



# The role of epistasis on the evolution of recombination in host–parasite coevolution

Roger D. Kouyos<sup>a,\*</sup>, Marcel Salathé<sup>b</sup>, Sarah P. Otto<sup>c</sup>, Sebastian Bonhoeffer<sup>a</sup>

<sup>a</sup> Institute of Integrative Biology, ETH Zürich, ETH-Zentrum CHN, 8092 Zürich, Switzerland

<sup>b</sup> Department of Biology, Stanford University, Stanford, CA 94305-5020, USA

<sup>c</sup> Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

## ARTICLE INFO

### Article history:

Received 5 May 2008

Available online 14 October 2008

### Keywords:

Red queen

Host–parasite interactions

Coevolution

Genetic systems

Sexual reproduction

Population genetics

## ABSTRACT

Antagonistic coevolution between hosts and parasites is known to affect selection on recombination in hosts. The Red Queen Hypothesis (RQH) posits that genetic shuffling is beneficial for hosts because it quickly creates resistant genotypes. Indeed, a large body of theoretical studies have shown that for many models of the genetic interaction between host and parasite, the coevolutionary dynamics of hosts and parasites generate selection for recombination or sexual reproduction. Here we investigate models in which the effect of the host on the parasite (and vice versa) depend approximately multiplicatively on the number of matched alleles. Contrary to expectation, these models generate a dynamical behavior that strongly selects against recombination/sex. We investigate this atypical behavior analytically and numerically. Specifically we show that two complementary equilibria are responsible for generating strong linkage disequilibria of opposite sign, which in turn causes strong selection against sex. The biological relevance of this finding stems from the fact that these phenomena can also be observed if hosts are attacked by two parasites that affect host fitness independently. Hence the role of the Red Queen Hypothesis in natural host parasite systems where infection by multiple parasites is the rule rather than the exception needs to be reevaluated.

© 2008 Elsevier Inc. All rights reserved.

## 1. Introduction

The interaction between hosts and parasites is one of the driving forces in evolution. In particular, it is well known that host–parasite interactions can affect the evolution of reproductive strategies that involve genetic shuffling. The so-called Red Queen hypothesis (RQH) claims that sex and recombination are maintained because they confer on the hosts an edge in the evolutionary arms race with various parasites (Jaenike, 1978). Although this hypothesis has been shown to work in principle (Hamilton et al., 1990; Peters and Lively, 1999; Schmid-Hempel and Jokela, 2002), the plausibility of the assumptions under which it works are hotly debated both in the theoretical (Gandon and Otto, 2007; Kouyos et al., 2007; Otto and Nuismer, 2004; Peters and Lively, 2007; Salathé et al., 2008) and the experimental literature (Fischer and Schmid-Hempel, 2005; Little, 2002; Lively, 1987; Lively et al., 1990, 1998).

It has been recognized early that the impact of Red Queen dynamics on the evolution of genetic shuffling depends crucially

on the specific type of fitness interaction between host and parasite (Bell and Smith, 1987). The subsequent theoretical discussion on this issue was dominated by the debate over the gene-for-gene and the matching-allele model (Agrawal and Lively, 2002; Frank, 1993a,b, 1994, 1996a,b; Parker, 1994, 1996): Whereas the plant-specific gene-for-gene model typically selects against genetic shuffling (Otto and Nuismer, 2004; Parker, 1994, 1996), most models in which the RQH has been shown to work belong to the matching-allele class; i.e. the fitness of host and parasite depend on the number of host loci matched by the parasite (Kouyos et al., 2007; Otto and Nuismer, 2004; Peters and Lively, 1999, 2007; Salathé et al., 2008; Schmid-Hempel and Jokela, 2002). However, even matching-allele models do not always select for higher rates of genetic shuffling; instead, the direction of evolution depends crucially on the fitness function; i.e. on how the number of matched loci determines the fitness of host and parasite (Kouyos et al., 2007; Salathé et al., 2008).

The building blocks of matching-allele models are pairs of loci (one such pair consists of one locus in the host and one corresponding locus in the parasite) that determine the outcome of infection. Usually, a host–parasite interaction is governed by several of those pairs. One important criterion to classify different matching-alleles (MA) models is the degree to which epistatic

\* Corresponding author.

E-mail address: [roger.kouyos@env.ethz.ch](mailto:roger.kouyos@env.ethz.ch) (R.D. Kouyos).

interactions between the different pairs determine the outcome of the host–parasite interaction. The most commonly used MA model assumes that the parasite can infect the host only if it matches the host at all interaction loci, thus assuming a very strong epistatic interaction between alleles at different loci. The other extreme is the so-called multiplicative matching-alleles model (MMA) where the number of matched alleles determines the fitness of host and parasite in a multiplicative way. This type of model is of particular relevance because it constitutes a “null-model” as it involves no implicit epistatic interactions between alleles at different loci. Other matching-alleles models can be obtained by adding an epistatic interaction to the MMA. Intuitively epistasis seems to play an important role for the RQH, because the hypothesis is based on the phase shift between the oscillations of epistasis and LD (Peters and Lively, 1999). Accordingly, several experimental studies have attempted to determine the amount of epistasis between the loci that govern the host parasite interaction (Kover and Caicedo, 2001; Wilfert et al., 2007a,b). Epistasis in host-parasite systems can come from two distinct sources. On one hand epistasis in the host population can be generated through statistical associations in the parasite population and vice versa. On the other hand epistasis can come from the direct genetic interactions between genes that determine the outcome of infection for an individual host and parasite combination. From a theoretical point of view, the relevance of the latter source of epistasis remains unclear and is the subject of this study.

Recently, we studied the evolution of recombination for a wide range of matching-alleles models, including both the MA and the MMA (Kouyos et al., 2007). We found that higher rates of genetic shuffling are strongly selected against in the MMA and closely related models. We related this effect to the observation that the MMA-systems seemed to converge to a state with strong positive or negative linkage disequilibrium (LD). In the example considered in this study, linkage between selected loci was strong and we showed that the sign of the LD depends on the initial conditions, but it does not change during the course of a simulation. By contrast, the only other theoretical study that considered the MMA (Otto and Nuismer, 2004), found that for this interaction model neither higher nor lower recombination rates were favored.

In multiplicative one-species models, both types of behaviors have been observed. On the one hand, Maynard-Smith (Smith, 1968) showed that in haploid systems with multiplicative selection no linkage disequilibrium is built up. Hence recombination has no effect in such systems. In diploid models with multiplicative selection, on the other hand, stable equilibria with strong LD have been described as high complementarity equilibria (HCE) (Bodmer and Felsenstein, 1967; Feldman et al., 1974; Franklin and Lewontin, 1970).

The potential relevance of HCE for the Red Queen Hypothesis is that, according to the reduction principle (Altenberg and Feldman, 1987), such strong and constant LD are expected to result in selection against genetic shuffling. Typically, host–parasite coevolution has been thought to result in oscillating LD, which can potentially favor the evolutionary maintenance of recombination. The HCE represent an alternative type of dynamic behavior, which has the opposite effect on the evolution of recombination. Assessing the impact of host–parasite interactions on the evolution of recombination therefore crucially requires knowledge of the conditions under which these opposite dynamical behaviors occur.

Although we mentioned the occurrence of such HCE in a previous publication (Kouyos et al., 2007), the conditions under which they occur has not been analyzed yet. In this article we extend (Kouyos et al., 2007) in several ways. First, we determine (i) the range of matching-alleles models where this behavior occurs and (ii) how mutation and recombination affect its occurrence and its strength. To this end we employ both an analytical approach

**Table 1**  
Model parameters.

Symbol	Explanation
$w_i^h$	Fitness of host $i$ interacting with parasite $j$
$w_j^p$	Fitness of parasite $i$ interacting with host $j$
$s_j$	Strength of selection on the host if $j$ loci are matched by the parasite
$m_h$	Mutation rate for the host
$m_p$	Mutation rate for the parasite
$r_h$	Recombination rate for the host
$r_p$	Recombination rate for the parasite
$n_{PG}$	Number of parasite generations per host generation

and numerical simulations. Second, we investigate the structural robustness of our findings, by considering both a discrete-time and a continuous-time version of our model. Third, we show that strong LD of constant sign can occur even when a host interacts with two different parasites that affect its fitness in a multiplicative way.

## 2. Models

Our model describes a haploid host interacting with a haploid parasite (see Table 1 for a description of the model parameters). The genomes of hosts and parasites have two biallelic loci with alleles 0 and 1. The state of the system is thus given by the host and parasite frequencies  $[h; p] = [h_1, h_2, h_3, h_4; p_1, p_2, p_3, p_4]$ , where the indices 1–4 correspond to the haplotypes 00, 01, 10, and 11 respectively. Statistical associations between the alleles at the two loci are measured by the linkage disequilibria (LD)  $D_h = h_1h_4 - h_2h_3$  in the host and  $D_p = p_1p_4 - p_2p_3$  in the parasite. Unless stated otherwise, the expression “LD” refers in the following to the LD in the host.

The processes of mutation and recombination are modeled in the following way: mutation rates (host:  $m_h$ , parasite:  $m_p$ ) and recombination rates (host:  $r_h$ , parasite:  $r_p$ ) are fixed. Mutation changes the allele at a given locus with a rate  $m$  (if not stated differently we assume  $m = 10^{-5}$ ) independently of the occurrence of mutations at other loci. Forward and backward mutation rate are equal (i.e. mutation is symmetric and bidirectional). The only effect of recombination is to reduce the linkage disequilibrium ( $D = h_1h_4 - h_2h_3$  for hosts and  $D = p_1p_4 - p_2p_3$  for parasites) by a factor  $(1 - r)$ , where  $r$  is the recombination rate. The process of selection is equivalent to the one described in Kouyos et al. (2007). For the sake of completeness, a full description of selection is given below.

### 2.1. Discrete-time model (DM)

In the discrete-time model, one step corresponds to one parasite generation. We describe different generation times for hosts and parasites, by replacing only a fraction  $\alpha = 1/n_{PG}$  of the host population after one parasite generation;  $n_{PG}$  corresponds thus to the number of parasite generations per host generation. If  $RMS$  denotes the successive action of selection ( $S$ ), mutation ( $M$ ), and recombination ( $R$ ) on the genotype frequencies (with a subscript  $h$  for host and  $p$  for parasite), we can express the change of host and parasite genotype in a time step as:

$$\begin{aligned} h(i+1) &= (1 - \alpha)h(i) + \alpha RMS_h([h(i); p(i)]) \\ p(i+1) &= RMS_p([h(i); p(i)]) \end{aligned} \quad (DM)$$

Note that formula (DM) implies that, at each time step, the whole parasite population (but only a fraction  $\alpha$  of the host population) experiences selection. Thus for  $\alpha < 1$ , parasites evolve faster than hosts. By contrast, Otto and Nuismer (2004) assume that selection acts on the entire host population.

## 2.2. Continuous-time model (CM)

The analogous equations in continuous time are given by

$$\begin{aligned} \frac{dh(t)}{dt} &= \alpha(RMS_h([h(t); p(t)]) - h(t)) \\ \frac{dp(t)}{dt} &= (RMS_p([h(t); p(t)]) - p(t)). \end{aligned} \quad (\text{CM})$$

These equations are obtained by taking the limit of infinitesimally small time steps  $dt$ . In each time step a fraction  $dt$  of parasites and  $\alpha dt$  of hosts is replaced (according to the discrete time equations); i.e.

$$\begin{aligned} h(t + dt) &= (1 - \alpha dt)h(t) + \alpha dt RMS_h([h(t); p(t)]) \\ p(t + dt) &= (1 - dt)p(t) + dt RMS_p([h(t); p(t)]) \end{aligned}$$

or

$$\begin{aligned} \frac{h(t + dt) - h(t)}{dt} &= \alpha(RMS_h([h(t); p(t)]) - h(t)) \\ \frac{p(t + dt) - p(t)}{dt} &= (RMS_p([h(t); p(t)]) - p(t)). \end{aligned}$$

For  $dt = 1$  this corresponds to the discrete model. In the limit  $dt \rightarrow 0$  (i.e. continuous time) this equation becomes the dynamical equation (CM).

## 2.3. Selection and interaction models

Selection is determined by the fitness matrices  $w_{ij}^H$  and  $w_{ij}^P$ . Specifically,  $w_{ij}^H$  denotes the fitness of a host-genotype  $i$  interacting with a parasite-genotype  $j$  and  $w_{ij}^P$  denotes the fitness of the parasite-genotype  $i$  interacting with a host genotype  $j$ . Since the interaction probability for host  $i$  and parasite  $j$  is proportional to their frequencies,  $h_i$  and  $p_j$ , the fitness of the host-genotype  $i$  reads

$$w_i^H = \sum_j w_{ij} p_j$$

and thus the genotype frequencies of the hosts that have undergone selection read

$$h'_i = S_h([h : p])_i = h_i \frac{w_i^H}{\sum_k w_k^H h_k}.$$

The impact of selection on the parasite frequencies is calculated analogously.

We consider host-parasite interactions of the generalized matching allele type (Kouyos et al., 2007; Salathé et al., 2008). These models assume that the outcome of the infection is determined by the number of matched alleles between host and parasite (i.e.  $w_{ij}^H = w_n^H = (1 - s_n)$ , where  $n$  is the number of matched loci) and that, for a given interaction, the fitness of host and parasite are inversely proportional ( $w_n^P = c/w_n^H$ ). For this generalized matching allele model, the fitness of a host matched at no, one or two loci is given by  $w_0^H = 1$ ,  $w_1^H = 1 - s_1$  and  $w_2^H = 1 - s_2$ . The corresponding fitnesses of the parasite are inversely related to that of the host and are given by  $w_0^P = 1 - s_2$ ,  $w_1^P = (1 - s_2)/(1 - s_1)$ , and  $w_2^P = 1$ . The parameters  $s_1$  and  $s_2$  specify a broad class of interaction types, including the matching-alleles model used in most RQH studies (Peters and Lively, 1999, 2007; Schmid-Hempel and Jokela, 2002) in which the parasite can infect the host only if it matches at all loci (i.e.  $s_1 = 0$ ) and the multiplicative matching allele model (given by  $w_2 = w_1^2$  or equivalently  $s_2 = 2s_1 - s_1^2$ ) in which the number of matched alleles determines host and parasite fitness in a multiplicative way (see section Models for a more detailed description of the models).

## 2.4. Independent parasites

In this model, a host with two biallelic loci interacts with two (independent) parasites, A and B, each having one biallelic locus. Parasite A interacts with the first host-locus, and parasite B with the second host locus. The state of the system is given by the genotype frequencies for hosts and parasites:

$$[h; pA; pB] = [h_1, h_2, h_3, h_4; pA_1, pA_2; pB_1, pB_2],$$

where, for the hosts, the indices (1, 2, 3, 4) correspond to the genotypes (00, 10, 01, 11), and for the parasites the indices (1, 2) correspond to genotypes (0, 1). Mutation and recombination are the same as in the single species model (with the only exception that parasites do not recombine because they have only one locus). Only the selection process is different. In the one-host/one-parasite model, a host of genotype  $ij$  interacts with a parasite of genotype  $kl$ . Here, in the two-parasite model, the host  $ij$  interacts with two parasites: A of genotype  $k$  and B of genotype  $l$ . A interacts with the first host locus, B with the second; i.e. A “matches” the host if  $i = k$  and B matches the host if  $j = l$ . The number of matched loci determines the host fitness in a similar way as for the MMA: if a host is matched by neither of the two parasites, then it has (maximal fitness) 1. If it is matched by one parasite but not by the other, it has fitness  $1 - s$ . If it is matched by both parasites it has fitness  $(1 - s)^2$ . The parasite, on the other hand, has (maximal) fitness 1 if it matches the host at the corresponding locus and fitness  $1 - s$  if does not match the host. Thus, the parasite frequencies determine the host fitness in the following way

$$w_i^H = \sum_{jk} (1 - s)^{A(i,j)+B(i,k)} pA_j pB_k$$

where the function  $A(i, j)$  is 1 if the parasite  $j$  matches the host  $i$  at the first locus and 0 else; and  $B(i, k)$  is 1 if the parasite  $k$  matches the host  $i$  at the second locus and 0 otherwise. The host frequencies in turn determine the parasite fitness

$$w_i^{PA} = \sum_j (1 - s)^{1-A(j,i)} h_j \quad \text{and} \quad w_i^{PB} = \sum_j (1 - s)^{1-B(j,i)} h_j.$$

## 3. Results

In the first part of this analysis, we focus on a continuous-time model (CM, see section Models), because its behavior allows a simpler interpretation. In a second step, we discuss the extent to which the corresponding discrete-time model (DM) exhibits a different behavior. Unless stated otherwise, we assume that hosts have a five times longer generation time than parasites ( $n_{pg} = 5$ , see Methods). We start by studying the high complementarity (HC) behavior for the simplest case in which only selection affects the gene frequencies and then successively consider the impact of mutation and recombination.

### 3.1. The high-complementarity region

The HC-behavior is strongly related to the existence and stability of the equilibria of the system. Because of its perfect symmetry, the system has the obvious “central” equilibrium

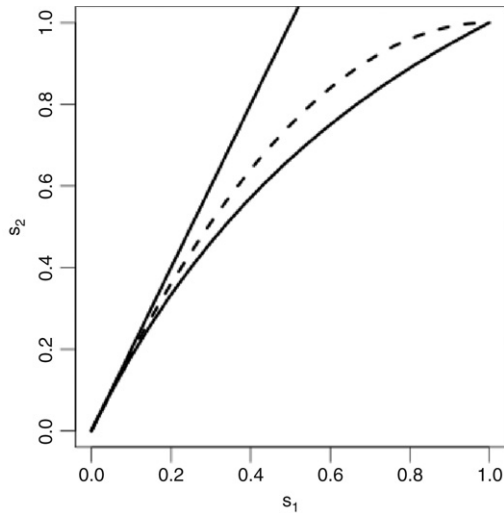
$$[h_c; p_c] = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}; \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right]$$

which corresponds to an LD of zero. In the absence of mutation and recombination, there are in addition two so-called “Highly-Complementary Equilibria” (HCE) given by

$$[h_{HCE+}; p_{HCE+}] = \left[ \frac{1}{2}, 0, 0, \frac{1}{2}; \frac{1}{2}, 0, 0, \frac{1}{2} \right]$$

and

$$[h_{HCE-}; p_{HCE-}] = \left[ 0, \frac{1}{2}, \frac{1}{2}, 0; 0, \frac{1}{2}, \frac{1}{2}, 0 \right],$$



**Fig. 1.** Illustration of the HC region. The HC region lies between the two solid curves (upper curve:  $s_2 = 2s_1$ ; lower curve:  $s_2 = 2s_1/(1 + s_1)$ ). The dashed curve corresponds to the MMA, i.e.  $s_2 = 2s_1 - s_1^2$  and always lies within the HC region..

which correspond to maximally positive and negative LD, respectively. There are further equilibria at the edges, e.g.  $[h; p] = [1, 0, 0, 0; 0, 0, 1, 0]$  but as these are unstable and disappear as soon as mutation is introduced, we do not discuss them further.

It is instructive to consider the stability of the HCE for different interaction models (given by  $s_1$  and  $s_2$ ; see Models). Linearization of the continuous-time model (CM) reveals three pairs of identical eigenvalues, whose real parts read

$$\text{Re}(\lambda_{1,2}) = \frac{s_2 - 2s_1}{2 - s_2}; \quad \text{Re}(\lambda_{3,4}) = \frac{2s_1 - s_2(1 + s_1)}{(1 - s_1)(2 - s_2)};$$

$$\text{Re}(\lambda_{5,6}) = 0.$$

These expressions allow us to distinguish between two regions in the parameter-space (see Fig. 1): If  $2s_1/(1 + s_1) < s_2 < 2s_1$ , then the HCE are neutrally stable. By contrast, the HCE are unstable for all interaction models outside the region between the lines  $s_2 = 2s_1$  and  $s_2 = 2s_1/(1 + s_1)$ . Thus, the HCE can persist only in the first region, which we will therefore refer to as the ‘‘HC region’’. Note that the MMA always lies in the HC region (Fig. 1). The central equilibrium, on the other hand, is neutrally stable outside the HCE region and unstable within the HCE region.

### 3.2. The impact of mutation

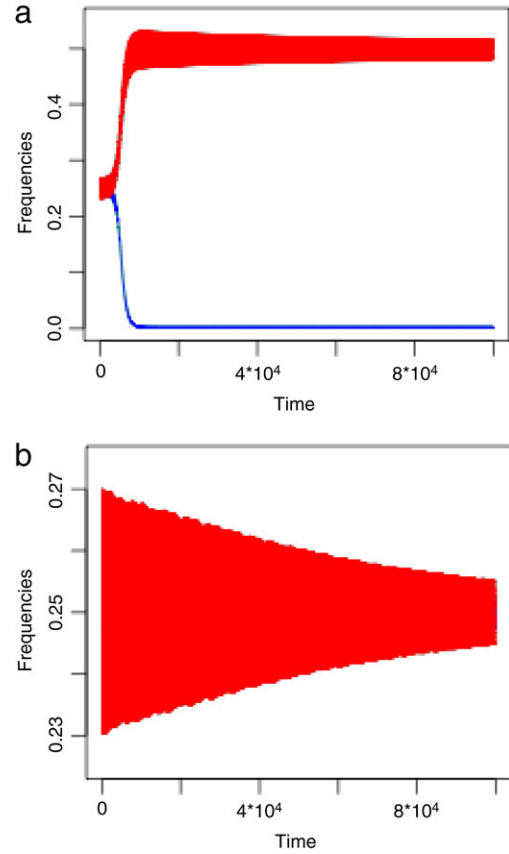
When mutation is introduced, only the central and, for some interaction models, the HC equilibria remain (see Appendix). Outside the HC region, the only equilibrium is the central equilibrium. If selection is strong in comparison to mutation (see Appendix), the system has three equilibria within the HC region: the central equilibrium and two HC equilibria, which lie in the vicinity of the HCE-points for no mutation ( $m = 0$ ). The HC equilibria are given by

$$[h_{HCE+}; p_{HCE+}] = \left[ \frac{1}{2} - \varepsilon_1, \varepsilon_1, \varepsilon_1, \frac{1}{2} - \varepsilon_1; \frac{1}{2} - \varepsilon_2, \varepsilon_2, \varepsilon_2, \frac{1}{2} - \varepsilon_2 \right]$$

and

$$[h_{HCE-}; p_{HCE-}] = \left[ \varepsilon_1, \frac{1}{2} - \varepsilon_1, \frac{1}{2} - \varepsilon_1, \varepsilon_1; \varepsilon_2, \frac{1}{2} - \varepsilon_2, \frac{1}{2} - \varepsilon_2, \varepsilon_2 \right],$$

where  $\varepsilon_1$  and  $\varepsilon_2$  are of the order of the mutation rate  $m$ . Importantly, these equilibria are stable throughout the HC region. By contrast, the central equilibrium is stable outside but unstable



**Fig. 2.** HCE in the continuous-time model. Panel (a) shows the dynamics of the host-genotype frequencies ( $h_1$  is cyan;  $h_2$  is green;  $h_3$  is blue;  $h_4$  is red) for a typical realization of the MMA model (with  $s = 0.1$ ); i.e. an interaction model within the HCE zone. Due to the overlap of the cyan (green) and the red (blue) lines only the red (blue) line is visible. For the initial conditions chosen the frequencies converge to the HCE point with positive LD. Panel (b) corresponds to a system outside the HC region ( $s_1 = 0.1$ ;  $s_2 = 0.21$ ). Due to the overlap of the lines, only the red line is visible. In this case the frequencies converge to the central equilibrium. Parameters:  $m = 10^{-5}$ ;  $r = 0$ ;  $n_{PG} = 5$ ; initial conditions: Allele frequencies 0.51 and 0.53 at the two host loci and 0.5 at the two parasite loci. The host and the parasite population are initiated in linkage equilibrium.

within the HC region. Thus, the stability analysis suggests that the system approaches the central equilibrium outside the HC region and one of the HCE equilibria within the HC region. This is in agreement with numerical simulations (see for example Fig. 2).

The above conclusions hold only if selection is not too weak compared with mutation. If mutation is too strong, then HCE do not exist. We analyzed this phenomenon for the MMA model (which lies in the center of the HC region). There, the criterion for the existence of HCE reads approximately  $s^2 > 16m$  (see Appendix); hence the squared selection coefficient must be considerably larger than the mutation rate. If selection is above the critical value then the LD at the HCE increases steeply as a function of  $s$  and converges to the maximal value  $\frac{1}{4}$  (see Fig. 3). Altogether, we find that the system exhibits HCE-behavior if selection is not too weak and the interaction model lies within the HC region.

### 3.3. The impact of recombination

Recombination acts to reduce LD and should thus weaken the strength and range of the HC behavior, as found in one-species models (Franklin and Lewontin, 1970). For the sake of simplicity, we restrict our analysis in this part to the MMA. Numerical simulations suggest that the behavior is qualitatively

















