

# **The evolution of Hutchinson's niche & mutualism**

*Román Ulises Zapién Campos*

**Supervisor: Dr David Murrell**

Centre for Biodiversity and Environment Research, Gower Street, UCL

A dissertation submitted in partial fulfillment

of the requirements for the degree of

**Master of Research in Biosciences**

of

**University College London.**

Division of Biosciences

University College London

September 15, 2016

I, Román Ulises Zapién Campos, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

# Abstract

**Background and Aims.** To understand the prevalence of mutualism in nature is a main challenge for the eco-evolutionary theory. In spite of the diversity of species in nature, the concept of ecological niche allows their comparison in common ground. Here we invoke it to study the evolutionary dynamics of mutualism. By decomposing the dynamics in ecological and evolutionary variables, we develop a mathematical model aiming to highlight possible mechanisms of mutualism persistence.

**Methods.** Our model is based on the exchange of commodities by populations evolving on a surface during several time-steps. Using analytic insights, we find out parameters regions sustaining mutualism, to then introduce a pipeline able to simulate the evolution of systems starting from multiple conditions. Ultimately we simulate a community of two populations that experience scenarios of null-, constitutive- or facultative-mutualism while exchanging two commodities.

**Results.** Our results indicate that ecological equilibrium is the main constraint of the process. This is determined by the fluxes of commodities between individuals. Only facultative fluxes, specially those functioning as public goods, make communities sufficiently resilient during their evolution. Conversely the stringency of constitutive fluxes lead the communities to collapse.

**Summary and Conclusions.** We have developed a quantitative model resembling the eco-evolutionary dynamics of mutualism. Inspired by a simple system, we have shown its potential and flexibility. However, its further application may provide an understanding of natural systems and insights to engineering synthetic ecologies.

# Contents

<b>Declaration of Ownership</b>	<b>2</b>
<b>Abstract</b>	<b>3</b>
<b>1 Introduction</b>	<b>9</b>
1.1 Ecological niche . . . . .	9
1.2 Mutualism . . . . .	11
1.3 Evolution of the fundamental niche . . . . .	13
1.4 Aims . . . . .	14
<b>2 Methods</b>	<b>16</b>
2.1 The model . . . . .	16
2.1.1 Ecological dynamics . . . . .	16
2.1.2 Evolutionary dynamics . . . . .	19
2.2 Parameters dependency . . . . .	20
2.2.1 Maximum probability ( $\gamma_f$ ) . . . . .	20
2.2.2 Steepness ( $\alpha_f$ ) . . . . .	20
2.2.3 x-fold commodity requirement ( $\beta_f$ ) . . . . .	22
2.2.4 Inherited deviations ( $\sigma_{R,k}$ , $\sigma_{l,k}$ ) . . . . .	23
2.3 Variables dependency . . . . .	23
2.4 Ecological equilibrium . . . . .	25
2.4.1 System of coupled equations . . . . .	25
2.4.2 Readdressing the $S_t(x,y)$ and $D_t(x,y)$ maximization . . . . .	28

<i>Contents</i>	5
2.4.3 Exact solutions . . . . .	29
2.4.4 An insight into symmetric spatial distributions . . . . .	30
2.5 Evolutionary equilibrium . . . . .	32
<b>3 Results</b>	<b>33</b>
3.1 Numerical implementation . . . . .	33
3.2 Null mutualism . . . . .	37
3.3 Constitutive mutualism . . . . .	41
3.4 Facultative mutualism . . . . .	43
<b>4 Discussion</b>	<b>48</b>
<b>5 Conclusions</b>	<b>56</b>
<b>Acknowledgements</b>	<b>57</b>
<b>Appendices</b>	<b>58</b>
<b>A Computational code</b>	<b>58</b>
A.1 Main functions . . . . .	58
A.2 Parameters . . . . .	61
<b>B Derivations</b>	<b>62</b>
B.1 Exact solutions: symmetric spatial distributions . . . . .	62
B.1.1 Unequal gathering . . . . .	62
B.1.2 Equal gathering . . . . .	63
B.2 Exact solutions: any spatial distribution . . . . .	64
<b>C Supplementary figures</b>	<b>66</b>
<b>D Colophon</b>	<b>71</b>
<b>References</b>	<b>78</b>

# List of Figures

2.1	Spatial structure in microbial communities . . . . .	16
2.2	Maximization of the fitness function . . . . .	25
3.1	Model assumptions . . . . .	34
3.2	Ecological parameters feasible region . . . . .	35
3.3	Ecological equilibrium . . . . .	36
3.4	Niche evolution: null mutualism (fixed leakage) . . . . .	38
3.5	Variables evolution: null mutualism (fixed leakage) . . . . .	39
3.6	Niche evolution: null mutualism (loose leakage) . . . . .	40
3.7	Variables evolution: null mutualism (loose leakage) . . . . .	41
3.8	Niche evolution: constitutive mutualism . . . . .	42
3.9	Variables evolution: constitutive mutualism . . . . .	43
3.10	Niche evolution: facultative mutualism (unmediated) . . . . .	44
3.11	Variables evolution: facultative mutualism (unmediated) . . . . .	45
3.12	Niche evolution: facultative mutualism (mediated) . . . . .	46
3.13	Variables evolution: facultative mutualism (mediated) . . . . .	47
C.1	Reduced niche evolution: null mutualism (fixed leakage) . . . . .	66
C.2	Reduced variables evolution: null mutualism (fixed leakage) . . . . .	67
C.3	Reduced niche evolution: null mutualism (loose leakage) . . . . .	67
C.4	Reduced variables evolution: null mutualism (loose leakage) . . . . .	68
C.5	Increased niche evolution: facultative mutualism (unmediated) . . . . .	68
C.6	Increased variables evolution: facultative mutualism (unmediated) . . . . .	69

C.7	Reduced niche evolution: facultative mutualism (mediated) . . . . .	69
C.8	Reduced variables evolution: facultative mutualism (mediated) . . . .	70

# **List of Tables**

2.1	Model variables and notation. . . . .	19
2.2	Model parameters and notation. . . . .	21

# **Chapter 1**

## **Introduction**

Ecological and evolutionary theories are cornerstones of our contemporary understanding of biological systems (Cordero and Polz, 2014). They emerged as independent fields and although an intimate interaction has been evident for long time, still much research is needed to unify them in a greater theoretical framework (Pelletier et al., 2009). Knowing that both theories have similarly benefited from a mathematical description, a successful integration will occur if new theory is derived following up this historic rigour.

Nowadays evolutionary ecology faces multiple fundamental challenges. One of them is to explain the abundance and persistence of mutualism in nature (Doebeli and Knowlton, 1998). From its conception this implies the explanation of an ecological interaction in evolutionary time-scales, so an integrative approach is mandatory.

### **1.1 Ecological niche**

The concept of ecological niche was formalized by G. Evelyn Hutchinson to describe the observable variety of living forms in a uniform framework (Holt, 2009). Hutchinson realized that each individual has a set of  $n$  requirements to satisfy in order to survive, and that they can be regarded as continuous variables of an  $n$ -dimensional hyperspace (Blonder et al., 2014). These requirements, that we will refer as “*commodities*”, include material (“goods”) and immaterial (“services”)

resources from biotic and abiotic sources (Ferrière et al., 2007). According to Hutchinson's description an individual will have a range on each of these variables that can sustain its survival, so when the  $n$  dimensions are considered, the concept of *fundamental niche* refers to the  $n$ -dimensional hypervolume constrained by such ranges (Blonder et al., 2014). Other definitions of ecological niche were developed independently by Joseph Grinnell and Charles S. Elton (Wiens et al., 2010); however in this work by ecological niche we will refer to the Hutchinsonian niche.

The fundamental niche concept is useful, nonetheless in nature these ranges are not always observable as the environment only provides a limited number of conditions to whom individuals have to adapt based on spatial and temporal constraints. The concept of *realized niche* is then required to refer to those combinations experimentally observable (Holt, 2009). This realized niche is contained in the fundamental niche and in contrast with it, it is continuously changing in ecological time-scales, while the fundamental niche only changes in evolutionary time-scales (Post and Palkovacs, 2009).

In this context a species is defined by its fundamental and realized niches. Moreover, according to the *competitive exclusion principle*, under constant abiotic conditions multiple species cannot share a niche without the eventual extinction of all but one of the species (Freilich et al., 2011). This has been thoroughly studied to prove that the winner of such competition will be the species whose fitness is the greatest, the so call “*survival of the fittest*” (Nowak, 2006). Hence the concept of *fitness* is referred to how fit a species is after receiving the commodities associated to one spot inside its niche. Several metrics like survival and reproductive rates, are used to this end but generically they regard the fitness as a measure of a species' adaptation to its environment (Lanfear et al., 2014). Indeed, fitness can be expressed as a function of the commodities, where some combinations provide fitter outcomes than others, so species are expected to strive in order to reach those spots.

A consequence arising from last point is that if species have an overlap in any

dimension of their realized niches, *conflict* will occur (Freilich et al., 2011). That is to say, populations will compete for the limited commodities their habitat provides; where *habitat* refers to the geographical and temporal occurrence of populations in the environment. The magnitude of such competition will scale up with the number of dimensions overlapped and their similarity. Therefore, it has been argued that in order to avert conflict, and its associated material and energetic costs, species will tend to avoid niche overlapping, and what is more they will seek to occupy all available niches (Blonder et al., 2014). In spite of that, vacant niches are found in nature and conflict does occur playing a central role in the eco-evolutionary dynamics of biological systems (Jones et al., 2012; Blonder et al., 2014).

Thereby only considering competition, the existence of vacant niches is a dilemma for the evolutionary ecology theory. However, it has been observed that other interactions that are widespread in nature (Jones et al., 2015), may shed light to the problem. In the subsection 1.2 I introduce mutualism, a cooperative interaction, in the context of ecological niche and eco-evolutionary dynamics.

## 1.2 Mutualism

Ecological observations have been traditionally studied by theoreticians either as top-down or bottom-up processes. In the former, changes in populations and communities, the higher trophic levels, are registered to then infer the state of lower levels; while in the latter, changes in food sources, the lowest level, are analysed to explain their effect on populations and communities (Gertsev and Gertseva, 2004).

Although a network description of ecological systems is thought to be contemporary, the idea of an interaction-based dynamics was introduced by de Bray since the 19th century. He proposed the term *symbiosis* to describe close encounters between pair of individuals changing their fitness dramatically (Momeni et al., 2011). According to the effect and directionality, nowadays this term includes competition, predation, antagonism, commensalism, mutualism and parasitism.

*Mutualism* occurs whenever the interacting partners get a mutual benefit out of the interaction. This is ubiquitous in nature, going from the directed exchange of commodities by pairs of individuals (Honegger, 2009; Zee and Bever, 2014) to the release of public goods within cooperative communities (Cordero et al., 2012; Campbell et al., 2015). What is more, since mutualism can be facultative, its influence may not be obvious until demanding circumstances align (Gilbert et al., 2015).

The essence of a mutualistic interaction may seem to contradict the selfishness of individuals. However, at its core, mutualism relies on competition through different means (Jones et al., 2012). Firstly, the limited amount of commodities to exchange causes competition between recipients. Secondly, exchanged commodities give mutualists a competitive advantage in comparison to non-mutualists. Finally, in the face of the emergence of exploiters (also known as freeloaders or cheaters) mutualists have to evolve a response to such challenge. Therefore, it has been argued that mutualism may be accountable for much of the dynamics and diversification of ecosystems (Pelletier et al., 2009; Campbell et al., 2016).

Nonetheless, several situations jeopardize mutualism constantly. Apart from being exposed to low quality interactions, where investments are high and benefits are minimal, mutualism is constantly threatened by exploiters (Sachs and Simms, 2006). These cheat mutualists by consuming commodities but no providing others in return, resulting in their greater relative fitness. Such advantage derives in the *tragedy of the commons*, a fate predicting the collapse of cooperating communities caused by the increasing density of cheaters (Waite and Shou, 2012). This is particularly critical for communities relying on the exchange of costly commodities (Momeni et al., 2013). Accordingly, to overcome this evolutionary threat different mechanisms have evolved: from the immediate as the spatial and kin selections, to the active as the partner fidelity feedbacks and partner choice, to the strategic as conditional rewards and sanctions (Foster and Wenseleers, 2006). Organisms in different kingdoms have evolved such mechanisms (Jones et al., 2015).

If mutualism is successful it can increase partners' habitat range; yet, depending on their closeness, this may cause the loss of autonomy of species as they become more reliant on complementary partners that may not reliably encounter (Hillesland et al., 2014). Interestingly, phylogenetic studies of lichens, one of the most successful examples of mutualism, where a photobiont (algae or cyanobacteria) and a fungi interact to inhabit almost any terrestrial ecosystem (Honegger, 2009), have shown their interaction has broken apart several times, not ending up in the extinction of the lichen but in the reversal to independence of the partners (Lutzoni et al., 2001; Sachs and Simms, 2006).

The spectrum of ecological and evolutionary dynamics introduced here makes mutualism a fascinating topic. This has already produced important results in game theory (Doebeli and Knowlton, 1998; Peña et al., 2016), however the recent interest to engineer synthetic ecologies demands a framework based on easily quantifiable variables (Momeni et al., 2011; Mee and Wang, 2012). Therefore, the analysis of mutualism from a commodities-exchange perspective, appealing to the ecological niche, may provide such framework.

### 1.3 Evolution of the fundamental niche

Knowing that mutualism may exert a strong influence in the evolutionary dynamics of ecosystems, now, we will discuss it in the context of ecological niche evolution. According to Nowak (2006), the elements of any evolving population are: reproduction, mutation, selection, random drift and spatial movement. These are found in every ecosystem, and although the ecological and evolutionary events occur in different time-scales their effects are intrinsically coupled (Karmon and Pilpel, 2016).

The evolutionary race is fed by novelty (Cordero and Polz, 2014). This originates from the mutations occurred during reproduction, to consequently be tested against others by natural selection. Scientists have witnessed this race through the fossil record (Yuan et al., 2005), and more recently, through phylogenetic studies

(Lutzoni et al., 2001) and experimental evolution assays (Lenski and Travisano, 1994; Lawrence et al., 2012). Remarkably, although mutations occur at random to generate a panoply of varieties (Freilich et al., 2011), they only get their quality as beneficial, harmful or neutral after the ongoing natural selection; this is known as the neutral theory of evolution (Hegreness et al., 2006).

During the evolutionary trajectory of populations, the future adequacy of a now-successful variety is uncertain, so to face a trade-off between adaptations is imminent (Ferrière et al., 2007). Furthermore, mutualistic communities are subject to strong co-evolution as each partner relies on another's performance to survive (Cordero et al., 2012). So, as selection changes, an additional trade-off between individual flexibility and mutualistic-adaptation is generated (Lawrence et al., 2012). Notoriously, due to these stresses, evolutionary rushes in response to changing selections have been observed (Hillesland et al., 2014).

The so mentioned interaction of ecological and evolutionary processes could be visualized directly in the dynamics of populations' ecological niche. To acknowledge the niche as the link between ecology and evolution may pave the way for a unified description of the eco-evolutionary dynamics (Post and Palkovacs, 2009). Particularly, in one direction such construct would allow us to analyse the influence of symbiotic interactions in the still intriguing niche constriction and vacant niches (Post and Palkovacs, 2009); and in an opposite direction to understand what evolutionary features allow the persistence of mutualism (Pelletier et al., 2009).

## 1.4 Aims

From an experimental perspective, recent experimental designs have relied more frequently on artificial systems to study the eco-evolutionary dynamics (Celiker and Gore, 2014; Escalante et al., 2014; Campbell et al., 2015). The increased controllability offered by these approaches (Momeni et al., 2011), allow us to test theoretical hypothesis more easily (Mee and Wang, 2012). To develop mathematical models of

these systems, may facilitate their design while enriching our understanding of the eco-evolutionary interplay (Porcar et al., 2013; Widder et al., 2016).

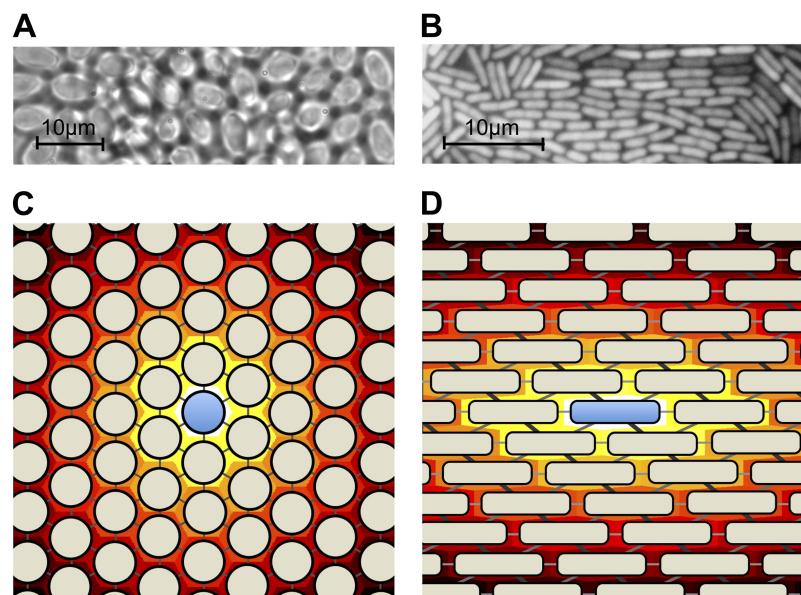
Therefore, to study the persistence of mutualism in nature we propose to develop a mathematical model based on the concept of ecological niche. Firstly, this should allow us to differentiate the ecological and evolutionary dynamics while keeping track of the simulated system; in such a way, we aim to get insights of the basic ecological principles required to engineer stable mutualisms in evolutionary times. Finally, inspired by the models of Mathiesen et al. (2011), Allen et al. (2013) and Peña et al. (2016) we aim to explore the exchange of commodities between individuals using a network description, such that, by constricting the fluxes (links), conditions of null-, constitutive- and facultative-mutualism could be analysed.

Ultimately, we want to use this model to study how stress imposed on mutualistic communities modifies their eco-evolutionary dynamics; and whether indeed depending on the nature of the interaction the niche evolution is constrained.

## Chapter 2

# Methods

### 2.1 The model



**Figure 2.1: Spatial structure in microbial communities.** (A) A culture of *S. cerevisiae* grows on a surface. (B) A culture of *E. coli* grows on a surface. (C) and (D) the graph representations of the spatial structure in A and B are shown, in both cases each cell experiences a senary neighbourhood. This figure was adapted from Allen et al. (2013).

#### 2.1.1 Ecological dynamics

Consider an ecological system like in the Figure 2.1. This depicts a microbial culture growing to form a single layer of cells. This can be modelled as a temporal and

spatial discrete system, where the space is a lattice  $\mathbf{E}|_t$  of dimensions  $(X, Y)$ , with each cell providing the same abiotic conditions but being either occupied or empty.

The organisms inhabiting these cells belong to  $NP$  populations differing in their commodities requirements (i.e. goods or services), but existing in a same niche space with  $K$  components that represent these commodities. Accounting for the natural variation of the populations, at time  $t$  the individual in every cell  $(x, y)$  has slightly different productions ( $P_{k,t}$ ) and requirements ( $R_{k,t}$ ). To satisfy the latter, productions are either maintained ( $M_{k,t}$ ) or leaked to the neighbours ( $L_{k,t}$ ):

$$M_{k,t}(x, y) + \eta L_{k,t}(x, y) = P_{k,t}(x, y) \quad (2.1)$$

Satisfying these equalities:

$$M_{k,t}(x, y) = m_{k,t}(x, y) P_{k,t}(x, y)$$

$$L_{k,t}(x, y) = l_{k,t}(x, y) P_{k,t}(x, y)$$

It can be shown that  $m_{k,t}(x, y) + \eta l_{k,t}(x, y) = 1$ , so  $0 \leq m_{k,t}(x, y) \leq 1$  and  $0 \leq l_{k,t}(x, y) \leq \frac{1}{\eta}$ . Where  $\eta = (\max N - \min N + 1)^2 - 1$ , being  $N$  the set of coordinates of neighbouring cells (e.g.  $N = \{-1, 0, 1\}$  for the immediate neighbourhood).

The production of the  $k$ -th commodity by the individual in  $(x, y)$  at  $t$ ,  $P_{k,t}(x, y)$ , is a function of its commodity requirement  $R_{k,t}(x, y)$ , a measure of its “physiological capability” (i.e. bigger organisms can produce more):

$$P_{k,t}(x, y) = f(R_{k,t}(x, y))$$

This function can have different shapes, however we will consider a positive linear relation determined by the initial conditions (intended to be a physiological constraint). The following equation is the requirement-based-unitary-production:

$$\hat{P}_{k,t}(x,y) = \left( \frac{P_{k,0}(x,y)}{R_{k,0}(x,y)} \right) R_{k,t}(x,y)$$

The effective production is then calculated knowing how much of the  $k$ -th requirement was actually satisfied:

$$P_{k,t}(x,y) = \hat{P}_{k,t}(x,y) \left( \frac{G_{k,t}(x,y)}{R_{k,t}(x,y)} \right) = \left( \frac{P_{k,0}(x,y)}{R_{k,0}(x,y)} \right) G_{k,t}(x,y) \quad (2.2)$$

Where  $G_{k,t}(x,y)$  indicates the gathering of the  $k$ -th commodity by maintained and neighbours-leaked commodities. We introduce the following definition:

$$G_{k,t}(x,y) = M_{k,t}(x,y) + \sum_{\substack{i,j \in N \\ \forall i,j \neq 0,0}} L_{k,t}(x+i,y+j)$$

So:

$$G_{k,t}(x,y) = m_{k,t}(x,y) P_{k,t}(x,y) + \sum_{\substack{i,j \in N \\ \forall i,j \neq 0,0}} l_{k,t}(x+i,y+j) P_{k,t}(x+i,y+j) \quad (2.3)$$

Then based on the commodities available for a cell, two updates from  $t$  to  $t+1$  can occur: (1) survival and (2) duplication. The following equation is introduced to describe their behaviour:

$$f_t(x,y) = \gamma_f \prod_{k=1}^K \left( \frac{1}{1 + e^{-\alpha_{f,k}(g_{k,t}(x,y) - 2^{-1}\beta_{f,k})}} \right)$$

Here  $g_{k,t}(x,y) = \frac{G_{k,t}(x,y)}{R_{k,t}(x,y)}$  is the requirement-normalized gathering of commodity  $k$ . This  $f_t(x,y)$  takes values from 0 to  $\gamma_f$  ( $\forall \gamma_f \in \{0, 1\}$ ), moving with steepness  $\alpha_{f,k}$  for commodity  $k$ , until it asymptotically approaches  $\gamma_f$  if  $g_{k,t}(x,y) = \beta_{f,k} \forall k$ . Note finally that  $\alpha_{f,k}, \beta_{f,k} \in IR$ ; being  $\beta_{f,k}$  the x-fold  $R_{k,t}$  necessary to reach  $\gamma_f$ .

Survival ( $S_t(x,y)$ ) and duplication ( $D_t(x,y)$ ) are particular functions of  $f_t(x,y)$

Variable	Notation
Populations spatial distribution	$\mathbf{E} _t$
Requirement of $k$ -th commodity at $t$	$\mathbf{R} _{k,t}$
Leakage of $k$ -th commodity at $t$	$\mathbf{l} _{k,t}$
Production of $k$ -th commodity at $t$	$\mathbf{P} _{k,t}$
Gathering of $k$ -th commodity at $t$	$\mathbf{G} _{k,t}$
Survival probability at $t$	$\mathbf{S} _t$
Duplication probability at $t$	$\mathbf{D} _t$

**Table 2.1:** Model variables and notation.

with parameters  $\alpha_{S,k}$ ,  $\beta_{S,k}$ ,  $\gamma_S$  and  $\alpha_{D,k}$ ,  $\beta_{D,k}$ ,  $\gamma_D$  respectively. So if cell  $(x,y)$  is occupied it will survive with probability  $S_t(x,y)$ , and then duplicate with probability  $D_t(x,y)$  if at least one immediate cell is empty. The values  $1 - \gamma_S$  and  $1 - \gamma_D$  correspond to the probability of no-surviving and of no-duplication, both of them independent of the commodities-gathering.

We will set  $\beta_{S,k} = 1$ , meaning that to reach the maximum survival probability  $\gamma_S$  each gathering ( $G_{k,t}$ ) must equal its requirement ( $R_{k,t}$ ). And we will focus on the case where  $\alpha_{S,k} = \alpha_S$ ,  $\alpha_{D,k} = \alpha_D$  and  $\beta_{D,k} = \beta_D$ ; so  $S_t(x,y)$  is defined by the parameters  $\alpha_S$ ,  $\gamma_S$  and  $D_t(x,y)$  by  $\alpha_D$ ,  $\beta_D$ ,  $\gamma_D$ .

$$S_t(x,y) = \gamma_S \prod_{k=1}^K \left( \frac{1}{1 + e^{-\alpha_S(g_{k,t}(x,y) - 2^{-1})}} \right) \quad (2.4)$$

$$D_t(x,y) = \gamma_D \prod_{k=1}^K \left( \frac{1}{1 + e^{-\alpha_D(g_{k,t}(x,y) - 2^{-1}\beta_D)}} \right) \quad (2.5)$$

These two equations determine the whole ecological dynamics. Starting from  $R_{k,t}(x,y)$  and  $l_{k,t}(x,y)$ ,  $S_t(x,y)$  and  $D_t(x,y)$  are mapped.

### 2.1.2 Evolutionary dynamics

We have described the state of the system at constant  $t$  so far. The transition in evolutionary time-scales occurs when a cell duplicates, allowing the daughter cells to evolve their  $K$  requirements and leakages randomly, sampling a normal distribution

with mean 0 and variance  $\sigma$ :

$$R_{k,t+1}(x,y) = R_{k,t}(x,y) + \mathcal{N}(0, \sigma_{R,k}) \quad (2.6)$$

$$l_{k,t+1}(x,y) = l_{k,t}(x,y) + \mathcal{N}(0, \sigma_{l,k}) \quad (2.7)$$

The assumption of how the evolution occurs is based on the work of Hegreness et al. (2006) about “An equivalence principle for the incorporation of favorable mutations in asexual populations”.

## 2.2 Parameters dependency

Although the parameters of the model were thought to be directly associated to *in-vitro/in-vivo* experimental measurements, an important pre-analysis is how these affect the model analytically. We introduced  $f_t(x,y)$  as a function depending only on three parameters,  $\alpha_f, \beta_f, \gamma_f \in IR$ . Moreover  $0 \leq \gamma_f \leq 1$  and  $0 \leq \alpha_f, \beta_f, g_{k,t}(x,y)$ . These ranges are useful to analyse how  $f_t(x,y)$  responds to each parameter.

### 2.2.1 Maximum probability ( $\gamma_f$ )

The effect of  $\gamma_f$  on  $f_t(x,y)$  is given by  $F_\gamma(x,y)|_t$ , with  $\alpha_f, \beta_f$  and  $g_{k,t}(x,y)$  constants:

$$F_\gamma(x,y)|_t = \frac{\partial f_t(x,y)}{\partial \gamma_f} \Big|_{\alpha_f, \beta_f, g_{k,t}(x,y)} = \prod_{k=1}^K \left( \frac{1}{1 + e^{-\alpha_f(g_{k,t}(x,y) - 2^{-1}\beta_f)}} \right)$$

So  $F_\gamma(x,y)|_t$  turns to be a constant  $0 < F_\gamma(x,y)|_t < 1, \forall \gamma_f \in \{0, 1\}$ .

It can be concluded then, that every value of  $\gamma_f$  impacts  $f_t(x,y)$  in the same way, so there are not preferable values for this parameter.

### 2.2.2 Steepness ( $\alpha_f$ )

The effect of  $\alpha_f$  on  $f_t(x,y)$  is given by  $F_\alpha(x,y)|_t$ , with  $\gamma_f, \beta_f$  and  $g_{k,t}(x,y)$  constants:

Parameter	Notation
Lattice horizontal size	$X$
Lattice vertical size	$Y$
Set of coordinates of neighbouring cells	$N$
Number of neighbouring cells	$\eta$
Number of populations	$NP$
Number of commodities	$K$
Initial $i$ -th population	$IP$
Initial production of $k$ -th commodity	$P_{k,0}$
Initial requirement of $k$ -th commodity	$R_{k,0}$
Initial leakage of $k$ -th commodity	$l_{k,0}$
Maximum survival probability	$\gamma_S$
Maximum duplication probability	$\gamma_D$
Steepness of survival probability	$\alpha_S$
Steepness of duplication probability	$\alpha_D$
Fraction to reach half-survival probability	$\beta_S$
Fraction to reach half-duplication probability	$\beta_D$
Mutation SD of $k$ -th commodity requirement	$\sigma_{R,k}$
Mutation SD of $k$ -th commodity leakage	$\sigma_{l,k}$

**Table 2.2:** Model parameters and notation.

$$\begin{aligned}
F_\alpha(x,y)|_t &= \frac{\partial f_t(x,y)}{\partial \alpha_f} \Big|_{\gamma_f, \beta_f, g_{k,t}(x,y)} \\
&= \gamma_f \prod_{k=1}^K (g_{k,t}(x,y) - 2^{-1}\beta_f) \left( \frac{e^{-\alpha_f(g_{k,t}(x,y) - 2^{-1}\beta_f)}}{\left(1 + e^{-\alpha_f(g_{k,t}(x,y) - 2^{-1}\beta_f)}\right)^2} \right)
\end{aligned}$$

This is a more complicated equation, however some qualitative aspects can be highlighted: 1) It's not strictly positive, because it will be negative if  $g_{k,t}(x,y) < 2^{-1}\beta_f$  for any  $k$ . 2) It can be zero if  $\gamma_f = 0$  or if  $g_{k,t}(x,y) = 2^{-1}\beta_f$  for any  $k$ ; the first case is senseless for our purposes, but the second does occur when  $G_{k,t}(x,y) = 2^{-1}\beta_f R_{k,t}(x,y)$  for any of the  $K$  commodities. Moreover, if we define  $z_{k,t}(x,y) = g_{k,t}(x,y) - 2^{-1}\beta_f$ , the numerator of the exponential factor cause  $F_\alpha(x,y)|_t$  to approach zero asymptotically for any large positive  $z_{k,t}(x,y)$ , and the denominator cause the same for any large negative  $z_{k,t}(x,y)$ . 3) The overall sign of

the equation depends on the number of  $z_{k,t}(x,y)$ 's being negative, so for even numbers  $F_\alpha(x,y)|_t$  will be positive, but it will be negative for odd numbers. 4) The effect of  $\alpha_f$  is only to potentiate the speed of asymptotic approach to zero, speeding up as  $K$  increases. So for sufficiently big positive values of  $\alpha_f$ ,  $F_\alpha(x,y)|_t$  will not suffer significant changes if  $\alpha_f$  is further increased; those changes occur whatsoever for positive values of  $\alpha_f$  approaching zero, leading the equation to move exponentially from  $F_\alpha(x,y)|_t \approx 0$ , to  $F_\alpha(x,y)|_t = \left(\frac{1}{4}\right)^K \gamma_f \prod_{k=1}^K z_{k,t}(x,y)$  at  $\alpha_f = 0$ .

It can be concluded then, that in order to modify  $f_t(x,y)$  the least sufficiently big values of  $\alpha_f$  are preferable (Figure 2.2).

### 2.2.3 x-fold commodity requirement ( $\beta_f$ )

The effect of  $\beta_f$  on  $f_t(x,y)$  is given by  $F_\beta(x,y)|_t$ , with  $\gamma_f$ ,  $\alpha_f$  and  $g_{k,t}(x,y)$  constants:

$$\begin{aligned} F_\beta(x,y)|_t &= \frac{\partial f_t(x,y)}{\partial \beta_f} \Big|_{\gamma_f, \alpha_f, g_{k,t}(x,y)} \\ &= \gamma_f (-2^{-1} \alpha_f)^K \prod_{k=1}^K \left( \frac{e^{-\alpha_f(g_{k,t}(x,y)-2^{-1}\beta_f)}}{\left(1+e^{-\alpha_f(g_{k,t}(x,y)-2^{-1}\beta_f)}\right)^2} \right) \end{aligned}$$

This is a similar but simpler equation with the following qualitative aspects:

1) It's not strictly positive, because it will be negative if  $K$  is odd. 2) It can be zero if  $\gamma_f = 0$  or if  $\alpha_f = 0$ , but both of them are senseless for our purposes. Nevertheless, the numerator of the exponential factor cause the equation to approach zero asymptotically for any large positive  $z_{k,t}(x,y)$ , and the denominator cause the same for any large negative  $z_{k,t}(x,y)$ . The change from positive to negative of  $z_{k,t}(x,y)$  occurs exactly at  $\beta_f = 2g_{k,t}(x,y)$ . 3) The effect of  $\beta_f$  on the equation is to move the maximum/minimum along each of the  $K$  components, going from  $F_\beta(x,y)|_t = \gamma_f (-2^{-1} \alpha_f)^K \prod_{k=1}^K \left( \frac{e^{\alpha_f g_{k,t}(x,y)}}{\left(1+e^{\alpha_f g_{k,t}(x,y)}\right)^2} \right)$  at  $\beta_f = 0$  (expected to be very close to zero because of the quadratic exponential denominator), to

$F_\beta(x,y)|_t = \left(\frac{1}{4}\right)^K \gamma_f (-2^{-1} \alpha_f)^K$  at  $\beta_f = 2g_{k,t}(x,y)$ , and finally to  $F_\beta(x,y)|_t \approx 0$  as  $\beta_f$  is further increased.

It can be concluded then, that  $\beta_f$  will modify  $f_t(x,y)$  importantly as it is expected to have values around  $\beta_f = 2g_{k,t}(x,y)$ .

### 2.2.4 Inherited deviations ( $\sigma_{R,k}$ , $\sigma_{l,k}$ )

The *error threshold* allowing the evolutionary adaptation proposed by Nowak (2006) is studied in our model changing  $\sigma_{R,k}$  and  $\sigma_{l,k}$  in the simulations.

## 2.3 Variables dependency

To maximize  $S_t(x,y)$  and  $D_t(x,y)$  let us study their most generic function  $f_t(x,y)$ , we have to analyse its change as a function of  $g_{h,t}(x,y) \forall h \in \{1, \dots, k\}$ . This is given by  $F_{g_h}(x,y)|_t$ , with  $\gamma_f$ ,  $\alpha_f$  and  $\beta_f$  constants:

$$\begin{aligned} F_{g_h}(x,y)|_t &= \frac{\partial f_t(x,y)}{\partial g_h} \Big|_{\gamma_f, \alpha_f, \beta_f} \\ &= \gamma_f \alpha_f \frac{e^{-\alpha_f(g_{h,t}(x,y)-2^{-1}\beta_f)}}{\left(1 + e^{-\alpha_f(g_{h,t}(x,y)-2^{-1}\beta_f)}\right)^2} \prod_{\substack{k=1 \\ k \neq h}}^K \left( \frac{1}{1 + e^{-\alpha_f(g_{k,t}(x,y)-2^{-1}\beta_f)}} \right) \end{aligned}$$

This equation has the following qualitative aspects: 1) It's strictly positive. 2) It can be zero if  $\gamma_f = 0$  or if  $\alpha_f = 0$ , but both of them are senseless for our purposes. Nevertheless, the numerator of the exponential factor cause the equation to approach zero asymptotically for a large positive  $z_{h,t}(x,y)$ , the denominator cause the same for a large negative  $z_{h,t}(x,y)$ , with the change from positive to negative of  $z_{h,t}(x,y)$  at  $\beta_f = 2g_{h,t}(x,y)$ . 3) The effect of  $\gamma_f$ ,  $\alpha_f$  and  $\beta_f$  on the equation is as it was described before. 4) For all the factors whose  $k \neq h$ , their value go from 0 for large negative  $z_{k,t}(x,y)$  to 1 for large positive  $z_{k,t}(x,y)$ , with  $\frac{1}{2}$  occurring at  $g_{k,t}(x,y) = 2^{-1}\beta_f$ .

It can be concluded that  $f_t(x,y)$  cannot be explicitly maximized. However  $F_{g_h}(x,y)|_t$  approaches zeros asymptotically for either large negative or posi-

tive values of  $z_{h,t}(x,y)$ , around a maximum value of  $F_{g_h}(x,y)|_t = \gamma_f \alpha_f \left(\frac{1}{2}\right)^{K+1}$  at  $z_{k,t}(x,y) = 0 \forall k \in \{0, \dots, K\}$ . To know if these are minima or maxima we calculate  $\frac{\partial}{\partial g_h} F_{g_h}(x,y)|_t$ :

$$\frac{\partial}{\partial g_h} F_{g_h}(x,y)|_t = \frac{\partial^2 f_t(x,y)}{\partial g_h^2} \Big|_{\gamma_f, \alpha_f, \beta_f} = \alpha_f F_{g_h}(x,y)|_t \frac{1 - e^{-\alpha_f(g_{h,t}(x,y) - 2^{-1}\beta_f)}}{1 + e^{-\alpha_f(g_{h,t}(x,y) - 2^{-1}\beta_f)}}$$

Now let us define  $c(x,y) = \gamma_f \prod_{\substack{k=1 \\ k \neq h}}^K \left( \frac{1}{1 + e^{-\alpha_f z_{k,t}(x,y)}} \right)$ . Such that:

$$f_t(x,y) = c(x,y) h_t(x,y)$$

$$\frac{\partial f_t(x,y)}{\partial z_{h,t}(x,y)} = c(x,y) \frac{\partial h_t(x,y)}{\partial z_{h,t}(x,y)}$$

$$\frac{\partial^2 f_t(x,y)}{\partial z_{h,t}(x,y)^2} = c(x,y) \frac{\partial^2 h_t(x,y)}{\partial z_{h,t}(x,y)^2}$$

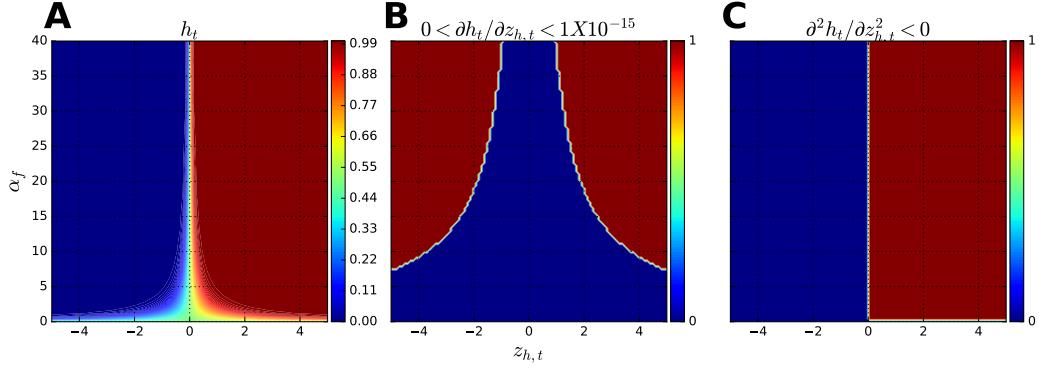
Where:

$$h_t(x,y) = \frac{1}{1 + e^{-\alpha_f z_{h,t}(x,y)}}$$

$$\frac{\partial h_t(x,y)}{\partial z_{h,t}(x,y)} = \alpha_f e^{-\alpha_f z_{h,t}(x,y)} h_t^2(x,y)$$

$$\frac{\partial^2 h_t(x,y)}{\partial z_{h,t}(x,y)^2} = \alpha_f^2 e^{-\alpha_f z_{h,t}(x,y)} \left( 1 - e^{-\alpha_f z_{h,t}(x,y)} \right) h_t^3(x,y)$$

Figure 2.2 shows that local maxima are found for  $z_{k,t}(x,y) > 0$  and that importantly these are closer to zero as  $\alpha_f$  gets bigger. Recalling that  $z_{k,t}(x,y) = \frac{G_{k,t}(x,y)}{R_{k,t}(x,y)} - 2^{-1}\beta_f$ , greater maxima appear as  $G_{k,t}(x,y)$  gets bigger. We will readdress this observation in the subsection 2.4.2.



**Figure 2.2: Maximization of the fitness function.** The simplified equations  $h_t$ ,  $\partial h_t / \partial z_{h,t}$  and  $\partial^2 h_t / \partial z_{h,t}^2$  are shown as functions of  $z_{h,t}$  and  $\alpha_f$ . (A)  $h_t > 0$  occurs mostly for  $z_{h,t} > 0$ ; however there is not a global maximum. (B) Local maxima are closer to  $z_{h,t} = 0$  for greater  $\alpha_f$ . (C) Maxima are exclusive for  $z_{h,t} > 0$ .

## 2.4 Ecological equilibrium

### 2.4.1 System of coupled equations

Starting from the definition of  $P_{k,t}(x,y)$  and substituting that of  $G_{k,t}(x,y)$ , the next equation explains the production of a  $k$ -th commodity in  $(x,y)$  at  $t$  by the conserved commodity and the influxes of neighbouring cells:

$$P_{k,t}(x,y) = \frac{P_{k,0}(x,y)}{R_{k,0}(x,y)} \left( m_{k,t}(x,y) P_{k,t}(x,y) + \sum_{\substack{i,j \in N \\ \forall i,j \neq 0,0}} l_{k,t}(x+i,y+j) P_{k,t}(x+i,y+j) \right)$$

On the other hand the consumption is defined by the outfluxes. Such that the equilibrium:  $\text{production}(x,y)|_{k,t} = \text{consumption}(x,y)|_{k,t}$  is satisfied:

$$\text{production}(x,y)|_{k,t} = \sum_{\substack{i,j \in N \\ \forall i,j \neq 0,0}} \frac{l_{k,t}(x+i,y+j)}{\frac{R_{k,0}(x,y)}{P_{k,0}(x,y)} - m_{k,t}(x,y)} P_{k,t}(x+i,y+j) = P_{k,t}(x,y)$$

$$\text{consumption}(x,y)|_{k,t} = (m_{k,t}(x,y) + \eta l_{k,t}(x,y)) P_{k,t}(x,y) = P_{k,t}(x,y)$$

The previous can be expressed as a homogeneous linear equation:

$$\left(1 - \frac{R_{k,0}(x,y)}{P_{k,0}(x,y)} - \eta l_{k,t}(x,y)\right) P_{k,t}(x,y) + \sum_{\substack{i,j \in N \\ \forall i,j \neq 0,0}} l_{k,t}(x+i,y+j) P_{k,t}(x+i,y+j) = 0 \quad (2.8)$$

$$A_{k,t}(x,y) P_{k,t}(x,y) + \sum_{\substack{i,j \in N \\ \forall i,j \neq 0,0}} A_{k,t}(x+i,y+j) P_{k,t}(x+i,y+j) = 0$$

Where  $A_{k,t}(x,y) = 1 - \frac{R_{k,0}(x,y)}{P_{k,0}(x,y)} - \eta l_{k,t}(x,y)$  and  $A_{k,t}(x+i,y+j) = l_{k,t}(x+i,y+j)$ .

This last equation can be arranged into a matrix form:

$$\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0} \quad (2.9)$$

Where  $\mathbf{A}|_{k,t}$  is a matrix of dimension  $(XY, XY)$  containing the coefficients  $A_{k,t}(x,y)$  and  $A_{k,t}(x+i,y+j)$  of each equation arranged in rows, and  $\mathbf{P}|_{k,t}$  is a vector of dimension  $(XY, 1)$  containing each  $P_{k,t}(x,y)$ .  $\mathbf{A}|_{k,t}$  is determined solely by  $l_{k,t}(x+i,y+j)$  and  $\frac{R_{k,0}(x+i,y+j)}{P_{k,0}(x+i,y+j)}$   $\forall x \in \{1, X\}$ ,  $\forall y \in \{1, Y\}$ ,  $\forall i, j \in N$  at  $k$  and  $t$  fixed. If  $\det(\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}) \neq 0$  then  $\mathbf{P}|_{k,t} = \mathbf{0}$  and the system is said to be extinct, however if  $\det(\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}) = 0$  there is at least one  $\mathbf{P}|_{k,t} \neq \mathbf{0}$  whose multiples are solutions as well. Numerical solutions are obtained using the Singular Value Decomposition (SVD),  $\mathbf{A}|_{k,t} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V}^T$ , where the matrix  $\mathbf{V}^T$  contains the null-space of  $\mathbf{A}|_{k,t}$  arranged into orthogonal column vectors, and matrix  $\boldsymbol{\Sigma}$  contains the singular values associated to each of the vectors in  $\mathbf{V}^T$  and  $\mathbf{U}$  in its diagonal.  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0}$  occurs iff  $\boldsymbol{\Sigma}$  has at least one zero singular value; if there are multiple zeros the solution is a linear combination of their associated  $\mathbf{V}^T$  vectors. In our model, the ecological equilibrium satisfying the requirements is calculated at each time-step.

Importantly, to be biologically meaningful  $\mathbf{P}|_{k,t} \geq 0$ . This implies that

$A_{k,t}(x,y)P_{k,t}(x,y) + \sum_{\substack{i,j \in N \\ \forall i,j \neq 0,0}} A_{k,t}(x+i,y+j)P_{k,t}(x+i,y+j) = 0$  iff  $A_{k,t}(x,y)$  and  $A_{k,t}(x+i,y+j)$  have opposite signs, however as  $A_{k,t}(x+i,y+j) = l_{k,t}(x+i,y+j) \geq 0$  by definition,  $A_{k,t}(x,y)$  must be strictly negative, occurring iff:

$$1 - \frac{R_{k,0}(x,y)}{P_{k,0}(x,y)} - \eta l_{k,t}(x,y) < 0$$

Also expressed as:

$$l_{k,t}(x,y) > \frac{1}{\eta} \left( 1 - \frac{R_{k,0}(x,y)}{P_{k,0}(x,y)} \right) \quad (2.10)$$

The right term of this inequality is negative if  $R_{k,0}(x,y) > P_{k,0}(x,y)$ , occurring for the “consumers” (those requiring more than its production), but not for the “producers” (those requiring less than its production). Since by definition  $0 \leq l_{k,t}(x,y) \leq \frac{1}{\eta}$ , during the evolutionary process consumers can have any value in this range, but no the producers whose range has to be then  $\frac{1}{\eta} \left( 1 - \frac{R_{k,0}(x,y)}{P_{k,0}(x,y)} \right) < l_{k,t}(x,y) \leq \frac{1}{\eta}$ . Such observation highlights the fact that our model considers the immediate consumption of commodities, avoiding any accumulation.

If the  $\mathbf{P}|_{k,t}$  fitted by the SVD contains positive and negative elements, a further numerical optimization is required to find the closest vector  $\mathbf{P}|_{k,t}^\dagger$  in the biologically feasible region (where  $\mathbf{P}|_{k,t}^\dagger \geq \mathbf{0}$ ). As long as  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \boldsymbol{\xi}|_{k,t}$ , where  $\boldsymbol{\xi}|_{k,t}$  is the vector of the residuals to fit the null vector  $\mathbf{0}$ , whenever  $\mathbf{P}|_{k,t}$  has positive and negative values,  $\boldsymbol{\xi}|_{k,t}$  can be used to point towards the direction of the optimization:

$$\text{Min } (\boldsymbol{\varepsilon}_- + \boldsymbol{\varepsilon}_+) |_{k,t}$$

$$\mathbf{A}_{\text{opt}}|_{k,t} \cdot \mathbf{P}_{\text{opt}}|_{k,t} = \boldsymbol{\xi}|_{k,t} \quad (2.11)$$

$$\mathbf{P}^\dagger|_{k,t}, \boldsymbol{\varepsilon}_-|_{k,t}, \boldsymbol{\varepsilon}_+|_{k,t} \geq 0$$

Where  $\mathbf{A}_{\text{opt}}|_{k,t} = (\mathbf{A}| - \mathbf{I}|\mathbf{I})|_{k,t}$  and  $\mathbf{P}_{\text{opt}}|_{k,t} = (\mathbf{P}^\dagger|\boldsymbol{\epsilon}_-|\boldsymbol{\epsilon}_+)|_{k,t}$ , being  $\mathbf{I}$  the identity matrix,  $\boldsymbol{\epsilon}_-|_{k,t}$  the negative and  $\boldsymbol{\epsilon}_+|_{k,t}$  the positive distances from  $\mathbf{P}|_{k,t}$  to  $\mathbf{P}^\dagger|_{k,t}$ , and  $\mathbf{P}^\dagger|_{k,t}$  the vector of productions within the biologically feasible region.

### 2.4.2 Readdressing the $S_t(x,y)$ and $D_t(x,y)$ maximization

Let us readdress the goal of maximizing  $S_t(x,y)$  and  $D_t(x,y)$ , now including the fact that  $\mathbf{P}^\dagger|_{k,t}$  is always positive. We had found that their maximization depend on the maximization of  $G_{k,t}(x,y)$ . In its matrix notation:

$$\text{Max } (\mathbf{G}|_{k,t})$$

$$\mathbf{G}|_{k,t} = \left( \mathbf{A}|_{k,t} + \mathbf{I} \cdot \text{vec} \frac{\mathbf{R}|_{k,0}}{\mathbf{P}|_{k,0}} \right) \mathbf{P}^\dagger|_{k,t}$$

Where  $\mathbf{G}|_{k,t}$  indicates the gathering of commodity  $k$  at  $t$ , and vec indicates the vectorization of a matrix resulting in a column vector. Analysing this equation we can see that as  $\mathbf{R}|_{k,0}$  and  $\mathbf{P}|_{k,0}$  are constants and  $\mathbf{P}^\dagger|_{k,t}$  depends on  $\mathbf{A}|_{k,t}$ , so the maximization depends solely on  $\mathbf{A}|_{k,t}$ .

We will focus on matrices  $\mathbf{A}|_{k,t}$  producing exact solutions to  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0}$ . SVD tackles this problem solving the equation  $\det(\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t} - \lambda \mathbf{I}) = 0$ , that can be expressed as  $\lambda(C_{XY}\lambda^{XY-1} + C_{XY-1}\lambda^{XY-2} + \dots + C_1) + C_0 = 0$  as well, to find  $XY$  eigenvalues ( $\lambda$ ) and their associated  $\mathbf{P}|_{k,t}$  eigenvectors ( $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \lambda \mathbf{P}|_{k,t}$ ). An exact solution solution exists if  $C_0 = 0$ , and as a consequence at least one  $\lambda = 0$ . What is more using the determinant equation we see that  $\det(\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t})|_{\lambda=0} = 0$  has to be true. A matrix of this kind is required to satisfy:  $\text{rank } \mathbf{A}|_{k,t} \leq XY$ , in other words to have at least one dependent row ( $\mathbf{A}|_{k,t}(i, ) = \sum_1^{XY} c_j \mathbf{A}|_{k,t}(j, ) \forall j \neq i$  where  $c_j \in IR$ ). This mathematical proof is summarized in the Appendix B.2.

Having described the generalities of  $\mathbf{A}|_{k,t}$ , let us propose a different approach. We had shown that  $\mathbf{A}|_{k,t}$  is determined by  $l_{k,t}(x,y)$  and  $\frac{R_{k,0}(x,y)}{P_{k,0}(x,y)}$ , but  $\mathbf{G}|_{k,t}$  depends

solely on  $l_{k,t}(x,y)$ . In section 2.4.4 we discuss the effect of  $l_{k,t}(x,y)$  on  $\mathbf{G}|_{k,t}$  and  $\mathbf{P}|_{k,t}$  for a symmetric spatial distribution, but first we highlight a crucial point about the  $G_{k,t}(x,y)$  derivatives: leakage to others cells is  $\frac{\partial G_{k,t}(x,y)}{\partial l_{k,t}(x,y)} = -\eta P_{k,t}(x,y) < 0$ , while leakage from other cells is  $\frac{\partial G_{k,t}(x,y)}{\partial l_{k,t}(x+i,y+j)} = P_{k,t}(x+i,y+j) > 0$ . So although receiving a  $k$ -th commodity benefits the receiver it damages the sender, therefore the maximization results from analysing the trade-off of leakages to and from others.

### 2.4.3 Exact solutions

Let us express  $\mathbf{A}|_{k,t}$  in the following way:

$$\mathbf{A}|_{k,t} = \mathbf{Y}|_{k,t} + \mathbf{\Gamma}|_{k,t} \cdot \mathbf{\Lambda}|_{k,t} - \mathbf{\Omega}|_{k,t} \quad (2.12)$$

Where:

-  $\mathbf{Y}|_{k,t}$  is a matrix of size  $(XY, XY)$  with diagonal element  $\mathbf{Y}|_{k,t}(iX + j, iX + j) = 1$  if cell  $\mathbf{E}|_t(i, j)$  is occupied or 0 if not, and 0 elsewhere.

-  $\mathbf{\Gamma}|_{k,t}$  is a matrix of size  $(XY, XY)$  whose row  $iX + j$  is the vectorization of matrix  $\mathbf{E}|_t$  focused around cell  $\mathbf{E}|_t(i, j)$ , with diagonal element  $\mathbf{\Gamma}|_{k,t}(iX + j, iX + j) = -\eta$  if  $\mathbf{E}|_t(i, j)$  is occupied or 0 if not, and non-diagonal elements either 1 or 0 indicating its living neighbours.

-  $\mathbf{\Lambda}|_{k,t}$  is a matrix of size  $(XY, XY)$ :  $\mathbf{\Lambda}|_{k,t} = \mathbf{I}$ ,  $\text{Diag}(\mathbf{\Lambda}|_{k,t}) = \text{vec } \mathbf{I}|_{k,t}$ .

-  $\mathbf{\Omega}|_{k,t}$  is a matrix of size  $(XY, XY)$ :  $\mathbf{\Omega}|_{k,t} = \mathbf{I}$ ,  $\text{Diag}(\mathbf{\Omega}|_{k,t}) = \text{vec } \frac{\mathbf{R}|_{k,0}}{\mathbf{P}|_{k,0}}$ .

If we replace this new expression in the production equation  $(\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0})$  and use the distributive property of the inner product:

$$\mathbf{Y}|_{k,t} \cdot \mathbf{P}|_{k,t} + (\mathbf{\Gamma}|_{k,t} \cdot \mathbf{\Lambda}|_{k,t}) \cdot \mathbf{P}|_{k,t} - \mathbf{\Omega}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0}$$

Now rewriting the gathering ( $\mathbf{G}|_{k,t}$ ) in this notation:

$$\mathbf{G}|_{k,t} = \mathbf{Y}|_{k,t} \cdot \mathbf{P}|_{k,t} + (\mathbf{\Gamma}|_{k,t} \cdot \mathbf{\Lambda}|_{k,t}) \cdot \mathbf{P}|_{k,t}$$

Requesting  $\mathbf{G}|_{k,t}$  to equal the x-fold requirements  $\Psi|_{k,t}$ , where  $\Psi|_{k,t} = \beta_f \mathbf{R}|_{k,t}$  (being  $\beta_f$  either  $\beta_S$  or  $\beta_D$ ), and rearranging:

$$(\Gamma|_{k,t} \cdot \Lambda|_{k,t}) \cdot \mathbf{P}|_{k,t} = \Psi|_{k,t} - \Upsilon|_{k,t} \cdot \mathbf{P}|_{k,t}$$

Substituting this equation in the production's:

$$\Psi|_{k,t} - \Omega|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0}$$

$$\Omega|_{k,t} \cdot \mathbf{P}|_{k,t} = \Psi|_{k,t} \quad (2.13)$$

We can solve this equation to find  $\mathbf{P}|_{k,t}$  and then rearrange the gatherings equation to find the leakages ( $\mathbf{l}|_{k,t}$ ):

$$\Phi|_{k,t} = \mathbf{I}, \text{Diag}(\Phi|_{k,t}) = \mathbf{P}|_{k,t}$$

$$\mathbf{l}|_{k,t} = \text{Diag}(\Lambda|_{k,t})$$

$$(\Gamma|_{k,t} \cdot \Phi|_{k,t}) \cdot \mathbf{l}|_{k,t} = \Psi|_{k,t} - \Upsilon|_{k,t} \cdot \mathbf{P}|_{k,t} \quad (2.14)$$

Solving  $\mathbf{l}|_{k,t}$ : this is a function of the constants  $R_{k,0}$ ,  $P_{k,0}$  and  $\beta_f$ , the variable  $\mathbf{R}|_{k,t}$  and their spatial distribution  $\mathbf{E}|_t$ .

#### 2.4.4 An insight into symmetric spatial distributions

A first approach to understand the effect of  $l_{k,t}(x,y)$  on  $\mathbf{G}|_{k,t}$  is to have a symmetric spatial distribution. Suppose a spatial distribution with two populations exists, where one is a consumer and other a producer. Each consumer is surrounded by  $\eta_{CP}$  producers,  $\eta_{CC}$  consumers and  $\eta_{CE}$  empty cells, while each producer is surrounded

by  $\eta_{PP}$  producers,  $\eta_{PC}$  consumers and  $\eta_{PE}$  empty cells. Such that:

$$\eta = \eta_{CC} + \eta_{CP} + \eta_{CE} = \eta_{PC} + \eta_{PP} + \eta_{PE}$$

If we denote  $l_C, \left(\frac{R_0}{P_0}\right)_C$  and  $l_P, \left(\frac{R_0}{P_0}\right)_P$  as the leakage and requirement/production ratio for consumer and producer respectively, the following equation states the relation of  $l_C$  and  $l_P$  resulting in an exact solution to  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0}$ :

$$\frac{\eta_{PC}\eta_{CP}}{\left(\left(\frac{R_0}{P_0}\right)_P - 1\right)\left(\left(\frac{R_0}{P_0}\right)_C - 1\right)} l_P l_C = \left( \frac{\eta_{PC} + \eta_{PE}}{\left(\frac{R_0}{P_0}\right)_P - 1} l_P + 1 \right) \left( \frac{\eta_{CP} + \eta_{CE}}{\left(\frac{R_0}{P_0}\right)_C - 1} l_C + 1 \right) \quad (2.15)$$

A case satisfying the previous and requiring the gathering of consumers and producers to be equal ( $G_C = G_P$ ), results in the following equation:

$$\left(\frac{R_0}{P_0}\right)_C ((\eta_{CP} + \eta_{PC} + \eta_{PE}) l_P - 1) = \left(\frac{R_0}{P_0}\right)_P ((\eta_{CP} + \eta_{PC} + \eta_{CE}) l_C - 1) \quad (2.16)$$

The derivation of these equations that are general for any symmetric spatial distribution is shown in the Appendix B.1. Two particular cases may be mentioned:

1) When the space is saturated ( $\eta_{CE} = \eta_{PE} = 0$ ) the equations become:

$$0 = \frac{\eta_{PC}}{\left(\frac{R_0}{P_0}\right)_P - 1} l_P + \frac{\eta_{CP}}{\left(\frac{R_0}{P_0}\right)_C - 1} l_C + 1$$

$$\left(\frac{R_0}{P_0}\right)_C ((\eta_{CP} + \eta_{PC}) l_P - 1) = \left(\frac{R_0}{P_0}\right)_P ((\eta_{CP} + \eta_{PC}) l_C - 1)$$

Which for former implies that  $l_P$  and  $l_C$  are linearly dependent and that  $\frac{\eta_{PC}}{\left(\frac{R_0}{P_0}\right)_P - 1} l_P > \frac{\eta_{CP}}{\left(\frac{R_0}{P_0}\right)_C - 1} l_C$ . The latter indicates this relation depends only on the parameters  $\left(\frac{R_0}{P_0}\right)_P$  and  $\left(\frac{R_0}{P_0}\right)_C$ , moreover knowing that  $\left(\frac{R_0}{P_0}\right)_C > \left(\frac{R_0}{P_0}\right)_P$  then  $l_P < l_C$ .

2) Whenever any two of these equalities exist:  $\eta_{CP} = \eta_{PP}$ ,  $\eta_{PC} = \eta_{CC}$ ,  $\eta_{PE} = \eta_{CE}$ , the following holds true:

$$-\frac{\eta_{CE}\eta}{\left(\left(\frac{R_0}{P_0}\right)_P - 1\right)\left(\left(\frac{R_0}{P_0}\right)_C - 1\right)}l_P l_C = \frac{\eta_{CC} + \eta_{CE}}{\left(\frac{R_0}{P_0}\right)_P - 1}l_P + \frac{\eta_{CP} + \eta_{CE}}{\left(\frac{R_0}{P_0}\right)_C - 1}l_C + 1$$

$$\left(\frac{R_0}{P_0}\right)_C (\eta l_P - 1) = \left(\frac{R_0}{P_0}\right)_P (\eta l_C - 1)$$

Which for the former as the left of the equation is strictly positive, this implies that  $\frac{\eta_{CC} + \eta_{CE}}{\left(\frac{R_0}{P_0}\right)_P - 1}l_P < \frac{\eta_{CP} + \eta_{CE}}{\left(\frac{R_0}{P_0}\right)_C - 1}l_C + 1$ ; while for the latter  $l_P = l_C = \frac{1}{\eta}$ .  
This last case represents the mean-field scenario of our model.

## 2.5 Evolutionary equilibrium

Each time-step of the model allows populations to explore their niche ( $R_{k,t}$  space). After some generations, if extinction is averted, the convergence to an evolutionary steady state might happen, depending on the response of individuals to deal with their physiological ( $R_{k,0}$ ,  $P_{k,0}$ ) and genetic constraints ( $R_{k,t}$ ,  $l_{k,t}$ ). To investigate how these responses originate and maintain mutualism is the objective of this project.

To contrast our results we will use the categories of niche evolution proposed by Tingley et al. (2009). According to them if comparing the historic state to the present state of a niche, three evolutionary dynamics might have been observed: 1) the *null model* if the centroid of the fundamental niche is conserved and the realized niches are occupied and emptied randomly; 2) the *static model* if the centroid is conserved but the realized niches are occupied and emptied at the borders; 3) the *dynamic model* if the centroid is changed and the realized niches are occupied and emptied in the regions of the contraction and the expansion respectively.

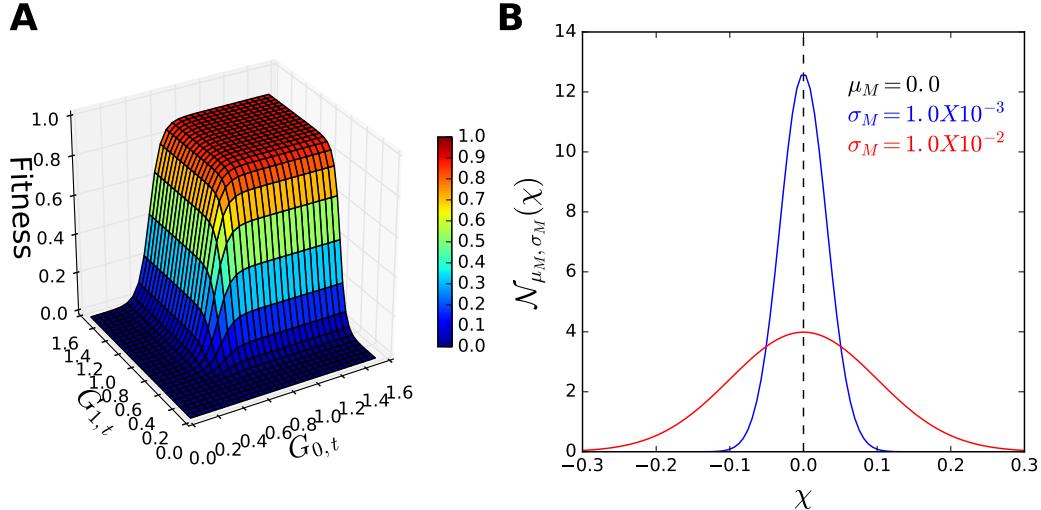
## Chapter 3

# Results

### 3.1 Numerical implementation

In this thesis we introduce a mathematical model to study the evolutionary dynamics of mutualism through the ecological niche. As we assert in Section 1, survival of individuals in nature is only possible if their commodities requirements are satisfied. If the environment does not provide a sufficient amount of these commodities, a mutualistic partnership may do it. This has proved to make some mutualisms successful (Yuan et al., 2005; Cordero et al., 2012). Thus, we hypothesize that under onerous conditions, if a community sustains its survival through a fulfilling mutualism, this interaction may constrain the further evolution of partner's niche.

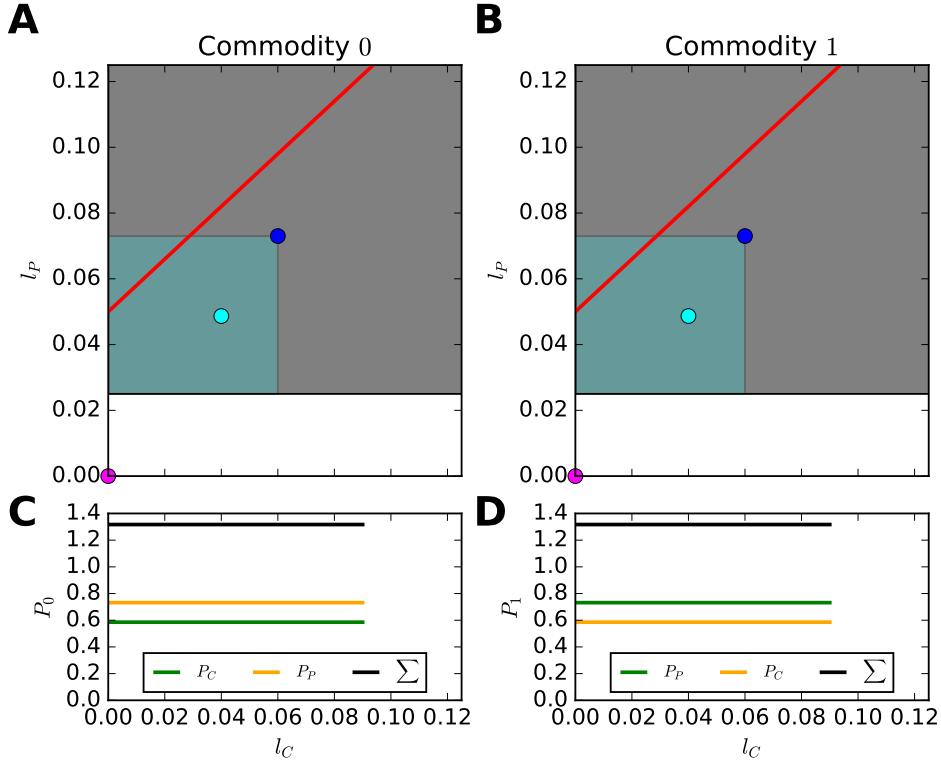
The model, that is thoroughly described in Section 2, has been implemented numerically to analyse the evolution of mutualism in a robust way (Appendix A). Starting from a set of parameters and initial conditions (Table 2.2), a spatial distribution of populations  $\mathbf{E}|_0$  is generated. At each time-step  $t$ , individuals have certain requirements  $\mathbf{R}|_{k,t}$  and leakages  $\mathbf{I}|_{k,t}$  of  $K$  commodities, these are employed to specify a system of homogeneous equations,  $\mathbf{A}|_{k,t}$ , that is solved to calculate the production of commodities  $\mathbf{P}|_{k,t}$  (equation 2.9). Using  $\mathbf{P}|_{k,t}$  the gathering of commodities by individuals  $\mathbf{G}|_{k,t}$  is calculated, finally resulting in the survival  $\mathbf{S}|_t$  and duplication  $\mathbf{D}|_t$  probabilities (Figure 3.3) that will define  $\mathbf{E}|_t$ ,  $\mathbf{R}|_{k,t}$  and  $\mathbf{I}|_{k,t}$  at the next time-step.



**Figure 3.1: Model assumptions.** **(A)** A sigmoid function  $f_t$  is assumed to describe the fitness response of individuals to the gathering of commodities ( $\mathbf{G}|_{k,t}$ ). Such  $f_t$  is parametrized by a maximum value  $\gamma_f$ , a steepness  $\alpha_f$ , and a fraction to reach half-maximum value  $\beta_f$ . This figure shows  $f_t$  for two commodities. **(B)** In agreement with the neutral theory of evolution where most mutations are neutral and the remaining either positive or negative, mutations are sampled from a normal probability density function with mean zero and standard deviation  $\sigma_M$ .

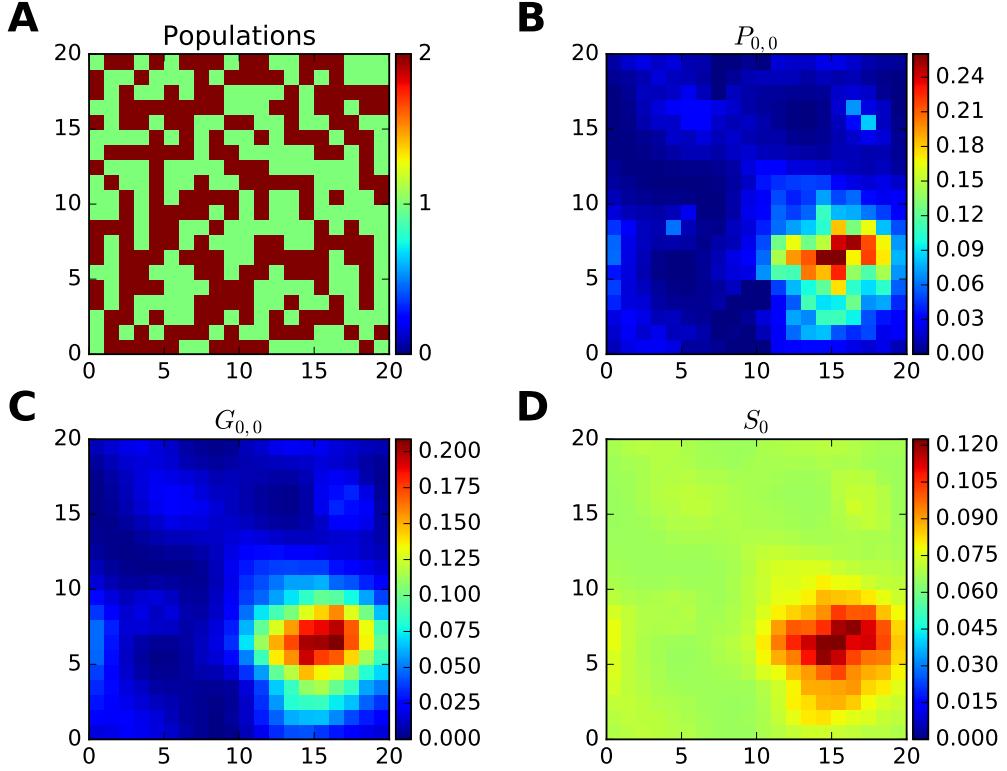
Mutations to  $\mathbf{R}|_{k,t}$  and  $\mathbf{I}|_{k,t}$  occur during duplication, and importantly as well as a sigmoid function to describe  $\mathbf{S}|_t$  (equation 2.4) and  $\mathbf{D}|_t$  (equation 2.5), in agreement with the neutral theory of evolution (Hegreness et al., 2006), the mutation density functions of  $\mathbf{R}|_{k,t}$  (equation 2.6) and  $\mathbf{I}|_{k,t}$  (equation 2.7) are assumed to be normal (Figure 3.1). These variables are summarized in Table 2.1.

This material balance approach allow us to study different categories of mutualism by changing the fluxes given by the leakages. In the simplest case mutualism is *null* ( $\mathbf{I}|_{k,t} = \mathbf{0}$ ), so recalling equation 2.3 survival and duplication only depend on the capacity of individuals to satisfy their own requirements. Then, in agreement with a genetic engineering terminology, other mutualisms can be either: *constitutive*, if individual leakages are fixed at  $t$ , or *facultative*, if aiming to maximize their survival, individual leakages adapt to their neighbours'. Figure 3.2 portrays these mutualisms for a system of two populations exchanging two commodities. There each population specializes in the production of one commodity. By catego-



**Figure 3.2: Ecological parameters feasible region.** As stated in Section 2.4 if  $R_{k,0}$  and  $P_{k,0}$  are given, the ecological equilibrium is determined by  $l_{k,t}$ . A system of two populations exchanging two commodities is shown: while for one commodity a population is a consumer ( $R_{0,0} = 0.5, P_{0,0} = 0.4$ ) another is a producer ( $R_{0,0} = 0.4, P_{0,0} = 0.5$ ), being these roles inverted for a second commodity ( $R_{1,0} = 0.4, P_{1,0} = 0.5; R_{1,0} = 0.5, P_{1,0} = 0.4$ ). Focusing on commodity 0, (A) shows in gray the biologically feasible region of leakage from producers ( $l_P$ ) and consumers ( $l_C$ ) specified by inequality 2.10, then the null-, constitutive- and facultative-mutualisms as the pink, blue and cyan dots, where the cyan area limits the facultative mutualism region; and the red line shows the parameters resulting in the exact solution of a homogeneous spatial distribution (equation 2.15). (C) The production of consumer ( $P_C$ ) and producer ( $P_P$ ) associated to the red line in A are indicated. These analyses are shown for commodity 1 in (B) and (D).

rizing each population either as consumer or as producer, the parameters space is divided into the regions specified by equations 2.10 and 2.15. This notion, that can be extended to a system of  $NP$  populations exchanging  $K$  commodities, reveal us several characteristics of the expected dynamics. Firstly, when  $\frac{R_{k,0}}{P_{k,0}} = 1$  for the producer any value in the range  $0 \leq l_P \leq \frac{1}{\eta}$  is biologically feasible, but instead as  $\frac{R_{k,0}}{P_{k,0}}$  gets lower, as consequence of a more adept producer, this region reduces (equation



**Figure 3.3: Ecological equilibrium.** (A) Starting from a random spatial configuration, populations inhabit a surface with periodic borders. (B) Individuals interact with their  $\eta$  immediate neighbours to produce  $K$  commodities. (C) Productions are exchanged through their leakage to other individuals whom gather them. (D) Based on such gatherings a fitness function is evaluated resulting in survival and duplication probabilities for the next time-step. The parameters used for both populations were:  $IP = 200$ ,  $\eta = 8$ ,  $l_{0,0} = l_{1,0} = 0.125$ ,  $\gamma_S = 0.9$ ,  $\alpha_S = 2$ ; and for population 0:  $R_{0,0} = P_{1,0} = 0.5$ ,  $R_{1,0} = P_{0,0} = 0.4$ , and population 1:  $R_{0,0} = P_{1,0} = 0.4$ ,  $R_{1,0} = P_{0,0} = 0.5$ .

2.10). Secondly, the subspace of exact solutions for symmetric spatial distributions is very limited (equation 2.15), therefore we do not expect these distributions to favourably sustain mutualism. Thirdly, while null mutualism is a solution for any spatial distribution, constitutive mutualism requires very specific leakages for every distribution, therefore this may be utterly susceptible to extinction. Finally, facultative mutualism has a greater range, thus it is expected to be amenable for different spatial distributions and consequently resilient during evolutionary times.

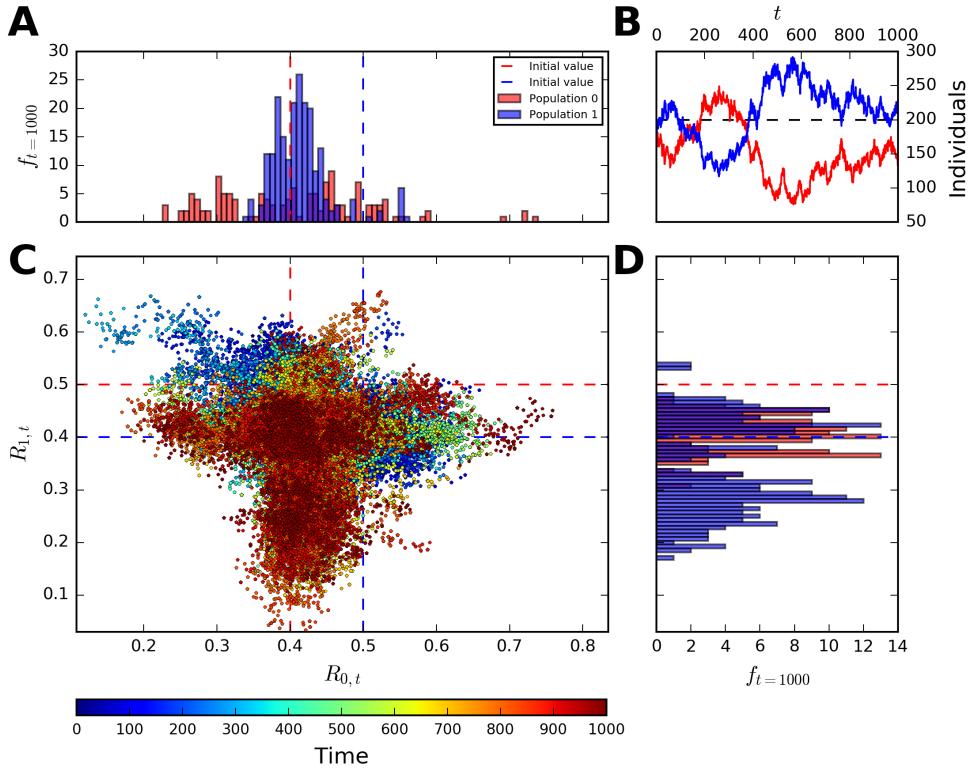
In order to simulate the different mutualisms, our model was coded into the scripts shown in Appendix A. A set of main functions accounts for each stage of the

simulation: the initial state (*SYS\_IS*), the current state (*SYS\_CS*) and the transition state (*SYS\_TS*). After the spatial distribution and variables are initialized by *SYS\_IS*, *SYS\_CS* and *SYS\_TS* are iterated. Apart from the numerical parameters, some string parameters are required by these functions. First, the *geometry* refers to the spatial distribution of individuals at the initial state, being either *homogeneous* for a distribution where all individuals per population have the same neighbouring cells (subsection 2.4.4) or *heterogeneous* for a random distribution. Then, *mutualism* can be either *null*, *constitutive* or *facultative*. Finally, *mediation* refers to the exchange of commodities when mutualism is facultative, being either *unmediated* if leakages are constrained as in Figure 3.2 or *mediated* if these are not constrained. An insight of the biological and numerical meaning of mediation is provided in subsection 3.4.

Although our code is sufficiently versatile to simulate a notorious variety of scenarios, due to time constraints we focused on a subset of salient biological cases. We proposed to simulate a mutualistic community of two populations ( $NP = 2$ ) exchanging two commodities ( $K = 2$ ). These populations grow on a  $20 \times 20$  lattice with periodic boundaries, where each cell interacts with its 8 immediate neighbours. At  $t = 0$  individuals of each population have identical requirements ( $R_{k,0}$ ), leakages ( $l_{k,0}$ ) and productions ( $P_{k,0}$ ), but while population 0 underproduces commodity 0 ( $R_{0,0} = 0.5$ ,  $P_{0,0} = 0.4$ ), population 1 overproduces it ( $R_{0,0} = 0.4$ ,  $P_{0,0} = 0.5$ ); being inverted for commodity 1. Such values portray a state where mutualism can favour survival, but whose absence does not cause the community collapse. The maximum survival and duplication probabilities are  $\gamma_S = \gamma_D = 0.9$ , requiring 0.5 times more commodities to duplicate than to survive ( $\beta_D = 1.5$ ). The initial leakages ( $l_{k,0}$ ) and the evolvabilities ( $\sigma_{R,k}$ ,  $\sigma_{l,k}$ ) are modified for each particular mutualism.

## 3.2 Null mutualism

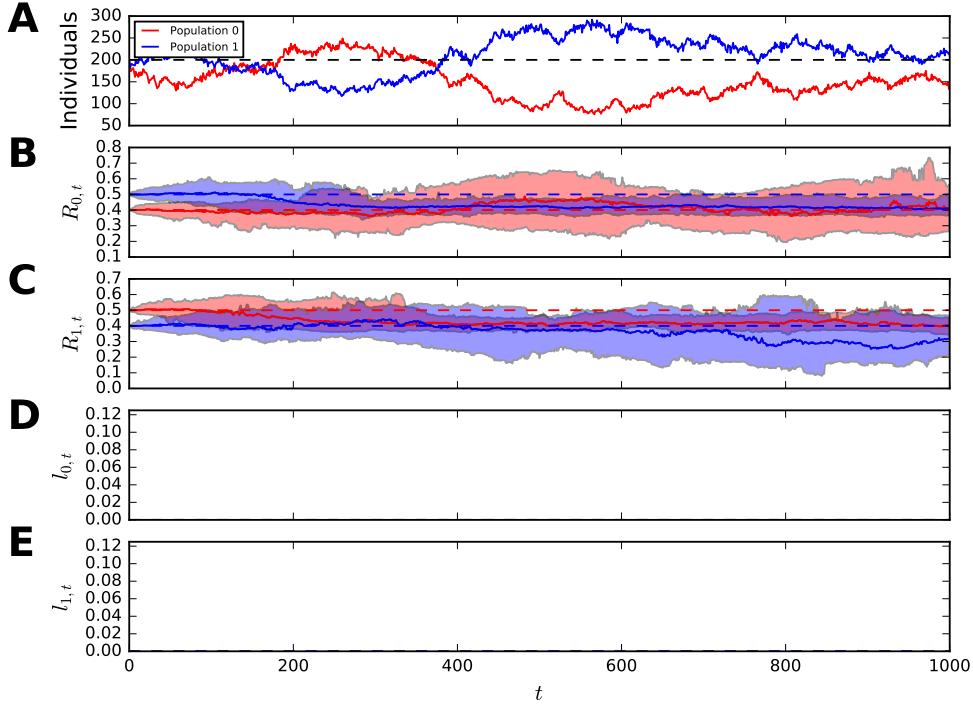
If mutualism is null, then commodities are not exchanged,  $l_{k,t} = 0$ , and consequently individual fitness only depends on oneself. Thereafter, this is the baseline to com-



**Figure 3.4: Niche evolution: null mutualism (fixed leakage).** Starting from a random spatial configuration the evolution of two populations producing two commodities during 1000 time-steps is shown. Individuals neither can self-sustain their requirements nor evolve their leakages. **(B)** The populations through time are shown. **(C)** The  $R_{0,t}$  and  $R_{1,t}$  of all individuals through time is collapsed. The distribution of **(A)**  $R_{0,t}$  and **(D)**  $R_{1,t}$  at  $t = 1000$  is compared with their values at  $t = 0$ . The parameters for both populations were:  $X = Y = 20$ ,  $IP = 200$ ,  $\eta = 8$ ,  $l_{0,0} = l_{1,0} = 0$ ,  $\gamma_S = \gamma_D = 0.9$ ,  $\alpha_S = \alpha_D = 20$ ,  $\beta_D = 1.5$ ,  $\sigma_{R,0} = \sigma_{R,1} = 1 \times 10^{-2}$ ,  $\sigma_{l,0} = \sigma_{l,1} = 0$ ; and for population 0:  $R_{0,0} = P_{1,0} = 0.5$ ,  $R_{1,0} = P_{0,0} = 0.4$ , and population 1:  $R_{0,0} = P_{1,0} = 0.4$ ,  $R_{1,0} = P_{0,0} = 0.5$ .

pare our subsequent results. Two possible subcases can occur, either leakage is physiologically constrained, so  $\sigma_{l,k} = 0$ , or leakage is physiologically possible but dysfunctional, so  $\sigma_{l,k} \neq 0$ . We refer to these possibilities as “*fixed leakage*” and “*loose leakage*” respectively. To study the evolutionary plasticity of communities, two values of  $\sigma_{l,k}$ , as in Figure 3.1 B, were used (Figures 3.4, 3.6, and C.1, C.3).

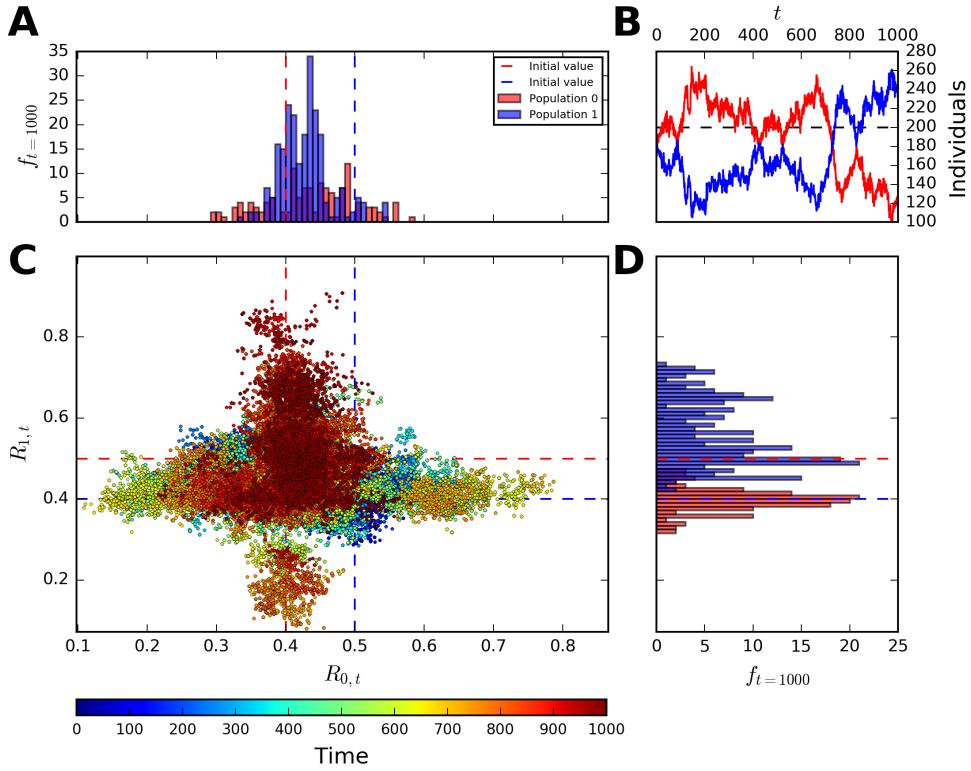
Having set a 1,000 time-steps simulation we found out that our computational code carried out the simulations without numerical shortcomings, and that populations were able to sustain their survival. Expectedly, to set out the leakage as fixed



**Figure 3.5: Variables evolution: null mutualism (fixed leakage).** (A) The populations through time along with the evolution of (B)  $R_{0,t}$ , (C)  $R_{1,t}$ , (D)  $l_{0,t}$  and (E)  $l_{1,t}$  of the simulation in Figure 3.4 are shown. The starting values are indicated by the dashed lines, the median by the continuous lines and the 5th to 95th percentile by the filled areas.

or loose, caused this variable not to be under selection as can be seen by its stillness in Figures 3.5 D-E and C.2 D-E, and its erratic evolution in Figures 3.7 D-E and C.4 D-E. The value of  $\sigma_{R,k}$  played a crucial role in the exploration of niche, so, independently of the leakage constraint, when  $\sigma_{R,k} = 1 \times 10^{-2}$  more spots within the  $R_{k,t}$  space had been tested (Figures 3.4 C and 3.6 C) when compared to  $\sigma_{R,k} = 1 \times 10^{-3}$  (Figures C.1 C and C.3 C). Either way, populations seem to have wandered around the niche space, eventually overlapping other population's niche for  $\sigma_{R,k} = 1 \times 10^{-2}$  (Figures 3.4 A, D and 3.6 A, D), and following a directionality as the median of  $R_{k,t}$  remained almost constant (Figures 3.5 B-C, 3.7 B-C, C.2 B-C and C.4 B-C).

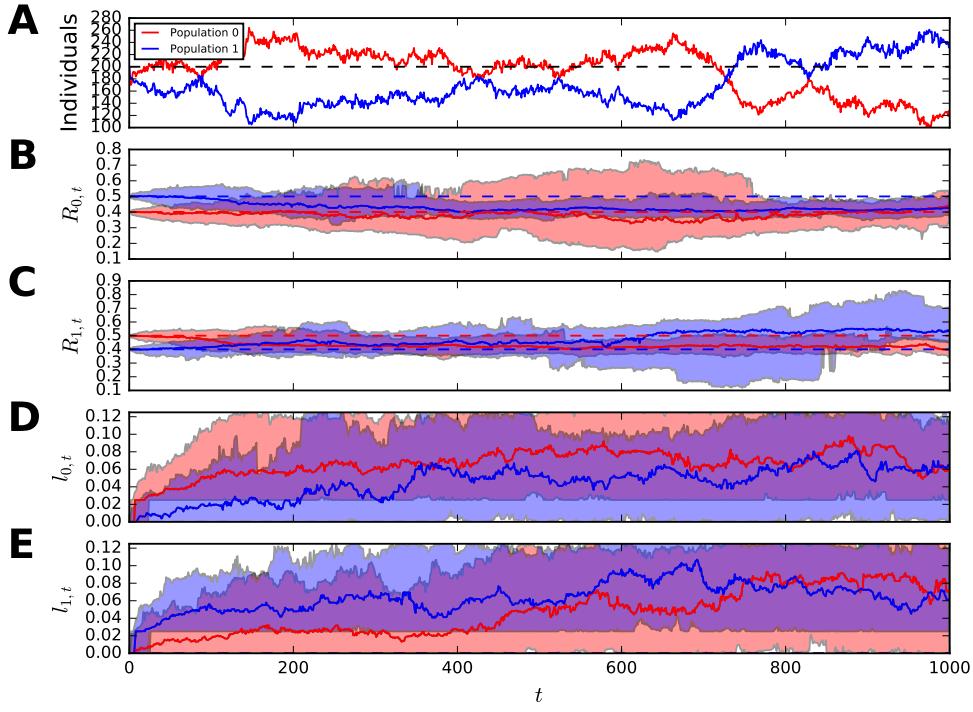
In every case, the population sizes changed drastically through time (Figures 3.4 B, 3.6 B, C.1 B and C.3 B). This seems to be related to the changes occurred



**Figure 3.6: Niche evolution: null mutualism (loose leakage).** This is a similar system to Figure 3.4, where individuals cannot self-sustain their requirements but now can evolve their leakages; nonetheless these are not functional and therefore not under selection. The same parameters as in Figure 3.4 were used, but instead  $\sigma_{l,0} = \sigma_{l,1} = 1 \times 10^{-2}$ . The notation is concomitant with Figure's 3.4.

to  $R_{k,t}$ . The  $R_{k,t}$  of the population with greater  $R_{k,0}$  moved away from its initial value until it was equal or smaller than another population's  $R_{k,0}$ . Furthermore, such variable appears to have been under selection as it was notoriously less spread from its median value through time (Figures 3.5 A-C and 3.7 A-C).

These results illustrate the feasibility of our model to track the ecological and evolutionary dynamics of an evolving community. In this subsection we enquired the evolution of  $l_{k,t}$  under free-selection conditions (Figures 3.5, 3.7, C.2 and C.4), and the effect of  $R_{k,t}$ 's evolvability (Figures 3.4, 3.6, C.1 and C.3). Regarding last point, notably, although the levels proposed differ greatly ( $\sigma_{R,k} = 1 \times 10^{-2}$  and  $\sigma_{R,k} = 1 \times 10^{-3}$ ), both communities were able to thrive during at least 1,000 time-steps, so an error threshold effect (Nowak, 2006) was not observed whatsoever.

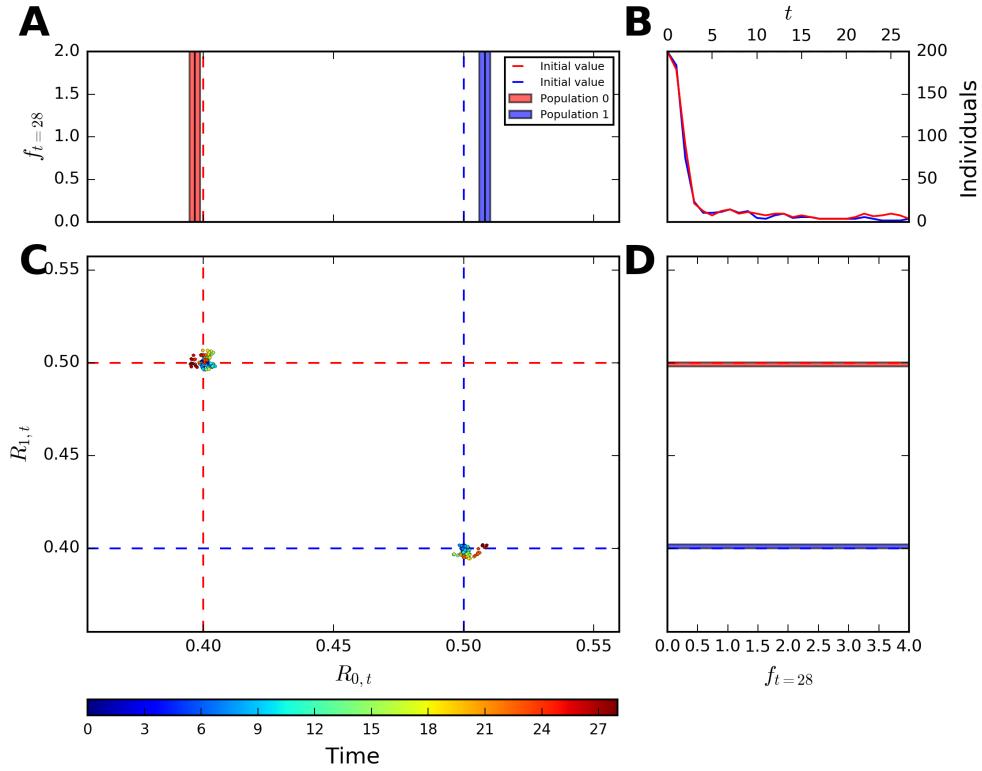


**Figure 3.7: Variables evolution: null mutualism (loose leakage).** The evolution of variables of the simulation in Figure 3.6 are shown. The notation is concomitant with Figure's 3.5.

### 3.3 Constitutive mutualism

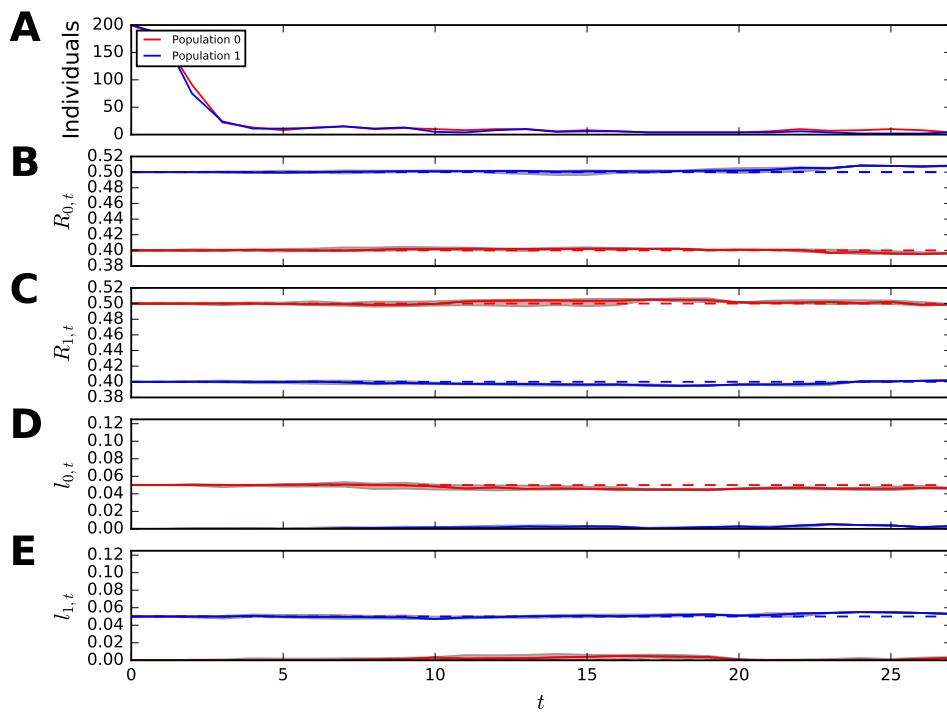
If mutualism is constitutive, then commodities are exchanged, and consequently individual fitness now depends on the community. In this case, at each time-step individuals leak commodities at a fixed value that is independent of their neighbourhood. Based on the analyses in subsections 2.4.4 and B.2 we know that leakages have to be suited to every spatial configuration. Therefore, the initial distribution,  $\mathbf{E}|_0$ , is critical to find a  $\mathbf{l}|_{k,0}$  resulting in a viable community. Under the premise of having two populations, only a symmetric distribution, where individuals have the same neighbourhood, will be viable. This contrasts with null mutualism where both homogeneous and heterogeneous distributions are effectively viable.

We proposed a saturated spatial distribution where every cell is surrounded by 4 individuals of each population (Figure 3.8). So based on equation 2.15, ini-



**Figure 3.8: Niche evolution: constitutive mutualism.** This is a similar system to Figure 3.4, where individuals cannot self-sustain their requirements but whose leakages are evolvable, being therefore under selection. The same parameters as in Figure 3.4 were used, but instead for both populations:  $\sigma_{R,0} = \sigma_{R,1} = \sigma_{l,0} = \sigma_{l,1} = 1 \times 10^{-3}$ ; and for population 0:  $l_{0,0} = 0.05$  and population 1:  $l_{1,0} = 0.05$ . The spatial distribution at  $t = 0$  was homogeneous. The notation is concomitant with Figure's 3.4.

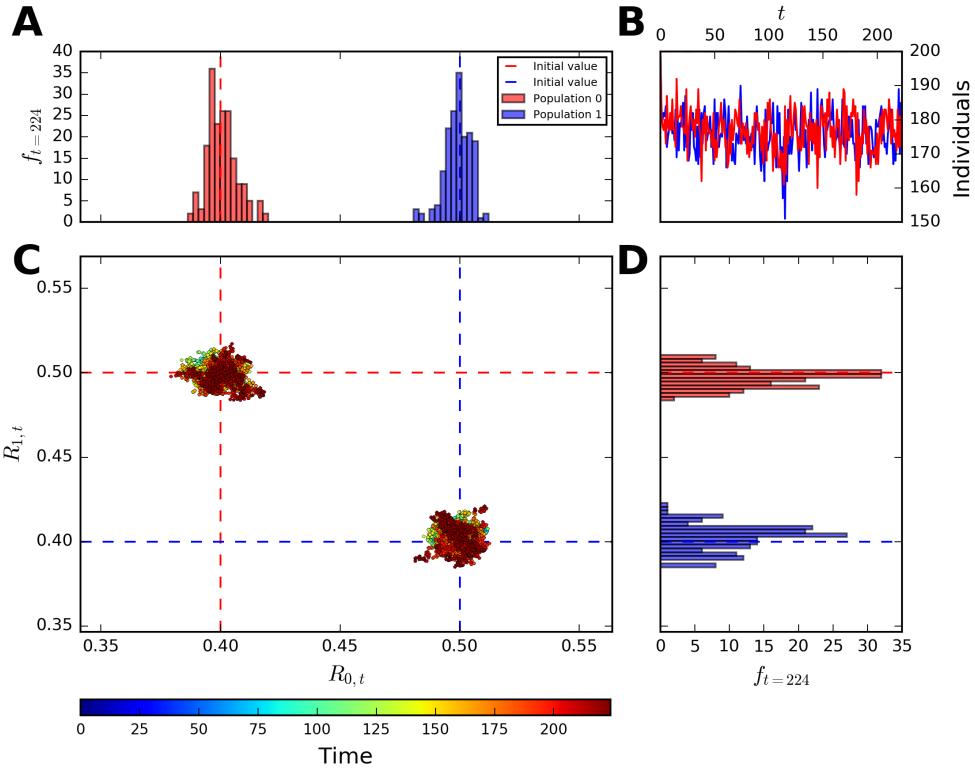
tial leakages for population 0 ( $l_{0,0} = 0.05$ ,  $l_{1,0} = 0$ ) and population 1 ( $l_{0,0} = 0$ ,  $l_{1,0} = 0.05$ ) were calculated. Initially, like in the null-mutualism scenario a high evolvability was proposed ( $\sigma_{R,k} = \sigma_{l,k} = 1 \times 10^{-2}$ ), however as such communities collapsed immediately,  $\sigma_{R,k} = \sigma_{l,k} = 1 \times 10^{-3}$  were used instead. Nonetheless, the simulated communities still collapsed very rapidly. Just after the second time-step, caused by spatial heterogeneities (empty cells and evolved  $R_{k,t}$ ,  $l_{k,t}$ ), populations sizes fell-off close to extinction (Figure 3.8 B). Survivors found themselves in locations where a suitable mutualism had occurred by chance. These continued evolving not far from their initial niche (Figures 3.8 A, C-D and 3.9), until their extinction was caused by the randomness of the survival probability.



**Figure 3.9: Variables evolution: constitutive mutualism.** The evolution of variables of the simulation in Figure 3.8 are shown. The notation is concomitant with Figure's 3.5.

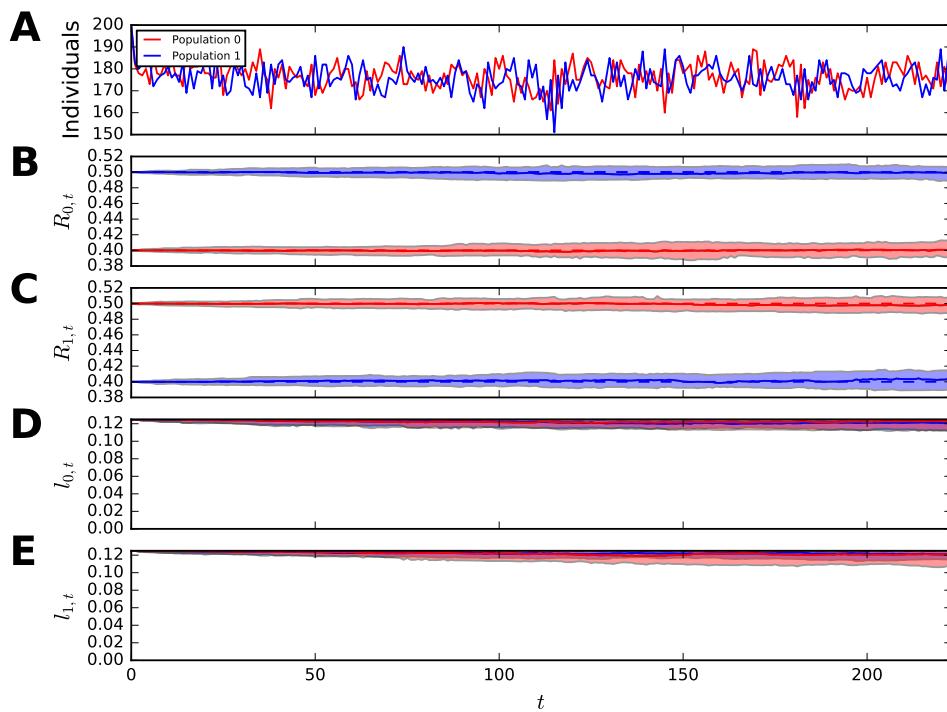
## 3.4 Facultative mutualism

If mutualism is facultative, then commodities are exchanged, and consequently individual fitness depends on the community as well. Now, however, at each time-step, in order to maximize the communal fitness, individual leakages are function of their neighbours' leakage. This flexibility allows the community to be responsive to their changing spatial structure and variables evolution. Nevertheless, this adaptation has to occur within physiological limits. Therefore, we regarded  $l_{k,t}$  as the maximum physiological leakage of individuals (Figure 3.2). In addition, in subsection 2.4.1 we found that  $l_{k,t}$  has lower and upper boundaries. The inequality 2.10, where the values defining whether a population is either a producer ( $R_{k,0} < P_{k,0}$ ) or a consumer ( $R_{k,0} > P_{k,0}$ ), determines the lower boundary. Whereas according to equation 2.1, to satisfy  $M_{k,t}, L_{k,t}, P_{k,t} \geq 0$ , the upper boundary has to be  $1/\eta$ .



**Figure 3.10: Niche evolution: facultative mutualism (unmediated).** This is a similar system to Figure 3.4, where individuals cannot self-sustain their requirements but whose leakages are evolvable and ecologically adaptable ( $\max l_{k,t} \leq 1/\eta$ ), being therefore under selection. The same parameters as in Figure 3.4 were used, but instead for both populations:  $\sigma_{R,0} = \sigma_{R,1} = \sigma_{l,0} = \sigma_{l,1} = 1 \times 10^{-3}$  and  $l_{0,0} = l_{1,0} = 0.125$ . The spatial distribution at  $t = 0$  was homogeneous. The notation is concomitant with Figure's 3.4.

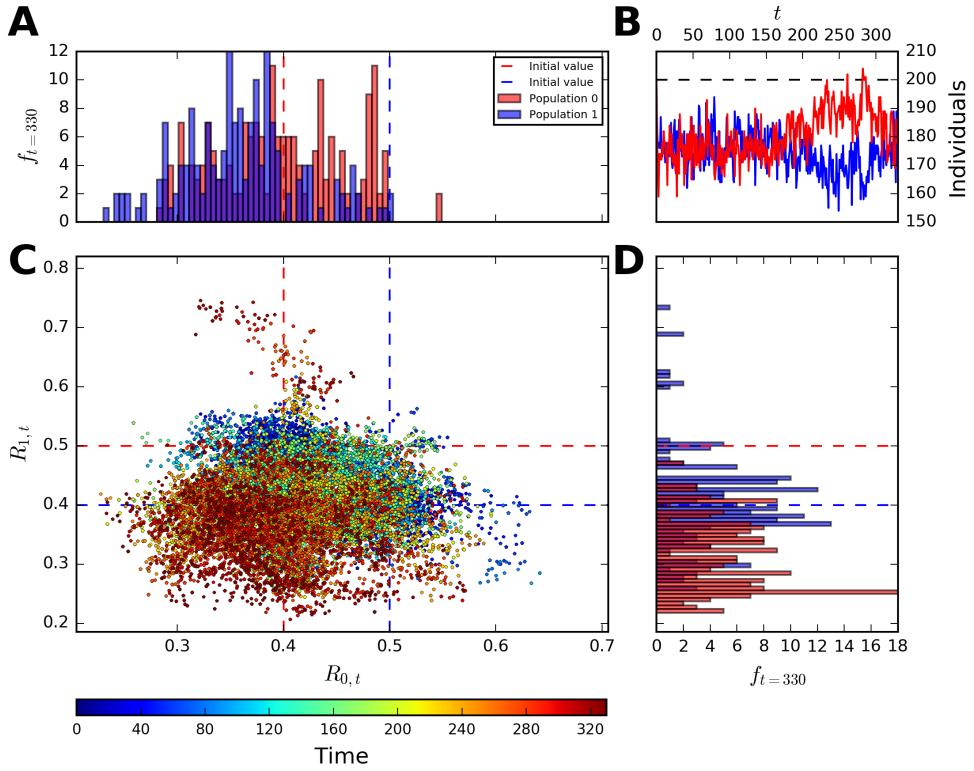
Surprisingly, after simulating communities satisfying the constraints discussed above (Figures 3.10 and C.5), we realized that some communities could not reach an ecological equilibrium, as in order to do so, some individuals required  $l_{k,t} > 1/\eta$ . Since this ends up in exact solutions to equation 2.9, recalling equation 2.8, as the left term becomes more negative, therefore, the right term becomes more positive. This implies that wherever  $l_{k,t} > 1/\eta$ , such individual is not receiving a benefit, as the extra commodities are immediately shared to others, but instead it is facilitating the transport of commodities from overproducers to underproducers (Figures 3.12 and C.7). We refer to these possibilities as “*unmediated*” if  $\max l_{k,t} \leq 1/\eta$  and “*mediated*” if  $\max l_{k,t} > 1/\eta$ , since cells can “*mediate*” the commodities exchange.



**Figure 3.11: Variables evolution: facultative mutualism (unmediated).** The evolution of variables of the simulation in Figure 3.10 are shown. The notation is concomitant with Figure's 3.5.

Once more, the initial spatial distribution is critical. If this is heterogeneous, as  $X$  and  $Y$  get bigger, more unfavourable neighbourhoods are found in  $E|_0$ , such that only facultative-mediated mutualism leads to communal ecological equilibria. On the other hand, we know that ecological equilibria are reached for any  $X$  and  $Y$  if the distribution is homogeneous (equation 2.15, Figure 3.2). We observed these outcomes whenever  $X = Y = 20$ . Thereafter, facultative-unmediated simulations started from homogeneous distributions (Figures 3.10 and C.5), whereas facultative-mediated started from heterogeneous distributions (Figures 3.12 and C.7).

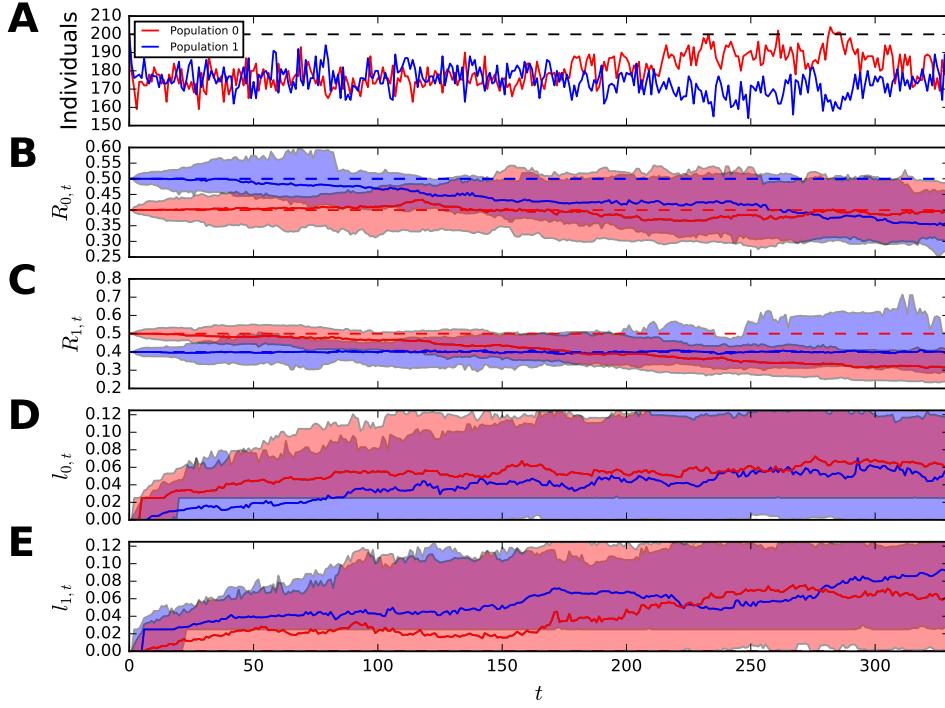
Initially, in agreement with previous cases, we used  $\sigma_{R,k} = \sigma_{l,k} = 1 \times 10^{-2}$  to simulate a highly evolvable facultative-unmediated mutualism (Figure C.5). Nonetheless, this led the community to collapse very rapidly (Figure C.6 A), a consequence of the increasing spatial heterogeneity and the evolution of variables



**Figure 3.12: Niche evolution: facultative mutualism (mediated).** This is a similar system to Figure 3.4, where individuals cannot self-sustain their requirements but whose leakages are evolvable and ecologically adaptable ( $\max l_{k,t} > 1/\eta$ ), being therefore under selection. The same parameters as in Figure 3.4 were used, but instead  $\sigma_{l,0} = \sigma_{l,1} = 1 \times 10^{-2}$ . Notation is concomitant with Figure's 3.4.

(Figure C.6 B-E). Moreover, by using  $l_{k,0} = 0.125$ , we set up the best case scenario, yet the evolvability caused an error threshold effect (Figure C.5 A, D). Proposing  $\sigma_{R,k} = \sigma_{l,k} = 1 \times 10^{-3}$  instead, effectively favoured the community survival (Figure 3.10). Now, the populations oscillated around a constant value (Figure 3.11 A). The niche exploration was reduced (Figure 3.10 C), and  $R_{k,t}$  and  $l_{k,t}$  remained attached to their initial values (Figure 3.10 A, D) with limited dispersal (Figure 3.11 B-E).

In contrast, facultative-mediated mutualism thrived for both high ( $\sigma_{R,k} = \sigma_{l,k} = 1 \times 10^{-2}$ ) and low ( $\sigma_{R,k} = \sigma_{l,k} = 1 \times 10^{-3}$ ) evolvability conditions (Figures 3.12 and C.7). In both cases, populations oscillated around an asymptotic value (Figures 3.13 A and C.8 A). A directionality in the evolution of niche through time was evident (Figures 3.12 C and C.7 C), with the highest  $R_{k,0}$  moving towards the lowest



**Figure 3.13: Variables evolution: facultative mutualism (mediated).** The evolution of variables of the simulation in Figure 3.12 are shown. The notation is concomitant with Figure's 3.5.

$R_{k,0}$  (Figures C.7 A, D and C.8 B-C) until their distribution eventually overlapped (Figures 3.12 A, D and 3.13 B-C). Finally, the leakages did not seem to be under selection, specially for the highly evolving community as their median value evolved irregularly and their distribution was completely spread over time (Figure 3.13 D-E). Thereafter, facultative-mediated mutualism seemed to provide the more fulfilling conditions for mutualism persistence.

Unfortunately, as shown in Figures 3.11, 3.13 and C.8, simulations stopped before the predefined 1,000 time-steps (Figures 3.5 and 3.7). Such numerical shortcomings were caused by particularities of the system being solved either at the SVD or linear programming stages. Hence, further numerical improvements are required to increase the robustness of our computational pipeline.

## Chapter 4

# Discussion

In this thesis we proposed a mathematical model to study the evolutionary dynamics of mutualistic communities. In agreement with the current interest to engineer synthetic ecologies (Mee and Wang, 2012; Widder et al., 2016), we based our approach in the quantitative description of a network of interacting individuals. Previous models have used a network description as well, however, being based on game theory, they reduced the interactions to abstract cost and benefit scores (Yamamura et al., 2004; Peña et al., 2016). Such approaches have offered important theoretical insights, but their applicability to engineering is not immediate (Escalante et al., 2014). Therefore, the variables and parameters in our model are material exchanges and measurable properties of the populations (Tables 2.1 and 2.2).

A description based on the ecological niche was fundamental, since it allows us to locate species inside a general  $n$ -dimensional space of commodities (Blonder et al., 2014). There, species are described as functions that for each set of coordinates receives a certain fitness ( $0 \leq f \leq 1$ ). In our model, all populations were described by the same fitness functions (equations 2.5 and 2.6), but with different parameters ( $\gamma_f, \alpha_f, \beta_f$ ). This fitness is not constant whatsoever, as the requirements ( $R_{k,t}$ ) and leakages ( $l_{k,t}$ ) of the populations evolve after duplication events. Then,  $R_{k,t}$  and  $l_{k,t}$  are used to estimate the gathering of commodities ( $G_{k,t}$ ) that individuals get after interacting with others (equation 2.1). Here, we have assumed that these

interactions resemble a lattice as shown in Figure 3.3.

Other lattice models of mutualism have been developed. Firstly, the pioneering work of Doebeli and Knowlton (1998) suggested the relevance of spatial structure and initial investments. Tainaka and Hashimoto (2016) studied obligate mutualisms, while Ezoe (2009) studied facultative mutualisms based on the Prisoners Dilemma. Ezoe and Ikegawa (2013) analysed how communities of mutualists and non-mutualists coexist. And finally, Iwata et al. (2011) and de Oliveira and Dickman (2014) have developed mean-field approximations of such models. Particularly, de Oliveira and Dickman (2014) generated a phase diagram of mutualism outcomes based on the death rate, birth rate and diffusion coefficient. Remarkably, as a result of the material exchange represented in our model, we could study several of the mutualisms discussed above, nonetheless we just focused on a few.

Firstly, in this project our objective was to study the evolution of niche in a mutualistic community. Thereafter, null-mutualism was the controlled condition to which other results would be contrasted. According to these simulations, null mutualistic communities were able to survive over time (Figures 3.4 and 3.6). This was expected since we had chosen the parameters and initial conditions to allow communities to barely survive without cooperation. Furthermore, the simulations where  $l_{k,t}$  was allowed to evolve, did not seem to put this variable under selection (Figure 3.6 and C.3). Interestingly, we observed that for the underproducers  $R_{k,t}$  evolved to equal  $R_{k,0}$  of the overproducers. This indicated an evolutionary selection.

Secondly, constitutive mutualistic communities were unable to thrive (Figure 3.8). Concretely, the spatial structure was determinant to such outcome. Only populations composed by identical individuals arranged in a symmetric spatial distribution would survive. Therefore, we argue based on our assumptions, this mutualism will rarely occur in nature. Instead, adaptations in the exchanges of commodities have to occur in order for the communities to reach an equilibrium (Maire and Youk, 2015). On the other hand, this result highlights a weakness of our model - the

ecological dynamics is not traced, instead an ecological equilibrium is assumed to exist and then calculated (Pérez-Escudero et al., 2009). For the constitutive mutualism this equilibrium is infinitesimal within the parameters space (Figure 3.2 A-B). An important observation during simulations was the error threshold effect (Figure 3.6). Using a model based on digital organisms Wilke et al. (2001) observed a similar effect, where high mutation rates cause individuals to be less fit. This occurs because less evolving individuals remain in specific regions of the niche space, while high evolving individuals are constantly changing their location, impeding them to remain in good spots (“survival of the flattest”).

Finally, facultative mutualism overcame the limitations of constitutive mutualism. Now, having a whole region of the parameters space (Figure 3.2 A-B), an equilibrium maximizing communal fitness can be found (Tikhonov, 2016). However, these simulations revealed us a detail disregarded by our analyses in section 3 - in order to sustain the survival of the whole community, apart from a facultative interaction, the mediation of commodities may be required. This condition agrees with a public goods interaction where individuals cooperate to sustain communal resources (Rauch et al., 2016). Nevertheless, as argued by Wakano et al. (2009), in this scenario individual selection favours cheaters. This seems to be true after comparing Figures 3.13 and C.8. In the first case,  $\sigma_{l,k} = 1 \times 10^{-2}$  causes the spread of  $l_{k,t}$  over time, thus low quality mutualists ( $l_{k,t} \ll 1/\eta$ ), cheaters ( $l_{k,t} = 0$ ) and exploited individuals ( $l_{k,t} \approx 1/\eta$ ) coexist; in the second case, the populations grow to the same level, but  $\sigma_{l,k} = 1 \times 10^{-3}$  limits the spread of  $l_{k,t}$  over time, thus neither cheaters nor exploited individuals are observed.

Now, we will discuss our assumptions. Regarding the spatial structure, other models are conventionally used (Dieckmann and Law, 2000), however each one is characterized by certain limitations. In our case, although lattices are widely used, these are only a particular case of a bigger modelling approach, graphs (Lieberman et al., 2005). Graphs allow the manipulation of the spatial structure by changing

the links between nodes. Peña et al. (2016) summarize their potential by comparing their results with the pairwise and diffusion approximations, finding that graphs capture more of the complexity arising from the spatial structure and interactions. Accordingly, by developing mechanistic models and compare them to its pairwise approximation Momeni and Shou (2016) found that pairwise approximations could produce misleading results as their assumptions may not hold true.

Regarding the initial conditions, some experimental results support our assumptions. Firstly, Campbell et al. (2015) have shown the emergence of cooperative communities exchanging metabolites in *Saccharomyces cerevisiae*. Furthermore, this may cause cellular heterogeneity within previously genetically-identical populations (Campbell et al., 2016). Rauch et al. (2016) have shown that the eco-evolutionary stability is essential to sustain a public goods cooperation; and importantly that communities require minimum viable population sizes to thrive. Finally, Klitgord and Segre (2010) developed genome-scale stoichiometric models of multiple microbial strains in order to find out which communities may cause a mutualistic interaction. Importantly, according to their results, environmental conditions may cause the interactions more readily than genetic modifications.

In agreement with our results, using a game graph model Peña et al. (2016) have shown that the evolution of cooperation in lattices could lead to more stringent conditions than in well-mixed populations. Additionally, Tainaka and Hashimoto (2016) have found that local interactions restrict the survival severely when compared to global interactions. Moreover, they argue a specific population ratio is required in order to survive (“ratio selection theory”). Finally, Allen et al. (2013) have used a graph model to describe the diffusion of public goods, deriving analytic functions for the success of mutualism based on the graph topology and diffusion. The work of Allen et al. (2013) resembles our project very neatly. Accordingly, we do consider a further investigation of the topology of the interactions network would offer a more comprehensive description of our results.

## 4.1 Significance of the research

Ecology and evolution have historically benefited from mathematical theory and models (Page and Nowak, 2002; Gertsev and Gertseva, 2004). Nowadays, allowed by current technologies, to engineer simple ecologies is possible (Shou et al., 2007). In order to exploit those approaches, a clear translation from theory to engineering is necessary (Klitgord and Segre, 2010). Our mathematical model may bring awareness to experimentalists about the relevant parameters and variables to engineer evolutionary resilient mutualisms (Escalante et al., 2014). Moreover, with the adequate parameters it could produce predictions of otherwise untraceable systems.

The material balance description of the mutualistic interactions in our model allow us to easily modify their nature by changing the fluxes. Moreover, to go from a simple system of 2 populations exchanging 2 commodities, to a high-dimensional system of  $NP$  populations and  $K$  commodities is straightforward. Thus, allowed by the current computational power, our model can simulate mutualistic communities in great detail, being useful to tackle theoretical and experimental enquiries.

## 4.2 Limitations

Although we successfully implemented our mathematical model to study a mutualistic community, limited by time, here we showed the results of individual simulations only composed by 400 cells during at most 1,000 time-steps. Caused by the probabilistic occurrence of survival and duplication, even simulations starting from the same initial conditions will follow different evolutionary trajectories. To overcome these limitations, we focused our analyses on the generalities by providing a comprehensive explanation of the outcomes observed. Nonetheless, to perform numerous simulations in order to quantify the variability is required if more specific enquiries are meant to be answer.

In order to simplify our analyses while contrasting different mutualisms, we fixed some of the initial conditions ( $R_{k,0}$ ,  $P_{k,0}$ ) and parameters ( $\eta$ ,  $\gamma_S$ ,  $\gamma_D$ ,  $\alpha_S$ ,  $\alpha_D$ ,

$\beta_D$ ). These were chosen to represent a community where a mutualistic partnership may increase the communal fitness, but still barely persistent without it. Each population was specialized to one commodity only, being either an underproducer ( $R_{k,0}/P_{k,0} = 0.8$ ) or an overproducer ( $R_{k,0}/P_{k,0} = 1.25$ ). Regarding  $\eta$ , neighbourhoods in lattice models are usually the immediate neighbours (Dieckmann and Law, 2000), here we attached to such standard, so  $\eta = 8$ . The maximum survival and duplication probabilities were associated to the fitness after commodities gathering, so to represent deaths independent of it we set up  $\gamma_S = \gamma_D = 0.9$ . Our sensitivity analysis in subsection 2.2.2 shows  $\alpha_S$  and  $\alpha_D$  to be less influential for sufficiently big values, so we set up  $\alpha_S = \alpha_D = 20$ . Finally, as duplication demands more commodities than survival we decided  $\beta_D = 1.5\beta_S$ . Although these values were based on theoretical analyses, still others could have been chosen. Therefore, we cannot generalize our results to be valid for any parameter combination. Moreover, individual's neighbourhood should be adapted to the geometry of the growing community (Figure 2.1).

Some fundamental assumptions about the populations were made. Firstly, the fitness received after the uptake of the  $K$  commodities was described by a sigmoid function parametrized by  $\gamma_f$ ,  $\alpha_{f,k}$  and  $\beta_{f,k}$ . To reduce their number, we assumed any commodity to affect the fitness equally, so  $\alpha_f = \alpha_{f,k}$  and  $\beta_f = \beta_{f,k}$ . This may not stand true for natural communities, as scarce commodities may have a major fitness effect (Zee and Bever, 2014). However, as long as this assumption was only intended to reduce the number of parameters discussed here, if the model is applied, this could be relaxed to better fit the experimental data. Secondly, having set a normal distribution to draw mutations from, the invariable nature of  $\sigma_{R,k}$ ,  $\sigma_{l,k}$  could be challenged, as Gingerich (2009); Lanfear et al. (2014) have found the mutation rate to react to the evolutionary dynamics. Thereafter, correlative functions adjusted to experimental data may better describe  $\sigma_{R,k}$  and  $\sigma_{l,k}$  dynamics.

Appraising our numerical implementation, this failed to find the ecological

equilibrium of some facultative mutualism systems. We found out that during the  $\mathbf{I}|_{k,t}$  optimization (line 91, Appendix A.1), based on the Simplex algorithm, convergence to the solution was impeded by a cycling during the pivoting. Additionally, during the  $\mathbf{P}|_{k,t}$  calculation (lines 96 – 98, Appendix A.1), the selection of the SVD’s singular values ( $\sqrt{\lambda}$ ) seldom led to erroneous results (Appendix B.2). What is more, these shortcomings may occur more frequently for larger systems. Alternative algorithms to the Simplex may be more robust and less computationally demanding, however this has to be investigated further.

### 4.3 Future directions

The very nature of our project has produced an important number of future directions to follow. These go from compiling more data, to propose new experimental designs, to eventually perform *in-vitro* or *in-situ* experiments.

Firstly, to support the findings in section 3 more data is necessary. Based on the single simulations we have described, we still cannot conclusively argue whether niche evolution is constrained by mutualism. However, after having more simulations’ results, by increasing the replicates, and specifically by simulating different mutualistic conditions from the same initial spatial distribution ( $\mathbf{E}|_0$ ), would allow us to provide a statistical description of our findings. Additionally, by increasing the spatial dimensions of communities ( $X, Y$ ), a closer approximation to an infinite system, that we resemble by imposing periodic boundaries, would be achievable.

Secondly, the exploration of more parameters would cement the relevance of our observations. In the first instance, to run simulations using multiple initial conditions and parameters is necessary to verify that conclusions withdrawn from our results are caused by the nature of the mutualistic interaction itself and not chance. In agreement, to simulate communities where  $NP > 2$  and  $K > 2$  would better represent natural ecosystems and possibly provide insights of persistence mechanisms unforeseeable in binary communities. In the final instance, regarding the fundamen-

tal niche, to perform a brute force search to simulate large parameters combinations would produce its most descriptive picture.

Thirdly, new experimental designs could be proposed. In the first instance, although we found the spatial structure to be crucial for the constitutive mutualism, we still need to analyse the effect of spatial disturbances on facultative mutualism. In the final instance, here we focused on the case where each population specializes in one commodity, like in lichens (Honegger, 2009) and synthetic communities (Shou et al., 2007). However, microbial communities may differ importantly, as multiple populations could specialize in the same or multiple commodities (Lenski and Travisano, 1994; Campbell et al., 2016). While simulating such system, demanding environmental conditions may required them to specialize, however, as they would be adept producers of multiple commodities, even from a single population multiple varieties may arise (Freilich et al., 2011).

Finally, *in-vitro* and *in-situ* experiments may confirm our theoretical findings. But the model as well may suggest parameters to engineer mutualistic communities. The methodologies of Campbell et al. (2015) and Hillesland et al. (2014) based on experimental evolution and those of Momeni et al. (2011) and Mee and Wang (2012) based on synthetic biology would be particularly appropriate. Interestingly, Ratzke and Gore (2016) and Celiker and Gore (2014) propose methodologies that only using Petri dish or 96-well plate cultures resemble our lattice simulations.

## **Chapter 5**

# **Conclusions**

We have developed a mathematical model describing the ecological and evolutionary dynamics of mutualism. Having defined populations by their ecological niches allowed us to propose a set of variables describing either ecological or evolutionary elements of the dynamics. Based on this separation we analysed the sensibility of the model to variables and parameters when interactions are either null, constitutive or facultative. According to our simulations facultative interactions could sustain a mutualistic partnership more readily. Importantly, we also provided a thorough analysis of the ecological equilibria and how spatial structure constrains them. Regarding the niche constriction, the simulations analysed here provide a glimpse of the expected dynamics, however a greater number of simulations and exhaustive statistical analyses are still needed. Finally, the categories of mutualism we have proposed agree with a genetic engineering description, moreover the variables and parameters could be adequately measured in a laboratory setting. Therefore, we expect this model to be applied to study natural and engineerable mutualisms.

# Acknowledgements

Firstly, I thank the wise support provided by my advisor, Dr David Murrell, whose combination of patience and empowerment was very formative from the beginning. I thank as well to my master and stream tutors, Dr Andrew Batchelor and Prof Ziheng Yang, for the invaluable advice received during this year, which along with their conciliation was vital for this outcome.

Secondly, this work could not have been possible without the support of my family, which, even in daunting times, has been there for me.

Finally, I acknowledge to the Mexican National Council for Science and Technology (CONACYT) for the funding provided to pursue my MRes Biosciences degree at University College London (scholarship number: 411162).

## Appendix A

# Computational code

## A.1 Main functions

```
1 #!/usr/bin/python
2 """ Niche Evolution: main functions """
3
4 # Import packages' functions and model's parameters
5 from numpy import argmax, array, diag, dot, exp, hstack, identity, ix_, nan_to_num
6     , ones, repeat, tile, unique, where, zeros
7 from numpy.random import choice, normal, shuffle
8 from numpy.linalg import pinv, svd
9 from scipy.optimize import linprog
10 from parameters import *
11
12 # Number of Neighbours
13 def fETA():
14     eta = pow(max(N)-min(N)+1,2)-1
15     return eta
16
17 # Neighbourhood Coordinates
18 def fN(dim, z):
19     lN = (len(N)-1)/2
20     if z==0: Nz = hstack((range(dim-lN,dim),range(0,lN+1)))
21     elif z==dim-1: Nz = hstack((range(-lN,1),range(-dim+1,-dim+lN+1)))
22     else: Nz = N
23     return array(Nz,dtype=int)
24
25 # Populations Initial Locations
26 def fiLOC(geometry):
27     if geometry == "homogeneous":
28         LOC, y = zeros((X,Y)), 0
29         while y < Y:
30             x = y%2
31             while x < X:
32                 LOC[x,y] = 1
33                 x = x+2
34             y = y+1
35         LOC = LOC.reshape(1,X*Y)[0]
36         LOC = hstack([where(LOC==1)[0],where(LOC==0)[0]])
37     if geometry == "heterogeneous":
38         LOC = range(X*Y)
39         shuffle(LOC)
40     return LOC
```

```

40
41 # Parameters
42 def fPAR(E):
43     G_S, G_D, A_S, A_D, B_S, B_D = zeros((X,Y)), zeros((X,Y)), zeros((X,Y)), zeros((X,Y)),
44     ((X,Y)), zeros((X,Y)), zeros((X,Y))
45     R0, P0 = [zeros((X,Y)) for k in range(K)], [zeros((X,Y)) for k in range(K)]
46     for np in range(NP):
47         loc = where(E==np+1)
48         G_S[loc], G_D[loc] = gamma_S[np][0], gamma_D[np][0]
49         A_S[loc], A_D[loc] = alpha_S[np][0], alpha_D[np][0]
50         B_S[loc], B_D[loc] = beta_S[np][0], beta_D[np][0]
51         for k in range(K): R0[k][loc], P0[k][loc] = r0[np][k], p0[np][k]
52     return R0, P0, G_S, G_D, A_S, A_D, B_S, B_D
53
54 # Production of Commodities
55 def fP(E, R, L, R0, P0, mutualism, mediation):
56     eta = fETA()
57     E_nz = where(E.reshape(1,X*Y)[0]!=0)[0]
58     l_nz, c_nz = len(E_nz), ix_(E_nz,E_nz)
59     P, L_mut = [zeros((X*Y,1)) for k in range(K)], [zeros((X*Y,1)) for k in range(K)]
60     for k in range(K):
61         Psi = B_D*R[k]
62         if mutualism == "null":
63             P[k] = nan_to_num(P0[k]/R0[k])*Psi
64             L_mut[k] = L[k]
65         elif mutualism == "constitutive" or mutualism == "facultative":
66             Upsilon, Gamma, Lambda, Omega, Phi = zeros((X*Y,X*Y)), zeros((X*Y,X*Y)),
67             zeros((X*Y,X*Y)), zeros((X*Y,X*Y)), zeros((X*Y, X*Y))
68             Pk_opt = zeros((X*Y,1))
69             Omega[range(X*Y),range(X*Y)] = nan_to_num(R0[k]/P0[k]).reshape(1,X*Y)
70             Pk_opt[E_nz] = dot(pinv(Omega[c_nz]),Psi.reshape(X*Y,1)[E_nz])
71             Phi[range(X*Y),range(X*Y)] = Pk_opt.T
72             Upsilon[range(X*Y),range(X*Y)] = 1
73             for x, y in zip(repeat(range(X),Y),tile(range(Y),X)):
74                 gamma = zeros((X,Y))
75                 Nx, Ny = fN(X,x), fN(Y,y)
76                 for i, j in zip(repeat(Nx,len(Nx)),tile(Ny,len(Ny))):
77                     if (i!=0 or j!=0) and E[x+i,y+j]!=0: gamma[x+i,y+j] = 1
78                     Gamma[x*X+y,:] = gamma.reshape(1,X*Y)
79             Gamma[E_nz,E_nz] = -eta
80             if mutualism == "constitutive":
81                 Lambda[range(X*Y),range(X*Y)] = L[k].reshape(1,X*Y)
82                 Ak = Upsilon + dot(Gamma,Lambda) - Omega
83                 L_mut[k] = L[k]
84             if mutualism == "facultative":
85                 a_f = hstack((dot(Gamma[c_nz],Phi[c_nz]),identity(l_nz),-identity(l_nz)))
86                 b_f = Psi.reshape(X*Y,1)[E_nz]-dot(Upsilon[c_nz],Pk_opt[E_nz])
87                 c_f = hstack((zeros((1,l_nz))[0],ones((1,2*l_nz))[0]))
88                 l_bound = (1-nan_to_num(R0[k]/P0[k]))/eta
89                 l_bound[l_bound<0] = 0
90                 l_bound = hstack((l_bound.reshape(1,X*Y)[0][E_nz],zeros((1,2*l_nz))[0]))
91                 if mediation == "unmediated": u_bound = hstack((L[k].reshape(1,X*Y)[0][E_nz],[None for i in range(2*l_nz)]))
92                 elif mediation == "mediated": u_bound = [None for i in range(3*l_nz)]
93                 lk_lp = linprog(c=c_f, A_eq=a_f, b_eq=b_f, bounds=zip(l_bound,u_bound),
94                                 options={"tol":1E-10, "maxiter":1E10, "bland":True})
95                 Lambda[E_nz,E_nz] = lk_lp.x[:l_nz].reshape(1,l_nz)
96                 Ak = Upsilon + dot(Gamma,Lambda) - Omega
97                 L_mut[k] = diag(Lambda).reshape(X,Y)
98                 ak = Ak[c_nz]
99                 u, s, v = svd(ak)

```

```

97     ind_ss = argmax(s[:-1]/s[1:])+1
98     pk = dot(v.T[:,ind_ss:],dot(pinv(v.T[:,ind_ss:]),Psi.reshape(X*Y,1)[E_nz
99         ])))
100    ind_nz = where(abs(pk)>1E-14)
101    if sum(pk[ind_nz]<0) > sum(pk[ind_nz]>0): pk = -pk
102    if len(unique(pk[ind_nz]>=0))==1 or len(unique(pk[ind_nz]<=0))==1: P[k][
103        E_nz] = abs(pk).reshape(l_nz,1)
104    else:
105        a = hstack((ak,identity(l_nz),-identity(l_nz)))
106        b = dot(ak,pk)
107        c = hstack((zeros((1,l_nz))[0],ones((1,2*l_nz))[0]))
108        opt_P = linprog(c=c, A_eq=a, b_eq=b, bounds=[0,None], options={
109            'maxiter':1E10})
110        P[k][E_nz] = opt_P.x[:l_nz].reshape(l_nz,1)
111    P[k] = P[k].reshape(X,Y)
112    return P, L_mut
113
114 # Gathering of Commodities
115 def fG(P, L, mutualism):
116     eta = fETA()
117     G = [zeros((X,Y)) for k in range(K)]
118     if mutualism == "null": L = [zeros((X,Y)) for k in range(K)]
119     for k in range(K):
120         for x, y in zip(repeat(range(X),Y),tile(range(Y),X)):
121             Nx, Ny = fN(X,x), fN(Y,y)
122             for i, j in zip(repeat(Nx,len(Nx)),tile(Ny,len(Ny))):
123                 if (i==0 and j==0): G[k][x,y] = G[k][x,y]+(1-eta*L[k][x,y])*P[k][x,y]
124                 else: G[k][x,y] = G[k][x,y]+L[k][x+i,y+j]*P[k][x+i,y+j]
125     return G
126
127 # Survival and Duplication Probabilities
128 def fSD(G, R, G_S, G_D, A_S, A_D, B_S, B_D):
129     S, D = G_S, G_D
130     for k in range(K):
131         gk = nan_to_num(G[k]/R[k])
132         S = S * (1./(1.+exp(-A_S*(gk-B_S/2.))))
133         D = D * (1./(1.+exp(-A_D*(gk-B_D/2.))))
134     return S, D
135
136 # System Initial State
137 def SYS_IS(geometry):
138     eta = fETA()
139     LOC = fiLOC(geometry)
140     E, R, L = zeros((1,X*Y))[0], [zeros((1,X*Y))[0] for k in range(K)], [zeros((1,X
141         *Y))[0] for k in range(K)]
142     for k in range(K):
143         cp = 0
144         for np in range(NP):
145             loc = LOC[cp:cp+IP[np][0]]
146             if k==0: E[loc] = int(np+1)
147             R[k][loc], L[k][loc] = r0[np][k], 10[np][k]
148             cp = cp+IP[np][0]
149             R[k], L[k] = R[k].reshape(X,Y), L[k].reshape(X,Y)
150     E = E.reshape(X,Y)
151     return E, R, L
152
153 # System Current State
154 def SYS_CS(E, R, L, mutualism, mediation):
155     R0, P0, G_S, G_D, A_S, A_D, B_S, B_D = fPAR(E)
156     P, L_mut = fP(E, R, L, R0, P0, mutualism, mediation)
157     G = fG(P, L_mut, mutualism)
158     S, D = fSD(G, R, G_S, G_D, A_S, A_D, B_S, B_D)
159     return S, D, P, G

```

```

156
157 # System Transition State
158 def SYS_TS(S, D, E, R, L):
159     eta = fETA()
160     R0, P0, G_S, G_D, A_S, A_D, B_S, B_D = fPAR(E)
161     X_nz, Y_nz = where(E!=0)
162     X_z, Y_z = where(E==0)
163     for x, y in zip(X_nz, Y_nz):
164         s = S[x,y]
165         if choice([0,1], 1, p=[1-s, s]) == 0:
166             E[x,y], D[x,y] = 0, 0
167             for k in range(K): R[k][x,y], L[k][x,y] = 0, 0
168         else: pass
169     for x_d, y_d in zip(X_z, Y_z):
170         Nx, Ny = fN(X, x_d), fN(Y, y_d)
171         I, J = tile(Nx, len(Nx)), repeat(Ny, len(Ny))
172         n = D[x_d+I, y_d+J]
173         if n.sum() != 0:
174             ind = choice(where(n==n.max()))[0]
175             x_p, y_p = x_d+I[ind], y_d+J[ind]
176             D[x_p, y_p] = 0
177             E[x_d, y_d] = E[x_p, y_p]
178             for k in range(K):
179                 s_R, s_l = sigma_R[int(E[x_p, y_p]-1)][k], sigma_l[int(E[x_p, y_p]-1)][k]
180                 if s_R == 0: R[k][x_p, y_p] = R[k][x_p, y_p]
181                 else: R[k][x_p, y_p] = R[k][x_p, y_p] + normal(0, s_R, 1)
182                 if s_l == 0: L[k][x_p, y_p] = L[k][x_p, y_p]
183                 else: L[k][x_p, y_p] = L[k][x_p, y_p] + normal(0, s_l, 1)
184                 L_min = (1. - R0[k][x_p, y_p]/P0[k][x_p, y_p])/eta
185                 if R[k][x_p, y_p] < 0: R[k][x_p, y_p] = 0
186                 if L[k][x_p, y_p] < 0: L[k][x_p, y_p] = 0
187                 elif L[k][x_p, y_p] > 1./eta: L[k][x_p, y_p] = 1./eta
188                 elif L[k][x_p, y_p] < L_min: L[k][x_p, y_p] = L_min + 1E-10
189                 R[k][x_d, y_d], L[k][x_d, y_d] = R[k][x_p, y_p], L[k][x_p, y_p]
190             else: pass
191     return E, R, L

```

Note: If mutualism="facultative", mediation="unmediated" and geometry="homogeneous", then "bland":False in line 90.

## A.2 Parameters

```

1 #!/usr/bin/python
2 """ Niche Evolution: parameters """
3
4 X, Y = 20, 20
5 N = [-1, 0, 1]
6 NP = 2
7 IP = [[200], [200]]
8 K = 2
9 #species -> commodity
10 p0 = [[0.4, 0.5], [0.5, 0.4]]
11 r0 = [[0.5, 0.4], [0.4, 0.5]]
12 l0 = [[0.0, 0.05], [0.05, 0.00]]
13 gamma_S = [[0.9], [0.9]]
14 gamma_D = [[0.8], [0.8]]
15 alpha_S = [[20], [20]]
16 alpha_D = [[20], [20]]
17 beta_S = [[1.0], [1.0]]
18 beta_D = [[1.5], [1.5]]
19 sigma_R = [[1E-2, 1E-2], [1E-2, 1E-2]]
20 sigma_l = [[1E-2, 1E-2], [1E-2, 1E-2]]

```

## Appendix B

# Derivations

## B.1 Exact solutions: symmetric spatial distributions

### B.1.1 Unequal gathering

Suppose a system at  $t$  is composed by empty cells ( $E$ ) and two populations, regarding a  $k$  commodity one is a consumer ( $C$ ) and other a producer ( $P$ ). Let  $\eta$  be:

$$\eta = \eta_{CC} + \eta_{CP} + \eta_{CE} = \eta_{PC} + \eta_{PP} + \eta_{PE}$$

Where the first digit of the subindex indicates the nature of the focus cell and the second digit the nature of its neighbours. Denote  $l_C$ ,  $\left(\frac{R_0}{P_0}\right)_C$  and  $l_P$ ,  $\left(\frac{R_0}{P_0}\right)_P$  as the leakage and requirement/production ratio for consumer and producer respectively. Then the equilibrium of production and consumption occurs when:

$$\begin{aligned} 0 &= \left(1 - \left(\frac{R_0}{P_0}\right)_C - \eta l_C\right) P_C + \eta_{CC} l_C P_C + \eta_{CP} l_P P_P \\ &= \left(1 - \left(\frac{R_0}{P_0}\right)_C - (\eta_{CP} + \eta_{CE}) l_C\right) P_C + \eta_{CP} l_P P_P \end{aligned}$$

$$\begin{aligned} 0 &= \left(1 - \left(\frac{R_0}{P_0}\right)_P - \eta l_P\right) P_P + \eta_{PC} l_C P_C + \eta_{PP} l_P P_P \\ &= \left(1 - \left(\frac{R_0}{P_0}\right)_P - (\eta_{PC} + \eta_{PE}) l_P\right) P_P + \eta_{PC} l_C P_C \end{aligned}$$

Solving for  $\frac{P_C}{P_P}$  in each case and equalling them we get:

$$\frac{\eta_{CP} l_P}{\left(\frac{R_0}{P_0}\right)_C + (\eta_{CP} + \eta_{CE}) l_C - 1} = \frac{\left(\frac{R_0}{P_0}\right)_P + (\eta_{PC} + \eta_{PE}) l_P - 1}{\eta_{PC} l_C}$$

After cross-multiplying, simplifying, dividing by  $\left(\left(\frac{R_0}{P_0}\right)_P - 1\right) \left(\left(\frac{R_0}{P_0}\right)_C - 1\right)$  and rearranging we finally find:

$$\frac{\eta_{PC} \eta_{CP}}{\left(\left(\frac{R_0}{P_0}\right)_P - 1\right) \left(\left(\frac{R_0}{P_0}\right)_C - 1\right)} l_P l_C = \left( \frac{\eta_{PC} + \eta_{PE}}{\left(\frac{R_0}{P_0}\right)_P - 1} l_P + 1 \right) \left( \frac{\eta_{CP} + \eta_{CE}}{\left(\frac{R_0}{P_0}\right)_C - 1} l_C + 1 \right)$$

### B.1.2 Equal gathering

Using the same notation and assumptions as before, if we use the equations of the gatherings ( $G$ ) and equal them we get:

$$(1 - (\eta_{CP} + \eta_{CE}) l_C) P_C + \eta_{CP} l_P P_P = (1 - (\eta_{PC} + \eta_{PE}) l_P) P_P + \eta_{PC} l_C P_C$$

Solving for  $\frac{P_C}{P_P}$  and equalling this equation to the left- and right-handside of the equation we found in the subsection B.1.1:

$$\frac{1 - (\eta_{CP} + \eta_{PC} + \eta_{PE}) l_P}{1 - (\eta_{CP} + \eta_{PC} + \eta_{CE}) l_C} = \frac{\eta_{CP} l_P}{\left(\frac{R_0}{P_0}\right)_C + (\eta_{CP} + \eta_{CE}) l_C - 1}$$

$$\frac{1 - (\eta_{CP} + \eta_{PC} + \eta_{PE}) l_P}{1 - (\eta_{CP} + \eta_{PC} + \eta_{CE}) l_C} = \frac{\left(\frac{R_0}{P_0}\right)_P + (\eta_{PC} + \eta_{PE}) l_P - 1}{\eta_{PC} l_C}$$

After cross-multiplying each one, simplifying, solving for  $l_C l_P$  and equalling them we finally find:

$$\left(\frac{R_0}{P_0}\right)_C ((\eta_{CP} + \eta_{PC} + \eta_{PE}) l_P - 1) = \left(\frac{R_0}{P_0}\right)_P ((\eta_{CP} + \eta_{PC} + \eta_{CE}) l_C - 1)$$

## B.2 Exact solutions: any spatial distribution

Let us start with a matrix  $\mathbf{A}|_{k,t}$  of dimension  $(XY, XY)$ . What properties of  $\mathbf{A}|_{k,t}$  make  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0}$  solvable through the Singular Value Decomposition (SVD)?

SVD relies on solving  $\det(\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t} - \lambda \mathbf{I}) = \mathbf{0}$ , to find  $XY$  eigenvalues  $\lambda$ . This allow us rewrite  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t}$  as  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \lambda \mathbf{P}|_{k,t}$ , that is to say  $\lambda \mathbf{P}|_{k,t} = \mathbf{0}$ . Here  $\mathbf{P}|_{k,t} = \mathbf{0}$  is always a solution, however we are interested in solutions where  $\mathbf{P}|_{k,t} \neq \mathbf{0}$ , implying that we need  $\mathbf{P}|_{k,t}$ s whose associated  $\lambda = 0$ . Re-expressing the determinant as the polynomial:  $\lambda(C_{XY}\lambda^{XY-1} + C_{XY-1}\lambda^{XY-2} + \dots + C_1) + C_0 = 0$ , we see that  $\lambda = 0$  if  $C_0 = 0$ . Moreover if that occurs:  $\det(\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}) = \mathbf{0}$ .

Three characteristics of  $\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}$  arise from this last point:

1. The columns of  $\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}$  are no independent.
2. The rows of  $\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}$  are no independent.
3.  $\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}$  is not invertible.

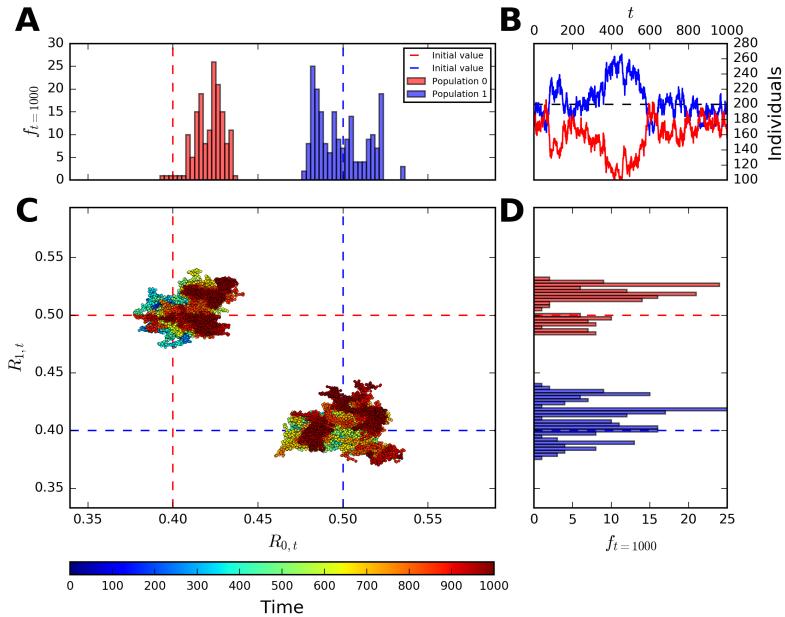
These points are referred to the product  $\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}$ , so to prove their validity on  $\mathbf{A}|_{k,t}$  allow us to introduce the following proof: let  $\mathbf{B}$  and  $\mathbf{C}$  be linearly independent square matrices and  $\mathbf{X}$  a column vector. Consider  $(\mathbf{B} \cdot \mathbf{C}) \cdot \mathbf{X} = \mathbf{0}$ . By the associative property of the inner product  $\mathbf{B} \cdot (\mathbf{C} \cdot \mathbf{X}) = \mathbf{0}$ . Now let  $\mathbf{C} \cdot \mathbf{X} = \mathbf{Y}$ , so  $\mathbf{B} \cdot \mathbf{Y} = \mathbf{0}$ , since

$\mathbf{B}$  is linearly independent that only occurs if  $\mathbf{Y} = \mathbf{0}$ , therefore  $\mathbf{C} \cdot \mathbf{X} = \mathbf{0}$ . Similarly since  $\mathbf{C}$  is linearly independent,  $\mathbf{C} \cdot \mathbf{X} = \mathbf{0}$  only occurs if  $\mathbf{X} = \mathbf{0}$ . So  $(\mathbf{B} \cdot \mathbf{C}) \cdot \mathbf{X} = \mathbf{0}$  has only the solution  $\mathbf{X} = \mathbf{0}$ , therefore  $\mathbf{B} \cdot \mathbf{C}$  is linearly independent. In our particular case  $\mathbf{B} = \mathbf{A}|_{k,t}^T$  and  $\mathbf{C} = \mathbf{A}|_{k,t}$ , then by negation of the previous proof, if  $\mathbf{A}|_{k,t}^T \cdot \mathbf{A}|_{k,t}$  is linearly dependent,  $\mathbf{A}|_{k,t}$  is linearly dependent as well.

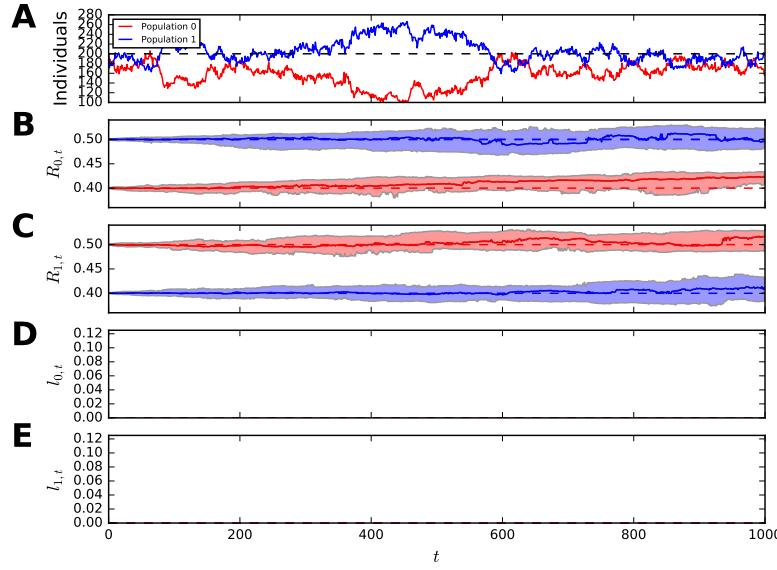
In conclusion, exact solutions to  $\mathbf{A}|_{k,t} \cdot \mathbf{P}|_{k,t} = \mathbf{0}$  only exist if:  $\text{rank } \mathbf{A}|_{k,t} \leq XY$ . After these analytic regards, you can see a numerical analysis referring to subsection 2.4.3, where a method based in a matrix decomposition is introduced.

## Appendix C

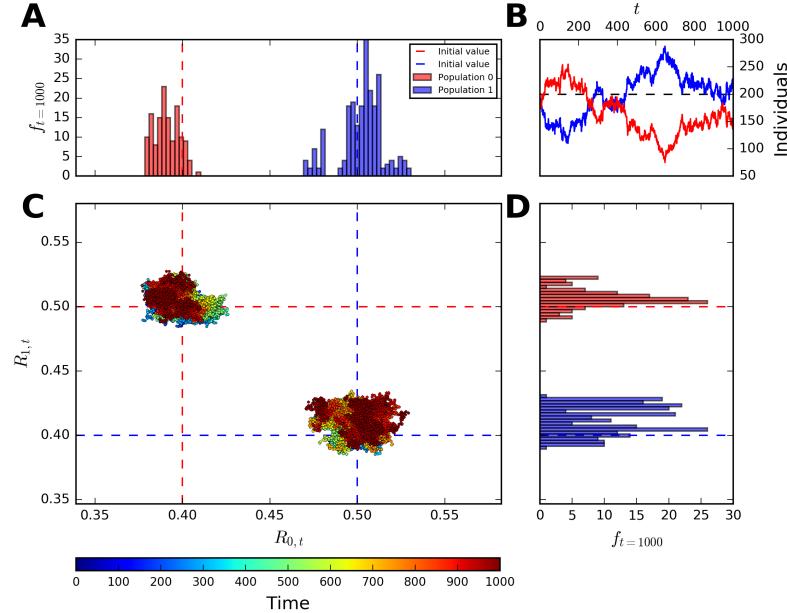
# Supplementary figures



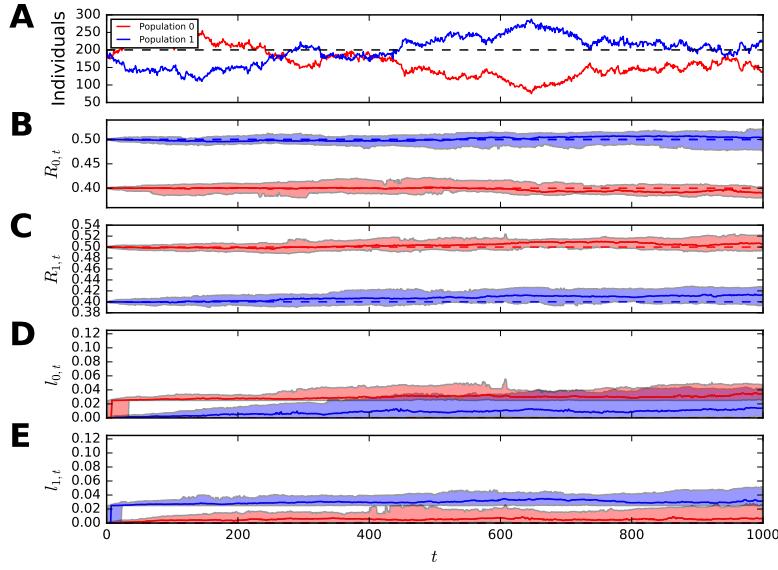
**Figure C.1: Reduced niche evolution: null mutualism (fixed leakage).** Starting from a random spatial configuration the evolution of two populations producing two commodities during 1000 time-steps is shown. Individuals neither can self-sustain their requirements nor evolve their leakages. **(B)** The populations through time are shown. **(C)** The  $R_{0,t}$  and  $R_{1,t}$  of all individuals through time is collapsed. The distribution of **(A)**  $R_{0,t}$  and **(D)**  $R_{1,t}$  at  $t = 1000$  is compared with their values at  $t = 0$ . The same parameters as in Figure 3.4 were used, but instead  $\sigma_{R,0} = \sigma_{R,1} = 1 \times 10^{-3}$ .



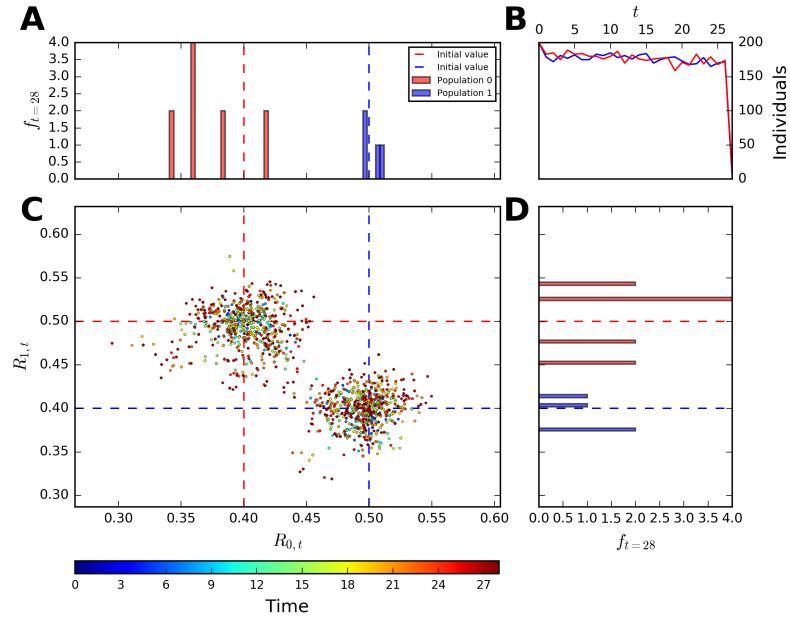
**Figure C.2: Reduced variables evolution: null mutualism (fixed leakage).** (A) The populations through time along with the evolution of (B)  $R_{0,t}$ , (C)  $R_{1,t}$ , (D)  $l_{0,t}$  and (E)  $l_{1,t}$  of the simulation in Figure C.1 are shown. The starting values are indicated by the dashed lines, the median by the continuous lines and the 5th to 95th percentile by the filled areas.



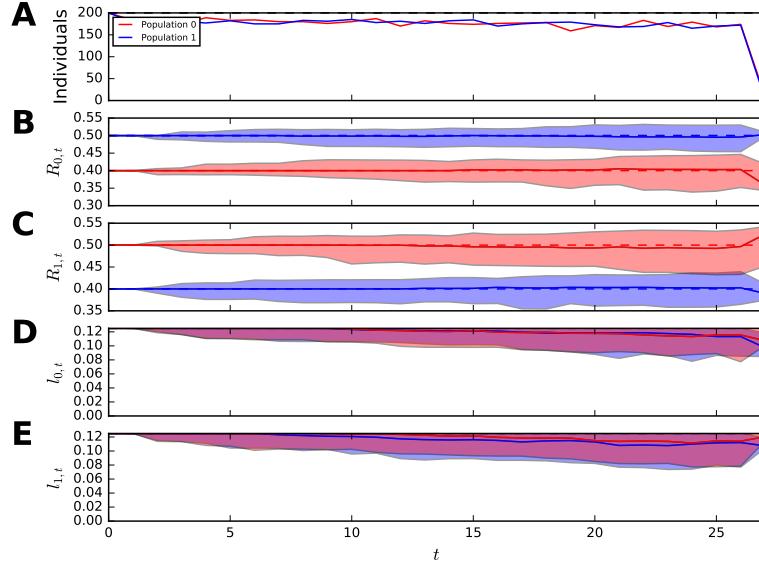
**Figure C.3: Reduced niche evolution: null mutualism (loose leakage).** This is a similar system to Figure C.1, where individuals cannot self-sustain their requirements but now can evolve their leakages; nonetheless these are not functional and therefore not under selection. The same parameters as in Figure 3.4 were used, but instead  $\sigma_{R,0} = \sigma_{R,1} = \sigma_{l,0} = \sigma_{l,1} = 1 \times 10^{-3}$ . The notation is concomitant with Figure's C.1.



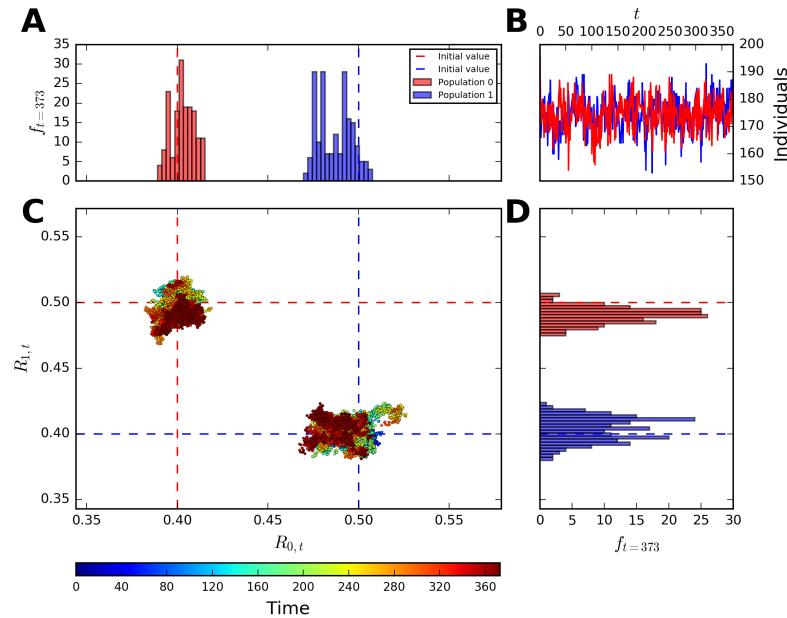
**Figure C.4: Reduced variables evolution: null mutualism (loose leakage).** The evolution of variables of the simulation in Figure C.3 are shown. The notation is concomitant with Figure's C.2.



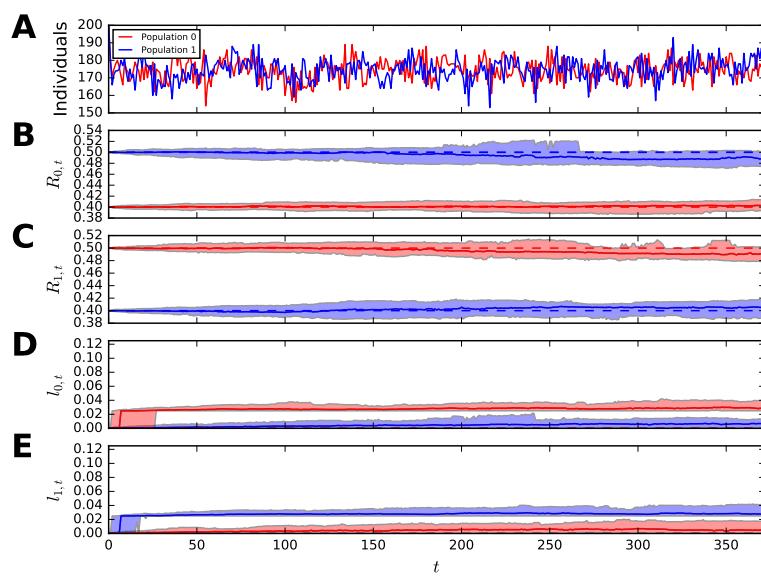
**Figure C.5: Increased niche evolution: facultative mutualism (unmediated).** This is a similar system to Figure C.1, where individuals cannot self-sustain their requirements but whose leakages are evolvable and ecologically adaptable ( $\max l_{k,t} \leq 1/\eta$ ), being therefore under selection. The same parameters as in Figure 3.4 were used, but instead for both populations:  $\sigma_{R,0} = \sigma_{R,1} = \sigma_{l,0} = \sigma_{l,1} = 1 \times 10^{-2}$  and  $l_{0,0} = l_{1,0} = 0.125$ . The spatial distribution at  $t = 0$  was homogeneous. The notation is concomitant with Figure's C.1.



**Figure C.6: Increased variables evolution: facultative mutualism (unmediated).** The evolution of variables of the simulation in Figure C.5 are shown. The notation is concomitant with Figure's C.2.



**Figure C.7: Reduced niche evolution: facultative mutualism (mediated).** This is a similar system to Figure C.1, where individuals cannot self-sustain their requirements but whose leakages are evolvable and ecologically adaptable ( $\max l_{k,t} > 1/\eta$ ), being therefore under selection. The same parameters as in Figure 3.4 were used, but instead  $\sigma_{l,0} = \sigma_{l,1} = 1 \times 10^{-3}$ . The notation is concomitant with Figure's C.1.



**Figure C.8: Reduced variables evolution: facultative mutualism (mediated).** The evolution of variables of the simulation in Figure C.7 are shown. The notation is concomitant with Figure's C.2.

## **Appendix D**

# **Colophon**

This document was set in the Times Roman typeface using L<sup>A</sup>T<sub>E</sub>X and BibT<sub>E</sub>X, being composed with the T<sub>E</sub>XMaker editor.

Computational code was written in the *Python* programming language with specific libraries: numerical processing was done using *Numpy*, optimizations were performed using *Scipy*, and figures were generated using *Matplotlib*.

# References

- Allen, B., Gore, J., and Nowak, M. A. (2013). Spatial dilemmas of diffusible public goods. *eLife*, 2:e01169.
- Blonder, B., Lamanna, C., Violle, C., and Enquist, B. J. (2014). The n-dimensional hypervolume. *Global Ecology and Biogeography*, 23(5):595–609.
- Campbell, K., Vowinckel, J., Mülleider, M., Malmsheimer, S., Lawrence, N., Calvani, E., Miller-Fleming, L., Alam, M. T., Christen, S., Keller, M. A., et al. (2015). Self-establishing communities enable cooperative metabolite exchange in a eukaryote. *eLife*, 4:e09943.
- Campbell, K., Vowinckel, J., and Ralser, M. (2016). Cell-to-cell heterogeneity emerges as consequence of metabolic cooperation in a synthetic yeast community. *Biotechnology Journal*, 11:1–10.
- Celiker, H. and Gore, J. (2014). Clustering in community structure across replicate ecosystems following a long-term bacterial evolution experiment. *Nature communications*, 5.
- Cordero, O. X. and Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology*, 12(4):263–273.
- Cordero, O. X., Ventouras, L.-A., DeLong, E. F., and Polz, M. F. (2012). Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proceedings of the National Academy of Sciences*, 109(49):20059–20064.

- de Oliveira, M. M. and Dickman, R. (2014). Phase diagram of the symbiotic two-species contact process. *Physical Review E*, 90(3):032120.
- Dieckmann, U. and Law, R. (2000). *The geometry of ecological interactions: simplifying spatial complexity*. Cambridge University Press.
- Doebeli, M. and Knowlton, N. (1998). The evolution of interspecific mutualisms. *Proceedings of the National Academy of Sciences*, 95(15):8676–8680.
- Escalante, A. E., Rebolledo-Gómez, M., Benítez, M., and Travisano, M. (2014). Ecological perspectives on synthetic biology: insights from microbial population biology. *Frontiers in microbiology*, 6:143–143.
- Ezoe, H. (2009). Dual lattice model of the evolution of facultative symbiosis with continuous prisoner’s dilemma game. *Journal of theoretical biology*, 259(4):744–750.
- Ezoe, H. and Ikegawa, Y. (2013). Coexistence of mutualists and non-mutualists in a dual-lattice model. *Journal of theoretical biology*, 332:1–8.
- Ferrière, R., Gauduchon, M., and Bronstein, J. L. (2007). Evolution and persistence of obligate mutualists and exploiters: competition for partners and evolutionary immunization. *Ecology Letters*, 10(2):115–126.
- Foster, K. R. and Wenseleers, T. (2006). A general model for the evolution of mutualisms. *Journal of evolutionary biology*, 19(4):1283–1293.
- Freilich, S., Zarecki, R., Eilam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R., and Ruppin, E. (2011). Competitive and cooperative metabolic interactions in bacterial communities. *Nature communications*, 2:589.
- Gertsev, V. and Gertseva, V. (2004). Classification of mathematical models in ecology. *Ecological Modelling*, 178(3):329–334.
- Gilbert, S. F., Bosch, T. C., and Ledón-Rettig, C. (2015). Eco-evo-devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics*, 16(10):611–622.

- Gingerich, P. D. (2009). Rates of evolution. *Annual Review of Ecology, Evolution, and Systematics*, 40:657–675.
- Hegreness, M., Shores, N., Hartl, D., and Kishony, R. (2006). An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science*, 311(5767):1615–1617.
- Hillesland, K. L., Lim, S., Flowers, J. J., Turkarslan, S., Pinel, N., Zane, G. M., Elliott, N., Qin, Y., Wu, L., Baliga, N. S., et al. (2014). Erosion of functional independence early in the evolution of a microbial mutualism. *Proceedings of the National Academy of Sciences*, 111(41):14822–14827.
- Holt, R. D. (2009). Bringing the hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences*, 106(Supplement 2):19659–19665.
- Honegger, R. (2009). Lichen-forming fungi and their photobionts. In *Plant relationships*, pages 307–333. Springer.
- Iwata, S., Kobayashi, K., Higa, S., Yoshimura, J., and Tainaka, K.-i. (2011). A simple population theory for mutualism by the use of lattice gas model. *Ecological modelling*, 222(13):2042–2048.
- Jones, E. I., Afkhami, M. E., Akçay, E., Bronstein, J. L., Bshary, R., Frederickson, M. E., Heath, K. D., Hoeksema, J. D., Ness, J. H., Pankey, M. S., et al. (2015). Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecology Letters*, 18(11):1270–1284.
- Jones, E. I., Bronstein, J. L., and Ferrière, R. (2012). The fundamental role of competition in the ecology and evolution of mutualisms. *Annals of the New York Academy of Sciences*, 1256(1):66–88.
- Karmon, A. and Pilpel, Y. (2016). Biological causal links on physiological and evolutionary time scales. *eLife*, 5:e14424.
- Klitgord, N. and Segre, D. (2010). Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol*, 6(11):e1001002.

- Lanfear, R., Kokko, H., and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology & Evolution*, 29(1):33–41.
- Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T., and Barraclough, T. G. (2012). Species interactions alter evolutionary responses to a novel environment. *PLoS Biol*, 10(5):e1001330.
- Lenski, R. E. and Travisano, M. (1994). Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences*, 91(15):6808–6814.
- Lieberman, E., Hauert, C., and Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316.
- Lutzoni, F., Pagel, M., and Reeb, V. (2001). Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, 411(6840):937–940.
- Maire, T. and Youk, H. (2015). Molecular-level tuning of cellular autonomy controls the collective behaviors of cell populations. *Cell systems*, 1(5):349–360.
- Mathiesen, J., Mitarai, N., Sneppen, K., and Trusina, A. (2011). Ecosystems with mutually exclusive interactions self-organize to a state of high diversity. *Physical review letters*, 107(18):188101.
- Mee, M. T. and Wang, H. H. (2012). Engineering ecosystems and synthetic ecologies. *Molecular BioSystems*, 8(10):2470–2483.
- Momeni, B., Chen, C.-C., Hillesland, K. L., Waite, A., and Shou, W. (2011). Using artificial systems to explore the ecology and evolution of symbioses. *Cellular and molecular life sciences*, 68(8):1353–1368.
- Momeni, B. and Shou, W. (2016). The validity of pairwise models in predicting community dynamics. *bioRxiv*, page 060988.
- Momeni, B., Waite, A. J., and Shou, W. (2013). Spatial self-organization favors heterotypic cooperation over cheating. *eLife*, 2:e00960.
- Nowak, M. A. (2006). *Evolutionary dynamics*. Harvard University Press.

- Page, K. M. and Nowak, M. A. (2002). Unifying evolutionary dynamics. *Journal of Theoretical Biology*, 219(1):93–98.
- Peña, J., Wu, B., Arranz, J., and Traulsen, A. (2016). Evolutionary games of multi-player cooperation on graphs. *PLoS Comput Biol*, 12(8):e1005059.
- Pelletier, F., Garant, D., and Hendry, A. (2009). Eco-evolutionary dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1523):1483–1489.
- Pérez-Escudero, A., Rivera-Alba, M., and de Polavieja, G. G. (2009). Structure of deviations from optimality in biological systems. *Proceedings of the National Academy of Sciences*, 106(48):20544–20549.
- Porcar, M., Latorre, A., and Moya, A. (2013). What symbionts teach us about modularity. *Frontiers in bioengineering and biotechnology*, 1.
- Post, D. M. and Palkovacs, E. P. (2009). Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1523):1629–1640.
- Ratzke, C. and Gore, J. (2016). Self-organized patchiness facilitates survival in a co-operatively growing bacillus subtilis population. *Nature Microbiology*, 1:16022.
- Rauch, J., Kondev, J., and Sanchez, A. (2016). A tradeoff between the ecological and evolutionary stabilities of public goods genes in microbial populations. *bioRxiv*, page 039784.
- Sachs, J. L. and Simms, E. L. (2006). Pathways to mutualism breakdown. *Trends in Ecology & Evolution*, 21(10):585–592.
- Shou, W., Ram, S., and Vilar, J. M. (2007). Synthetic cooperation in engineered yeast populations. *Proceedings of the National Academy of Sciences*, 104(6):1877–1882.
- Tainaka, K.-I. and Hashimoto, T. (2016). A theory of ratio selection-lattice model for obligate mutualism. *Open Journal of Ecology*, 6(06):303.

- Tikhonov, M. (2016). Community-level cohesion without cooperation. *eLife*, 5:e15747.
- Tingley, M. W., Monahan, W. B., Beissinger, S. R., and Moritz, C. (2009). Birds track their grinnellian niche through a century of climate change. *Proceedings of the National Academy of Sciences*, 106(Supplement 2):19637–19643.
- Waite, A. J. and Shou, W. (2012). Adaptation to a new environment allows cooperators to purge cheaters stochastically. *Proceedings of the National Academy of Sciences*, 109(47):19079–19086.
- Wakano, J. Y., Nowak, M. A., and Hauert, C. (2009). Spatial dynamics of ecological public goods. *Proceedings of the National Academy of Sciences*, 106(19):7910–7914.
- Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., Brown, S. P., Momeni, B., Shou, W., et al. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME journal*.
- Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., Damschen, E. I., Jonathan Davies, T., Grytnes, J.-A., Harrison, S. P., et al. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, 13(10):1310–1324.
- Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., and Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412(6844):331–333.
- Yamamura, N., Higashi, M., Behera, N., and Wakano, J. Y. (2004). Evolution of mutualism through spatial effects. *Journal of Theoretical Biology*, 226(4):421–428.
- Yuan, X., Xiao, S., and Taylor, T. N. (2005). Lichen-like symbiosis 600 million years ago. *Science*, 308(5724):1017–1020.

- Zee, P. C. and Bever, J. D. (2014). Joint evolution of kin recognition and cooperation in spatially structured rhizobium populations. *PLoS one*, 9(4):e95141.