Coarse Graining in an Evolutionary Algorithm with Recombination, Duplication and Inversion

C.R. Stephens
Dept. of Computer Science, University of Essex
Wivenhoe, CO4 3SQ, UK
Instituto de Ciencias Nucleares, UNAM
A. Postal 70-543, México, D.F. 04510
csteph@essex.ac.uk

R. Poli
Dept. of Computer Science
University of Essex
Wivenhoe, CO4 3SQ, UK
rpoli@essex.ac.uk

Abstract- A generalised form of recombination, wherein an offspring can be formed from *any* of the genetic material of the parents, is analysed in the context of a two-locus recombinative GA. A complete, exact solution, is derived, showing how the dynamical behaviour is radically different to that of homologous crossover. Inversion is shown to potentially introduce oscillations in the dynamics, while gene duplication leads to an asymmetry between homogeneous and heterogeneous strings. All non-homologous operators lead to allele "diffusion" along the chromosome. We discuss how inferences from the two-locus results extend to the case of a recombinative GA with selection and more than two loci.

1 Introduction

Over the last few years coarse-grained formulations of the dynamics of evolutionary algorithms (EAs) have been seen to proffer many advantages relative to "microscopic" ones based on the string/chromosome degrees of freedom. These benefits have been exhibited, not only in the standard GA [12], but also in variable-length GAs/linear GP and GP itself [8]. The main advantage is the simpler and deeper understanding of the role of homologous recombination they provide [12, 8], wherein the most appropriate effective degrees of freedom for describing the evolution of the system are not strings/chromosomes, but coarse-grained "building blocks", with which the EA builds optimal solutions. The form that the building blocks take depends on the representation used. For instance, in GAs they are a particular subset of schemata that form an alternative and more appropriate basis - the Building Block Basis (BBB) [2]. In the case of variable-length strings and trees, they are generalisations of those found in fixed length GAs - Building Block Hyperschemata [8].

Building Block schemata are the most appropriate effective degrees of freedom for GAs for any mask-based (homologous) recombination operator.[1] However, in nature there are many more ways of combining parental genetic material into an offspring than just homologous crossover, many of which have been used in EAs. Gene duplication, for example, has been studied in biology [3] as well as in the context of GAs [10] and GP [7], while inversion was one of the operators used by Holland [6] in the original formulation of the GA. Additionally, there seems to have been little to no theoretical analysis in the EC literature concerning inversion and duplication, at least not based on an underlying exact dynamical model.

In [9], an exact model for a fixed length GA was derived and studied that introduced a generalised form of recombination that could account for *any* redistribution of parental genes into the offspring. This required the generalisation of the concept of a crossover mask to that of a Generalised Crossover Mask (GCM), with an associated Generalised Recombination Distribution (GRD). This generalised recombination subsumes many other forms, including homologous crossover and inversion as well as fixed-length versions of gene duplication and deletion. It was shown that a coarse grained version of the dynamical equations led to much greater transparency and simplicity of the underlying dynamics, leading to the possibility of a better theoretical understanding of the intrinsic biases associated with these operators which could then be turned into recipes for practitioners.

In this paper we study this new formulation in the context of a two-locus model. Of course, one might question to what extent a two-locus model can illuminate the more complicated multi-locus case. It is wise to remember however, that in population biology such models have played a crucial role permitting the qualitative, and sometime quantitative, analysis of a host of important phenomena (see for instance, [1] and references therein). Even in EC, such models have made important appearances, such as in the deceptive two-bit problem [5] and in previous analyses of the effects of recominbation and mutation [11]. The model we will present has the advantage of being exactly soluble, while at the same time being quite transparent. Additionally, all the interesting phenomena observed in the present model are also present in the case of multi-locus models where analysis is much more complicated, although a more formal mathematical analysis [9] yields many results of interest for the asymptotics of such EAs.

2 Generalised Recombination Distributions

Standard recombination can be succinctly modelled using the concept of a recombination mask, which is used to indicate from which parent to take an allele for each available locus. A mask, $\mathbf{m}$, for strings of length ℓ, can be represented by an ℓ-dimensional vector $\mathbf{m} = (m_1, m_2, \ldots, m_\ell)$,

[1] Mutation also looks simpler in the BBB than in the string basis, though not as simple as in the Walsh basis.

where $m_i = 0,\ 1$ indicates from which parent the ith allele is taken - 0 meaning take it from the ith locus of the first parent and 1 from the ith locus of the second parent. The total number of possible masks is 2^ℓ. Associated with them is a recombination distribution, denoted by $p_c(\mathbf{m})$. If we take the probability to implement crossover as p_{xo}, then $p_c(\mathbf{m})$ is the conditional probability for choosing the mask $\mathbf{m}$ *given*, i.e. conditioned on, the fact that crossover was chosen in the first place. Hence, $\sum_{\mathbf{m}} p_c(\mathbf{m}) = 1$ and $p_{xo} \times p_c(\mathbf{m})$ is the probability to crossover using the mask $\mathbf{m}$.

Recombination masks are sufficient to model homologous genetic operators. However, to describe more general operators we need to consider when an allele in one particular locus of the offspring comes from a different locus in either one of the parents. In the context of a fixed length representation this new level of generality can be obtained by introducing a generalisation of a crossover mask - *a generalised crossover mask* (GCM), $\mathbf{v}$, which is an ℓ-dimensional vector that specifies the origin of the alleles in the offspring. To take into account that the ith allele in the offspring could, in principle, come from any locus in the parents, each v_i can take values from $\mathcal{N}_{2\ell} = \{1, \cdots, 2\ell\}$, values from 1 to ℓ denoting that the allele originated in the first parent and values between $\ell+1$ and 2ℓ signifying that it came from the second parent. Thus, for example, for $\ell = 3$, $(1, 5, 3)$ represents a GCM where the first gene of the offspring came from the first gene of the first parent, the second gene from the second of the second parent and the last from the last of the first parent. As the "mask alphabet" is of cardinality 2ℓ rather than two, as in the case of normal crossover masks, the total number of GCMs is $(2\ell)^\ell$. The associated distribution of probabilities, $p_c(\mathbf{v})$, then determines the generalisation of the recombination distribution - the *Generalised Recombination Distribution* (GRD).

Standard recombination masks are associated with crossover vectors where the genes in the offspring are ordered, in that genetic loci i and j in the offspring originated uniquely from loci i and j in the parents. Inversions can be realised by permutations of some or all of the elements of the string ordering $\{v_1, \ldots, v_\ell\}$. For example, $\{5, 1, 3\}$ represents a GCM where the first gene of the offspring came from the second gene of the second parent, the second gene from the first of the first parent and the last from the last of the first parent. There are two forms of duplication: one parent duplication - duplication from the same locus in the same parent - and two-parent duplication - duplication from the same locus but in different parents. The former is manifest in the corresponding crossover vector by the repetition of an element, e.g. $\{1, 1, 6\}$ gives an offspring where both the first and second genes came from the first locus of the first parent. Duplication from different parents can be seen in the crossover vector $\{1, 2, 5\}$, where the second and third genes of the offspring came from the second gene of the first parent and the second gene of the second parent respectively.

Inheritance of a gene from a parent via a recombination vector $\mathbf{v}$ can be denoted by ${\delta_{I_i}}^{J \otimes K(v_i)}$, where δ is the Kronecker delta, which means that the ith locus of the offspring of genotype I is inherited from the locus associated with the ith component of the recombination vector $\mathbf{v}$, $J \otimes K$ representing the 2ℓ-dimensional vector whose first ℓ components represent the loci of the first parent, J, and the second ℓ those of the second parent, K. As the components of $\mathbf{v}$ range over values $[1, 2\ell]$ $J \otimes K(v_i)$ picks out the corresponding component of $J \otimes K$. For instance, for two-bit strings and a recombination vector $\mathbf{v} = (2, 3)$, ${\delta_{I_1}}^{J \otimes K(v_1)} = {\delta_{I_1}}^{J_2}$, i.e. the first bit of the offspring was inherited from the second locus of the first parent, while ${\delta_{I_2}}^{J \otimes K(v_2)} = {\delta_{I_2}}^{K_1}$, i.e. the second bit of the locus was inherited from the first bit of the second parent.

For two bits we can represent the GRD by a set $\{p_{ab}\}$, where the indices a and b take values from one to four, one and two corresponding to the first and second loci of the first parent and three and four the corresponding loci of the second parent. Thus, for example $\{p_{13}\}$ gives the probability for applying the GCM associated with finding the first locus of the offspring from the first locus of the first parent and the second locus from the first locus of the second parent.

3 Evolution Equation in the String Basis

We first write down the exact, finite population equation in the string basis with selection and generalised recombination

$$E(P_I(t+1)) = (1 - p_{xo})P'_I(t) + p_{xo} \sum_{\mathbf{v}} p_c(\mathbf{v}) \sum_J \sum_K {\lambda_I}^{JK}(\mathbf{v}) P'_J(t) P'_K(t) \quad (1)$$

where $P_I(t)$ is the proportion of genotype I at generation t, $E(P_I(t+1))$ being the *expected* proportion of genotype I in generation $t+1$ given $P_I(t)$. Here, I is a multi-index, $I = (I_1, I_2, \ldots, I_\ell)$, P'_I is the probability to select I and ${\lambda_I}^{JK}(\mathbf{v})$ is the conditional probability that the offspring I is formed given the parents J and K and a GCM $\mathbf{v}$. ${\lambda_I}^{JK}(\mathbf{v}) = 0,\ 1$ as either the offspring is formed or it isn't. The first term in (1) arises from the cloning of I while the second term represents all the ways in which I may be constructed from other strings via generalised recombination.

Equation (1) gives the most general way of recombining genetic material from two parents, ranging from a locus by locus partition, such as is the case for a standard recombination mask, to a complete duplication of one particular allele value associated with a particular genetic locus in a particular parent. The most complicated part of equation (1) is the ${\lambda_I}^{JK}(\mathbf{v})$. They can be formally written as

$${\lambda_I}^{JK}(\mathbf{v}) = \prod_{i=1}^{\ell} {\delta_{I_i}}^{J \otimes K(v_i)} \quad (2)$$

As an example, consider for $\ell = 3$ the recombination vector $\mathbf{v} = (3, 1, 5)$. In this case ${\lambda_I}^{JK}(3, 1, 5) = {\delta_{I_1}}^{J_3} {\delta_{I_2}}^{J_1} {\delta_{I_3}}^{K_2}$.

Note that equation (1) is functionally identical to that for the case of standard mask-based crossover [2], the only difference being the different recombination distribution, and

hence the different set of ${\lambda_I}^{JK}(\mathbf{v})$ that are non-zero. As in the standard crossover case for binary strings we have 2^ℓ coupled, first-order difference equations to solve. The chief problem however, is the fact that on the right hand side we have $2^\ell \times 2^\ell \times (2\ell)^\ell = (8\ell)^\ell$ possible contributing terms. For example, for two bits there are sixteen GCMs denoted by $\{(v_1, v_2)\}$, where v_1 and v_2 run over the values $1,\ 2,\ 3$ and 4. The sums over the strings J and K run over the values 1 to $\mathcal{A}^\ell$ for an alphabet of cardinality $\mathcal{A}$. Thus, for an arbitrary GRD, even at the two bit level there are $16 \times 4 \times 4 = 256$ ${\lambda_I}^{JK}(\mathbf{v})$ to compute for a given string I. Symbolically however, the terms are quite simple

$$\begin{aligned}
\text{cloning} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(1,2) = \delta_{I_1 J_1}\delta_{I_2 J_2} && (3)\\
\text{inversion} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(2,1) = \delta_{I_1 J_2}\delta_{I_2 J_1} && (4)\\
\text{crossover} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(1,4) = \delta_{I_1 J_1}\delta_{I_2 K_2} && (5)\\
\text{crossover+inversion} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(4,1) = \delta_{I_1 K_2}\delta_{I_2 J_1} && (6)\\
\text{duplication 1} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(1,1) = \delta_{I_1 J_1}\delta_{I_2 J_1} && (7)\\
\text{duplication 1} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(2,2) = \delta_{I_1 K_1}\delta_{I_2 K_1} && (8)\\
\text{duplication 2} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(1,3) = \delta_{I_1 J_1}\delta_{I_2 K_1} && (9)\\
\text{duplication 2} \quad & {\lambda_{I_1 I_2}}^{J_1 J_2 K_1 K_2}(2,4) = \delta_{I_1 J_2}\delta_{I_2 K_2} && (10)
\end{aligned}$$

where we show only those GCMs that correspond to creation of genotype I using J as the first parent. The corresponding second parent terms can be found by interchanging J and K on the right hand side of (3-10) and letting letting $(v_1, v_2) \rightarrow (v_1', v_2')$ where $v_i' = v_i + 2 \bmod 2$. The meaning of these terms, as alluded to in equations (3-10), is the following: the terms represented by GCMs $(1,2)$ (cloning of first parent) and $(3,4)$ (cloning of second parent) are cloning terms due to the application of a trivial standard crossover mask, where both offspring alleles come from the corresponding loci of only one of the parents. The inversion term is represented by GCMs $(2,1)$ (inversion of first parent) and $(4,3)$ (inversion of second parent). The GCMs $(1,4)$ and $(3,2)$ represent the results of standard one-point crossover, while the GCMs $(4,1)$ and $(2,3)$ represent the results of standard one-point crossover followed by an inversion (or vice versa). The terms denoted by duplication 1 - one-parent duplication - are associated with the GCMs $(1,1)$, $(2,2)$, $(3,3)$ and $(4,4)$ and represent duplication of an allele from a single locus of a single parent. Finally, the duplication 2 - two-parent duplications - GCMs $(1,3)$, $(2,4)$, $(3,1)$ and $(4,2)$ represent gene duplication as well, but where the two genes of the offspring come from the same locus but in different parents.

Substituting these expressions in (1), computing all terms and setting $I_1 = i$ and $I_2 = j$ for conciseness, one finds

$$\begin{aligned}
E(P_{ij}(t+1)) = {} & (1-p_{xo})P'_{ij} + p_{xo}[(p_{11}+p_{33}+p_{22}+p_{44})P'_{ij}\delta_{ij} \\
+ {} & ((p_{11}+p_{33})P'_{i\bar{j}} + (p_{22}+p_{44})P'_{\bar{i}j})\delta_{ij} + (p_{12}+p_{34})P'_{ij} \\
+ {} & (p_{13}+p_{31}+p_{14}+p_{32}+p_{23}+p_{41}+p_{42}+p_{24})P'_{ii}P'_{jj} \\
+ {} & (p_{21}+p_{43})P'_{ji}(t) + (p_{23}+p_{41}+p_{13}+p_{31})P'_{ii}P'_{j\bar{j}} \\
+ {} & (p_{13}+p_{31}+p_{14}+p_{32})P'_{i\bar{i}}P'_{jj} + (p_{13}+p_{31})P'_{i\bar{i}}P'_{j\bar{j}} \\
+ {} & (p_{24}+p_{42}+p_{14}+p_{32})P'_{ii}P'_{\bar{j}\bar{j}} \\
+ {} & (p_{14}+p_{32})P'_{i\bar{i}}P'_{\bar{j}j} + (p_{24}+p_{42}+p_{23}+p_{41})P'_{i\bar{i}}P'_{jj} \\
+ {} & (p_{41}+p_{23})P'_{\bar{i}i}P'_{j\bar{j}} + (p_{42}+p_{24})P'_{\bar{i}i}P'_{\bar{j}j}] \qquad (11)
\end{aligned}$$

where, for simplicity, we are restricting attention to a binary alphabet and $\bar{i}$ signifies the bit complement mod 2 of i. The first term on the right hand side is a cloning term due to the fact that with probability $(1 - p_{xo})$ strings are copied without recombination, P'_{ij} being the probability to select the genotype ij. The meaning of the different terms in (11) is inherited from the meaning of the corresponding terms of the GRD, the p_{ab} being the notation for the GCM probability associated with the GCM (a, b). The Kronecker delta, $\delta_{ij} = 1,\ i = j;\ \delta_{ij} = 0,\ i \neq j$, ensures that the contribution from gene duplication from a single parent is only present for homogeneous offspring, i.e. those with both allele values the same. Note that of the 256 possibilities there are only 44 non-zero terms in (11). However, in order to compute which ones are non-zero all have to be computed. By way of comparison, the canonical GA with one-point crossover, where $p_{14} = p_{32} = 1/2$ with all other GCMs zero, has only 8 of the $2 \times 4 \times 4$ possible terms non-zero.

4 Coarse Grained Evolution Equation

For both homologous and generalized recombination it is clear that there is a great deal of redundancy in the string representation. In the case of homologous recombination it has been found that a coarse grained representation in terms of Building Block schemata makes the dynamics much more transparent, partly due to the fact that the number of terms on the right hand side of (1) reduces to 2^ℓ in the case of homologous crossover which, when compared to 8^ℓ in the string basis, is a substantial reduction in complexity. For a particular type of recombination distribution, such as one-point crossover, where there are only $(\ell - 1)$ non-zero masks, the simplification is even greater, from 8^ℓ to $(\ell - 1)$. One is naturally inclined to ask whether an appropriate simplification can be effected in this more general case.

In [9] it is shown how a coarse graining greatly simplifies (1) by removing the string sums, effectively reducing $(8\ell)^\ell$ terms to $(2\ell)^\ell$. One can illustrate how the coarse graining is effected by considering a particular GCM, such as recombination and inversion with the GRD $\{p_{41}, p_{23}\}$ (for simplicity we put $p_{xo} = 1$). From equation (1)

$$\begin{aligned}
E(P_{I_1 I_2}(t+1)) &= p_{41}\sum_{J_1 J_2}\sum_{K_1 K_2}\delta_{I_1 K_2}\delta_{I_2 J_1}P'_{J_1 J_2}P'_{K_1 K_2} \\
&+ p_{23}\sum_{J_1 J_2}\sum_{K_1 K_2}\delta_{I_1 J_2}\delta_{I_2 K_1}P'_{J_1 J_2}P'_{K_1 K_2} \\
&= p_{41}P'_{I_2 *}P'_{*I_1} + p_{23}P'_{*I_1}P'_{I_2 *} \qquad (12)
\end{aligned}$$

where $*$ is the standard wildcard symbol that represents the fact that the allele value at that locus has been summed over. Similarly, in the general two-locus case one obtains

$$\begin{aligned}
E(P_{ij}(t+1)) = {} & (1-p_{xo})P'_{ij} + p_{xo}[(p_{11}+p_{33})P'_{i*}\delta_{ij} \\
& +(p_{22}+p_{44})P'_{*i}\delta_{ij} + (p_{12}+p_{34})P'_{ij} + (p_{21}+p_{43})P'_{ji} \\
& +(p_{14}+p_{32})P'_{i*}P'_{*j} + (p_{23}+p_{41})P'_{*i}P'_{j*} \\
& +(p_{13}+p_{31})P'_{i*}P'_{j*} + (p_{42}+p_{24})P'_{*i}P'_{*j}] \qquad (13)
\end{aligned}$$

Notice now that there are only 16 terms corresponding to the 16 terms of the GRD. Just as in the case of homologous crossover the sums over J and K have disappeared leading to a reduction from $(8\ell)^\ell$ to $(2\ell)^\ell$ terms, in this case from 256 to 16. This makes manifest that in the case of generalised recombination, just as in the case of homologous crossover, Building Block Schemata and not strings are the appropriate effective degrees of freedom, and that a GA based on such recombination also builds solutions from lower order building blocks. The difference when compared to homologous crossover is simply that there are more building blocks. Normally, with homologous crossover, formation of ij is via the Building Block schemata $i*$ and $*j$. Here, however, the relevant building blocks are $i*$, $*j$, $*i$ and $j*$.

5 Two Locus Solution

One is naturally led to ask: if and how the dynamics of this more general class of GAs differs from that of the canonical GA with homologous recombination? This is naturally a more complicated question, given that the array of genetic operators that we are considering under the rubric of generalised recombination is quite large, including standard mask-based recombination, inversion, different types of gene duplication and combinations of these different operators. In order to understand in this more general context the different biases associated with these operators we investigate the solutions to equation (13) in the infinite population limit in the absence of selection, i.e. $P'_{ij}(t) = P_{ij}(t)$, and where, for simplicity, we set $p_{xo} = 1$.

In the infinite population limit $E(P_I(t+1)) \to P_I(t+1)$ and then (1) and the equations derived from it become deterministic equations for the string and/or schema proportions and describe the corresponding dynamical system. The results derived from the infinite population model neglect the variance inherent in the dynamics due to limited sampling, an effect which is expected to vary as $N^{-1/2}$, where N is the population size. Hence, for large population sizes or short runs one would expect the analysis below to be an accurate representation of what actually occurs. For small populations one would have to consider directly the Markov chain for this model. The transition matrix elements that enter in this case would still be obtained by using a multinomial distribution with success probabilities given by the right hand side of equation (1). Similarly, one would expect our subsequent analysis to give a good qualitative description of the biases engendered by generalized recombination even if selection is included as long as the selection is weak, as will illustrated later.

5.1 Building Block Dynamics

In order to solve (13), just as in the case of standard homologous recombination [12], we need to hierarchically solve for the dynamics of the Building Blocks of the genotype ij. These are: $i*$ and $*i$. From (13) we can determine the equations for the Building Block schemata by considering $P_{i*}(t) = \sum_{j=0,1} P_{ij}(t)$ and $P_{*j}(t) = \sum_{i=0,1} P_{ij}(t)$. The equations for the one-schemata are[2]

$$P_{i*}(t+1) = (p_{1*} + p_{3*})P_{i*}(t) + (p_{2*} + p_{4*})P_{*i}(t) \quad (14)$$
$$P_{*i}(t+1) = (p_{*1} + p_{*3})P_{i*}(t) + (p_{*2} + p_{*4})P_{*i}(t) \quad (15)$$

where we are using a schema-like notation here also for the GCMs, e.g. $p_{1*} = \sum_{i=1}^{4} p_{1i}$. Equations (14) and (15) form a coupled linear system, similar to that of mutation in a one-locus system. To solve these equations we need to determine the eigenvalues and eigenvectors of the matrix

$$\mathbf{W} \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} (p_{1*} + p_{3*}) & (p_{2*} + p_{4*}) \\ (p_{*1} + p_{*3}) & (p_{*2} + p_{*4}) \end{pmatrix} \quad (16)$$

In the case of pure mask-based recombination, only p_{12}, p_{34}, p_{14} and p_{32} are non-zero. Hence, the evolution of the schema $i*$ is independent of the schema $*i$, i.e. the two equations decouple, giving as solution $P_{i*}(t) = P_{*i}(0)$ and $P_{*i}(t) = P_{*i}(0)$. More generally, the eigenvalues of matrix (16) are

$$\lambda_\pm = \frac{(a+d)}{2} \pm \frac{1}{2}((a-d)^2 + 4bc)^{1/2} \quad (17)$$

with corresponding normalised eigenvectors

$$\mathbf{e}_+ \equiv \begin{pmatrix} e_{+1} \\ e_{+2} \end{pmatrix} = N_+ \begin{pmatrix} b \\ (\lambda_+ - a) \end{pmatrix} \quad (18)$$
$$\mathbf{e}_- \equiv \begin{pmatrix} e_{-1} \\ e_{-2} \end{pmatrix} = N_- \begin{pmatrix} b \\ (\lambda_- - a) \end{pmatrix} \quad (19)$$

where $N_+ = ((\lambda_+ - a)^2 + b^2)^{-1/2}$ and $N_- = ((\lambda_- - a)^2 + b^2)^{-1/2}$ are normalisation constants. The transformation matrix $\mathbf{\Lambda} \equiv (\mathbf{e}_+\mathbf{e}_-)^{-1}$, formed from the eigenvectors (18) and (19), diagonalises $\mathbf{W}$ and rotates the vector $\mathbf{P}(t) = (P_{1*}P_{*1})^T \longrightarrow (\tilde{P}_+(t)\tilde{P}_-(t))^T$ such that

$$\tilde{P}_+(t+1) = \lambda_+ \tilde{P}_+(t) \quad (20)$$
$$\tilde{P}_-(t+1) = \lambda_- \tilde{P}_-(t) \quad (21)$$

which can be immediately integrated to yield

$$\tilde{P}_+(t) = \lambda_+^t \tilde{P}_+(0) \quad (22)$$
$$\tilde{P}_-(t) = \lambda_-^t \tilde{P}_-(0) \quad (23)$$

Rotating back now to the original schema basis one finds

$$P_{i*}(t) = Det(\Lambda^{-1})((e_{+1}\lambda_+^t e_{-2} - e_{-1}\lambda_-^t e_{+2})P_{i*}(0) + (-e_{+1}\lambda_+^t e_{-1} + e_{-1}\lambda_-^t e_{+1})P_{*i}(0)) \quad (24)$$
$$P_{*i}(t) = Det(\Lambda^{-1})((e_{-2}\lambda_+^t e_{+2} - e_{+2}\lambda_-^t e_{-2})P_{i*}(0) + (e_{+1}\lambda_+^t e_{-2} - e_{-1}\lambda_-^t e_{+2})P_{*i}(0)) \quad (25)$$

from which one may determine, for instance, the asymptotic behaviour as $t \to \infty$. As P_{i*} and P_{*i} are both probabilities, then $\lambda_\pm \leq 1$ and, in fact, one eigenvalue must be unity due to the row stochasticity of the matrix $\mathbf{W}$, as can be seen from equation (16) noting that, as $\sum_{i=1}^{4} p_{i*} = \sum_{i=1}^{4} p_{*i} =$

[2]Note that i and j are purely symbolic values so that P_{i*} and P_{*i} are suffi cient to cover the 4 possibilities P_{1*}, P_{0*}, P_{*1} and P_{*0}.

1, then $a+b=c+d=1$. Substituting these constraints into (17) one finds

$$\lambda_+ = 1 \tag{26}$$

$$\lambda_- = \frac{1}{2}(a+d-b-c) = \frac{1}{2}[(p_{14}-p_{41}) + (p_{32}-p_{23}) + (p_{34}-p_{43})] \tag{27}$$

with corresponding eigenvectors

$$\mathbf{e}_+ = 2^{-1/2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{28}$$

$$\mathbf{e}_- = (b^2+c^2)^{-1/2}\begin{pmatrix} b \\ -c \end{pmatrix} \tag{29}$$

Hence, the solutions (24) and (25) simplify to

$$P_{i*}(t) = \frac{(cP_{i*}(0)+bP_{*i}(0))}{(b+c)} + \frac{b(a-c)^t}{b+c}(P_{i*}(0)-P_{*i}(0)) \tag{30}$$

$$P_{*i}(t) = \frac{(cP_{i*}(0)+bP_{*i}(0))}{(b+c)} - \frac{c(a-c)^t}{b+c}(P_{i*}(0)-P_{*i}(0)) \tag{31}$$

From this solution we can examine the fixed point. As $|a-c| \leq 1$, except for $a=1$, $c=0$ or $a=0$ and $c=1$, the time dependent term vanishes asymptotically, giving as fixed point

$$P^*_{i*} = \lim_{t\to\infty} P_{i*}(t) = \frac{(cP_{i*}(0)+bP_{*i}(0))}{(b+c)} \tag{32}$$

$$P^*_{*i} = \lim_{t\to\infty} P_{*i}(t) = \frac{(cP_{i*}(0)+bP_{*i}(0))}{(b+c)} \tag{33}$$

Interestingly, in this case the fixed point is the same for the schemata $i*$ or $*i$, though this proportion is approached from opposite directions. The behaviour of the transients on approaching the fixed point also depends sensitively on the value of $(a-c)$.[3] If $(a-c) > 0$ then the fixed point is approached monotonically. However, for $(a-c) < 0$ the factor $(-1)^t$ implies the presence of oscillations. However, as $|a-c| \leq 1$ these oscillations are damped and vanish asymptotically. We will now consider some particular cases of interest.

1. $b=c=0$ - in this case $a=d=1$, $\lambda_\pm = 1$ and $P^*_{i*} = P_{i*}(0)$, $P^*_{*i} = P_{*i}(0)$; thus the initial proportions are preserved. This type of recombination is homologous and preserves gene frequencies at a given locus.

2. $a=d=0$ - here, $b=c=1$, $\lambda_\pm = 1, -1$ and the associated eigenvectors are $\mathbf{e}_+ = (1/\sqrt{2})(1\ 1)^T$ and $\mathbf{e}_- = (1/\sqrt{2})(1\ \ -1)^T$. There is now no fixed point, but rather a cycle of period two, where $P_{i*}(t) = P_{i*}(0)$ for t even and $P_{*i}(0)$ for t odd. Similarly, $P_{*i}(t) = P_{*i}(0)$ for t even and $P_{i*}(0)$ for t odd. This leads to lateral diffusion of alleles along the string from one genetic locus to another.

3. $a=b=c=d=1/2$ - in this case $\lambda_\pm = 1,\ 0$ with the same eigenvectors as for case 2.. Now there are no oscillations ($(a-c)=0$) and the fixed point $P^*_{i*} = P^*_{*i} = (P_{i*}(0)+P_{*i}(0))/2$ is reached after only one generation

4. $P_{i*}(0) = P_{*i}(0)$ - when this condition holds, irrespective of the generalised recombination probabilities, this remains a fixed point. This condition is satisfied both at the centre of the simplex [13] as well as at its vertices. It is equivalent to having equal proportions for heterogeneous genotypes.

We have discussed the asymptotic behaviour in terms of the four parameters a, b, c and d. However, we wish to understand the dynamics in terms of the generalised recombination probabilities, p_{ab}. For case 1. above $(p_{1*}+p_{3*}) = (p_{*2}+p_{*4}) = 1$ and $(p_{*1}+p_{*3}) = (p_{2*}+p_{4*}) = 0$, the latter being equivalent to there being no duplication or inversion or any combination that includes them. This means that there are no genetic operators that lead to lateral diffusion of alleles along the string from one genetic locus to another. The resultant fixed point for the one schemata is on the Robbins/Geiringer manifold [4]. Similarly, for case 2. we have $(p_{1*}+p_{3*}) = (p_{*2}+p_{*4}) = 0$ and $(p_{*1}+p_{*3}) = (p_{2*}+p_{4*}) = 1$. Under these conditions the only non-zero terms are those associated with inversion. There is no homologous recombination or duplication. Thus, pure inversion without duplication or homologous crossover leads to periodic behaviour.

We may also investigate the biases of a particular genetic operator, investigating the solutions in the absence of the other operators. Thus, for instance, for duplication from one parent, then $p_{ii} \neq 0$ while all other generalised recombination probabilities are zero. In this case $a=c=(p_{11}+p_{33})$ and $b=d=(p_{22}+p_{44})=(1-(p_{11}+p_{33}))$. Additionally, we have $\sum_{i=1}^4 p_{ii} = 1$, i.e. $b+c=a+d=1$. Hence, there is no transient term and the fixed point is

$$P^*_{i*} = P^*_{*i} = P_{i*}(0) + (p_{22}+p_{44})(P_{*i}(0)-P_{i*}(0)) \tag{34}$$

which is reached after one generation. For cloning, p_{12} and p_{34} are the only non-zero GCMs, hence, $a=d=1$ and $b=c=0$. In this case the fixed point is trivially

$$P^*_{i*} = P_{i*}(0) \tag{35}$$

$$P^*_{*i} = P_{*i}(0) \tag{36}$$

For inversion, the only non-zero probabilities are p_{21} and p_{43}. In this case $a=d=0$ and $b=c=1$, and the asymptotic behaviour is governed by the two cycle of 2. above with

$$P^*_{i*} = \frac{1}{2}((1+(-1)^t)P_{i*}(0) + (1-(-1)^t)P_{*i}(0)) \tag{37}$$

$$P^*_{*i} = \frac{1}{2}((1-(-1)^t)P_{i*}(0) + (1+(-1)^t)P_{*i}(0)) \tag{38}$$

[3] Note that due to the identities $a+b=1$ and $c+d=1$ this is equivalent to $(b-d)$.

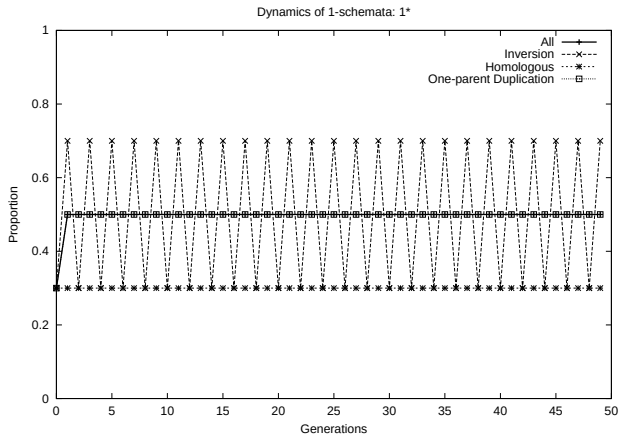

Figure 1: Dynamical evolution of the one-schema $1*$ for different GRDs

For two-parent duplication, the appropriate non-zero recombination probabilities are p_{13}, p_{24}, p_{31} and p_{42}. Hence, $a = c = (p_{13} + p_{31})$ and $b = d = (p_{24} + p_{42})$ with $b + c = a + d = 1$. As $a = c$ there are no transients and the fixed point

$$P^*_{i*} = P^*_{*i} = P_{i*}(0) + (p_{24} + p_{42})(P_{*i}(0) - P_{i*}(0)) \quad (39)$$

is reached after one generation just as in the case of one-parent duplication. Note that this fixed point is of the same form as that found for one parent duplication. For homologous crossover, the non-zero recombination distributions are p_{14} and p_{32}. In this case $a = d = 1$, $b = c = 0$ and the fixed point is the same as that for cloning. Finally, for crossover and inversion the corresponding recombination distributions are p_{41} and p_{23} which implies $a = d = 0$ and $b = c = 1$. In this case the fixed point is the same as that for inversion above.

In Figure 1, using equation (30), we see a graph of the evolution of the one-schema $1*$ for different GRDs. The direct integration of equations (14) and (15) yields exactly the same curves, as expected. The initial condition used is an asymmetric one, where $P_{11}(0) = P_{00}(0) = 0.1$, $P_{01}(0) = 0.6$ and $P_{10}(0) = 0.2$; hence, $P_{1*}(0) = 0.3$. The fixed point behaviour described in points 1. - 4. and equations (34-39) is clearly visible. For one-parent duplication the fixed point is reached after one generation at a value $P^*_{1*} = P_{1*}(0) + P_{*1}(0)/2 = 0.3 + 0.7 = 0.5$. For inversion, one sees the characteristic oscillations between the values 0.3 and 0.7 associated with $P_{1*}(0)$ and $P_{*1}(0)$. For homologous crossover the fixed point is the initial proportion $P_{1*}(0) = 0.3$, i.e. the allele frequency at a given locus is preserved. Finally, considering all GCMs with equal probability - the All curve in Figure 1 - one sees that the system reaches a fixed point in one generation.

The features and fixed points we have just delineated for $\ell = 2$, are also present for $\ell > 2$ and represent qualitatively new phenomena when compared to the normal homologous forms of crossover with which we are familiar. The lateral diffusion of alleles, relative to the homologous case, leads to a fixed point, where for a given offspring locus, the allele frequency at that locus depends not only on the allele frequency at the same parental loci but also on the allele frequencies at other genetic loci. For $\ell > 2$, instead of a pair of linear coupled equations for the one-schemata one has ℓ coupled equations whose solution can be found by solving the corresponding eigensystem.

5.2 Solution for Strings

With the solutions for the one-schemata in hand we can now proceed to determine the solutions for the strings themselves from (13) by substituting in the solutions (30) and (31), which are the contributions from the Building Blocks, to yield

$$P_{ij}(t+1) = (1 - p_{xo})P_{ij} + p_{xo}(p_{12} + p_{34})P_{ij}(t) + p_{xo}(p_{21} + p_{43})P_{ji}(t) + p_{xo}F_{ij}(t) \quad (40)$$

where

$$F_{ij}(t) = (C_{ij} + D_{ij}(a - c)^t + E_{ij}(a - c)^{2t}) \quad (41)$$

and the matrices C_{ij}, D_{ij} and E_{ij} are given by

$$C_{ij} = [(p_{14} + p_{41} + p_{32} + p_{23} + p_{13} + p_{31} + p_{24} + p_{42})A_iA_j + (p_{11} + p_{22} + p_{33} + p_{44})A_i\delta_{ij}] \quad (42)$$

$$\begin{aligned} D_{ij} = [&((p_{14} + p_{32})(bA_jB_i - cB_jA_i) \\ &+(p_{23} + p_{41})(bA_iB_j - cB_iA_j) \\ &+(p_{13} + p_{31})b(A_jB_i - B_jA_i) \\ &-(p_{24} + p_{42})c(A_jB_i - B_jA_i)) \\ &+((p_{11} + p_{33})b - (p_{22} + p_{44})c)B_i\delta_{ij}] \end{aligned} \quad (43)$$

$$E_{ij} = ((p_{13} + p_{31})b^2 + (p_{24} + p_{42})c^2 - (p_{14} + p_{41} + p_{32} + p_{23})bc)B_iB_j \quad (44)$$

where

$$A_i = \frac{(cP_{i*}(0) + bP_{*i}(0))}{(b + c)} \quad (45)$$

$$B_i = \frac{(P_{i*}(0) - P_{*i}(0))}{(b + c)} \quad (46)$$

To solve (40) we need to also have the equation for P_{ji}, which is just (40) with $i \leftrightarrow j$. The matrices C_{ij} and E_{ij} are both symmetric matrices, D_{ij} however is not. Hence, $P_{ji}(t)$ satisfies

$$P_{ji}(t+1) = (1 - p_{xo})P_{ji} + p_{xo}(p_{12} + p_{34})P_{ji}(t) + p_{xo}(p_{21} + p_{43})P_{ij}(t) + p_{xo}F_{ji}(t) \quad (47)$$

where

$$F_{ji}(t) = (C_{ij} + D_{ji}(a - c)^t + E_{ij}(a - c)^{2t}) \quad (48)$$

Equations (40) and (47) are linear coupled inhomogeneous first order difference equations and can be solved in an analogous fashion to that of equations (14) and (15) by determining the corresponding eigensystem. Putting $p_{xo} = 1$ the relevant matrix is

$$\mathbf{W}' = \begin{pmatrix} (p_{12} + p_{34}) & (p_{21} + p_{43}) \\ (p_{21} + p_{43}) & (p_{12} + p_{34}) \end{pmatrix} \quad (49)$$

whose eigenvalues and eigenvectors are

$$\lambda_{\pm} = (p_{12} + p_{34}) \pm (p_{21} + p_{43}) \tag{50}$$

$$\mathbf{e}_{+} = 2^{-1/2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{51}$$

$$\mathbf{e}_{-} = 2^{-1/2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{52}$$

In the eigenvector basis $\mathbf{P}(t) = (P_{ij} \;\; P_{ji})^T \rightarrow (\tilde{P}_{+}(t) \; \tilde{P}_{-}(t))^T$ such that

$$\tilde{P}_{+}(t+1) = \lambda_{+}\tilde{P}_{+}(t) + \tilde{F}_{+}(t) \tag{53}$$

$$\tilde{P}_{-}(t+1) = \lambda_{-}\tilde{P}_{-}(t) + \tilde{F}_{-}(t) \tag{54}$$

where

$$\tilde{F}_{+}(t) = \frac{1}{2^{1/2}}(F_{ij}(t) + F_{ji}(t)) \tag{55}$$

$$\tilde{F}_{-}(t) = \frac{1}{2^{1/2}}(F_{ij}(t) - F_{ji}(t)) \tag{56}$$

which can be immediately integrated to yield

$$\tilde{P}_{\pm}(t) = \lambda_{\pm}^{t}\tilde{P}_{\pm}(0) + \sum_{n=0}^{t-1} \lambda_{\pm}^{t-n-1}\tilde{F}_{\pm}(n) \tag{57}$$

Rotating back to the original basis one finds

$$\begin{aligned} P_{ij}(t) = {} & \frac{1}{2}(\lambda_{+}^{t} + \lambda_{-}^{t})P_{ij}(0) + \frac{1}{2}(\lambda_{+}^{t} - \lambda_{-}^{t})P_{ji}(0) \\ & + \frac{1}{2}\sum_{n=0}^{t-1}(\lambda_{+}^{t-n-1}(F_{ij}(n) + F_{ji}(n)) \\ & + \lambda_{-}^{t-n-1}(F_{ij}(n) - F_{ji}(n))) \end{aligned} \tag{58}$$

There now only remains to do the summations to obtain the final answer

$$\begin{aligned} P_{ij}(t) = {} & \frac{1}{2}(\lambda_{+}^{t} + \lambda_{-}^{t})P_{ij}(0) + \frac{1}{2}(\lambda_{+}^{t} - \lambda_{-}^{t})P_{ji}(0) \\ & + \frac{1}{2}\left[2C_{ij}\left(\frac{1-\lambda_{+}^{t}}{1-\lambda_{+}}\right) + 2E_{ij}\left(\frac{(a-c)^{2t} - \lambda_{+}^{t}}{(a-c)^{2} - \lambda_{+}}\right)\right. \\ & + (D_{ij} + D_{ji})\left(\frac{(a-c)^{t} - \lambda_{+}^{t}}{a-c-\lambda_{+}}\right) \\ & \left. + (D_{ij} - D_{ji})\left(\frac{(a-c)^{t} - \lambda_{-}^{t}}{a-c-\lambda_{-}}\right)\right] \end{aligned} \tag{59}$$

Note how this solution has been created - hierarchically, as in the case of homologous crossover [12]. One can solve first for the order one Building Blocks, which then serve as a "source" for construction of order 2 Building Blocks, which serve as a source for the order 3 etc. until one arrives at the strings themselves. The difference here is that inversion can couple different Building Blocks of the same order, unlike the homologous case where they are decoupled.

In the asymptotic limit $t \rightarrow \infty$, in the case where the cloning or inversion probabilities are less than one, the fixed point of (60) is

$$P_{ij}^{*} = \lim_{t\rightarrow\infty} P_{ij}(t) = \frac{C_{ij}}{(1-\lambda_{+})} \tag{60}$$

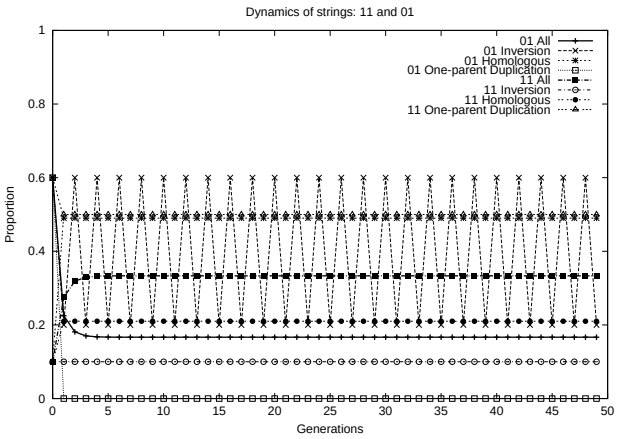

Figure 2: Dynamical evolution of the strings 11 and 01 for different GRDs

Explicitly, in terms of the GRD

$$\begin{aligned} P_{ij}^{*} = {} & \left[\frac{(p_{14} + p_{41} + p_{32} + p_{23} + p_{13} + p_{31} + p_{24} + p_{42})}{(1 - p_{12} - p_{34} - p_{21} - p_{43})}\right. \\ & \times \frac{((p_{*1} + p_{*3})P_{i*}(0) + (p_{2*} + p_{4*})P_{*i}(0))}{((p_{*1} + p_{*3}) + (p_{2*} + p_{4*}))} \\ & \left. \times \frac{((p_{*1} + p_{*3})P_{*j}(0) + (p_{2*} + p_{4*})P_{j*}(0))}{((p_{*1} + p_{*3}) + (p_{2*} + p_{4*}))}\right] \\ & + \left[\frac{(p_{11} + p_{22} + p_{33} + p_{44})}{(1 - p_{12} - p_{34} - p_{21} - p_{43})}\right. \\ & \left. \times \frac{((p_{*1} + p_{*3})P_{i*}(0) + (p_{2*} + p_{4*})P_{*i}(0))}{((p_{*1} + p_{*3}) + (p_{2*} + p_{4*}))}\delta_{ij}\right] \end{aligned} \tag{61}$$

To get some intuition about the nature of this fixed point we will consider some limits of interest associated with different initial populations and different recombination probability distributions. Beginning with a random initial population, where $P_{ij}(0) = 1/4$, the fixed point becomes

$$\begin{aligned} P_{ij}^{*} = {} & \frac{(p_{14} + p_{41} + p_{32} + p_{23} + p_{13} + p_{31} + p_{24} + p_{42})}{4(1 - p_{12} - p_{34} - p_{21} - p_{43})} \\ & + \frac{(p_{11} + p_{22} + p_{33} + p_{44})}{2(1 - p_{12} - p_{34} - p_{21} - p_{43})}\delta_{ij} \end{aligned} \tag{62}$$

Only in the case where there is no one-parent gene duplication, i.e. $p_{ii} = 0$, is the centre of the simplex a fixed point. In the presence of one-parent gene duplication homogeneous strings are favoured over their heterogeneous counterparts. For instance, for GCMs with equal probabilities of $1/16$, the asymptotic proportions of homogeneous and heterogeneous strings are $1/3$ and $1/6$ respectively. Thus, homogeneous strings have higher *effective* fitness [12] than heterogenous ones.

In Figure 2 we see a graph of the solution (60) for the strings 11 and 01 for the same asymmetric initial conditions used for Figure 1, and for the same GCMs. Notice the presence of four different fixed points (two-cycle in the case of inversion) for each string type. This is a much richer behaviour than in the case of simple homologous crossover where the Geiringer limit $P_{ij}^{*} = P_{i*}(0)P_{*j}(0)$

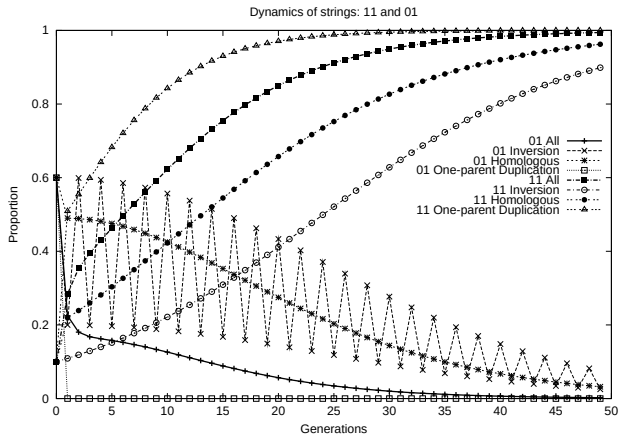

Figure 3: Dynamical evolution of the strings 11 and 01 for different GRDs in a onemax landscape

holds. The Geiringer limits for 11 and 10, with the previously stated initial conditions, are $P^*_{11} = 0.3 \times 0.7 = 0.21$ and $P^*_{01} = 0.7 \times 0.7 = 0.49$, both of which agree with the asymptotic limits seen in Figure 2. For one-parent duplication the expected fixed points for 11 and 10 are from (62) $(0.7 + 0.3)/2$ and 0 respectively, once again in agreement with the graph.

6 Discussion and Conclusions

We have seen the appearance of novel phenomena and have been able to describe them exactly in the present two-locus context in the absence of selection and for an infinite population. One is prompted to wonder what happens for $\ell > 2$ and when selection is present.

For the case of selection we illustrate the effects in Figure 3 for a non-epistatic landscape where $f_{11} = 12$, $f_{10} = f_{01} = 11$ and $f_{00} = 10$, where the two-locus equations with selection have been iterated. What is clearly seen is a similar bias as found from the no selection case, but superimposed on a selection dominated "trend". The oscillations for inversion only are clearly visible. However, here, as only heterogeneous strings can oscillate, and as the optimal string is homogeneous, the oscillations diminish in amplitude. For one-parent duplication, the proportion of 01 strings vanishes after one generation, as in the no selection case. For 01 with all GCMs present, as in the no selection case, there is a sharp initial decrease. In distinction to that case though, in the presence of selection, the proportion continues to diminish, but at a much reduced rate. As the selection pressure diminishes, the curves of Figure 3 will imitate those of Figure 2 ever more closely, while for increasing selection pressure the phenomena due to inversion and duplication will be less and less noticeable. We see then that, although we have only exactly solved the no selection case, observed phenomena such as lateral allele diffusion, oscillations, preference for homogenous strings etc. are also present in the presence of selection.

For $\ell > 2$ exact solutions are naturally more complicated. However, all the observed phenomena - oscillations, homogeneous/heterogeneous asymmetry, lateral allele diffusion - all appear for $\ell > 2$ - as has been explicitly checked by iterating the microscopic equations (1) for $\ell = 3,\ 4$. In this case though the various phenomena can occur simultaneously, for instance for $\ell = 4$ one might have inversion restricted to the first two loci leading to oscillations there while restricting duplication to the last two loci and having a preference for homogeneous alleles there.

We have shown that coarse grained formulations, which are so powerful in the context of homologous crossover, can be generalised to take into account other genetic operators, such as inversion and duplication, thereby showing that Building Block schemata are also the appropriate effective degrees of freedom for generalised recombination. We believe that the present formalism allows for a much deeper theoretical analysis of many different types of recombination above and beyond the well studied homologous type.

Acknowledgements

CRS and RP thank the ESPRC for financial support (grant number GR/T24616/01). CRS also thanks DGAPA of the UNAM for a Sabbatical Fellowship and Conacyt project 30422-E.

Bibliography

[1] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester, UK, 2000.

[2] C. Chryssomalakos and C. R. Stephens. What basis for genetic dynamics? In K. Deb, editor, *Proceedings of GECCO 2004*, pages 1018–1029, Berlin, Germany, 2004. Springer Verlag.

[3] A. Clark. Invasion and maintenance of a gene duplication. *Proc. Nat. Acad. Sci.*, 91:2950–2954, 1994.

[4] H. Geiringer. On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics*, 15(1):25–57, March 1944.

[5] D. E. Goldberg. Simple genetic algorithms and the minimal deceptive problem. In L. Davis, editor, *Genetic Algorithms and Simulated Annealing*, pages 74–88, London, UK, 1987. Pitman.

[6] J. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, USA, 1975.

[7] J. R. Koza. Gene duplication to enable genetic programming to concurrently evolve both the architecture and work-performing steps of a computer program. In *IJCAI-95 Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, volume 1, pages 734–740, Montreal, Quebec, Canada, 20-25 Aug. 1995. Morgan Kaufmann.

[8] R. Poli. Exact schema theory for genetic programming and variable-length genetic algorithms with one-point crossover. *Genetic Programming and Evolvable Machines*, 2(2):123–163, 2001.

[9] R. Poli and C. R. Stephens. Theoretical analysis of generalised recombination. In *CEC-2005*, 2005. Accepted.

[10] H. Sawai and S. Adachi. A comparative study of gene-duplicated GAs based on pfGA and SSGA. In D. Whitley, D. Goldberg, E. Cantu-Paz, L. Spector, I. Parmee, and H.-G. Beyer, editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000)*, pages 74–81, Las Vegas, Nevada, USA, 10-12 July 2000. Morgan Kaufmann.

[11] W. M. Spears. The equilibrium and transient behavior of mutation and recombination. In W. Spears and W. Martin, editors, *FOGA 6*, pages 74–88, San Francisco, USA, 2001. Morgan Kaufmann.

[12] C. R. Stephens and H. Waelbroeck. Schemata evolution and building blocks. *Evolutionary Computation*, 7(2):109–124, 1999.

[13] M. D. Vose. *The simple genetic algorithm: Foundations and theory*. MIT Press, Cambridge, MA, 1999.