

Biologically Inspired Non-Mendelian Repair for Constraint Handling in Evolutionary Algorithms

Amy FitzGerald
Dept. Of Computer Science
NUI Maynooth
Ireland
+353-1-7086099
Amy.fitzgerald@nuim.ie

Diarmuid P. O'Donoghue
Dept. Of Computer Science
NUI Maynooth
Ireland
+353-1-7083851
Diarmuid.odonoghue@nuim.ie

ABSTRACT

This paper examines a repair technique that enables evolutionary algorithms to handle constraints. This repair technique, known as GeneRepair, repairs invalid individuals so that all problem constraints are met by every individual in the population. GeneRepair is based on the repair technique used by the *Arabidopsis thaliana* plant which was proposed by Lolle *et al* in 2005. This controversial repair method uses information inherited from ancestors previous to the parent (non-Mendelian inheritance) as a repair template to fix errors or invalidities in the current population. We compare the use of three different ancestors as repair templates and investigate the effects of various biological parameters on the choice of repair template to use.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search – *Heuristic Methods*; G.1.6 [Numerical Analysis]: Optimization – *Constrained Optimization*

General Terms

Algorithms, Performance, Design, Experimentation, Theory,

Keywords

Evolutionary Algorithms, Biologically inspired, Genetic Repair

1. INTRODUCTION

Mendelian inheritance is a corner stone of our understanding of natural and simulated evolutionary systems. But recent discoveries in plant genetics suggest that exceptions to this fundamental tenet may serve to correct genetic defects [1]. This genetic repair process appears to act as an additional inheritance mechanism that is only activated when genetic defects are detected. Furthermore, this additional inheritance mechanism appears to make use of ancestral genomic information that is not detected by current techniques. This process has been referred to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'10, July 7–11, 2010, Portland, Oregon, USA.
Copyright 2010 ACM 978-1-4503-0073-5/10/07...\$10.00.

as the “parallel path of inheritance” (Robert Pruitt, Washington Post, March 23rd, 2005).

The work described in this paper is a direct result of the recent discovery of a genetic repair process. Lolle *et al* [1] identified a Non-Mendelian genetic repair mechanism in the *Arabidopsis thaliana* plant. This repair mechanism uses information which was not present in either parent to repair invalid individuals. We have adapted this repair technique to correct invalid solutions in an evolutionary algorithm by using ancestors as repair templates. In this paper we view this proposed genetic repair mechanism as means of enforcing constraints on the search space that is explored by an evolutionary system. As the proposed repair mechanism is still poorly understood, this paper explores a variety of influences on this novel genetic repair process.

Thus far, there have been four distinct approaches to imposing constraints upon evolutionary search spaces. These four approaches are: penalizing invalid individuals, using modified representations and operators that prohibit infeasible solutions, multi-objective optimization separating fitness and feasibility metrics and finally the genetic repair approach. Of these, the final genetic repair approach has received relatively little attention and is the focus of this paper.

There have been a number of investigations into the use of genetic repair operators within evolutionary algorithms. These include: harmonization [2], 2-repair [3] GENOCOP III [4] DER [5] and CleanUp/GeneRepair [6] as well as [7] and [8]. (For a recent review of genetic repair approaches see [9]. Many of these employ problem specific heuristics [10] to transform invalid solutions into similar valid ones.

In our previous work we looked at a biologically inspired GeneRepair process [11]. The main differences in this paper are that we explore a range of mutation rates, several population sizes and we also investigate the impact of GeneRepair during very early evolution. All results in this paper were generated on three new, larger problem instances.

To our knowledge no other authors have adapted a genetic repair mechanism and applied it as a constraint handling mechanism. Our approach enjoys the benefit that it can, in principle, be applied across many disparate problem domains. This flexibility arises because our GeneRepair approach uses corrective templates to direct the repair process, with these templates being sourced

from previous generation of valid solutions. The main requirements are that genetic errors can be detected and that a repair template can serve to correct these errors.

This paper focuses on adapting the standard evolutionary algorithm (EA) to incorporate this newly discovered repair technique. The hope was that this would make the EA capable of handling constraints in an efficient yet domain independent manner. Our primary focus is to explore the parameters which lead to near optimal solutions to constrained problems in a reasonable amount of time. This paper shows how the different biological parameters in the EA can effect the choice of template to use with this repair strategy. We look at adapting of population size, mutation rate and number of generations as well as adapting the repair strategy itself to optimize the solutions produced.

2. EA & CONSTRAINTS

EAs operate on the theory of survival of the fittest. A random population is produced and certain individuals are selected depending on their fitness levels. These individuals are crossed over to produce a new population, which is mutated at a pre-chosen rate and this is repeated for a set number of generations. While this is the basic evolutionary algorithm, there is huge scope for fine tuning of these operators. For the experiments described in this paper we have chosen to use truncation selection for its simplicity [12]. There are many crossover methods available of which I have chosen to use single point. Various mutation rates are compared in a later section.

As previously mentioned, the operators used by the standard EA (crossover and mutation) are blind to the constraints that exist in many problem domains. Thus, their use frequently results in the generation of invalid individuals that do not correspond to valid solutions to the given problem. Four approaches have previously been adopted to enforce constraints on EA searches [13]. Firstly, modified operators and representations ensure invalid solutions are never produced. Secondly, penalty points are used to favour individuals that do not possess any constraint violations. Third, multi-objective optimization treats fitness and constraint violations as independent factors influencing survivability. Finally, invalid solutions can be somehow repaired in order to preserve the diversity and validity of future generations. The use of modified operators reduces the biological integrity of EAs and as they were first modeled on biology this could be seen as contradictory going forward. The biological plausibility of the penalty infliction approach lies in the fact that the less valid the individual is (ie the more constraints an individual breaks) the less chance it has to procreate which is similar to the idea of survival of the fittest. A significant problem with this approach however is that it can lead to a large reduction in diversity across the population which can in turn force the individuals towards a local optimal as opposed to the desired global optimal. We investigate the third option of GeneRepair which is modeled on the repair mechanism of the *Arabidopsis thaliana*. As this repair technique is found in nature it is fitting that it should be used with EA which are modeled on natural evolution.

3. GENEREPAIR

Arabidopsis thaliana (thale cress) is a model plant used for a wide variety of detailed studies and was the first plant genome to be fully sequenced. Lolle *et al* [1] investigated *A. thaliana* plants with an organ fusion mutation on the Hothead gene (*hth*), which resulted in an abnormal formation of the plants flower. The studies of Lolle *et al* revealed that two plants with the *hth* mutation can produce offspring without this abnormality. . Surprisingly, approximately 10% of the offspring of two mutant plants were found to revert to the normal form of the hothead gene [14] Thus, the resultant offspring have the normal form of the hothead gene (*HTH*), even though this information was present in neither of the parental genomes. This high rate of reversion is far higher than can be explained by random mutations of these specific alleles, which would be of the order of 1 per billions [15] per allele per generation. It is thought that this phenomenon is not exclusive to the *Arabidopsis thaliana* but occurs in a range of organisms, including the Soya bean. It was found that these revertant genomes all appeared to inherit genetic information from their grandparent's genome, which did have the normal (*HTH*) form. Thus, genetic information appeared to skip a generation, reappearing in a subsequent generation. This has been referred to as a "parallel path of inheritance", which appears to occur in addition to standard Mendelian inheritance. In essence, a corrective template is used to correct broken or damaged sequences of DNA, possibly in response to stress placed on the plant due to the presence of the genetic mutation.

While Lolle's [1] controversial explanation relies on a cache of RNA inherited from previous generations, our approach is more similar to the explanation offered by Ray [16] but is still compatible with Lolles findings. Rays explanation relies on an archival form of DNA that serves to store the ancestral DNA, but which is not detected by the processes used to sequence the regular encoding of DNA.



Figure 1 *Arabidopsis thaliana*

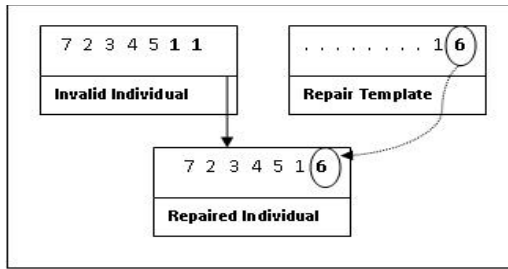


Figure 2 Template Repair

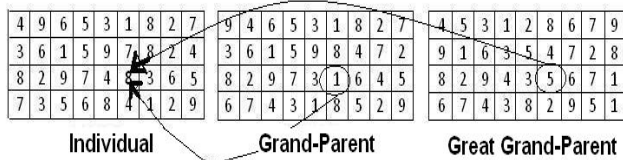


Figure 3 Ancestral Template Repair

4. GENEREPAIR TEMPLATES

As shown in Figure 3 above, GeneRepair replaces an invalid allele with a selected allele from the repair template provided. Current research investigates which ancestral repair template is best to use and to do this we continuously compare the use of the parent, grandparent and great grandparent templates while adjusting the various other biological parameters such as problem size, population size, mutation rate and number of generations.

The results illustrated in this paper were all produced using the 532 Traveling Salesman Problem [17] dataset. The same results were produced using TSPs of different sizes as shown in Section 8 at the end of the paper. The experimental set up for the results outlined in the sections below can be described as follows; the problem being used is the 532 city Traveling Salesman Problem (att532 TSP), with a population of 50 unless otherwise stated, running for 500,000 generations, using the truncation selection method and single point crossover. Swap mutation is used throughout this paper to simplify our analysis because we wished to limit the source of genetic errors to crossover.

The experiments in this paper were produced by carrying out GeneRepair in a random and changing direction. To explain this point, we make use of the illustration in Figure 4. The defect in the erroneous individual concerns repetition of city 1 on the corresponding TSP tour. One of these duplicates must be repaired and this direction parameter is partly responsible for deciding which duplicate is replaced. In left-to-right repair the left most duplicate is replaced, resulting in the generation of “repaired individual a”. A similar process occurs for right-to-left repair, resulting in the generation of “repaired individual b”. We note that the two different directions can produce two different individual using the same repair template. Thus, this repair process corrects all errors in the population, unlike [8] who propose a much lower rate of repair.

Randomly changing repair dynamically chooses between these two repair directions for each individual genetic defect. This repair strategy allows us to guarantee the validity of all repaired

individuals while maintaining the repair template (parent, grandparent or great-grandparent) as the primary controller of the repaired individuals.

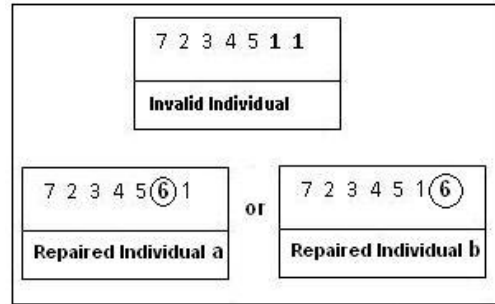


Figure 4 Direction of Repair

Figure 5 compares the results produced by parent, grandparent and great-grandparent templates with a mutation of 0.1%. The results indicate the tour lengths produced when each ancestor was used for genetic repair across 26 experiments. We can see that grandparent outperforms both parent and great-grandparent by producing shorter tour lengths for these parameters. A Mann-Whitney analysis of these results show that grandparent template outperforms parent with a confidence level of $p < 0.0001$ and great-grandparent with a confidence level of $p < 0.0001$.

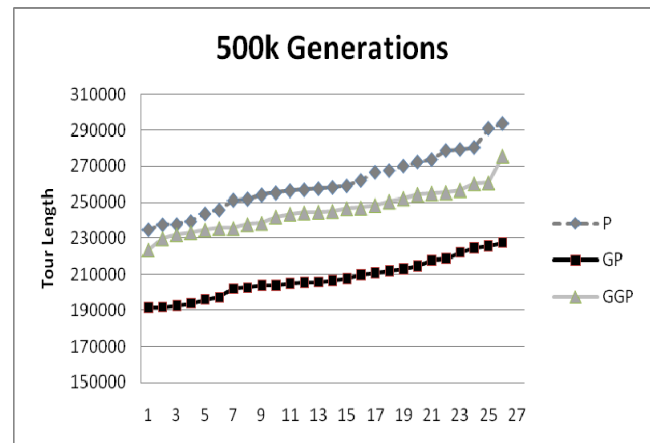


Figure 5

5. EARLY EVOLUTION

To investigate further how the results above were produced we can examine what happens when the individual templates begin to take effect in the early generations.

To begin the evolutionary algorithm a population is randomly generated. Each individual in this population obeys the problem constraints. For the first generation this randomly generated population is copied into the parent location. These parents are then crossed over to create a new population. The only biological repair template available to use is the parent and thus the parent is used to carry out GeneRepair on the current population in order to repair errors introduced at crossover. To create the second generation the parent population is copied into the grandparent location and the current population is copied into the parent

location. The parents are then crossed over to create a new population.

To create the third generation the grandparent population is moved into the great-grandparent location, the parent population is moved into the grandparent location and the current population is copied into the parent location. The parents are then crossed over to create a new population. There are now three template choices to use for GeneRepair; the parent, the grandparent or the great-grandparent. The parents have been used to create the current population and are therefore less contrasting to the population than the older templates. The grandparent has the ability to introduce greater diversity into the population as it was not directly crossed over to produce the current population. The great-grandparent is the most removed from the population and therefore has the ability to introduce the widest amount of diversity.

Figure 6 illustrates the results produced at generation three. This could be seen as the first full evolutionary cycle of the system as each template has been updated with biological data. The lines in the graph compare the results produced by each repair template when the experiment was run 26 times. The problem set is the 532 city TSP which means that the lower the tour length (Y-AXIS) the better the result. The best result is produced when the grandparent is used as a repair template. A Mann-Whitney analysis of these results show that grandparent template outperforms parent with a confidence level of $p < 0.0001$ and great-grandparent with a confidence level of $p < 0.0001$.

We can see that the results produced by using the great-grandparent repair template do not produce tour lengths as low as that of the grandparent but are closer in length to those produced by the grandparent template than those produced by the parent template. We can also see that at this early point in evolution there is a clear and distinct difference between using the parent template and the grandparent template. The difference between the parent template and the current population that is being repaired is not large enough to improve the population as significantly as the grandparent population.

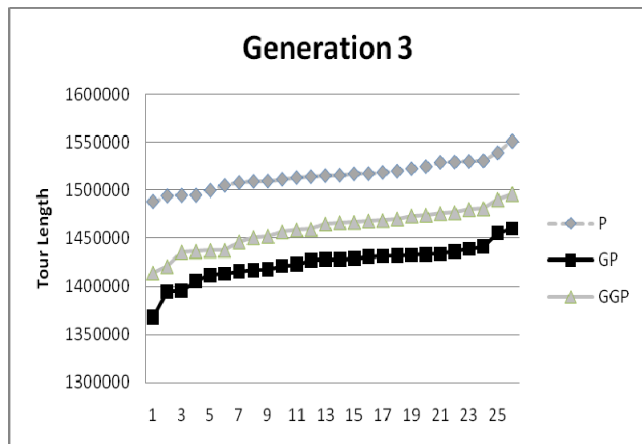


Figure 6

Allowing this experiment to evolve further the difference between the results produced by the grandparent and great-grandparent repair template becomes more significant as the tour lengths decrease. Figure 7 shows that while the results produced by parent

and great-grandparent repair templates have improved slightly between generation 3 (Figure 6) and generation 10 (Figure 7) the results produced by the grandparent repair template have improved drastically with much shorter tour lengths being produced.

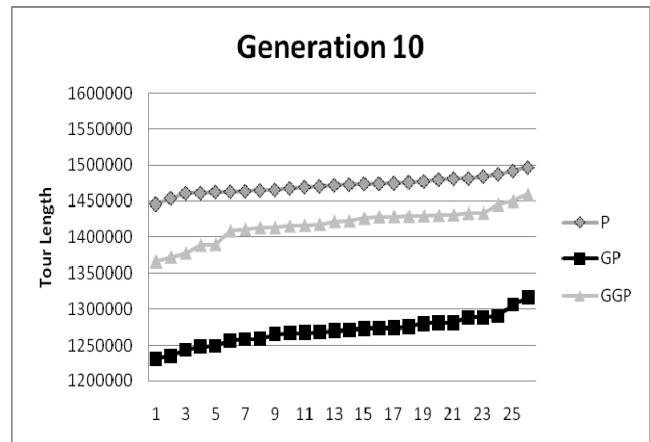


Figure 7

From this brief analysis it is clear that the difference between repair strategies emerges quite quickly. For these experimental parameters, it appears that the grandparent strategy offers the best compromise between exploiting the existing population, while using the mutagenic effects of GeneRepair to explore new regions of the problem space.

6. MUTATION

For all of the experiments described in this paper swap mutation has been used as it does not cause solutions to become invalid, this allows us to restrict the introduction of invalid solutions to crossover. Figure 5 was produced using a mutation rate of 0.1%. We now investigate the effect of raising this mutation to 1% and 2%. As shown in the Figure 8, 9 and 10 a higher mutation rate does not affect the rank of the templates but the lower tour length was produced by the mutation rate of 0.1%

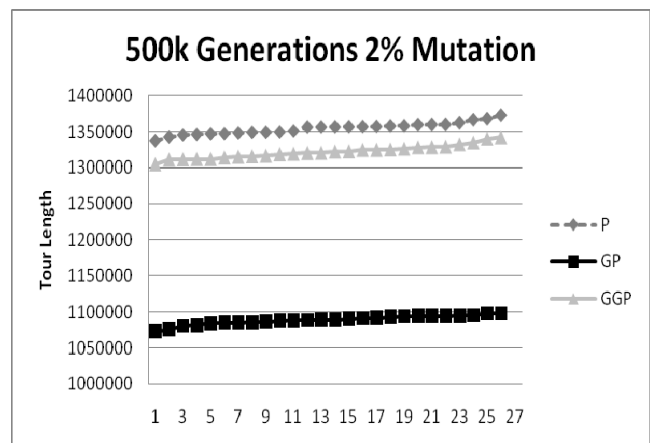


Figure 8

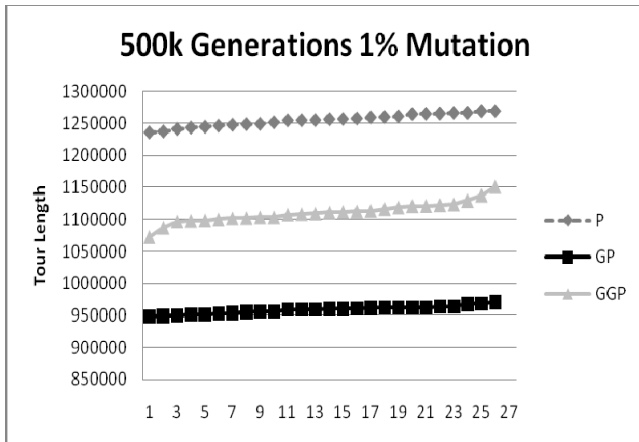


Figure 9

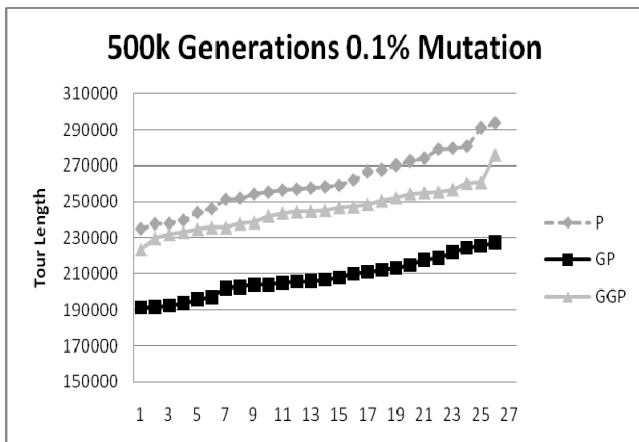


Figure 10

7. POPULATION

A population of 50 individuals has been used for all previous experiments illustrated above. We now investigate the results produced when a population of 100 is used. Figure 11 shows that while the lowest tour length produced by the grandparent, 188697 compared to 191533 with a population of 50, is not significantly different than that produced using a population of 50, the highest result (ie the worst result) produced is higher than that produced when using a population of 50, 238307 compared to 227546 with a population of 50. We can also see that the results produced by the parent and great-grandparent are superior to those produced by a population of 50. The graph illustrates that there is less of a significant difference between using the parent and great-grandparent repair templates. This could be due to the fact that there is greater diversity across the population due to the larger population size so the individuals are less vulnerable to diversity introduced by the repair templates.

There is a significant computation and therefore time cost involved in doubling the population from 50 to 100 and the payback being a decrease of 2836 in the lowest result produced does not warrant this cost.

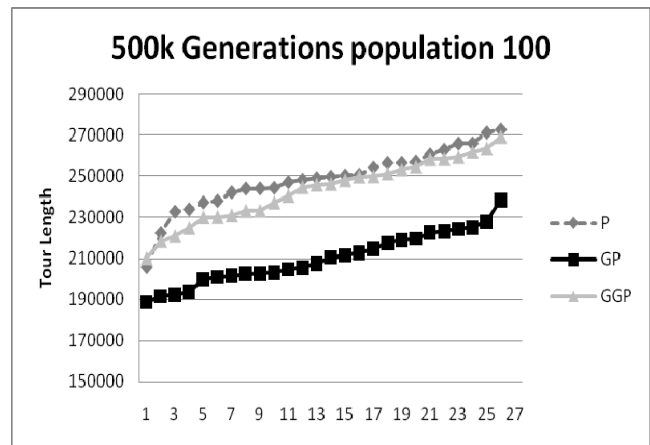


Figure 11

8. PROBLEM SIZE

All of the experiments above were run on a 532 city TSP (att532). If we run this experiment on a 1379 city TSP the following Figure is produced.

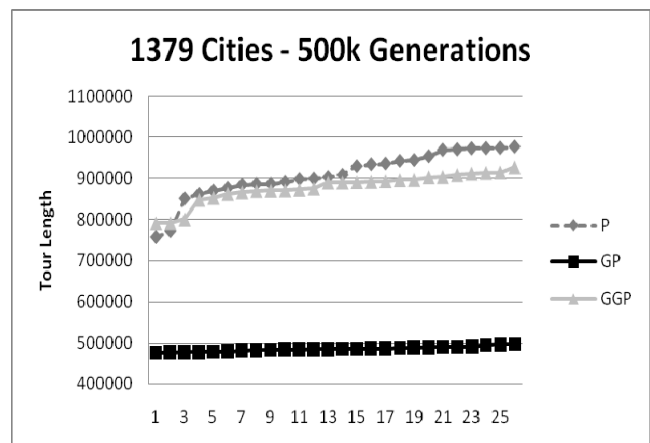


Figure 12

This experiment was run for 10,000 generations, the population was 100 and as above in Figure 5 the mutation was 0.1%. We can theorise from Figure 12 that parent introduces too little diversity into the population while great-grandparent produces too much diversity both leading to longer tour lengths than those produced using the grandparent repair template. This can also be seen in Figure 13 where the 76 city TSP was used for 10,000 generations with a population of 100.

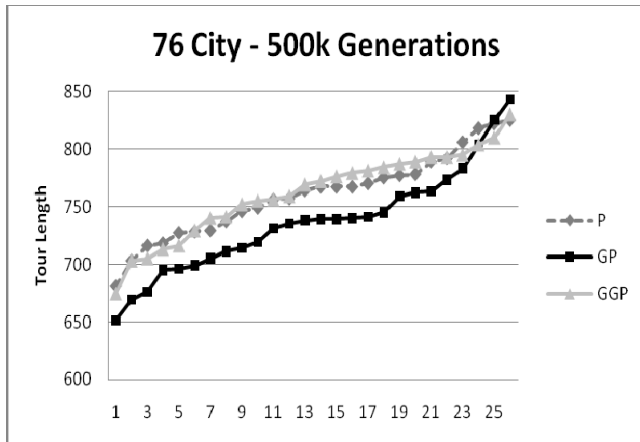


Figure 13

9. FUTURE WORK

In this paper and previous publications we have shown how GeneRepair can be used in conjunction with evolutionary algorithms to handle constraints using the TSP. We are currently conducting experiments across a range of complex problems to show that this repair technique is not problem specific. We are also investigating the use of different problem representations as well as alternative mutation and crossover methods.

10. CONCLUSION

In 2005 Lolle *et al* published a paper [1] proposing to have found a Non-Mendelian repair mechanism in the *Arabidopsis thaliana* plant. This repair mechanism used inherited information which was not present in either parent to repair individuals in the current population. We have mirrored this repair technique to repair invalid solutions in an evolutionary algorithm using ancestral repair templates.

This paper focuses on adapting a standard EA by incorporating the repair technique found in the plant *Arabidopsis thaliana* thus making the EA capable of handling constraints. Our primary focus is to explore some of the parameters which yield near optimal solutions to constrained problems in a reasonable amount of time. This paper examined how the different biological parameters in the EA can affect the choice of template to use with this repair strategy.

We compared the use of the parent, grandparent and great-grandparent repair templates. The original experiment used a population of 50 with a mutation rate of 0.1%. We found that the grandparent template far outperformed the use of the parent or great-grandparent template. This could be due to the fact that the parent template does not introduce enough diversity into the population while the great-grandparent template introduces too much diversity into the population.

We went on to investigate the biological parameters used in this experiment by examining what happens when the mutation rate is lifted from 0.1% to 1% and further to 2%. At 1% mutation the results degraded significantly with tour lengths of almost 5 times the length of those produced by a mutation rate of 0.1%. While

the results were drastically different the template ranking remained the same with the superior template being the grandparent. There is however a more significant difference between the use of the great-grandparent and parent templates as the great-grandparent would introduce even greater diversity than before with the mutation set to a higher rate. At 2% mutation the results degraded again with the grandparent remaining the superior template choice. The difference between the use of the parent and great-grandparent template was significantly smaller as the diversity introduced by the high mutation rate would make the population less vulnerable to the diversity introduced the repair template.

The final biological parameter investigated in this paper was population size. It was found that the grandparent remained the superior template to choose. While a slightly better result was produced when a population of 100 was evolved the time and computation costs were not economical for such a small deviation in the results.

In this paper we have illustrated a constraint handling mechanism for EAs which mirrors that found in the *Arabidopsis thaliana* plant. We have compared the use of three ancestral repair templates and shown that while a variety of biological parameters were investigated the grandparent template remained the superior choice to use with GeneRepair to enforce constraints in an EA.

11. ACKNOWLEDGMENTS

We would like to acknowledge Susan J. Lolle and Marianne Kaup Hopkins, Department of Biology, University of Waterloo, Canada for information relating to the proposed genetic repair process in *Arabidopsis thaliana* and Marianne Kaup Hopkins, for the images of *A. thaliana*. This research is funded by the Irish Research Council for Science Engineering and Technology (IRCSET). The authors acknowledge the Computer Science Dept of NUIM for access to the CREIG cluster to run the above. The authors acknowledge the Irish Center for High End Computing (ICHEC) for the use of their cluster to run the above.

12. REFERENCES

- [1] Lolle, S.J., Victor, J., Young, J., Pruitt, R.: Genome-wide non-Mendelian inheritance of extra-genomic information in *Arabidopsis*. *Nature* 434(1) (2005) 505–509
- [2] Nakano, R., Yamado, T.: Conventional genetic algorithm for job shop problems. *Proc. 4th International Conference on Genetic Algorithms* (1991) 474–479
- [3] Gorges-Schleuter, "ASPARAGUS: An Asynchronous Parallel Genetic Optimization Strategy", *Proc. of ICGA-89*, pp.422-427
- [4] Michalewicz Z. & Nazhiyath G. *Genocop III: A Co-evolutionary Algorithm for Numerical Optimization Problems with Nonlinear Constraints*. *Proceedings of IEEE International Conference on Evolutionary Computation*. 1995
- [5] T. Walters, "Repair and Brood Selection in the Traveling Salesman Problem," *Proceedings of the 5th International*

- Conference on Parallel Problem Solving from Nature - PPSN V, pp. 813-822, 1998.
- [6] Mitchell, G.G. , O'Donoghue, D. Trenaman, A New Operator for Efficient Evolutionary Solutions to the Travelling Salesman Problem, Applied Informatics, Innsbruck, Austria, pp 771-774, 2000,
- [7] Arroyo, J.M. Conejo, A.J. A Parallel Repair Genetic Algorithm to Solve the Unit Commitment Problem, IEEE Transactions on Power Systems, Vol. 17, no. 4, November 2002.
- [8] Orvosh, D. & Davis L. Shall We Repair? Genetic Algorithms, Combinatorial Optimization and Feasibility Constraints. Proc. International Conference on Genetic Algorithms, Volume 5, 1993.
- [9] Salcedo-Sanz, S. A survey of repair methods used as constraint handling techniques in evolutionary algorithms, Computer Science Review 3, pp 175-192, 2009.
- [10] Ray, S.S. Bandyopadhyay, S. Pal, S.K. Genetic operators for combinatorial optimization in TSP and microarray gene ordering, Applied Intelligence 26:183–195, 2007
- [11] Fitzgerald, A., O'Donoghue, D. P., Liu, Xinyu. Genetic Repair Strategies inspired by *Arabidopsis thaliana*, Lecture Notes in Artificial Intelligence, forthcoming (2010)
- [12] Blickle, T. & Thiele, L. A comparison of selection schemes used in evolutionary algorithms, *Evolutionary Computation* Volume 4, Issue 4 (Winter 1996) Pages: 361-394. MIT Press
- [13] Coello, C.C.: Theoretical and numerical constraint handling techniques in evolutionary algorithms: A survey. *Computer Methods in Applied Mechanics and Engineering* 191 (2002) 1245–1287
- [14] Coghlan, A.: Rogue weeds defy rules of genetics. *New Scientist* 2492 (26 March 2005)
- [15] Weigel, Jurgens: Hothead healer. *Nature* 434 (2005) 443
- [16] Ray, A.: Plant genetics: Rna cache or genome trash? *Nature* 437(E1-E2) (2005)
- [17] Reinalt, G.: *Tsplib - a travelling salesman problem library*. OSRA Journal of Computing 3, 376 – 384, (1991)