Investigating the Influence of Population and Generation Size on GeneRepair Templates

Amy FitzGerald
Department of Computer Science
National University of Ireland, Maynooth
Kildare, Ireland
amy.fizgerald@nuim.ie

Diarmuid P. O'Donogue
Department of Computer Science
National University of Ireland, Maynooth
Kildare, Ireland
diarmuid.odonoghue@nuim.ie

Abstract— **In 2005 Lolle *et al* published controversial findings showing that the *Arabidopsis thaliana* plant repairs invalid genetic information using the grandparent as a kind of repair template. We have previously shown how a genetic repair operator (GeneRepair) can be used to correct invalid individuals in an evolutionary strategy. It has been shown that superior results are produced when the individual's grandparent is used as the repair template in comparison to using the individual's parent. This paper investigates whether the results produced by GeneRepair templates are affected by parameters of population size and number of generations. The results indicate that the grandparent template outperforms the parent template regardless of population or generation size. These findings further supports the controversial theory of Lolle *et al.***

Keywords-component GeneRepair; evolutionary strategy; non-mendelian inheritance;population size, generation size, Arabidopsis thaliana

I. Introduction

Evolutionary Strategies (ES) are based on the Darwinian principle of survival of the fittest. Their expected success is based on the fact that they mirror biological evolution. ES have been shown to be successful when used on problems with large search spaces, but traditional ES are ill-suited to constraint based problems [1]. Three approaches have been adopted to enforcing constraint upon an ES: penalty points [2], modified ES operators and genetic repair. In this paper we evaluate a modified ES that incorporates a genetic repair process.

In 2005 Lolle *et al* published a controversial paper showing that the *Arabidopsis thaliana* plant used a genetic repair process to correct invalid genetic information [3]. The corrected individuals appeared to be repaired using information originating in the grandparent generation – information that *apparently* by-passed the parent generation. This radical form of non-Mendelian inheritance was treated with much skepticism by some of the scientific community.

In this paper we modify a traditional ES to mirror this genetic repair technique. We compare the effect of population and generation size on the results produced, using both the parent and grandparent template to repair invalid individuals. Invalid individuals are those that do not satisfy the constraints of a given problem – the biological equivalent of producing a viable individual/phenotype. For this paper we have used the TSPLIB eil51 51 city Travelling Salesman's Problem (TSP) to evaluate our hypothesis and compare results. The results show that as the evolutionary strategy is modified to more closely mirror biology, the grandparent repair becomes a superior candidate for use as a repair template than the parent.

II. GeneRepair

GeneRepair [4] is an operator which repairs invalid individuals produced by crossover or mutation in an evolutionary strategy. An example of an invalid individual for the TSP would be an individual with duplicate cities. The GeneRepair operator would ensure that this individual satisfies the problem constraints by replacing the duplicate city with any missing city.

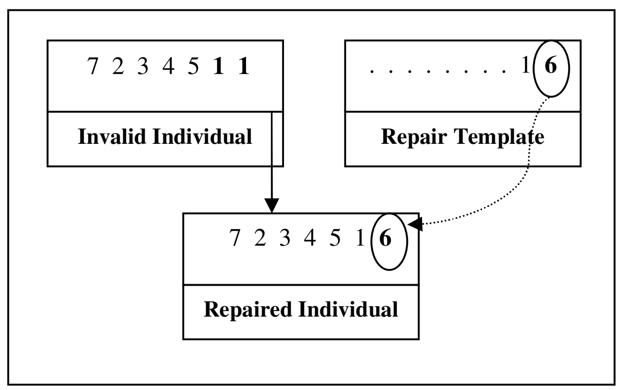

Figure 1. An Individual with an invalid duplicate gene

As we can see in Fig. 1 the individual is invalid as it has a duplicate of city 1. GeneRepair is invoked to repair this error by replacing the duplicate with a missing city. In order to decide which city to replace GeneRepair uses a template. Our research compares the use of the individual's parent and the grandparent as possible repair templates.

III. Results

We ran an experiment to compare grandparent and parent GeneRepair with a population of ten for five hundred thousand generations on the eil51 TSP to investigate the consequences of change in generation and population size on the success of the GeneRepair templates. We ran this experiment 260 times. Our results are presented in two separate subsections. In the first subsection we present the results produced by the 260 runs. We sample these results at 100,000, 250,000 and 500,000 generations and at each point we compare the use of parent and grandparent templates. In the second subsection we analyze the best result produced using the parent template and compare it to the best result produced using the grandparent template. We sample both of these files at 50, 500, 5000, 50000 and 500000 generations and compare the best parent solution and the best grandparent solution at each point.

A. *Comparison of Final Results Produced by Parent and Grandparent Templates*

In previous research [5] we have shown how the grandparent can be used as a template for repair in an ES to successfully find near optimal solutions to constraint based problems. In this paper we investigate whether change in population size or number of generations has an effect on those findings. We have used a population of ten for this set of experiments which is a tenth of the population used in our previous publication. The mutation rate set to two for all experiments. This mutation rate is based on the findings from the investigation carried out in mutation rates by Mitchell [6].

In this first set of results we compare the final results produced by each of the 260 experiments carried out using the parent template to repair invalid individuals to the 260 experiments carried out where the grandparent template was used. Each of the experiments was run for 500,000 generations. In Fig. 2 below you can see that each of the grandparent experiments produced a smaller tour length and so a better result that its parent counterpart.

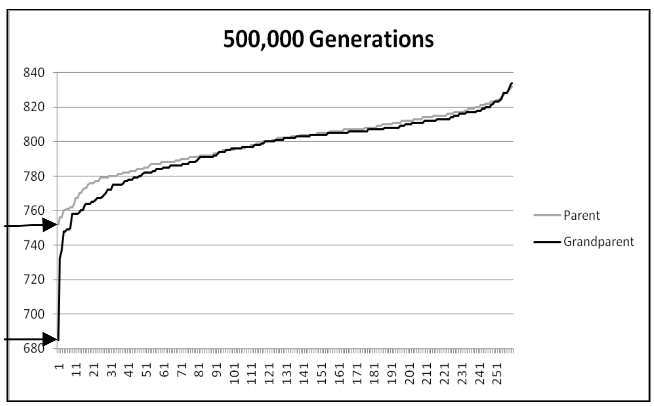

Figure 2. Results for 260 experiments after 500,000 Generations

We went on to sample each of these results at 100,000 and 250,000 generations. In Fig. 3 the result of each experiment is shown after 100,000 generations. It can be clearly seen that there is a significant gap between the tour length produced using the grandparent template and the tour length produced using the parent template for the majority of the results.

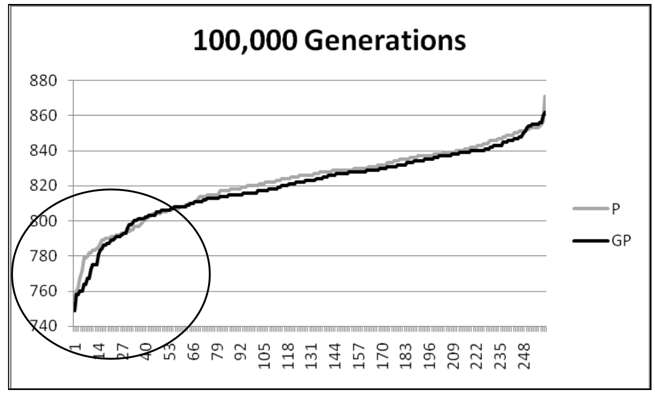

Figure 3. Results for 260 experiments after 500,000 Generations

In Fig. 4 below we can see that the gap between the results produced when using the parent and grandparent templage has grown when the results were sampled at 250,000 generations. The grandparent produces superior results to its parent template for every experiment. You can also see that the best results produced by the grandparent, as highlighted in Fig. 4 are signicantly better than those produced by the parent template.

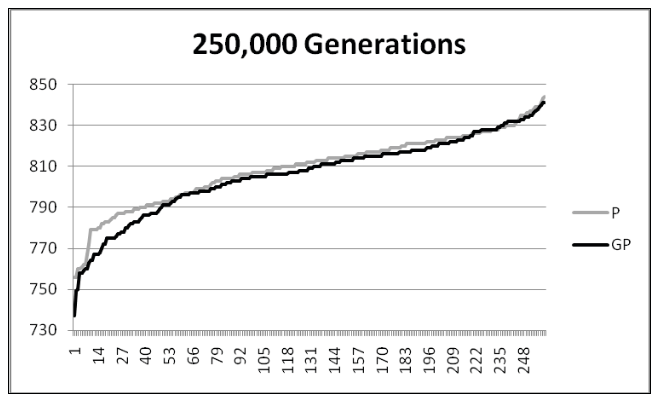

Figure 4. Results for 260 experiments after 500,000 Generations

If we analyze the results produced by the experiment as illustrated in Fig. 2, it is clear that grandparent is superior to parent as a candidate for a GeneRepair template even though the population size was one tenth of the size in previously published results. By not only looking at the final result but also sampling the experimental results at two separate points we can see that the grandparent template produces superior results to the parent template regardless of generation size. If we go on to analyze these results further we can see that while the average grandparent has a lower tour length than the average parent the standard deviation across the results produced by the grandparent template is much wider (See Table 1). Perhaps it is this diversity within the grandparent that allows us to produce better results through wider exploration of the search space.

TABLE I. EXPERIMENTAL RESULTS FOR 260 RUNS

Template	*Tour Length*	
	Average	*Standard Deviation*
Parent	799.2769	16.06984
Grandparent	796.2654	19.63673

B. *Analysis of Best Parent and Best Grandparent Results*

This second section of our results compares and analyzes the best result produced when using the parent template compared to the best result produced when using the grandparent template as highlighted in Fig. 2. We have sampled these results at five different points in the experiment to compare parent and grandparent GeneRepair at five different generation sizes. The results in Fig. 5 compare best results produced by both grandparent and parent GeneRepair after 50 generations. The lines in the graph below indicate the tour length (Y-axis) produced for the eil51TSP at that particular generation (X-axis). We can see that only after 50 generations a small but significant gap has formed between the use of grandparent and parent templates. We can also identify that using the parent template has caused the decrease in tour length to somewhat plateau in comparison to the steady improvement of results when using the grandparent template.

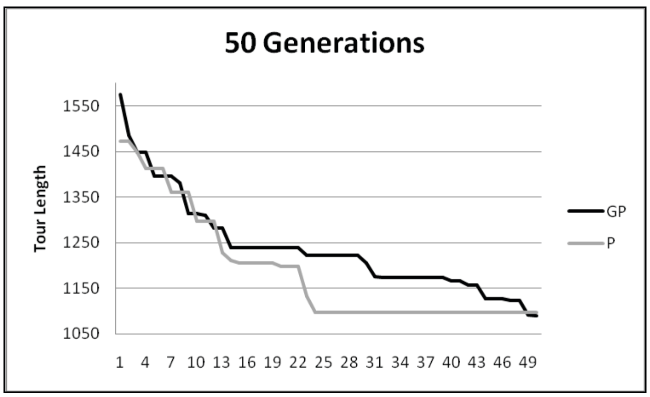

Figure 5. Results after 50 Generations

In Fig. 6 below we can see that the gap between the results produced by the grandparent and parent template has grown significantly after 500 generations. Grandparent appears to continue to evolve while once again the parent template has caused a plateau in the results.

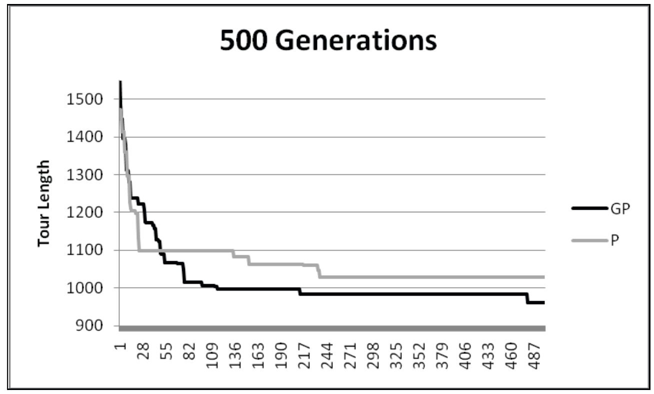

Figure 6. Results after 500 Generations

We have sampled the experiment at 5 different points; 50, 500, 5000, 50,000 and 500,000 generations. The only time that parent produces better results is at 5,000 generations but even at this point the difference between grandparent and parent is minimal in comparison to the difference between them at each of the other sampled points. (See Table 2 & Fig. 7)

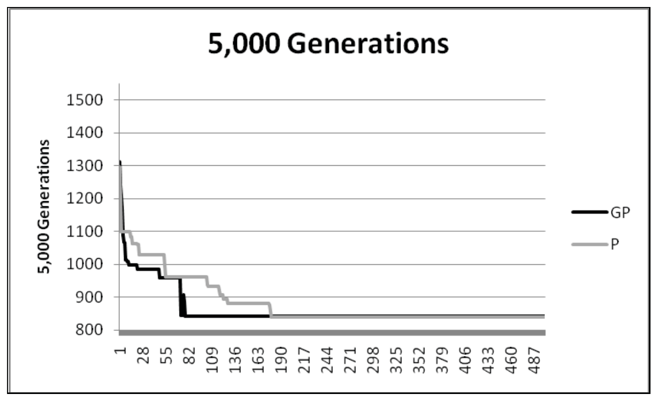

Figure 7. Results after 5,000 Generations

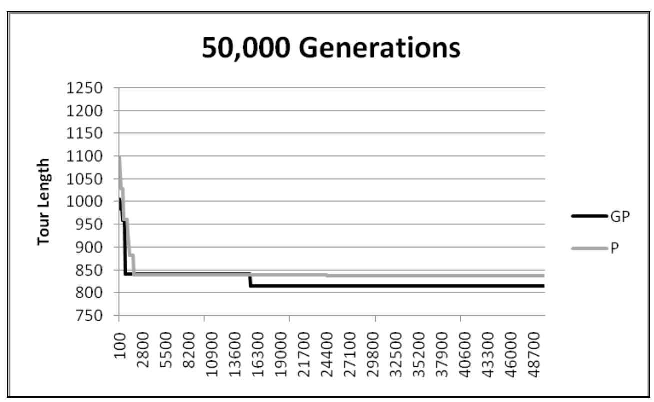

Figure 8. Results after 50,000 Generations

When the results are next sampled at 50,000 generations grandparent is significantly better than parent again which concretes the finding that the result at 5,000 generations was a definite outlier for the parent template in comparison with the constant positive results for the grandparent template (Fig. 8)

The last sample of the results was taken at 500,000 generations and the gap between the parent and grandparent template has once again grown. The final results shown in Fig. 9 illustrate that on completion of the experiment the grandparent template has produced a significantly lower tour length than the parent template.

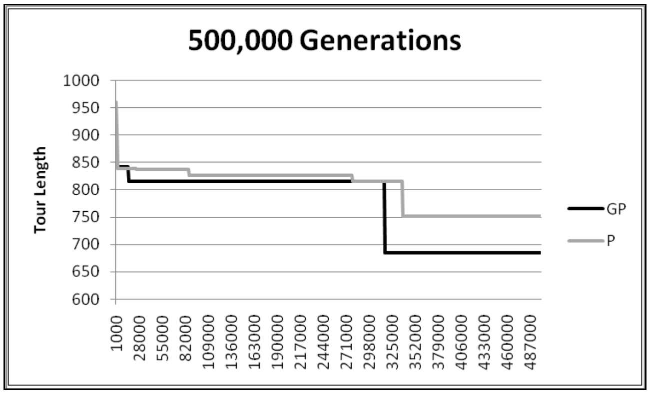

Figure 9. Results after 500,000 Generations

TABLE 2 provides a summary of the results as sampled at the five points explained above. We can see that the difference between the grandparent and parent template becomes more apparent as you increase the number of generations in the experiment. Not only is grandparent a superior repair template, it is impervious to both population size and the number of generations. We can see that it is resistant to changes in the population as this experiment gives the same conclusion as previously published results [5] even though the population used for the results shown in this paper is a tenth of what it was for previous experiments. We have also shown that it is resistant to changes in the number of generations by sampling the results at five different points and comparing the parent template to the grandparent template at each of these points (See Table 2).

TABLE II. FINAL RESULTS

Number of Generations	*Repair Template*	
	Parent	*Grandparent*
50 Generations	1098	1092
500 Generations	1028	960
5000 Generations	839	842
50000 Generations	837	816
500000 Generations	752	685

IV. CONCLUSION

Previous results [5] have illustrated that the grandparent is a superior repair template to the parent when used by GeneRepair in an ES. This paper goes on to show that the superiority of the grandparent repair template is invariant across many population and generation sizes. The population used for the experiments illustrated in this paper was set to 10 in comparison with a population of 100 in previously published results [5]. The final 260 experiments were analyzed at three different generation sizes. The results showed that at each of the sampled points the grandparent template outperformed the parent as a repair template. The results also showed that the grandparent template continuously produced superior results and so at each sampled point in the experiment it was a new best result produced by the grandparent being compared to the parent. The grandparent continually produced superior results as opposed to one superior outlier.

We then went on to examine the best result produced by the parent template in comparison to the best result produced by the grandparent template. The experiment was sampled at five different generation sizes and four out of five of these samples were positive that that the grandparent is a superior template. The one sample that was not positive was too weak to suggest that the parent template was superior. It is also to be expected with a stochastic method such as ES that a small proportion of the results will be unreliable due to the strong influence of outliers on the overall results.

The results presented in this paper strongly support the controversial findings of Lolle *et al* [3] where they show that in the biological environment of the *Arabidopsis thaliana* plant the grandparent template is successfully used as the repair template to correct invalid genetic information. This paper concludes that not only is it possible to use the grandparent as a template for repair, it is shown to be superior to using that of the parent regardless of population and generation size.

ACKNOWLEDGMENTS

The research of Amy FitzGerald carried out as part of her PhD is funded by IRCSET. The authors acknowledge the CS Dept of NUIM for access to the CREIG cluster to run the above.

REFERENCES

[1] Coello Coello, C., “Theoretcial and Numerical Constraint Handling Techniques in Evolutioanry Algorithms: A Survey”, Comp. Methods in App. Mathematics and Engineering,Vol 191(11,12),pp.1245–1287 (2002)

[2] Kalyanmoy, D., “An Efficient Constraint Handling Method for Genetic Algorithms”, Comp. Methods in Applied Mechanics & Engineering, Vol 186(2-4), pp. 311-338 (2000)

[3] Lolle, S.J., Victor, J.L., Young, J.M., Pruitt, R.E., “Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis” Nature, vol.434, pp. 505–509, 2005

[4] Mitchell, G.G., O’Donoghue, D.P., Trenaman, A., “A New Operator for Efficient Evolutionary Solutions to the TSP” Applied Informatics vol. 0-88986-280-X, pp. 771-774, 2000.

[5] FitzGerald, A., O’Donoghue, D.P. *“Genetic Repair for Optimization under Constraints”,* 10th Intl Conference on Parallel Problem Solving From Nature (PPSN2008),Dortmund,Germany,pp399-408, 2008.

[6] Mitchell, G.G., “Evolutionary computation applied to Combinatorial Optimization Problems”, PhD Thesis, Dublin City University, Dublin Ireland, 2007