

Assortative mating drastically alters the magnitude of error thresholds

Gabriela Ochoa¹ and Klaus Jaffe²

¹ INRIA – COMPLEX Team,
Domaine de Voluceau BP 105 78153 Le Chesnay Cedex – France
Gabriela.Ochoa.@inria.fr

² Laboratorio de Comportamiento,
Universidad Simón Bolívar, Caracas 89000, Venezuela
kjaffe.@usb.ve

Abstract. The *error threshold* of replication is an important notion of the *quasispecies* evolution model; it is a critical mutation rate (error rate) beyond which structures obtained by an evolutionary process are destroyed more frequently than selection can reproduce them. With mutation rates above this critical value, an error catastrophe occurs and the genomic information is irretrievably lost. Recombination has been found to reduce the magnitude of the error threshold in evolving viral quasispecies. Here, through a simulation model based on genetic algorithms, we incorporate assortative mating and explore its effect on the magnitude of error thresholds. We found, consistently on four fitness landscapes, and across a range of evolutionary parameter values, that assortative mating overcomes the shift toward lower error threshold magnitudes induced by recombination, on the other hand, disassortative mating drastically reduces the error threshold magnitude. These results have implications to both natural and artificial evolution: First, they support the hypothesis that assortative mating by itself may overcome some of the evolutionary disadvantages of sex in nature. Second, they suggest a critical interaction between mutation rates and mating strategies in evolutionary algorithms.

1 Introduction

Quasispecies theory was derived by Eigen and Schuster [3, 5], to describe the dynamics of molecular evolution under the influence of mutation and selection. The theory was originally developed for studying pre-biotic evolution, but in a wider sense it describes any population of reproducing organisms. The *error threshold* of replication is an important notion of the quasispecies model; it is a critical mutation rate (error rate) beyond which structures obtained by an evolutionary process are destroyed more frequently than selection can reproduce them. With mutation rates above this critical value, an error catastrophe occurs and the genomic information is irretrievably lost [14, 21]. Therefore, studying the factors that alter this threshold has important implications in the study of evolution. The quasispecies model, as stated originally, considered infinite asexual

populations (i.e no recombination) on a single peak (needle in a haystack) landscape. Later on, Nowak and Schuster [14] extended the calculations of the error threshold on this landscape to *finite* populations. Finite populations lose grip on the solitary spike of superior fitness more easily because of the added hazard of natural fluctuations. Another extension by Boerlijst et al. [1] included recombination; the authors studied two abstract simple landscapes: the isolated peak and a *plateau* landscape (see section 2) and found that recombination shifted error thresholds toward lower values, and made the transition sharper (for an explanation to this phenomenon, see section 3.1). Thus, recombination is advantageous only if the landscape is correlated and if the mutation rate is sufficiently small. Results obtained using infinite population models cannot be expected to automatically apply to the more realistic case of finite populations. To investigate this latter case, Ochoa and Harvey [16], reproduced the experiments in [1] but used a genetic algorithm (GA) [7] – and hence finite populations – instead of the quasispecies model as the underlying model of evolution. The empirical results mirrored qualitatively those produced by Boerlijst et al. for infinite populations. Notably, error thresholds for finite populations were, in all scenarios, lower than for the infinite case. The work by Wright et al. [25, 26] used the Vose [24] dynamical system model to show that a simple GA can exhibit bistability on a single peak and double peak fitness landscapes. They also found that recombination creates catastrophic error threshold transitions as mutation was increased, and confirmed that recombination decreased the mutation rate at which the error threshold occurred.

The work of Bonhoeffer and Stadler [2] described an empirical approach for locating thresholds on complex landscapes (see section 2.2). In [15] this method is borrowed and adapted. Instead of the quasispecies model a GA is used as the underlying model of evolution, thus a method for estimating error thresholds in GAs is devised. In addition, *consensus sequence plots* (see section 2.2) are suggested as tools for visualising the structure of fitness landscapes. A later contribution [19] confirmed the existence of the error threshold feature on a wider selection of complex landscapes including real-world domains, the study also considered the effect of modifying the most prominent evolutionary parameters on the magnitude of error thresholds.

The recombining model by Boerlijst et al. considered random mating. In nature, however, mating is rarely random, and mate selection may be as important in guiding evolution as natural selection [10]. Theoretical studies of mate selection using agent-based simulations [20, 9] suggest that some mating strategies confer higher fitness on individuals, and produce higher evolutionary stability than random mating. Assortative mating is a form of non-random mating common in nature, where individuals of similar phenotype mate more (or less) often than expected by chance. It is positive if similar organisms mate more often, and negative (or dissortative) if dissimilar organisms mate more often.

In this paper we incorporate non-random mating through a GA-based simulation model, and study the effects of assortative mating on the magnitude of the error threshold. We argue that this study is relevant to both natural and

artificial evolution. In evolutionary biology models, the mating strategy modelled has been shown to be fundamental in determining whether sexual reproduction emerges and is maintained in the simulated populations [8]. In evolutionary computation, the notion of error threshold has been related to the idea of having an “optimal” mutation rate, since this threshold is intuitively related to the idea of an optimal balance between *exploitation* and *exploration* [17, 19]

2 Methods

We studied four fitness landscapes. First, the two simple landscapes proposed by Boerlijst et al. [1]: a single peak and a plateau landscape. The single peak landscape has much neutrality (almost all flat), but a correlation measure of the whole landscape indicates it is highly uncorrelated; the plateau is still simple, with much neutrality but slightly more correlated. Second, two families of more correlated and complex landscapes: Royal Staircase functions [23], and NK landscapes [11]. A description of these landscapes, along with the particular instances selected, is given below:

Single peak landscape: In this scenario, only one sequence F has an increased fitness. This single bit string has fitness $R_F = 5$, whereas all other sequences have fitness $R_i = 3.5$.

Plateau landscape: Here, the single peak landscape is modified so that the fitness of sequences close by the fittest string F is increased to $R_{H_1} = 4.8$, and $R_{H_2} = 4.6$, where H_1 is the set of all sequences with a Hamming distance of 1 from the fittest string F , and H_2 the set of all sequences with a Hamming distance of 2 from F .

Royal Staircase: The Royal Staircase family of functions was proposed for analysing epochal evolutionary search, it is a simple class of functions that allows neutrality to be modelled and tuned. Genotypes are specified by binary strings of length $L = MB$, where M is the number of blocks and B the number of bits per block. Any completely set block (with all bits set to 1) contributes a fitness component, but blocks that are only partially set (with one or more bits at zero) contribute no fitness. Royal Staircase functions are always unimodal, but we can increase the landscape neutrality by enlarging the size of the blocks. Modifying the number of blocks also alters the overall shape and ruggedness of the landscape. The selected instance ($M = 3$, $B = 14$) corresponds to a rugged, neutral, unimodal landscape.

NK landscape: The NK family of landscapes is a problem-independent model for constructing multimodal landscapes that can gradually be tuned from smooth to rugged. In the model, N refers to the number of genes in the genotype (i.e. the string length) and K to the number of genes that influence a particular gene. By increasing the value of K from 0 to $N - 1$, NK landscapes can be tuned from smooth to rugged. The selected NK landscape instance ($N = 24$, $K = 10$) corresponds to a multimodal rugged landscape.

The default experiment setting used a generational GA with fitness proportional selection and a population of size 100. The genetic operations were uniform crossover [22] and the standard bit mutation. The GA was run in four modes: (a) using mutation only (*Asexual*), (b) using both mutation and recombination with random mating (*Random Mating*), (c) implementing assortative mating (*Assortative Mating*), and (d) implementing dissortative mating (*Dissortative Mating*). Assortative mating was implemented as follows: when selecting two individuals for a crossover, the first parent was selected as usual (based on fitness). For choosing the second parent, two potential partners were selected using the GA fitness-based selection method. Thereafter, the similarity between the two potential partners and the first parent was computed. For negative assortative (dissortative) mating, the genotype with less similarity was chosen. For positive assortative mating, the genotype closer to the first parent was selected as the second parent. We used Hamming distance as the similarity measure. Although in nature assortative mating is based on phenotypes, here we select a similarity measure based on genotypes given the simplicity of the model and landscapes as compared to nature. Furthermore, the phenotypic traits of organisms in nature are an expression of their genotypes.

2.1 Estimating error thresholds on simple landscapes

On the single peak and plateau landscape, we studied the steady state structure of the population, using the GA model described above, for a range of mutation rates. The structure of the population is characterised by the proportion of each error (or Hamming distance) class. We used the evolutionary parameters explored in [1]: string length of 15 and recombination rate of $r = 0.5$. In both landscapes there is a single optimum, F , we set it as the string of all 0s (000000000000000) with no loss of generality. Any other bit string is referred to as a ‘mutant’, and belongs to one of the Hamming distance classes H_i , where i is the Hamming distance to F . In the simulations, the initial population was generated differently for each landscape. For the single peak landscape, around 50% of the population was set on the peak and the rest was randomly generated. For the plateau landscape, 25% was set on the peak, 25% in the H_1 compartment, 25% in the H_2 compartment, and the rest was randomly generated. The per bit mutation rate p was varied from $p = 0.000$ up to $p = 0.05$, with a step size of 0.001. The number of generations per GA run was 1000. This value was empirically selected; the distribution of sequences was fairly stable by this point in all cases. Each experiment was run 50 times and the results were averaged.

2.2 Estimating error thresholds on complex landscapes

Bonhoeffer and Stadler (1993) studied the evolution of quasispecies on two correlated fitness landscapes (the Sherrington Kirkpatrick spin glass and the GraphBipartitioning landscape), and described an empirical approach for locating thresholds on complex landscapes. The approach is to calculate and plot the consensus sequence at equilibrium for a range of mutation rates. The consensus sequence

in a population is defined as the sequence of predominant symbols (bits) in each position; it is plotted as follows: if the majority of individuals has a ‘1’ or ‘0’ in a position i the field is plotted white or black, respectively. The field is plotted Gray if the position is undecided. The *equilibrium state* is reached when the proportion of different sequences in the population is stationary. This happens when evolution is simulated for a large enough number of generations. In practise, it is considered that the equilibrium is reached when several parameters of the population (e.g. the maximal and average fitness) reach equilibrium. According to Bonhoeffer and Stadler (1993) the error threshold may be approached from *below* or *above* with both methods producing similar results. For approaching the error threshold from above, the simulation starts with a random population. Then the population is allowed to reach equilibrium at a constant predefined maximum for the mutation rate. Afterwards, the mutation rate is decreased by a fixed small step and the computation continues with the current population. This process is repeated until the mutation rate is 0.0. Therefore, the consensus sequence in the population is calculated and plotted at the end of each simulation cycle for each mutation step. The error threshold is characterised by the loss of the consensus sequence, i.e. the genetic information of the population. Beyond the error threshold the consensus sequence is no longer constant in time (see Figure 2).

Previous studies [18], confirmed that: (i) error thresholds approached from below and above produce similar results, (ii) the error threshold magnitude is independent of the particular initial population; and (iii) the error threshold is similar for different instances of an NK landscape with fixed N and K . Hence, the approach followed here is to approach error thresholds from above using a fixed random seed for generating the initial population in all cases; and for the NK landscape, selecting a single landscape instance. For the experiments reported here, the recombination rate was set to 1.0 when recombination is used. Mutation rates per locus ranged from 0.05 to 0.0 with a step of 0.001. For each mutation rate the simulation lasted 10,000 generations, this number was found empirically to equilibrate the population maximal and average fitness.

2.3 Varying Evolutionary Parameters

In order to explore the robustness of the results, the most relevant evolutionary parameters were varied from the default setting described above, on two selected landscapes: the single peak and NK landscapes. In particular, we explored the effect of modifying the population size, chromosome length, and, on the single peak landscape, the relative fitness (or fitness difference) between the peak and the rest.

3 Results

Figure 1 show the steady state distribution of sequences on the plateau landscapes, for a range of mutation rates, and the four reproductive strategies. Similar plots (not reported here) were also produced for the single peak landscape.

Sequences are classified in error classes: all sequences of i errors from the master are members of class i . The master sequence F (thick line) and error classes H_1 and H_2 are highlighted in the plots. The error threshold can be identified visually as the mutation rate just before the error classes become distributed as for a completely random population (i.e. the lines become parallel). Similar plots are commonly used to visualise error thresholds in quasispecies studies (see for instance [13], pp. 48). Assortative mating, on both the single peak and plateau landscapes, increases considerably the error threshold as compared to both random mating and no recombination. Moreover, assortative mating is advantageous for the population, because it increases the abundance of F , and makes the population more stable as the error threshold moves to higher values. Notice that on the plateau landscape (Figure 1) the proportion of the master sequence F for assortative mating is greater than twice the corresponding proportion for random mating, and about three times that proportion for the asexual population.

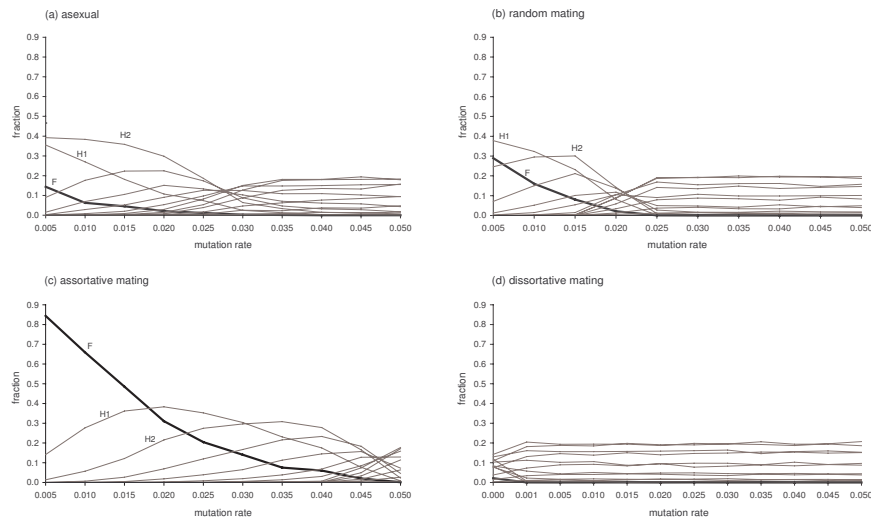


Fig. 1. Distribution of sequences for a range of mutation rates on the plateau landscape and the four reproductive strategies. Sequences are classified in error classes: all sequences of i errors from the master are members of class i . The master sequence F (thick line), error classes H_1 and H_2 are highlighted in the plots, and the other light lines correspond to the remaining error classes. The Y axis shows population fractions, and the X axis shows mutation rates (per bit). Error thresholds can be identified visually as the mutation rate just before the error classes become distributed as for a random population (the lines become flat).

Figure 2 shows the consensus sequence plots on the NK landscape. Similar plots were produced for the Royal Staircase Landscape. The plots show a clear error threshold; there is a distinguishable transition between an “ordered” (selection-dominated) regime, and a “disordered” (mutation-dominated) one. There is a stable consensus sequence for mutation rates below the error thresh-

old. On the NK landscape (Figure 2), the consensus sequence in each case is different and corresponds to one of the many NK landscape’s local optima; whereas on the Royal Staircase, the consensus sequence is always the single optimum in the landscape (the string of all 1s). Clearly assortative mating produces the highest error threshold, whereas asexual reproduction produces the lowest.

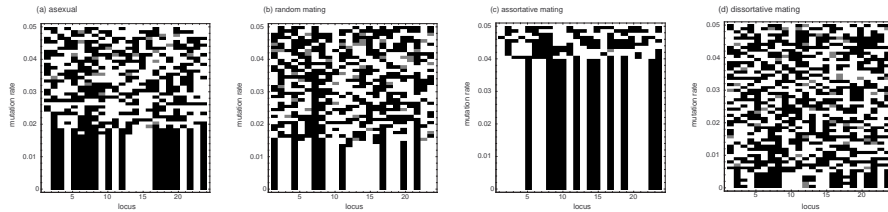


Fig. 2. Consensus sequence plots on the NK landscape and the four reproductive strategies. The horizontal axis shows the consensus bit for each locus, the vertical axis shows per bit mutation rates. The error threshold is characterised by the loss of the consensus sequence (one local optima for the NK Landscape).

	<i>Single Peak</i>	<i>Plateau</i>	<i>Royal Staircase</i>	<i>NK</i>
Asexual	0.017	0.030	0.018	0.018
Random Mating	0.010	0.018	0.004	0.015
Assortative Mating	0.032	0.045	0.024	0.040

Table 1. Approximate error thresholds on the four landscapes studied.

Table 1 summarises the approximate error thresholds values on the four landscapes studied, and the default evolutionary parameters. Results for dissortative mating were not included as they were generally null or close to zero. Additionally, tables 2 and 3, report the approximate error thresholds after altering the population size and chromosome length, respectively, on the single peak and NK landscapes. Results suggest that assortative mating increases considerably the error threshold as compared to random mating, these findings were found to be robust across a range of evolutionary parameter values. Error thresholds with assortative mating are even higher than for the populations without recombination. Figure 3, shows approximate error thresholds on the single peak landscape for a range of fitness difference values (between the peak and the rest), again assortative mating increases error threshold values as compared to random mating, the increase surpasses the error threshold of the asexual population for small and moderate fitness differences. For large fitness differences the asexual population has the higher error threshold. This is consistent with other authors observation that recombination may be advantageous for evolving populations, the critical requirement being that the locations of local optimal carry mutual information about where other good optima are located [1, 11]. Finally, assorta-

tive mating was implemented selecting a mate from a pool of only two potential candidates, if the size of this pool is increased, empirical results confirmed that the effects on the error threshold are augmented (i.e. the error threshold is even higher).

	<i>Single Peak</i>				<i>NK</i>			
	50	100	200	500	50	100	200	500
Asexual	0.013	0.017	0.019	0.021	0.013	0.018	0.020	0.028
Random Mating	0.009	0.010	0.011	0.011	0.012	0.015	0.015	0.016
Assortative Mating	0.028	0.032	0.034	0.035	0.031	0.040	0.052	0.059

Table 2. Approximate error thresholds on the Single Peak and *NK* landscapes for a range of population sizes.

	<i>Single Peak</i>				<i>NK</i>			
	10	15	30	60	16	20	24	28
Asexual	0.031	0.017	0.015	0.008	0.028	0.024	0.018	0.012
Random Mating	0.017	0.010	0.004	0.002	0.023	0.018	0.015	0.011
Assortative Mating	0.047	0.032	0.018	0.014	0.063	0.049	0.040	0.030

Table 3. Approximate error thresholds on the Single Peak and *NK* landscapes for a range of chromosome lengths.

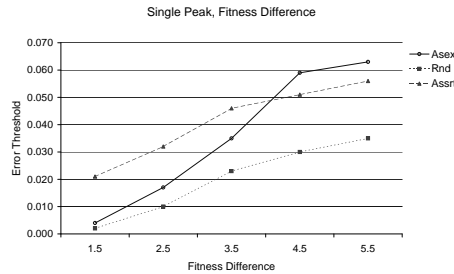


Fig. 3. Approximate error thresholds on the Single Peak landscape for several values of the fitness difference between the peak and the rest. The Y axis shows error thresholds, and the X axis shows fitness difference values.

3.1 Discussion

We explored the effect of including assortative mating on the magnitude of error thresholds on four different landscapes. Additionally, the robustness of the results was tested on a range of values of the most significant evolutionary parameters.

Remarkably in all scenarios, recombination shifted the error threshold to lower mutation rates and it made the transition sharper. Near the error threshold, without recombination, the fittest sequence only makes up a small percentage of the total population [4]. Under such conditions, recombination acts as a diverging operator, driving the population beyond the error threshold. In this scenario assortative mating may be helpful for the evolving population by counteracting this diverging effect. Experiments including mate selection showed that assortative mating considerably increases the error threshold value, even beyond the corresponding value for asexual reproduction on correlated landscapes. Moreover, assortative mating increases the abundance of the master sequence, and makes the population more stable in the presence of higher mutation rates.

As Kauffman [11] suggests, recombination appears to be a powerful strategy on a wide range of rugged fitness landscapes. The critical requirement appears to be that local optima carry mutual information about the location of other good or better optima. Thus, although our results suggest that assortative mating increases the value of sex as an evolutionary strategy, sex even with assortative mating may be sub-optimal under extreme conditions. Caution must be also taken when setting evolutionary parameters, as there seems to be a strong interaction between mating strategies and mutation rate.

Regarding the implications to natural evolution, our results suggest that recombination with random mating reduces the population stability and abundance of the fittest individuals; and thus may hinder the average fitness of the whole population. However, assortative mating eliminates this negative effect of sex on evolutionary stability and is even capable of improving it over the asexual dynamics on correlated landscapes. This supports the conclusion by Jaffe [9] that assortative mating by itself may overcome some of the evolutionary disadvantages of sex, thus casting a new light on the dilemma posed by Fisher [6] of why sex exists.

References

1. M. C. Boerlijst, S. Bonhoeffer, and M. A. Nowak, *Viral quasi-species and recombination*, Proc. Royal Soc. London B **263** (1996), 1577–1584.
2. S. Bonhoeffer and P. Stadler, *Error thresholds on correlated fitness landscapes*, J. Theor. Biol. **164** (1993), 359–372.
3. M. Eigen, *Self-organization of matter and the evolution of biological macromolecules*, Naturwissenschaften **58** (1971), 465–523.
4. M. Eigen, J. McCaskill, and P. Schuster, *Molecular quasi-species*, J. Phys. Chem. **92** (1988), 6881–6891.
5. M. Eigen and P. Schuster, *The hypercycle: A principle of natural self-organization*, Springer-Verlag, 1979.
6. R. A. Fisher, *The genetic theory of natural selection*, Clarendon Press, Oxford, 1930.
7. J. H. Holland, *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, 1975.
8. Klaus Jaffe, *On the adaptive value of some mate selection strategies*, Acta Biotheoretica **1** (1999), no. 47, 29–40.

9. ———, *Emergence and maintenance of sex among diploid organisms aided by assortative mating*, *Acta Biotheoretica* **2** (2000), no. 48, 137–147.
10. ———, *On sex, mate selection and evolution: an exploration*, *Comments on Theoretical Biology* **2** (2002), no. 7, 91–107.
11. Stuart A. Kauffman, *The origins of order*, Oxford University Press, Oxford, 1993.
12. J. Maynard Smith, *The evolution of sex*, Cambridge University Press, 1978.
13. J. Maynard Smith and E. Szathmary, *The major transitions in evolution*, Oxford University Press, 1995.
14. M. Nowak and P. Schuster, *Error thresholds of replication in finite populations: Mutation frequencies and the onset of Muller's ratchet*, *J. Theor. Biol.* **137** (1989), 375–395.
15. G. Ochoa, *Consensus sequence plots and error thresholds: Tools for visualising the structure of fitness landscapes*, *Parallel Problem Solving from Nature - PPSN VI 6th International Conference*, *Lecture Notes in Computer Science*, vol. 1917, Springer Verlag, 2000, pp. 129–138.
16. G. Ochoa and I. Harvey, *Recombination and error thresholds in finite populations*, *Foundations of Genetic Algorithms 5* (W. Banzhaf and C. Reeves, eds.), Morgan Kaufman, 1998.
17. G. Ochoa, I. Harvey, and H. Buxton, *Error thresholds and their relation to optimal mutation rates*, *Proceedings of the 5th European Conference on Advances in Artificial Life (ECAL-99) (Berlin)* (Dario Floreano, Jean-Daniel Nicoud, and Francesco Mondada, eds.), *LNAI*, vol. 1674, Springer, September 13–17 1999, pp. 54–63.
18. Gabriela Ochoa, *Error thresholds and optimal mutation rates in genetic algorithms*, Ph.D. thesis, School of Cognitive and Computing Sciences, University of Sussex, UK, 2001.
19. ———, *Error thresholds in genetic algorithms*, *Evolutionary Computation Journal* **2** (2006), no. 14.
20. Gabriela Ochoa and Klaus Jaffe, *On sex, mate selection and the red queen*, *J. Theor. Biol.* **1** (1999), no. 199, 1–9.
21. J. Swetina and P. Schuster, *Self-replication with errors, a model for polynucleotide replication*, *Phys. Rev. A [15] Stat. Phys.* **45** (1992), 6038–6050.
22. Gilbert Syswerda, *Uniform crossover in genetic algorithms*, *Proc. 3rd International Conference on Genetic Algorithms (George Mason University)* (J. David Schaffer, ed.), Morgan Kaufmann, 1989, pp. 2–9.
23. E. van Nimwegen and J. P. Crutchfield, *Optimizing epochal evolutionary search: Population-size dependent theory*, *Tech. Report Preprint 98-06-046*, Santa Fe Institute, 1998.
24. Michael D. Vose, *The simple genetic algorithm: foundations and theory*, MIT Press, Cambridge, MA, 1999.
25. Alden H. Wright, Jonathan E. Rowe, and James R. Neil, *Analysis of the simple genetic algorithm on the single-peak and double-peak landscapes*, *Proceedings of the 2002 Congress on Evolutionary Computation CEC2002*, IEEE Press, 2002, pp. 214–219.
26. Alden H. Wright, Jonathan E. Rowe, Christopher R. Stephens, and Riccardo Poli, *Bistability in a gene pool GA with mutation*, *Foundations of Genetic Algorithms 7* (Kenneth A. De Jong, Riccardo Poli, and Jonathan E. Rowe, eds.), Morgan Kaufmann, San Francisco, 2003, pp. 63–80.