dashed lines: unstable equilibria, dotted lines: max/min limit cycle, dots: bifurcations ((h): estimated location of possible homoclinic bifurcation, H: Hopf, F_1 and F_2: Fold, T: transcritical bifurcation). **Left**: general figure, parameters: $\lambda_P = 0, \mu = 0.4, a_{ps} = a_{xs} = 0.75, a_{xp} = 0.8$. **Centre**: changing μ and λ_P (green: μ=0.3, blue: μ=0.4; dark colour: $\lambda_P = 0$, light colour $\lambda_P = 0.02$), other parameters: $a_{ps} = a_{xs} = 0.7, a_{xp} = 0.775$. **Right**: highest values of μ (green: μ=0.45, red: μ=0.52; dark colour: $\lambda_P = 0$, light colour $\lambda_P = 0.02$), other parameters as before. Parameters common to all three figures: $a_{xx} = 0.9, d = 0.03, \kappa = 1, \Theta = 1$. For all parameters combinations, the extinction state $(\overline{0})$ is a stable equilibrium.

scheme of the system reads:

$$
\begin{aligned}
2\,\mathrm{X} \; &\underset{b_{xx}}{\overset{a_{xx}}{\rightleftharpoons}}\; \mathrm{C}_{xx}
\begin{cases}
\xrightarrow{\kappa,\theta,(1-\mu)} 3\,\mathrm{X}\\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_S} 2\,\mathrm{X}+\mathrm{S}\\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_P} 2\,\mathrm{X}+\mathrm{P}\\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} 2\,\mathrm{X}+\mathrm{J}
\end{cases}\\[4pt]
\mathrm{P}+\mathrm{X} \; &\underset{b_{px}}{\overset{a_{px}}{\rightleftharpoons}}\; \mathrm{C}_{px}
\begin{cases}
\xrightarrow{\kappa,\theta,(1-\mu)} 2\,\mathrm{P}+\mathrm{X}\\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_S} \mathrm{P}+\mathrm{X}+\mathrm{S}\\
\xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} \mathrm{P}+\mathrm{X}+\mathrm{J}
\end{cases}\\[4pt]
\mathrm{X}+\mathrm{S} \; &\underset{b_{xs}}{\overset{a_{xs}}{\rightleftharpoons}}\; \mathrm{C}_{xs}\\[4pt]
\mathrm{P}+\mathrm{S} \; &\underset{b_{ps}}{\overset{a_{ps}}{\rightleftharpoons}}\; \mathrm{C}_{ps}\\[4pt]
\mathrm{X},\mathrm{S},\mathrm{P},\mathrm{J},&\mathrm{C}_{xx},\mathrm{C}_{px},\mathrm{C}_{xs},\mathrm{C}_{ps}\; \xrightarrow{d}\; \theta,
\end{aligned}
\tag{4.2}
$$
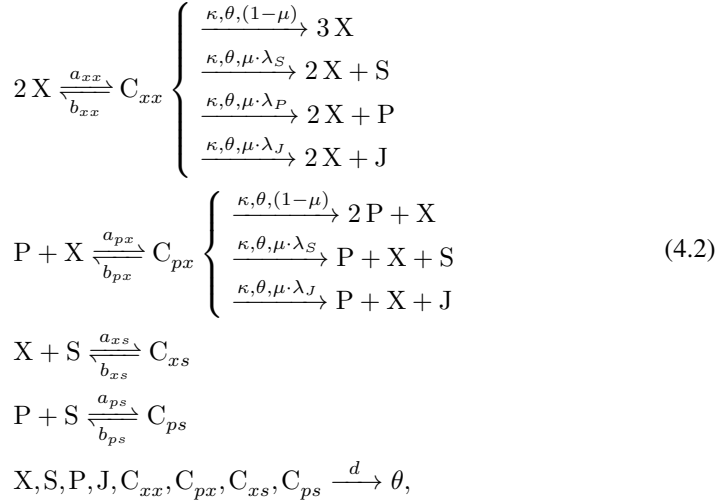
where the names of variables and parameters are assigned as above. The ODE system derived from this reaction scheme, as well as a summary of the bifurcation analysis, can be found in Supplementary Material.

The model formulated here is very similar to those studied in (Takeuchi and Hogeweg, 2007b). However, the main focus there was on understanding how the interplay of mutation rates with the affinity of the parasites for replicators affected the stability of the system (and in that sense, the comparison with the spatial system was made at a mesoscale level). Here, we are interested in the average behavior of the system at a small scale, where units of replication, stallers and parasites have similar dangling ends, making the rate of the reactions $\mathrm{X}+\mathrm{S}\xrightarrow{a_{xs}}\mathrm{C}_{xs}$ and $\mathrm{P}+\mathrm{S}\xrightarrow{a_{ps}}\mathrm{C}_{ps}$ comparable to $\mathrm{P}+\mathrm{X}\xrightarrow{a_{px}}\mathrm{C}_{px}$ and rather high. The important difference between the master sequence and the other units of replication (at high Hamming Distance) is that the latter have more stallers in the mutational neighbourhood than the former, i.e. the probability that a staller is produced (λ_S) is higher.

The system displays a wide range of behaviors. Fig. 4.10 gives an overview of the concentration of units of replication, X, as a function of the rate at which stallers are produced, λ_S, for different combinations of other parameters. For lower mutation rates, in the absence of parasites ($\lambda_P = 0$), units of replication are minimally hindered by stallers (uppermost line in the left and center pane of Fig. 4.10). However, if the fraction of mutants that turn into stallers is too low ($\lambda_S < \lambda_S(h)$), parasites can invade and lead the system to extinction. Increasing λ_S weakens progressively more the exploitation of units of replication by parasites (stable lines in the middle of left and center pane of

Fig. 4.10) until the point ($\lambda_S = \lambda_S(F_1)$) where parasites cannot invade any longer. A similar result holds if parasites are generated as mutants of units of replication (Fig. 4.10, center pane). A large production of stallers at higher mutation rates becomes deleterious for units of replication, and may lead the system to extinction ($\lambda_S > \lambda_S(F_2)$, Fig. 4.10, right pane).

In conclusion, stallers are unavoidable in the full model, especially at higher mutation rates, and they constitute a hinder for units of replication if abundant in the mutational neighbourhood, while providing a form of defense against parasites. The master sequence has a limited fraction of them in its mutational neighbourhood, which exposes it to the risk of invasion by parasites. However, mutant units of replication have a high fraction of stallers in their mutational neighbourhood, which protect the system from parasites by sequestering such units into complex with stallers, or by decreasing the chance that contacts happen between parasites and units of replication. In practice, such strategy favors the master sequence while hindering everybody else in the long run.

4.3.7 Global picture of quasispecies: a *functional* ecosystem

After having elucidated the functional role of non viable mutants, results can be summed up to present the general picture of the quasispecies. In Fig. 4.11 (upper pane), the quasispecies is presented as a network of mutationally adjacent units of replication. The only clearly abundant units of replication are those closest in Hamming Distance to the master sequence (Fig. 4.11, lower left pane): this is understandable because, on the one hand, they are often generated from the master sequence, on the other, and more importantly, their mutational neighbourhood is the most similar to that of the master sequence. Upon inspection, it is clear that already at $HD = 2$ no sequences are as abundant. The abundance of each unit of replication seems to correlate more with the abundance of helpers and with the scarcity of stallers in their mutational neighbourhood, rather than with the replication rate of the sequence (e.g. cf. the two units of replication at $HD = 1$ in orange at the bottom, with the neutral ones at the top, in yellow). The lower right pane of Fig. 4.11 shows the progression of mutant units of replication towards the periphery (in genotype space) of the quasispecies. The more Hamming Distance increases, the higher the fraction of stallers in the mutational neighbourhood becomes. This explains why units of replication closer to the master sequence are selected (in the short run), as seen in the field (Fig. 4.8), why the peak of the distribution of helpers is one step after (in Hamming Distance) from that of units of replication, and why the peak of the distribution of stallers is at higher Hamming Distances.

Altogether, the evolved quasispecies is under the mutational control of the master sequence, which minimizes the competition with other units of replication by its low degree of neutrality, minimizes the hinder from stallers while it maximizes the help received.
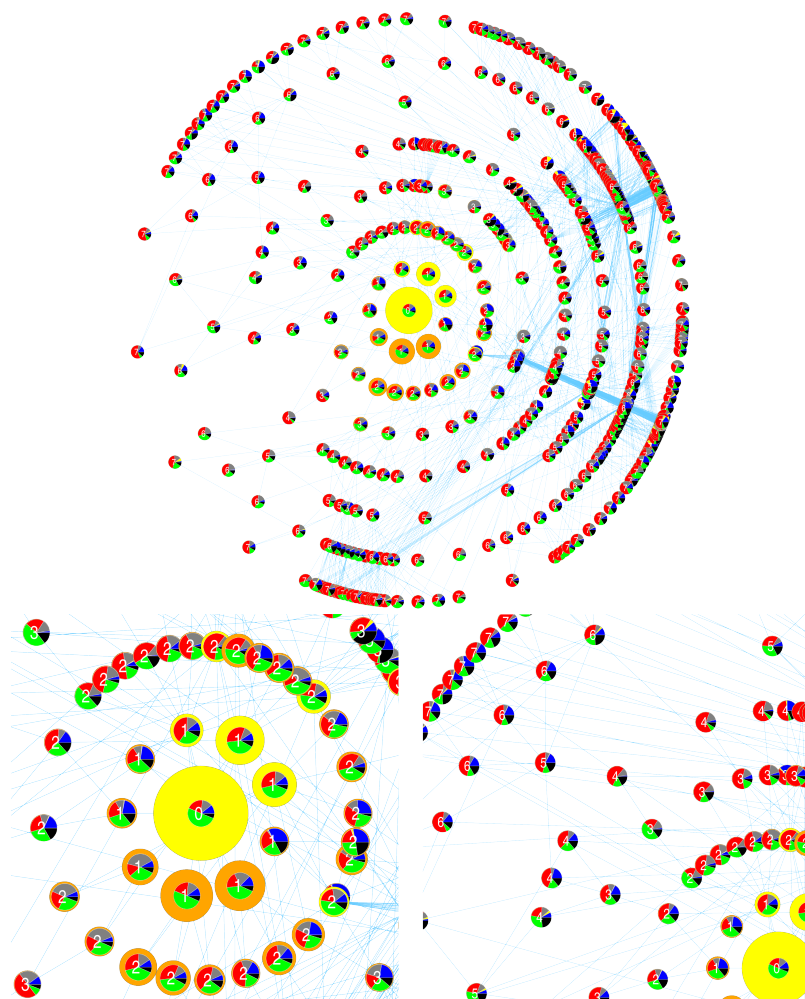
Figure 4.11. Overview of the units of replication in the quasispecies (top) and close-ups (bottom). Units of replication are generated by following the procedure explained in the caption of Fig. 4.4. Each node represents a unit of replication and each edge connects two units of replication if they are at Hamming Distance = 1 from each other. The number in the centre of the node is the Hamming Distance from the master sequence. Each node shows the mutational neighbourhood of the unit of replication it represents (the pie chart in the centre, colour coding as in Fig. 4.4), and its abundance (the size of the ring around), which is coloured according to whether it is a core neutral mutant of the master sequence (in yellow) or not (orange). The radial layout is meant to stress the central role the master sequence has in shaping the rest of the quasispecies. In order to make the figure clearer, units of replication are shown only up to Hamming Distance = 7. The network is visualised with Cytoscape 2.8.

Other units of replication are selected in the short run provided their mutational neighbourhood has a high enough fraction of helpers. However, at higher Hamming Distances the fraction of stallers in their mutational neighbourhood makes them *effectively* non viable. Helpers have been shown to be necessary for the survival of the whole system, whereas stallers contribute to the global stability of the quasispecies (i.e. the master sequence) by sequestering units of replication and by limiting their accessibility to parasites. No particular selection pressure is found to act on the frequency or the production of inert molecules (junk). In conclusion, the quasispecies behaves functionally like an ecosystem, where different emergent functions are acquired by non viable sequences and have particular and defined roles.

4.3.8   Generality of the results, steep vs. flat quasispecies

The initial step to evolve units of replications to higher mutation rates was repeated both here and in (Takeuchi and Hogeweg, 2008). Including the one described so far, a total of eight units of replication were evolved. These sequences can survive to mutation rates within the range $0.014 < \mu < 0.0165$ (see table 4.1). In most cases (6/8), these sequences exploit a catalytic strategy based on C's (as the unit of replication described above), and their mutational neighbourhood is characterized by a lower degree of neutrality, a higher fraction of helpers and a minimal fraction of parasites. Less frequently (2/8 cases), sequences exploit a catalytic strategy that seems to rely more on A's than other nucleotides, and, more importantly, they have a higher degree of neutrality and a lower fraction of helpers in their mutational neighbourhood, although parasites are still minimized.

The units of replication belonging to the steep quasispecies not always display a unique master sequence: in some cases (4., 5., and 6. in table 4.1) a small group of core neutral mutants with almost identical mutational neighbourhood substitutes it. However, the cases that resist to the highest mutation rates do have a unique master sequence (1., 2. and 3. in table 4.1).

Survival of the flattest   The two sequences with a higher degree of neutrality (7. and 8. in table 4.1) seem to be a qualitatively different outcome at high mutation rates. Their sequence composition is similar to the units of replication that evolve at lower mutation rates (the "A-catalyst" observed in (Takeuchi and Hogeweg, 2008)). In both cases, the quasispecies have no persisting master sequence, and the sequence variability in the field is high. Moreover, parasites are much more abundant in the field than in the case of steep quasispecies (see Supplementary Material Fig. 4.15). The characteristics of these two quasispecies clearly point to a survival of the flattest effect.

With this in mind, and given that the overall replication rate of steep and flat quasispecies

are comparable, we performed all the pairwise competitions between the flat quasispecies and the steep ones. In all cases but one the steep quasispecies outcompete the flat ones in little time (notice that in the case in which the steep quasispecies is outcompeted, its Error Threshold is very close to the mutation rates used for the competition experiment).

4.4 Discussion

In this study we investigated the eco-evolutionary dynamics at high mutation rates of interacting replicators with an explicitly defined genotype (RNA-like sequences), and a phenotype that depends on a complex genotype-to-phenotype map (RNA folding). We show that a quasispecies evolves which is steep and in which emergent functionalities are associated to non viable individuals, and critically contribute to the overall stability and long term persistence of the system. The information on the structure of the *functional* ecosystem is coded on the master sequence, and stochastically decoded through mutations.

A key feature of the model is that neither an interaction structure is preconceived, nor it is predefined which sequences are most fit (i.e. no fitness function is imposed on the system). All that is implemented are simple "chemistry" rules, i.e. sequence recognition for complex formation, which selection may exploit. Both complex formation and replication are local processes, which means that individuals interact mostly with identical copies of themselves as well as, inevitably, with their close mutants. This implies that interactions, when happening, are usually strong (i.e. if the appropriate dangling ends are present, the probability of complex formation and replication is high). Co-localization and strong interaction are at the basis of the observed evolutionary (positive) feedback: units of replication are selected (at least in the short run) if their close mutants replicate them more than the mutants themselves get replicated. This happens in two ways: close mutants (both in genotype and in physical space) are exploited for giving replication without being able to replicate (helpers), further away mutants (stallers) block competing units of replication and parasites. So, although genotypically very close, these sequences are phenotypically and functionally different. Altogether, the evolutionary structuring of this mutational process establishes the functional linkages within the quasispecies. The extent of the mutational control the master sequence reaches is surprisingly deep (Hamming Distance ≈ 10, where it becomes indistinguishable from random).

This model was initially studied in (Takeuchi and Hogeweg, 2008), where it was argued that the evolution of ecological complexity is limited at greater mutation rates, where a single, invariant quasispecies persists. Here we see that low diversity at sequence level enables a specific mutational neighbourhood and therewith a large functional diversity. The evolution of new lineages is prevented by the structured mutational neighbourhood at high mutation rates. At low mutation rates, the amount of functionally different mutants generated within the quasispecies is reduced, and separate lineages, e.g. parasites, readily

	mut. nei.	μ_{max}	replic. rates		competition	
			+/+	+/-	7	8
Steep quasispecies						
1.		0.0164	0.902 0.914	0.858 0.831	✓	✓
2.		0.0154	1.000 0.932	0.878 0.854	✓	✓
3.		0.0151	1.000 0.870	0.777 0.744	✓	✓
4.		0.0145	1.000 0.866	0.817 0.777	✓	✓
5.		0.0151	1.000 0.818	0.777 0.731	✓	✓
6.		0.0143	1.000 0.858	0.777 0.729	x	x
Flat quasispecies						
7.		0.0154	0.725 0.892	0.817 0.798		
8.		0.0149	0.902 0.872	0.817 0.792		

Table 4.1. Properties of units of replications that survive at $\mu \geq 0.014$. 1st column: pie charts of the HD=1 mutational neighbourhood of the master sequence (for case 4. and 5. no single master sequence is present, the pie chart is an average of the common ancestors along the line of descent collected every $5*10^4$ time steps. Analogously for case 6., where the most abundant individuals were collected every $5*10^4$ time step); 2nd column: value of μ between the maximum μ for which the system doesn't go extinct and the minimum μ for which it does, confidence interval: ± 0.0001; 3rd and 4th columns: Upper row: probability of complex formation of the master sequence (or the first common ancestor) with itself, lower row: average probability of complex formation for units of replication within the quasispecies. Left: catalytic strands with other catalytic strands, right: with complementary sequences (calculated by sampling $5*10^2$ units of replication, determining all vs. all probabilities, then averaging); 5th and 6th columns: competition experiment with *flat* quasispecies (7. and 8.). ✓: the steep one outcompetes the flat one, x: *vice versa* (the experiments are performed by initializing half of the field with one quasispecies and half with other, both taken from runs at sufficiently late time steps).

evolve.

In conclusion, mutation rate "automatically" classifies different eco-evolutionary dynamic regimes, even though the genotype-phenotype-interaction map is identical.

One-to-many genotype-to-phenotype map   The outcome of the evolutionary dynamics consists of a sequence which stores the information for the entire *functional* ecosystem. Functions are unfolded through minimal variations (mutations). In our model, substitutions are the only mechanism that can allow the spatial and functional control required to evolve the complex interaction structure observed, hence the need for high mutation rates. Altogether, this represents an example of a one-to-many genotype-to-phenotype map.

Other mechanisms to achieve multiple functionalities with limited coding resources (e.g. in proximity of the Information Threshold) have been explored *in silico* by (Hogeweg and Hesper, 1992), and more recently within the context of RNA folding in e.g. (Ancel and Fontana, 2000), (Fontana, 2002), (de Boer and Hogeweg, 2012) and (de Boer, 2012) (the phenomenon occurs frequently *in vivo* (Trifonov, 1989); see (Tuck and Tollervey, 2011) for a recent review about RNAs). While in those models the evolution of a one-to-many genotype-to-phenotype map was the object of study, in our model it was completely unexpected. Similarly to (de Boer and Hogeweg, 2012), our results make the point that mutation rate should be interpreted as a structural degree of freedom rather than only a limiting factor to information accumulation.

Steep vs. flat quasispecies   From a quasispecies-theoretical perspective, evolution exploits the very defining feature of quasispecies *sensu strictu*, namely, that mutations happen often and surely. In our models this produces an interaction structure. In combination with a complex genotype to phenotype map, some control over these mutants can emerge.

This cannot happen if the only selectable trait is replication rate: when mutations happen frequently (with large enough populations), populations evolving on neutral networks automatically increase their mutational robustness (Van Nimwegen *et al.*, 1999). Quasispecies models that display the survival of the flattest effect (originally (Schuster and Swetina, 1988), later emphasized by (Wilke *et al.*, 2001)) suffer from the same limitation. Indeed, the spatially extended versions of these models show that the range of mutation rates in which flat quasispecies outcompete steeper ones is larger than in the well mixed case (Sardanyés *et al.*, 2008). Multiple selection steps at different life stages can lead to favoring antirobust individuals (as they purge deleterious mutations faster), but this is not the case at higher mutation rates (Archetti, 2009).

In contrast, our model shows that the evolution of an interaction structure and the lack of neutrality are closely linked. This explains why the increase in mutational robustness is not observed in most cases, and why, for the cases in which flatter solutions evolve, they

are most often outcompeted by quasispecies with a highly structured mutational neighbourhood. In other words, survival of the flattest does not happen. Sequences in flat quasispecies are, to some extent, under the same selection pressure as the steep ones, i.e. selection for a structured mutational neighbourhood (which can be seen from the moderate fraction of helpers and the low fraction of parasites in their mutational neighbourhood, in Table 4.1). In this sense, it seems that flat quasispecies are somewhat deregulated solutions in comparison to the steep ones. Nonetheless, flatter quasispecies retain a high maximum replication rate, comparable to that of steep solutions (in contrast to (Elena *et al.*, 2007)), while the average replication rate is even higher (in Table 4.1 the average replication rate is calculated taking into account only units of replication, and not non viable individuals). This means that higher neutrality in the quasispecies does not come about due to a less effective selection pressure, as is the case in (Krakauer and Plotkin, 2002) as well as in the Royal Road Genetic Algorithm ((Mitchell *et al.*, 1992),(van Nimwegen and Crutchfield, 2000)). The flat quasispecies are outcompeted because they lack the (evolved) properties of the steep ones, namely, a structured mutational neighbourhood which establishes a functional linkage within the quasispecies.

It has been shown (Huynen, 1993) that in predator-prey dynamics with RNA-like organisms (in which the prey is not eaten if it is not "recognized", i.e. if the predator's phenotype does not match that of the prey), populations evolve to steeper regions of the phenotype landscape and the increase in mutational robustness does not happen. Selection acts to rapidly change phenotypes so that individuals that can change the fastest (i.e. those that are in the least neutral portion of the neutral network for a given secondary structure) will have an advantage.

The evolutionary dynamics in that model is very different from ours. In that case it is Red Queen dynamics ((Van Valen, 1973)), while in our case a steady state-like solution is evolved (moreover we do not preconceive any interaction structure). However both models make the point that interacting individuals characterised by a phenotype determined with a complex map (e.g. RNA folding) may evolve to lower degrees of mutational robustness at high mutation rates, as these solutions are more versatile.

The effect of lethal mutants   In our model (and conceivably with RNA in general), the loss of viability as a result of mutations manifests at the phenotype level, when the sequence has already been generated. This means that lethal phenotypes (in a broad sense of the word) do not die the moment they are born (which is, instead, the customary modelling approach in quasispecies theory, e.g. in (Takeuchi and Hogeweg, 2007a), (Kirakosyan *et al.*, 2010), but not in (Tejero *et al.*, 2010)), while it still remains that the fitness advantage of the individual that produced them is infinite. The consequences of this (in our model) are important: 1) viable individuals evolve to increase the fraction of lethal (non viable) sequences in their mutational neighbourhood, with the latter acquiring novel functionalities, 2) the exploitation of non viable mutants by units of replication to outcompete other viable individuals becomes more extreme the higher the mutation

rates. A remarkable side effect is that neither delocalization nor an Error Threshold *sensu strictu* happen. For higher mutation rates, the quasispecies hits an extinction threshold when the viable individuals do not produce enough viable offsprings.

In conclusion, the presence or absence of delocalization depends on the evolutionary dynamics of the quasispecies. This provides a counter example to the claim made in (Holmes, 2010), that the "Extinction Threshold" does not rely on quasispecies dynamics, although our model is obviously not explicitly meant to address the quasispecies dynamics of RNA viruses.

Evolution and persistence of diversity   Since the original work on the Hypercycle (Eigen and Schuster, 1978), the problem posed by the Information Threshold has often been addressed as the problem of maintaining (functional) diversity, i.e. in terms of persistence of independent lineage (e.g. (Happel and Stadler, 1998), (Szathmáry, 1991), (Boerlijst and Hogeweg, 1991c)) or evolving it (e.g. (Hogeweg, 1994)).

It was recently shown that genotypic diversity could be maintained if RNA replicators exploited different nucleotide compositions (Szilágyi *et al.*, 2013), although in that case coexistence seems to be very sensitive to single substitutions. In (Ma and Hu, 2012), a system of functional molecules which cooperate to perform the various steps needed for (protocell) self replication, could spread without loosing diversity. In (Könny and Czárán, 2013) a large number of different species are shown to coexist in a spatially extended system if they are all necessary to the (local) production of resources they all (locally) exploit, even if parasites evolve (depending on the diffusion rates of resources and of replicators). However, (to the best of our knowledge) the evolutionary maintenance of the metabolic diversity is unclear. Nevertheless, within the framework, parasites were shown to acquire replicase activity, thus increasing the overall complexity of the system (Könnyű *et al.*, 2008). The stabilisation of replicators/parasites interactions in spatially extended systems is mediated by spatial pattern formation. The eco-evolutionary dynamics of such systems have been studied in depth in (Takeuchi and Hogeweg, 2009), where it was shown that travelling waves constitute a higher level of selection than that of individual replicators and parasites. In (de Boer and Hogeweg, 2010) spatial pattern formation mediated the evolution of an ecosystem based information-processing system at high mutation rates. In our case, differently functional individuals and their interactions are emergent phenomena. However, different functions are performed by similar genotypes, hence diversity is phenotypic rather than genotypic.

Division of labour   From a more general biological perspective, the ecoevolutionary dynamics presented here represent a form of (partial) reproductive division of labour. The evolution of division of labour in an RNA-like system, from a self replicating molecule to a transcription-like mechanism (i.e. with templates and polymerases), has been shown to evolve because it confers an increased resistance to parasites (Takeuchi *et al.*,

2011). In this respect, we present a different mechanism to achieve this. What we observe is twofold: 1) a line of descent of (almost) identical genotypes (the master sequences) carrying the information for their own survival, as well as that for the functions and spatial organization of the mutants and 2) a multitude of mostly non viable genotypes very similar to those along the line of descent, with very different phenotypes, which come from and aid the further propagation of the master sequences. In this sense, our results are reminiscent of the reproductive division of labour in social insects or that of germline and soma in (possibly early evolutionary stages of) developmental processes, although in those cases regulation instead of mutation underlies the differentiation process.

4.5 Conclusions

Although the details of the results presented here are specific to the arbitrary chemistry implemented, we maintain that the conclusions are general in that:

- A quasispecies evolves which behaves like an ecosystem,

- The emergent functions are carried out by non viable sequences,

- It constitutes an individually coded, but stochastically decoded ecosystem based solution,

- It exploits the frequently arising mutations by evolving to regions of the genotype-to-phenotype map where small genotypic change produce large phenotypic differences.

4.6 Supplementary material

4.6.1 Global distribution of functional classes in the field vs. genotype space

4.6.2 Evolution to high mutation rates

As mentioned in the main text, initializing the system at high mutation rates ($\mu = 0.014$ or higher) with randomly generated units of replication always leads to a quick extinction. Here, results are reported about the evolutionary run that leads to sequences able to sustain high mutation rates. A homogeneous population of units of replication is used to initialise the system. The mutation rate (μ) assigned is sligthly greater than zero. μ determines the fidelity by which a unit of replication copies another sequence. When mutations occur, μ is changed as well, by adding a tiny but strictly positive number.
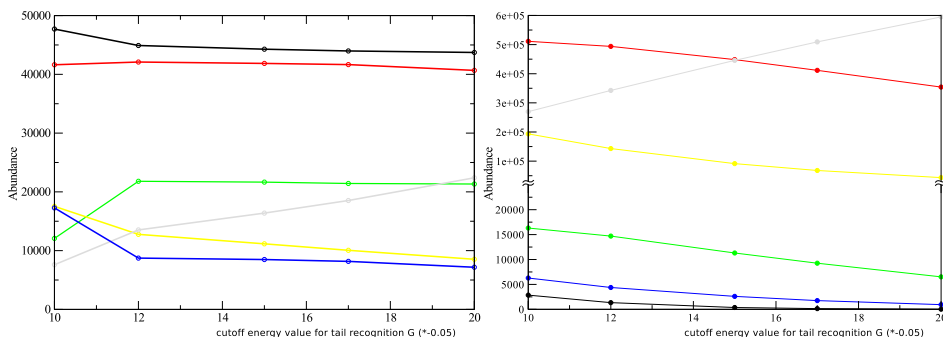
Figure 4.12. (In)sensitivity of functional classes to the cutoff parameter. Global distribution of functional classes in the CA (left) and in entire genotype space (right). The different abundances in functional classes are shown as function of the cutoff score used to determine presence/absence of dangling ends, in the main text the value corresponding to 15 is used as default (see Models and Methods). Left pane: at one time point of a standard simulation (as described in the main text), all sequences are collected to produce the figure. Right pane: 10^6 random sequences are generated for the figure. Notice that the scale of the two y axis is different.

An alternative method consists of imposing on all sequences an external mutation rate and setting it to progessively higher values (in little steps). Results do not differ more between runs wich implements different methods than they differ for two runs with the same method. In Fig. 4.13, one such pre-evolution runs is shown. The distribution of mutation rates is plotted in time for both Units of replication (red) and Parasites (black).

During the pre-evolution step, we distinguish three time-interval groups, on the basis of the dynamic regime the system displays. These three intervals can be characterized by the mutation rates of the individual lineages present during such interval. Patterns are distinguished by clustering the sequences and analyzing their phenotypes, which, as will become clear in the following, corresponds to characterize their ecological role. The three time periods (in arbitrary time units) are 1) Red Queen dynamics: $0 < time < 600 * 10^3$, 2) Multi-species dynamics $600 * 10^3 < time < 1300 * 10^3$, 3) Single (quasi-)species dynamics $time > 1300 * 10^3$. Time boundaries are somewhat arbitrary, as it is not possible to uniquely locate in time when the system changes dynamic regime.

Red Queen dynamics   For the first $600 * 10^3$ time steps, the average mutation rate for the catalytic sequences is low ($0 < \mu < 0.003$). As can be seen from Fig. 4.13 (black box), at time $500 * 10^3$ three different lineages of Units of replication coexist with a parasitic one. All of them can be characterized by a particular nucleotide preference on the 5' dangling end of the catalytic strand (depicted in the figure), or, for the Parasite, the 3' dangling end. All the catalytic lineages have about the same mutation rate, and the variability in the field is limited (see the sequence Logo's). However, the system is not

stable and these lineages are not persisting, with the exception of the unit of replication class characterized by 'C' nucleotide. Since Parasites descend from Units of replication, they are assigned a mutation rate, which they inherit generation after generation. Such parameter does not have any actual use for Parasites, because they are not catalytic, but it still mutates with the same algorithm by which a μ assigned to a functional molecule would. Therefore, given that mutation rate is biased to only increase, the exponentially increasing curve of the parasitic branch represents drift. Altogether, the evolutionary dynamics seems to be dominated by a Red Queen-like effect (Van Valen, 1973), with different catalytic strategies rapidly changing in response to the arise of parasitic lineages.

Multi-species dynamics   In the second time period, between $600 * 10^3$ and $1300 * 10^3$ time steps, the nucleotide preferences in the lineages become more predictable. When four species are present, the 5' dangling end of all units of replication are composed either by a majority of 'C' or 'A', while the parasites have 3' dangling ends composed either by 'G' or by 'U's. The C-unit of replications are exploited by G-parasites, while the A-units are exploited by the U-parasites (Fig. 4.13 red box). In the CA this gives rise to mutually invading travelling waves, where each catalyst grows in the empty space left by the parasite it is not exploited by. When the A-units of replication reach high mutation rates, first the U-parasite goes extinct, then the A-units disappear. The C-units maintain a much lower mutation rate and, since the G-parasite is still present, this opens a niche for a new A-unit of replication to evolve, quickly increase the mutation rate, reach the Error Threshold and disappear again. The cycle of evolution and extinction of the A-catalytic lineage repeats a few times. This can be explained as follows: while the C-unit/G-parasite system is present, mutations that turn the C nucleotide preference to the A nucleotide are selected, as they allow a progressively lower exploitation by the G-parasite. However, since mutation rate is biased to increase upon substitutions, the same process that drives the evolution of an A-lineage, also drives it quickly to hit its Error Threshold and make it disappear, opening the niche again for a new A-lineage. However, when mutation rate becomes sufficiently high for the C-unit of replication, new A-catalytic lineages cannot appear any longer and only two species coexist (Fig. 4.13 blue box).

Single (quasi-)species dynamics   In the last time period ($time > 1300 * 10^3$) mutation rate is greater than 0.012, and the system is characterized by a single quasispecies. The structure, the selection pressures and the survival mechanism of this quasispecies are the concern of the main text of the paper.

4.6.3   Snapshot of the field for the quasispecies described in the main text

See figure 4.14.
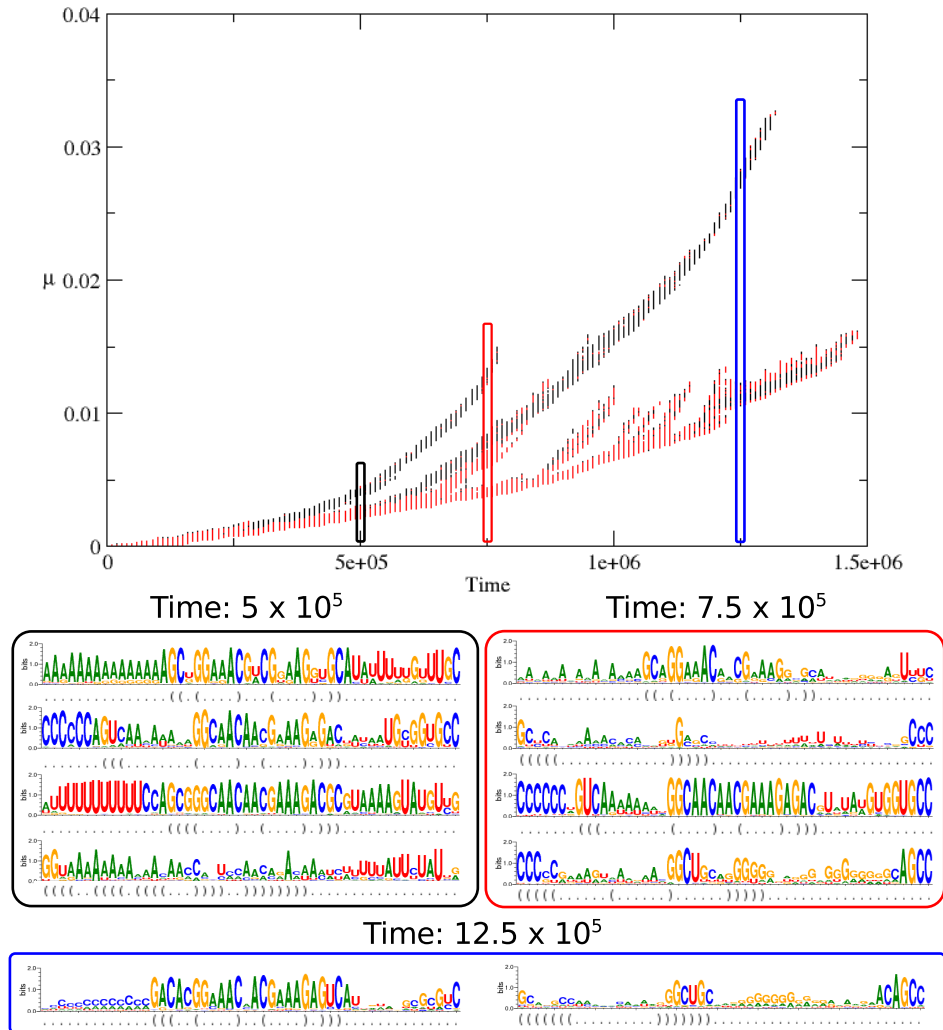
Figure 4.13. Distribution of mutation rates in time and species composition at different time points. Top pane: every 10^4 timesteps, the mutation rate of each sequence is collected and plotted. Red: catalytic sequences, black: non catalytic sequences. At three different time points, corresponding to different dynamic regimes, a random sample of sequences of the same strand (chosen so that the units of replication display the catalytic one) are clustered and the sequence logo is obtained for each quasispecies (see caption of Fig. 4.3 in the main text).
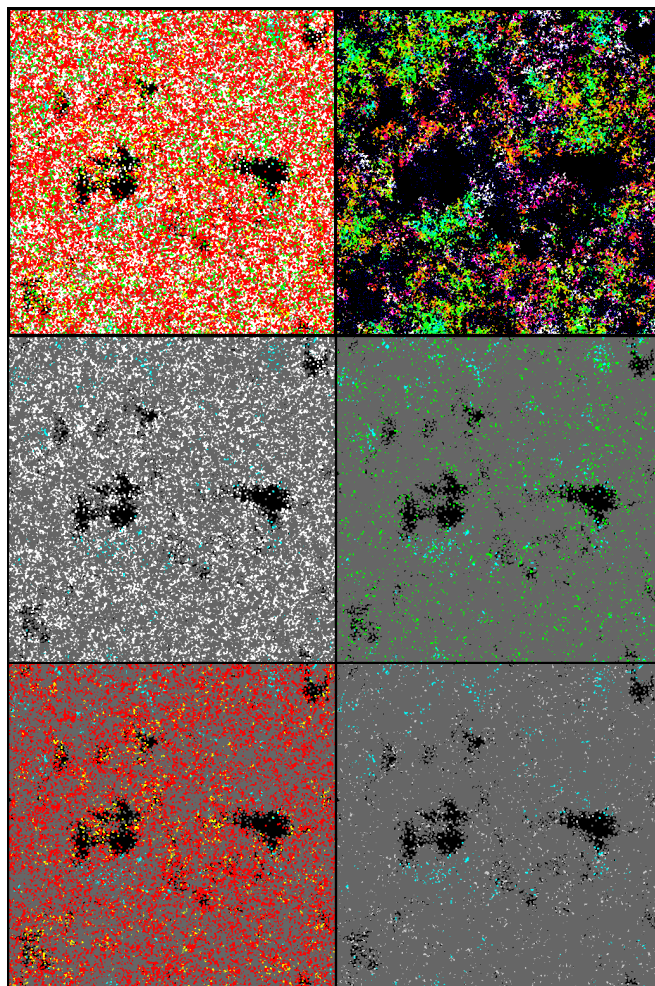
Figure 4.14. A snapshot of the field: all figures are at the same time point, but coloured differently for each functional class. **Upper-left**: the full field; colours as in the main text. **Upper-right**: The Hamming Distance of all sequences from the master sequence (cyan) is depicted. The colours follow the gradient cyan-green-yellow-red-magenta, corresponding to Hamming Distance 0 to 7. In white, Hamming Distance > 7. Blue: parasites. For all that follows: copies of the master sequence are in cyan, all the classes not explicitly mentioned are in dark grey, background is in black. **Centre-left**: units of replication (white). **Centre-right**: helpers (green). **Bottom-left**: stallers (red), parasites (yellow). **Bottom-right**: junk (light grey). Colouring is as follows: for each cell the functional classes in the Moore neighbourhood are determined (excluding empty cells, except in the upper right figure), and the whole neighbourhood is coloured according to the (functional or Hamming Distance) class in the majority. Notice that the density in the field seems higher than what it actually is.

4.6.4 ODE system with helpers

The ODE system derived here follows the reaction scheme presented in the main text (reaction scheme 4.1). For all second order reactions, mass action is assumed:

$$
\begin{aligned}
\dot{x} &= -2a_{xx}x^2 + (2b_{xx} + 3\kappa\theta(1-\mu) + 2\kappa\theta\mu)c_{xx} \\
&\quad -a_{xh}xh + (b_{xh} + 2\kappa\theta(1-\mu) + \kappa\theta\mu)c_{xh} \\
&\quad -a_{px}px + (b_{px} + \kappa\theta)c_{px} - dx \\
\dot{h} &= \kappa\theta\mu\lambda_h c_{xx} - a_{xh}xh + (b_{xh} + \kappa\theta(1-\mu) + \kappa\theta\mu(1+\lambda_h))c_{xh} \\
&\quad -a_{ph}ph + (b_{ph} + \kappa\theta)c_{ph} - dh \\
\dot{p} &= \kappa\theta\mu\lambda_p c_{xx} + \kappa\theta\mu\lambda_p c_{xh} - a_{px}px + (b_{px} + 2\kappa\theta)c_{px} \\
&\quad -a_{ph}ph + (b_{ph} + 2\kappa\theta)c_{ph} - dp \\
\dot{j} &= \kappa\theta\mu\lambda_j c_{xx} + \kappa\theta\mu\lambda_j c_{xh} - dj \\
\dot{c}_{xx} &= a_{xx}x^2 + b_{xx}c_{xx} - \kappa\theta c_{xx} - dc_{xx} \\
\dot{c}_{xh} &= a_{xh}xh + b_{xh}c_{xh} - \kappa\theta c_{xh} - dc_{xh} \\
\dot{c}_{px} &= a_{px}px + b_{px}c_{px} - \kappa\theta c_{px} - dc_{px} \\
\dot{c}_{ph} &= a_{ph}ph + b_{ph}c_{ph} - \kappa\theta c_{ph} - dc_{ph}
\end{aligned}
$$

where $\theta = 1 - (x+h+p+j+c_{xx}+c_{xh}+c_{px}+c_{ph})/\Theta$, $\lambda_h + \lambda_p + \lambda_j = 1$, $b = 1-a$ and $\kappa = 1$. For all μ, $(\overrightarrow{0})$ is a stable equilibrium. Progressively increasing μ, the non trivial, stable equilibrium (solid line in Fig. 4.9) is at first a fixed point, then a supercritical Hopf bifurcation occurs and a stable limit cycle appears. Further increasing μ, the limit cycle disappears (most likely via a homoclinic bifurcation, when the minimum of the limit cycle touches the line of unstable equilibria) and the only stable equilibrium remaining is $(\overrightarrow{0})$, i.e. the system goes extinct (but notice that the difference between the Hopf bifurcation and the homoclinic bifurcation is always very small). Different fraction of mutants turning into helpers λ_H change the bifurcation plot quantitatively. Removing parasites from all the equations (Fig. 4.9, right pane), changes the bifurcation plot both quantitatively and qualitatively in that both Hopf and homoclininc bifurcations disappear and extinction at high μ is now mediated by a fold bifurcation.

4.6.5 ODE system with stallers

Analogously to the case with helpers, the ODE system derived for stallers (from the reaction scheme 4.2) reads:

$$
\begin{aligned}
\dot{x} &= -2a_{xx}x^2 + (2b_{xx} + 3\kappa\theta(1-\mu) + 2\kappa\theta\mu)c_{xx} \\
&\quad -a_{px}px + (b_{px} + \kappa\theta)c_{px} - a_{xs}xs + b_{xs}c_{xs} - dx \\
\dot{s} &= \kappa\theta\mu\lambda_s(c_{xx} + c_{xp}) - a_{xs}xs + b_{xs}c_{xs} - a_{ps}ps + b_{ps}c_{ps} - ds \\
\dot{p} &= \kappa\theta\mu\lambda_p c_{xx} - a_{px}px + (b_{px} + 2\kappa\theta(1-\mu) + \kappa\theta\mu)c_{px} \\
&\quad -a_{ps}ps + b_{ps}c_{ps} - dp \\
\dot{j} &= \kappa\theta\mu\lambda_j(c_{xx} + c_{xp}) - dj \\
\dot{c}_{xx} &= a_{xx}x^2 + b_{xx}c_{xx} - \kappa\theta c_{xx} - dc_{xx} \\
\dot{c}_{xs} &= a_{xs}xs + b_{xs}c_{xs} - dc_{xs} \\
\dot{c}_{px} &= a_{px}px + b_{px}c_{px} - \kappa\theta c_{px} - dc_{px} \\
\dot{c}_{ps} &= a_{ps}ps + b_{ps}c_{ps} - dc_{ps}
\end{aligned}
$$

where $\theta = 1 - (x + s + p + j + c_{xx} + c_{xs} + c_{px} + c_{ps})/\Theta$, $\lambda_S + \lambda_P + \lambda_J = 1$, $b = 1 - a$ and $\kappa = 1$. For all parameters combinations, the extinction state $((\overrightarrow{0}))$ is a stable equilibrium.

Higher a_{**} values, $\lambda_P = 0$ See Fig. 4.10, left pane. For high λ_S, only one nonzero stable equilibrium is present, which corresponds to a situtation where parasites are out-competed and go extinct. The point F_1 marks a fold bifurcation, after which, following the lower branch, a stable equilibrium emerges and parasites can coexist with replicators. Until the transcritical bifurcation (T), the system is bistable. The upper unstable branch departing from T corresponds to a situation where the system can be invaded by parasites (hence its instability), but is otherwise stable without (the other stable branch of the transcritical bifurcation is not included in the figure because equilibrium values for parasites become negative). For lower values of λ_S a Hopf bifurcation happens (H), with units of replication and parasites coexisting on a stable limit cycle, followed by a (possible) homoclinic bifurcation (h), when the minimum of the limit cycle hits the lower unstable manifold. After this value, depending on initial conditions (i.e. if parasites are present or not), the system will either converge to the state where no parasites are present, or will go extinct.

Lower a_{**}, different mutation rates and $\lambda_P \neq 0$ See Fig. 4.10, middle pane. Changing a_{**} and μ produce only quantitative differences, except at lower λ_S where the Hopf and the homoclinic bifurcation may disappear. However, setting $\lambda_P > 0$ changes the bifurcation plot qualitatively: both the transcritical and the fold bifurcation disappear,

as well as the (unstable) equilibrium with no parasites at low λ_S, because parasites are always produced from units of replication. Despite this, parasites are present in very low numbers at higher λ_S, meaning that stallers effectively limit them.

High μ   See Fig. 4.10, right pane. Pushing the system to higher μ has the effect of making $\bar{X}$ more dependent on λ_S as stallers are produced more often. F_2 marks the limit fraction of stallers that can be produced after which the system goes extinct (via fold bifurcation). For the highest μ parasites, even if arising as mutants of units of replication, cannot invade.

4.6.6   Snapshot of the field for a flat quasispecies
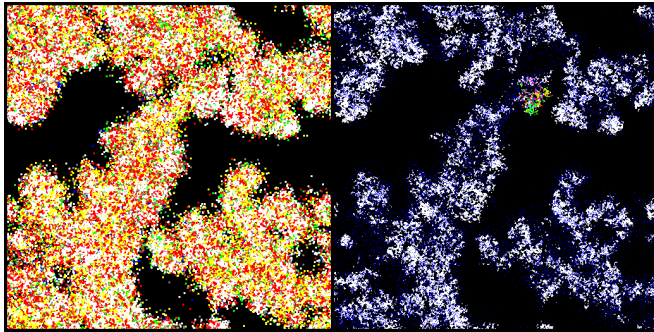
Figure 4.15. A snapshot of the field for the neutral quasispecies (7. in Table 4.1, in the main text): the figures correspond to the first two in Fig. 4.14. The colouring is the same. For the master sequence, the most abundant sequence in the field is used.

4.6.7   Sequences mentioned in Table 1

1. CCCCCCCCCCCCCGACACGGAAACGACGUGAGAGUCAUUAGAUAGGUGUC

2. CCCCCCCCCCCCCCGGCCGGAAACAACGUAAGAGCCAUUGUGUGGAUGCC

3. CCCCCCCCCCCAGCACCGGAAACAACGAAAGUACGCUGAAUGAGUGGUGC

4. ACCCCCCCCCCAGGCAACGAAAGACGAAAGAACGCCAUUGAGUGUGUGCC

5. CCCCCCCCCCGGCACAGGAAACAAAGCUAAAAUGCAGCCAUGCGGUGUGC

6. CCCCCCCCCCCGCAACGGCAACCGGAAAUCCAGCGACUUGAGUGUGUGCC

7. AACACUUCUACCCAGGCAAGGAAACACGGAAACAGCCAUCUUUUACUGCC

8. GAAAAAAAGAACAGAGCCAGGGGACACGUAAGGAGGCAUUAUACUAUGGC

Acknowledgment

5

Increased rate of duplications and deletions prevents evolutionary deterioration

Understanding the mutational dynamics of yeast rDNA

Enrico Sandro Colizzi, Paulien Hogeweg (2016)

In preparation

Abstract

Clashes between transcription and replication on heavily transcribed genes can lead to point mutations as well as chromosome rearrangements. In yeast, transcription-replication conflicts lead to copy number variation in the ribosomal genes. Surprisingly, their occurrence seem to reflect physiological strategies which are beneficial only in the long (evolutionary) term.

In order to study how the mutational dynamics observed in yeast affect genomes' evolutionary stability, we developed an *in silico* evolutionary simulation system where single-cell organisms undergo mutations more frequently when their transcriptional load is larger. We show that mutations induced by high transcriptional load are beneficial when biased towards duplications and deletions because they decrease mutational load, even though they increase the overall mutation rates. Moreover, evolving a larger proportion of transcriptional duplications allows organisms to maintain high fitness in the presence of random, life-history independent high rates of deletions and deleterious mutations, as is the case for yeast rDNA.

Our results show that the mutational dynamics observed in yeast are beneficial for the long term stability of the genome, and pave the way for a theory of evolution of mutational biases.

5.1 Introduction

In both Bacteria and Eukaryotes, genes that must be actively transcribed throughout a cell's life time may become the stage for clashes between transcription and replication complexes (Brambati *et al.*, 2014, Kim and Jinks-Robertson, 2012, Merrikh *et al.*, 2012). These clashes are known to destabilise genome and epigenome, e.g. by inducing single and double strand breaks (Takeuchi *et al.*, 2003), by generating R loops (mRNA-DNA hybrids) (Brambati *et al.*, 2014) and interfering with the correct depositions of histones (Castellano-Pozo *et al.*, 2013, Herrera-Moyano *et al.*, 2014). Several strategies have evolved to correctly schedule transcription and replication so that collisions are avoided. Some examples involve slowing down the replication fork or stalling it at the beginning of actively transcribed genes by means of Replication Fork Barriers (Labib and Hodgson, 2007); alternatively, heavily transcribed genes may be translocated closer to nuclear pores to release their mRNAs faster (Bermejo *et al.*, 2011), or removing transcription complexes altogether (including degrading the nascent mRNA) (Brambati *et al.*, 2014).

This notwithstanding, conflicts do occur and often results in mutations, ranging from single nucleotide substitutions to duplications and deletion due to recombination (Kim and Jinks-Robertson, 2012, Sankar *et al.*, 2016, Takeuchi *et al.*, 2003).

One of the best characterised example of transcription-replication conflicts comes from the ribosomal RNA genes of *Saccharomyces cerevisiae*. Ribosomal RNA genes in yeast are organised in several identical repeats - about 150 to 180 in wild-type - on the XII chromosome. Their expression is independent of copy number, and is instead under control of a dedicated polymerases (PolI) (French *et al.*, 2003). Indeed, deleting a large number of rDNA copies either "physiologically" by homologous recombination (Sinclair and Guarente, 1997), or artificially does not alter yeast's growth rate (French *et al.*, 2003). Because few rDNA copies impose a larger transcriptional load, transcription-replication conflicts occur more frequently (Ide *et al.*, 2010, Kim *et al.*, 2007). The outcome of these mutations is often recombination, mediated by a Replication Fork Barrier (Takeuchi *et al.*, 2003). Interestingly, recombination events lead frequently to duplications, which restore the wild-type rDNA copy number distribution after several generations (Jack *et al.*, 2015, Kobayashi *et al.*, 1998).

It has been recently discovered that the TOR pathway (a ubiquitous nutrient signalling pathway) controls ribosomal RNA gene duplications in yeast during caloric excess (Jack *et al.*, 2015). The signalling cascade leads to a larger rate of double strand breaks, which are then repaired by non-homologous recombination and results into a larger rate of duplications when the rDNA copy number is small.

Altogether, yeast evolved a nutrient-dependent mechanism which causes mutations that increase the number of rDNA repeats in the absence of short-term selection. As such, this is a prime experimental example of evolution of evolvability (Crombach and Hogeweg, 2008, Hindré *et al.*, 2012).

With yeast rDNA dynamics in mind, we make an *in silico* evolutionary model which integrates the effects of metabolism and mutations, in order to understand the functional significance of the observed mutational regime in yeast.

We show that mutations induced by high transcriptional load reduce mutational load when they are biased towards duplications and deletions, even though they increase the overall mutation rates. Moreover, we show that evolving a larger proportion of transcriptional duplications allows organisms to maintain high fitness when random deletions and deleterious mutations occur with high frequency, as is the case for yeast rDNA.

Results are organized as follows: In the first part, we study the evolutionary dynamics of the system with large or small influx of resources, including two different mutational schemes: either only life-history independent (background) mutations, or in combination with mutations caused by large transcription rates. By comparing these systems we show how selection for growth in a richer environment in combination with higher mutation rates leads to a large genome composed mostly of inactive genes and low fitness.

Next, we incorporate the effects of intergenic fragile sites (genomic break point), which increase mutation rates and bias mutational outcomes towards duplication and deletions. We make the case that intergenic break points can be selected because they reduce mutational load.

Finally, by studying a system where transcription-induced mutations compensate biases in background mutations, we show how the mutational dynamics observed in yeast's rDNA contribute to long-term genome stabilty and fitness (Jack *et al.*, 2015, Kim *et al.*, 2007, Takeuchi *et al.*, 2003).

5.2   Model

The model is loosely inspired from (Scott *et al.*, 2010), and consists of a population of single-cell asexual individuals. Fig. 5.1 shows a cartoon of population and cell dynamics. Each cell has a genome, a proteome and is capable of regulation. Cells uptake resources from the environment and convert them into aminoacids, which are used as building blocks for proteins, as well as signalling molecules for regulation of gene expression. Growing cells must reach a target volume (an increasing function of genome size, specified below), at which they can replicate (i.e. they are divided in two daughter cells), to begin the cycle anew. Fitness is based on inter-division time and internal homeostasis and population size is kept constant.

The cellular proteome consists of enzymes, housekeeping proteins and ribosomal proteins. Enzyme convert resources into aminoacids, housekeeping proteins must be kept at homeostatic concentration while (partially) determining transcription rates and cell fitness, and ribosomes -the combination of ribosomal proteins and ribosomal RNA- trans-

late mRNAs into their respective proteins.

Transcription is regulated by a linear function that combines basal transcription rates and metabolite- (i.e. aminoacid-) dependent transcription rates, it is reduced when the cell is far from homeostasis in housekeeping proteins, and is capped by a maximum per-gene transcriptional load.

There are four types of genes, one for each macro-molecule. The genome is a string of genes, which can be "active" or mutated into a non-functional state.

The sum of proteins and ribosomal RNA determines the volume of the cell. Genome size determines the target volume at which a growing cell divides, producing two cells. For the sake of simplicity, only one of the two cells is mutated. Two types of mutational events affect the genome: random mutations independent of the cell life-history, and mutations due conflicts between transcription and replication complexes, occurring on highly transcribed genes. The effect of a mutation can be gene duplication, deletion and inactivation. Once inactivated, genes cannot become active again. Mutations of the regulation parameters also occur during replication, by adding a small random number taken from a uniform distribution centred in zero. Where specified, the relative proportion of duplications, deletions and inactivations can also mutate.

Overview of intracellular dynamics   The intracellular dynamics are modelled in terms of Ordinary Differential Equations (ODEs), as follows:

$$
\begin{aligned}
\dot{S} &= S_{in} - d_s S - F_{\text{met}} \\
\dot{A} &= -d_a A + F_{\text{met}} - F_{\text{transl}}(m_T) - F_{\text{transl}}(m_Q) - F_{\text{transl}}(m_{R_p})
\end{aligned}
$$

$$
\begin{aligned}
\dot{R}_r &= F_{\text{transcr}}(R_r) - d_{R_r} \\
\dot{m}_T &= F_{\text{transcr}}(T) - d_{m_T} \\
\dot{m}_Q &= F_{\text{transcr}}(Q) - d_{m_Q} \\
\dot{m}_{R_p} &= F_{\text{transcr}}(R_p) - d_{m_{R_p}}
\end{aligned}
$$

$$
\begin{aligned}
\dot{T} &= F_{\text{transl}}(m_T) - d_t T \\
\dot{Q} &= F_{\text{transl}}(m_Q) - d_q Q \\
\dot{R}_p &= F_{\text{transl}}(m_{R_p}) - d_{R_p} R_p
\end{aligned}
$$

where S represents resources, A aminoacids (products of metabolism), R_r ribosomal RNA and R_p ribosomal proteins, T enzymes, Q housekeeping proteins, and m_X the mRNA corresponding to each macromolecules.
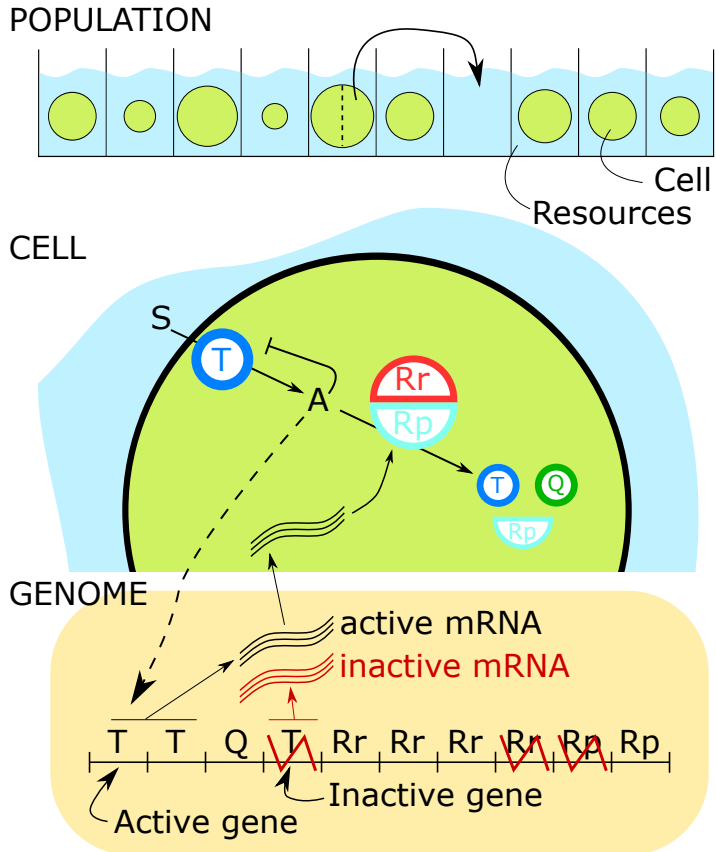
Figure 5.1. A cartoon of the model. See Methods for details. S: resources, A aminoacids, T: enzymes, Q housekeeping proteins, Rr ribosomal RNA, Rp ribosomal proteins.

Functions are defined as follows:

- Metabolism, the enzymatic conversion of resources to aminoacids, is defined as $F_{\text{met}} = k_t T \frac{\alpha_t S}{\alpha_t S + \beta_t A + 1}$, where S are resources, A are aminoacids, T are enzymes, and the functional form of metabolism assumes Michaelis-Menten kinetics with product-inhibition.

- Transcription of a gene X (where X can be enzymes, housekeeping proteins, ribosomal RNA or ribosomal proteins) depends on regulation parameters, amino acids and homeostasis requirements, and is capped at a maximum per-gene transcription rate, as follows. Every integration step, a "Transcriptional demand" t_X is generated as a linear function of aminoacids $t_X = k_0 + A * k_1$, where k_0 and k_1 are regulation parameters, they are evolvable and take real values. The transcriptional demand is scaled by the "health" of a cell, i.e. a function h ($0 < h < 1$) that measures the distance of Housekeeping protein concentration from an arbitrary homoeostatic optimum. If $[Q]$ denotes the concentration of housekeeping proteins in a cell volume V (the sum of the proteins and the ribosomal RNA) $h(Q) = \exp(-k_h([Q]-Q_{\text{opt}})^2)$. Transcription is bounded between zero and a maximum transcription rate which depends on the number of genes $t_{\text{max}} = k_{\text{max}} n_X$, assuming n_x is the copy number of genes of type X. Altogether, the transcriptional load per time step is $l_X = \min(t_X \cdot h(Q), t_{\text{max}})$, and the amount of mRNA m_x is the proportion of the load transcribed from the active genes: $F_{\text{transcr}}(X) = \frac{n_{\text{active}X}}{n_X} l_X$.

- Translation of mRNAs into the respective proteins is carried out by the Ribosome, in the presence of aminoacids, according to the function $\frac{m_x R A}{\epsilon + s_M m_x + s_R R + s_A A}$, where the ribosome R is formed by the association of ribosomal RNAs R_r and proteins R_p, assumed in steady state: $R = k_{R_p \bullet R_r} \min(R_p, R_r)$

Each generation, the cell dynamics described above is run for all individuals of the population. Cells grow until they reach their target volume, which is calculated from the genome size $\mathcal{G}$ as $V_t = \kappa \mathcal{G}^\lambda$ (with $\kappa = 3.5$ and $\lambda = 0.9$).

When all cells have reached their target volume (or after a maximum integration time elapsed), they are ranked on the basis of their inter-division time (the number of integration steps from birth to reaching target volume), which is used to calculate the probability of replication. A fraction of the cells are culled from the population, with cells that have lost all their active genes of any type being preferentially removed. If the fraction of cells with zero fitness exceeds 10%, these are the only cells removed, otherwise cells are sampled randomly until 10% of the population is reached. The fittest cells (up to stochasticity) are duplicated and the offspring mutated.

Random mutations occur with a per-gene probability μ_R and are independent of a cell's life history. In addition, mutations may be due to high transcriptional load, which interferes with replication. These mutations occur with larger frequency when genes are

transcribed more. To this end, we calculate the life-time average, per-timestep transcriptional load $\langle l_x \rangle$ (separately for every gene). The closer this is to the maximum transcription rate, the higher the chances of mutations will be. To be precise, transcription induced mutation rates $\mu_T = \frac{\mu_{max}}{1 + (h_\mu \frac{\langle l_x \rangle}{t_{max}})^\rho}$, where $\rho = 10$ ensures that mutations occur rarely with lower transcriptional load.

Finally, the number of proteins, RNAs and aminoacids is divided by two, and the populations enters the next generation. Notice that external resources S, are not halved.

The numerical value of parameters that remain the same throughout all simulations can be found in Appendix 5.5.1. Relevant parameter values are specified where needed.

A number of assumptions are introduced in the model:

- Target volume is directly determined by genome size, according to the relation $V_t = \kappa \mathcal{G}^\lambda$, with $\lambda = 0.9$ (see van Hoek and Hogeweg (2009) and references therein). This assumption introduces fitness differences by penalising slower growth or larger genomes.

- Cells live in environments that are independent from each other. Thus, individuals are only competing for growth rates and ecological strategies are not possible (e.g. deplete resources without actually consume them).

- Dosage effects do not occur automatically upon gene duplication or deletions, because transcription is assumed to be under control of polymerase (if a cell double its genes, half of the polymerases will be available on each gene).

- Much of intracellular dynamics are modelled in terms of phenomenological functions.

- Aminoacids behave both as signalling molecules for gene expression (Jewell *et al.*, 2013) and as building blocks for proteins.

- Mutations do not diversify genes (only inactivate them).

5.3 Results

5.3.1 Evolutionary dynamics with background mutations only

We first study the evolutionary dynamics of the system when mutations occur independently of transcription. The aim here is to set a baseline for the case with transcription-induced mutations.

The proportion of duplications, deletions and inactivations is set to be equal, $p_\mu^{\text{BG}} = (1/3, 1/3, 1/3)$, and we compare results at lower, intermediate and higher mutation rates, i.e. for $\mu_{\text{BG}} = 0.01$, 0.084 and 0.23.

Fig. 5.2a shows such comparison when cells have evolved under a small constant influx of resources $S_{in} = 1$, and Fig. 5.2b when cells evolved in a richer environment, i.e. $S_{in} = 10$. Notice that because the decay rate of resources is $d_S = 0.1$, the steady state values of resources in the absence of cellular metabolism would be, respectively, $\bar{S}_{\text{in}=1} = 10$ and $\bar{S}_{\text{in}=10} = 100$. Lower values for S are due to faster metabolism. Similarly, lower values for aminoacids A indicate faster translation.

The differences in the effect mutation rates make when evolving in a rich or poor environment are striking. When resources are scarce there is little difference between high or low mutation rates with respect to genome size, resource exploitation and aminoacid usage. Inter-division time increases with higher mutation rates, i.e. fitness is lower, presumably due to the slightly smaller number of coding genes.

In contrast, in the rich environment genome size is much larger when mutations are more frequent because of the accumulation of inactive genes. Consequently, inter-division time increases. In addition, a decrease in coding genes for ribosomal RNA and enzymes can be observed with increasing mutation rates.

Moreover, cells evolved in richer environments have 10 to 20 times larger genomes than those evolved in poor environments under the same mutation rates. Most of the difference is caused by inactive genes, which constitute between 50% and 90% of the total copy number of each gene type, while the copy number of coding genes is only 2 to 3 times larger. The genes for ribosomal RNA are in the largest copy number in all cases (rRNA is present in large copy number in eukaryotes (Long and Dawid, 1980) and to a lesser extent in bacteria (Stoddard *et al.*, 2015)). This is because the abundance of ribosomal RNAs is proportional to transcriptional load, whereas protein counts can be increased by translating the same mRNA multiple times.

Despite a larger genome, cells replicate faster when resources are abundant, as long as mutation rates are not too large.

Fig. 5.3a shows the evolutionary trajectory along the line of descent for the first 16000 generations, in a simulation with richer environment ($S_{in} = 10$) at intermediate mutation rates ($\mu_{\text{BG}} = 0.084$).

We start from small, random, initial genomes, with metabolic and regulatory parameters set so that homeostasis is already achieved (but can still evolve). We observe that inter-division time decreases quickly, and reaches a stable minimum within 5000 generations. This is a consequence of the fitness criterion, which rewards faster cell growth (provided homeostasis in housekeeping gene expression). The decrease in inter-division
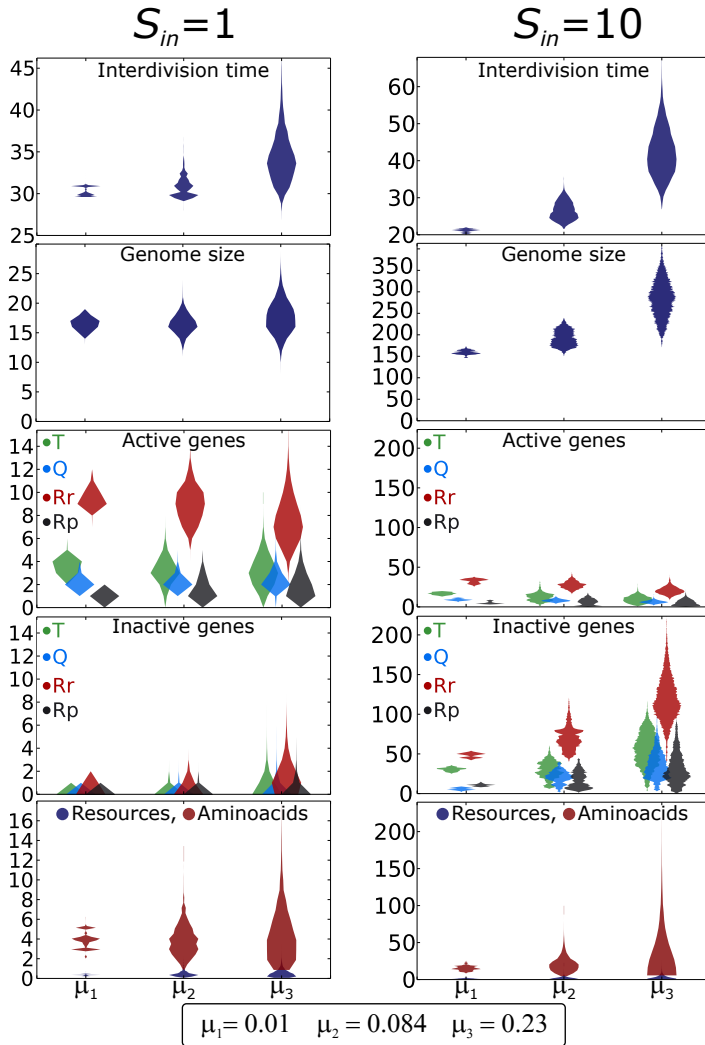
Figure 5.2. Simulations with Background mutations only, evolved in environments with **a** $S_{in} = 1$ and **b** $S_{in} = 10$. Colour coding for active and inactive genes: T (green) enzymes, Q (blue) housekeeping proteins, Rr (red) ribosomal RNA, Rp (black) ribosomal proteins. For all datasets, data are collected from the ancestor's line of descent for the last $250 \cdot 10^3$ generations of a simulation, after evolutionary steady state is reached (to ensure this, in all cases we collected data after at least $3 \cdot 10^6$ generations). For each subplot the x-axis is the background mutation rate.

time seems to be due to the increase in ribosomal RNA gene copy number, accompanied by an increase in its expression. This allows faster growth because it allows a larger exploitation of resources (resources decrease to almost zero in a comparable time-scale).

Interestingly, while duplications are positively selected in this phase, the number of inactivating mutations that go to fixation (those in the ancestral linage) is much smaller, even though they occur with the same probability. This is likely due to a stronger competition for shorter inter-division time in the first phase of the evolutionary experiment, where inactivations are doubly deleterious because they increase the target volume without contributing to faster growth.

In contrast, inactive genes do accumulate in the long evolutionary term, when active genes reach a stable distribution, as shown in Fig. 5.3b. An expansion of the inactive genome leads to a larger target volume without increasing growth and should not be favoured. It is, however, accompanied by a larger usage of aminoacids, implying that cells partially compensate a larger mutational load by becoming more efficient.

Over long evolutionary time scales, inactive genes reach an overall stable steady state, even though large scale fluctuations in genome size occur frequently (Fig. 5.3b).
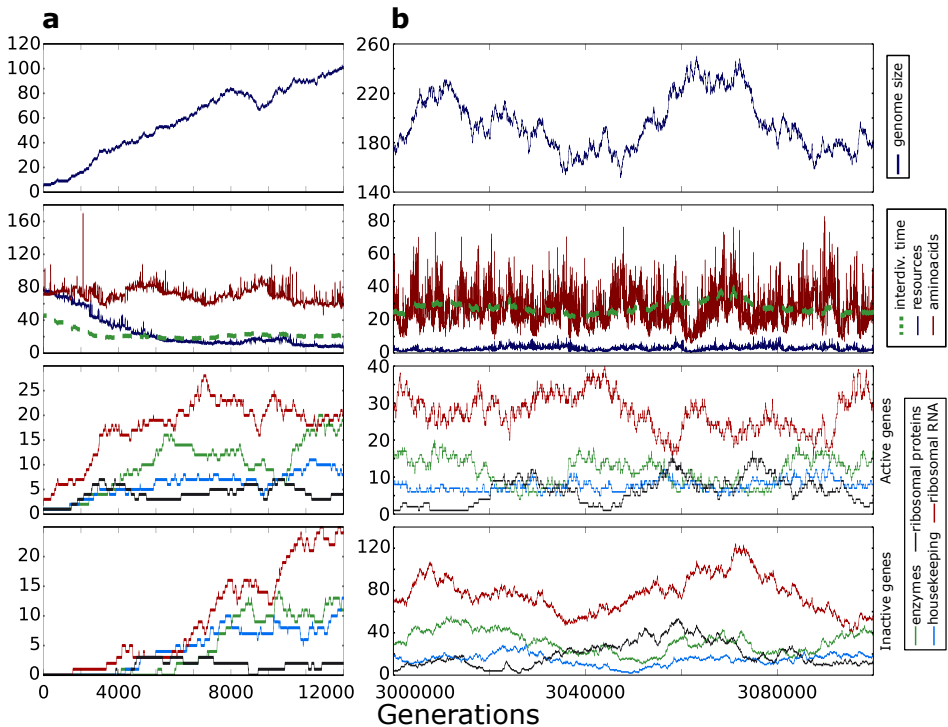

Figure 5.3. Early and late evolution along the line of descent. Background mutations $\mu_{BG} = 0.084$, resource influx $S_{in} = 10$. Notice that both x and y axis differ between a and b.

143

In summary, the initial evolution of the system is driven by selection for reducing inter-division time, achieved by increasing the number of active genes in the genome; later, cells become more efficient at using aminoacids, while inactive genes accumulate.

The initial genome expansion is limited by the maximum amount of inactivating mutations that can be weeded out by selection. Once the mutational limit for the genome is reached, duplications of active genes may still increase growth rate, thus being selected in the very short run. This would be the case especially in the richer environment, where larger metabolism leads to a larger consumption of resources. As genome size becomes large, the effect of each inactivating mutation becomes less deleterious, even though mutations occur more often in absolute terms. The cell responds by evolving regulation parameters to increase trancription, which can be again satisfied by a larger active genome, and the cycle repeats. This runaway process leads to the evolutionary deterioration of the genome, especially at higher mutation rates.

5.3.2 Evolutionary dynamics with transcription-induced mutations

In the previous paragraph we have observed that genome expansion in a rich environment (to accommodate a larger transcriptional load) leads to the accumulation of inactive genes. In this paragraph we show that including mutations induced by transcription in the same proportion as background mutations leads to a worse evolutionary steady state, with lower fitness and a larger inactive genome, even when mutation rates are lower.

In the model discussed so far mutations occur as random mistakes during genome replication. It is known, however, that a larger transcriptional load can lead to conflicts between the transcription machinery and the replication apparatus which increase the chances of mutations (Kim and Jinks-Robertson, 2012). We include this observation in the system by assuming that the per-gene probability of mutation depends on the total amount of transcription of that gene type over the life span of a cell. For example, if ribosomal RNA genes are few in copy number but heavily transcribed, the chances of mutation on ribosomal RNA genes will be higher. Inactive genes can have a new role in this system, i.e. they buffer against higher mutation rates as they reduce the per-gene transcriptional load. This models the sequestering of RNA-polymerases by inactive genes.

As a first step, we assume that transcription-induced mutations amplify the chances of background mutations, with the relative proportion of duplications, deletions and inactivation being equal: $p_\mu^{\text{TR}} = p_\mu^{\text{BG}} = (1/3, 1/3, 1/3)$. In Fig. 5.4 we compare the results with only background mutations and the system that incorporates transcription-induced mutations. For the sake of the comparison between the two mutational regimes, we set the sum of background and transcriptional mutation rates in the latter to being, at most, as high as the background mutation rate in the former.

Evolution in the presence of transcriptional mutations leads to a longer inter-division time, due to a larger non-coding genome. The number of coding genes is instead slightly smaller than in the case of higher rates of background mutations.

Even though the overall mutation rate is lower when transcription-induced mutations are introduced, cells perform worse than in the case of a life history-independent mutations (i.e. when only background mutations are present, even at high rates) because mutations affect those genes that are most transcribed (i.e. most needed), thereby having a larger effect. Indeed, transcriptional load is kept maximal throughout a cell's lifespan both at early and at late evolution when only background mutations occur (Appendix 5.5.2a). In contrast, cells evolve to a lower transcriptional load for all genes when exposed to transcription-induced mutations (Appendix 5.5.2b). As a side effect, the growth curve is roughly exponential in the former case, and sub-exponential in the latter, even though resources and aminoacid consumption follow approximately the same curves (notice, however, that the shape of the transcriptional load over a cellular lifespan may change considerably over long evolutionary times).

As mentioned above, a larger mutational load is due to the inability of a genome to purge inactivated genes, while still selecting for a larger active genome because of increasing transcriptional demand ($t_X = k_0 + A * k_1$, see Methods). Including transcription-induced mutations strengthens this process because mutations target the most actively transcribed genes, and a larger fraction of active genes is beneficial in the short run, even if selection cannot maintain it in the face of mutations. This can be seen in Fig. 5.5 by the fact that ancestors have a larger fraction of active genes than the rest of the population in which they lived (in 99% of the cases ancestral genomes are larger than the median of the population, and in 46% of the cases they are above the 95th percentile, in the last $1.5 \cdot 10^6$ generations, sampled every 5000).

Selection pressure for a larger active genome leads through mutations to an increase in inactive genes. This in turn causes lower RNA polymerase occupancy (lower transcriptional load) and therewith to selection pressure for higher transcription rates. As regulation parameters evolve to increase transcriptional load, the chances of transcription-induced mutations rise. Those mutations that increase the active genomes are selected, and the cycle repeats.

Thus, this evolutionary feedback results in a large non-coding genome.

To further clarify this, we compare evolutionary steady state genomes when regulation does not evolve (Fig. 5.6). We extract regulation parameters from the simulation that produces μ_{BG+TR} in Fig. 5.4, at an early evolutionary time and at a later one ($5 \cdot 10^3$ and $2 \cdot 10^6$). Inactive genes do not accumulate in the long run when "early" parameters are used, whereas evolution increases mutational load with "late" parameters (Fig. 5.6). Inter-division time mirrors this (notice that the number of coding genes is similar in the two cases).

In conclusion, if the sole effect of transcription-induced mutations is to amplify the effect
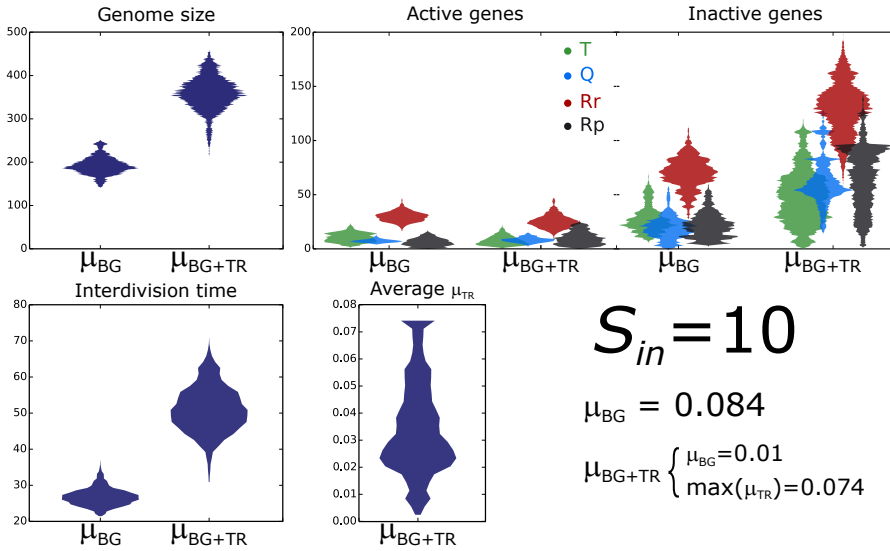
Figure 5.4. Transcription-induced mutations worsen mutational load despite being, at most, as frequent: a comparison with background mutations only (identical to Fig. 5.2b, μ_2). For $\mu_{\mathrm{BG+TR}}$ background mutations are the same as μ_1 in Fig. 5.2b, and have the same proportion of duplications, deletions and inactivations for transcription-induced mutations. Colour coding in the figure as above.
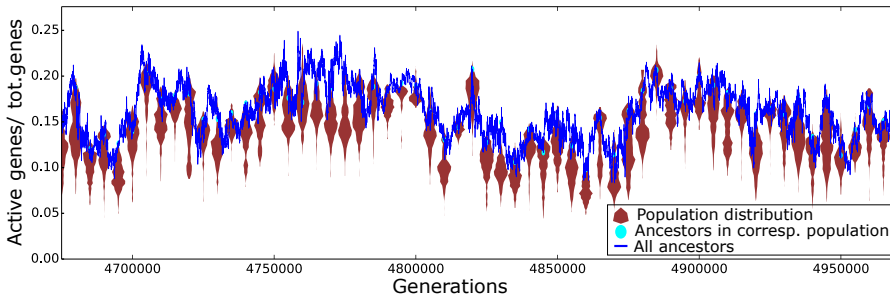
Figure 5.5. Evolving with transcription-induced mutations, cells in the ancestral lineage have a larger fraction of active genes than the population in which they lived. $\max(\mu_{\mathrm{TR}}) = 0.074$; $\mu_{\mathrm{BG}} = 0.01$; $p_\mu^{\mathrm{TR}} = p_\mu^{\mathrm{BG}} = (1/3, 1/3, 1/3)$.

of background mutations, evolution along the line of descent leads to lower fitness than when only background mutations are present even when the mutation rate is, on average, lower. As in the case without transcription-induced mutations, selection cannot purge inactive genes effectively when genome size is large. This is more so in this case, because inactive genes buffer against mutations by lowering transcriptional load, even though they do not contribute to the transcription pool.

Although the intuitive solution to the problem of mutational deterioration would be to lower mutation rates, in the next paragraph we show how a coding-rich genome evolves when mutation rates are higher and mutational effects biased, as is the case for yeast's rRNA genes.
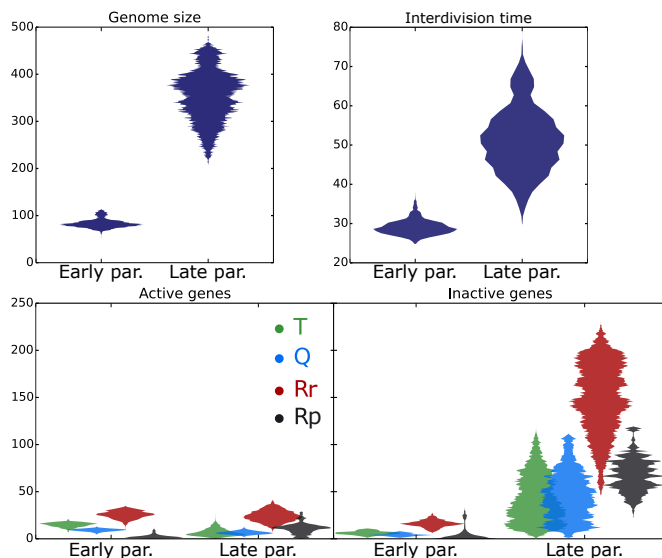
**Figure 5.6.** Mutational load increases with larger transcriptional demands. Two ancestral cells, at an early and at a later stage of evolution (from the same dataset as Fig. 5.4, at generation number $5 \cdot 10^3$ and $2 \cdot 10^6$), are used to seed a homogeneous population which cannot mutate its regulation parameters (other mutational dynamics are as before). After long-term evolution, the last $50 \cdot 10^3$ generations are collected for both to generate the figure. Values of "early" transcription parameters k_0, k_1: Enzymes (0.099, 0.014), Housekeeping proteins (0.11, 0.0078), Ribosomal RNA (0.12, 0.031), Ribosomal proteins (0.11, 0.025). Values of "late" parameters: Enzyme (0.54, 0.39), Housekeeping proteins (-0.24, 0.33), Ribosomal RNA (0.13, 0.69), Ribosomal proteins (0.14, 0.21).

5.3.3 Larger rates of duplications and deletions are beneficial despite increasing mutation rates

Fragile sites, regions of a genome most affected by mutations, are more frequently located outside coding parts of genes (Lemaitre *et al.*, 2009). An example of fragile sites are those portion of genomes where Replication Fork Barriers (RFB) are located: RFBs stall the replication machinery, which may in turn slip or detach entirely from the template it is copying and re-attach somewhere else, often causing mutations (Rothstein *et al.*, 2000). RFB-mediated stalling in yeast rDNA can cause recombination, resulting in duplications and deletions if transcriptional load is large enough (Takeuchi *et al.*, 2003).

As above, we do not model a specific mechanism but limit ourselves to model the effects of mutations, i.e. their outcomes, when every gene is flanked by a fragile site (as is the case for ribosomal RNA genes in yeast), and assume that mutational effects are limited to a single gene downstream. We combine low rates of background mutations (as before $\mu_{BG} = 0.01$, $p_\mu^{BG} = (1/3, 1/3, 1/3)$), with a much larger maximal rate of transcription-induced mutations ($\max(\mu_{TR}) = 0.074$) with a bias to duplications and deletions.

In Fig. 5.7 we show that a larger bias towards transcription-induced duplications and deletions contributes to a more compact genome and a shorter inter-division time. Genomes evolved with only transcription-induced duplications and deletions ($\mu_{TR\,3}$ Fig. 5.7) have a larger active genomes and hardly any inactive gene. These cells maintain a large transcriptional load and therewith a high frequency of duplications and deletions, whereas cells evolved with no bias in their mutations decrease their transcriptional load which results in transcription-induced mutations occurring less frequently. Cells evolved with a larger transcriptional mutation rate biased to duplications and deletions reach a fitter evolutionary steady state than those evolved only with rare background mutations (compare $\mu_{TR\,3}$ and μ_{BG} Fig. 5.7). Moreover, when the relative proportions of transcription-induced mutations evolve, we observe a strong selection pressure towards minimising inactivations and/or increasing duplications and deletions, keeping the proportion of the latter two at the same level (Appendix 5.5.3). A consequence of a steady and large transcriptional load is that mutations happen at a similar (large) rate every generation. This suggests that we could make a "conceptual" simplification of the model by introducing this bias directly in background mutations (i.e. disregarding that biased mutations are induced by transcription). The simplified model leads to similar results as above (Appendix 5.5.3).

As a side note, if transcription-induced duplications occur more frequently than deletions, genomes accumulate inactive genes without bound, even in the absence of transcription-induced inactivations (See Appendix 5.5.4).

To conclude, even though conflicts between transcription and replication might be unavoidable and frequently result in mutations, we find that biasing mutations to duplications and deletions may turn this "scorch" into an asset, leading to the evolutionary stability of fitness and smaller genomes containing mostly active genes. In the next paragraph

we show that transcription-induced mutations can evolve to counter-balance pre-existing biases in background mutations.
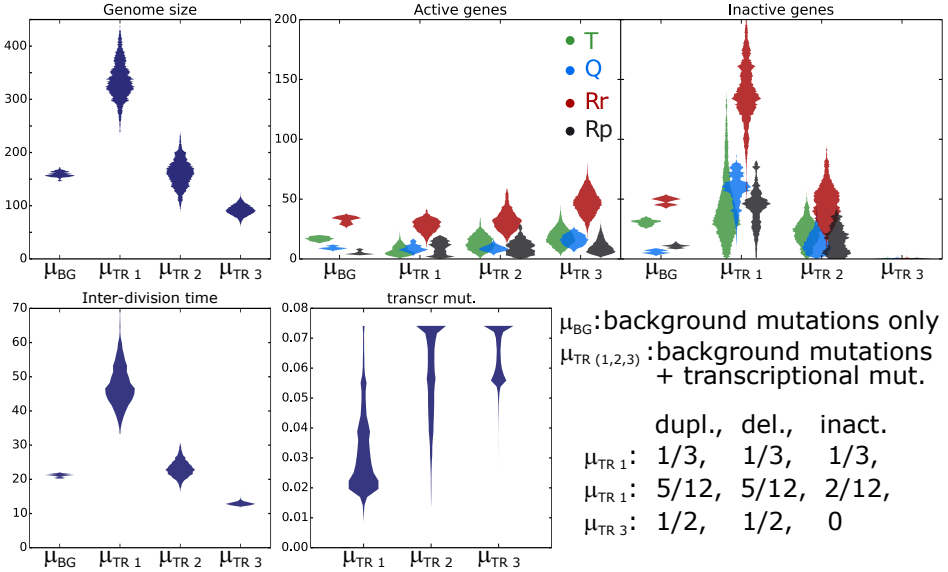
Figure 5.7. The beneficial effect of larger rates of transcription-induced mutations, when biased to duplications and deletions. Data collected along the line of descent for $250 \cdot 10^3$ generations after evolutionary steady state is reached. Background mutations are the same in all four cases: $\mu_{\text{BG}} = 0.01$, $p_\mu^{\text{BG}} = (1/3, 1/3, 1/3)$. The maximum mutation rate caused by transcription is $\max(\mu_{\text{TR}}) = 0.074$.

5.3.4 Evolution biases transcription-induced mutations in response to skewed background mutations

As mentioned in the introduction, yeast controls its rDNA mutational dynamics by means of the TOR pathway (a nutrient signalling pathway) (Jack *et al.*, 2015). In particular, rDNA gene are more likely to duplicate under caloric excess if rDNA copy number is small, even though no short-term selective advantage is gained by duplicating genes ((Ide *et al.*, 2010)). The regulation of this duplication bias is not entirely understood, but it likely involves double strand breaks at replication fork barriers (every ribosomal gene copy has one) due to transcription/replication conflicts and non-homologous recombination.

We study genome stability and mutational load assuming that background mutations result in deletions and inactivations more frequently than duplications (as in yeast), with and without transcription-induced mutations. In the latter case we let the proportion of duplications, deletions and inactivations caused by transcription evolve.

In Fig. 5.8 we show that background mutations skewed towards deletions and inactivations (μ_{DI}, Fig 5.8) lead to compact genomes with almost no inactive genes. Despite the absence of inactive genes, cells are less fit than when evolved with an equal proportion of duplications, deletions and inactivations (compare inter-division time with μ_2, Fig. 5.2b). In contrast, evolution reaches a larger final fitness when evolvable transcription-induced mutations are incorporated (μ_{TRev}, Fig 5.8), even though inactive genes are purged less successfully. The rate of transcription-induced duplications evolves to a slightly larger value than the rate of transcriptional deletions (Fig.5.9a, the median of the differences between duplications and deletion is 0.046), to balance the continuous loss of genes due to background mutations, while the rate of transcription-induced inactivations approaches zero. We also observe that there is no direct benefit in the high copy numbers of rRNA genes: Fig. 5.9b shows that a considerable portion of rRNA genes must be removed from an evolved genome in order to increase inter-division time (notice that more rRNA genes can be removed without altering fitness if the initial genome has a higher copy number). This is in qualitative agreement with yeast rDNA genome dynamics, where removing copies of ribosomal genes does not affect short-term fitness (growth rate), but removing too many leads to a rapid deterioration of genome stability and, consequently, fitness (Ide *et al.*, 2010).

Altogether, evolution balances the biases in life-history independent mutations by maintaining a high transcription-induced mutation rate and a skewed mutational outcome. The process leads to a larger fitness despite a two-fold increase in overall mutation rates.

We conclude that the observed features of the yeast's rDNA organisation and mutational processes, i.e. transcription-replication conflicts and biases towards recombination of rRNA repeats, are beneficial and contribute to its long term physiological and evolutionary stability.

5.4   Discussion

In this chapter we have studied the evolutionary dynamics of genomes in single-cell organisms by constructing a minimal (computational) model in which metabolism, regulation and mutational dynamics are integrated. Several theoretical models have been developed to study genome dynamics (Crombach and Hogeweg, 2008, Cuypers and Hogeweg, 2012), some at a greater resolution than what we do here (Knibbe *et al.*, 2007). However, the effects of large transcriptional load on genome dynamics have received no attention (to the best of our knowledge), despite being extensively studied experimentally. Here we made a first step in this direction.

We find that cells quickly adapts to fast growth (inter-division time being the fitness criterion), but their long term evolution depends strongly on the mutational dynamics. Genomes accumulate inactive genes in the long evolutionary term when background mutational outcomes are equally divided among duplications, deletions and inactivations,
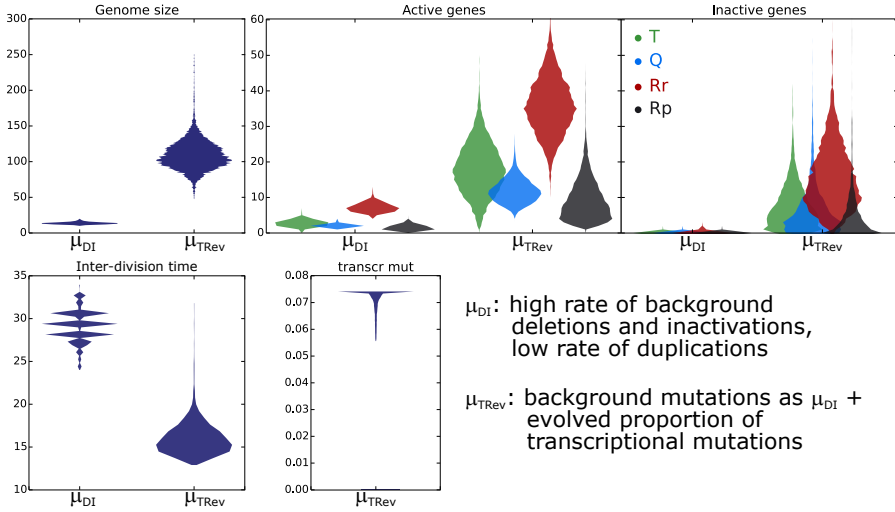
Figure 5.8. A large transcriptional load is maintained and is beneficial when background mutations are biased towards deletions and inactivations. $\mu_{DI} = 0.084$, $p_\mu^{DI} = (1/25, 12/25, 12/25)$; Background mutations in μ_{TRev} are the same as μ_{DI}, transcription-induced mutations can evolve their relative proportions (see Fig. 5.9a); the maximum mutation rate caused by transcription is $\max(\mu_{TR}) = 0.074$.
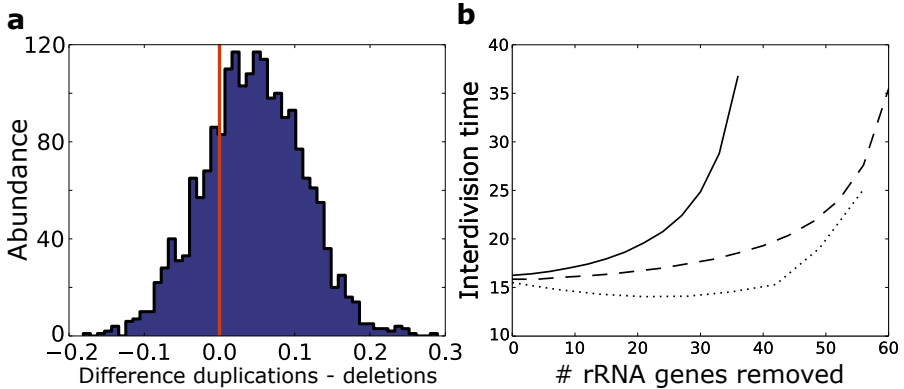
Figure 5.9. a Histogram of the differences between duplications and deletions for the system μ_{TRev} in Fig. 5.8. Positive values indicate a bias towards duplication, negative ones towards deletions (inactivations approach zero in all cases). Red line indicates zero. Data collected along the line of descent every 1000 generations. **b** Growth rate is minimally affected when several ribosomal RNA genes are removed (cf. French *et al.* (2003)). Three cells are taken at a later evolutionary stage from the same dataset as Fig. 5.8 μ_{TRev}, along the line of descent. Ribosomal RNA genes are removed maintaining the proportion between active and inactive ones and inter-division time is recorded. The original copy number of rRNA genes in the genomes used for these experiments is (total, active, inactive): 39, 27, 12 (full line); 66, 49, 17 (dashed line); 71, 60, 11 (dotted line).

because duplications of active genes are selected in the short run, but inactivations cannot be effectively purged. Mutational dynamics caused by large transcriptional load worsens this problem when the proportion of mutational outcomes has no biases, even though the system evolves to low mutation rates. The accumulation of inactive genes caused by failure of selection to purge them is is reminiscent of an Error-threshold problem (Eigen, 1971). It is most surprising, then, that higher mutation rates can overcome it.

Genome integrity (as well as fitness) is restored when transcription-induced duplications and deletions occur more frequently than inactivations, even when the overall mutation rate is increased. Inactive genes do not accumulate despite transcription being maintained at large rates, and thus transcription-induced mutations occur more often. High rates of duplications and deletions increase the variability in the population (compare active and inactive genome distributions for μ_{BG} and $\mu_{TR\ 1\text{-}2\text{-}3}$ in Fig. 5.7), which increases the selection pressure for a larger active genome and a smaller inactive one. This overcomes the long term mutational flow from active to inactive genes, and suppresses the evolutionary feedback that leads to the deterioration of the genome in the long term. Note, however, that the short term effect of duplications and deletions is not the direct cause of increased fitness, as these mutations are almost neutral. Their effect lies in halting the flow to inactivations by accelerating gene dynamics.

Altogether, increased rates of neutrality (mutational robustness) and a faster mutational dynamics (which leads to larger adaptability) prevent the evolutionary deterioration of the genome by slightly deleterious mutations.

A similar situation is observed in yeast, where no short-term benefit is apparent despite the frequent occurrence of duplications and deletions in the rRNA gene cluster (French *et al.*, 2003). In this respect, our work suggests that the observed mutational dynamics in yeast are beneficial for long-term genome integrity. Moreover, the large neutral variability in yeast's rRNA copy number is a long term evolutionary side-effect of the observed mutational dynamics.

In summary, the short-term benefit of duplications of active genes lead to long-term deleterious consequences, i.e. the inability of purging inactive genes, even at low mutation rates. Evolution has solved this problem in yeast by increasing mutation rates and biasing the mutational outcome towards duplications and deletions.

The evolutionary dynamics described in this study applies at evolutionary steady state only for genes existing in large numbers of identical copies, such as the rRNA gene. However, it is known that genome expansion can lead to long-term fitness benefits because the excess genetic material provides more degrees of freedom for evolution to adapt (Cuypers and Hogeweg, 2012, de Boer and Hogeweg, 2010). The evolutionary dynamics discussed here can form the substrate on which neo-functionalisation works.

Few mutational operators are defined in our model because we coarse-grained genomes to the gene level. As a consequence, genomes have little freedom in the coding structure they can evolve. While this study shows the effects of the mutational dynamics observed in yeast, future work will address the evolution of the mechanism that increases transcription-induced gene-duplications and deletions, by incorporating finer details of genomes, such as evolvable fragile sites.

Finally, aminoacids are used as signalling molecules in our minimal model of regulation, in accordance with their prominent role in the signalling cascade in the TOR pathway (Jewell *et al.*, 2013, Kim, 2009). In future developments of the model we will explore the evolutionary consequences of richer regulation-mutation dynamics, such as those observed in yeast.

Nevertheless, this study makes the case that mutations induced by large rates of duplications and deletions induced by transcriptional load can prevent long term genome deterioration.

5.5 Appendix

5.5.1 Numerical values of parameters used in the simulations

Parameter	Value	Description
d_s	0.1	Decay resources
d_a	0.005	Decay amino acids
d_T	0.005	Decay Enzymes
d_Q	0.005	Decay Housekeeping proteins
d_{R_p}	0.005	Decay Ribosomal proteins
d_{R_r}	0.005	Decay Ribosomal RNA
d_{m_T}	0.05	Decay mRNA of enzymes
d_{m_Q}	0.05	Decay mRNA Housekeeping pr.
$d_{m_{R_p}}$	0.05	Decay mRNA Ribosomal pr.
k_t	2.5	Metabolism of resources
α_t	0.5	Metabolism of resources
β_t	1.0	Metabolism of resources
k_h	15	Homoeostatic parameter
k_h	0.3	Optimal concentration of Housekeeping pr.
$k_{\max}$	0.2	Max transcription rate time step
ϵ	10^{-4}	Translation parameter
s_m	10	Translation parameter
s_R	1	Translation parameter
s_A	3	Translation parameter
$k_{R_p \bullet R_r}$	0.9	Ribosomal RNA and pr. association
h_μ	0.9	Transcription-induced mut. parameter
ρ	10	Transcription-induced mut. parameter
λ	0.9	Genome to volume scaling
κ	3.5	Genome to volume factor

5.5.2 Details of the intracellular dynamics with background and transcription-induced mutations

In Fig. 5.10, we compare the life cycles of cells at different time points with background mutations only (a), or with un-biased transcription-induced mutations (b).
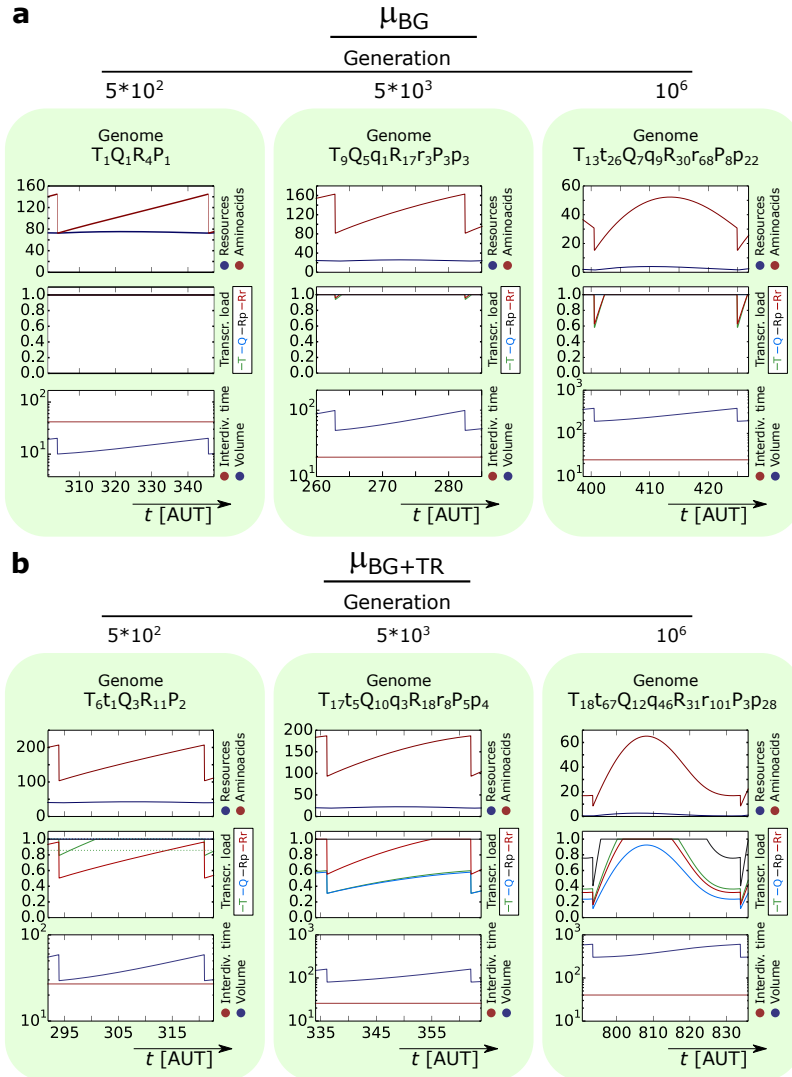
Figure 5.10. A comparison of intracellular dynamics during one life cycle at different time points. **a** background mutations only (μ_{BG} in Fig.5.4); **b** both background mutations and transcriptional mutations (μ_{BG+TR} in Fig.5.4). The internal dynamics of ancestral cells at chosen time points (from dataset used for Fig. 5.2) are run in the absence of mutations and parameters are plotted for one generation. Abscissa in all plots: integration time (arbitrary units). Genomes are represented with upper and lower case letters indicating the respectively the active and inactive form of genes (T: enzymes, Q: housekeeping, R: ribosomal RNA, P: ribosomal proteins), lower scripts indicate copy number. Transcriptional load is represented as the fraction of mRNA production relative to the maximum gene occupancy (see Methods). Notice log y-axis for volume plots.

155

5.5.3 Larger rates of duplications and deletions in background mutations

Fig. 5.11 shows that a bias towards duplications and deletions in background mutations restores fitness despite an increased overall mutation rates.
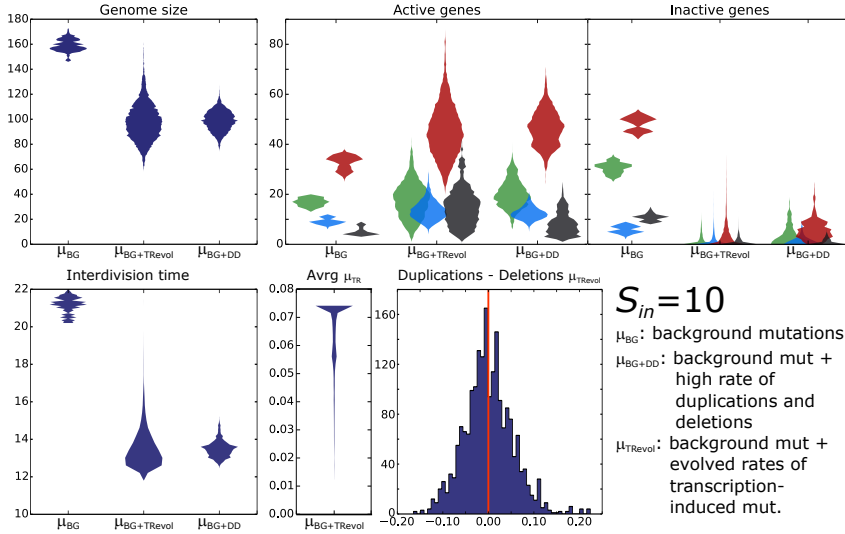
Figure 5.11. The beneficial effect of larger rates of background transcription-induced mutations, when biased to duplications and deletions. $\mu_{\mathrm{BG}} = 0.01, p_\mu^{\mathrm{BG}} = (1/3, 1/3, 1/3)$; $\mu_{\mathrm{BG+DD}} = 0.084, p_\mu^{\mathrm{DD}} = (12/25, 12/25, 1/25)$. Data collected along the line of descent for $250 \cdot 10^3$ generations after evolutionary steady state is reached.

5.5.4 Deleterious consequences of duplications being more frequent than deletions

Fig. 5.12 shows that a bias towards duplications leads to inordinate integration of non-coding genes in the line of descent of a population.
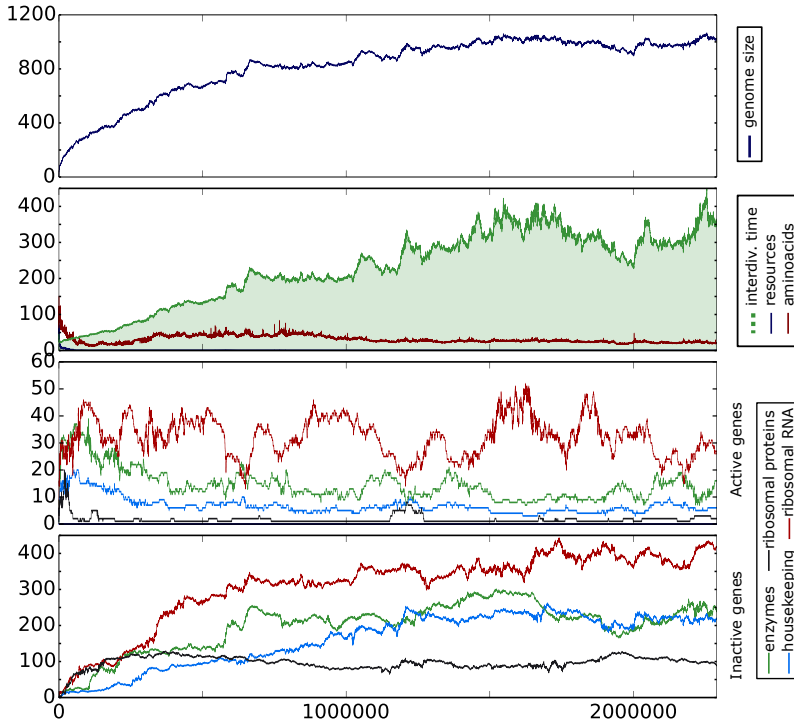
Figure 5.12. A larger proportion of transcription-induced duplications leads to unbounded genome expansion, due to the accumulation of inactive genes. $\mu_{\text{BG}} = 0.01$, $p_\mu^{\text{BG}} = (1/3, 1/3, 1/3)$; $\max(\mu_{\text{TR}}) = 0.074$, $p_\mu^{\text{TR}} = (1/2, 1/3, 1/6)$.

6

Discussion

6.1 Summarising Discussion

In Chapter 2 we found that strong parasites generate the spatial conditions in which replicators can become more cooperative, in a prebiotic evolutionary model. We extended these results in Chapter 3 by studying the evolutionary dynamics of cooperation when individuals must produce a costly public good to survive. We observed the evolution of two lineages at high cost, a cooperative and a parasitic one, each enhancing the evolution of the other. In Chapter 4 we analysed the interplay of spatial pattern formation and coding structure in an RNA-like model system. We found that a master sequence arises at high mutation rates that encodes the information for an array of novel functions beneficial to the master sequence and associated to genotypes that cannot be replicated. In Chapter 5 we studied the interplay of mutations induced by transcription, background mutations and genome dynamics in a cell model. Large mutational load is accumulated in the genome when gene duplications, deletions and inactivations occur at similar rates, because mutations that are beneficial in the short run cannot be maintained by selection in the long evolutionary term, and slightly deleterious mutations accumulate. In contrast, an excess of gene duplications and deletions reduces the mutational load of the cell, improves long-term genome stability and increases fitness. Yeast's rRNA genes are known to frequently undergo duplications and deletions due to conflicts between transcription and replication complexes without leading to differences in growth rates. Our results suggest that the

mutational dynamics observed in yeast is the result of a long-term evolutionary pressure to maintain genome integrity.

In the following paragraphs we expand the discussion of the results presented in this thesis, and we outline several directions for future research.

In Chapter 2 we observed two evolutionary strategies depending on travelling waves dynamics: selection for fecundity is observed with weaker parasites, selection for longevity with stronger ones. Both outcomes can evolve when either parasite strength or replicators' association rate is set to evolve. When both replicators and parasites evolve, the outcome depends on the time-span needed to complete replication, i.e. how costly replication is.

In replicator-parasite systems encapsulated in vesicles (which constitute a higher level of organisation) selection always favoured longevity (Takeuchi and Hogeweg, 2009). However, two alternative evolutionary strategies were also observed, depending on mutation rates: lower mutation rates led to the evolution of weaker parasites - analogously to the chaotic waves regimes in Chapter 2 and (Takeuchi and Hogeweg, 2009); higher mutation rates, instead, destabilised the coexistence of replicator and parasites, to which vesicle-level selection responded by slowing down their internal dynamics - a solution that has no analogue in the self-organised system.

In the self-organised system, evolution chooses either of two strategies by the feed-back process with the spatial organisation studied in Chapter 2. However, due to the inherent stability of travelling waves, it is unlikely that changing mutation rates can change the direction of evolution, as happens in the vesicles (nevertheless the experiment remains to be done).

Altogether, even if local (microscopic) rules are the same for parasites in the two systems, they become functionally incomparable (qualitatively different) due to the higher levels of organisation.

The evolutionary mechanism that allows the increase in cooperation observed in Chapter 2 and Chapter 3 rests on the presence of uninterrupted empty space generated by strong parasites. However, a second species of replicators can directly invade that space if it is not affected by the same parasite (Takeuchi and Hogeweg, 2008). This suggests a slightly different mechanism for the evolution of cooperation, which does not depend on empty space: the direct competition for space between a replicator and the parasite of a different species leads both of them to increase replication rate, the former by becoming more cooperative, the latter by exploiting its replicator more strongly. Importantly, this mechanism works as a selection pressure to diversify the second replicator species (Takeuchi and Hogeweg, 2008), but would work as well if the two pairs of replicators

and parasites had no genetic relation at all. Therefore cooperation could evolve as a side effect of spatial competition with multiple species (von der Dunk *et al.*, 2016).

Moreover, in Chapter 2 and Chapter 3 we observed that the interplay between parasites and replicators leads to increasing the rate at which complexes form - which is an altruistic trait - when costs are larger. Parasites evolve quickly in the interacting RNAs model of Chapter 4 and (Takeuchi and Hogeweg, 2008), suggesting that the same interplay between evolution and self-organisation can explain the increase of binding affinity in the master sequence, which seems optimised for replication.

At high mutation rates, replicators "solve" the problem of selfish mutants (and parasites) by evolving the complex coding structure presented in Chapter 4. Most mutants of the evolved master sequence cannot be replicated (they are non-viable). This increases the selective advantage of the master sequence, whose competition is local, i.e. mostly with its kins. However, the most striking result is that these non-viable mutants are functional, in that they enhance the replication of the master sequence (we named these sequences "helpers") or inhibit the replication of competitors as well as the rare parasites (we named these "stallers"). Altogether, several different functions are united into a single template, as is true for chromosomes.

Two questions follow from the fact that the system functionalises mutations, i.e. it exploits the characteristic property of quasispecies. First, is there an Error Threshold? Secondly, what happens if sequence length is not fixed and consequently a variable volume of genotype space can be exploited by the master sequence?

The answer to the first question can be obtained straightforwardly by tracking the mutational neighbourhood of the master sequence as mutation rate is increased. One could expect that extinction ensues when sequences begin to de-localise because variability becomes so large that complex formation cannot occur. Fig. 6.1 shows that extinction does not occur via an Error Threshold, i.e. the quasispecies does not de-localise. As mutation rate increases, the process that stabilises the master sequence simply becomes more intense until the master sequence is replicated correctly so infrequently that the system goes extinct. We conclude that whether sequences delocalise or not at high mutation rates is a side effect of the evolved properties of a quasispecies.

The second question does not have a definite answer, instead. As mentioned in the Introduction, the protocol for RNA evolution to target secondary structure in a well-mixed system can be easily extended to incorporate mutations that change sequence length (structure must be suitably coarse-grained to assign the same fitness to identical motifs with different length). Fig. 6.2 shows results in the spatially extended system at high mutation rates. First, the master sequence becomes shorter to reduce mutational load (from 50 to about 40). Because the mutational neighbourhood of the shorter master sequence does not generate the same distribution of mutants, parasites evolve and organise with replicators in travelling waves. These parasites are optimised to exploit the master

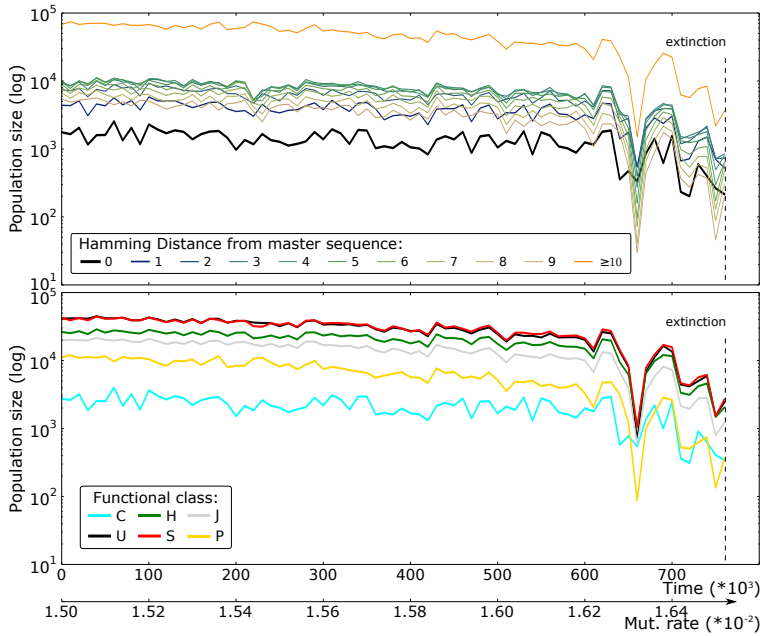**Figure 6.1.** Lack of delocalisation in the interacting RNAs system of Chapter 4, when mutation rate is increased to larger values. Per-base mutation rate is increased of 10^{-4} every $50 * 10^3$ time steps until global extinction occurs. Upper pane: Hamming Distance from master sequence; Lower pane: functional composition in the system (see Chapter 4 for the definition of functional classes.).

sequence (the 5' end of both strands consists mostly of G nucleotide), and are very short (20 nucleotides). In the long evolutionary run, however, a novel sequence appears which 1) cannot be characterised by a particular replicating strategy, 2) outcompetes the previously established lineages of replicators and parasites, 3) has a surprisingly large degree of neutrality (more than 60% and in some tested sequences up to 75%).

The fact that the original master sequence disappears is not unexpected once considered that it is fine-tuned to the effects of substitutions and not insertions and deletions. These results should be compared with sequences that undergo long-term evolution with these mutational operators. The significance of such large increase in neutrality, however, deserves further investigation: can the neutral evolution of robustness lead to this mutational neighbourhood, or is there a novel evolutionary pressure behind it?

In Chapter 5 we have encountered the well known problem of the long term accumulation of deleterious mutations (Kimura *et al.*, 1963, Muller, 1964). In our case, however, the unavoidable accumulation of mutational load is caused by long-term accumulation of mutations that are beneficial (for fitness) in the short run. The conflict between short and long term fitness is inherent to the dynamics of altruistic replication (Chapter 2, 3 and 4). Much less obvious is the occurrence of such conflict in the lack of multilevel selection, and even less so its resolution by larger mutation rates (Chapter 5). The occurrence of this effect is limited to repeated genes in our model, even though different lines of research hint to its generality (Rutten *et al.*, 2016).

6.2   Towards a theory of constructive evolution

At the beginning of this thesis we stated that the evolutionary and functional significance of novel traits is best understood from a multilevel perspective. Chapter 2 to 5 illustrate how new functions emerge in multilevel evolutionary models. A multilevel approach has been recently found to be necessary for the emergence of (any) function in a dramatically more open-ended model at the interface between Artificial Life and Artificial Chemistry (Hickinbotham and Hogeweg, 2016).

Although most functional traits are in some sense emergent (emergence should perhaps be the default expectation), here we mention three examples that are clear-cut enough to be amenable to theoretical investigation: 1) the repression of reproduction in soma-like cells of Volvox carteri achieved by germline cells by inducing an environmental stress response (Nedelcu, 2009), 2) the over-expression of metabolic and regulatory proteins for their optical properties in the crystallin (Piatigorsky, 2003) and 3) the regulation of ribosomal gene copy number under caloric excess by exploitation of mutation-inducing transcriptional dynamics (Jack *et al.*, 2015) (the topic of Chapter 5).

Clearly, challenges remain in understanding the constructive nature of Evolution. Nev-
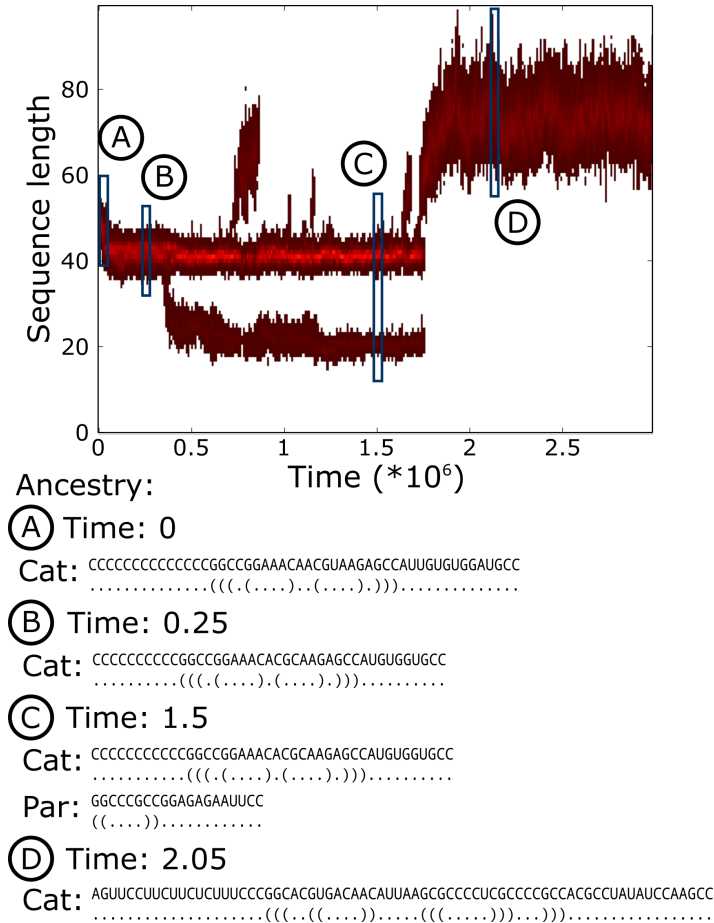
Figure 6.2. The evolution of sequence length in the interacting RNAs system (see Chapter 4) with nucleotide insertions and deletions. Sequence and structure of ancestral catalysts and parasites at selected time points are indicated with A,B,C and D. Brighter colours indicate larger proportions of the population. 10^4 individuals are sampled every 10^4 time steps to generate the plot.

ertheless, we are confident that the approach we have taken here will pave the way to a richer understanding of the evolutionary principles that govern how novel functions emerge.

Bibliography

Adami, C., Ofria, C., and Collier, T. C. 2000. Evolution of biological complexity. *Proceedings of the National Academy of Sciences*, 97(9): 4463–4468. (Cited on page 5.)

Ancel, L. W. and Fontana, W. 2000. Plasticity, evolvability, and modularity in rna. *Journal of Experimental Zoology*, 288(3): 242–283. (Cited on pages 16 and 121.)

Archetti, M. 2009. Survival of the steepest: hypersensitivity to mutations as an adaptation to soft selection. *Journal of evolutionary biology*, 22(4): 740–750. (Cited on page 121.)

Attwater, J., Wochner, A., and Holliger, P. 2013. In-ice evolution of rna polymerase ribozyme activity. *Nature chemistry*, 5(12): 1011–1018. (Cited on pages 4 and 35.)

Baake, E., Baake, M., and Wagner, H. 1997. Ising quantum chain is equivalent to a model of biological evolution. *Physical Review Letters*, 78(3): 559. (Cited on page 13.)

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J., and Frey, B. J. 2010. Deciphering the splicing code. *Nature*, 465(7294): 53–59. (Cited on page 16.)

Barve, A. and Wagner, A. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, 500(7461): 203–206. (Cited on page 18.)

Bermejo, R., Capra, T., Jossen, R., Colosio, A., Frattini, C., Carotenuto, W., Cocito, A., Doksani, Y., Klein, H., Gómez-González, B., *et al.* 2011. The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. *Cell*, 146(2): 233–246. (Cited on page 135.)

Bianconi, G. and Rahmede, C. 2011. Unified framework for quasispecies evolution and stochastic quantization. *Physical Review E*, 83(5): 056104. (Cited on page 13.)

Boerlijst, M. and Hogeweg, P. 1991a. Self-structuring and selection: Spiral waves as a substrate for prebiotic evolution. *Artificial life*, 2: 255–276. (Cited on page 51.)

Boerlijst, M. C. and Hogeweg, P. 1991b. Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physica D: Nonlinear Phenomena*, 48(1): 17–28. (Cited on page 23.)

Boerlijst, M. C. and Hogeweg, P. 1991c. Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physica D: Nonlinear Phenomena*, 48(1): 17–28. (Cited on pages 35, 51, 71, 96, and 123.)

Boerlijst, M. C. and Van Ballegooijen, W. M. 2010. Spatial pattern switching enables cyclic evolution in spatial epidemics. *PLoS Comput Biol*, 6(12): e1001030. (Cited on pages 48 and 51.)

Boerlijst, M. C., Bonhoeffer, S., and Nowak, M. A. 1996. Viral quasi-species and recombination. *Proceedings of the Royal Society of London B: Biological Sciences*, 263(1376): 1577–1584. (Cited on page 13.)

Brambati, A., Colosio, A., Zardoni, L., Galanti, L., and Liberi, G. 2014. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for dna stability. *Frontiers in genetics*, 6: 166–166. (Cited on page 135.)

Bresch, C., Niesert, U., and Harnasch, D. 1980. Hypercycles, parasites and packages. *Journal of Theoretical Biology*, 85(3): 399–405. (Cited on pages 22, 35, and 51.)

Carlson, M. and Botstein, D. 1982. Two differentially regulated mrnas with different 5' ends encode secreted and intracellular forms of yeast invertase. *Cell*, 28(1): 145–154. (Cited on page 79.)

Castellano-Pozo, M., Santos-Pereira, J. M., Rondón, A. G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., and Aguilera, A. 2013. R loops are linked to histone h3 s10 phosphorylation and chromatin condensation. *Molecular cell*, 52(4): 583–590. (Cited on page 135.)

Colizzi, E. S. and Hogeweg, P. 2015. Videos of the evolutionary dynamics for high costs, and of travelling waves. http://bioinformatics.bio.uu.nl/enrico. (Cited on pages 73, 75, 76, and 77.)

Crespi, B. J. 2001. The evolution of social behavior in microorganisms. *Trends in ecology & evolution*, 16(4): 178–183. (Cited on pages 72 and 79.)

Crick, F. H. 1958. On protein synthesis. 12(138-63): 8. (Cited on page 4.)

Crombach, A. and Hogeweg, P. 2008. Evolution of evolvability in gene regulatory networks. *PLoS Comput Biol*, 4(7): e1000112. (Cited on pages 3, 18, 135, and 150.)

Cuypers, T. D. and Hogeweg, P. 2012. Virtual genomes in flux: an interplay of neutrality and adaptability explains genome expansion and streamlining. *Genome biology and evolution*, 4(3): 212–229. (Cited on pages 18, 21, 150, and 152.)

Cuypers, T. D. and Hogeweg, P. 2014. A synergism between adaptive effects and evolvability drives whole genome duplication to fixation. *PLoS Comput Biol*, 10(4): e1003547. (Cited on page 18.)

Cwynar, L. C. and MacDonald, G. M. 1987. Geographical variation of lodgepole pine in relation to population history. *American Naturalist*, pages 463–469. (Cited on page 50.)

Darwin, C. 1872. *The origin of species*. (Cited on page 2.)

Datta, M. S., Korolev, K. S., Cvijovic, I., Dudley, C., and Gore, J. 2013. Range expansion promotes cooperation in an experimental microbial metapopulation. *Proceedings of the National Academy of Sciences*, 110(18): 7354–7359. (Cited on pages 71, 76, and 78.)

de Aguiar, M. A., Schneider, D. M., do Carmo, E., Campos, P. R., and Martins, A. B. 2015. Error catastrophe in populations under similarity-essential recombination. *Journal of theoretical biology*, 374: 48–53. (Cited on page 13.)

de Boer, F. K. 2012. *Coding Flexibility or How Evolution Shapes the Structure and Complexity of Coding*. Ph.D. thesis, Utrecht University, Theoretical Biology and Bioinformatics. (Cited on pages 20 and 121.)

de Boer, F. K. and Hogeweg, P. 2010. Eco-evolutionary dynamics, coding structure and the information threshold. *BMC evolutionary biology*, 10(1): 361. (Cited on pages 20, 21, 123, and 152.)

de Boer, F. K. and Hogeweg, P. 2012. Less can be more: Rna-adapters may enhance coding capacity of replicators. *PloS one*, 7(1): e29952. (Cited on pages 17 and 121.)

de Boer, F. K. and Hogeweg, P. 2014. Mutation rates and evolution of multiple coding in RNA-based protocells. *Journal of molecular evolution*, 79(5-6): 193–203. (Cited on page 17.)

de Boer, R. J. and Pagie, L. 2005. Grind, ver. 2.09. Last accessed: 14-07-2015. (Cited

on page 114.)

de Boer, R. J. and Staritsky, A. D. 2000. Cash. http://bioinformatics.bio.uu.nl/rdb/software.html. (Cited on pages 36, 52, 81, and 97.)

Diggle, S. P., Griffin, A. S., Campbell, G. S., and West, S. A. 2007. Cooperation and conflict in quorum-sensing bacterial populations. *Nature*, 450(7168): 411–414. (Cited on page 79.)

Doebeli, M. and Hauert, C. 2005. Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecology Letters*, 8(7): 748–766. (Cited on page 79.)

Doebeli, M., Hauert, C., and Killingback, T. 2004. The evolutionary origin of cooperators and defectors. *Science*, 306(5697): 859–862. (Cited on pages 71, 78, and 80.)

Domingo, E., Martínez-Salas, E., Sobrino, F., de la Torre, J. C., Portela, A., Ortín, J., López-Galindez, C., Pérez-Breña, P., Villanueva, N., Nájera, R., *et al.* 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevanceâĂŤa review. *Gene*, 40(1): 1–8. (Cited on page 13.)

Draghi, J. and Wagner, G. P. 2008. Evolution of evolvability in a developmental model. *Evolution*, 62(2): 301–315. (Cited on page 18.)

Draper, W. E., Hayden, E. J., and Lehman, N. 2008. Mechanisms of covalent self-assembly of the azoarcus ribozyme from four fragment oligonucleotides. *Nucleic acids research*, 36(2): 520–531. (Cited on page 35.)

Eigen, M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10): 465–523. (Cited on pages 6, 7, 9, 13, 95, and 152.)

Eigen, M. and Schuster, P. 1977. A principle of natural self-organization. *Naturwissenschaften*, 64(11): 541–565. (Cited on page 22.)

Eigen, M. and Schuster, P. 1978. The hypercycle. *Naturwissenschaften*, 65(1): 7–41. (Cited on pages 21, 51, 95, 96, and 123.)

Elena, S. F., Wilke, C. O., Ofria, C., and Lenski, R. E. 2007. Effects of population size and mutation rate on the evolution of mutational robustness. *Evolution*, 61(3): 666–674. (Cited on page 122.)

Fletcher, J. A. and Doebeli, M. 2009. A simple and general explanation for the evolution of altruism. *Proceedings of the Royal Society B: Biological Sciences*, 276(1654): 13–19. (Cited on pages 71, 72, 78, and 88.)

Fontana, W. 2002. Modelling "evo-devo" with rna. *BioEssays*, 24(12): 1164–1177. (Cited on page 121.)

Fontana, W., Konings, D. A., Stadler, P. F., and Schuster, P. 1993. Statistics of rna secondary structures. *Biopolymers*, 33(9): 1389–1404. (Cited on page 95.)

Freeland, S. J., Knight, R. D., Landweber, L. F., and Hurst, L. D. 2000. Early fixation of an optimal genetic code. *Molecular Biology and Evolution*, 17(4): 511–518. (Cited on page 16.)

French, S. L., Osheim, Y. N., Cioci, F., Nomura, M., and Beyer, A. L. 2003. In exponentially growing saccharomyces cerevisiae cells, rrna synthesis is determined by the summed RNA polymerase i loading rate rather than by the number of active genes. *Molecular and cellular biology*, 23(5): 1558–1568. (Cited on pages 32, 135, 151, and 152.)

Gilbert, W. 1986. Origin of life: The rna world. *Nature*, 319(6055). (Cited on pages 4 and 35.)

Grüner, W., Giegerich, R., Strothmann, D., Reidys, C., Weber, J., Hofacker, I. L., Stadler, P. F., and Schuster, P. 1996. Analysis of RNA sequence structure maps by exhaustive enumeration i. neutral networks. *Monatshefte für Chemie/Chemical Monthly*, 127(4): 355–374. (Cited on page 15.)

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. 1983. The rna moiety of ribonuclease p is the catalytic subunit of the enzyme. *Cell*, 35(3): 849–857. (Cited on page 35.)

Hamilton, W. D. 1964. The genetical evolution of social behaviour. i. *Journal of theoretical biology*, 7(1): 1–16. (Cited on pages 71 and 79.)

Happel, R. and Stadler, P. F. 1998. The evolution of diversity in replicator networks. *Journal of theoretical biology*, 195(3): 329–338. (Cited on page 123.)

Hazen, R. M., Griffin, P. L., Carothers, J. M., and Szostak, J. W. 2007. Functional information and the emergence of biocomplexity. *Proceedings of the National Academy of Sciences*, 104(suppl 1): 8574–8581. (Cited on page 5.)

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. 2013. Transcription-replication encounters, consequences and genomic instability. *Nature structural & molecular biology*, 20(4): 412–418. (Cited on page 32.)

Hermisson, J., Redner, O., Wagner, H., and Baake, E. 2002. Mutation–selection balance: Ancestry, load, and maximum principle. *Theoretical population biology*, 62(1): 9–46. (Cited on pages 9 and 100.)

Herrera-Moyano, E., Mergui, X., García-Rubio, M. L., Barroso, S., and Aguilera, A. 2014. The yeast and human fact chromatin-reorganizing complexes solve r-loop-mediated transcription–replication conflicts. *Genes & development*, 28(7): 735–748. (Cited on page 135.)

Hickinbotham, S. and Hogeweg, P. 2016. In preparation. (Cited on page 163.)

Higgs, P. G. and Lehman, N. 2014. The rna world: molecular cooperation at the origins of life. *Nature Reviews Genetics*. (Cited on pages 4, 52, and 79.)

Hildenbrandt, H., Carere, C., and Hemelrijk, C. K. 2010. Self-organized aerial displays of thousands of starlings: a model. *Behavioral Ecology*, 21(6): 1349–1359. (Cited on page 2.)

Hindré, T., Knibbe, C., Beslon, G., and Schneider, D. 2012. New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews Microbiology*, 10(5): 352–365. (Cited on pages 32 and 135.)

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. 1994. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie/Chemical Monthly*, 125(2): 167–188. (Cited on pages 14 and 96.)

Hogeweg, P. 1993. As large as life and twice as natural: Bioinformatics and the artificial life paradigm. *Complex systems: From biology to computation*, pages 2–11. (Cited on page 19.)

Hogeweg, P. 1994. Multilevel evolution: replicators and the evolution of diversity. *Physica D: Nonlinear Phenomena*, 75(1): 275–291. (Cited on pages 71 and 123.)

Hogeweg, P. 2007. From population dynamics to ecoinformatics: Ecosystems as multi-level information processing systems. *Ecological Informatics*, 2(2): 103–111. (Cited on page 3.)

Hogeweg, P. 2012. Toward a theory of multilevel evolution: long-term information integration shapes the mutational landscape and enhances evolvability. pages 195–224. (Cited on pages 3 and 32.)

Hogeweg, P. and Hesper, B. 1992. Evolutionary dynamics and the coding structure of sequences: multiple coding as a consequence of crossover and high mutation rates. *Computers & chemistry*, 16(2): 171–182. (Cited on pages 16 and 121.)

Hogeweg, P. and Takeuchi, N. 2003. Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. *Origins of Life and Evolution of the Biosphere*, 33(4-5): 375–403. (Cited on pages 35, 36, 38, 51, 71, 73, and 96.)

Holmes, E. C. 2010. The rna virus quasispecies: fact or fiction? *Journal of molecular biology*, 400(3): 271–273. (Cited on page 123.)

Hordijk, W., Hein, J., and Steel, M. 2010. Autocatalytic sets and the origin of life. *Entropy*, 12(7): 1733–1742. (Cited on page 4.)

Horning, D. P. and Joyce, G. F. 2016. Amplification of RNA by an RNA polymerase ribozyme. *Proceedings of the National Academy of Sciences*, page 201610103. (Cited on page 4.)

Hud, N. V., Cafferty, B. J., Krishnamurthy, R., and Williams, L. D. 2013. The origin of RNA and âĂIJmy grandfatherâĂŹs axeâĂİ. *Chemistry & biology*, 20(4): 466–474. (Cited on page 4.)

Huvet, M. and Stumpf, M. P. 2014. Overlapping genes: a window on gene evolvability. *BMC genomics*, 15(1): 721. (Cited on page 16.)

Huynen, M. 1993. *Evolutionary Dynamics and Pattern Generation in the Sequence and Secondary Structure of RNA*. Ph.D. thesis, Utrecht University, Theoretical Biology and Bioinformatics. (Cited on pages 20 and 122.)

Huynen, M. A. 1996. Exploring phenotype space through neutral evolution. *Journal of molecular evolution*, 43(3): 165–169. (Cited on pages 15 and 95.)

Huynen, M. A., Konings, D. A., and Hogeweg, P. 1993. Multiple coding and the evolutionary properties of rna secondary structure. *Journal of theoretical biology*, 165(2): 251–267. (Cited on page 95.)

Huynen, M. A., Stadler, P. F., and Fontana, W. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. *Proceedings of the National Academy of Sciences*, 93(1): 397–401. (Cited on page 15.)

Ide, S., Miyazaki, T., Maki, H., and Kobayashi, T. 2010. Abundance of ribosomal RNA gene copies maintains genome integrity. *Science*, 327(5966): 693–696. (Cited on pages 32, 135, 149, and 150.)

Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., *et al.* 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*, 333(6040): 348–353. (Cited on page 16.)

Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., and Serrano, L. 2008. Evolvability and hierarchy in rewired bacterial gene networks. *Nature*, 452(7189): 840–845. (Cited on page 18.)

Iwasa, Y., Michor, F., and Nowak, M. A. 2004. Stochastic tunnels in evolutionary dynamics. *Genetics*, 166(3): 1571–1579. (Cited on page 15.)

Jack, C. V., Cruz, C., Hull, R. M., Keller, M. A., Ralser, M., and Houseley, J. 2015. Regulation of ribosomal dna amplification by the TOR pathway. *Proceedings of the National Academy of Sciences*, 112(31): 9674–9679. (Cited on pages 32, 135, 136, 149, and 163.)

Janas, T., Janas, T., and Yarus, M. 2006. Specific RNA binding to ordered phospholipid bilayers. *Nucleic acids research*, 34(7): 2128–2136. (Cited on page 4.)

Jewell, J. L., Russell, R. C., and Guan, K.-L. 2013. Amino acid signalling upstream of mTOR. *Nature reviews Molecular cell biology*, 14(3): 133–139. (Cited on pages 140 and 153.)

Jones, E. I., Afkhami, M. E., Akçay, E., Bronstein, J. L., Bshary, R., Frederickson, M. E., Heath, K. D., Hoeksema, J. D., Ness, J. H., Pankey, M. S., *et al.* 2015. Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecology letters*. (Cited on page 80.)

Joyce, G. F. 2007. Forty years of in vitro evolution. *Angewandte Chemie International Edition*, 46(34): 6420–6436. (Cited on page 4.)

Killingback, T., Doebeli, M., and Knowlton, N. 1999. Variable investment, the continuous prisoner's dilemma, and the origin of cooperation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1430): 1723–1728. (Cited on pages 71, 72, and 79.)

Killingback, T., Bieri, J., and Flatt, T. 2006. Evolution in group-structured populations can resolve the tragedy of the commons. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1593): 1477–1481. (Cited on page 71.)

Kim, E. 2009. Mechanisms of amino acid sensing in mTOR signaling pathway. *Nutrition research and practice*, 3(1): 64–71. (Cited on page 153.)

Kim, N. and Jinks-Robertson, S. 2012. Transcription as a source of genome instability. *Nature Reviews Genetics*, 13(3): 204–214. (Cited on pages 135 and 144.)

Kim, N., Abdulovic, A. L., Gealy, R., Lippert, M. J., and Jinks-Robertson, S. 2007. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of dna replication. *DNA repair*, 6(9): 1285–1296. (Cited on pages 135 and 136.)

Kimura, M., Maruyama, T., and Crow, J. F. 1963. The mutation load in small populations. *Genetics*, 48(10): 1303. (Cited on page 163.)

Kirakosyan, Z., Saakian, D. B., and Hu, C.-K. 2010. Evolution models with lethal mutations on symmetric or random fitness landscapes. *Physical Review E*, 82(1): 011904. (Cited on page 122.)

Knibbe, C., Coulon, A., Mazet, O., Fayard, J.-M., and Beslon, G. 2007. A long-term evolutionary pressure on the amount of noncoding dna. *Molecular biology and evolution*, 24(10): 2344–2353. (Cited on pages 18 and 150.)

Kobayashi, T., Heck, D. J., Nomura, M., and Horiuchi, T. 1998. Expansion and contraction of ribosomal dna repeats in saccharomyces cerevisiae: requirement of replication fork blocking (fob1) protein and the role of RNA polymerase i. *Genes & development*, 12(24): 3821–3830. (Cited on page 135.)

Koella, J. C. 2000. The spatial spread of altruism versus the evolutionary response of egoists. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1456): 1979–1985. (Cited on page 80.)

Komarova, N. L., Sengupta, A., and Nowak, M. A. 2003. Mutation–selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Journal of theoretical biology*, 223(4): 433–450. (Cited on page 15.)

Könny, B. and Czárán, T. 2013. Spatial aspects of prebiotic replicator coexistence and community stability in a surface-bound rna world model. *BMC evolutionary biology*, 13(1): 204. (Cited on page 123.)

Könnyű, B., Czárán, T., and Szathmáry, E. 2008. Prebiotic replicase evolution in a surface-bound metabolic system: parasites as a source of adaptive evolution. *BMC*

evolutionary biology, 8(1): 267. (Cited on pages 36, 51, and 123.)

Konovalova, A., Petters, T., and Søgaard-Andersen, L. 2010. Extracellular biology of myxococcus xanthus. *FEMS microbiology reviews*, 34(2): 89–106. (Cited on page 79.)

Korolev, K. S. 2013. The fate of cooperation during range expansions. *PLoS Comput Biol*, 9(3): e1002994. (Cited on pages 50, 71, 76, and 78.)

Krakauer, D. C. and Plotkin, J. B. 2002. Redundancy, antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences*, 99(3): 1405–1409. (Cited on page 122.)

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. 1982. Self-splicing rna: autoexcision and autocyclization of the ribosomal rna intervening sequence of tetrahymena. *cell*, 31(1): 147–157. (Cited on page 35.)

Kun, A., Santos, M., and Szathmáry, E. 2005. Real ribozymes suggest a relaxed error threshold. *Nature genetics*, 37(9): 1008–1011. (Cited on pages 11 and 12.)

Kuznetsov, Y. A. 1999. Content, ver. 1.5. Last accessed: 09-08-2013. (Cited on page 114.)

Labib, K. and Hodgson, B. 2007. Replication fork barriers: pausing for a break or stalling for time? *EMBO reports*, 8(4): 346–353. (Cited on page 135.)

Lemaitre, C., Zaghloul, L., Sagot, M.-F., Gautier, C., Arneodo, A., Tannier, E., and Audit, B. 2009. Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. *BMC genomics*, 10(1): 1. (Cited on page 148.)

Letunic, I. and Bork, P. 2011. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research*, 39(suppl 2): W475–W478. (Cited on page 101.)

Leuthäusser, I. 1986. An exact correspondence between eigenâĂŹs evolution model and a two-dimensional ising system. *The Journal of chemical physics*, 84(3): 1884–1885. (Cited on page 13.)

Lincoln, T. A. and Joyce, G. F. 2009. Self-sustained replication of an rna enzyme. *Science*, 323(5918): 1229–1232. (Cited on page 35.)

Lindgren, K., Eriksson, A., and Eriksson, K.-E. 2004. Flows of information in spatially extended chemical dynamics. In *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, pages 456–460. (Cited on page 5.)

Long, E. O. and Dawid, I. B. 1980. Repeated genes in eukaryotes. *Annual review of biochemistry*, 49(1): 727–764. (Cited on page 141.)

Lorenz, R., Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. 2011. ViennaRNA package 2.0. *Algorithms for Molecular Biology*, 6(1): 1. (Cited on page 13.)

Luján, A. M., Gómez, P., and Buckling, A. 2015. Siderophore cooperation of the bacterium pseudomonas fluorescens in soil. *Biology letters*, 11(2): 20140934. (Cited on page 79.)

Ma, W. and Hu, J. 2012. Computer simulation on the cooperation of functional molecules during the early stages of evolution. *PloS one*, 7(4): e35454. (Cited on page 123.)

Maree, A. F., Panfilov, A. V., and Hogeweg, P. 1999. Phototaxis during the slug stage of dictyostelium discoideum: a model study. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1426): 1351–1360. (Cited on page 2.)

Marek, M. S., Johnson-Buck, A., and Walter, N. G. 2011. The shape-shifting quasispecies of RNA: one sequence, many functional folds. *Physical Chemistry Chemical Physics*, 13(24): 11524–11537. (Cited on page 17.)

Marín, A., Tejero, H., Nuño, J. C., and Montero, F. 2012. Characteristic time in quasispecies evolution. *Journal of theoretical biology*, 303: 25–32. (Cited on page 8.)

Martin, L. L., Unrau, P. J., and Müller, U. F. 2015. Rna synthesis by in vitro selected ribozymes for recreating an rna world. *Life*, 5(1): 247–268. (Cited on page 35.)

Más, A., López-Galíndez, C., Cacho, I., Gómez, J., and Martínez, M. A. 2010. Unfinished stories on viral quasispecies and darwinian views of evolution. *Journal of molecular biology*, 397(4): 865–877. (Cited on page 13.)

Mast, C. B., Schink, S., Gerland, U., and Braun, D. 2013. Escalation of polymerization in a thermal gradient. *Proceedings of the National Academy of Sciences*, 110(20): 8030–8035. (Cited on page 4.)

Maynard Smith, J. 1979. Hypercycles and the origin of life. *Nature*, 280: 445–446. (Cited on pages 35 and 51.)

Maynard Smith, J. 2000. The concept of information in biology. *Philosophy of science*, pages 177–194. (Cited on page 13.)

Maynard Smith, J. and Szathmary, E. 1995. *The major transitions in evolution*. Oxford, UK: WE Freeman. (Cited on page 22.)

McCaskill, J. S., Füchslin, R. M., and Altmeyer, S. 2001. The stochastic evolution of catalysts in spatially resolved molecular systems. *Biological chemistry*, 382(9): 1343–1363. (Cited on page 51.)

Merrikh, H., Zhang, Y., Grossman, A. D., and Wang, J. D. 2012. Replication–transcription conflicts in bacteria. *Nature reviews Microbiology*, 10(7): 449–458. (Cited on page 135.)

Miethke, M. and Marahiel, M. A. 2007. Siderophore-based iron acquisition and pathogen control. *Microbiology and Molecular Biology Reviews*, 71(3): 413–451. (Cited on page 79.)

Mills, D., Peterson, R., and Spiegelman, S. 1967. An extracellular darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences of the United States of America*, 58(1): 217. (Cited on page 35.)

Mitchell, M., Forrest, S., and Holland, J. H. 1992. The royal road for genetic algorithms: Fitness landscapes and ga performance. In *Proceedings of the first european conference on artificial life*, pages 245–254. Cambridge: The MIT Press. (Cited on page 122.)

Mitteldorf, J. and Wilson, D. S. 2000. Population viscosity and the evolution of altruism. *Journal of Theoretical Biology*, 204(4): 481–496. (Cited on page 71.)

Momeni, B., Waite, A. J., and Shou, W. 2013. Spatial self-organization favors heterotypic cooperation over cheating. *Elife*, 2: e00960. (Cited on pages 71, 76, and 78.)

Monod, J. 1970. *Le hasard et la nécessité: essai sur la philosophie naturelle de la biologie moderne*. (Cited on page 2.)

Muller, H. J. 1964. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1): 2–9. (Cited on page 163.)

Müller, M. J., Neugeboren, B. I., Nelson, D. R., and Murray, A. W. 2014. Genetic drift opposes mutualism during spatial population expansion. *Proceedings of the National*

Academy of Sciences, 111(3): 1037–1042. (Cited on page 71.)

Nahum, J. R., Harding, B. N., and Kerr, B. 2011. Evolution of restraint in a structured rock–paper–scissors community. *Proceedings of the National Academy of Sciences*, 108(Supplement 2): 10831–10838. (Cited on page 71.)

Nedelcu, A. M. 2009. Environmentally induced responses co-opted for reproductive altruism. *Biology letters*, 5(6): 805–808. (Cited on page 163.)

Niesert, U., Harnasch, D., and Bresch, C. 1981. Origin of life between scylla and charybdis. *Journal of Molecular Evolution*, 17(6): 348–353. (Cited on pages 35 and 51.)

Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481): 920–930. (Cited on page 35.)

Nowak, M. A. and May, R. M. 1992. Evolutionary games and spatial chaos. *Nature*, 359(6398): 826–829. (Cited on page 71.)

Ohtsuki, H., Hauert, C., Lieberman, E., and Nowak, M. A. 2006. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092): 502–505. (Cited on pages 30, 71, 72, and 79.)

Olendorf, R., Getty, T., and Scribner, K. 2004. Cooperative nest defence in red–winged blackbirds: reciprocal altruism, kinship or by–product mutualism? *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1535): 177–182. (Cited on page 79.)

Paul, N. and Joyce, G. F. 2002. A self-replicating ligase ribozyme. *Proceedings of the National Academy of Sciences*, 99(20): 12733–12740. (Cited on page 35.)

Payne, J. L. and Wagner, A. 2014. The robustness and evolvability of transcription factor binding sites. *Science*, 343(6173): 875–877. (Cited on page 18.)

Penn, A. S., Conibear, T. C., Watson, R. A., Kraaijeveld, A. R., and Webb, J. S. 2012. Can simpson's paradox explain co-operation in pseudomonas aeruginosa biofilms? *FEMS Immunology & Medical Microbiology*, 65(2): 226–235. (Cited on page 80.)

Piatigorsky, J. 2003. Gene sharing, lens crystallins and speculations on an eye/ear evolutionary relationship. *Integrative and comparative biology*, 43(4): 492–499. (Cited on page 163.)

Powner, M. W., Gerland, B., and Sutherland, J. D. 2009. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459(7244): 239–242. (Cited on page 4.)

Rajamani, S., Ichida, J. K., Antal, T., Treco, D. A., Leu, K., Nowak, M. A., Szostak, J. W., and Chen, I. A. 2010. Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *Journal of the American Chemical Society*, 132(16): 5880–5885. (Cited on page 13.)

Reynolds, C. W. 1987. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH computer graphics*, 21(4): 25–34. (Cited on page 2.)

Robertson, M. P. and Joyce, G. F. 2012. The origins of the RNA world. *Cold Spring Harbor perspectives in biology*, 4(5): a003608. (Cited on page 4.)

Rothstein, R., Michel, B., and Gangloff, S. 2000. Replication fork pausing and recombination or "gimme a break". *Genes & Development*, 14(1): 1–10. (Cited on page 148.)

Rutten, J. P., Hogeweg, P., and Beslon, G. 2016. In preparation. (Cited on page 163.)

Samal, A., Wagner, A., and Martin, O. C. 2011. Environmental versatility promotes modularity in genome-scale metabolic networks. *BMC systems biology*, 5(1): 1. (Cited

on page 18.)

Sankar, T. S., Wastuwidyaningtyas, B. D., Dong, Y., Lewis, S. A., and Wang, J. D. 2016. The nature of mutations induced by replication–transcription collisions. *Nature*, 535(7610): 178–181. (Cited on page 135.)

Santos, F. C. and Pacheco, J. M. 2005. Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters*, 95(9): 098104. (Cited on page 71.)

Sardanyés, J. and Solé, R. V. 2006. Ghosts in the origins of life? *International Journal of Bifurcation and Chaos*, 16(09): 2761–2765. (Cited on page 48.)

Sardanyés, J., Elena, S. F., and Solé, R. V. 2008. Simple quasispecies models for the survival-of-the-flattest effect: The role of space. *Journal of theoretical biology*, 250(3): 560–568. (Cited on page 121.)

Savill, N. J., Rohandi, P., and Hogeweg, P. 1997. Self-reinforcing spatial patterns enslave evolution in a host-parasitoid system. *J. theor. Bio.*, 188: 11–20. (Cited on pages 24 and 51.)

Scheuring, I., Czárán, T., Szabó, P., Károlyi, G., and Toroczkai, Z. 2003. Spatial models of prebiotic evolution: soup before pizza? *Origins of life and evolution of the biosphere*, 33(4-5): 319–355. (Cited on page 51.)

Schrettl, M., Bignell, E., Kragl, C., Joechl, C., Rogers, T., Arst, H. N., Haynes, K., and Haas, H. 2004. Siderophore biosynthesis but not reductive iron assimilation is essential for aspergillus fumigatus virulence. *The Journal of experimental medicine*, 200(9): 1213–1219. (Cited on page 79.)

Schultes, E. A. and Bartel, D. P. 2000. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science*, 289(5478): 448–452. (Cited on page 17.)

Schuster, P. and Stadler, P. F. 2008. Early replicons: origin and evolution. *Origin and Evolution of Viruses*, 2: 1–42. (Cited on page 6.)

Schuster, P. and Swetina, J. 1988. Stationary mutant distributions and evolutionary optimization. *Bulletin of mathematical biology*, 50(6): 635–660. (Cited on pages 10, 95, and 121.)

Schuster, P., Fontana, W., Stadler, P. F., and Hofacker, I. L. 1994. From sequences to shapes and back: a case study in rna secondary structures. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1344): 279–284. (Cited on pages 15 and 95.)

Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. 2010. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007): 1099–1102. (Cited on page 136.)

Serganov, A. and Nudler, E. 2013. A decade of riboswitches. *Cell*, 152(1): 17–24. (Cited on pages 4 and 17.)

Sherwood, A. V. and Henkin, T. M. 2016. Riboswitch-mediated gene regulation: Novel RNA architectures dictate gene expression responses. *Annual Review of Microbiology*, 70(1). (Cited on page 4.)

Shine, R., Brown, G. P., and Phillips, B. L. 2011. An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences*, 108(14): 5708–5711. (Cited on page 50.)

Sinclair, D. A. and Guarente, L. 1997. Extrachromosomal rdna circlesâĂŤa cause of aging in yeast. *Cell*, 91(7): 1033–1042. (Cited on page 135.)

Smaldino, P. E., Schank, J. C., and McElreath, R. 2013. Increased costs of cooperation help cooperators in the long run. *The American Naturalist*, 181(4): 451–463. (Cited on pages 71, 79, and 80.)

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3): 431–432. (Cited on page 101.)

Stadler, B. M. and Stadler, P. F. 2003. Molecular replicator dynamics. *Advances in Complex Systems*, 6(01): 47–77. (Cited on page 22.)

Stadler, P. F. 1991. Dynamics of autocatalytic reaction networks iv: Inhomogeneous replicator networks. *Biosystems*, 26(1): 1–19. (Cited on page 22.)

Stewart, P. S. and Costerton, J. W. 2001. Antibiotic resistance of bacteria in biofilms. *The lancet*, 358(9276): 135–138. (Cited on page 79.)

Stich, M. and Manrubia, S. C. 2011. Motif frequency and evolutionary search times in RNA populations. *Journal of theoretical biology*, 280(1): 117–126. (Cited on page 8.)

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R., and Schmidt, T. M. 2015. rrndb: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, 43(D1): D593–D598. (Cited on page 141.)

Suweis, S., Simini, F., Banavar, J. R., and Maritan, A. 2013. Emergence of structural and dynamical properties of ecological mutualistic networks. *Nature*, 500(7463): 449–452. (Cited on page 71.)

Szabó, P., Scheuring, I., Czárán, T., and Szathmáry, E. 2002. In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature*, 420(6913): 340–343. (Cited on pages 25, 36, 50, 51, 71, and 80.)

Szathmáry, E. 1991. Simple growth laws and selection consequences. *Trends in ecology & evolution*, 6(11): 366–370. (Cited on pages 22 and 123.)

Szathmáry, E. and Demeter, L. 1987. Group selection of early replicators and the origin of life. *Journal of theoretical biology*, 128(4): 463–486. (Cited on pages 28, 35, 51, and 96.)

Szilágyi, A., Zachar, I., and Szathmáry, E. 2013. Gause's principle and the effect of resource partitioning on the dynamical coexistence of replicating templates. *PLoS computational biology*, 9(8): e1003193. (Cited on page 123.)

Szolnoki, A., Antonioni, A., Tomassini, M., and Perc, M. 2014. Binary birth-death dynamics and the expansion of cooperation by means of self-organized growth. *EPL (Europhysics Letters)*, 105(4): 48001. (Cited on pages 71, 79, and 80.)

Takeuchi, N. and Hogeweg, P. 2007a. Error-threshold exists in fitness landscapes with lethal mutants. *BMC Evolutionary Biology*, 7(1): 15. (Cited on pages 12, 100, and 122.)

Takeuchi, N. and Hogeweg, P. 2007b. The role of complex formation and deleterious mutations for the stability of RNA-like replicator systems. *Journal of molecular evolution*, 65(6): 668–686. (Cited on pages 13, 25, 38, 105, and 115.)

Takeuchi, N. and Hogeweg, P. 2008. Evolution of complexity in RNA-like replicator systems. *Biol. Direct*, 3: 20. (Cited on pages 27, 37, 50, 96, 100, 104, 113, 118, 119, 160, and 161.)

Takeuchi, N. and Hogeweg, P. 2009. Multilevel selection in models of prebiotic evolution ii: a direct comparison of compartmentalization and spatial self-organization. *PLoS computational biology*, 5(10): e1000542. (Cited on pages 27, 28, 36, 38, 40, 48, 50,

51, 53, 65, 71, 73, 96, 123, and 160.)

Takeuchi, N. and Hogeweg, P. 2012. Evolutionary dynamics of rna-like replicator systems: a bioinformatic approach to the origin of life. *Physics of life reviews*, 9(3): 219–263. (Cited on pages 12, 24, and 79.)

Takeuchi, N., Poorthuis, P. H., and Hogeweg, P. 2005. Phenotypic error threshold; additivity and epistasis in RNA evolution. *BMC Evolutionary Biology*, 5(1): 9. (Cited on page 10.)

Takeuchi, N., Hogeweg, P., and Koonin, E. V. 2011. On the origin of dna genomes: evolution of the division of labor between template and catalyst in model replicator systems. *PLoS computational biology*, 7(3): e1002024. (Cited on pages 25, 36, 51, and 123.)

Takeuchi, N., Kaneko, K., and Hogeweg, P. 2016. Evolutionarily stable disequilibrium: endless dynamics of evolution in a stationary population. In *Proc. R. Soc. B*, volume 283, page 20153109. The Royal Society. (Cited on page 28.)

Takeuchi, Y., Horiuchi, T., and Kobayashi, T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rdna. *Genes & development*, 17(12): 1497–1506. (Cited on pages 135, 136, and 148.)

Tejero, H., Marín, A., and Montero, F. 2010. Effect of lethality on the extinction and on the error threshold of quasispecies. *Journal of theoretical biology*, 262(4): 733–741. (Cited on pages 6 and 122.)

Tokuriki, N. and Tawfik, D. S. 2009. Protein dynamism and evolvability. *Science*, 324(5924): 203–207. (Cited on page 16.)

Torney, C., Neufeld, Z., and Couzin, I. D. 2009. Context-dependent interaction leads to emergent search behavior in social aggregates. *Proceedings of the National Academy of Sciences*, 106(52): 22055–22060. (Cited on page 2.)

Travis, J. M. and Dytham, C. 2002. Dispersal evolution during invasions. *Evolutionary Ecology Research*, 4(8): 1119–1129. (Cited on page 50.)

Trifonov, E. N. 1989. The multiple codes of nucleotide sequences. *Bulletin of Mathematical Biology*, 51(4): 417–432. (Cited on page 121.)

Trifonov, E. N. 2011. Thirty years of multiple sequence codes. *Genomics, proteomics & bioinformatics*, 9(1): 1–6. (Cited on page 16.)

Tuck, A. C. and Tollervey, D. 2011. Rna in pieces. *Trends in Genetics*, 27(10): 422–432. (Cited on page 121.)

Van Baalen, M. and Rand, D. A. 1998. The unit of selection in viscous populations and the evolution of altruism. *Journal of Theoretical Biology*, 193(4): 631–648. (Cited on pages 30, 71, and 79.)

van Ballegooijen, W. M. and Boerlijst, M. C. 2004. Emergent trade-offs and selection for outbreak frequency in spatial epidemics. *Proceedings of the National Academy of Sciences*, 101(52): 18246–18250. (Cited on pages 24, 36, and 51.)

Van Der Laan, J. D. and Hogeweg, P. 1995. Predator-prey coevolution: interactions across different timescales. *Proceedings of the Royal Society of London B: Biological Sciences*, 259(1354): 35–42. (Cited on page 19.)

Van Dyken, J. D., Müller, M. J., Mack, K. M., and Desai, M. M. 2013. Spatial population expansion promotes the evolution of cooperation in an experimental prisonerâĂŹs dilemma. *Current Biology*, 23(10): 919–923. (Cited on pages 30, 71, 76, and 78.)

van Hoek, M. J. and Hogeweg, P. 2009. Metabolic adaptation after whole genome du-

plication. *Molecular biology and evolution*, 26(11): 2441–2453. (Cited on page 140.)

Van Nimwegen, E. and Crutchfield, J. P. 2000. Metastable evolutionary dynamics: crossing fitness barriers or escaping via neutral paths? *Bulletin of mathematical biology*, 62(5): 799–848. (Cited on page 15.)

van Nimwegen, E. and Crutchfield, J. P. 2000. Optimizing epochal evolutionary search: population-size independent theory. *Computer Methods in Applied Mechanics and Engineering*, 186(2): 171–194. (Cited on page 122.)

Van Nimwegen, E. and Crutchfield, J. P. 2001. Optimizing epochal evolutionary search: Population-size dependent theory. *Machine Learning*, 45(1): 77–114. (Cited on page 8.)

Van Nimwegen, E., Crutchfield, J. P., and Huynen, M. 1999. Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences*, 96(17): 9716–9720. (Cited on pages 10, 11, 95, and 121.)

Van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory*, 1: 30. (Cited on pages 19, 122, and 126.)

Vasas, V., Fernando, C., Santos, M., Kauffman, S., and Szathmáry, E. 2012. Evolution before genes. *Biology Direct*, 7(1): 1. (Cited on page 4.)

von der Dunk, S., Colizzi, E. S., and Hogeweg, P. 2016. In preparation. (Cited on page 161.)

Wagner, A., Andriasyan, V., and Barve, A. 2014. The organization of metabolic genotype space facilitates adaptive evolution in nitrogen metabolism. *Journal of Molecular Biochemistry*, 3(1). (Cited on page 18.)

Wagner, G. and Krall, P. 1993. What is the difference between models of error thresholds and muller's ratchet? *Journal of Mathematical Biology*, 32(1): 33–44. (Cited on page 12.)

Wakano, J. Y., Nowak, M. A., and Hauert, C. 2009. Spatial dynamics of ecological public goods. *Proceedings of the National Academy of Sciences*, 106(19): 7910–7914. (Cited on pages 71 and 80.)

Ward, W. L., Plakos, K., and DeRose, V. J. 2014. Nucleic acid catalysis: metals, nucleobases, and other cofactors. *Chemical reviews*, 114(8): 4318–4342. (Cited on page 4.)

West, S. A., Diggle, S. P., Buckling, A., Gardner, A., and Griffin, A. S. 2007. The social lives of microbes. *Annual Review of Ecology, Evolution, and Systematics*, pages 53–77. (Cited on page 72.)

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., and Adami, C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412(6844): 331–333. (Cited on pages 10 and 121.)

Wilson, D. S. 1975. A theory of group selection. *Proceedings of the national academy of sciences*, 72(1): 143–146. (Cited on pages 71 and 79.)

Wilson, D. S. 1979. Structured demes and trait-group variation. *American naturalist*, pages 606–610. (Cited on pages 78 and 88.)

Wilson, D. S., Pollock, G., and Dugatkin, L. A. 1992. Can altruism evolve in purely viscous populations? *Evolutionary ecology*, 6(4): 331–341. (Cited on page 30.)

Wilson, R. C. and Doudna, J. A. 2013. Molecular mechanisms of RNA interference. *Annual review of biophysics*, 42: 217–239. (Cited on page 4.)

Wochner, A., Attwater, J., Coulson, A., and Holliger, P. 2011. Ribozyme-catalyzed transcription of an active ribozyme. *Science*, 332(6026): 209–212. (Cited on page 35.)

Woese, C. R. 1965. On the evolution of the genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, 54(6): 1546. (Cited on page 16.)

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. 1. (Cited on page 12.)

Zhang, S., Blain, J. C., Zielinska, D., Gryaznov, S. M., and Szostak, J. W. 2013. Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. *Proceedings of the National Academy of Sciences*, 110(44): 17732–17737. (Cited on page 6.)

Samenvatting

In dit proefschrift onderzoeken we hoe nieuwe functies kunnen ontstaan door Darwiniaanse Evolutie. Van het begin af aan, ongeveer vier miljard jaar geleden, heeft evolutie nieuwe eigenschappen, vormen en functies gegenereerd. In dit vroege evolutionaire stadium bestond er nog geen cellulair leven, en volgens de "RNA Wereld" hypothese functioneerden in plaats daarvan RNA moleculen zowel als informatie-opslagmedium (nu de taak van DNA) en als katalysator voor chemische reacties (de huidige rol van eiwitten). Omdat zulke RNA moleculen elkaar's replicatie katalyseerden (replicatie is dus een altruïstische eigenschap), hadden parasieten, die wel gerepliceerd werden maar die geen tijd besteedden aan het repliceren van anderen, een evolutionair voordeel. Werd de overlevingskans van replicatoren hierdoor bedreigd? Hoe konden betere replicatoren evolueren als parasieten een selectief voordeel hadden? In hoofdstuk 2 laten we zien dat sterke parasieten indirect de evolutie van betere replicatoren stimuleren, dankzij een terugkoppeling via de ruimtelijke zelforganisatie van deze twee soorten.

Het probleem van de evolutie van altruïstische en coöperatieve eigenschappen beperkt zich niet tot de RNA wereld, maar geldt ook voor bestaande organismen. Een voorbeeld is de productie van gedeelde maar kostbare stoffen - zogenaamde publieke goederen - door meerdere microben. Zelfzuchtige individuen die geen publieke goederen produceren, kunnen wel profiteren van de coöperatieve individuen, maar het ecosysteem valt uiteen als niemand publieke goederen produceert. Is er minder selectie voor publiek-goedproductie als de productiekosten hoger zijn, omdat dan de selectie voor zelfzucht hoger is? In hoofdstuk 3 laten we zien dat zelfzuchtige individuen inderdaad evolueren als de productiekosten hoog zijn, en dat ze ruimtelijke patronen vormen met coöperatieve individuen. Deze patronen beïnvloeden de evolutie van publiek-goedproductie door coöperatieve individuen, die meer gaan samenwerken, terwijl zelfzuchtige individuen dat juist minder gaan doen.

In hoofdstuk 4 geven we RNA-achtige replicatoren een genotype en een fenotype, namelijk een nucleotide-sequentie en een secundaire structuur door het vouwen van het RNA. We bestuderen de evolutionaire gevolgen van zo'n complexe genotype-fenotype vertaling bij hoge mutatiesnelheid. We zien dat een functioneel ecosysteem ontstaat, waarin nieuwe functies worden uitgevoerd door RNA sequenties die niet gerepliceerd kunnen worden. Deze sequenties worden daarom alleen gevormd via replicatie en mutatie van een mastersequentie, die dus de informatie voor het hele ecosysteem bevat.

Dit ecosysteem verhoogt weer de overlevingskansen van de mastersequentie.

Het stabiliserende effect van mutaties is experimenteel waargenomen in het rRNA gen-cluster van gist, waar mutaties als gevolg van transcriptie-replicatieconflicten worden uitgebuit om het aantal rRNA genen te verhogen, afhankelijk van de beschikbaarheid van voedingsstoffen. In hoofdstuk 5 modelleren we de mutatiedynamica van gist rRNA en zien dat hogere mutatiesnelheden zorgen voor een stabieler genoom, mits de mutaties gebiast zijn richting genduplicatie en -deletie. Dit mechanisme werkt ondanks dat het korte-termijneffect van deze mutaties vrijwel neutraal is, net als voor de gist rRNA genen.

Ten slotte benadrukken de resultaten in dit proefschrift de voordelen van een meerlagige ("multilevel") aanpak om de eindeloze vindingrijkheid van evolutie te begrijpen.

Compendio

L'oggetto di ricerca di questa tesi è il modo in cui nuove funzioni emergono da un processo di evoluzione darwiniana. Fin dal suo principio, circa quattro miliardi di anni fa, l'evoluzione ha generato nuovi tratti e funzioni; in quella fase la vita era molto diversa da come la conosciamo oggi. Le cellule non esistevano ancora e gli acidi ribonucleici (RNA), secondo l'ipotesi del mondo a RNA, svolgevano un duplice ruolo: di conservazione dell'informazione genetica (funzione che oggi svolge il DNA) e di catalizzatori di reazioni chimiche (funzione oggi svolta dalle proteine). Dato che tali RNA catalizzavano la replicazione gli uni degli altri (in un processo di "replicazione reciproca" come forma di altruismo), la selezione naturale avrebbe dovuto favorire l'evoluzione dei parassiti, ovvero di quegli RNA replicati in gran numero ma che non replicano altri RNA.

La sopravvivenza dei replicatori è minacciata dalla presenza dei parassiti? Come possono gli RNA accrescere le loro capacità riproduttive se la selezione naturale favorisce i parassiti? Nel capitolo 2 si mostra che parassiti più aggressivi generano (in modo indiretto) le condizioni ottimali affinché i replicatori si riproducano velocemente a seguito dell'innesco di un meccanismo a feedback mediato dall'auto organizzazione nello spazio delle due specie.

Le dinamiche evoluzionistiche di altruismo e cooperazione non si limitano al mondo a RNA, ma sono presenti tutt'oggi negli organismi moderni. Un esempio è la produzione da parte dei micro-organismi di sostanze condivise ma energeticamente costose, i cosiddetti "beni pubblici". Gli individui egoisti che non apportano alcun bene, possono sopravvivere approfittando di quei micro-organismi che cooperano, ma se nessuno coopera il sistema può arrivare al collasso. Dato che la selezione favorisce l'egoismo quando i costi (energetici) di produzione di un bene condiviso sono elevati, come puÃš evolvere la produzione di un bene comune? Nel capitolo 3 si mostra che con alti costi gli organismi egoisti evolvono organizzandosi nello spazio assieme gli organismi che cooperano. L'auto-organizzazione di individui che cooperano e di individui egoisti genera le condizioni ideali per aumentare la produzione di bene comune tra gli organismi che cooperano, e per diminuirla tra quelli egoisti.

Nel capitolo 4 è assegnato ai replicatori genotipo e fenotipo, rispettivamente, una sequenza di nucleotidi e la sua (conseguente) struttura secondaria. Il capitolo affronta le conseguenze evoluzionistiche di un alto tasso di mutazioni su una mappa genotipo-fenotipo complessa (il ripiegamento dell'RNA). L'evoluzione genera un ecosistema in

cui nuove funzioni emergono come fenotipi di sequenze che non possono essere replicate, e che sono invece prodotte per mutazione di una sequenza "master". Dunque tale sequenza codifica l'informazione per l'intero ecosistema, e l'informazione viene decodificata dal processo mutazionale. L'ecosistema, dal canto suo, è di importanza critica per la sopravvivenza della sequenza master.

Le mutazioni genetiche possono quindi essere determinanti nel mantenimento del livello di fitness individuale.

Per esempio i lieviti sfruttano le mutazioni risultanti dai conflitti tra trascrizione e replicazione del DNA per accrescere il numero di copie del gene per l'rRNA, con un meccanismo dipendente dall'abbondanza di risorse. Nel capitolo 5, si studia un modello di dinamiche mutazionali osservate in lievito e si mostra che le mutazioni possono essere benefiche per l'integrità del genoma nel lungo periodo quando si trovano con una maggiore frequenza nelle duplicazioni e delezioni di geni. Questo meccanismo funziona nonostante tali mutazioni siano quasi neutrali nel breve-periodo, in accordo con le osservazioni sul lievito.

In conclusione, i risultati della tesi evidenziano i vantaggi di un approccio multi-livello nella comprensione dei meccanismi dell'Evoluzione e della sua infinita inventività.

Curriculum vitæ

Enrico Sandro Colizzi was born in Monselice, Italy on May 23, 1986. He attended the Liceo Scientifico (high school with scientific specialisation) E. Curiel, where he graduated in 2005. From 2005 to 2008 he studied Biotechnology at the University of Padova, where he obtained his Bachelor's degree. After a year studying Industrial Biotechnology at the University of Padova, in 2009 he moves to Utrecht to study Theoretical Biology, where he obtained his Master's degree in 2011. The day after completing his master's, he began his postgraduate research under the supervision of Prof. Dr. Paulien Hogeweg. The results of his PhD are described in this thesis.

List of Publications

E.S. Colizzi, P. Hogeweg. Evolution of functional diversification within quasispecies. *Genome biology and evolution*. 6(8); 2014.

E.S. Colizzi, P. Hogeweg. High cost enhances cooperation through the interplay between evolution and self-organisation. *BMC evolutionary biology*. 16(1); 2016.

E.S. Colizzi, P. Hogeweg. Parasites sustain and enhance RNA-like replicators through spatial self-organisation. *PLoS Computational Biology*. 12(4); 2016.

E.S. Colizzi, P. Hogeweg. Increased rate of duplications and deletions prevents evolutionary deterioration. *In preparation*. 2016.

Acknowledgments

We make models that teach us about Nature.

I learnt from my supervisor - Paulien Hogeweg - that the results of a valuable model extend beyond what they are originally meant to teach. Could we say, then, that the teachings of a valuable teacher extend beyond the subject she teaches? I learnt a lot in these years and, perhaps more importantly, I learnt how to learn further.

Thank you Paulien, thank you so much.

I wish to thank the members of the reading committee for the time they spent reading this thesis: Guillaume Beslon, Nicholas Hud, Susanna Manrubia, Eörs Szathmáry, Andreas Wagner.

Most of the work in this thesis (and most of that which did not become part of it) required great computational resources as well as great patience from our system administrator, Jan-Kees van Amerongen, to whom I am sincerely grateful.

I thank my colleagues (past and present), teachers and friends at the Theoretical Biology and Bioinformatics group, both for all I learnt and for all the fun times. A big thank to each and all my friends, that willing or not *had* to discuss with me the latest science-thing I had found on the Internet. Please know that carrying out this PhD would have been nothing short of impossible without your support.

I would like to explicitly acknowledge those friends and colleagues that were most significant for me "scientifically": Nobuto Takeuchi, Folkert de Boer, Thomas Cuypers, Chris van Dorp, Bram van Dijk, Hilje Doekes. Double thanks to Thomas and Bram for being my paranimfen, to Renske Vroomans for translating the summary to Dutch and to Francesca Sara Colizzi for her help with the Italian version of the summary.

Un grande grazie a mia madre per il supporto - anche materiale - in tutti questi anni (tu più di tutti sai da quanto lontano venga il mio amore per la scienza).

Finally, dear Renske, thank you for blurring with me the line between scientific collaboration and everyday life.

Yet all experience is an arch wherethro'
Gleams that untravell'd world whose margin fades
For ever and for ever when I move.

Alfred, Lord Tennyson