Diversity Maintenance on Neutral Landscapes:
An Argument for Recombination

Nicholas Geard[1] (nic@itee.uq.edu.au)
Janet Wiles[1, 2] (j.wiles@itee.uq.edu.au)

[1]School of Information Technology and Electrical Engineering
[2]School of Psychology
The University of Queensland, Brisbane Q 4072, Australia

Abstract – It has been demonstrated that several standard evolutionary computation test problems can be solved by a simple hill climbing search algorithm – often more efficiently than by a population based evolutionary algorithm.

There remain some classes of problems, however, for which maintaining a genetically diverse population is essential in order to discover the optimal solution. In biological populations, diversity maintenance is important to enable populations to adapt to rapidly changing environments and to exploit environmental niches.

We demonstrate that on a neutral landscape recombination allows a population to maintain a significantly greater level of genetic diversity through the transition between two fitness layers. Recombination may therefore have a role to play in maintaining population diversity across fitness transitions.

I. INTRODUCTION

The metaphor of a rugged and hilly fitness landscape over which an evolving population moves has become such a dominant image in evolutionary computation that it is easy to forget that there is more to evolutionary computation than just climbing hills. The preoccupation with landscape traversal is possibly due to the large amount of research into evolutionary computation that is focused on function optimisation, in which fitness landscapes are generally static searchable domains. In other situations however, issues such as maintaining genetic diversity in the population can also be crucially important. The importance of diversity is especially significant in biological models, such as those of coevolving populations, where the fitness landscape may be highly dynamic. In these cases, a population that has converged on a fitness peak at one stage may subsequently find itself deep in a valley.

If function optimisation is the goal of an evolutionary algorithm it has been demonstrated that, in many cases, a simple hill climbing algorithm will outperform a population based evolutionary algorithm – frequently with fewer computational overheads. For example, some of the Royal Road problems, although explicitly designed to demonstrate the strengths of genetic recombination, were solved more quickly by a 'random-mutation' hill climber [3]. There remain some classes of problem, however, for which the existence of a population is desirable, if not essential. Examples include Watson & Pollock's HIFF function [13], Holland's hyperplane-defined functions [5] and Pelikan and Goldberg's hierarchical trap functions [10]. These are all problems that require a diverse population to be maintained

in order for the optimal solution to be discovered. In biology, diversity is important for several reasons; it allows a population to adapt more rapidly to changing environments (e.g., in a host-parasite situation [9]) and it allows the exploitation of multiple environmental niches (e.g., in the immune system [11]).

Within the domain of population based evolutionary algorithms, a further issue is the relative importance of the mutation and recombination operators. The initial interpretation, often referred to as the Building Block Hypothesis, was that recombination was the primary operator due to its role in constructing advantageous combinations of genes [3, 4]. Mutation was considered a 'background' operator, whose role was to introduce new genetic material and prevent the total convergence of the population [4]. An alternative view is that mutation is the dominant operator and recombination is relegated to the role of an error correction mechanism or macro-mutation operator [2].

One way of investigating the role these operators may play is to observe their effect on diversity maintenance in a population undergoing the transition between two fitness layers on a neutral landscape. In this study, neutrality per se is not central to the phenomena of interest, but it does enable the change in diversity to be demonstrated very clearly. The critical feature is the 'punctuated equilibria' dynamic displayed by populations evolving on neutral landscapes [1, 6, 8].

II. METHODOLOGY AND RESULTS

The Royal Staircase model [12], a variation of the well-known Royal Road functions [7], is a simple landscape with clearly defined neutral layers. Genotypes are specified by binary strings consisting of N blocks of length K. Each genotype has a base fitness of 1, to which is added 1 fitness point for each consecutive block (counting from the left) formed completely of ones. The fitness is thus an integer in the range $[1, N + 1]$.

For the simulations described here, a Royal Staircase landscape with $N = 4$ and $K = 8$ was used (i.e. 4 blocks of 8 bits each = genotype of length 32 bits). In order to provide the clearest demonstration of the phenomena, the methodology of Newman and Engelhardt [8] was followed and a population of size 100 was used, initially converged upon a random genotype. Trial simulations with a non-converged initial population were also run and similar

phenomena were observed. The reason for this similarity is that, after a block has been discovered and the population fitness has undergone a significant increase, the population tends to converge upon the individual who discovered the new block. A mutation rate of 0.001 and fitness proportional selection was used. In one condition, crossover was not used; in the other condition, single point crossover with a crossover rate of 0.7 was used.

"Diversity" in this study was defined to be the number of unique genotypes in a population. Other more complex measures of diversity are possible (e.g. entropy as defined in [1] or [8]), however, we were interested in explicit diversity rather than variance within a population. At each generation, the population diversity at each fitness level and the total population diversity were recorded. The difference between total population diversity before (g_{old}) and after (g_{new}) a fitness transition was used to measure the proportion of diversity lost as the population made a transition to the new fitness level. The loss in diversity was given by ($g_{old} - g_{new})/g_{old}$. For example, if there were half as many distinct genotypes after the transition, the loss in diversity would be 0.5.

For simplicity, only the first fitness transition was taken into account (i.e., the rise in fitness when the first block was discovered). The start of a fitness transition was taken to be the generation in which the increase in average population fitness exceeded 0.01 and the end of a fitness transition was the generation in which the increase in average population fitness dropped below 0.01. By using a genotype with multiple blocks, a reasonable amount of diversity in non-converged blocks of the genotype could be maintained at each fitness level.

Two sets of simulations were performed, one in which the population evolved using mutation only, and another in which the population evolved using recombination as well as mutation. Each set consisted of 20 independent runs and the results from each of these were averaged to obtain the mean diversity loss (Table I).

TABLE I

DROP IN DIVERSITY ACROSS A FITNESS TRANSITION

	Average (Std. Dev.) over 20 runs
Mutation only	0.48 (0.17)
Crossover and mutation	0.22 (0.12)

To investigate the dynamics of a population undergoing a transition between two fitness layers more closely, the diversity of the sub-populations on the lower fitness level ('the old population') and the higher fitness level ('the new population') were plotted (Figure 1 shows the population diversity over a typical simulation run).

III. ANALYSIS

The simulations illustrate some of the differences between the dynamics of populations evolving using only mutation and those evolving using mutation and recombination (Figure 1). The rise in diversity during the first 20 generations is similar in both cases. As the initial population was converged, recombination will not have a significant effect until mutation has introduced a certain level of diversity.

After the first 20 generations, the differences between the two runs are striking. Whereas the diversity of a population using only mutation stabilised at around 20 unique genotypes, the diversity of a population using recombination and mutation continued to rise until the majority of individuals had unique genotypes.
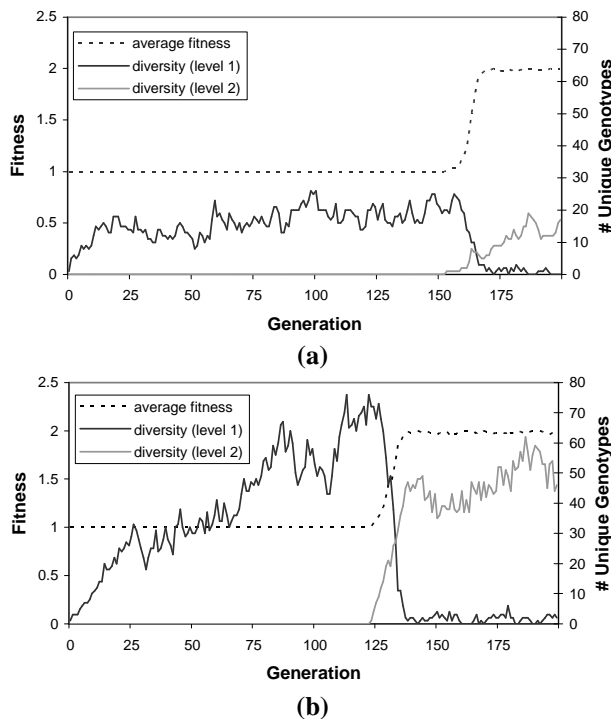

(a)

(b)

Figure 1: Diversity changes across fitness level transitions. An initially converged population of 100 evolving on a Royal Staircase landscape (N = 4, K = 8) with a mutation rate of 0.001 and a crossover rate of (a) 0.0 and (b) 0.7. Initial growth in diversity is similar; however with mutation only, a ceiling on diversity is soon reached. With crossover, a much higher level of diversity is possible. In both cases, diversity is lost during the transition from a lower fitness level to a higher fitness level. However, this diversity is recovered much more rapidly by the population using crossover.

The next notable difference occurred at the transition point between two fitness levels. The population using recombination was able to maintain a much greater level of diversity across the fitness transition (Table 1). Whereas the population using only mutation lost almost half of its diversity on average, the population using mutation and recombination lost around 20% of its diversity on average. Furthermore, it recovered this lost diversity more quickly.

IV. DISCUSSION

The observed behaviour is consistent with expectations. One possible reason for the higher level of diversity in the population using recombination and crossover is that random mutation alone is likely to create new genotypes whose Hamming distance from the original genotype is very low (on average they will differ at only one bit position). There is a distinct possibility of 'back mutations' in future generations restoring previously mutated genotypes to the dominant population genotype. Recombination allows the creation of new genotypes whose Hamming distance from the population average is much greater, and hence whose probability of undergoing back mutation is much lower. On a neutral layer, convergence due to drift will also limit the diversity in both conditions.

At the transition point, a population using only mutation loses all the diversity in the old population as these individuals are rapidly selected against. Typically only one individual in the population ('the founder') discovers the new fitness level. Every member of the new population will be descended from the founder and the population must 'rediscover' the lost diversity via mutation. The growth of diversity in the new population is likely to be somewhat slower than the initial growth of diversity in the old population, as a certain proportion of mutations will either occur to individuals in the old population, or will act to lower the fitness of an individual in the new population (i.e. by destroying the assembled block). The maximum level of diversity possible in the new population is also likely to be lower, as the first block will remain converged in the population.

The population using recombination, however, is able to 'transfer' much of the diversity from the old population to the new population. So long as recombination does not interfere with a newly assembled block, a cross between an individual from the new population and an individual from the old population will result in the offspring in the new population carrying genetic material *not* inherited from the founder. The growth in diversity at this point is therefore much more rapid than the initial growth in diversity. Once again, the maximum possible level of diversity will be lower due to the convergence of the first block in the new population.

IV. CONCLUSION

While recombination, and indeed populations, may only be of benefit to optimisation problems in certain cases, they should be more useful in problem domains where diversity maintenance is necessary. In this study, we found that a population using recombination as well as mutation was able to maintain a consistently higher level of genetic diversity across an evolutionary run on a neutral landscape than a population using only mutation.

These results raise several issues in relation to the role of recombination in biological populations. It may plausibly be argued that a population in which every fitness increase was accompanied by dramatic genetic convergence would be far more susceptible to invasion by parasites that evolved to exploit the most common genotype in the population. By allowing diversity from a lower fitness level to be transferred to a higher fitness level, recombination enables populations to guard more carefully against exploitation by a parasite species.

REFERENCES

[1] Barnett, L., *Tangled webs: Evolutionary dynamics on fitness landscapes with neutrality*, MSc diss., School of Cognitive and Computing Sciences, University of Sussex, 1997.

[2] Barnett, L., Netcrawling – Optimal Evolutionary Search with Neutral Networks, In *Proceedings of the 2001 Congress of Evolutionary Computation (CEC2001), May 27-31, Seoul, Korea*, pp. 30-37, Piscataway, NJ: IEEE Press, 2001.

[3] Forrest, S. & Mitchell, M., Relative building-block fitness and the building block hypothesis, In D. Whitley (Ed.), *Foundations of Genetic Algorithms 2*, pp. 109-126, San Mateo, CA: Morgan Kaufmann, 1993.

[4] Holland, J. H., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, Ann Arbor, MI: University of Michigan Press, 1975.

[5] Holland, J. H., Building Blocks, Cohort Genetic Algorithms, and Hyperplane-Defined Functions, *Evolutionary Computation* Vol. 8:4, pp. 373-391, 2000.

[6] Kimura, M., *The Neutral Theory of Molecular Evolution*, Cambridge, UK: Cambridge University Press, 1983.

[7] Mitchell, M., Forrest, S., and Holland, J. H., The royal road for genetic algorithms: Fitness landscapes and GA performance, In F. J. Varela and P. Bourgine (Eds.), *Toward A Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life,* pp. 3-10, Cambridge, MA: MIT Press/Bradford Books, 1992.

[8] Newman, M. & Engelhardt, R., Effect of neutral selection on the evolution of molecular species, *Proc. R. Soc. London B.*, Vol. 256, pp. 1333-1338, 1998.

[9] Ochoa, G. & Jaffe, K., On Sex, Parasites, and the Red Queen, *Journal of Theoretical Biology*, Vol. 199, pp. 1-9, 1999.

[10] Pelikan, M. & Goldberg, D. E., *Escaping Hierarchical Traps with Competent Genetic Algorithms*, IlliGAL Technical Report 2001003, 2001.

[11] Smith, R. E., Forrest, S., and Perelson, A. S., Searching for Diverse, Cooperative Populations with Genetic Algorithms, *Evolutionary Computation*, Vol. 1:3, pp. 191-211, 1993.

[12] van Nimwegen, E. & Crutchfield, J. P., Optimising epochal evolutionary search: Population-size independent theory, *Computer Methods in Applied Mechanics and Engineering,* Vol. 186:2-4, 2000.

[13] Watson, R. A., & Pollack, J. B., Hierachically-Consistent Test Problems for Genetic Algorithms, In Angeline et. al., (Eds.), *Proceedings of the 1999 Congress of Evolutionary Computing*, pp. 1406-1413, IEEE Press, 1999.