Spatiotemporal modelling of ecological and evolutionary problems

DSc. dissertation

Czárán Tamás
MTA-ELTE Ökológiai és Elméleti Biológiai Kutatócsoport
H-1117 Budapest, Pázmány Péter sétány 1/C

2006. Budapest
Contents

Preface ................................................................................................................................5
The structure of the dissertation..........................................................................................7
Acknowledgements.............................................................................................................9

Part I. Models of Spatiotemporal Dynamics .................................................................10
  I.1. Introduction .............................................................................................................11
  I.2. Simplifying assumptions of classical population dynamics .................................13
  I.3. Translating individual states (i-states) to population states (p-states) ..................15
  I.4. Spatial extensions of the classical approach: considering spatial constraints .........17
  I.5. The object-based paradigm: neighbourhood modelling ......................................19
  I.6. Recent applications of spatiotemporal models ....................................................23
  I.7. Classification and characterisation of spatiotemporal models .............................24
  I.8. The components of spatial effects in CA – mean-field and configuration-field
      approximations ........................................................................................................34

Part II. Population Dynamics ......................................................................................38
  II.1. Killer-Sensitive coexistence in metapopulations of microorganisms ....................39
  II.2. Spatial constraints decrease the invasion rate of fitter mutants in asexual
       populations .............................................................................................................54

Part III. Microbial Evolution ......................................................................................63
  III.1. Evolution of sexual asymmetry .........................................................................64
  III.2. Chemical warfare between microbes promotes biodiversity .............................80
  III.3. A spatial model of the evolution of quorum sensing .........................................90
  III.4. Quorum sensing: the evolutionary benefit of dishonest signalling ....................103

Part IV. Prebiotic Evolution ......................................................................................117
  IV.1. Coexistence of metabolically co-operating replicators in a cellular automaton:
       the importance of space without mesoscopic structure ......................................118
  IV.2. Prebiotic replicase evolution in a metabolic system ..........................................128

Epilogue ..........................................................................................................................144
References: ......................................................................................................................145
Preface

Writing a dissertation is a peculiar genre of scientific activity. It requires a measure of determination on behalf of the candidate, much patience from the family involved, and a lot of hardly rewarded work from all the colleagues taking part in the evaluation process. In addition, the product itself – the dissertation – is inevitably meant for a rather narrow audience: usually it is neither published nor otherwise made accessible to professionals apart from the evaluating committee and perhaps a small group of immediate colleagues. I have to admit I was convinced for these reasons that it is not much more than just an obligatory, uncomfortable and lonely display of one’s scientific achievements – until I actually started writing it. Now I am partly converted to the opinion that, even though all the above apply, it might still be a useful experience for at least one reason: it forces the candidate to look at his/her own work from a more general, wider perspective compared to the one taken when writing a paper on a specific scientific issue. Such a wider perspective helps to see the merits and the limits of the approaches and the actual methods one has ever applied to answer what one considered interesting scientific questions, and it may even shed light on possible new paths of research to follow in the future.

In my case the dissertation project helped me putting my own field of research into context within what can be called a generalized spatiotemporal syndynamics. All three terms of this – admittedly somewhat pompous, yet I also think rather accurate – name deserve a short explanation. First, my way of thinking about problems of population coexistence and coevolution is definitely syndynamical (sensu Juhász-Nagy, 1986). This means that the questions I pose always concern patterns of coexistence on the community level. However, my approach is bottom-up: I consider populations interacting under specific environmental conditions and expect them to assemble and form new units on higher levels of organization (such as plant or animal communities). Obviously, such population-level mechanisms must be responsible for shaping the symmorphology of the higher units, but it might be very difficult to find direct causation between the elementary mechanisms of population interactions and the emergent symmorphological patterns (cf. Bartha, 2004; Bartha et al., 1995, 1997, 2004). Second, my approach is not only temporal but spatiotemporal from the outset, which fact is due to my early and deep conviction that spatial structure plays a crucial, in many cases even decisive role in the outcome of virtually any dynamical process conceivable in
biology in general, and in syndynamics in particular. Ample proof has been accumulated for this statement by now, both in the theoretical and in the experimental part of the ecological literature (e.g., Kröel-Dulay & Coffin, 1998; Gassmann et al., 2000; Arditi et al., 2001; Grimm & Berger, 2002; Johansson & Sumpter, 2003; Mizera & Meszéna, 2003; Namba & Hashimoto, 2004; Hilker et al., 2004; Magyar & Kertész, 2004; Mágori et al., 2005) – up to the point of it becoming a triviality of the discipline. Finally, my view of syndynamics is generalized in two different respects. One is the object of research: the actual range of problems I have studied stretches from syndynamical problems of plant communities to those of microbes and even prebiotic replicators (e.g., Czárán, 1985, 1989, 1992, 1993; Czárán & Bartha, 1989; Czárán et al., 2002; Czárán & Hoekstra, 2003, 2004; Czárán & Szathmáry, 2000). All the problems I have dealt with, irrespective of their actual objects, are related to ecology in the sense that the dynamical models I have applied in my studies are basically ecological, both in spirit and in the methods used. The other aspect of generalization steps over disciplinary borders: besides problems of coexistence in the population dynamical sense, I have studied evolutionary and coevolutionary processes in bacterial communities and in prebiotic systems as well. These can also be regarded as natural extensions to syndynamics, as long as the basic question remains that of emergent patterns of coexistence in a set of interacting populations, albeit on a longer timescale and in an evolving genotype space.
The structure of the dissertation

The dissertation consists of four parts. Part One is largely methodological: it is a short typology of the most frequently used models of spatiotemporal dynamics in general, based on an earlier monograph (Czárán, 1998). This part also includes the qualitative description of a new method I have developed for the assessment of the dynamical role that spatial constraints play in lattice models (Czárán & Hockstra, 2003). The method is called the configuration-field approximation, and it is a model type between spatially explicit stochastic cellular automata (SCA) and their completely nonspatial mean-field approximations.

Part Two contains a few of my recent models on spatiotemporal population and community dynamics in yeast and bacterium strains. These models demonstrate that spatial constraints are sometimes the key for understanding the coexistence of populations that would exclude one another in a well-mixed, completely homogeneous environment. In Chapter II.1 the configuration-field approximation is calculated in detail for a two-species system of competing yeast populations.

Part Three is devoted to evolutionary dynamics in interacting microbial populations. Chapter III.1 tries to give an answer to the old question why sex is binary, i.e., why are there two genders (mating types) in almost any sexual organism? The second chapter in this part (Chapter III.2) deals with the microevolutionary buildup of extreme biodiversity in bacteriocin-producing strains of bacteria and killer yeasts in homogeneous environments – a puzzling phenomenon that can be explained easily with spatial modelling, but hardly any other way. Chapters III.3 and III.4 provide tentative answers and even more questions regarding another puzzle of microbial evolution, namely the appearance and maintenance of quorum sensing, a way of communication between cooperating bacteria.

Part Four is about prebiotic evolutionary dynamics. It contains my contribution to the theory of surface evolution in a metabolic RNA enzyme system. Chapter IV.1 is the „prototype” model demonstrating that a replicator system of a few types of RNA-like macromolecules, each possessing a specific metabolic enzyme activity, can maintain genetic information beyond the error threshold, due to the coexistence of cooperating replicators on a surface. The model provides an ecological
solution to the old chicken-or-egg problem of prebiotic information integration. Chapter IV.2 develops the idea further, letting the „parasite” replicator of the metabolic system mutate and thus gain a weak RNA-replicase activity. The „converted” parasite starts working for the common good by helping the replication of all the replicators of the system. At the end of the chapter I also sketch a plausible scenario for the appearance of the first metabolic proto-cell (including metabolism and the replicase, and dressed in a self-produced membrane vesicle) based on this model.

In the Epilogue I briefly sketch my research plans for the near future.
Acknowledgements

Even though writing a dissertation is an isolating experience, the actual scientific work that it is meant to account for is inspiring and social. It is in fact difficult, if not impossible, to do effective and enjoyable research without the constant input of, and detailed discussions with, fellow scientists. I have had the privilege of always enjoying the company of colleagues and friends open for scientific and non-scientific discussions – I am truly grateful to them all.

My first and most important inspiration came from Pál Juhász-Nagy, my mentor during my university years and later, whose broad mind and exceptional personality had always been, and has remained after his passing, an essential motivation of my scientific work and my way of looking at the world. I know we are many in this country feeling the same.

I wish to thank my friend, Sanyi Bartha, for the exciting discussions we have had on very diverse topics during the years since our graduation – these had a significant impact on my thinking about science in general and specific scientific problems in particular.

I have published many of my papers with co-authors (S. Bartha, Sz. Bokros, M. Habets, R. Hoekstra, J. Hofbauer, E. Jablonka, Gy. Károlyi, É. Kisdi, I. Molnár, B. Oborny, L. Pagie, J. Podani, I. Scheuring, Gy. Szabó, P. Szabó, E. Szathmáry), all of whom contributed not only the papers themselves but my intellectual habit as well. Thanks for that.

A number of science funding organisations have sponsored the work presented in this dissertation. Financial support from The Hungarian Scientific Research Fund (OTKA), Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), NATO Science Program, and Wageningen University are gratefully acknowledged.

Last, but not least: my family had to practice a considerable amount of patience with me during the past few decades, and especially during the past few months while I was writing the dissertation. A father sitting by the computer all day and sometimes half night too is neither the most entertaining nor the most useful member of a family. Thanks to my wife, Judit, and my children, Dorka, Domi and Bogi, for their love and care all along these years – I could not have managed without them.
Part One

Models of Spatiotemporal Dynamics
I.1. Introduction

The mutual dependence of spatial structure and temporal dynamics counts as a slumberous commonplace in recent ecology textbooks. This being the case, the lack of a broad interface between structural and dynamical studies in ecology - which has been, and partly remains, a serious bother of the discipline - might seem puzzling. The overwhelming majority of structural research has been directed towards revealing the spatial pattern and the topographical-topological relations of co-occurring populations by statistical means, while most dynamical studies omitted spatial pattern altogether, allocating considerable effort into the purely temporal description of population processes.

The structural-dynamical dichotomy originated - at least in part - from differences in the attitudes and the methodologies of plant and animal ecology. Taking the Humboldtian viewpoint, most plant ecologists were primarily motivated by the need to discover the structural rules behind the seemingly static patterns of vegetation on many different spatial scales, whereas animal ecology inquired into the more obvious temporal changes of abundance in animal populations and communities, which do not usually seem to preserve a static spatial structure for a long time. The former attitude gave rise to the structural-static-statistic discipline of biosociology (coenology), whereas the latter resulted in the development of the classical, spatially astructural-dynamic-analytic approach of population dynamics, following in the pioneering steps of Malthusian demography.

Ever since the Humboldtian start, biosociology has remained closely tied to field practice, theoretical efforts directed mostly towards the related statistical methodologies. The interest of biosociologists has been mainly focussed on the spatial structure of an existing population, or that of a real community. By now, sophisticated (and computerized) oligo- and multivariate statistical methods help discovering spatial and diversity patterns in much detail (e.g., ter Braak, 1986, 1994; Podani, 1989, 2001, 2005; Tóthmérész, 1994, 1996), but – despite repeated attempts to relate community structure to community dynamics on a statistical basis (e.g., Wilson, 1991, 1994) – these methods alone do not seem sufficient to reveal specific dynamical mechanisms producing specific spatial patterns. Biosociology does not fully explain how and why a certain pattern develops in a community, but it can provide a detailed phenomenological description of the spatial and temporal pattern of communities in statistical
terms, which may give hints on what kind of dynamical rules might have acted in producing a certain spatial structure or diversity pattern (cf. Gallé, 1991; Padisák, 1998; Fekete & Virágh, 2000; Izsák & Papp, 2000; Newstrom and Armstrong, 2003; Remmel & Csillag, 2003; Sapoukhina et al., 2003).

On the other hand, amidst the exciting endeavour to discover and explain mechanisms leading to different temporal abundance patterns of populations and communities, classical population dynamics have largely given up the requirement of direct reference to actual field situations. The object of a classical population dynamical model can be a fictitious population or a number of them, each assigned only a few attributes representing general properties such as birth and death rates, interaction coefficients etc. - the attributes that are considered dynamically most relevant in a given context. The typical questions these models are suitable to answer regard the qualitative properties of the differential or difference equations that govern the dynamics of the model system; typical answers express criteria for the existence of equilibria, and specify their stability properties if they exist. Such dynamical models are called strategic, their aim being not to reproduce or to predict a concrete field situation, but to delimit the range of possible consequences for a certain set of simple assumptions, initial and boundary conditions - that is, for an ecological mechanism devoid of effects other than those defined in the model.

Strategic models are the means for obtaining a deep understanding of the relatively simple ecological mechanisms they are meant to represent - but these are hardly ever found in their pure form in nature. It is, therefore, very difficult to cite examples for simple, low-dimensional population dynamical models that proved really helpful in reproducing or predicting actual field or experimental situations. For that end, it is almost always necessary (but, to be honest, not always sufficient) to consider many more biologically relevant details of the specific process to be simulated. The model thus obtained is a tactical representation of the concrete situation.

Apart from a very limited number of exceptions, classical theoretical population dynamics is purely strategic in the above sense. Different strategic models could produce almost any temporal coexistence pattern that appears on the field, but comparing these to actual field data, one could never be sure that the mechanism specified in the model is really an accurate representation of the one acting on the field. Completely different mechanisms might produce the same phenomenon - just think of the textbook example of the oscillating hare-lynx system, which can be explained very convincingly with at least three fundamentally different models. To decide - or at least to guess on an acceptable basis - which is
the actual mechanism behind a certain temporal pattern, one has to compare 
the model parameters to those measured on the field. This is practicable 
only if the model parameters are relatively easily measured, which is 
unfortunately not the case for most strategic population dynamical 
representations.

The gap between the field-oriented practice of biosociology and the 
model-oriented theory of population dynamics has become discouragingly 
wide in this century, both disciplines having followed their own attitudes 
and methodologies for a long time, without much reference to each other’s 
results. Now that the structural aspects of population and community 
dynamics on the one hand, and the dynamical constraints on community 
pattern formation on the other are clear to most ecologists, the need to 
bridge the gap seem to have become urgent for both sides. Spatial and 
temporal patterns of populations and communities cannot be reliably 
interpreted one without the other any more.

1.2. Simplifying assumptions of classical population 
dynamics

The strategic models of classical, nonspatial population dynamics are based 
on three fundamental assumptions:

1) populations consist of large numbers of individuals (abundance 
assumption);

2) all individuals of the same population are identical in every 
dynamically relevant respect (uniformity assumption); and

3) the movement of the individuals is such that the population as a 
whole can be treated as an ergodic\(^1\) system (ergodicity assumption).

Assumptions 1-3 are equivalent to the postulates of statistical mechanics 
and chemical kinetics - disciplines spectacularly successful in explaining 
many aspects of the macroscopic behaviour of gases and well-mixed liquid 
solutions, starting from the microscopic properties of the particles they 
consist of. The “mass action” type rules of chemical reaction kinetics are 
directly derived from the stoichiometry of chemical reactions and the 
models of statistical mechanics, the latter built on the general assumptions 
1-3, the laws of classical mechanics, and statistical principles. Individual

\(^1\) The ergodicity condition roughly refers to perfect spatial mixing in this context: each particle (molecule, 
individual) “feels” the same environment around it, because the particles move fast and independently 
one from the other, thus they collide (interact) with probabilities proportional to the products of their 
densities - or, in more chemical terms, their concentrations.
particles being indistinguishable, the fate of any single particle is irrelevant and impossible to follow.

Relying on the general assumptions in common with those of statistical mechanics and chemical kinetics, many methods from the powerful mathematical arsenal of physico-chemical disciplines were naturally adapted to attack population dynamical problems. This led to a proliferation of phenomenological population models, the best known classical examples of which are the Malthusian model of exponential population growth, the Pearl-Verhulst model of logistic growth, and the Lotka-Volterra models of population interactions. These, and a lot of other models, yielded deep theoretical insight into many aspects of population processes, which, beyond doubt, could not ever have been achieved without them.

On the other hand, however, the lack of a broader interface between theoretical population dynamics and field ecology can be in a large part attributed to the fact that assumptions 1-3 happen not to apply to biological objects as well as they do to molecule populations. The uniformity assumption is violated in every actual field case, almost without exception, in population biology: individual variation within biological populations is much higher than that of the particles in a usual physical or chemical system. One can say that, in fact, every individual of a community should be regarded unique for its combination of dynamically relevant genetical, physiological and environmental properties, even when comparing individuals of the same species (Huston et al., 1988; DeAngelis and Gross, 1992). Ecological objects also violate the abundance and the ergodicity criteria: the number of individuals is fewer by tens of orders of magnitude, even in a very dense community, than the number of particles in a test tube, and their motion is usually very far from ergodic on all spatiotemporal scales. The ergodicity condition is spectacularly violated by plants and benthic animals, which do not move at all through most of their lifetime: they can interact with close neighbours only, and practically they do not feel the presence of individuals outside a limited spatial interval around them. But even populations of animals capable of very fast movement do not mix perfectly: individuals might stick to home ranges, mates or resource patches to some extent, thus they experience more or less neighbourhood effect as well, albeit possibly on a larger spatial scale and in a fuzzier manner than sessile organisms do. Many of the abundant cases of even qualitative misfits between the predictions of classical population dynamical models and actual field data can be attributed to the violation of the ergodicity assumption. Of course, neither statistical mechanics nor chemical kinetics suffer so much from the discrepancy between theoretical
expectations and empirical observations so much, because their usual objects approximately satisfy assumptions 1-3: they are really huge sets of identical particles, which can be safely assumed to mix well and move fast enough to “average out”, on the macroscopic spatiotemporal scale, the effects of the local context on the microscopic level.

I.3. Translating individual states (i-states) to population states (p-states)

Assumptions 1-3, which proved to be sufficiently realistic in models of simple physico-chemical systems are, in fact, imposed by a methodological constraint on population dynamics. Mass-action type population interaction rules are necessary to postulate, because these allow for mathematical representations simple enough for analytical treatment. Even then, the number of interacting populations, that is, the dimensionality of the system, remains a critical methodological problem of classical population dynamics. Analytical tractability requires this number to be either small (not larger than 3 or 4 in most cases), or else infinitely large. For the number of populations to be acceptably low, many individuals are to be lumped together, based on their similarity regarding a dynamically relevant characteristic, or a combination of such characteristics. What individual characteristics make the basis of the grouping is determined by the question the model is expected to answer. It can be the species identity, age, developmental stage, size, or the spatial position of the individuals, for example, among many more. At one extreme, all individuals can be assumed identical, which is the case in the single-species, unstructured models of Malthusian or logistic population growth, for example. Following the conceptual framework and the terminology of Metz and Diekman (1986) and Caswell and John (1992), we say that all individuals are in the same *individual state* or *i-state* in these models.

A Lotka-Volterra type competition model, for example, represents interactions among populations that differ in species identity. This is the only relevant *i-state variable* (descriptor) of the model, and it takes $s$ different values ($s$ being the number of species considered), so that the *i-state space* of such models is one-dimensional and discrete. All individuals belonging to the same species are assumed identical, forming a homogeneous mass - in fact, the individuals as discrete entities are not considered in such classical models at all. The basic objects are the populations themselves, labelled by their species names as the corresponding *i-state* values. It is meaningless to ask anything about certain individuals; the relevant dynamical questions regard the *abundances* of the
competing species. All mechanisms of population growth and competitive interaction are formulated in terms of species abundances, and they affect species abundances directly. This situation is the analogue of a well-mixed chemical system in a fluid medium, in which $s$ different compounds interact (react) with each other. The abundance distribution on the $i$-states is called the population state or $p$-state, which is, in the case of a Lotka-Volterra competition model, represented by the $s$-dimensional vector $\mathbf{X}$ of species abundances. Note that the dimensionality of the $p$-state space (in this example, $s$) is equal to the number of possible $i$-states considered, and not the dimensionality of the $i$-state space (which is 1 in the example).

The dynamical rules of the Lotka-Volterra model can be formulated on the $p$-state level, as a system of ordinary differential equations, of which the population densities $x_i$ are the state variables:

$$\frac{dx_i}{dt} = r_i x_i - \sum_{j=1}^{s} \alpha_{ij} x_j x_i \quad (i = 1,\ldots,s).$$

The species interact through the “mass action” type rules of classical population dynamics, which are analogous to those of chemical kinetics: the intensity of the interaction between populations $i$ and $j$ is proportional to the $x_i x_j$ product of their densities (concentrations). The parameters of the model are $p$-state parameters, which characterise the speed of population growth ($r_i$) for species $i$ when it is not constrained by competition, and the competitive effect ($\alpha_{ij}$) that a unit density of population $i$ suffers in the presence of a unit density of population $j$. The relevant questions these models can address are the stability properties of the $p$-state, that is, the stationarity conditions of the abundance distribution on the $i$-states. This is why Caswell and John (1992) and Maley and Caswell (1993) call such systems $i$-state distribution models, suggesting this term as a more precise substitute for the “state variable model” concept of Huston et al. (1988).

Based on a somewhat less abstract property that is common in all $i$-state distribution models, namely that they define the dynamical relations of the populations in terms of mass-interaction, I shall use the more intuitive term “mass-interaction model” as a synonym for “$i$-state distribution model” in the forthcoming.

If more than one $i$-state descriptors make the difference among individuals (for example, both species identity and larval stage are dynamically relevant), than a particular $i$-state can be thought of as a point in the $i$-state space (which is then two-dimensional and discrete in both
dimensions), spanned by the \( i \)-state descriptors as axes. Therefore the \( p \)-state representation is not a vector distribution, but a higher dimensional distribution of abundances (a 2D matrix distribution in the example). This makes a difference in analytical tractability, as the number of possible \( i \)-states, and thus the number of simultaneous equations might be large, but the resulting model is still a system of ordinary differential equations (ODE’s). If, however, at least one of the \( i \)-state descriptors take a continuum of values (e.g., age in continuous time), then the \( i \)-space is also continuous in the corresponding dimension, and the \( i \)-state distribution (mass interaction) problem becomes infinite dimensional. The appropriate mathematical tool to attack this is a system of partial differential equations (PDE’s).

I.4. Spatial extensions of the classical approach: considering spatial constraints

The consideration of spatial aspects in population dynamics allows for the relaxation of at least one of the biologically unrealistic 1-3 postulates: spatial population models never satisfy the ergodicity assumption on all scales. In the most general terms, this means that overall spatial mixing, thus the spatially homogeneous growth and interaction of the populations is replaced by a spatially constrained mechanism.

Populations obviously interact one with the other through the actual interactions of their individuals. In fact it is not the whole population of blue tits that competes with the whole population of great tits for their insect meals, just as it is not the population of wolves that chases and kills the deer population. Individual tits compete for resources, and individual wolves hunt for deer individuals. All such events are spatially constrained: only nearby individuals can engage in any kind of interaction. Moreover, the spatial dispersion of the offspring of an individual is usually also seriously constrained – this is very obvious in plants propagating with short stolons, for example, but it is to a large extent true for most other plants and even to most animals as well, even if on different spatial scales. These simple facts may have profound consequences on the outcome of interactions on the population level. The local nature of interactions and the spatial limitations applying to offspring dispersion are the spatial constraints of population dynamics.

Spatial constraints can be considered in many different ways in population dynamical models. A straightforward way is to extend a nonspatial mass-interaction model - for example, a Lotka-Volterra type
model - so that the $i$-states of the subpopulations include spatial position as well. If space is continuous, then the $i$-state descriptors of spatial position are Cartesian coordinates, i.e., continuous variables; if space is discrete, the corresponding $i$-states are discrete identifiers of locality (index names of islands, habitat islands). In the former case, the model is a reaction-diffusion system, represented by a system of PDE’s. In the latter, we have a patch-abundance model, represented by a system of ODE’s, one equation for each $i$-state. Reaction-diffusion and patch-abundance models are both mass-interaction systems, since it is still the abundance (density) of the populations that is directly affected by the dynamics - but the local, not the overall abundance, unlike in nonspatial models. In fact the interactions of individuals within the localities are ergodic in these systems as well, only the large-scale movements between the localities are spatially constrained.

If space is not in any explicit way included among the relevant $i$-states, but the model implicitly assumes a spatial structure of the habitat, affecting the abundance dynamics of the resident populations, we have a spatially implicit mass-interaction model. Metapopulation (i.e., patch-occupancy) and aggregated interaction (mainly host-parasitoid) models are the basic categories of this class. Metapopulation models fit into the nonspatial mass-interaction framework, with the conceptual modification that it is the habitat patch that takes the role of the individual, and the whole set of patches is analogous to the population. Thus, the fate of a single habitat patch is irrelevant and impossible to trace in a metapopulation model. The analogy can be so close that metapopulation models are often formally equivalent with certain nonspatial population models. Examples for $i$-state descriptors (variables) in a metapopulation model are resident species combination and patch age, to take just the most obvious ones; it is the abundance distribution of the patches on the $i$-state space that is directly modelled.

Aggregated interaction models - the other type of spatially implicit mass-interaction systems - represent a very different approach. These incorporate the effects of spatial habitat structure by manipulating the interaction terms of originally nonspatial models, so that the dynamical consequences of the aggregation of interactions can be assessed without any explicit reference to space. Kareiva (1990) call these models “pseudospatial”.
I.5. The object-based paradigm: neighbourhood modelling

A common feature of all mass-interaction models - spatial and nonspatial alike - is that they conform to the general assumptions of the high abundance and the uniformity of individuals within any possible $i$-state. It is quite a recent development of theoretical ecology that these assumptions are not unavoidable any more. There is a fast developing class of models directly based on the fact that any two individuals do differ in at least one aspect from each other: even if they are - unrealistically - postulated to be identical for example in their genetic constitution, physiology or environment, the similarity of their $i$-states is always constrained spatially and temporally. To be short, two individuals cannot be present on the same place simultaneously.

Given the inherently local manner of the interactions among individuals, the positional aspect of their $i$-states might be of much dynamical relevance, especially on smaller spatial scales. This is an aspect of $i$-state that is necessarily omitted in mass-interaction models, which always assign a large number of individuals to a single spatial point - this is most obviously the case in nonspatial systems, but the same applies to patch models and diffusion models as well. If spatial position is dynamically important even on a scale close to the size of a single organism, the individuals cannot be legitimately lumped together into large, internally homogeneous groups, and no such groups can be regarded as the units of interaction. On the contrary, the subdivision of the population must go down close to the level of single individuals in order to find the ultimate, internally homogeneous unit of interaction. Then each such interacting unit must be characterised by its own $i$-state, differing from those of all the others. The most extreme case is when the interacting unit is the individual itself, so that the number of $i$-states equals the number of individuals. But the interacting unit needs not be the solitary individual at all. It can be, for example, a small group of individuals belonging to the same family, a small colony occupying the same habitat patch, or a tussock of grass, the ramets of which belong to the same genet (clone). $i$-state distribution approaches are obviously inadequate to represent the dynamics in such situations, because the criteria for the application of the statistical principles underlying the mass-interaction formalism are seriously violated, e.g. by the unicity of the interacting units. At this level of dynamical resolution, the rules of interaction must be defined among the unique objects themselves, not among the densities of internally uniform groups of individuals. That is, the dynamical rules act on the $i$-state level, altering the
i-states of the interacting objects directly, unlike in mass-interaction models, where the dynamical rules act on the abundances of the i-state classes, i.e., on the p-states.

Caswell and John (1992) call the i-state-based interaction rules constraint functions, and the population-level representation of i-states, in a certain environmental setting, an i-state configuration (see also Maley and Caswell, 1993.). The i-state configuration of a given system is a list of the individual objects it includes, together with their i-states. A simple example for an i-state configuration is a map showing the spatial positions of all individuals, each assigned all the i-state attributes (e.g., species identity and/or age, size etc.) that are considered as dynamically relevant in the model of interest.

An i-state configuration type p-state is obviously more informative than is an i-state distribution type p-state for the same set of individuals, since the number of i-states considered is equal - or at least it is close - to the number of individuals in the former, whereas it is much less in the latter case. Therefore it is always possible to create a number of i-state distributions from an i-state configuration by simply omitting some i-state descriptors and lumping individuals together differing in those only, but the reverse cannot be done without relying on additional information. In the above example, omitting the spatial coordinates of the individuals would yield an i-state distribution, with individuals of the same species (age, size, etc.) belonging to the same group.

The constraint function is the set of rules defining the way how i-state configurations are transformed in time. It can be a function (either deterministic or stochastic) of any one, or any combination, of the i-state descriptors. As stated before, the constraint function acts on the level of the individual objects, by altering i-states directly. Such dynamical systems are called i-state configuration models by Caswell and John (1992), as an extension of the less general concept of individual-based models (Huston et al., 1988). I propose the more intuitive but equally precise “object-interaction model” synonym for the term “i-state configuration model”.

If the argument of the constraint function includes spatial position, the object-interaction model is explicitly spatial; if spatial proximity is one of the factors determining the strength of the interaction between the objects, one has a neighbourhood model in hand. That is, neighbourhood models are a subset of the spatial subset of object-interaction models: an object-interaction model need not be spatial, and a spatial object-interaction model need not be a neighbourhood model. For example, morphological constitution, physiological status, age, developmental stage, motivational
state, or any other \( i \)-state descriptor might determine interaction among individuals, without any direct reference to spatial position. Even if there is spatial reference in the transformation rules, it might not be limited to a certain neighbourhood around the individual objects. Thus, there is a wealth of potentially different object-interaction systems. It is a sign of the importance of spatial proximity in ecological interactions that by far the most object-interaction approaches in ecology belong to the specific class of neighbourhood models.

Neighbourhood models developed so far can be naturally and conveniently classified into four major groups, according to their spatiotemporal setup and the regarding constraint functions:

1) cellular automata (interacting particle systems);
2) tessellation models;
3) distance models.

A rough characterisation of these model types can be based on the way they define neighbourhoods and the nature of their interaction rules. Fig. 1.1 provides examples for some of the most common neighbourhood definitions (cf. Czárán and Bartha, 1992; Czárán, 1998).

From the theoretical viewpoint, interacting particle systems and cellular automata differ from tessellation models and distance models in at least two important respects: 1) IPS and CA models are spatially discrete, whereas tessellations and distance models are spatially continuous; 2) in fact the real interacting object of IPS and CA models is the discrete spatial
unit (site) within the grid of such units, whereas the interacting unit is the individual itself in the other two types of neighbourhood models.

The second difference is a fundamental one: in IPS and CA it is always a unit of space, that is, the site (or cell) to which different i-states are assigned, and it is only in specific cases that this i-state is an attribute of a single individual occupying the site. It is the fate of the site, not of the individual, that is followed in IPS and CA models; site states can be - and in most models they are - different from the state of a single individual. Thus, in the strict sense, the name “interacting particle systems” is somewhat misleading: the physical applications of IPS models are not particle based, and the ecological applications are not individual based. State descriptions and state transitions of IPS and CA apply to localities directly, therefore the i-state variable can be different from anything applicable to a single individual. It can be, for example, the species composition occupying the site, if the spatial resolution of the model requires that, like in CA implementations of metapopulation models (Caswell and Etter, 1993). In some other systems the relevant state of a site is the (small) number of individuals it actually contains (e.g., Weiner and Conte, 1981; Czárán, 1989; Palmer, 1992). Even if the site-to-individual correspondence is unambiguous at a certain point of time, it needs not stay the same by the next generation: a mobile organism might move from one site to another (as in the model of De Roos et al., 1991, for example), whereby it alters the i-state of two sites simultaneously: the one it leaves, and the one it arrives to. It is also possible that an individual occupies more than one site of the grid at the same time, as happens to be the case in models of clonal plant growth (e.g., Ford, 1987; Callaghan et al., 1990; Oborny, 1994a). That is, the mapping of individuals onto the lattice may be variable both in space and in time.

Unlike in IPS and CA models, the basic interacting object of a TM or a DM is always the solitary individual, the i-state of which always includes its spatial position. This difference serves as the basis for classifying interacting particle systems and cellular automata as site-based neighbourhood models, while tessellation models and distance models are discussed under the heading individual-based neighbourhood models in the sequel. There is also a difference between these two types in the mathematical representation of space: site-based models are spatially discrete, whereas the overwhelming majority of individual-based models is spatially continuous. Tessellation models allocate a part of the habitat space to every individual, whose dynamical performance is a direct function of the area and the shape of the captured habitat space segment. The allocation of space can be implemented in many different ways, depending on the
spatial, temporal and physiological constraints of the organisms considered. In distance models, the pairwise interaction between any two individuals is a direct function of their distance in space - but the actual interaction functions can be of very different forms.

Neighbourhood models got established as standard theoretical tools in practically all branches of natural sciences in the past decade, from theoretical physics through geography to many branches of biology. There are strong signs of their penetration into economics and the humanities - sociology, for example - as well, wherever the spatial interactions of unique entities can be an issue of interest. Besides their popularity as model systems in many branches of science, interacting particle systems, cellular automata and tessellations are eminent fields of research for a new, thoroughly computerized mathematical discipline called experimental mathematics, which combines analytical and numerical techniques in the study of the formal structure of these (and other) models.

Existing neighbourhood models are very diverse both conceptually and structurally. There is one methodological point in which they are similar, however, and that methodological similarity reflects a deep structural homology. Principally owing to the fact that the fate of each individual object is kept track of through time in neighbourhood models, there is hardly any chance to analyse them by pure analytical techniques - the most adequate (and almost always the only possible) means of discovering the dynamical behaviour of a neighbourhood system is computer simulation. Analytical approximations are feasible in specific cases, but these presume the omission of very important spatial aspects (e.g., the variation in neighbourhood composition) usually. The consequence of the loss of analytical tractability is an inevitable loss in the generality of results and thus of the theoretical insight obtainable (Murdoch et al., 1992; Lomnicki, 1992; Metz and DeRoos, 1992), but many of the recent theoretical problems of population and community dynamics cannot even be formulated, let alone solved, without an object-interaction modelling framework (cf. Judson, 1994; Wolff, 1994; Dunning et al., 1995; Turner et al., 1995; Conroy et al., 1995; Holt et al., 1995).

I.6. Recent applications of spatiotemporal models

Due to considering spatial population structure on a wide scale of different approaches, spatiotemporal modelling is perhaps the most adequate means of bridging the methodological gap between theoretical and field ecology. Reaction-diffusion and patch-abundance models - being direct extensions of the classical nonspatial systems - are ideally suited for assessing the
genuine effects of spatial structure on population dynamics by comparing
the predictions of the spatiotemporal systems to those of their nonspatial
counterparts, all other things kept equal as much as possible. The reaction-
diffusion and the patch-abundance analogues of certain nonspatial systems
might qualitatively differ in their dynamical behaviour both from the
nonspatial version and from each other; the explicit inclusion of the spatial
aspect often leads to more realistic coexistence patterns. Metapopulation
and aggregated interaction models (i.e., spatially implicit systems) are
routinely tested against field data, but at the same time many of them
provide valuable theoretical insights as well. Almost the same applies to
site-based neighbourhood models, the important difference being that while
a spatially implicit model can often be relevant theoretically and practically
alike, almost any concrete cellular automaton or interacting particle system
model may be quite securely categorised as either theoretically or
practically motivated (that is, strategic or tactical). It is the class of
individual-based neighbourhood models that is closest to field practice both
in its characteristic spatiotemporal scale and in the potential realism of the
biological mechanisms that a distance model or a tessellation model can
incorporate. The majority of existing individual-based models are tactical
for the time being, albeit the possibility of applying them to more general
theoretical problems is open. With distance models and tessellation models,
population dynamics has reached the dynamical resolution level where
computer-simulated processes are directly comparable to actual field
processes on the smallest relevant spatiotemporal scale. Some applications
of spatiotemporal population dynamics to microbial communities are
discussed in Part Two.

Spatiotemporal modelling approaches have recently been extended to
models of evolutionary dynamics in many different levels of organisation
(from chemical evolution to the evolution of languages) as well – Parts
Three and Four of this dissertation will discuss some of these in some
detail.

I.7. Classification and characterisation of
spatiotemporal models

Fig. I.2. is a diagrammatic classification of spatiotemporal population
dynamical models, constructed in accordance with the theoretical aspects
emphasized above. The structure of discussing the models in this part of the
dissertation corresponds to this classification in the first place.
**Mass-interaction** type spatial approaches can be divided into two broad classes: *spatially explicit* and *spatially implicit* models. Spatially explicit models are reaction-diffusion systems and patch-abundance models. The two classes of spatially implicit mass-interaction models are metapopulation systems and aggregated interaction models.

Spatial **object-interaction** systems, i.e., neighbourhood models can be classified as *site-based* approaches (cellular automata, or, more generally, interacting particle systems which also include percolation models as specific cases). The other class of object-interaction models is that of *individual-based* neighbourhood models (tessellation models and distance models), is the most diverse as for the number of different model types. The family of distance models incorporates three different approaches: fixed radius neighbourhood models, zone of influence models and ecological field models.

The order of the short descriptions of these model types below roughly follows the historical order of their appearance. Along this “historical gradient”, a few canonical trends can be detected: the approximate decrease in the spatiotemporal scale of the processes of interest from regional to local; the increase in the number of biological details that can be incorporated; and the increasing role of numerical techniques in the analysis of the models. Mass-interaction systems are mainly strategic, relatively large-scale in both space and time, and analytical, whereas neighbourhood models are mostly tactic, spatiotemporally small-scale and numerical. Of course, there are a few exceptions to these statements in each model category, but the general trend is this.

Some general features are shared by all models of the dynamics of spatially distributed populations. Specifically, they assume that births, deaths and interactions are local events, and that the spatial range of
dispersal - roughly speaking, the measure of how far the offspring of an individual can get from the parent- is limited. But they differ in the assumptions regarding the spatial structure of the habitat, in the specific form of the vital attributes (fecundities, death rates, age or stage structure, interaction parameters, dispersal mode, etc.) for the populations involved, and consequently in the mathematical framework applied. Many of the conclusions they yield on vegetation dynamics seem quite robust in the sense that different representatives of the three kinds of model give the same, or at least very similar, answers for persistence/coexistence problems; in some other respects they predict quite different dynamical behaviour. But they support, almost without exception, the general conclusion that the spatial structure of plant populations is a key factor in plant community dynamics.

**Reaction-diffusion models**

When searching for methods to treat the spatial aspects of population interactions theoretically, one reasonable option is to extend the analogy (or, in fact, homology) between the models of chemical kinetics and population dynamics, including diffusion processes in population models. The multitude of reaction-diffusion systems in population dynamics proposed mainly in the past four decades, have come a long way from being simple analogues of physicochemical models of reaction-diffusion kinetics. The main assumptions behind the partial differential-equation formulation are: (i) populations are large enough for stochastic effects not to be taken into account; (ii) individuals are identical in their population dynamical attributes; (iii) the vital attributes and the external variables may be explicitly dependent on spatial position, either directly or indirectly, via the local abundances of the populations; (iv) individuals (or, for sessile organisms, generations of offspring) can move in a diffusive way; (v) ‘reaction’ is represented by local births, deaths and interactions (competitive, mutualistic, predative, etc.).

Reaction-diffusion models have mainly been applied to animal population dynamics (e.g., Kierstead and Slobodkin, 1953; Possingham et al., 1990; Diekmann, 1978; Murray et al., 1986; Yachi et al, 1989; Pech and McIlroy, 1990; Ludwig et al., 1979; Fleming, R.A. et al., 1982; Kareiva, 1983; Kareiva and Shigesada, 1983; Okubo et al., 1989; Andow et al, 1990; Lawton and Godfray, 1990), but since all plants have mobile propagule-states during certain stages of their lives, reaction-diffusion approaches can be (and are) often applied to plants as well. Persistence and coexistence problems are inseparable from the questions of spatial pattern
that emerge in reaction-diffusion systems: passive but density-dependent diffusion has been shown to facilitate the spatial segregation, and thus the regional coexistence, of similar competing species in heterogeneous environments. In environmentally homogeneous, more-than-one-dimensional space, multispecies competitive systems may produce travelling waves of population densities, so that the competitors coexist despite the temporally changing spatial pattern of their abundances. Excellent monographs on reaction-diffusion modelling are Okubo (1980) and Murray (1989).

**Patchy-environment models**

This class of models addresses a very wide range of spatiotemporal scales, from biogeographic to landscape to habitat-patch models, from diurnal to seasonal to evolutional time perspectives. At the scale of plant population and community dynamics, habitat-patch models are the most relevant. These provide a framework within which spatiotemporal dynamical problems are relatively easy to handle, both conceptually and formally.

The common assumptions of patch models are that populations grow and interact in a number of finite topographical regions (islands, habitat islands), separated by a continuous area that is uninhabitable but more or less ‘penetrable’ for them. This area, which is sometimes called the ‘bath’, represents a barrier to dispersal and migration, that is, it reduces the exchange of individuals between patches. The state variables may be either population abundances or the fraction of patches occupied by the populations. If the state variables are population abundances, then the autonomous dynamics on a single patch are usually represented by one of the classical models of closed systems, such as the logistic model or a variant of Lotka-Volterra models; in some studies, the populations are assumed to be age structured. The dynamics within the patches are coupled via dispersal terms, allowing for the flow of individuals directly from any one patch to any other, and/or indirectly through the bath (e.g., Comins and Blatt, 1974; Roff, 1974; Levin, 1974, 1976, 1978; Hamilton and May, 1977; DeAngelis et al., 1979; Hastings, 1982, 1991, 1993; Crowley, 1981; Comins and Noble, 1985; Diekmann et al., 1988; deRoos et al., 1991; Nisbet et al., 1992; McLaughlin and Roughgarden, 1992; for a field case, see Papp, 2003).

Without going into details, some general conclusions of patch models on spatiotemporal dynamics may be summarized as follows: The increased propensity of multipatch systems for persistence and coexistence, as compared with single-patch systems of similar structure, is usually
demonstrated by both single-species and multispecies models. The stabilizing effect may be attributed to the partial isolation of local habitat patches, which may asynchronise local dynamics, producing local source and sink populations. This is true even if the patches are identical environmentally, because the bath itself represents a different environment for the populations, so the whole system is heterogeneous for the environmental variables. It is this heterogeneity that has the stabilizing effect on regional dynamics.

**Metapopulation models**

These are very often confused with patch models, because the basic structure of the environment they assume are similar: the habitat of metapopulations is composed of a set of distinct patches, just like in patch models. However, there is a fundamental difference justifying a separate treatment for these two model types: metapopulation systems use the *fractions of occupied patches* as state variables, instead of local population abundances. The relevant processes are colonizations and extinctions in a set of islands or habitat islands, not population growth or decline on each patch. That is, in metapopulation models there is no explicit formulation for the abundance dynamics of the populations within the patches, nor for dispersal between them. Instead, the rates of colonization and extinction are functions of the fraction of islands already colonized. The patches are not treated explicitly as spatial objects: neither spatial coordinates nor spatial indices make them distinguishable among themselves – hence they belong to the class of spatially implicit models.

Besides the difference in the state variables of patch abundance and patch occupancy models, there are two other serious reasons why to treat them under separate headings. One of these has to do with the difference in assumptions concerning the spatiotemporal pattern of the environment. In patch-abundance models, the environment within the patchwork is usually assumed constant in time, and in many cases also in space. Neither spatial nor temporal homogeneity is inevitable, of course; patches might differ in their local parameters like the carrying capacities, the growth rates and the interaction parameters of the subpopulations, and the dynamics can be driven by temporal changes in the environment. In metapopulation models, the heterogeneity of the environment both in space and in time is inherent in the assumption of the random occurrence of local extinctions, thus it is a core feature of the approach. The other reason - closely tied to the previous one - for the separate treatment of patch abundance and patch occupancy models is the difference in their assumptions on local dynamics. Patch-
occupancy models define non-equilibrium dynamics on the local scale, assuming recurrent local cycles of externally driven extinctions (local catastrophes) and recolonisations on the patches, unlike most patch-abundance models (Levin and Paine, 1974; and Paine and Levin, 1981 are two of the few exceptions), in which local extinction, if it occurs at all in finite time, is a consequence of the internal dynamics (for example, demographic stochasticity) of the system (cf. Verboom et al., 1991; Adler and Neurnberger, 1994).

The metapopulation concept was originally applied for single-species systems (Levins, 1969), but it has been soon extended to multispecies situations, whereby the two categories of presence and absence for a single species on a patch have been replaced with the 2^s categories of possible presence-absence combinations for s species (local biotas – cf. Juhász-Nagy, 1978). The state variables of the resulting metacommunity models are the frequencies of species combinations; also in this case no explicit information on local population abundances are used or available.

The metapopulation approach has proven one of the most successful of theoretical population dynamical methodologies in predicting or explaining real field patterns of abundance, in both animal and plant communities (e.g., Hanski, 1983, 1985, 1986, 1994; Gilpin and Hanski, 1991). The reason for the success might be related to two simple facts. One is that the metapopulation paradigm uses a relatively simple state variable (i.e., patch occupancy) that is easy to observe and record on the field but still comprises the most relevant population events. The other is that it is built on an inherently non-equilibrium dynamics on the local level, unlike most classical approaches. Any actual locality goes through cycles of different occupancy states, ultimately resulting in a constantly moving spatial pattern, even if the regional abundances of the populations involved are relatively static.

**Aggregated interaction models**

Besides reducing the representation of local dynamics, the crucial simplifying assumption of the metapopulation approach as compared to patch abundance models is the omission of the actual spatial allocation of patches, thus making the spatial reference implicit in the model. This modelling strategy works for single species and for a small number of interacting species as well. Under certain constraints on the structure of the model, we can go even further with the simplification of our assumptions regarding the spatial structure of the habitat. Specifically: as far as only pairwise interactions of species are concerned, one can even give up the
assumption of a necessarily patchy habitat structure and consider inhomogeneity in the spatial distribution of the species in a rather general sense, without any direct reference to space. The trick is to assume biotic interactions to be inhomogeneously distributed, so that the probability of interaction between individuals of the two species might be locally disproportional to the product of the actual overall abundances.

The simplest such system considers spatially aggregated interactions of parasitoids with their hosts; it is a modification of the classical Nicholson-Bailey (1935) model. The basic idea of including nonrandom interactions into the Nicholson-Bailey system was proposed a long time ago by Bailey et al. (1962) and Griffiths and Holling (1969). The idea developed into a ramified family of models, assuming some kind of patchy habitat structure first in a “semi-explicit” spatial setting, by the specification of the joint distributions of the interacting populations on the patchwork habitat (e.g. Hassell and May, 1973, 1974). The literature of aggregated host-parasitoid interactions has grown very large since then, leading ultimately to the spatially implicit phenomenological model of May (1978), which accounts for non-random parasitisation regardless of the spatial mechanisms that actually realise it. Shorrocks et al. (1979), Atkinson and Shorrocks (1981) and Ives (1988, 1991, 1992) adapted the same idea to competitive situations. Pacala et al. (1990) review some aspects of host-parasitoid aggregation, and the early monograph by Hassell (1978) and a more recent one by Godfray (1994) present a broad spectrum of both spatial and nonspatial extensions to the classical Nicholson-Bailey model, with field examples and further references.

A less well-known class of single-species models assuming spatial population structure in an implicit manner is that of spatial inhibition processes (Matérn, 1960; Strauss, 1975; Ripley, 1977; Ripley and Kelly, 1977; see Kenkel, 1995, for a review of actual and possible applications to clonal plant growth). Spatial inhibition models are a specific type of stochastic point processes with a relatively small range of biological applicability (see Kenkel, 1993 for examples); I do not treat them in more detail here.

**Interacting particle systems and cellular automata**

The name “interacting particle system” (IPS) comes from the pioneering applications of stochastic discrete-event models in physics (e.g., spin-glass models; Wolfram, 1986) to systems in which the interacting units were spatially fixed particles. The particle-to-site correspondence is perfect in such physical models: it was unnecessary to emphasize that the interacting
objects of cellular automata and IPS’s are in fact spatially defined. In ecological applications, the individual-to-site correspondence is fuzzier; it can change with time, for example, and the spatial scale of an individual and a site need not be the same. The ultimate object of interaction is obviously the site in any ecological IPS, therefore the name of the model class is somewhat misleading in ecological context.

Cellular automata (CA) are a specific type of interacting particle systems. Originally, CA were deterministic, synchronously updated IPS in which interaction neighbourhoods are defined on a topographical basis; in short, neighbours are spatially adjacent sites. Recently this distinction of the concepts of IPS and CA seems to fade – any IPS can be called CA, regardless of the nature of the updating algorithm and the neighbourhood definition.

Population dynamical CA consist of grids of ‘cells’ or ‘sites’, usually in a quadratic or hexagonal arrangement, each representing a small area to be occupied by a small number of individuals (usually one), or a part of a clone (Fig.1.1.a). The neighbourhood of a cell is composed of nearby cells, in one or possibly more concentric zones. The fate of the individuals within each cell is then determined according to the rules of neighbourhood interaction, which yield the next-generation cell-occupancy pattern. The rules may be stochastic in that they define only the probabilities of birth, growth, death, dispersal and interaction events, so that the pattern resulting from a previous state of the grid cannot be unambiguously predicted. In fact, this is usually the case. Such situations are far too complicated to be tractable analytically, so cellular-automaton models are always implemented as stochastic, discrete-state computer simulations (Monte Carlo systems).

There is an inherent similarity between patch models and cellular automata, which should not be missed when discussing the dynamical properties of the latter. Namely, the boundary lines of adjacent cells represent singular places regarding birth, growth, death, dispersal and interaction alike, as do the contour lines of the patches in patch models. The boundaries separate and couple within-cell processes, thus promoting persistence or coexistence. But the patch effect may be artificial in this case, since in most real situations there are normally no such singularities. This effect is minimized by choosing each cell to be a ‘site’ or a ‘microsite’ of only one individual or a part of a clone, with the population-dynamical parameters set accordingly. This means a finer spatial resolution of the grid, which is expected to represent some ‘continuous’ field situations better.
Distance models

Distance models are built directly on the assumptions of sessility, local interactions of individuals (clones) and limited dispersal. Consequently, these provide the best fit for plant population dynamical problems (e.g., Pacala & Silander, 1985, 1987; Pacala, 1986a,b, 1987; Czárán, 1985, 1986a,b, 1987; Czárán & Bartha, 1989; Wu et al., 1985; Walker et al., 1989). Their critical step is the definition of a dynamically reasonable neighbourhood for individual plants, since it is not predefined by artificial grid structures, as in cellular automaton models. Distance models are diverse with respect to this definition. In all cases, however, geometric relations (distances and angular dispersion) of the individuals are important, and have a definite role in the functions describing interactions between them.

The simplest method is to define a neighbourhood as a set of individuals within a given distance from the focal individual, and to assume that each of these exerts an unweighted effect on the mortality and the fecundity of the focal plant (Fig. 1b). The choice of the neighbourhood radius may be the result of some biological considerations. But even if this is so, it is easy to see that this approach inherits one of the artificial assumptions of cellular-automaton models: the perimeter of each neighbourhood area is singular regarding plant-to-plant interactions. In addition, for multispecies situations it is difficult to explain why the neighbourhood radius is the same for any pair of species.

A different approach uses pairwise distances between individuals to define the neighbourhoods. In most pairwise distance models, it is assumed that neighbouring plants affect each other depending on some measure of the overlap between their zones of influence. This zone is a circular segment of space around an individual (clone), the area of which correlates with the size of the plant; the neighbourhood is the set of individuals having zones of influence overlapping that of the focal individual (Fig. 1c). As plants grow, the dynamically effective neighbourhood relations may change in time. Interactions may be mutual or one-sided. For the study of the joint dynamics of many populations, the zones of influence must be dependent not only on plant sizes but also on the ordered pairs of species (since species A may have a different radius of influence on species B than species B has on A). Interaction may affect any of the dynamically relevant parameters of the populations, from seedling mortality to fecundity to dispersal. But the relative versatility of these models is paid for by the loss of analytical tractability: this kind of model is completely computer oriented, using spatially explicit simulation methods.
**Tessellations**

Spatial tessellation is a special way of defining neighbourhoods. Ecological tessellation involves ‘tiling’ the habitat plane according to the positions of individuals, so that each of them has a polygonal area around it. There are many different tessellation algorithms; the best-known type is the Dirichlet tessellation, which determines the polygon as the set of points closer to the focal individual than to any other (Upton and Fingleton, 1985; Okabe et al., 1992). Most modifications are built on this type, and usually include some kind of weighting according to individual size differences within the population. The dynamically effective neighbourhood consists of the individuals with contacting polygons (Fig.1.1.d), and the fate of an individual is determined as a function of the area of the polygon around it. Tessellation models have so far mainly been applied to single-species size-structure problems in plant ecology (e.g., Mead, 1966, 1967, 1971; Fischer & Miles, 1973; Mack & Harper, 1977; Liddle et al., 1982; Watkinson et al., 1983; Mithen et al., 1984; Hutchings & Waite, 1985; Matlack & Harper, 1986; Firbank & Watkinson, 1987; Kenkel, 1988; Kenkel et al., 1989; Owens and Norton, 1989; Aguilera & Lauenroth, 1993), but clearly they could be adapted to multispecies coexistence problems to study vegetation dynamical problems.

Tessellation models are based on implicit or explicit assumptions concerning habitat space capture by individuals through either individual growth or territorial behaviour. If the dynamical assumptions are implicit, the system is a tessellation assignment model, in which a part of the habitat and the resources within are assigned to every individual, the size of each habitat tile depending on the spatial positions, and possibly also on a few biological parameters, of neighbouring individuals. Prediction of the fate (survival, vitality, reproduction) of each individual can be based on the size and shape of the habitat tile it captured, which is assumed to be in direct relation with the share of resources available to the individual. The same approach provides an attractively simple, and also rather general explanation for the -3/2 power law of self-thinning in plant monocultures, one of the most robust relationships in ecology.

Temporally explicit tessellations are applied in ecology as models of territoriality and plant growth. In synchronous territory establishment models the individuals arrive at the same time, and they share the habitat more or less evenly by the random tessellation adjustment process. In asynchronous territory models, the individuals arrive sequentially, and each tries to find a sufficiently large tile among the established territories to
maintain itself and its offspring. Simple plant growth models are derived from the simultaneous isotropic growth (SIG) model (which leads to the Voronoi tessellation of the habitat space) by relaxing either or both the simultaneity and the homogeneity of the speed of growth assumptions of the SIG process. Tessellation models of long-term population dynamics are yet to be constructed.

I.8. The components of spatial effects in CA – mean-field and configuration-field approximations

As might be seen from the foregoing description, the formal structure of IPS models is very simple on the local (site) level. The constraints and the rules of their local behaviour are easy to state, easy to understand, easy to implement on a computer and easy to change, but the consequences of the local structure on the formation of lattice level patterns is almost always impossible to infer by analytical means. This is true in the literal sense regarding deterministic CA, for which analytical approaches are nearly out of question. The updating rules and the initial configuration of a CA tell almost nothing about how the system will behave if left alone for a number of generations, except if it follows very specific or rather uninteresting rules like unconditional extinction (whatever the neighbourhood configuration was, the focal site switches “off”). One exception, a specific updating rule for which it is possible to infer the outcome without numerical simulation, but nevertheless it is interesting, is the so-called parity rule, which is capable of multiplying any initial pattern.

There are rough characterisations for the possible outcomes of some elementary types of deterministic CA (Wolfram, 1986), based mostly on patterns produced by the simplest possible system (one dimensional, deterministic, Boolean, 2-neighbour), but there is no obvious correspondence between the type of the updating rule or that of the initial configuration, and the type of the actual outcome. One has to let the automaton work to see what state configuration it evolves to from a given initial pattern. The global (lattice-level) properties of CA are not transparent from, albeit they are of course determined by, the local rules.

Given the difficulties of almost any analytical approach to deterministic CA, surprising it may be, nevertheless it is true that there exists an increasing arsenal of sensible analytical methods for the study of stochastic IPS. It is also true in general, however, that most theorems on IPS concern oversimplified cases from the viewpoint of ecological applications (but see Durrett and Levin, 1994, and references therein, for
exceptions where the existing body of theory could in fact have been applied to models that were studied numerically by their authors originally). One of the most effective simplifications, frequently leading to some analytical tractability at least, is what physicists call the mean-field approximation. This means ignoring the actual neighbourhood configurations in the next-state function, and assuming the next-state of each cell to depend on the expected (average) neighbourhood (mean-field) state only. This simplification eliminates the spatial information from the system altogether, and it degenerates the model from an object-interaction (i-state configuration) to a mass-interaction (i-state distribution) type: it becomes either nonspatial or spatially implicit. Thus the price we pay for analytical insight is a complete loss of spatial detail, which may or may not affect the qualitative behaviour of the IPS considered.

Besides mean-field approximations, there is another useful method for exploring IPS dynamics - an option in between the numerical simulation of the original IPS system and the analytical discussion of its mean-field approximation. The idea on which it is based is the following: It is always possible in a stochastic cellular automaton to calculate the probability of any imaginable neighbourhood configuration from the frequencies of the cell states at time \( t \) (that is, the state frequency vector \( \mathbf{x}(t) \)), provided that

i) the grid is infinitely large and

ii) the cell states are randomly reallocated every time unit after updating.

Since it is the neighbourhood configuration at time \( t \) that determines the state transition of the focal cell by time \( t + 1 \) according to the updating rule, and the probability of each possible neighbourhood can be calculated, a state transition matrix \( \mathbf{T}(\mathbf{x}) \) can be computed for any state frequency vector \( \mathbf{x}(t) \). Thus the IPS model is reduced to a nonlinear Markov chain. Such approximations might be termed configuration-field systems (Czárán, 1998), a detailed example for which is given in Chapter II.1 below. The main difference between a mean-field and a configuration-field approximation is that the latter preserves more from the assumptions of the original IPS. The effect of neighbourhood size is easy to consider in a configuration-field model; it is possible, for example, to calculate the configuration-field approximation with a Moore- and a Neumann-neighbourhood for the same IPS, and to ask about possible differences in these two cases, which distinction cannot be made in a mean-field model.
On the other hand, a configuration-field system offers the possibility to treat the problem in closed form like a frequency-dependent transition matrix - which does not usually mean analytical tractability at the same time, of course. The temporal trajectories of the state vector have to be calculated numerically for most actual models.

With both the mean-field and the configuration-field approximation available, different layers of the contribution of space to the actual dynamics of the CA system at hand can be explored. The mean-field approximation eliminates all spatial effects, since it is the nonspatial mass-interaction equivalent of the CA itself. However, the configuration-field approximation retains part of the spatial constraints of the original CA. Namely, it retains the local (i.e., within-neighbourhood) manner of interactions while eliminating the effect of limited offspring dispersion. The latter is responsible for the development of mesoscopic spatiotemporal patterns in the CA such as aggregated patches of individuals from a single species, or even more complex ones like rotating spiral waves. That is, comparing the results of the original CA with those of the configuration-field approximation, one can directly see the dynamical effect of mesoscopic pattern formation due to limited dispersion in finite populations, whereas comparing the configuration-field approximation to the mean-field results sheds light on the effect of local interactions within finite neighbourhoods but without the effect of limited dispersal. Fig.I.3. illustrates this point.

Even very simple conjectures on the behaviour of IPS might be very awkward, if at all possible, to prove rigorously if no mean-field or configuration-field approximation is applicable to a problem. This can happen, for example, if the updating rules depend on spatial direction. The analytical treatment of IPS can be very difficult even if an approximation is straightforward, however. For illustrations of the difficulties and the
occasional beauty of rigorous proofs for interacting particle systems, the reader is referred to Liggett (1985). The transition from discrete to continuous time (that is, from synchronous to asynchronous updating) helps somewhat in making certain IPS analytically tractable. This effect is analogous to that of the transition from difference- to differential equations in calculus.
Part Two

Population Dynamics
II.1. Killer-Sensitive coexistence in metapopulations of microorganisms

Introduction: Toxic killing in microbes

In the light of many recent observations and laboratory experiments toxic interactions appear to be common and ecologically important among microorganisms. The excretion of antimicrobial compounds effective against related species or conspecifics is known to be widespread among bacteria (Chao and Levin, 1981; Dykes, 1995; Riley, 1998) and yeasts (Tipper and Bostian, 1984; Starmer et al., 1987; Jacobs and Van Vuuren, 1990; Abranches et al., 1997; Schmitt and Breinig, 2002). In mycelial fungi somewhat comparable phenomena occur as hyphal interference between different species (Berdy, 1974). Toxin excretion in Paramecium has also been described (Grun, 1976), and even multicellular metazoaons like sponges have been shown to produce toxic substances against each other (Thompson et al, 1985). Antagonistic effects resulting from the excretion of biologically active metabolites are known in plants as well (Rice, 1984), but the general ecological importance of this so-called allelopathy is unclear.

For obvious practical reasons the toxins produced by bacteria of medical or industrial importance and their biochemical mechanisms of action are particularly well studied. Toxins excreted by bacteria against similar bacterium species or conspecific strains are called bacteriocins. The best known types of bacteriocin are the colicins from *E.coli* (Pugsley, 1984) and nisins from lactic acid bacteria (James et al. 1991; Vuyst and Vandamme, 1994). Experimental data based on the analysis of strains in *E.coli* collections (Riley and Gordon, 1992) suggest that about 35% of the strains are colicinogenic. Data on human *E.coli* isolates indicate even higher (50%) prevalence of colicin production (Achtman et al., 1983, Riley and Gordon, 1992). The genes coding for the toxin and for the corresponding immunity factor (necessary to avoid suicide in the killer phenotype) are commonly carried by plasmids, both genes located on the same plasmid.

The killer phenotypes of *Saccharomyces cerevisiae* and a few other yeasts have also been studied in depth (Wickner, 1992). The best known killer system in *S. cerevisiae* is controlled by two separate, multi-copy
dsRNA virus-like particles in the cytoplasm (Schmitt & Breinig, 2002). The genetic backgrounds of bacteriocin production and yeast killing are different, but their effects are very similar in ecological terms: the toxin produced eliminates competitor strains from the habitat. Estimates of killer activity among wild yeasts from various habitats suggest that between 5% and 30% of the strains can kill a standard sensitive *Candida glabrata* strain (Starmer et al., 1987; Abranches et al., 1997). However, this is almost certainly an underestimate because assays of toxin production are highly dependent on the choice of the sensitive strains and the appropriate culture conditions.

The elimination of competitors means either literal killing or severe fitness reduction of sensitive strains through blocking one or a few specific, vital biochemical reactions in their metabolic pathways or punching a hole into the cell membrane. Whatever actual form it takes, toxin production is a clear case of *interference competition* (Ganter & Starmer, 1992) - a type of ecological interaction differing from resource competition in that it involves a direct, active suppression of the rival population by means other than just exhausting the common limiting resources. Active interference with the biochemical machinery of the competitor allows the toxin-producing population to take a disproportionally large share of the common resources at the relatively low cost of toxin synthesis and release, i.e., toxin production may ensure a large fitness advantage over sensitive strains. It is possible even without specific „targeting devices” in microorganisms because diffusion is sufficient to mediate the toxic effect to the target cells through the common liquid medium of the habitat. This explains - in evolutionary terms - why toxic interaction is common in microorganisms but not in larger organisms: diffusion cannot deliver a concentrated toxic effect to distances much larger than the body size of a microbe.

Between populations with otherwise similar environmental requirements, the advantage of toxin release may be decisive for the outcome of competition - the toxin-producer might be absolutely dominant over the corresponding sensitive strains in sympatric situations (Adams et al., 1979; Chao and Levin, 1981). In spite of the asymmetry of dominance relations the coexistence of toxin-producing and sensitive populations (and sometimes even the exclusion of toxic strains by sensitive ones) has been observed in different natural and artificial microbial communities on different spatiotemporal scales (Chao and Levin, 1981; Jacobs and Van Vuuren, 1990; Abranches et al., 1997), which facts call for an ecological explanation.
Methods: Metapopulation models for toxic interference

We suggest a spatial non-equilibrium mechanism to explain the coexistence of killer and sensitive strains relying on simple and biologically plausible assumptions. The models we used share the following characteristics:

1. there are two interacting strains, one (the killer strain) dominant over the other (sensitive strain) by means of interference competition;
2. the common habitat of the competing strains consists of many discrete patches, each of which offers enough resources to maintain viable local populations of any one of the strains;
3. the interference competitive effect is in a weak trade-off with population growth rate: a killer strain grows slightly slower than a sensitive strain;
4. the strains are ecologically similar in the sense that, separately placed into the same habitat, they behave alike in terms of local population dynamics (i.e., they follow identical growth trajectories);
5. both strains placed in the same undisturbed local habitat the killer strain always excludes the sensitive one;
6. the habitat patches are ephemeral; each patch (local habitat) goes through cycles of externally driven, random local extinctions and recolonisations from occupied patches;
7. the extinction rates and the dispersal parameters are the same for all populations.

These assumptions were implemented in stochastic cellular automata (SCA) in the first place; with a little modification of the system, we have derived the corresponding configuration-field approximations (Czárán, 1998) to the SCA models.

Stochastic cellular automata

The arena of competition is a 100x100 toroidal lattice of local habitats (cells); time is discrete. The local habitat state transition rules are as follows:
We label the sensitive strain „s” and the killer strain „k”. Three local abundance states are defined for both strains: absent (0), sparse (s and k) and abundant (S and K). That is, a local habitat can be in one of 9 possible abundance combination states (site states) at any point of time: 00, s0, 0k, sk, S0, 0K, Sk, sK, SK. Transitions among these may occur in three ways: 1) colonisations from neighbouring patches, 2) extinctions due to local catastrophes and 3) internal transitions by population growth and competitive interaction. A diagrammatic representation of permissible colonisations and internal state transition routes is shown in Figure 1.

Colonisations: Empty (00) patches can be colonised from within the Moore neighbourhood (8 adjacent patches) centered on them, but only source habitats containing many individuals (that is, sites in any of the states S0, Sk, sK, 0K or SK) send out colonisers. The simple transition rule of colonisation is that Sx sites emit „s” type propagules and xK sites send „k” type ones, in each case with a dispersal probability d onto each of the 8 patches around them. A colonisation attempt can have a dynamically relevant effect only on local habitats where the coloniser type is not yet present: we assume that the propagulum is small, and it does not contribute significantly to the dynamics of a previously inhabited patch. A „k”-type colonisation is always successful on patches with no killers present, but for an „s”-type colonisation to be successful, the recipient patch must be empty. That is, s0 and S0 sites can be overcolonised by the killer, but 0k and 0K sites cannot be invaded by the sensitive - this is one aspect of the dominance of „k” over „s”. Note that the overcolonisation of an S0 site implies structured local habitats: Chao and Levin (1981) show that killers cannot invade an established sensitive population in a liquid culture.
Empty sites can be colonised simultaneously and independently by both types of propagules. For example, if a 00 patch has three source sites: two sK and one SK in its Moore neighbourhood, then the probability that it remains empty is \((1 - d)^4\), which is the likelihood that none of the neighbouring coloniser populations (three K and one S) send propagules there. The chance that this empty patch will be colonised by strain „s“ or strain „k“ alone is \(d(1 - d)^3\) and \((1 - d)[1 - (1 - d)^3]\) respectively; the probability that the site will be colonised by both species so that it turns to state sk is \(d[1 - (1 - d)^3]\). Equal dispersivity for „s“ and „k“ is inherent in the assumption that \(d\) is the same for both strains.

**Local extinctions.** Any patch in a state different from 00 will be reset to the empty state with a probability \(e\), which is the chance of a local catastrophe (disturbance) to occur on a site in a time unit. This parameter is called the extinction probability; it is essentially the same as the extinction parameter \(e\) in metapopulation models.

**Internal transitions.** Patches neither colonised nor disturbed will either stay as they were (this applies to 00, S0 and 0K sites) or they step forward deterministically each to a specified next state. These internal transitions represent the local dynamics on an undisturbed patch: population growth on single-strain patches, and population growth and competitive interaction on patches with both strains present. „s“ and „k“ are both capable of reaching high local abundances in single-strain situations with the same speed (i.e., the transitions s0 → S0 and 0k → 0K require a single time step for both). The trade-off relation between population growth and competitive ability shows up in the internal state transitions of sites inhabited by both „s“ and „k“ (Fig. II.1.2): the presence of „s“ slightly impairs the growth of „k“.

An undisturbed sk patch allows the sensitive population to grow faster initially than the killer (sk → Sk → SK), but later the killer eliminates the
sensitive from the patch through interference competition (SK \rightarrow sK \rightarrow 0K). Note that 0K is a sink state: all the routes of colonisations and internal transitions on Fig.II.1.1 end there, which means that the corresponding undisturbed (equilibrium) system converges to the spatially uniform exclusion of the sensitive population.

The parameters of internal habitat dynamics being wired into the state transition rules, the model has two parameters altogether: the probability of neighbourhood dispersal \( d \) and the chance of local extinction, \( e \). Both range from 0 to 1, so scanning the two-dimensional parameter space - for regions of extinction, competitive exclusion and coexistence - by simulations at an arbitrarily fine resolution is a matter of computing time.

*Configuration-field approximations*

At the price of adopting two additional, slightly restrictive assumptions while keeping all other postulates of the SCA models unchanged, a semi-analytical approximation (a configuration-field approximation; c.f. Czárán, 1998) to the SCA system is straightforward to develop. In principle, it is possible to construct the configuration-field approximations for any single- or multispecies model - we have developed them for the one- and the two-species cases.

To develop the semi-analytical configuration-field approximation for the one- and two-species SCA models we assume (in addition to the postulates of the SCA specified above) that:

i) the number of local habitats is infinitely large and

ii) colonisations can take place from any 8 sites of the infinite lattice, not just from the 8 immediate neighbours of the focal site.

The first assumption excludes the contingencies of stochasticity arising from finite lattice size, whereas the second one filters out the effects of any mesoscale spatial pattern on the dynamics of the system. In this sense, a configuration-field approximation can be used as a reference when determining the actual contribution of the finite-size effect and that of spatial patterning to the qualitative or the quantitative behaviour of the corresponding SCA system.

With i) and ii) in effect, and the relative frequency vector \( \mathbf{X}(t) \) of the patch states at time \( t \) given, the probabilities of all possible colonisation
events at \( t + 1 \) can be determined in closed algebraic form, using rather simple probability calculus, as follows:

**Expected colonisation probabilities:** In the single-strain, \((n+1)\)-stage case, let the state frequency vector be \( \mathbf{X}(t) = \{x_0(t), \ldots, x_n(t)\} \), in which \( x_0(t) \) is the fraction of empty patches, and \( x_n(t) \) is that of the most crowded ones. The expected probability \((1 - C(t))\) that an empty patch in the focus of an 8-patch colonisation neighbourhood remains empty by time \( t+1 \) is the sum of the probabilities that \( i \) of the neighbours are coloniser patches, but all \( i \) fail to establish a new propagulum on the focal patch, i.e.,

\[
1 - C(t) = \sum_{i=0}^{8} \binom{8}{i} p_c^i (1 - p_c)^{(8-i)} (1 - d)^i = (1 - dp_c)^8
\]

where \( d \) is dispersal probability (the same as in the SCA systems) and \( p_c \) is the fraction of coloniser habitat patches (those emitting propagules) among all the patches at time \( t \). In all models we assumed that only the most abundant local populations send out dispersers, that is, \( p_c = x_n(t) \). The actual probability of colonisation on an empty patch is

\[
C(t) = 1 - (1 - dp_c)^8
\]

Note that a change in the colonisation neighbourhood size is simple to implement in the single-strain model: neighbourhood size is the exponent in Eq.2.

In order to derive the analogous colonisation probabilities for the two-strain model, the 9 different patch states should be grouped into 4 coloniser types, in accordance with the colonisation rules specified earlier:

1) 00, s0, 0k, sk patches are non-colonisers;
2) S0 and Sk patches are „s”-colonisers;
3) 0K and sK patches are „k”-colonisers;
4) SK patches are „s-k” colonisers.

Denoting the relative frequency vector of the site states at time \( t \) by
\( X(t) = \{ x_{00}(t), x_{s0}(t), x_{0k}(t), x_{sk}(t), x_{S0}(t), x_{0K}(t), x_{SK}(t), x_{SK}(t) \} \),

the fractions of patches in the four coloniser types at time \( t \) are

- **non-colonisers:** \( p(t) = x_{00}(t) + x_{s0}(t) + x_{0k}(t) + x_{sk}(t) \);
- **"s"-colonisers:** \( q(t) = x_{S0}(t) + x_{Sk}(t) \);
- **"k"-colonisers:** \( r(t) = x_{0K}(t) + x_{sK}(t) \);
- **"s-k" colonisers:** \( s(t) = x_{SK}(t) \);

where \( p(t) + q(t) + r(t) + s(t) = 1 \). We use the notations \( p(t) = p, q(t) = q, r(t) = r, s(t) = s \) below for brevity. The probability \( P_{j_0, j_s, j_k, j_{sk}}(t) \) that a neighbourhood assembled from 8 patches randomly chosen from the infinite lattice consists of exactly \( j_0 \) non-colonisers, \( j_s \) "s"-colonisers, \( j_k \) "k"-colonisers and \( j_{sk} \) "s-k" colonisers is the corresponding term of the multinomial distribution:

\[
P_{j_0, j_s, j_k, j_{sk}}(t) = \frac{8!}{j_0! j_s! j_k! j_{sk}!} p^{j_0} q^{j_s} r^{j_k} s^{j_{sk}},
\]

with \( j_0 + j_s + j_k + j_{sk} = 8 \). Given the \( (j_0, j_s, j_k, j_{sk}) \) neighbourhood configuration, the probabilities of no colonisation at all \( (Q^0_{j_0, j_s, j_k, j_{sk}}) \), colonisation by "s" only \( (Q^s_{j_0, j_s, j_k, j_{sk}}) \), by "k" only \( (Q^k_{j_0, j_s, j_k, j_{sk}}) \) and simultaneous colonisation by both species \( (Q^{sk}_{j_0, j_s, j_k, j_{sk}}) \) on the focal site can be calculated as

\[
\begin{align*}
Q^0_{j_0, j_s, j_k, j_{sk}} &= (1 - d)^{j_0 + j_s + 2 j_{sk}} \\
Q^s_{j_0, j_s, j_k, j_{sk}} &= \left[1 - (1 - d)^{j_s + j_{sk}}\right]\left[1 - d\right]^{j_k + j_{sk}} \\
Q^k_{j_0, j_s, j_k, j_{sk}} &= \left[1 - d\right]^{j_s + j_{sk}}\left[1 - (1 - d)^{j_k + j_{sk}}\right] \\
Q^{sk}_{j_0, j_s, j_k, j_{sk}} &= \left[1 - (1 - d)^{j_s + j_{sk}}\right]\left[1 - (1 - d)^{j_k + j_{sk}}\right].
\end{align*}
\]

Thus the total colonisation probabilities of empty sites are
\[ C_0(t) = \sum_{j_0+j_s+j_k+j_{sk}=8} Q_{j_0,j_s,j_k,j_{sk}}^0 \cdot P_{j_0,j_s,j_k,j_{sk}}(t) \]
\[ C_s(t) = \sum_{j_0+j_s+j_k+j_{sk}=8} Q_{j_0,j_s,j_k,j_{sk}}^s \cdot P_{j_0,j_s,j_k,j_{sk}}(t) \]
\[ C_k(t) = \sum_{j_0+j_s+j_k+j_{sk}=8} Q_{j_0,j_s,j_k,j_{sk}}^k \cdot P_{j_0,j_s,j_k,j_{sk}}(t) \]
\[ C_{sk}(t) = \sum_{j_0+j_s+j_k+j_{sk}=8} Q_{j_0,j_s,j_k,j_{sk}}^{sk} \cdot P_{j_0,j_s,j_k,j_{sk}}(t) \] (5)

\( C_0(t), C_s(t), C_k(t) \) and \( C_{sk}(t) \) are the expected probabilities of no colonisation, colonisation by „s”, by „k”, and by both, respectively, on a randomly chosen empty site with a random neighbourhood configuration around it (\( C_0(t) + C_s(t) + C_k(t) + C_{sk}(t) = 1 \)). These colonisation terms cannot be composed from the simple single-strain form \( C(t) \), because the colonisation neighbourhood configurations are mutually constrained in both strains by the presence of the other: the colonisation events are interdependent. The smaller the colonisation neighbourhood the stronger this interdependence.

The probability that an s0 or an S0 type local habitat does not go extinct and it gets at least one „k” propagule (i.e., the chance that an s0 or S0 patch turns to state Sk for time \( t + 1 \) through an overcolonisation event) is \( A = (1 - e)(C_k(t) + C_{sk}(t)) \). The probability that an s0 or S0 patch survives, but it does not get overcolonised by „k” (i.e., the chance that it turns to state S0 by time \( t + 1 \)) is \( B = (1 - e)(1 - C_k(t) - C_{sk}(t)) \).

Nonlinear transition matrices: Once the colonisation (and overcolonisation) probabilities are determined, they can be plugged into a nonlinear state transition matrix, in which the colonisation terms are dependent on the actual state vector \( X(t) \). The single-strain transition matrix for a 3-stage model is

\[
M[X(t)] = \begin{pmatrix}
1-C & e & e \\
C & 0 & 0 \\
0 & 1-e & 1-e
\end{pmatrix}, \tag{6}
\]

where \( X(t) = \{x_0(t), x_1(t), x_2(t)\} \) is the state vector at time \( t \), and \( C = C[X(t)] \) is the actual chance of colonisation as given by Eq. 2: \( C = 1 - (1 - d \cdot x_2(t))^8 \).
The two-strain transition matrix - based on the state transition graph of Fig.II.1.1 - is given in Table 1.

\[
M[X(t)] = \begin{pmatrix}
C_0 & e & e & e & e & e & e & e \\
C_e & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C_k & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C_{sk} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & B & 0 & 0 & B & 0 & 0 & 0 \\
0 & 0 & e & 0 & 0 & 1 & e & 0 \\
0 & A & 0 & e & A & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & e & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

Table II.1.1. The state transition probability matrix \(M[X(t)]\) for the configuration-field approximation of the two-species model (see text for the specification of nonlinear elements \(A, B\) and \(C\)).

Equilibria: In both one- and two-strain models, one step of the state transition is represented by the nonlinear system of difference equations

\[
X(t+1) = M[X(t)] \cdot X(t) .
\]

(7)

The parameter space of the configuration-field models is the same as that of the SCA systems: it is spanned by \(e\) and \(d\), the probabilities of local extinction and propagule dispersion, both ranging from 0 to 1.

For the single-strain model, the feasible fixed points of Eq.7 can be numerically found from the equilibrium condition

\[
\dot{X} = M[\dot{X}] \cdot X
\]

(8)

at any specified point of the parameter space. We used Mathematica for Windows to locate the fixed points. Numerical results indicate that the single-strain system can
have either one or two fixed points on the $n$-dimensional standard simplex: the trivial equilibrium $\hat{X}_1 = \{1,0,...,0\}$ (only empty sites: extinction) and possibly another one with $0 \leq \hat{x}_0 < 1$ (persistence). At parameter values admitting two fixed points, the trivial equilibrium is always repelling, whilst the other is stable; the trivial equilibrium is an attractor from within the positive orthant otherwise.

Fig.II.1.3 shows the persistence- and the extinction-domains of the parameter space for the 3, 4 and 5-stage single-strain systems.

In principle at least, the equilibria of the two-strain model can also be calculated as numerical solutions of the corresponding fixed point equations (analogues of Eq.8), but in fact this is impracticable - the dimensionality of the system is too high, calculations across the parameter space would last for weeks even with the fastest computers. The only useful method left is to iterate Eq.7 until the state frequency vector $X(t)$ remains numerically stationary.

**Results: Persistence and coexistence in time and space**

The SCA has been run in 4 replicates with each parameter combination, each run lasting for 10,000 generations. Systematic simulations across the e - d parameter space reveal that, at intermediate local extinction probabilities ($0.2 \leq e \leq 0.8$) and sufficient colonisation ($d \geq 0.2$), the sensitive population can co-exist with the killer and, at high enough extinction rates ($0.6 \leq e \leq 0.8$), the sensitive type even takes over, excluding the killer from the habitat patchwork completely (Fig. II.1.4a).
The part of the parameter space where the sensitive survives extends to more than half of the corresponding domain of single-strain persistence (cf. Fig. II.1.3), even though the only disadvantage of the killer type is that its population grows a little slower in the presence of the sensitive; the solitary growth rates and the dispersal abilities of the two types are the same, and the killer always wins on undisturbed patches.

What roles the possibly emerging mesoscale patterns and the finite size of the habitat universe play in the coexistence of the two populations can be judged from a comparison of the SCA results with those of the configuration-field approximation (Fig.4a and b). The qualitative behaviour of the configuration-field model turns out to be quite similar to that of the SCA in that the same types of persistence, coexistence and extinction domains obtain in its parameter space. The results are different in quantitative terms, however: the parameter domain of coexistence is much larger in the configuration-field model, at the expense of each persistence domain and the extinction domain. The explanation involves both the finite-size effect and the pattern effect. First: on the borderline of two parameter domains, one of the populations is always sparse at equilibrium, and that one might go extinct by chance in the SCA model - the smaller the lattice, the easier it is to lose the last cell occupied by the sparse type. As the state of extinction is a sink (apart from the initial input, no external introduction of propagules is allowed into the habitat patchwork), the coexistence domain is diminished from all sides by the finite size effect. For the same reason, the extinction domain expands at the expense of the persistence domains. Second: since the neighbourhood relations of the habitat patches are frozen in the SCA, short-range dispersion can produce small regions of monodominant (single-type) persistence at certain parameter combinations. In such a single-type habitat aggregation, most of the propagules emitted by a patch within the clump are wasted, probably landing on local habitats already occupied by the same type. For this reason, the effective rate of colonisation is considerably smaller from clumped patches than it would be in a non-aggregated situation (like the configuration-field approximation), which fact explains why the SCA system goes extinct close to the outer margins of the persistence domains (i.e., close to the extinction domain) of the configuration-field model. The same effect works against the invasion of the sensitive population by clumps of the killer in the SCA, resulting in a larger persistence domain for the sensitive type at the expense of the coexistence domain. That is, the mesoscale pattern emerging in the SCA model at certain parameter combinations seems to help the sensitive strain to get rid of the killer.
Fig. II.1.5 compares the time series of the SCA and the matrix model in this part of the parameter space.

In all, the pattern effect decreases

1) the coexistence domain at boundaries common with the persistence domains; and
2) both persistence domains in favour of the extinction domain.

With lattices of the size used in the simulations, the pattern effect proves to be much more pronounced than the finite-size effect: decreasing the size of the lattice to a quarter of the original (from 100x100 to 50x50) had no qualitative effect on the SCA results at all – the persistence-coexistence map of the parameter space for the smaller system is the same as that of the larger one in Fig. II.1.4.a.

Discussion

In relation to the theoretical problem of the coexistence of interference competitors with a dominance hierarchy in equilibrium situations, we draw the following general conclusions:

1. The regional coexistence of ecologically similar, but locally exclusive interference competitors is possible in a surprisingly large part of the biologically feasible range of dispersal and local extinction parameters, provided that the dynamics of the system is non-equilibrium on the local scale. A low probability of the occurrence of local, asynchronous disturbance events is sufficient to maintain coexistence.

2. Unlike in metapopulation models and the corresponding mean-field approximations published earlier (Caswell and Cohen, 1991; Caswell and Etter, 1993), the coexistence of the dominant (killer) and the subordinated (sensitive) types is not dependent
on differences in dispersal abilities: in fact the probability of
neighbourhood dispersion was the same for all types in all the
models studied. The only feature favouring the subordinated
type was a slight trade-off between local population growth rate
and competitive ability.

3. Even if the trade-off relation is very weak, the subordinated
strain enjoys the full advantage of local disturbances: the model
assumes that the growth rate of the sensitive population is a little
larger than that of the killer only if they co-occur on a habitat
patch, otherwise the growth rates are equal.

4. Comparison of the SCA results to those of the configuration-
field approximations shows that, at moderately large lattice
sizes, there is no considerable finite-size effect in the long run.

5. The effect of mesoscale pattern formation can change the results
in a quantitative sense: in the SCA models, at intermediate
disturbance probabilities, the spatial aggregation of local habitat
states due to short-range dispersal helps the subordinated
population to outcompete the dominant regionally.

These conclusions are fairly general in the sense that they apply to
any competitive metapopulation system obeying the same rules of local
interactions and colonisation. One specific field of its application is the
ecological problem of the widespread coexistence of killer and sensitive
strains in microorganisms. Regular, asynchronous local habitat destruction
could well be common in microbial communities, whether the local
habitats are host animals or plants, fallen fruits or cactus rots, to mention
just a few. All such local habitats or habitat patches are temporary, capable
of maintaining small local communities for a certain period of time, then
disappearing and re-emerging again. This implies nonequilibrium dynamics
on the local scale, interrupted by local extinctions, reset by local
colonisations - and hence metapopulation dynamics on the regional scale.

Note, however, that the models we applied here are based on the
condition that there are only killer and sensitive strains. This seems mostly
ture for killer yeasts usually defending themselves against their own toxin
by producing an immunity factor, which is a slightly modified and
intracellularly produced toxin molecule. The toxin and the immunity factor
are basically coded by the same genes and the metabolic pathways
producing them are almost completely identical (Wickner, 1992). It is not
always true in bacteriocin systems, however, which often use resistance
factors to prevent suicide. Resistance is due to mutant membrane-bound transport proteins not accepting the toxin as a substrate. Thus the toxin and the resistance „factor” are coded by different genes - although often located on the same plasmid - which allows decoupling of toxin production and resistance by mutations. Of course a mutant having lost the functionality of the immunity factor gene alone is not viable: such a mutant commits suicide by producing a toxin that it is not resistant to. Losing the toxin gene while retaining resistance may be a viable mutation, however: the result is a resistant strain which does not itself produce the toxin. Assuming a metabolic cost to toxin production, the resistant form may be more efficient in resource competition than the killer phenotype. The possibility of the rise of a resistant mutant may result in qualitatively different dynamics, because once the resistant population takes over, sensitive strains also get a chance to spread: they do not even pay the cost of resistance and may thus be superior to the resistant mutant in resource competition. This means that a killer-resistant-sensitive system may constitute a circle of competitive dominance: killers are dominant over sensitives by interference competition (toxic killing), but resistsants take over the killer strains and sensitives take over the resistsants through resource competitive exclusion. This offers a completely different mechanism for maintaining the coexistence of genetically different strains (Durrett and Levin, 1998; Pagie and Hogeweg, 1999; Frean and Abraham, 2001; Kerr et al., 2002), which we have explored elsewhere in more detail (Czárán et al., 2002). Unlike in the metapopulation system we discussed here, coexistence in the intransitive competition cycle critically depends on localized dispersion.
II.2. Spatial constraints decrease the invasion rate of fitter mutants in asexual populations

Introduction

In search for mechanisms explaining the emergence and maintenance of biodiversity, many ecologists and evolutionary biologists have been interested in the role of spatial structure on adaptation. In a number of ways spatial structure can contribute to the emergence of adaptive radiation, by maintaining the coexistence of competing genotypes. Spatially structured habitats are often heterogeneous environmentally, providing additional niches which can lead to the diversification of a resident population (Rainey and Travisano, 1998; Korona, 1994). Spatial structure can also contribute to genetical diversification by dividing large populations into many small subpopulations. Once these become isolated, they can evolve independently one from the other. Moreover, due to limited offspring dispersion and localized interactions among neighbouring individuals (i.e., due to the spatial constraints of population dynamics), beneficial mutants or invading, ecologically superior genotypes may coexist with the resident population much longer in most natural habitats than in a well-mixed fluid medium, thus maintaining genetic diversity on a longer time-scale. It is this effect of spatial constraints on the invasion of a beneficial mutant that we investigate in the present study.

Besides generating more diversity, spatial structure can influence adaptation in different ways. Previously, we have reported a slowing down of the rate of adaptation in spatially structured environments with a rigid population structure. One single genotype of E.coli B was serially transferred for 900 generations in three different environments; a well-mixed homogeneous environment and two spatially structured environments that differed in only one aspect; in one treatment the population structure was kept intact; in the other treatment the population was mixed before transfer to new medium. Relative fitness of the latter populations proved to be significantly higher after 300 and after 900 generations of evolution than fitness of the populations that had evolved with an intact population structure. Since both treatments only differed in this one aspect – mixing – keeping structure intact apparently had lead to a slowing down of adaptation.

The adaptation of asexual species is believed to consist of sequential substitutions of beneficial mutations. When a beneficial mutation arises in a
genetically homogeneous, well-mixed population, it is expected to sweep through the population following an exponential dynamics until it is fixed. After fixation the population waits, in a genetically static phase, for the next beneficial mutation to arise and sweep through. We hypothesize that with spatial constraints on population dynamics (i.e., with limited dispersal and localized competition) the invasion of a new beneficial mutant may be much slower. The reason for this expectation is simple: Limited dispersal creates compact colonies of invaders surrounded by a continuous colony of the ancestral strain. Cells of *E. coli* grown on agar plates indeed form colonies with a radial growth that is at least an order of magnitude faster than the vertical growth rate at the centre of the colony (Grimson and Barker, 1991). Within the invading patch, most competition takes place among mutant clonemates, because interactions are localized and thus affect only very close neighbours. It is only on the perimeter of the mutant colony where the invader can take advantage of its competitive superiority and displace the ancestral strain. The radius of the invading patch increases at a constant rate, which means a quadratic increase in time for the number of mutant cells. Comparing this to the – faster – exponential increase of mutant frequency in a well-mixed fluid medium, we might say that instead of *sweeping*, beneficial mutations will *creep* to fixation under spatial constraints.

Below we attempt to show both experimentally and theoretically that the sweep-creep difference produces conspicuous divergences in fixation dynamics, the proximate reason for which is the decline of effective mutant fitness due to spatial constraints on population interactions.

**Materials & methods**

*Bacteria, media and experimental design*

In a previous experiment, 36 populations derived from the *E. coli* B strains REL 606 or REL 607, were propagated for 900 generations in either a homogeneous environment or a spatially structured environment (??). Of each population, three clones were isolated and analyzed. In this experiment we used two clones of two independent populations – populations that had evolved in 1/10 LB in a spatially structured environment differing only in the method of transfer; in one case, the population was replicated by velvet –leaving the population structure intact–, in the other case, the population was mixed before each transfer. Both clones have a fitness advantage of about 50% relative to their ancestral clone REL 607. The ancestral clone(s) have been used frequently
in serial transfer experiments and have been described previously (Lenski, 1991).

The evolved strains were competed against the ancestor for five days in an environment identical to the one they had adapted to: Petri plates (diameter 60 mm) of 10 ml’s of 1/10 LB agar. Prior to competition, the competitors were preconditioned to the same environment they would compete in, to make sure they were both in the same physiological state. At the beginning of the competition the ancestor was spread on the agar surface, while the evolved clone (the superior competitor) was introduced in three different ways: in the first treatment (A), the superior competitor was introduced in one spot in the middle of the plate (1µl); in the second treatment (B) the cells were introduced in 20 different spots (1 µl each), and in the third and fourth treatment (C and M) we mixed both competitors before spreading them evenly on the plate. The ratio and absolute number of the inferior versus superior competitor at the start of the competition was the same for all treatments: 500:1. Because the evolved clone is introduced in low numbers, the cell density on the plate is of the same magnitude everywhere. Every day, the populations were transferred to new medium; in the first three treatments (A, B, and C) the populations of both competitors were replica plated every day by velvet; in the fourth treatment (M) cells were scraped off the plate and diluted and mixed in 10 ml’s of saline before transfer, similar to the way we transferred the mixed plate treatment in the long-term evolution experiment. In all treatments the populations were diluted 200-fold before transfer, which means a transfer of 2.5 *10^6 cells per ml. Carrying capacity of the medium is 5*10^8 cells per ml, which means populations grew 8-fold before the next transfer. At day 1, 2, 3 and 5 of the competition, the relative frequency (and fitness) of the superior competitor was measured by plating a dilution of the population on indicator plates. Method to measure the fitness has been described before (Lenski, 1991).

Two evolved clones were used in order to exclude the possibility that any of the clones had evolved specifically to the way of transfer, which would add a difference in results between mixing and replica plating. Both clones were competed in all four different ways; every treatment was replicated 6 fold.

Model
We used a simple stochastic cellular automaton (CA) model to imitate the 4 different treatments of the experiment: 1 spot (A), 20 spots (B), fixed random initial pattern (C) and mixed everyday (M). The CA space is a 300
x 300 square grid of sites representing the surface of the agar medium in a Petri dish. Each site is assumed to be capable of harbouring a single bacterium cell. Of the 90,000 sites, 10% were inoculated by bacteria at a mutant to ancestor ratio of 1:500 at time 0. The ancestral strain was dispersed on the plate at random. For treatment A the mutant cells were clumped into a single spot in the middle of the plate; in B they were grouped into 20 spots and the spots spaced out evenly on the plate; in C and M the mutants were dispersed over the plate at random. A generation consists of 90,000 independent, random updating steps, so that each site is updated once on average every generation. An updating step starts with the random choice of a focal site and one of its neighbouring sites. If the focal site contains a bacterium and the neighbouring one is empty, then the bacterium may put a copy of itself into the empty site with a probability equal to its basic fitness $f_i$, a parameter of the strain the focal cell belongs to. We used $f_{\text{ancestor}} = 0.67$ and $f_{\text{mutant}} = 1.00$ to maintain a relative fitness of 1.5 for the mutant. Every 8th generation a 5% sample of the bacteria present on the plate are transferred to a new plate. In treatments A, B and C the bacteria in the sample keep their previous site on the new plate; in treatment M the sample is reshuffled and dispersed on the new plate at random. Using this updating algorithm we have recorded the relative frequency and the (effective) relative fitness of the mutant strain over 500 generations.

Note that the number of simulated cells is about 5 orders of magnitude smaller than that in the Petri dish experiments described above, because the fates of astronomical numbers of cells are impossible to follow even with the most powerful of recent computers. Therefore a quantitatively correct simulation of the dynamics on a Petri dish is not feasible; what we expect from the simulation model is a qualitative fit to the experimental data.

**Results**

We investigated the rate of invasion of a superior competitor in a spatially structured environment, when mixed to a different degree with a dispersed competitor. The rate of invasion of the superior competitor enhances significantly with an increase in the mixing of the two competitors (repeated measure; df = 4; $F = 17.14$; $P \leq 0.001$) as is seen in Figure II.2.1. When the superior competitor is introduced in one spot, the increase of the relative frequency over time is minimal. This is predicted by our hypothesis; the superior competitor will be more in competition with itself then with the other competitor and the relative fitness of the superior competitor is lower.
competitor will therefore be lower. The hypothesis assumes that growth rates in a spatially structured environment are dependent on the growth rate of neighboring cells. We have shown indirectly that the growth rate of the superior competitor is indeed slower in treatment A and B when compared to treatment C and M. This was done by repeating the experiment, but this time without the ancestor. The superior clone R was introduced in the same way as done before, and all treatments were replicated 3-fold. In treatment B, C and M, the cells reached about carrying capacity after 24 hours; there was however a slight but significant difference between treatment B and M at \( t=1 \) until \( t=3 \). Treatment A had at all time points a significantly lower amount of cells compared to any other treatment (ANOVA, Tukey HSD).

Initial density does not influence the competition. When the ancestor is in competition with itself instead of with the superior clone, none of the treatments shows an increase of any one of the competitors, nor is there a treatment effect (ANOVA: \( \text{df}=3; \ F=0.173; \ P=0.912 \)).

**Fitness:** In evolution experiments with microorganisms, the relative fitness in competition is measured by the ratio of the different growth rates of the two competitors. This way of measuring fitness assumes an independent growth rate. On plates, however, the invasion of a genotype is dependent on the local conditions, as the results show. We assume this is caused by a difference in growth rate of the superior competitor. The growth rate probably depends on local competitors, because the diffusion rate of nutrients in agar is relatively slow. Based on the results of the experiment, we expect the fitness of a superior competitor to decline with increasing invasion. We tested this hypothesis by measuring the fitness of the superior competitor in the invasion experiment. Relative fitness was calculated at 3 different time points (cf. Figure II.2.2).

Results suggest that the relative fitness of the clone is influenced in two ways; first the fitness is significantly different when the initial distribution of the superior competitor is different; the more mixed the
competitors are, the higher the fitness of the superior competitor. (ANOVA on relative fitness at time t=5 for the different treatments: df= 3; F=147.57; p< 0.001). Second, when the competitors are not evenly mixed every day, there is a significant fitness difference over time (ANOVA on both clones; treatment effect: A: p=0.023; B: p< 0.001; C: p=0.004; M: p=0.386). The fitness of the clones is significantly different from each other in treatment B and C.

This shows that fitness indeed declines when the competitor increases in frequency, leading to a higher competition between superior competitors locally, reducing their relative growth rate. The more the competitor is mixed, the less the difference in fitness over time. (ANOVA on average difference between t=5 and t=2 of the different treatment: p=0.011).

However, in this scenario one would expect treatment A (superior competitor introduced in one spot) to have the most significant decline in fitness over time, which is not the case, as Fig.II.2.3. shows beyond doubt. This is probably due to
the fact that population growth in this treatment is already very slow even at \( t = 2 \), i.e. there is no room for spectacular fitness decrease later.

**Model results**

Simulations of the four treatments (A, B, C and M) with the CA model produced the time series of relative mutant frequencies shown on Fig.II.2.4.

![Figure II.2.4. Simulated relative frequencies of the superior competitor over time. Left panel: simulation for 500 generations; Right panel: the first 100 generations of the same run of the simulation model. Parameters: Basic fitnesses: \( f_{ancestors} = 0.67, f_{mutants} = 1.00 \); Initial ancestor to mutant frequency ratio: 500:1; Transfer dilution: 20-fold; Transfer frequency: every 8th generation.](image)

It is obvious from the simulations that the mutant strain takes over sooner or later. The time to exclusion of the ancestral strain depends on the spatial mixing treatment alone, roughly the same way as in the experiments. Intensive mixing (treatment M) gives the largest advantage to the mutant, and a rigid spatial population structure with a single invasion centre (treatment A) delays the exclusion process considerably. Keeping the rigid spatial structure but increasing the number of invasion centres (in treatments B and C) gives the mutant sufficient competitive advantage, so that a relatively short time is enough for it to complete the takeover. It is interesting that even the fixed random initial pattern (treatment C), which supplies many invasion centres, leads to significantly slower exclusion than regular mixing (treatment M). This is also what we observed in the experiments (even though the exclusion-time difference was smaller there than it is in the simulations). The explanation for the difference lies in the development of pure mutant colonies in the undisturbed habitat of treatment C, which are regularly mixed with ancestral cells in treatment M. Because of the fixed proximity of parent to daughter cells in the undisturbed habitat, ancestral cells tend to be surrounded by ancestral cells and mutants are the neighbours of other mutants in most cases. That is, the majority of competitive interactions on the plate take place within-strain,
and only a small fraction of competitions are inter-strain (between ancestors and mutants). This implies a relative decrease in the realized competitive advantage of the mutant, because it is only a minor share of its competitive interactions that leads to a (positive) change in its relative frequency. This is what we can observe in Fig.II.2.5. The effective (realized) fitness of the mutant significantly decreases with time in all but the mixed case (treatment M), since increasing within-strain aggregation prevents the bulk of mutants getting in contact with the inferior competitor ancestral population, and this constrains mutant reproduction.

**Discussion**

In a homogeneous environment, the difference in growth rate between the ancestral strain and the new beneficial mutation is absolute. When grown in a spatially structured environment, however, the growth rate of any genotype will depend on the growth rate of neighboring genotypes, since there is only a limited amount of nutrients in each spot. Moreover, space also counts as a limiting factor on agar plates, just as it is in the case of any sedentary organism like higher plants or benthic animals. The individual capturing a site earlier wins all the resources associated to the site, excluding all other individuals from there for its lifetime (spatial pre-emption effect). Without dispersal, a mutant with a higher fitness will not be able to exploit its advantage to the full, because it is mostly in competition with itself. The rate of invasion will consequently be reduced if the dominant competitor is constrained spatially, since the increase in that subpopulation will be slower when compared to a well-mixed environment. Due to slower invasions, we expect not only the ecology, but also the evolution of the community to be influenced: since inferior competitors are present in the community for longer time, they have more chance to obtain new mutations, some of which might be beneficial. In short, spatially constraints maintain more genetic variation, which might help the population to adapt to changing environments. We have shown previously that in populations which had evolved in a spatially structured
environment with a rigid population structure, more genotypes coexisted than either in a liquid culture or in a structured but regularly reshuffled environment. Thus we arrive at the – somewhat paradoxical – conjecture that, by slowing down the exclusion dynamics between competing strains, spatial constraints are expected to speed up evolutionary changes in bacterial populations. This idea requires empirical proof yet, however.
Part Three

Microbial Evolution
III.1. Evolution of sexual asymmetry

Introduction

One of the most general rules in biology seems to be that sex involves the fusion of gametes (sometimes of other specialised structures) of different type. In most taxa this sexual asymmetry is reflected in the male/female distinction between mating partners and/or between mating sex cells. This paper aims to help understand why sex is asymmetric.

The primary difference between male and female is anisogamy, the differential size and mobility of gametes. Anisogamy is thought to have evolved from a more primitive condition of isogamy (for reviews see Hoekstra, 1987; Randerson & Hurst, 2001; see also Bulmer & Parker, 2002). In isogamous species without apparent male-female differentiation, like unicellular green algae (e.g. *Chlamydomonas*) and fungi (e.g. yeast), the asymmetry in sexual fusion and subsequent development are regulated by a binary mating type system. Mating is only possible between cells of different mating type. Molecular analysis has revealed a remarkable and complex genetic mating type structure (Herskowitz, 1988; Ferris et al., 2002). The two mating types in a species consist of so-called idiomorphs (Glass et al., 1990), non-homologous complexes of closely linked genes that occupy homologous positions at the same chromosomal locus. They behave as alleles in being mutually exclusive in meiotic segregation. A similar binary mating type system exists in many filamentous ascomycetous fungi (Coppin et al., 1997), which however often also exhibit male/female differentiation. Only matings between individuals of different mating type are allowed. Thus in mycelia that can function both as male and as female self-mating is prevented. Mating in such species is heterothallic, that is, always between different individuals. However, many ascomycetes are homothallic, i.e. can complete the sexual cycle in a single individual. Homothallic species may lack mating types, such as *Aspergillus nidulans*, or may consist of individuals that are heterokaryotic for mating type (carry nuclei of both mating types) such as *Podospora anserina*. In the latter case sexual fusion is between different mating types at the nuclear level, but can occur within a single individual mycelium.

In basidiomycetous fungi, morphological sexual differentiation is absent, but mating is regulated by complex mating systems, generating in some cases large numbers of different mating types. Also here, the mating type genes control sexual fusion and post-fusion development (Casselton,
2002). Again, mating cannot occur between individuals of the same mating type.

In other taxa still other genetic systems exist that control sexual fusion, sometimes in addition to the male-female difference. In monoecious higher plants often self-incompatibility systems occur that effectively exclude self-mating (Silva & Goring, 2001; Nasrallah, 2002). Among ciliates, several variations on the theme of mating type differentiation exist, which are not further detailed here.

All these different mating systems have one characteristic in common: mating is always asymmetric. When gender differences exist, mating involves the fusion of a male and a female cell; this may occur when the male and female functions are in different individuals, or when a single individual possesses both male and female functions. When gender differentiation is absent, mating type systems guarantee that sexual fusions are between different types. However, the absence of both gender and mating type differentiation has never been observed. This would imply symmetric sexual fusion: a species in which every sex cell could potentially fuse with any other sex cell. Because gender differences starting with anisogamy most likely evolved from pre-existing isogamy, we should consider the evolution of mating types in an isogamous species to understand why sex is asymmetric.

Functional explanations of the evolution of a binary mating type system have been explored in theoretical models by Hoekstra (1982), Hurst & Hamilton (1992), Hutson & Law (1993) and Hurst (1995). These models differ in their biological assumptions. According to Hurst & Hamilton (1992) and Hutson & Law (1993), mating types have evolved to suppress harmful conflicts between cytoplasmic elements, while Hoekstra (1982) suggests that mating type loci have evolved in response to polymorphisms for genes involved in gamete recognition. It is still not possible to conclusively decide between the alternative biological scenario’s (Charlesworth, 1994). However, all models envisage as a starting point an initially undifferentiated population in which every gamete can mate with any other gamete, and derive conditions for the evolution of two mating types that exclusively mate with each other and have lost the ability to mate with their own type. A general conclusion emerging from the models is that mating types may invade the initially undifferentiated population under fairly broad conditions, but that the removal of the undifferentiated type requires very strong selective forces. It is this latter aspect which in our view still forms a problem, because it is difficult to see why the original
type should be so disadvantageous compared to the differentiated mating types. The mentioned models assume a homogeneous population in which random encounters lead to mating. However, this assumption is likely to be very unrealistic if vegetative reproduction is much more frequent than sexual reproduction, like it is in present-day protists, and if the mobility of the cells is low. Since the motion of cells or gametes in water is characterized by a Reynolds number (the ratio of the inertial forces to the viscous forces) smaller than one (Purcell, 1977), clonally related cells will tend to remain in each others vicinity, and therefore a clonal distribution of cells and gametes is expected, rather than a well-mixed homogeneous population. This implies that mating types will have a smaller chance of finding a suitable mating partner than in a homogeneous population, since they are unable to mate within their clone, while the undifferentiated gamete type has no reduced opportunity for mating, although most matings will be intra-clonal. As shown in a theoretical study by Iwasa and Sasaki (1987) the “mating kinetics” may strongly influence the optimality of a sexual system.

In order to investigate the effects of spatial population structure on the evolution of mating types, we have analysed this process in a cellular automaton model and compared the results it yields to those of the corresponding non-spatial (mean-field) approximation. Such a comparison allows precise consideration of the kinetics of gamete encounters in the model system and emphasizes the role that spatial aspects of the kinetics might have played in mating type evolution. For detailed descriptions of both models see the Methods section.

**Methods**

*The Mating Type Competition System*

The basic setup of our model is similar to that of Hoekstra (1982). The model organism is an aquatic unicellular ‘alga’ with a haplontic life cycle. Three different types of haploid cells compete for space and reproduce both vegetatively and sexually. During the periods between instances of sexual reproduction, the cells multiply vegetatively, producing genetically identical daughter cells. When entering the sexual cycle, a vegetative cell turns into a gamete that can fuse with another gamete. In their gamete stage the three types of cells differ in their mating capacities as represented by different configurations of recognition molecules on the cell surface, as shown on Fig.III.1.1.
The first gamete type $G_1$ is ‘pan-sexual’ and can mate with any potential partner including its own type, while the other two, $G_2$ and $G_3$, are mating types, unable to mate with their own kind. Thus the system allows four kinds of matings: $G_1.G_1$, $G_1.G_2$, $G_1.G_3$ and $G_2.G_3$ of which only the last one involves both mating types.

In this basic model we furthermore apply the following assumptions. The fitness of a vegetatively produced daughter cell is equal to (or lower than, see below) that of its parent. Sexual fusion produces a dormant zygote which upon germination gives rise to haploid vegetative cells through meiotic division, in which the parental gamete types segregate as if determined by a mendelian pair of alleles. To these meiotic products – “post-zygote” vegetative cells – a higher fitness, i.e., a higher division rate and/or a lower death rate, is attributed than to “pre-zygote” vegetative cells not having gone through a sexual cycle in the near past. That is, we assume that sexual offspring have an immediate short-term fitness advantage over asexually derived daughter cells. The actual advantage may be dependent on whether the zygote has been produced by “outbreeding” (with at least one of the gametes involved belonging to one of the two mating types) or “inbreeding” (both gametes pan-sexual). In general we may, but need not, assume that inbred zygotes yield vegetative cells of somewhat less (but still positive) fitness advantage than outbred zygotes. Note that here “outbreeding” and “inbreeding” mean mating between different and identical gamete types, respectively, i.e., we assume – without specifying the precise nature of this outbreeding advantage – that mating between different gamete types may result in fitter offspring on average than mating between cells of the same (pan-sexual) gamete type. The simplest possible genetic mechanism with this effect might be the production of recombinant offspring carrying fewer (slightly) deleterious alleles than both parental genotypes. This mechanism will be operative more often in heterotypic than in homotypic matings, because among the latter a larger proportion will involve selfing (mating between genetically almost identical genotypes). The fitness advantage of sexually derived vegetative cells fades away in time during successive rounds of vegetative reproduction (fitness erosion due to the accumulation of harmful mutations), but it can be regained through another sexual event. This means that post-zygote cells
return to the pre-zygote state when they are not involved in a new sexual cycle for a sufficiently long time.

As for the ecology of the system, we assume that the habitat consists of a limited amount of sites that cells can occupy, and that the three cell types are competing for these sites. Death events leave empty sites behind, which can be occupied later by new offspring. The chance of a newborn cell to settle is proportional to its division rate and the number of empty sites available. In accordance with what has been said earlier about the fitness advantages of sex, three different division rates and death rates are possible: one for pre-zygote, the second for inbred post-zygote, and the third for outbred post-zygote vegetative cells. The straightforward fitness order of these three types is: $W_{\text{pre-zygote}} < W_{\text{post-zygote,inbred}} \leq W_{\text{post-zygote, outbred}}$.

The fusion of two gametes produces a zygote of double size compared to a gamete, and the zygote enters a dormant state with zero rates of division and death. Zygotes leave dormancy at a constant rate, giving rise to post-zygote vegetative cells which inherit the mating type of the gametes they are produced by, and gain fitness according to whether the mating was of the inbreeding or the outbreeding type.

Fig. III.1.2 is a diagram of the possible state transitions in the mating type system. The number of possible states for a site is 12 (including the empty state), according to the type of the cell occupying the site. Thus a site can be in any one of the 3 types of pre-zygote vegetative, 4 types of different zygote, 4 types of post-zygote vegetative, and the empty state.
The Nonspatial Model

Based on Fig.III.1.2, the mathematical formulation of the nonspatial (mean-field) model for the competitive mating type system is straightforward; the differential equations for the 12 site-states are:

$$
\dot{x} = (r \cdot E - d) \cdot x - \sigma \cdot x \cdot (y + Y + p + P + Q) + \phi \cdot X
$$

$$
\dot{y} = (r \cdot E - d) \cdot y - \sigma \cdot y \cdot (x + X + p + P + Q) + \phi \cdot Y
$$

$$
\dot{p} = (r \cdot E - d) \cdot p - \sigma \cdot p \cdot (x + X + y + Y + p + P + Q) + \phi \cdot (P + Q)
$$

$$
\dot{Z}_{xy} = 2\sigma \cdot (x + X) \cdot (y + Y) - 2g \cdot Z_{xy}
$$

$$
\dot{Z}_{xp} = 2\sigma \cdot (x + X) \cdot (p + P + Q) - 2g \cdot Z_{xp}
$$

$$
\dot{Z}_{yp} = 2\sigma \cdot (y + Y) \cdot (p + P + Q) - 2g \cdot Z_{yp}
$$

$$
\dot{Z}_{pp} = \sigma \cdot (p + P + Q)^2 - 2g \cdot Z_{pp}
$$

$$
\dot{X} = (R \cdot E - D) \cdot X - \sigma \cdot X \cdot (y + Y + p + P + Q) - \varphi \cdot X + g \cdot (Z_{xy} + Z_{xp})
$$

$$
\dot{Y} = (R \cdot E - D) \cdot Y - \sigma \cdot Y \cdot (x + X + p + P + Q) - \varphi \cdot Y + g \cdot (Z_{xy} + Z_{yp})
$$

$$
\dot{P} = (R \cdot E - D) \cdot P - \sigma \cdot P \cdot (x + X + y + Y + p + P + Q) - \varphi \cdot P + g \cdot (Z_{xp} + Z_{yp})
$$

$$
\dot{Q} = (R \cdot E - D) \cdot Q - \sigma \cdot Q \cdot (x + X + y + Y + p + P + Q) - \varphi \cdot Q + 2g \cdot Z_{pp}
$$

$$
\dot{E} = d \cdot (x + y + p) + D \cdot (X + Y + P) + D \cdot Q - E \cdot [r \cdot (x + y + p) + R \cdot (X + Y + P) + R \cdot Q]
$$

where $x$, $y$, and $p$ are the numbers of sites occupied by pre-zygote vegetative cells ($x$ and $y$: mating types, $p$: pan-sexual type), $Z_{xy}$, $Z_{xp}$, $Z_{yp}$ are the sites of outbred, and $Z_{pp}$ are those of inbred zygotes. Similarly, $X$, $Y$, and $P$ are sites of outbred, $Q$ are those of inbred post-zygote vegetative cells. $E$ is the number of empty sites within the habitat. The parameters of the model are listed and described in Table 1.

The right-hand side of the differential equations for the sites occupied by pre-zygote vegetative cells ($x$, $y$ and $p$) has three terms. The first defines the vegetative fitness of the corresponding cell type (divisions and deaths under the competitive effect of all cell types present in the habitat), the second is the outflow from the pre-zygote vegetative state due to sex, and the third is the inflow due to the fitness erosion of post-zygote vegetative cells. Zygotes have no vegetative fitness; the first term in their differential equations is the inflow due to sex, the second is the outflow due to germination. Post-zygote vegetative cells have a vegetative fitness different from that of pre-zygotes (first term); they form zygotes fusing (after induction to sexual competence) with both pre- and post-zygote cells matching in mating type (second term); their fitness advantage erodes at a
constant rate resulting in an outflow into the pre-zygote state (third term), and the germination of dormant zygotes maintains an inflow from the zygote states (fourth term). The number of empty sites is increased by the deaths of vegetative cells (first three terms) and decreased by the number of sites taken by newborn vegetative offspring (fourth term). The total number of sites does not change in time, so the 12 time derivatives sum up to zero.

Analytical solutions to this nonlinear model are out of question. We have chosen to find equilibria via numerical solutions, in order to be able to compare the results to those of the spatial model (see below). In all numerical calculations the initial populations were 10 pre-zygote vegetative cells of both mating types and the pan-sexual type, all other states had 0 initial abundances.

**The Spatial Simulation Model**

With assumptions as similar to the nonspatial system as possible, we have implemented a site-based (cf. Czárán, 1998.), spatially explicit stochastic cellular automaton model to which the nonspatial system above is a mean-field approximation. The arena of the spatial model is a set of sites arranged in a 300x300 square grid of toroidal topology to avoid edge effects. Each site can be occupied by any one of the 11 cell types (3 pre-zygote, 4 post-zygote vegetative types and 4 types of zygote) or it can be empty. Zygotes occupy two adjacent sites.

The pattern is updated one randomly chosen site at a time, i.e., we use an asynchronous random updating algorithm. Any site chosen for update can be empty, occupied by a vegetative cell, or occupied by a zygote. We specify the algorithm for each of these cases in turn. A schematic diagram of a single step of updating is given in Fig.III.1.3.

<table>
<thead>
<tr>
<th>Parameters of the non-spatial model:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$ pre-zygote birth rate</td>
</tr>
<tr>
<td>$R$ post-zygote birth rate (outbred)</td>
</tr>
<tr>
<td>$R'$ post-zygote birth rate (inbred)</td>
</tr>
<tr>
<td>$d$ pre-zygote death rate</td>
</tr>
<tr>
<td>$D$ post-zygote death rate (outbred)</td>
</tr>
<tr>
<td>$D'$ post-zygote death rate (inbred)</td>
</tr>
<tr>
<td>$\sigma$ sex rate</td>
</tr>
<tr>
<td>$g$ germination rate</td>
</tr>
<tr>
<td>$\varphi$ erosion rate of post-zygote fitness advantage</td>
</tr>
</tbody>
</table>

**Table III.1.1. Parameters of the mean-field approximation model**
Empty site update: After updating, an empty site can be occupied by one (and only one) of the vegetatively produced offspring of the cells in the 8 neighbouring sites (i.e., the Moore neighbourhood of the focal site), or it remains empty. Each vegetative neighbour $i$ has a chance $p_i$ to put a daughter cell into the empty site. $p_i$ depends on the vegetative reproduction parameter $\beta_i$ ($0 \leq \beta_i \leq 1$) of neighbour $i$. $\beta_i$ is the spatial analogon of $r_i$ in the mean-field model, and it takes one of three possible values depending on whether $i$ is in the pre-zygote, the inbred or the outbred post-zygote state.

Specifically, the chance of the empty site to remain empty is

$$p_e = \prod_{i=1}^{8}(1 - \beta_i),$$

so the probability that the offspring of neighbour $i$ takes the site is

$$p_i = \frac{\beta_i}{1 - p_e}.$$

The rationale behind this formalism is that each neighbour attempts putting an offspring into the empty site with a probability $\beta_i$, but only one of the candidate offspring survives. The chance of survival is proportional to the reproduction parameter of the mother cell.

Vegetative site update: Updating a site occupied by a vegetative cell may result in four possible outcomes: turn the site into the empty state (death),
leave it as it was (survival maintaining fitness), change the vegetative status of the resident cell from post-zygote to pre-zygote (survival with fitness erosion), or produce a zygote (sex). The probability of a death event depends on the death probability $\delta$ of the cell occupying the site, which in turn depends on its vegetative status (pre-zygote, inbred or outbred post-zygote). With a mating partner in one of the neighbouring sites, a surviving vegetative cell may enter the sexual cycle with probability $s$ turning itself and a randomly chosen, suitable neighbour into gametes, and mate. The result is a dormant zygote occupying the two neighbouring sites of the fused gametes. A survivor skipping sex may keep its original fitness, or – if it was a post-zygote cell – it can lose its fitness advantage with a probability $f$ (which is the spatial analogon of the fitness erosion rate $\phi$ in the mean-field model).

**Zygote site update:** A zygote can do two things: remain dormant (with probability $1 - \gamma$) or germinate (with probability $\gamma$). A germinated zygote yields two vegetative cells, the mating types of which are the same as those of the gametes which produced the zygote. The vegetative status of the cells thus obtained is post-zygote, and they can be either inbred or outbred, depending on the parental gamete type combination. The daughter cells are positioned at random into the two sites the zygote had occupied.

At time 0 we have populated 2% of the sites by pre-zygote vegetative individuals of both mating types and the pan-sexual type, assigning individuals to sites at random. All other sites were empty at time 0. The simulations were run for 10,000 generations.
Results

The specific questions we address with both the mean-field model and the cellular automaton are the following:

a) Are there reasonable parameter values that allow the coexistence of the mating types and the pan-sexual type?

b) Under what (if any) circumstances is it possible that the mating types exclude the pan-sexual type?

c) Does spatial structure play an important role in the outcome of the mating type competition system?

Coexistence of the two Mating Types and the Pan-Sexual Type

Numerical solutions to the mean-field model and simulations with the cellular automaton reveal that the system admits a single stable equilibrium state both in the non-spatial and in the spatial setting (Fig.III.1.4). The actual equilibrium densities depend on the parameters, i.e., on the vegetative growth rates \( r, R, R' \), the vegetative death rates \( d, D, D' \), the germination rate \( g \), the sex rate \( \sigma \) and the finess erosion rate \( \varphi \) in the mean-field, and the corresponding probability parameters in the cellular automaton model. Having explored a broad range of the parameter space - with straightforward constraints on the fitness parameters (birth and death rates), i.e., with \( D \leq D' \leq d < r \leq R' \leq R \) - we found that it is the strength of the inbreeding effect (the difference of \( D \) and \( D' \) and that of \( R' \) and \( R \)) and the rate of fitness erosion \( \varphi \) that has the most interesting effects on coexistence. Changing the remaining parameters – the sex rate and the germination rate – within reasonable limits (\( \sigma > 0, g > 0 \)) does not affect the results in a qualitative sense.

We have scaled the inbreeding effect into a single parameter \( \xi \), defined by the equations

\[
D_\xi = D + \xi(D' - D) \\
R_\xi = R + \xi(R' - R)
\]

\( D' \) and \( R' \) have been replaced by \( D_\xi \) and \( R_\xi \) both in the mean-field model and in the spatial simulations, with \( \xi \) changing from 0 to 1 along the “inbreeding effect” axis of the graphs in Fig.III.1.5 and Fig.III.1.6. \( \xi = 0 \) represents no inbreeding effect (i.e., the vegetative cells germinated from
outbred zygotes have the same fitness as those produced by inbred zygotes), and \( \xi > 0 \) means a fitness difference in favour of out bred offspring.

Fig.III.1.5. shows the equilibrium densities of the mating types and the pan-sexual type, the zygotes and the empty cells across a range of the \( \xi - \phi \) projection of the parameter space, for both the mean-field model (Fig.III.1.5.a) and the cellular automaton (Fig.III.1.5.b). It is obvious from the graphs that the sum of mating types, pan-sexual and zygote equilibrium densities (and thus the equilibrium density of empty sites) is almost unaffected by the focal parameters, but the relative frequencies of the mating types and the pan-sexual type vary across the \( \xi - \phi \) plane. This applies to both the mean-field and the spatial model.

**Figure III.1.5.** Simulation results:

A) mean-field
   - fitness erosion rate range \( \phi : 0.0 \rightarrow 20.0 \);
   - inbreeding effect range \( \xi : 0.0 \rightarrow 1.0 \);
   - abundance range \( N : 0 \rightarrow 90.000 \)

B) cellular automaton
   - fitness erosion probability range \( \phi : 0.0 \rightarrow 1.0 \);
   - inbreeding effect range \( \xi : 0.0 \rightarrow 1.0 \);
   - abundance range \( N : 0 \rightarrow 90.000 \)

**Role of space**

Fig.III.1.5.a and Fig.III.1.5.b might look quite alike at first sight, suggesting that spatial constraints like short-range interactions and limited dispersion might not play a decisive role in the dynamics of the gamete
type competition system. Upon closer inspection of the data, however, this impression turns out to be wrong. Even though the general shapes of the 3D graphs are similar for the non-spatial and the spatial model, there are important differences between them affecting mainly the persistence of the pan-sexual population.

One of these differences shows up in the biologically significant case of very small $\zeta$ and $\varphi$ values. In the mean-field model, at $\zeta = 0$, that is, at no fitness advantage for outbreeding, the pan-sexual strain excludes the mating types for any positive rate of fitness erosion ($\varphi > 0$). At $\varphi = 0$ (no fitness loss during vegetative multiplications), on the other hand, it is the mating types who win for any $\zeta > 0$. At $\zeta = 0 = \varphi$, the mating types and the pan-sexual type coexist, and the same applies to any parameter combination satisfying $\zeta \neq 0 \neq \varphi$. Thus we can say that the non-spatial (mean-field) model allows coexistence for almost any parameter combination, except for the biologically less feasible margins of the parameter plane. It predicts in general that both the mating types and the pan-sexual type should have persisted, even if at variable relative frequencies. The cellular automaton model yields a different prediction, admitting the exclusion of the pan-sexual type, i.e., the victory of the two mating types on a considerable section of the parameter plane, including the $\zeta = 0 = \varphi$ point and its close (and biologically the most realistic) neighbourhood (cf. Fig.III.1.4).

![Simulation results with 40% sex rate (sex probability) reduction in the pan-sexual strain. Other parameters as in Fig. III.1.5.](image)
**Alternative adaptations?**

One might guess that in the spatial model the ultimate exclusion of the pan-sexual strain – wherever it happens – is the result of its producing too many dormant zygotes. This would mean that the pan-sexual cells are too frequently induced to become sexually competent and that the resulting high mating frequency impairs their ecological competitiveness. With this hypothesis, a logical next question to ask is: can the pan-sexual strain prevent its elimination by lowering its sensitivity to the induction of sexual competence? With modified versions of both the mean-field model and the cellular automaton we have simulated the effect of such an “adaptation” (Fig. III.1.6). The only modification made to the original models was the reduction by 40 percent of the chance that a pan-sexual cell gets induced by a neighbouring gamete resulting in mating. As it is obvious from a comparison of Figs. III.1.5 and III.1.6, this does not solve the problem of the pan-sexual strain – to the contrary, the chances of the mating types to displace the pan-sexual are even slightly better in the modified models for the largest part of the parameter space. In the mean-field model the relative frequency of the pan-sexual population at equilibrium is smaller almost everywhere except for small nonzero values of the inbreeding effect (compare Figs. III.1.5.a and III.1.6.a). In the cellular automaton the pan-sexual strain does a little better for very high values of both the inbreeding effect and the fitness erosion rate, but suffers more everywhere else compared to the original model without sex rate reduction (compare Figs. III.1.5.b and III.1.6.b).

**Discussion**

There are a few conclusions that apply to any simulation regardless of its being non-spatial or spatial. Not surprisingly, increasing the fitness advantage of outbreeding $\xi$ favours the mating types, because all their sexual interactions produce outbred offspring, while part of the matings of pan-sexual gametes always produces inbred offspring with a smaller fitness. Less obviously, increasing the fitness erosion rate $\varphi$ benefits the pan-sexual type in general, because its effective sex rate is higher: every mating attempt of a pan-sexual gamete can be successful, unlike for the mating types which refuse inbreeding. Therefore the pan-sexual type has more chance than the mating types to reset its eroded fitness to the post-zygote level through mating. The faster the fitness erosion, the more pronounced the advantage of being pan-sexual, hence the more frequent the pan-sexual strain becomes.
In the mean-field model the coexistence of mating types and the pan-sexual type at $\xi = 0 = \phi$ is a spatially unrobust phenomenon. It is highly dependent on the assumption that the system is well-mixed, i.e., each cell encounters other cells of each type with a probability exactly proportional to the relative frequency of that particular type within the whole habitat. It is the breaking of this interaction symmetry in the cellular automaton that gives the mating types a definite advantage compared to the pan-sexuals, even at $\xi = 0 = \phi$ (see Fig.III.1.4). The detailed mechanism is as follows: At $\xi = 0$ it makes no difference whether the mating is inbred or outbred, and at $\phi = 0$ the fitness advantage once obtained in a single event of sex cannot be lost. Since dormant zygotes do not die, empty sites can only be produced by the death of vegetative cells, but the death rates are all equal and independent of gamete type, because (after a short transient period) every vegetative cell is in the post-zygote state. For the same reason the birth rates are also equal for all the vegetative cells, so the only factor that can make a difference between the cell types is the availability of empty sites: the limiting “resource” for reproduction. In the mean-field model the empty sites are equally available to any cell, so the growth rates of the pan-sexual and the mating type strains are identical in the long run, hence their coexistence. In the cellular automaton, however, each strain develops patches. The mating type strains do not have sex at all within their own patch, only at the interface with the patches of other strains. The pan-sexual strain has sex all the time everywhere in the habitat, therefore a larger part of its population is in the dormant zygote state. It is for this reason that at the interface with the mating type patches the pan-sexual strain has a smaller supply of vegetative invaders and thus a smaller chance to capture an empty site there. This results in a travelling front between a mating type patch and a pan-sexual patch and ultimately in the demise of the pan-sexual population altogether. This effect can even overcompensate a small disadvantage for the mating types arising from increasing the rate of fitness erosion $\phi$ slightly above 0, therefore the close neighbourhood of the $\xi = 0 = \phi$ point on the parameter plane belongs to the mating types as well. We think that it is exactly this mechanism that makes the mating types victorious in the spatial model at many parameter combinations that allowed for coexistence in the mean-field approximation. The elementary events at the interfaces between patches of different gamete types have a profound effect on the ultimate outcome of their competition at the larger spatial scale of the whole habitat.

An alternative explanation for the difference of mean-field and cellular automaton results could be that it is the finite size effect that kills off the pan-sexual population from the spatial model at many parameter
combinations. Indeed, the cellular automaton is a finite system, the margins of the state space of which are sinks, but looking at the striking difference of the behaviours of the frequency trajectories at $\xi = 0 = \varphi$ for example (or anywhere else where the mating types take over) in the two models proves that it is not stochastic drift but a real dynamical trend that eliminates the pan-sexual strain in the cellular automaton (see Fig.III.1.4). The equilibrium value for the pan-sexual type is so far from zero in the mean-field model and its decrease to zero so steady in the cellular automaton that drift as the cause of the difference can be safely ruled out. Moreover, if the pan-sexual strain could be drifted to extinction, so could the mating types, but in fact we have never obtained ambiguous outcomes: sufficiently long replicate simulations always yield the same result. This applies to the whole range of the parameter space.

In order to explain the net effect of sex rate reduction on the fitness, and thus on the survival chances of the pan-sexual population one has to consider two different aspects. On the one hand, sex rate reduction decreases the relative fitness of the pan-sexual strain, because it decreases the frequency of both its inbred and outbred matings, the means of keeping fitness high. This negative fitness effect is most pronounced at high rates of fitness erosion $\varphi$. On the other hand, less frequent sex yields fewer zygotes, i.e., fewer dormant cells with 0 growth rate (recall that zygotes do not multiply and do not die). If the populations are viable, i.e., if they have a vegetative growth rate higher than 0, then less frequent mating (dormancy) is beneficial in terms of the average fitness of the pan-sexual population. This effect dominates at low values of $\varphi$, where the fitness advantage of sex does not vanish too fast. A comparison of Figs. III.1.5 and III.1.6 shows that neither these effects are strong, but both are detectable. The net influence on the mean-field model is quantitative, the size of the parameter domain of coexistence is not much affected. In the cellular automaton model the overall effect of sex rate reduction is a slightly larger domain of coexistence: the pan-sexual strain cannot exclude the mating types at high fitness erosion rates, and it is somewhat more persistent at medium values of $\varphi$. In all, it is quite obvious that sex rate reduction is not an efficient strategy for the pan-sexual strain to avoid exclusion by the mating types.

There is a logical possibility that asymmetric cell fusion has evolved for other reasons than and prior to sex and has subsequently been incorporated in the evolution of a full sexual cycle (the sequence of syngamy, karyogamy and meiosis). In that case sex would have been asymmetric from the start. This speculative idea has been analysed theoretically by Hoekstra (1990) (see also Bell, 1993). The present analysis
clearly does not apply to that scenario, but implicitly explains why sexual asymmetry did not disappear once evolved.

**Conclusions**

Assuming that sexual reproduction confers some average fitness advantage compared to simple clonal multiplication, and also supposing that the more genetically different the fusing gametes are the bigger the fitness benefit of the offspring can be, we show that a population consisting of two mating types can displace a pan-sexual population which is otherwise similar to the mating types in all other respects. In the most realistic domain of its parameter space (i.e., at low rates $\phi$ of the erosion of sexually gained fitness, and very slight extra fitness benefits for heterothallic – outbred – matings, $\xi$) our spatial (cellular automaton) model shows the evolution towards exclusively two mating types, whereas the non-spatial model of the same system with the same parameters predicts the coexistence of the mating types and the pan-sexuals. Thus, taking for granted that sex is profitable in evolutionary terms, we offer a basically ecological answer to the question why two mating types can be better than just one. This is, however, only a solution to half of the problem of the optimal number of mating types. Could a third, a fourth, a fifth etc. mating type invade the same system? These questions arise on a very general level in relation to the origin of sexual asymmetry, and they call for a more extended theoretical approach in the future.
III.2. Chemical warfare between microbes promotes biodiversity

Introduction

The past 10 years have seen great progress in measuring bacterial diversity by the application of several molecular approaches (Pace, 1997). Many studies (e.g., Amann et al., 1995) conducted in very diverse microbial habitats have arrived at the same general conclusion: only a very small fraction (less than 1%) of the bacterial species present can be recovered by standard cultivation techniques, but the actual biodiversity in microbial communities is enormous in many cases (Szabó, 1988). For a striking example, a 30-gram soil sample from a Norwegian forest has been estimated to contain about half a million species, although this estimate is necessarily based on many ad hoc assumptions (Torsvik et al. 1990; Dykhuizen, 1998). It is difficult to see how such astronomical species numbers could fit into the conventional resource competition framework, even if forest soil can be considered a highly structured habitat.

Widespread Antibiosis in Microbial Communities

Microorganisms very commonly produce and excrete antibiotic compounds that inhibit or kill sensitive strains from their own or from closely related species. The excretion of antimicrobial substances is known to be widespread among bacteria (Reeves, 1972), yeasts (Starmer et al., 1987) and other fungi (Berdy, 1974). Particularly well studied systems comprise some bacteriocins (antimicrobial toxins produced by bacteria) like the colicins from *E. coli* and nisins from lactic acid bacteria (James et al., 1991). Experimental data based on the analysis of strains in *E. coli* collections (Riley & Gordon, 1992; Achtman et al., 1983) suggest that at least 35% of the strains are colicinogenic. Most strains were sensitive to at least one of the 20 different types of colicin tested but multiple resistance was common, 22% of the strains being resistant to all of them. Also the killer phenotypes of *Saccharomyces cerevisiae* and a few other yeasts have been studied in depth (Wickner, 1992). Estimates of killer activity among wild yeasts from various habitats suggest that between 5% and 30% of the strains can kill a standard sensitive *Candida glabrata* strain (Starmer et al., 1987; Abranches et al., 1997).
Effects of Excreted Toxins on Diversity: Inconclusive Experimental Evidence

In view of the widespread occurrence of antibiosis in microbial communities, the diversity conundrum appears to be particularly confusing: multiply toxic environments are the least expected to maintain many different species. Pioneering work (Adams et al., 1989; Chao & Levin, 1991) on the ecological and evolutionary aspects of antimicrobial toxin excretion has demonstrated that competition between a colicinogenic and a sensitive strain of *E. coli* results in the final exclusion of one or the other depending on their initial proportions. Theoretical work (Levin, 1988; Iwasa et al., 1998) on competition between colicin producing and sensitive strains appears to be in line with these results, but other sets of empirical data show a very different picture: small-scale coexistence of toxin producing and sensitive strains has been reported in a number of natural and laboratory systems (Starmer et al., 198; Abranches et al., 1997; Ruiz-Barba et al., 1994). One theoretical explanation for killer – sensitive polymorphism within a single species invokes micro-scale habitat segregation, such that sensitive strains do better in poor quality habitats and toxin producers do better in rich habitats (Frank, 1994). However, because of the requirement of a very special habitat structure and spatial distribution of different toxicity types, this explanation seems unlikely to be of general validity.

Cyclic Dominance Hierarchy of Killer, Resistant and Sensitive Strains May Favor Polymorphism

Another theoretical approach assumes an interplay of interference and resource competition (Durrett & Levin, 1997, 1998; Pagie & Hogeweg, 1999; Szabó & Czárán, 2001), defining a cyclic dominance structure of
Killer (K), Resistant (R) and Sensitive (S) strains within a species. The new element of this theory is the inclusion of resistance, a common empirical observation for bacteriocin systems left out from previous models altogether. Resistant strains are immune to the corresponding toxin without actually producing it, due to the damage or the loss of the toxin gene and the presence of intact immunity genes. Since toxin production involves a certain metabolic cost, resistant (R) strains are assumed to be superior to killers (K) in resource competition. Similarly, sensitive (S) strains are supposed to outcompete resistan ts (R) because they do not pay the metabolic cost of producing the immunity factor. The remaining sensitive-killer (S-K) interaction is settled to the advantage of the killer strain by means of the interference competitive effect of toxic killing – a mechanism fundamentally different from resource competition, involving the active suppression of the concurrent population by means other than resource exhaustion. This cyclic interaction pattern (K beats S beats R beats K) can be naturally cast into the game theoretical framework of the Rock-Scissors-Paper (RSP) game (Maynard Smith, 1982; Hofbauer & Sigmund, 1998). However, the mean-field (nonspatial) model for RSP dynamics admits neutrally stable periodic trajectories on the state space of the system at best (Fig.III.2.1.a), and assuming a small fitness cost of interaction makes the situation even worse: the system becomes unstable, all trajectories approaching the margin of the state space gradually through all limits (Fig. III.2.1.b), so that eventually only one of the 3 strategies survives (Hofbauer & Sigmund, 1998). The spatial (cellular automaton) version of the same model is very robustly stable (Fig. III.2.1.c) (Tan & Riley, 1997).

An extension of the spatial cyclic exclusion approach to a multi-type eco-evolutionary model (Pagie & Hogeweg, 1999) of the dynamics of nine different colicin plasmids in *E. coli* populations showed that a high diversity of colicin plasmid types is easily maintained in a single niche. This high diversity is expressed in one of two possible modes, dependent on the metabolic cost associated with resistance. For high resistance cost, persistent strains show “multi-toxicity”, different strains maintaining different combinations of colicin genes. For low metabolic costs the community approaches the state of „hyper-immunity”, in which most bacteria are immune to most toxins but produce only a few, in good agreement with experimental data (Riley & Gordon, 1992).
Methods

A Model of Antibiotics Excretion Shaping the Structure and Diversity of Microbial Communities

We have studied a spatially explicit model of a multi-species microbial community using a multi-type RSP game implemented in a randomly updated cellular automaton. The arena is an environmentally uniform 180 x 180 square grid of cells with a toroidal topology. Each cell of the grid represents the site of a single microbial clone such as a bacterial colony, which is characterized by its ability to produce particular anti-microbial toxin(s) and the corresponding resistance factors. All sites are always occupied, i.e., a colony can only be replaced by another one or remain locally persistent. The model specifies up to 14 different toxins and for each one a clone is either 'Killer', i.e. it can excrete the toxin and is also resistant to it, 'Resistant', i.e. resistant to the toxin but unable to produce it, or 'Sensitive', i.e. unable to produce both the toxin and the resistance factor. Thus the maximum number of strains with different toxicity-resistance patterns is $3^{14}$. Toxin production and resistance are assumed to be costly: metabolic costs are ordered as $S < R < K$. The update of a single cell comprises of the following steps. First, with a probability $m$, the colony in the cell mutates a randomly chosen toxin-type into the subsequent dominant state, i.e., Killer $\rightarrow$ Resistant, Resistant $\rightarrow$ Sensitive and Sensitive $\rightarrow$ Killer. The first two transitions require the mutational inactivation of a gene and are therefore 'normal' mutations. The transition from Sensitive to Killer requires the acquisition of a novel toxin production gene as well as the corresponding resistance gene and is probably best viewed as a horizontal transfer (e.g., by transformation) of a genetic element (e.g. a plasmid or virus RNA) carrying both genes or as some other rare sequence of events resulting in the evolution of a novel toxin system. When mutation does not occur the colony undergoes recombination, with probability $r$, with a neighboring colony which is chosen randomly from the 4 orthogonal neighbors. We have modeled recombination in the simplest possible way by not specifying any type of segregation: the strains simply incorporate the toxin and resistance genes of the other strain into their genome, in addition to its own toxin and resistance genes, resulting in two identical colonies. If neither mutation nor recombination occurs, the two interacting colonies determine if one colony is dominant over the other. Colony A interacting with a neighboring colony B will invade and replace B if (i) A can kill B, but B cannot kill A, or (ii) neither can kill the other but A has a smaller metabolic burden, or (iii) both can kill the other but A has a smaller metabolic burden. The update rule for a single site can be represented in pseudo-code as follows:
It is clear that the time scale of an update step of the model is comparable to the time scale of local population dynamics – a complete colony replacement process can be accomplished during such a time unit. One generation is a number of such update steps equal to the number of sites in the grid, so that on average each site is updated once in every generation.

**Results**

*Frozen State* Quasi-Equilibrium

First we consider how a microbial community in which initially no toxin is excreted evolves towards a state of widespread toxin production. Starting from a random community close to the uniform all-Sensitive state ($S(0)$ close to 1.0), allowing small $S \rightarrow K$, $K \rightarrow R$, and $R \rightarrow S$ mutation rates, and setting the metabolic cost of $K$ (toxin production + resistance) equal to twice the cost of $R$ (just resistance), the system evolves to a fine-grained distribution of small patches consisting of monotoxic strains. All 14 possible toxins persist in the system, but the great majority of the colonies carry just 1 toxin gene plus the corresponding resistance gene. Fig. III.2.2 illustrates the main findings in this case. After a few hundreds of generations the spatial pattern appears ‘frozen’, with very little subsequent change going on. We interpret the quasi-frozen mosaic pattern as a ‘deadlock’ in which the great majority of neighboring colonies hold each other in check because their (single) toxin is either the same, in which case they are resistant to each other, or their toxins are different, in which case neither can invade the other because both have the same metabolic cost. Colonies that have acquired an additional toxin gene will spread within the monotoxic patch (group of neighboring cells containing identical colonies) in which they appeared because they are effective against the ‘parent’ strain, but will then with a high probability encounter a neighboring patch consisting of monotoxic colonies with a third type of toxin to which they are not resistant. These monotoxic neighbors will defeat the bitoxic mutant strain because of the higher metabolic burden the latter carries: the patch of the mutant is invaded by monotoxic colonies from all sides and finally it
disappears – the pattern becomes frozen again. Remarkably, the quasi-frozen spatial distribution of small patches of monotoxic strains does not occur in the simulations when the number \( n \) of possible different toxins is smaller than 4. A plausible explanation for this limit invokes the famous Four-Color-Map Theorem (Appel & Haken, 1977) of mathematical topology, which states that for any map 4 colors are sufficient to color neighboring countries always different. Our grid containing patches (coherent groups of identical monotoxic colonies) can be viewed as a map where each patch is a different country. With less than 4 different toxins, a (monotoxic) patch will with a high probability border to a patch of the same type. In other words, at start the grid will then actually contain a labyrinth-like percolation network of connected patches for each monotoxic type. Let us consider the \( n=3 \) case. A bitoxic mutant \( A \) has a large ramified habitat to colonize, giving it more time to survive. During this longer survival time, also a neighboring patch carrying a different toxin has a good chance to obtain a second toxin to form mutant \( B \). The two bitoxic strains play draw against each other, and the time for both to pick up the third toxin is again long enough. With 4 or more different toxins, the likelihood of forming extended patches is small from the outset (this is the point where the Four-Color-Map Theorem becomes relevant), therefore mutations remain local, confined to relatively small patches, resulting in quick elimination by monotoxic neighbours.

**Figure III.2.2.** Multi-toxin RSP game: „Frozen state“ (a) Time series of the relative frequencies of Killer, Resistant and Sensitive phenotypes on an average toxin locus, with different initial states; the yellow dotted line is the approximate „separatrix“ between the „basins of attraction“ for the frozen state and the generic hyper-immunity state; (b) Distribution of the numbers of Killer and Resistant phenotypes per strain at different generations for the all-Sensitive initial state. Simulation type: 14 toxin loci; 180 x 180 grid; metabolic cost for killing is twice as large as for just resistance; all mutation rates are \( m = 10^{-5} \); no recombination \( (r = 0) \)
“Hyper-Immunity” Quasi-Equilibrium

Starting the simulations from a community composition in which the average strain already possesses a few toxin and resistance genes produces convergence to a very different "hyper-immunity" (Pagie & Hogeweg, 1999) quasi-equilibrium characterized by low toxicity, high resistance and a very dynamic pattern. Apparently with the average initial strain possessing more than one toxin plus resistance factor, the likelihood of a draw against all neighboring patches is so small that a frozen pattern is prevented to emerge. The patches persist long enough to collect more and more toxin + resistance genes by mutation and the system evolves towards multiple toxicity. However, once this state has developed, losing some toxin genes while maintaining resistance to all the toxins becomes advantageous due to the reduction of metabolic costs. Eventually the system settles at the generic quasi-equilibrium where the average colony harbours a few (typically 0 to 5 in the 14-toxin system) toxin genes and many (10 to 14) resistance genes (Figure III.2.3). The diversity of toxicity/resistance types is very high at the quasi-equilibrium – we have found up to a thousand different strains in a 180x180 grid, and this number increases with grid size. Although the average levels of toxicity and resistance remain within rather narrow limits, the spatial configuration of the community continues to change while the average patch size is much greater than in the frozen pattern.

Frozen State Prevented by Recombination and by Fast Interference
The frozen state is also prevented by frequent recombination. If the frequency of recombination events is higher than $10^{-4}$ per colony per generation, the system converges to the hyper-immune quasi-equilibrium. Recombination creates multiply toxic strains on the border of two different patches that can invade both parental patches, thus acquiring a relatively large territory. With a sufficiently high rate of recombination this may be repeated frequently enough at the newly established borders with other patches to offset the invasion by strains with at least one different toxin and a smaller metabolic cost. In this way the average number of toxins per strain continues to rise, until mutant strains appear with increasing numbers of resistances. Eventually the generic hyper-immune quasi-equilibrium is established with low toxin and high resistance multiplicity. This typical sequence of events is depicted in Figure 3. Even the all-Sensitive start and negligible recombination are not enough to maintain the quasi-frozen monotoxicity state if the rate of exclusion is much slower for resource competition than it is for toxic killing (i.e., interference competition). If interference competition is at least 4 times faster than resource competition in excluding a competitor strain, the terminal state is always “hyper-immunity”.

**Discussion**

Several conclusions follow from these simulation studies. First, local interference competition resulting from the excretion of antibiotic compounds and the resource competitive effects due to the associated metabolic costs, may produce stable coexistence of huge numbers of different strains or species even in a temporally constant and spatially homogeneous environment if the fitness relations between killer, resistant and sensitive types conform to those of the Rock-Scissors-Paper game. The self-organized spatial pattern (30) of the system always plays a crucial role in defining the resulting community composition. Interestingly, the key idea underlying our work has already been proposed 25 years ago to explain observed patterns of sessile invertebrates on coral reefs. Jackson & Buss (1975) suggested that a cyclic competitive hierarchy among species (A eliminates B, B eliminates C, but C eliminates A directly) may maintain diversity in space-limited systems in the absence of high levels of predation or physical disturbance. However, another study (Connell, 1976) failed to find evidence in support of this hypothesis. Apparently in following years ecologists lost interest in the idea, presumably because convincing examples were lacking. A second conclusion from our work is that the rates of processes like recombination or horizontal genetic transfer, which enable the reallocation of genetic toxicity and resistance
determinants among different strains are relevant for the ultimate composition and spatial pattern of the system. With sufficiently frequent recombination ($r > 10^{-4}$) a community evolves in which the average strain produces only a few toxins but is resistant to many, whereas for smaller $r$ the system remains in the monotoxic frozen state in which the average strain produces just one toxin and is sensitive to all others. Recombination rate behaves as the control parameter of a phase transition here, with the critical value $r_c$ being close to $10^{-4}$.

**Species Diversity in a Community or Polymorphic Species?**

In our model the organisms (‘strains’) are specified by their toxin production profiles and their resistances and sensitivities to toxins. Depending on the rates of the evolution of novel toxin – resistance systems and of their transmission between strains relative to the rates of genetic diversification in other traits and of speciation, different interpretations of the model ‘strains’ are possible. At one extreme, every strain represents a different species in a community. This interpretation is applicable if novel toxicity systems appear at a very low rate and horizontal transmission between species is very rare. At the other extreme, all strains represent variants within a species. This interpretation applies in cases of relatively fast evolution of toxicity systems and high rates of between-strain transfer (either horizontally or by some recombination process). Intermediate interpretations apply to a community consisting of different species, where species may share one or more toxin systems with some other species, but may also exhibit some intra-specific polymorphism in toxin systems. The most natural interpretation of the quasi-frozen ensemble of strains in our model is a multispecies community with low rates of recombination between species. In the hyper-immunity state the many ($\sim 1000$) different strains in the grid can be interpreted as belonging to the same species, but, since plasmid transfer often occurs also between strains from different bacterial species, the game model can also be viewed as that of a multi-species bacterial community. It is interesting in this connection to contrast the available evidence on bacterial systems and yeast communities. Many bacterial systems are characterized by relatively frequent horizontal genetic transfer of toxin and resistance genes, particularly when these are located on conjugative plasmids, and the evidence on bacterial toxin systems indeed suggests that the majority of strains typically produce only a few toxins but are resistant to many (Riley & Gordon, 1992). On the other hand, among yeast strains transfer of genes coding for toxin production and resistance (often located on symbiontic dsRNA viruslike particles) is probably very rare, and here the evidence – although not so extensive as for
bacterial systems - points to monotoxicity and absence of multiple resistances (Starmer et al., 1987; Wickner, 1992; Abranches et al., 1997). It would be very interesting to find out empirically if yeast communities may in fact exhibit a spatially quasi-frozen mosaic pattern. The relative time-scales of interference and resource competition can play a decisive role in shaping the outcome of interaction on the community level, which prediction also calls for experimental verification. It is fortunate that the class of organisms to whom we think our model applies best, i.e. micro-organisms, provide systems that are quite suitable for experimental testing under controlled conditions in the laboratory. This is testified by recent work on the relationships between diversity and productivity (Buckling et al., 2000) and between diversity and disturbance (Kassen et al., 2000) using the bacterium *Pseudomonas fluorescens*. It will be a challenge to design experiments that can discriminate between various proposed explanations of diversity, whether they are based on environmental heterogeneity in some form or another (Frank, 1994), or should apply even in homogeneous environments (Huisman and Weissing, 1999, and the present model), although of course they do not exclude each other.
III.3. A spatial model of the evolution of quorum sensing

Introduction

Many species of bacteria exhibit quorum sensing, the ability to release and respond to signalling molecules (Miller & Bassler, 2001; de Kievit & Iglewski, 2000; Crespi, 2001). Typically, cells produce an extracellular autoinducer molecule – often a peptide or an acyl homoserine lactone – and simultaneously sense its concentration in their immediate environment. Once the concentration exceeds a threshold, the cell is induced to undertake some action, for example expressing genes leading to the excretion of specific enzymes, toxins, or other functional products. This type of regulation has been named quorum sensing because a widely accepted interpretation holds that it allows the bacteria to sense population density. Many bacterial functions, like biofilm differentiation, swarming, toxin excretion in interstrain competition, and virulence factor production are only optimally expressed above a certain critical population density. Regulation by quorum sensing would allow the cells to express the appropriate behaviour only when it is effective, thus saving resources under low density conditions.

The evolutionary stability of quorum sensing as a social communication system among bacteria is somewhat problematic. If cooperative behaviour involves a fitness cost for the cooperating individual, the system becomes vulnerable to cheaters. For example, in a quorum sensing population mutant cells that do not produce the signalling molecules may obtain the benefits without having to pay the costs involved in producing the signal. For recent treatments of evolutionary aspects of cooperative behaviour in microbial populations, see Velicer (2003) and Travisano & Velicer (2004). Because of the potential evolutionary instability of quorum sensing as a social communication system, Redfield (2002) has proposed that its main function may be the detection of the diffusive properties of the local environment of the cell. Since this would be of immediate benefit to the individual cell, the evolutionary stability problems associated with altruistic behaviour would not arise.

Brookfield (1998) and Brown & Johnstone (2001) have analysed models of the evolution of bacterial quorum sensing. Although differing in modeling approach, both have studied the evolution of quorum sensing in the context of explicit two-level selection, where selection at the individual level operates against cooperation, while selection at the group level
favours quorum sensing. Both studies conclude that under fairly broad conditions stable polymorphism may arise between presence and absence of quorum sensing.

In this chapter we model the evolution of quorum sensing using a Cellular Automaton approach. We concentrate on one particular bacterial phenotype, namely toxin excretion in bacterial interstrain competition. Anticompetitor toxin excretion is known to be regulated by quorum sensing in E.coli and other species (Riley & Wertz, 2002). In contrast to the models by Brookfield (1998) and Brown & Johnstone (2001) we do not explicitly assume two-level selection among individuals and among groups, but consider all effects of competition and quorum sensing to occur between neighbouring cells. Our analysis suggests that quorum sensing is unlikely to have evolved as a regulating mechanism of anticompetitor toxins in bacteria.

**Methods**

*The Quorum-sensing and the Bacteriocin system*

We consider bacterial populations polymorphic for genes controlling the production of and immunity to bacteriocin (a excreted toxin that may kill sensitive cells) and for genes controlling the quorum sensing (QS) system. The locus of *toxin production* and the corresponding *immunity factor* gene constitute the bacterio-cin system, whereas the quorum-sensing (QS) system consists of a *signal locus* and a *response module*. The toxin gene expresses a protein that, at sufficiently high local concentrations, is toxic to related strains lacking the immunity factor (the product of the intact immunity gene). The quorum signal gene codes for an autoinducer molecule, which is excreted into extracellular space. The quorum response module consists of at least two closely linked loci, one of
which codes for a membrane-bound receptor of the quorum signal. The other is a cytoplasmic signal transfer peptide that is modified by the receptor-signal complex so that it can trigger the transcription of the toxin gene (Fig.III.3.1).

The bacteriocin system and the QS system interact through the toxin-production and the QS response genes, to ensure that in quorum-sensing strains bacteriocin production occurs only if the quorum signal is of sufficient concentration in the extracellular space, i.e., at an efficient quorum (local density) of potential toxin producing cells. This postulate is in line with the consideration that it is economical in terms of metabolic efforts to keep the toxin gene silent until the local density of potential toxin producers exceeds a critical value. The critical density of toxin producers is called the effector threshold of the toxin, at and above which the toxin is of sufficiently high local concentration to kill sensitive individuals in the close neighbourhood.

To allow for the quorum sensing strains to take the fitness advantage of conditional toxin production we assume that bacteria lacking a functional QS response module produce the toxin constitutively. More precisely, the toxin gene (if present) is expressed if either the QS response module is inactive or the QS response module is functional and the extracellular concentration of the quorum signal exceeds a critical level (the QS threshold) in the close proximity of the cell. For a plausible realization of the simplest possible case of a quorum-sensing regulated bacteriocin system, see Fig.III.3.1.

Based on these assumptions we analyze a model to answer the following question: „Is it theoretically possible that the quorum sensing system has evolved and is maintained as an adaptive regulatory circuit of bacteriocin production?“.

Metabolic costs and functional thresholds

Each of the four functions is assumed to involve a fitness cost due to the metabolic burden of gene expression – a cost that individuals harbouring the inactive allele of the same locus do not pay. An obvious constraint on the relative magnitudes of these costs is that the QS system should be significantly cheaper to run than the bacteriocin system, since QS is itself hypothesized to be an adaptation for reducing toxin costs.

Another assumption is that the toxin gene should be switched on at a local density of potential toxin producers which is just sufficient to kill nearby sensitives. At quorums lower than the toxin threshold the
production of the toxin is just wasting resources. That is, the QS threshold should not be very different from the effector threshold – optimally, they should be equal.

*The cellular automaton (CA) model*

The cellular automaton model is staged on a square lattice of toroidal topology which sets periodic boundary conditions for the habitat. Each cell of the lattice harbours a single bacterium. Individuals are genetically identical, except possibly at the loci controlling the bacteriocin system and the QS system. With respect to the bacteriocin system we consider the following genotypes:

- toxin-"ON" and immunity-"ON" (killer, K);
- toxin-"OFF" and immunity-"ON" (resistant, R);
- toxin-"OFF" and immunity-"OFF" (sensitive, S).

The fourth logical possibility (toxin-"ON" and immunity-"OFF") is non-viable, because such a genotype would commit suicide well before reproduction, thus having no chance to spread in the community. In the QS system we have the following four genotypes:

- signal-"ON" and response-"ON" (signaler-responder, sr);
- signal-"ON" and response-"OFF" (signaler, s0);
- signal-"OFF" and response-"ON" (responder, 0r);
- signal-"OFF" and response-"OFF" (deaf-mute, 00).

With respect to the four loci considered we have $3 \times 4 = 12$ possible allele combinations (genotypes), each paying a different fitness cost. We seed the system with a homogeneous population of deaf-mute sensitives (S00) and let them mutate and interact in a stochastic manner according to the following rules:

1. Each one of the 4 functions (i.e., bacteriocin production, immunity to the bacteriocin, QS signalling and QS responding) is controlled by an...
independent locus (or a closely linked group of genes: a functional locus);

2. Each of the 4 functions can be acquired (switched ON) or lost (switched OFF), either by mutation, or transformation, or simply by invasion by a different genotype – more details below);

3. The QS signalling, the QS detection and the immunity functions are always "ON" if the corresponding allele is present in the genome;

4. Bacteriocin production (if the toxin producing allele is present in the genome) is always "ON" if QS responding is "OFF" (constitutive toxin production in the absence of a working QS system);

5. Bacteriocin production is conditionally "ON" if the QS responding is "ON" and the number of QS-signaler neighbours exceeds the QS threshold;

6. The toxin is effective only above a local threshold concentration – i.e., if the number of toxin-"ON" neighbours exceeds the toxin threshold;

7. Each of the 4 functions (the expression of the corresponding „ON” allele) carries a associated fitness cost $C$, with the ordering of costs

\[
C_{\text{toxin}} > C_{\text{immunity}} \gg C_{\text{signaling}} > C_{\text{responding}}
\]

8. Interactions are occurring between random pairs of lattice neighbours. The outcome involves the site of the opponent: the winner replaces the loser with a – possibly mutated – offspring of its own;

9. Toxin producing (K) bacteria always beat sensitives (S) if they have enough K neighbours to pass the toxin threshold;

10. If an interaction does not result in killing, the contest is settled on the basis of fitness costs (resource competition), in a stochastic manner: the contestant with the smaller fitness cost has a higher chance to take over the site of its opponent.

We introduce genetic variation into the population by means of „mutations“, but the actual biological mechanism of changing one genotype to another is not specified. It is assumed that each of the functional loci has a chance to flip to the opposite state (from ON to OFF or vice versa) at reproduction with specific „mutation” rates $m_{k}$ and $m_{k+}$ at locus $k$. The mutational changes at the four functional loci are independent
in all cases except for toxin and immunity loci: to prevent the occurrence of suicidal phenotypes, losing (switching OFF) the immunity function in a killer (K) genotype means losing the toxin gene too.

The output of the CA model is simply the frequency distribution of the 12 genotypes on the lattice. We have sampled the parameter space of the model by changing the crucial input parameters: fitness costs, mutation rates, recombination rate, QS threshold and toxin threshold. Simulations over the reasonable parts of the parameter space show that the CA is very robust in terms of the evolutionary fates of the four loci studied: apart from relatively small quantitative differences, all runs of the model predict the same. The results are presented as time series for 1) the genotype frequency distributions and 2) the allele frequencies at the four functional loci.

The nonspatial (mean-field) model

On biological assumptions as similar to those of the CA (points 1.-10. in the previous section) as possible we have built the mean-field approximation of the same system. The only difference between the two models is that neighbourhood interactions are replaced by mass interactions, i.e., the population is of infinite size and the competitive interactions are dependent on overall genotype frequencies instead of local ones. Thus we obtain a nonspatial dynamical system consisting of 12 coupled ODE’s, one for the frequency of each feasible genotype:

\[
\dot{x}_i = x_i \sum_j x_j \left[ \text{Mut}_{ij} + \text{Comp}_{ij} (\mathbf{X},C,qt,tt) \right] , \quad (i = 1,\ldots,12)
\]

where \( \mathbf{X} \) is the genotype frequency vector with elements \( x_i \), \( \text{Mut}_{ij} \) is the mutation rate from genotype \( j \) to \( i \) (calculated from the locus-level rates of mutations \( m_{k+} \) and \( m_k \) on locus \( k \)), and \( \text{Comp}_{ij} \) is the competition coefficient specifying the overall competitive effect of genotype \( j \) on \( i \). Competition has two components: resource competition due to differences in fitness costs \( C \), and interference competition due to toxic killing of sensitive (S) strains by killers (K). \( qt \) is the QS threshold, in this model specifying the total frequency of quorum signalling individuals at and above which toxin production sets in in Ksd and K0d genotypes. \( tt \) is the toxin threshold, the total frequency of actual toxin producing killer (K) genotypes (both constitutive and quorum signal-induced) above which the bacteriocin is effective in the culture. Here we do not specify the actual forms of the mutation and the competition rates – they follow assumptions
1.-10. strictly, but they are rather complicated and uninstructive formally, mainly due to the conditional expression of the toxin gene and the resulting variable fitness costs.

The mean-field model can be solved numerically to obtain time series of genotype frequency distributions.

**Results**

**Mean-field model**

The spatially homogeneous, unstructured mean-field model yields rather sobering conclusions. Depending on the actual values of the input parameters (fitness costs of gene expressions $C$, mutation rates $m_k$, QS threshold $q_t$ and toxin threshold $t_t$) and on the initial genotype frequency distribution used, either all four functions deteriorate resulting in the overwhelming dominance of the S00 genotype (Fig.III.3.2.b), or it is only the bacteriocin system that persists, with the QS system practically eliminated (Fig. III.3.2.a).

To understand these results, first the bacteriocin system has to be studied separately, with the QS system switched off. The three feasible genotypes of the bacteriocin system are killers (K), resistants (R) and sensitives (S). Based on fitness costs alone, S is the fittest of the three types, because it does not pay the costs of toxin and immunity factor production. R is inferior to S, because it pays the cost of resistance, but superior to K which carries the costs of both toxin production and resistance. However, K has the interference competition advantage over S by toxic killing which outweighs its metabolic handicap. Thus the interaction pattern of the three genotypes forms an intransitive cycle like the strategies of the rock-scissors-paper game: K beats S beats R beats K (Czárán et al., 2003). In the spatially homogeneous model this amounts to neutral oscillations (assuming no mutations and no QS). With positive mutation rates the neutrally stable periodic solution of the system becomes damped onto an apparent fixed point due to the net mutational „flow” from all-time frequent genotypes into rare ones. This is why we see damped oscillations on Fig. III.3.2.a and III.3.2.c.

Fig.III.3.2 also shows that the dynamics of the system is hardly affected by coupling the QS system to the bacteriocin machinery, provided that the QS threshold and the toxin threshold are set sufficiently low. This is not surprising given that the sole function of quorum sensing is supposed to be switching off toxin production at frequencies of K below $t_t$ in order to save metabolic costs. At toxin thresholds close to zero this function is
useless, so even the small cost of running the QS system is wasted and bacteria expressing it are selected against. This results in the mutation-selection equilibrium frequencies for the QS alleles.

Since the dynamics of the bacteriocin system is periodic, in simulations started far from the equilibrium state killers closely approach

Figure III.3.2. Numerical solutions of the mean-field model. Input parameters: mutation rates: $10^{-6}$; basic fitness cost for all genotypes: 100 (arbitrary units), metabolic costs: QS responding 1, QS signalling 3, bacteriocin immunity 10, bacteriocin production 30.

Left panels: relative frequencies for the genotypes of the bacteriocin system (K:red, R:blue and S:green). Right panels: relative allele frequencies for QS response (genotypes sr + 0r): yellow, QS signal (genotypes sr + s0): green, bacteriocin immunity: blue and bacteriocin production: red.

A: Solution with zero QS and toxin thresholds, initial state: uniform S
B: Solution with QS and toxin thresholds 0.03 (3 % rel. frequency), initial state: uniform K
C: Solution with QS and toxin thresholds 0.03, initial state: 1/6 K, 2/3 R, 1/6 S
the extinction state again in each period. However, once the frequency of killers shrinks below $tt$, there is no way for it to return, because the bacteriocin is inefficient at killer phenotype frequencies under the toxin threshold. Since killers devoid of the toxin effect are the least fit of all types, they will be outcompeted by R and S. Ultimately, both K and R will be swept away by S (Fig.III.3.2.b). The only way to save the persistence of the bacteriocin system with a toxin threshold is to keep the frequency of K safely above $tt$ at all times. If we initiate the population with the S genotype only, this is unlikely to happen, because the system returns very close to the pure S state in the next period. Keeping K away from the toxin threshold is possible by initiating the system with a state sufficiently close to its actual fixed point, but this is of very low probability for random initial conditions.

More important, and even more surprising: Fig.III.3.2.c shows that even if $qt$ and $tt$ are relatively high and the initial state is close to the fixed point of the bacteriocin system, quorum sensing does not evolve: the frequencies of the signaler allele and the responder allele do not exceed their mutation-selection equilibria based on mutation rates and fitness costs alone. That is, the QS system itself behaves neutrally in the context of bacteriocin production: apart from selection due to metabolic fitness costs, QS alleles undergo neither positive nor negative selection pressures in the mean-field model. We suspect that the immediate reason for this is that the cheap QS system is just as good a mediator of dishonest signals as it can be of honest ones. We shall return to this interpretation in more detail later.

**Cellular automaton**

Spatial interactions change many of these conclusions. First: unlike the mean-field system, the CA is stable in terms of the relative frequencies of K, R, and S genotypes even without the damping effect of mutations: any mixed initial state (with all three genotypes present) quickly converge to a stationary frequency distribution. The equilibrium thus approached is very static within the lattice as a whole, but very dynamic at the local scale: each site of the lattice experiences the endless [S to K to R to S] state transition cycle. On the mesoscale we see patches of killers chasing patches of sensitives chasing patches of resistants etc. (Czaran et al., 2002). Second: unlike the mean-field model, the CA predicts the persistence of the bacteriocin system even at quite high toxin threshold values. Starting from a uniform S population and letting them mutate, in time K type mutants will pop up. They are at a disadvantage compared to the „wild-type” S in terms of fitness costs, yet they still have a positive chance to reproduce.
Wherever they happen to reach the toxin threshold frequency locally they form a „beach-head” patch, from which they invade the whole lattice in a very short time. Thus local demographic stochasticity maintains killers and leads to the ultimate equilibrium K-R-S frequency distribution.

**Figure III.3.3.** Time series of a typical run of the cellular automaton model. Input parameters: mutation rates: $10^{-4}$; basic metabolic burden for all genotypes: 100 (arbitrary units), metabolic costs: QS responding 0.1, QS signalling 0.2, bacteriocin immunity 5.0, bacteriocin production 10.0; QS and toxin thresholds: 3 individuals within the Moore neighbourhood of the focal individual.

Upper panel: genotype frequencies for the bacteriocin system (K red, R blue, S green)

Lower panel: relative allele frequencies for QS response (genotypes sr + 0r): yellow; QS signal (genotypes sr + s0): green; bacteriocin immunity: blue and bacteriocin production: red. The straight lines are the corresponding mutation-selection equilibria calculated from the mutation rates and the metabolic costs of the working alleles.
From this mechanism it is obvious that killers are in a desperate need of cooperation among themselves for survival and spread, and they do exploit the benefits of cooperation very efficiently in the spatial setting. Adopting the quorum sensing “communication” system could apparently make the cooperation even more profitable for them, saving the toxin costs whenever it would be wasted at densities below the toxin threshold. Based on this expectation one is tempted to predict that the functioning QS alleles should become associated with the K genotype and go to fixation very fast, but simulations reveal that this is not what actually happens (Fig.III.3.3). The frequency of the responder allele remains low while the signaler allele slowly spreads through all genotypes and reaches high frequencies. No clear allelic association builds up between bacteriocin production and quorum sensing – Cramer’s V-index suggests a very weak cross-dependence between the bacterio-cin system and the QS modules (V = 0.069 for the 3x4 contingency table of genotype frequency averages – Tab. III.3.1 – over the last 1000 gene-rations of Fig.III.3.3). Changing the crucial input parameters (fitness costs, mutation rates, QS and toxin threshold values) between reasonable limits does not affect the main conclusion: quorum sensing does not evolve as a means of communication among cooperating bacteriocin-producing bacteria to save unnecessary toxin costs.

**Discussion**

The mean-field version of our model does not permit the evolution of quorum sensing as a regulatory mechanism of bacteriocin production. This may in part result from the absence of spatial structure in this model, which prevents the formation of locally high concentrations of bacteriocin producing cells, which might profit the most from regulation by quorum sensing. However, the results from the Cellular Automaton version of the model suggest another, more important explanation. There are three conspicuous facts in the outcome of the CA model calling for explanation with respect to the temporal pattern of quorum sensing allele frequencies: a) the responder allele is kept at a low frequency; b) the quorum signalling

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>S</th>
<th>R</th>
<th>K</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>3 210</td>
<td>27 803</td>
<td>2 191</td>
<td>33 204</td>
</tr>
<tr>
<td>0r</td>
<td>220</td>
<td>1 789</td>
<td>193</td>
<td>2 211</td>
</tr>
<tr>
<td>s0</td>
<td>3 630</td>
<td>44 631</td>
<td>2 118</td>
<td>50 379</td>
</tr>
<tr>
<td>sr</td>
<td>593</td>
<td>3 183</td>
<td>430</td>
<td>4 206</td>
</tr>
<tr>
<td>Σ</td>
<td>7 653</td>
<td>77 415</td>
<td>4 932</td>
<td>90 000</td>
</tr>
</tbody>
</table>

*Table III.3.1.* Contingency table of a simulated genotype distribution. Cell values are the frequencies of the 12 possible genotypes averaged over the last 1000 generations (from gen. 9.001 to 10.000) of the run producing Fig.III.3.3. Cramer’s V-index for the mutual dependence of the bacteriocin system and the quorum sensing system is 0.069 for these data, showing a very weak association between bacteriocin production and quorum sensing.
allele spreads over all bacteriocin genotypes; and c) there is no association between those alleles of the bacteriocin system and the QS modules that are expected to benefit from cooperation. The common reason for these facts is cheating. The quorum sensing system, just like any other form of communication, is exposed to cheating strategies – in our case, to mutant genotypes promising cooperation (issuing the quorum signal) thereby convincing others to cooperate (produce toxin), yet not cooperating when it comes to investment into toxin production. Cheaters are at an obvious advantage while they are rare: toxin production is costly, but the cheap quorum signal is sufficient to induce toxin production in cooperating neighbours, which means they have a good chance to enjoy the benefit of the bacteriocin without paying for it. The most successful of cheaters is obviously Rs0, the resistant type which is signalling but not responding (i.e. producing bacteriocin). It is the resistant type that pays the least yet gets most help from others. Rs0, the most efficient parasite is also by far the most frequent genotype in the quasi-stationary states of the simulated communities (cf. Tab.III.3.1). As cheaters become common, their fitness advantage vanishes, but the reliability of the QS „communication system” is ruined by then: most of the strains have gotten rid of the „input channel” (the QS responder allele) to avoid the disadvantage of being cheated (i.e., producing the toxin in vain). If the quorum signal is cheap to produce, the slight advantage of cheating the few remaining cooperators will spread it over almost all genotypes. This might be prevented by a higher metabolic cost of QS signalling, but it does not help the QS responder allele to attain a higher stationary frequency, because the vast majority of the signalling alleles occur in the Rs0 (parasitic) genotype anyway. The overall result is the demise of the QS system due to cheating.

Our result that quorum sensing is not expected to evolve as a regulatory circuit of bacteriocin production raises the question why in present-day bacteria toxin production appears to be regulated by quorum sensing (Riley & Wertz, 2002). A possible answer may be that quorum sensing has evolved and is being maintained for other reasons. We are presently theoretically exploring this possibility, using the same modeling approach under a set of more general assumptions.

Another interesting question concerns the basic difference between our results and those of earlier theoretical analyses of the evolution of quorum sensing by Brookfield (1998) and Brown & Johnstone (2001). Although using different models, both these studies considered two-level selection: acting on individual cells and on well-defined groups of cells (colonies), such that colonies containing quorum sensing members enjoyed selective advantage compared to colonies in which quorum sensing is
It is this aspect of group selection which allows the coexistence of quorum sensing and non-quorum sensing strains (see also Szathmary & Demeter 1987). In our approach there is also selection at the individual level opposing selection at the level of groups of cells (namely groups formed by the individual cell considered and its immediate neighbours), but this group structure is much more diffuse and apparently less effective in compensating the selective disadvantage suffered individually by quorum sensing cells. We conclude therefore, that the precise spatial organization of bacterial communities is likely to be a key factor in the evolution and maintenance of quorum sensing.
III.4. Quorum sensing: the evolutionary benefit of dishonest signalling

Introduction

In the past few years many bacterial strains have been shown to excrete certain molecules (typically, but not exclusively, small peptides or acyl-homoserine-lactones, AHLs) which they can also pick up from extracellular space as a signal inducing the expression of specific effector genes in their own genome and in that of their clonemates (Visick & Fuqua, 2005). Once outside the cell, the signal molecule can bind to a receptor sitting in the cell membrane, and the receptor passes the signal to the chromosome through a specialized signal transduction system. The transcription of the effector genes is induced only if the concentration of the signal molecule exceeds a certain threshold in the close vicinity of the cell. Therefore the actual function of this specific biochemical machinery called quorum sensing is widely believed to be the detection of local frequency (quorum) of signalers through measuring the local concentration of the signal compound (cit!). Information on the actual local frequency of potential cooperators may be useful for the members of a population if, compared to solitary actions, a coordinated action involving synchronized, massive gene expression in many individuals is rewarded in terms of fitness. The expression of virulence factors in pathogens, the formation of biofilms, or the production of toxic compounds like bacteriocins are just a few of the many different examples (e.g. Bassler, 1999) in which it pays for all participants to wait for the threshold quorum before switching on the effector genes at once. E.g., expressing virulence factors by pathogenic bacteria are necessary for the invasion of their hosts, but it also provokes immune responses. Solitary virulents are therefore easily detected and defeated by the immune system, yet a large enough pathogen population expressing virulence factors at once can overwhelm the immune defence of the host, thus opening the way for rapid population growth of the pathogen and the development of the manifest disease.

For such seemingly obvious reasons quorum sensing has been interpreted as a form of chemical communication among bacteria (Miller & Bassler, 2001) – a feature supposed to be so widespread because of the inevitable adaptive value it confers on the strains adopting it. However, as Redfield (2002) points out, the evolution and the maintenance of quorum sensing as an adaptive means of bacterial communication and cooperation
is far from obvious. Like any other communication system, quorum sensing is exposed to cheating. Liars, i.e., mutants emitting the quorum signal, thus inducing the expression of cooperation genes in nearby individuals, but themselves not contributing to the common good, still enjoy the benefits of cooperation by others and may prevail. Cheating might ruin the reliability – and thus the adaptivity – of this form of chemical communication. In a previous work (Czárán & Hoekstra, 2006 – Chapter III.3. in this dissertation) we have shown that for exactly this reason it is very improbable that quorum sensing has been evolved as a communication system coordinating bacteriocin expression in the first place, even though we know that recent bacteriocin-releasing strains do use quorum sensing for synchronizing toxin production (e.g., Van der Ploeg, 2005). This means that either quorum sensing had evolved for different purposes and became associated with the bacteriocin system later or there must be an efficient way to avoid (or punish) cheaters in quorum sensing bacteriocin producer strains. There is no known solution to this puzzle yet.

In our present attempt to find a feasible evolutionary scenario for the acquirement and maintenance of a working quorum sensing machinery, we explore another simulation model using more general assumptions with respect to the fitness effect of cooperation and its regulation by quorum sensing. Namely, we assume that the expression of the effector gene(s) is of immediate fitness benefit for bacteria which happen to be in the company of a sufficient number of other cooperating clonemates also expressing the same gene(s) – but this kind of cooperation is costly. Therefore it is not worth expressing the effector until the sufficient quorum (exceeding the effector threshold) of cooperators congregate within a small neighbourhood. The actual mechanism regulating the expression of the cooperating genes might be operon-like, involving a transcription repressor that blocks the start region of the effector module at low concentrations of the quorum signal and releases it above the effector threshold. In such a regulatory system the quorum sensing signal can be used as a cheap „promise” of cooperation. Cells possessing a response module (receptor and transduction system) for the signal and the intact effector gene(s) will start expressing the effector module, provided that the concentration of the quorum signal is sufficiently high around them. The signal is produced by a single gene, whereas the response module (the receptor and the signal transduction system) is coded by a functionally linked group of genes (a functional locus). The cooperating module (effector) is also a functional locus of possibly many genes. These 3 „loci” can each have a working and an inactive „allele”. „Mutations” from the working to the inactive allele forms and vice versa are possible.
By running the simulation model with many different parameter sets we try to answer the following questions: Does cooperation (i.e., a working effector) evolve in such a population? If so, will the cooperators maintain a fully functional quorum sensing machinery to spare the costs of unnecessary effector production at local densities below the effector threshold? What parameters determine the equilibrium frequencies of cooperators and quorum sensors?

**Model**

The model we use is a two-dimensional cellular automaton (CA) of toroidal lattice topology. Each of the 300 x 300 grid-points of the square lattice represent a site for a single bacterium; all the sites are occupied all the time, i.e., bacteria may replace each other, but may not leave empty sites. The inhabitants of the sites are haploid and genetically identical except for 3 functional loci: one for cooperation (locus „A”), the other two for quorum sensing (locus „B” for producing the signal molecule and locus „C” for the signal detection module). Each of these loci can harbour one of two functional alleles: a working (ON) or an inactive (OFF) form. We label working alleles by capital letters (A, B and C); the inactive functional alleles are denoted by small letters (a, b, and c). With 3 loci and 2 alleles on each locus the bacteria can have 2^3 = 8 different genomes (Table II.4.1).

**Fitness effects of cooperation**: The product of the cooperating A allele may increase the fitness of a bacterium, provided there are at least n_e bacteria (possibly, but not necessarily, including itself) within its 3 x 3 neighbourhood expressing the A allele as well; n_e is the effector threshold of cooperation. Less than n_e cooperators in the neighbourhood have no benefit in terms of fitness. The precise mechanism of cooperation and the way the fitness advantage thereof comes about is not specified – we just assume there is a certain measure of fitness benefit in successful cooperation. On the other hand, cooperation carries a fitness cost which is always paid by the cooperator whether or not it enjoys the benefits of cooperation. The cost of cooperation is the metabolic burden associated with the production of the effector. That is, cooperation (expressing A)
carries an inevitable fitness cost and a conditional fitness benefit. Of course for cooperation to be feasible at all the benefit has to outweigh the cost.

**Fitness effects of quorum sensing:** The working allele \( B \) on the first quorum sensing locus „B” produces the quorum signal molecule, whereas the working allele \( C \) on the second locus „C” codes for the response module (the quorum sensing receptor and the signal transduction peptides). Both of these imply a fitness cost as well, because producing the signal and running the response machinery obviously takes metabolic resources, even if less than cooperation itself. The fitness benefit of a quorum sensing system is an indirect one: communication using a signalling system may spare unnecessary costs of futile attempts to cooperate whenever the local density of potential cooperators is lower than the effector threshold \( n_e \). For this communication benefit to be feasible, the quorum sensing machinery altogether has to be much cheaper (in terms of metabolic costs) than cooperation itself, otherwise constitutive (unconditional and permanent) cooperation would be a better option for the bacterium, and resources invested into quorum sensing would be wasted. Thus the ordering of the metabolic fitness costs of cooperation and quorum sensing are assumed to be \( m_A \gg m_B \geq m_C \). The inactive alleles \( a, b \) and \( c \) carry no metabolic cost: \( m_a = m_b = m_c = 0 \).

**The effect of the the quorum sensing genes on the cooperation gene:** The quorum signal is supposed to be the regulator of cooperation: bacteria with an \( A.C \) genome (i.e., those harbouring a working cooperation allele \( A \) and a working response module \( C \) will actually express the effector gene (i.e., cooperate) only if there is a sufficient quorum \( n_e \) of signallers („B” individuals) within their neighbourhood. That is, quorum sensor cooperators wait for a number of „promises” of cooperation in their neighbourhood before they cooperate themselves. \( A.c \) genotypes do not have a functioning response module, therefore they are constitutive cooperators.

**Selection:** Individuals compete for sites. Competition is played out between randomly chosen pairs of neighbouring cells, on the basis of the actual net metabolic burdens \( M(1) \) and \( M(2) \) they carry. The net metabolic burden \( M(i) \) of an individual \( i \) is calculated as the sum of the basic metabolic load \( M_0 \) carried by all individuals and the total metabolic cost \( m_e(i) \) of the actual „A”“, „B” and „C” gene expressions of the genotype of \( i \) (see Table 1), multiplied by the cooperation reward parameter \( r \) if \( i \) is surrounded by a sufficient quorum of cooperators:
$M(i) = [M_0 + m_e(i)]$ if # of cooperators in neighbourhood is below the effector threshold $n_e$

$M(i) = [M_0 + m_e(i)](1 - r)$ otherwise $(0 < r < 1)$. 

Thus, successful cooperation reduces total metabolic burden in a multiplicative fashion. The relative fitness of individual $i$ compared to its competitor $j$ is simply $M(j)/M(i)$. In practice, the outcome of competition is determined by a random draw, with chances of winning weighted in proportions of relative fitnesses. The winner takes the site of the loser, replacing it by a copy of itself.

**Mutations:** During the takeover of a site by the winner of the competition the invading cell, i.e., the copy of the winner occupying the site of the loser, can change one of its 3 functional alleles (chosen at random) from working to inactive or vice versa. We call these allele changes „mutations”, but in fact they can be due to either mutation or some other process like transformation or even the immigration of individuals carrying the „mutant” allele. The point in allowing allele changes both ways (losing and obtaining them) is to maintain the presence of all six different genes ($A, a, B, b, C, c$) in the population so that the system doesn’t get stuck in any particular genetic state because of the lack of alternative alleles. Thus, each of the six possible allele changes has a positive probability. Allele swaps are independent on the three loci – e.g., the quorum signal gene $B$ can be lost without losing the response module $C$ at the same time; the resulting mutant will be „dumb” yet still able to respond to quorum signals.

**Diffusion:** Each competition step may be followed by a number ($D$) of diffusion steps. One diffusion step consists of the random choice of a site, and the 90° rotation of the 2x2 subgrid with the randomly chosen site in its upper left corner. Rotation occurs in clockwise or anticlockwise direction with equal probability (Toffoli and Margolus, …). $D$ is the diffusion parameter of the model: it is proportional to the average number of diffusion steps taken by a cell per each competitive interaction it is engaged in. Larger $D$ means more intense mixing in the population.

**Initial genome distributions and output:** At $t = 0$ the lattice is „seeded” by the „Ignorant” ($abc$; simulation 1.), the „Liar” ($aBc$; simulation 2.) or the „Blunt” ($Abc$; simulation 3.) genotype on all sites. We simulate pairwise competitive interactions, mutations and diffusive movements for $N$ generations. One generation consists of a number of competition steps equal to the number of sites in the lattice, so that each site is updated once
per generation on average. The number of diffusion steps per site per generation is $D$, of course. In each run of the simulation we record and plot the time series of the 8 different genome frequencies, from which the frequencies for the working alleles of the three functional loci can be calculated and plotted against time.

We compare the stationary allele frequencies to their respective mutation-selection equilibria based on the metabolic burdens $m_e$, $m_s$ and $m_r$ and the (uniform) mutation rate $\mu$. Relative frequencies $q$ above the mutation-selection equilibrium $\hat{q} = \mu/s$ indicate beneficial effects and thus, on average, positive selection for the allele. Relative frequencies below the mutation-selection equilibrium suggest deleterious effects other than just the metabolic burden associated to the function.

**Simulations:** We investigate three basically different scenarios:

1. **Both cooperation and quorum sensing are optional:** all the three loci can be occupied by a working or an inactive allele, i.e., all of them have positive mutation rates in both directions (back and forth). Each of the three working alleles have a positive cost of expression, constrained by the relation $m_j \gg m_s \geq m_c$ (see above). The initial genome is the ,,Ignorant” one (abc), and we follow the evolution (the change in allele frequencies) on the functional loci for both cooperation and the two components of quorum sensing.

2. **Quorum signalling is inevitable:** the quorum signal molecule is a metabolic waste that has to be eliminated from the cell anyway. Thus, producing the signal is for free (the signalling cost is zero), and signallers cannot mutate back to the ,,dumb” (.b.) state. We check if the response module and cooperation evolves starting from a monomorphic initial ,,Liar” (aBc) population.

3. **Cooperation gene constitutive:** the cooperation module cannot be switched OFF completely by simply losing it because it carries another essential function at low expression level. High expression levels represent cooperation, but the quorum response module (if present in the genome) downregulates expression to the basic level in the absence of sufficient quorum signal. The extra cost of high-level effector gene expression is the cost of cooperation, but this cost is not paid by individuals with a silenced cooperation module. Back-mutation (i.e., losing the cooperation module altogether) is not allowed because of the pleiotropy of the gene. We start the
simulations with an all-“Blunt” (Abc) initial state and see if the quorum sensing loci are evolved to obtain their working allele forms.

In all these cases we check if varying the fitness reward of cooperation ($r$), the metabolic costs of cooperation and quorum sensing ($m_c, m_s$ and $m_r$), the intensity of diffusive mixing ($D$) and the effector threshold ($E$) has any qualitative effect on the evolution of cooperation and quorum sensing. The simulations have been run until the relative frequencies of the three focal alleles (A, B and C) approached their quasi-stationary values. This could be achieved within 10,000 generations in most cases. The first few simulations have been repeated 3 times with each parameter setting, using different random number arrays, but since variation in the results was negligible at a lattice size of 300 x 300 in all cases, and each run took a long time to finish, we stopped producing replica runs.

**Results**

**Cooperation and quorum sensing both optional**

With all the three functional loci free to mutate in both directions (to and from the working allele forms) by probability $10^{-4}$, the ON-state alleles of cooperation and quorum sensing have very limited chances to evolve to significant frequencies in the population. A typical run of the simulation model (Fig.III.4.1) with a medium effector threshold ($E = 3$), a moderate cooperation reward ($r = 0.5$) and slow diffusion ($D = 1$) shows that the
cooperating A allele reaches a stationary frequency of roughly 10% in the population, whereas the quorum sensing alleles (B and C) climb to about 20% of relative frequency each. These are quite modest achievements: neither cooperation nor quorum sensing becomes dominant in the population. However, the relative frequencies of the alleles for cooperation (A) and quorum sensing (B and C) are still well above the mutation-selection equilibria calculated on the basis of the mutation rates and the corresponding metabolic burdens $m_e$, $m_s$ and $m_r$. This suggests that both cooperation and quorum sensing have some adaptive value for at least a part of the population, but under the conditions specified by the parameters $E$, $r$ and $D$ the whole population cannot manifest those advantages to the full.

The average benefit of carrying the cooperation genes is a result of rewarded honest cooperation. This outcome seems to contradict Redfield’s (2002) expectation, but it is very sensitive to the spatial structure of the population. More intensive mixing, i.e., faster diffusion ($D = 10$) pushes the relative frequencies of all the three alleles very close to their mutation-selection equilibria (Fig.2.a), which suggests that whenever cheats have

![Figure III.4.2. The effect of diffusion on cooperation and quorum sensing. Parameters as in Figure III.4.2, except for diffusion: A) $D = 10$; B) $D = 0$; C) The pattern of genotypes at generation 10,000 for the simulation with $D = 0$.](image)
better access to their victims they are successful in exploiting and finally almost completely eradicating them. The lack of diffusion ($D = 0$) helps cooperation, of course, by keeping cooperators and their offspring clumped and thus somewhat protected from cheats (Fig. III.4.2.b). Within homogeneous clumps of cooperators, the benefit of cooperation directly translates to a fitness increase of the cooperating individuals themselves, not to that of cheaters. This can help both the cooperating and the quorum sensing genes to attain relative frequencies up to 40%. However, a considerable proportion of the quorum signalling alleles $B$ present in the population are used for cheating even in the zero diffusion case, as they are not always associated with the cooperating allele $A$. Fig. III.4.2.c shows that cheats (i.e., non-cooperating genotypes $a..$ : „Ignorants”, „Voyeurs”, „Liars” and „Lames”) are tightly associated with cooperating genotypes – in fact cheaters appear to „chase” the clumps of cooperators in space.

An increased fitness reward for cooperation has a peculiar evolutionary effect (Fig. III.4.3). Changing $r$ from 0.5 to 0.9 increases the relative frequency of cooperators from 10% to about 15%, which is by no means a conspicuous improvement. However, it gives a real boost to cheaters: the relative frequency of the quorum signalling allele $B$ shoots over 50%, but more than two thirds of the signalers belong to the „Liar” ($aBc$) genotype. This suggests that a very high fitness reward for cooperation increases the temptation for cheating much more efficiently than it increases the drive for cooperation. Changing fitness rewards in the other direction has much worse effect, however: decreasing the fitness reward below $r = 0.2$ ruins
cooperation and the quorum sensing system altogether; the relative frequencies of all the three alleles fall to their mutation-selection equilibria.

**Constitutive quorum signalling**

In this set of runs the quorum signal is always produced (there is no mutant without the quorum signal gene), and the signal is for free (no metabolic burden is associated with the production of the signal molecules). Of course in this case the signal itself is meaningless in terms of promising cooperation, because every individual produces it regardless of its possession or lack of the cooperating allele. Since the signal does not carry information at all, we cannot expect any positive selection for the response module either, i.e., evolution towards a functional quorum sensing system as such is out of question. What is however indeed surprising is that all other parameters left unchanged, the constitutive production of the signal stops cooperation altogether: the relative frequency of the effector allele $A$ falls to its mutation-selection equilibrium (Fig. III.4.4.a). The explanation to this strange effect might be that, even though quorum sensing is quite unreliable and exposed to cheating in the original model, it still turns the fitness balance of the cooperating allele $A$ into positive by preventing the waste of metabolic resources on futile cooperation in many cases. That is, even an unreliable quorum sensing system is worthwhile maintaining because it might have a marginal fitness benefit compared to no quorum sensing at all. This explanation seems to be supported by the fact that any other help to the cooperation gene – either increasing the fitness reward of cooperation ($r = 0.9$ – Fig. III.4.4.b) or stopping diffusion ($D = 0$ – Fig. III.4.4.c) – pulls the stationary frequency of the cooperating allele above its mutation-selection equilibrium, just like the use of a quorum sensing system. Note that operating the quorum sensing system is of some metabolic cost for the bacteria, which means that the fitness benefit of quorum sensing exceeds the fitness handicap due to the metabolic burden of signal production and response. Moreover, even if quorum sensing is to a very large extent prone to cheating, it is still far more efficient in terms of fitness yield to operate the quorum sensing machinery than, for example, increasing the fitness reward of cooperation (compare the stationary frequencies of the cooperating allele in Figs. III.4.1.b and III.4.4.b). A more direct evidence to the beneficial effect of quorum sensing on the maintenance of cooperation is that switching off both quorum sensing genes ($B$ and $C$) in the simulation of Fig.III.4.1 (in which both cooperation and quorum sensing evolve) also prevents cooperation altogether.
Constitutive cooperation

In all the cases studied so far it was assumed that the cooperation gene can be switched ON or OFF by mutation. This means that cooperation and quorum sensing have interacted both ways: the fate of the cooperating allele determined that of the quorum sensing genes, and *vice versa*. However, our primary concern is the evolution of quorum sensing as a regulating mechanism of bacterial cooperation, not the two-way interaction of cooperation and quorum sensing. Therefore now it is necessary to ask: what happens to the quorum sensing alleles if the presence of the cooperation gene is inevitable, i.e., if the cooperating allele \( A \) cannot be lost? Does quorum sensing evolve under this assumption? If so, does it help cooperation? To answer these questions we have set the mutation rates on the „A“ locus to zero, and started the simulations from the all – ”Blunt“ (\( \text{Abc} \)) state. Initially, all other parameters were left unchanged compared to the run in Fig.1; later we have gradually changed other parameters as well.

Before going into details of the simulations it has to be noted that the fixed presence of the cooperating allele does not necessarily mean phenotypic cooperation in all bacteria: the cooperating gene can be down-regulated to a low expression level, and thus the metabolic cost of cooperation spared, by the quorum response allele \( C \) in the absence of sufficient quorum signal. If this were not the case, the quorum sensing machinery would be obviously

\[\text{Figure III.IV}.4.4\text{. The evolution of cooperation and quorum response if the quorum signal is a cue (always produced). Parameters as in Fig.III.1, except for the mutation rates for locus „B“ (these are set to 0), and either the fitness reward: } r = 0.9 \text{ (Panel B), or the diffusion parameter: } D = 0 \text{ (Panel C).}\]
superfluous: it would not have any effect on cooperation because all bacteria are bound to keep the cooperating genotype due to the pleiotropy of the effector gene (i.e., its essential function at a low expression level).

The simulation with constitutive cooperation gives the result on Fig. III.4.5. The most striking feature of this run is the immediate sweep of the quorum response allele C through the whole population. The C allele is obviously used by the bacteria as a „cheating device” in the first place: they adopt it to silence their cooperating gene, and successfully so, since there is hardly any quorum signal producer around yet to force them to cooperate. However, as soon as the „Shy” (AbC) genotype becomes dominant, the production of the quorum signal becomes profitable, because all the „Shy” neighbours of a small clump of „Honest” (ABC) mutants will join in the cooperation and contribute the fitness benefit of the signallers just like their own. It is the metabolic cost of the signal that stops further spread of the B allele, because the fitness benefit of signalling follows a „diminishing returns” pattern: the more signallers in a neighbourhood, the smaller the benefit of joining in signalling. The resulting equilibrium between the „Shy” and the „Honest” genotypes is hardly affected by the few surviving unconditional cooperators („Blunt”, Abc; and „Altruist”, ABc), because these are at a serious metabolic disadvantage against both dominants wherever the quorum of signallers is below the effector threshold.

The evolutionary dynamics of the quorum sensing genes are far less sensitive to spatial population structure in constitutive cooperators than they are when the quorum sensing system is coupled to an optional cooperation locus. However fast diffusion we set (we tried it up to $D = 100$), the result remains qualitatively the same, with somewhat fewer „Honest” and a little more „Shy” genotypes present. Stopping diffusion ($D$...
= 0) has a weak effect in the opposite direction. This indicates that it is cooperation itself that is sensitive to diffusion – with the cooperation gene fixed, quorum sensing is far less affected by spatial dispersal. Higher fitness rewards for cooperation do not significantly change the results either.

**Discussion**

Apart from the metabolic burden associated to the expression of functioning effector and quorum sensing alleles (which, of course, always decreases fitness by itself) there are additional beneficial and deleterious effects stemming in cooperation and cheating. For the locus of cooperation, the beneficial effect is obviously the fitness reward that successful cooperators enjoy. For quorum signallers, the benefit may be the fitness reward obtained by triggering close neighbours to switch on the cooperating effector allele. The quorum signal is the „promise” of cooperation: in neighbourhoods with a sufficient number of quorum signallers present, all the conditional cooperators (i.e., those with functioning signal detector and effector alleles) switch to cooperating mode. However, responding to quorum signals is risky: there is ample room for fake promises in the quorum sensing system. A dishonest signaler can convince a conditional cooperator to switch its effector gene ON even if the signaler does not possess the effector allele itself. Such a liar can only gain by so doing: if the cheated cooperator expresses the effector gene, it may help the neighbouring cheat as much as it helps itself and other cooperators, but the liar does not carry the metabolic burden of expressing the effector genes. With this handicap, cooperators probably lose the competition for empty sites when playing against cheats.

However, we have shown that even under such seemingly hopeless circumstances the genes responsible for cooperation and quorum sensing do evolve to stationary frequencies high above their corresponding mutation-selection equilibria in a spatial model. Both the cooperation and the quorum sensing loci will be polymorphic in the stationary state, with an intermediate equilibrium frequency ratio of their working and silent allele forms. The main reason for the persistence of the non-cooperating allele is cheating, of which there are two, basically different forms in this model specifically. The first form of cheating is silent and passive: it is simply non-cooperation, the result of a knock-out mutation on the cooperating allele. The second is explicit lying: issuing a signal that promises cooperation, and then refraining from joining in the coordinated action for
common benefit when it comes to actual cooperation. The first form of non-cooperation is a kind of passive parasitism, the second is real, active cheating.

We are convinced that the same two kinds of cheating strategies necessarily emerge in any system of cooperation that is coordinated via some kind of communication. Parasitism occurs wherever the fitness effects of cooperation (or that of defection) are not inevitably and precisely fed back to the actor (cooperator or defector) itself in the form of fitness rewards and/or punishments. Lying develops in any setting whereby the signal can be issued independent of the action of cooperation. If the signal is cheap, there is a lot to gain by lying; if it is expensive, then it is worthwhile to get rid of the communication system altogether. In the specific case of the quorum sensing communication system, the responding module can even be used as a cheating device to silence the cooperation gene if for some reason (e.g., the pleiotropy of the locus) there is no way to get rid of the working allele completely.

In summary: our results suggest that a cheap quorum sensing system is advantageous to develop and maintain, both on the level of individual selection as well as for the whole population, provided that spatial constraints on the dynamics of the population, namely the limited spatial dispersion and local interactions of individuals, apply.
Part Four

Prebiotic Evolution
IV.1. Coexistence of metabolically co-operating replicators in a cellular automaton: the importance of space without mesoscopic structure

Introduction

The role of spatial population structure in promoting cooperation and mutualism has received much interest recently (e.g. Nowak & May 1992; Hammerstein & Hoekstra 1995; Killingback & Doebeli 1996). It has also been emphasized in the context of the origin of life, for essentially the same reason: it is a means to establish coexistence of potentially competing template replicators (see Maynard Smith & Szathmáry 1995 for review). There are three known, detailed model approaches: (1) structured deme type models (Wilson 1980; Michod 1983; Szathmáry 1992); (2) replication-diffusion systems as modelled by cellular automata (Boerlijst & Hogeweg, 1991); and group selection of replicators encapsulated in compartments (Szathmáry 1986; Szathmáry & Demeter 1987; Maynard Smith & Szathmáry 1993).

The motivation of these studies originates with the seminal paper of Eigen (1971), arguing (i) that primitive genomes must have been segmented (consisting of physically unlinked genes); (ii) that these unlinked genes must have had the tendency to compete with one another; and, as a consequence, (iii) that some mechanism ensuring their coexistence was needed. He saw the hypercycle to fulfill this role: a system of cyclically interacting molecular mutualists. It turned out to be the case that the hypercycle in a spatially homogeneous setting is vulnerable to parasitism: a cheating replicator that does not give catalytic aid to any member of the cycle can kill it off, provided it receives more help from the cycle than the member that it competes with in the first place (Fig. IV.1.1). As it was realized, compartmentation is an efficient means to separate bad from good genes (Maynard Smith 1979; Eigen et al. 1981). The stochastic corrector model demonstrated that once we have compartments, their internal organization

![Figure IV.1.1. Cooperation of replicators (I₁,...,I₄) in the hypercycle. Circular arrows: replication; solid arrows: direct catalytic help in replication.](image-url)
need not be hypercyclic: a so-called metabolic coupling of replicators is sufficient (Fig. IV.1.2), whereby genes contribute to the good of the compartment by catalyzing its metabolism at various points (Szathmáry & Demeter 1987). The stochastic corrector models assumes that there is an optimal template composition of compartments, which gives the highest protocell division rate (reviewed by Szathmáry & Maynard Smith 1995). Variation between compartments is generated by the stochastic effects in template reassortment upon cell division as well as in replication. Natural selection between the compartments is acting on this variation generated by stochasticity.

The structured deme type models essentially led to the same conclusion: interaction with only neighbours promotes coexistence. Michod (1983) showed the resistance of the hypercycle against parasites in this setting. The viability of the hypercyclic (Fig. IV.1.1) and metabolic systems (Fig. IV.1.2) was also demonstrated in such a context (Szathmáry 1992).

The cellular automaton approach applied so far to the problem of information integration (Boerlijst & Hogeweg 1991) is rather different from both the stochastic corrector and the structured deme framework. It is basically a discretised reaction-diffusion system: replication and diffusion of templates is imagined to take place on an adsorbing surface, without compartmentalisation. Resistance of a hypercycle against parasites is possible in such a reaction-diffusion system, provided the number of replicators exceeds four. The reason for this is that spiral waves emerge as spatial manifestations of the temporal limit cycle trajectory, itself the immediate consequence of the intransitive circle of mutualistic interaction assumed in the hypercycle model. Without the spirals, e.g. with fewer replicators, this particular system collapses if a parasite invades. Cronhjort and Blomberg (1995) have studied numerically the partial differential equation model of the same reaction-diffusion system, and found that the section of parameter space allowing for parasite resistance is smaller than it is in the cellular automaton of Boerlijst & Hogeweg (1991). The spirals are spontaneously emerging self-organised units of selection in the reaction-diffusion system.
approach, which to a certain extent play the role of the compartments in the other two. Spiral waves as units of selection are of course much less definite than the compartments of the structured deme and the stochastic corrector models: whether a replicator molecule belongs to a certain spiral is not always easy to decide.

The fundamental difference of our cellular automaton model as compared to that of Boerlijst and Hogeweg (1991) is exactly in what the stochastic corrector differs from the hypercycle: the dynamical link among the replicator types is realised through a common metabolism, instead of the direct, intransitive hypercyclic coupling. Moreover, unlike in the cellular automaton approximations to reaction-diffusion systems, the corpuscular appearance of the replicator macromolecules is not a methodological compromise in our model, but an essential feature of the object to be investigated: populations of macromolecules can be hardly imagined as fluids consisting of an infinite number of sizeless mass-points. We will show that corpuscularity can be, and in our model it is, of profound dynamical importance (cf. also Durrett & Levin 1994).

Using the cellular automaton model of the metabolic system, our aim was to show that (i) metabolic coupling can lead to coexistence of replicators in spite of an inherent competitive tendency; (ii) parasites cannot easily kill the whole system; (iii) complexity can increase by natural selection; (iv) varying the critical neighbourhood size and the diffusion rate, one can approximate the behaviour of different other models.

The model

The model is a stochastic cellular automaton (Wolfram 1984; Czárán & Bartha 1992; Czárán, 1998) consisting of a 300 x 300 grid of sites with a toroidal topology (wrap-around margins) to avoid edge effects. Each site of the grid can contain at most one replicator molecule. At $t = 0$, half of the sites are occupied by $n$ types of replicators; the types are equally abundant in the initial pattern. One generation consists of three essential component processes: replication, decay and diffusion. We discuss these in turn.

For a replication event of any template $s$ into a neighbouring empty site to occur, $s$ must be complemented by all other types present in its neighbourhood of a given size (Fig. IV.1.3). The claim of template $s$ to replicate into an adjacent empty site is:
\[ C(s) = k_s \left[ \prod_{i=1}^{n} f(i) \right] \]

where \( f(i) \) is the copy number of replicator type \( i \) within the neighbourhood of \( s \) and \( k_s \) is its specific replication rate. Notice that \( C(s) \) is proportional to the geometric mean of the within-neighbourhood replicator frequencies.

The chance of \( s \) to replicate into the empty site is:

\[ P_s = \frac{C(s)}{C_e + \sum_l C(l)} \]  

where \( l \) are the four orthogonal nearest neighbours (the Neumann neighbourhood) of the empty site, and \( C_e \) is the claim of the empty site to remain empty. Thus the probability that the empty site remains empty is:

\[ P_e = \frac{C_e}{C_e + \sum_l C(l)} \]  

Note that the spatially homogeneous dynamics (that is, a possible continuous mean-field approximation - cf. Durrett ..... for this system would be

\[ \frac{dx_i}{dt} = x_i [k_i M(x) - \varphi(x)] \]

where \( x_i \) is the concentration of replicator \( i \), of which \( k_i \) is the growth rate. \( M(x) \) expresses the effect of metabolism on the replication rate. It is a common multiplicative function of the replicator concentrations \( x \), so that each replicator type needs the presence of all the others to be able to replicate, but the metabolic help received is aspecific. \( \varphi \) is the outflow function (cf. Eigen 1971) acting as a density (concentration) dependent
selection constraint. Although all replicators must be present for $M$ to be positive, we know that this does not preclude competitive exclusion of all other types by the fittest (of largest $k_i$) replicator (Eigen & Schuster 1977), since $M$ is the same in all equations of (4), and thus even very small positive concentrations of the competitively inferior types maintain the advantage of the dominant, in terms of the speed of replication. That is, we analyze the beneficial effect of spatial structure on coexistence on the basis of a worst-case assumption, similar to that of the stochastic corrector model (Szathmáry & Demeter 1987).

Decay is aspecific, defined as a constant probability for an occupied site to become empty in time $t + 1$, irrespective of what type of replicator it harbored in time $t$. The competitive (or chance) exclusion from the grid of any one replicator type results in the disappearance of all the other types as well, since they cannot replicate any more but decay continues.

Following Boerlijst and Hogeweg (1991), we modelled the diffusive movement of replicators by the algorithm of Toffoli and Margolus (1987), which preserves particle number and frequency distribution within the grid. One diffusion step is complete by executing the following three instructions: (i) divide the grid into 2 x 2 subgrids within a fixed frame; (ii) rotate each subgrid 90° clockwise or anticlockwise with equal probability; (iii) shift the grid frame one site diagonally with isotropic probability. Diffusion rate increases with the number of diffusion steps between two replication phases.

The update procedure of the cells in the replication phase can be either synchronous or asynchronous: since this is known to affect the arising spatial patterns as well as coexistence (Huberman & Glance 1993), we used both, and for the latter we chose a random update: one site was updated at a time, and that one was chosen at random. To ensure that each site was updated once on the average, the replication phase consisted of a number of updates equal to that of the sites in the grid.
The computer program implementing the cellular automaton was written in MS-FORTRAN 5.1, and run on an ALR Evolution Dual 6 (Dual Pentium Pro 133) machine. The program needs a lot of computing power, especially in terms of CPU time, therefore we could not run the program many times with all parameter sets. We produced replicate runs for a few points of the parameter space; the differences among replicate runs were negligible in all cases.

**Results**

The most important, and also somewhat surprising result of the simulations is that the cellular automaton is capable of producing coexistence in a large part of its parameter space. No conspicuous mesoscopic pattern, similar to spiral waves, arises in any experiment, since a relatively homogeneous spatial distribution of the replicator types is necessary for many neighbourhoods to contain a metabolically sufficient set of macromolecules. Without this, there would not be enough replication events to compensate for the loss due to spontaneous decay. That is, a persistent system cannot show an aggregated pattern.

![Figure IV.1.4.](image)

**Figure IV.1.4.** Time series of simulation results with different system sizes (NR), metabolic neighbourhood sizes (NHS) and diffusion (DIFF). Other parameters: \(C_e\) (vote weight of empty sites) is 2.0, \(p_d\) (decay probability of the replicators) is 0.2 in all cases; \(k_1 = 2.0, k_2 = 4.0\) and \(k_3 = 6.0\) if NR (system size) is 3; \(k_1 = 2.0, k_2 = 10/3, k_3 = 14/3\) and \(k_4 = 6.0\) if NR is 4.
From the viewpoint of template coexistence, the most relevant parameters of the model are diffusion rate, neighbourhood size and system size (the latter being the number of different, metabolically necessary template types). We take a closer look at the effects of these and also of the introduction of parasites below. Fig. IV.1.4. summarizes the ecology of the replicator system for different parameter sets.

**Diffusion.** Increased diffusion rate promotes coexistence. Boerlijst and Hogeweg (1991) were surprised to see that this did not lead to the reappearance of the spatially homogeneous dynamics in their model either: we give an explanation in the Discussion.

**Neighbourhood size.** For a fixed system size, there is an optimum neighbourhood size below which the chance that it contains a metabolically complete set of replicators is too small or even zero (metabolism "does not fit in") and above which replicators start to "feel" the overall population density rather than a strictly local one. That is, too large a metabolic neighbourhood shifts the dynamics towards that of the mean-field model (4), i.e., the metabolic system collapses (Fig. IV.1.5).

**System size.** For any fixed combination of neighbourhood size and diffusion rate, an increase in system size ultimately leads to the collapse of the system for the reason discussed above in relation to neighbourhood size reduction: neighbourhood size is the upper limit of system size. Decreasing system size makes coexistence more likely in any parameter setting, but it is to be noted that we do not consider the absolute efficiency of metabolism to be a function of system size in this model. This would be reasonable to assume, however, since more replicators might be more efficient in catalysing metabolism, giving more chance of survival for the larger system.
Parasites. By definition, parasite molecules are replicated by the others through metabolism, but they do not contribute to it. We found that a coexistent system cannot be killed by a parasite, even if its replication rate exceeds that of the fastest "altruistic" macromolecule type. If the parasite is very efficient, it can depress the concentration of the metabolically active replicators by simply occupying most of the surface of the substrate, but even then, local neighbourhoods containing fewer copies of the parasite will be positively selected. This gives an advantage for the altruistic members of the system against the parasite: wherever there are too many cheaters, the system slows down in relative terms, while localities of comparatively low parasite concentrations speed up in replication. The result is a system stably coexisting with the parasite, as shown on Fig. IV.1.6.

Discussion

The result that there is coexistence without any mesoscopic emergent pattern is robust and counter-intuitive. It is due to the inherent discreteness (i.e., the corpuscular nature of the replicator molecule populations) and spatial explicitness of the model, which grasp essential features of the living world in general, and macromolecular replicator systems in particular. An inferior (that is, slowly replicating) molecule type does not die out since there is an advantage of rarity in the system: a rare template is more likely to be complemented by a metabolically sufficient set of replicators than a common one (cf. Fig. IV.1.3). This effect stems in the joint effect of the discrete and the spatial nature of the system: neither is sufficient without the other.
What would happen in a spatial, but still continuous system could be best demonstrated by turning to the reaction-diffusion equivalent of (4). Obviously this would not help the system to become coexistent: the competitive dominance order of the replicators would be the same at every spatial location $u$, because the metabolic function $M(x, u)$ is the same for all replicators at any location. The reaction-diffusion version of the model would collapse like (4), because the macromolecule type of largest replication rate $k_i$ would dominate everywhere everytime, thus driving the whole system to extinction (cf. Cronhjort & Blomberg 1995).

Allowing for corpuscularity alone without spatial inhomogeneity would not enhance coexistence either. This situation would best be modelled by our cellular automaton with a neighbourhood size equal to the size of the grid. In such a model, all possible neighbourhoods would be of the same composition, independently of spatial position. Increased neighbourhood size mimics, in contrast to increased diffusion rate, better mixing, whereby global rather than local densities determine metabolic efficiency. As demonstrated on Fig. IV.1.5, increasing neighbourhood size is detrimental for coexistence, which also means at the same time that abolishing spatiality would have the same effect.

Increased diffusion rate promotes coexistence because by it the system approaches the following dynamics: local replication -- random reassortment of groups -- local replication and so on. This is practically a so-called "trait group" model sensu Wilson (1980), for which template coexistence has been demonstrated analytically for the hypercyclic as well as the metabolic systems (Szathmáry 1992). Neighbourhood interaction represents a kind of temporary compartmentalisation, which helps maintaining the complete set of metabolically active replicators. Diffusion does not homogenise the compositions of the neighbourhoods below the resolution level of one site, therefore the advantage of rarity does not vanish with the intensity of diffusion getting larger. On the contrary, diffusion drives the two copies of a rare template type apart in space after replication, thus giving each a chance to replicate further in separate neighbourhoods.

Whether complexity can increase in such a system is always a pertinent question. Since the parasite cannot kill the system, and the system cannot eliminate the parasite, once appeared, the cheater can be around for a long time and mutate freely. Thus there is the possibility for a positive conversion to occur: the system can incorporate a mutant parasitic sequence into the metabolic machinery, and make it work for the common good. The beneficial effect might be facultative at first, giving only some (local) replication advantage to neighbourhoods containing the converted parasite,
but later it can become an essential part of the metabolism. Once the interaction is obligate, we have a system one member larger, and metabolically more effective than the original. This scenario will be checked by future simulation experiments. For now, we have checked if such a converted parasite can coexist with the 3-member metabolic system – and it turned out to do so (Fig. IV.1.7)

A general importance of surface dynamics seems more and more important for the origin of life in general: as Wächtershäuser (1988) pointed out, chemical evolution leading to more and more complicated networks, is likely to have taken place on the surface, especially on that of pyrite. Chemical dynamics on a surface occurs essentially in 2D. This has important thermodynamic and kinetic consequences. For example, an appropriate surface can act as a catalyst for the reactions in question. Water, liberated from the surface following condensation reactions leading to larger molecules, renders the reaction favourable, due to the increased entropy of the system as a whole. Surface dynamics of replicators with indefinite heredity is a natural outgrowth of this "primordial pizza" dynamics (cf. Maynard Smith & Szathmáry 1995). In this paper we have shown that the spatial and discrete nature of the metabolic replicator system on a surface crucially change the outcome of selection.

**Figure IV.1.7.** A possible scenario for the evolution of system size. (a) a persistent 3-replicator system with parameters $N_{HS} = 5 \times 5$, $DIFF = 100$, $C_p = 2.0$, $p_d = 0.2$, $k_1 = 2.0$, $k_2 = 4.0$, $k_3 = 6.0$; (b) the same system with a parasite of $k_p = 8.0$ introduced; (c) the same system with the parasite converted to a cooperative (just as essential as the other 3) member of the system.
**IV.2. Prebiotic replicase evolution in a metabolic system**

*Introduction*

The majority of recent theories on prebiotic evolution agree that even the most primitive forms of life must have been cellular, with the first proto-cells including at least three interacting subsystems: metabolism, genetic material, and membrane (Gánti 1970, Maynard Smith & Szathmary 1995, Szostak et al. 2001). At present, none of the existing models of chemical evolution offer any actual scenario that concludes in a fully functional proto-cell starting from what the primeval ocean could deliver. Of course one of the reasons for the lack of such scenarios is that we do not even know for certain what the starting point was: the issue of the chemical composition of the atmosphere and the oceans of prebiotic Earth is still debated (Lazcano et al. 1988, Lazcano & Miller 1996, Ferris et al 1978). This being the case, research on the actual series of events in early evolution is constrained to speculations to a large extent. Moreover, due to the inevitable lack of a prebiotic fossil record we have very limited hope for a solid proof of any conceivable natural history of chemical evolution ever. What we might expect at best is some empirical (chemical and physical) evidence supporting or rejecting any scenario researchers may come up with.

What we shall treat in more detail in this paper is a part of the prebiotic scenario: the evolutionary process that might have led to the coexistence of a few different types of replicator enzymes driving a primitive metabolism, and a replicase catalyzing their replication on mineral surfaces. We assume that both the metabolic enzymes and the replicase are RNA molecules or other polymers of similar chemical structure – even though experimental data are ambiguous on this matter (c.f. Nelson et al. 2000).

The core of the model is a non-evolving metabolic replicator system (Czárán and Szathmáry, 2000. – see Chapter IV.1.), with the addition of an evolving replicator that acts as a replicase on itself and on the metabolic replicator-enzymes. Specifically, we aim to show that the spatial metabolic system is capable of developing and supporting a replicase that evolves to increase both its own replication rate and that of the metabolic replicators, thus increasing the fitness of the whole replicator system.
The model

The metabolic model without replicase

Czárrán and Szathmáry (2000) consider a number of relatively short replicator macromolecules, each of which is capable of catalysing one (and only one) essential reaction of a metabolic reaction network $M$. Metabolism is specified neither in stoichiometric nor in topological terms, we only assume that $M$ produces the monomers for the replication of the macromolecules themselves, and that the catalytic help of the replicators is essential for monomer production. In other words: metabolism aspecifically supports the replicators by providing them with monomers, and the replicators specifically support metabolism by catalysing certain reactions of it (Fig. IV.1.2). The extinction of any one of the replicators $I_i$ results in the collapse of metabolism and thus the demise of the whole replicator system. The simplest mathematical model for the temporal dynamics of a nonspatial system with these properties is

$$\frac{dx_i}{dt} = x_i\left[k_i \cdot M(x) - \phi(x)\right]$$

(1)

where $x_i$ is the concentration, $k_i$ is the growth rate of replicator $i$. $M(x)$ expresses the effect of metabolism on the replication rate. It is a common multiplicative function of the replicator concentrations $x$, so that each replicator type needs the presence of all the others to be able to replicate, but the metabolic help received is aspecific. $\phi$ is an outflow function acting as a concentration dependent selection constraint, keeping the total concentration of the replicators constant. Although all replicators must be present for $M$ to be positive, we know that this does not preclude competitive exclusion of all other types by the fittest (of largest $k_i$) replicator (Eigen & Schuster 1977). Since $M$ is the same in all equations of (1), even very small positive concentrations of the competitively inferior types maintain the advantage of the dominant in terms of the speed of replication. Therefore the replicator of largest $k_i$ will always multiply the fastest of all, excluding the slower (of smaller $k_i$) ones, and with the excluded type missing from the metabolic network the system ultimately collapses with all replicator types going extinct. That is, the metabolic model is not capable of maintaining the community of replicators – at least not in a well-mixed medium.

Czárrán and Szathmáry (2000) show that the very same metabolic system is robustly persistent and coexistent in a spatially explicit cellular
automaton model even if the replication rates of the replicators are different. The crucial assumption of the spatial model is that any replicator needs the presence of all the other types within a small section of space (called the metabolic neighbourhood) around itself for being able to replicate, because local monomer supply depends on local synthesis – diffusion cannot deliver monomers at sufficient concentrations to longer distances. Thus the somewhat surprising result of stable coexistence in the cellular automaton is due to the effect of group selection within small sections of space for metabolically complete replication neighbourhoods. Rare replicators are at a relative advantage compared to common ones, because they have more chance to be complemented by all the common types within a small metabolic neighbourhood – more common replicators have less chance to find at least one rare type molecule nearby and thus to get copied.

Besides its persistence and coexistence in itself, the spatial metabolic model has been shown to be resistant to its only conceivable parasite as well. A parasitic replicator of the metabolic system is one that uses the monomers provided by metabolism for its own replication, but does not itself contribute to monomer production at all (Fig. IV.2.1). Such parasites are unable to kill off the spatial metabolic system even if their replication rate $k_p$ is much larger than those of the cooperating replicators. The simple reason for this is that wherever the parasite becomes abundant, the metabolic system breaks down locally, therefore any further parasite replication becomes impossible, while neighbourhoods devoid of parasitises still produce the cooperating replicator types. The overall effect of this spatial regulation is that the parasite coexists with the metabolic system,
albeit at a relatively low frequency. It is definitely not able to ruin the metabolic system altogether.

**The metabolic model with an evolving replicase**

Once the parasites are around, neither doing real harm to the metabolic system as a whole nor going extinct, they are free to mutate. Mutants may be even more harmful than the original parasite they are derived from, or they may obtain traits that are of help for the survival of the metabolic system – we shall discuss these possibilities in more general terms in below. For now it is sufficient to note that harmful mutants are quickly eliminated from the system by the very same mechanism that keeps the original parasite rare: it kills off nearby metabolic replicators thus committing suicide itself. What we explore below in more detail is the case when the parasite evolves to a beneficial function for the metabolic system by gaining replicase activity.

The basic metabolic machinery is the same cellular automaton as the one described by Czárán and Szathmáry (2000) (Chapter IV.1.): it consists of a 300 x 300 grid of sites with a toroidal topology to avoid edge effects. Each site of the grid can contain at most one replicator molecule, which may be either one of the metabolic „enzymes” (replicators) or a parasite.

**Replication and decay**

Metabolic replicators can do two things: replicate, or decay. For a replication event of any template \( s \) into a neighbouring empty site to occur,

\[ s \] must be comple-mented by all the metabolically act-ive types present in its metabolic neighbourhood of a certain size (Fig. IV.2.2.b). Note that the parasite requires all the \( n \) metabolic types around for its replication, but the
metabolic replicators do not need the presence of the parasite in their metabolic neighbourhood for their reproduction.

The claim of template \( s \) to replicate into an adjacent empty site is:

\[
C(s) = r_s \cdot k_s \cdot M_s,
\]

where \( r_s \) is the replicase support (see below), \( k_s \) is the reproduction rate and \( M_s \) is the metabolic support for replicator \( s \).

\[
M_s = \left[ \prod_{i=1}^{n} f(i) \right]^{1/n},
\]

in which \( f(i) \) is the copy number of metabolic replicator type \( i \) within the metabolic neighbourhood of \( s \). \( n \) is system size (the number of metabolic replicator types). Observe that the parasite, which is type \( n+1 \), is not counted here. Notice also that \( M_s \) is proportional to the geometric mean of within-neighbourhood metabolic replicator frequencies – if any one of the \( n \) metabolic replicator types is missing from the metabolic neighbourhood, then \( M_s \), and thus also \( C(s) \), is zero. The chance of \( s \) to replicate into the empty site is:

\[
P_s = \frac{C(s)}{C_e + \sum_l C(l)}
\]

where \( l \) are the four orthogonal nearest neighbours (the replication neighbourhood, Fig. IV.2.2.a) of the empty site, and \( C_e \) is the claim of the empty site to remain empty. Thus the probability that the empty site remains empty is:

\[
P_e = \frac{C_e}{C_e + \sum_l C(l)}
\]
Decay is aspecific, defined as a constant probability $p_d$ for any occupied site to become empty in time $t + 1$, irrespective of what type of replicator it harboured in time $t$.

**Mutations**

Parasites can also replicate and decay, and $P_s$ is calculated exactly the same way for them as for the metabolic replicators. The difference between metabolic replicators and parasites is that the parasites are not needed for the replication of any template (because they do not produce monomers), they can mutate whenever they replicate, and the mutants may occasionally obtain some replicase activity. Specifically, mutation affects two crucial traits of parasites: their replication rate $k_p$ and their aspecific replicase activity $r_p$. These traits are in a trade-off relation one with the other: if a mutation happens to increase the replication rate of a parasite, it will decrease its replicase activity, and vice versa.

The algorithm of the mutation process is the following: If a parasite is chosen for replication from a replication neighbourhood (Fig. IV.2.2.a), we draw a random number ($d_k$) from a Gaussian distribution of mean 0 and standard deviation $\sigma_k$. $d_k$ determines the mutated replication parameter $k_p'$ of the „daughter“-parasite according to the

$$k_p' = k_p \left(1 + d_k \frac{k_{\text{max}} - k_p}{k_{\text{max}}}\right)$$  \hspace{1cm} (6)

equation, in which $k_p$ is the replication parameter of the „mother“ and $k_{\text{max}}$ is the upper limit for the replication parameter. To avoid mutations producing nonsense (i.e., negative, or above-limit) replication parameters, (6) scales the actual value of $k_p'$ into the $[0,k_{\text{max}}]$ range for any reasonable (i.e., not irrealistically large) value of $d_k$.

Once the mutation change $d_k$ for the replication parameter is specified, the trade-off relation

$$d_e = -a \cdot d_k$$  \hspace{1cm} (7)
determines the expected value of the mutation change in replicase activity, \( d_e \). \( a \) is the trade-off parameter – the larger it is, the more severe the trade-off between replicase activity and reproduction rate. To allow for some „wobbling” in the trade-off relation, we add a Gaussian noise term \( \xi(0,\sigma_r) \) of 0 mean and \( \sigma_r \) standard deviation to the \( d_e \) value, to obtain the actual replicase activity parameter \( d_r \):

\[
d_r = d_e + \xi(0,\sigma_r).
\]  

(8)

\( \sigma_r \) specifies the plasticity of the trade-off relation – the larger \( \sigma_r \) the softer the trade-off, i.e., the more the relation of the two parameters can deviate from the trade-off line (7). \( r'_p \), the actual replicase activity of the mutant „offspring” is determined using \( d_r \), \( r_p \) and \( r_{\text{max}} \) by rescaling in a way completely analogous to (6). Fig. IV.2.2.3 illustrates how \( d_k \) and \( d_r \) are calculated using the trade-off function.

The „daughter-copy” of the parasite is the mutant, the „mother-copy” keeps its original phenotype (i.e., replication rate and replicase activity).

---

**Fig. IV.2.3.** The replicase activity – replication rate trade-off. \( d_k \) is the change in replication rate due to a single mutation. \( d_k \) is randomly drawn from a Gaussian distribution (dashed line) of mean 0 and standard deviation \( \sigma_k \). \( d_r \) is the change in replicase activity due to the same mutation. The solid line is the trade-off curve of parameter \( a \). In panel A \( d_k \) is simply calculated from the \( d_k = -a d_e \) linear trade-off function (hard trade-off case). In panel B a noise component of Gaussian distribution (dashed-dotted line) with mean 0 and standard deviation \( \sigma_r \) is added to \( d_k \) to allow for some „wobbling” in the trade-off relation (soft trade-off). \( \sigma_r \) is the plasticity parameter of the trade-off.
Replicase support

A parasitic replicator “r” either provides help for the replication of neighbouring template molecules or it hampers template replication, depending on its actual replicase activity \( r_r \). If \( r_r \) exceeds 1.0 (the basic catalytic activity of the hosting surface), then \( r \) acts as a real replicase, i.e., it helps replication, but a parasite with \( r_r < 1.0 \) acts as a “poison” for surface catalysis: it binds the template and does not even let the mineral surface help the replication process. Replicase \( r \) that supports (or suppresses) template \( s \) is selected by a draw from within the catalytic neighbourhood of \( s \). The chance \( P_r \) of \( r \) to bind template \( s \) is proportional to its replicase activity \( r_r \):

\[
P_r = \frac{r_r}{\sum_{l=1}^{k} r_{l}} \quad (9)
\]

where \( r \) is a member of, and \( l \) runs through, the set of replicases within the catalytic neighbourhood (Fig. IV.2.2.c) of \( s \). That is, replicator \( s \) is most probably helped by the best replicase in its catalytic neighbourhood. If there is no replicase around then reproduction takes place at the basic rate (of value 1.0) provided by the pyrite surface itself. Of course a parasite (replicase) molecule cannot help its own replication, it needs another replicase in its catalytic neighbourhood for that.

Updating and diffusion

The cellular automaton is updated one randomly chosen site at a time. If the site chosen is occupied by a replicator, it becomes empty with a probability \( p_d \). If it is empty, then all the replicators (metabolic and parasitic) in its replication neighbourhood compete for occupying the empty site with a copy of themselves, according to the stochastic rules (2) – (5). One generation consists of a number of such updates equal to the number of sites (90,000) in the lattice, so that each site is updated once per generation on average.

The diffusive movement of replicators is modelled by the algorithm of Toffoli and Margolus (1987), which preserves particle number and
frequency distribution within the grid. The intensity of diffusive mixing depends on the number of diffusion updates \((D)\) per generation.

**Results**

*Coexistence and parasite resistance*

With \(\sigma_k = 0.0\), i.e., with parasite mutation banned, the model is identical to that of Czárán and Szathmáry (2000). The most important result with this setting is that the cellular automaton is capable of producing coexistence in a large part of its parameter space. No conspicuous mesoscopic patterns like monotypical patches or spiral waves (Boerlijst and Hogeweg, 1991) arise in any simulation, since a relatively homogeneous spatial distribution of the metabolic replicator types is necessary for many neighbourhoods to contain a metabolically sufficient set of macromolecules and thus for the replicators to survive. That is, a persistent system cannot show an aggregated pattern.

From the viewpoint of template coexistence, the most relevant parameters of the model are diffusion rate \((D)\) neighbourhood size \((h)\) and system size \((n)\). Fig. IV.1.4 illustrates the effects of these parameters on the ecology of the metabolic system without parasites and mutation.

*Diffusion.* Increased diffusion rate promotes coexistence in any case of sufficient replicator density. However, very sparse systems are killed by fast diffusion, because potentially cooperating replicators are dispersed apart in space and thus they have little chance to be metabolically complemented. Diffusion helps parasites in spreading, but they cannot drive the metabolic system extinct even at very high diffusion rates: the spatial regulation mechanism works well also at fast diffusion.

*Metabolic neighbourhood size.* For a fixed system size, there is an optimum neighbourhood size below which the chance that it contains a metabolically complete set of replicators is too small or even zero (metabolism "does not fit in") and above which replicators start to "feel" the overall population density, and the results are similar to those of the nonspatial model (1). Note that metabolic neighbourhood size corresponds to the distance within which metabolite concentrations do not shrink too low, i.e., indirectly it depends on metabolite diffusion rates.
System size. For any fixed combination of neighbourhood size and diffusion rate, increasing system size ultimately leads to the collapse of the system for the reason discussed above in relation to neighbourhood size reduction: neighbourhood size is the upper limit of system size. However, this is partly due to an artificial effect of coarse spatial resolution: assuming more sites within a metabolic neighbourhood of the same physical size, the system size effect could be weaker.

Decreasing system size makes coexistence more likely in any parameter setting, but it is to be noted that we do not consider the absolute efficiency of metabolism to be a function of system size in this model. This would be reasonable to assume, however, since more replicators might catalyse a more efficient metabolism, giving more chance of survival for the larger system. That is, we apply a worst-case assumption here.

Parasites. Parasite molecules benefit from the presence of metabolic cooperators which drive metabolism (the source of monomers for replication), but they do not themselves contribute to monomer production at all. As in Czárán and Szathmáry (2000), we found that a coexistent metabolic system cannot be killed off by such a parasite, even if its replication rate exceeds that of the fastest cooperating macromolecule type. Extremely efficient parasites can reduce the concentration of the metabolically active replicators by simply occupying most of the surface available, but even then local neighbourhoods containing fewer or no copies of the parasite will be at an advantage and thus increase the relative frequency of cooperators. The result is a persistent metabolic system which is coexistent with its parasite. At whichever parameter set the system is persistent without the parasite it is also persistent with it in most cases. The system collapses if it would collapse without the parasite, too (Fig. IV.2.4).
Replicase evolution: direct and indirect selection

By allowing for mutation changes in the parasite according to trade-off relations (6) – (8), evolution either towards higher replication rates alone or to both higher replication rates and a subsequent increase in replicase activity occurs (Figs. IV.2.5 and IV.2.6). Which of the two comes about depends on the strength and the plasticity of the trade-off between replication speed and replicase activity. Too strong and too hard a trade-off – i.e., large \( a \) in (7) and small \( \sigma_r \) in (8) – means that a small increase in replication rate causes a large decrease in replicase activity. Since higher replication rates have a direct positive effect on fitness, the system evolves towards increasingly efficient parasites, but the steep slope and the small wobbling of the trade-off relation do not give a chance for replicase activity to catch up even when replication rate has reached its maximum (Fig. IV.2.5).
In contrast, decreasing $a$ and/or increasing $\sigma_r$ (i.e., weaker and/or softer trade-off relations) result in a fast increase in replication rate first, followed by a slower increase in replicase activity later (Fig. IV.2.6). The pattern and the timing of these evolutionary changes can be explained in terms of direct and indirect selection forces acting on the parasites of the metabolic system.

The more obvious of the two cases is direct selection acting in favour of faster reproduction: a larger replication rate is clearly of a direct fitness increasing effect – this is the cause of the steep increase in average replication rate among parasites. However, replicase activity itself has no direct effect on the fitness of the parasite, because parasite molecules with a higher replicase activity cannot help their own replication, only that of their neighbours. In spite of this we find slower, yet steady evolution towards increased replicase activity $r$. The reason for this is indirect selection: better replicase molecules increase the fitness of their neighbours. The benefitted
neighbours are either metabolic replicators or parasites. With limited spatial dispersion of the offspring, neighbouring parasites are with a good chance “relatives” of the one that gives catalytic help: what we see is a typical case of kin selection at work. This in itself results in a fitness increase for parasites with a higher replicase activity in a poorly mixed system, but there is also another indirect beneficial fitness effect which does not require limited spatial mixing. Metabolic neighbours reciprocate the catalytic help they get from the replicase, by providing monomers for its own reproduction. An efficient replicase has more chance to be surrounded by a complete set of metabolic replicators (and thus being copied) simply because the local density of replicators around it is higher than around less efficient ones. This indirect positive selection effect does not depend on spatial mixing (i.e., on diffusion), and even if so, it definitely does not require slow (or no) diffusion.

As for the real influence of diffusion rate $D$ on replicase evolution, we have performed simulations with different diffusion intensities (up to $D = 50$) and found no qualitative difference compared to lower rates of diffusion (Fig. IV.2.6). A slight improvement in the overall performance of the system is observed in the course of replicase evolution: the total number of replicators increases at the expense of empty sites (Fig. IV.2.7).

**Discussion: Metabolic parasites as preadaptations**

As shown above, metabolic parasites can mutate to different functions. Some of the mutants might be even more harmful than just a non-cooperating parasite, directly damaging the metabolic system – but these mutants are doomed to fast extinction, because they kill their „hosts” (the cooperating replicators) before they could enjoy the benefits of having
them around, and thus they die out themselves too. Most mutants will be neutral, i.e., just as parasitic as their ancestors, doing no more and no less harm to the system than just tapping the metabolism for monomers and using them for their own replication. Neutral mutants will diversify and thus „scan“ the sequence space, and they will all coexist with the hosting system just as their ancestors do. This means that many neutral mutants of different sequences accumulate within the metabolic system as new mutations occur. Finally, some of the many neutral mutants might mutate to something that potentially carries some utility for the metabolic system itself – and that something might be many different things.

We have explored the case when mutants can show a little better replicase activity than the very basic catalytic help to replication given by the (probably pyrite) surface harbouring the metabolic system (Fig. IV.2.8). This has been shown to directly benefit the mutant itself, and obviously it is positively selected on the level
of the whole metabolic system too, because it increases the replication rates of all the replicators present. That is, the replicase will spread, and the relative fitness of the mutant system, compared to one with a weaker replicase activity of the „converted” parasite, increases.

Some other mutants may be useful in the metabolic reaction network, possibly catalyzing one or another reaction better than the previous replicator, or even opening new and useful reaction routes. In any case the new, more efficient mutant spreads, and the metabolic system itself also benefits (Fig. IV.2.9).

Still other mutants may open yet another metabolic route by converting some of the intermediate metabolites or waste compounds to small amphipatic molecules, and those may become the first building blocks of a primitive membrane structure wrapping the metabolic system into small vesicles. Such a mutant would be of incredible selective value to the whole replicator community, because it would result in the encapsulation of the – thus far surface-bound – metabolic system, providing it with a more efficient method for group selection: the stochastic corrector mechanism (Szathmáry & Demeter, 1987) (Fig. IV.2.10).

In all, the metabolic parasites of the surface-bound metabolic system may represent preadaptations to virtually any possible catalytic function in a future protocell, including better metabolic enzyme functions, membrane production, and replicase evolution. Membrane production would ultimately lead to the „prebiotic takeoff” of the metabolic system, detaching it from the surface, thus allowing it to enter a new environment and a more efficient selection regime in a proto-cellular structure. All these benefits depend critically on the presence and ceaseless mutations of metabolic parasites on a hosting surface – and these circumstances seem
inevitably be present in the spatial metabolic system anyway. We have explored in some detail the case when mutants gain replicase activity, but both the adoption of some mutants as new metabolic replicators by the metabolic system and the hypothetical case of „prebiotic takeoff” require detailed modelling for more theoretical support.
Epilogue

I am planning to continue research on all the four topics of this dissertation:

1. For the methodical part: I will formalize and algorithmize the configuration-field approximation method, and compare it to pair- (and, in more general terms, cluster-) approximation techniques. Under certain specific conditions, cluster-approximations and configuration-field approximations may give the same results, but in general they are sensitive to different aspects of spatial constraints. I wish to specify the conditions for both cases.

2. In spatiotemporal population dynamics I will continue comparing experimental and field cases to models specifically designed to explain the dynamics of spatially interacting populations. We have concrete plans to do so in experiments with microbial (bacterial and yeast-) populations, but I am sure there is a multitude of phenomena in terrestrial macrophyton community dynamics that cannot be explained without an explicitly spatial approach (cellular automata, distance models or tessellations).

3. Experimental evolution is one of the hot topics of contemporary evolutionary biology. We have supplied ample proof in a number of studies presented in this dissertation to the statement that many of the evolutionary changes not accessible with mass-interaction models can be simply explained using spatiotemporal methodology. This applies to detailed tactical models designed for very specific experimental situations and strategic models of more general, usually mainly theoretical, interest. In cooperation with Rolf Hoekstra and Arjan de Visser in Wageningen University, we continue work on both these fields.

4. We have just started a longer project on prebiotic evolution, using the surface-bound metabolic system as an initial model, with the aim of constructing a relatively detailed, physico-chemically plausible scenario for the origin of the first protocell (i.e., a set of metabolic ribozymes and a replicase, wrapped in a membrane vesicle built by the metabolism itself). The results so far are promising, but the chemical details of the model need to be worked out.
References:


