

Genetic Repair Strategies inspired by *Arabidopsis thaliana*

Amy FitzGerald, Diarmuid P. O'Donoghue, Xinyu Liu

National University of Ireland, Maynooth, Co. Kildare
{amy.fitzgerald; diarmuid.odonoghue; xinyu.liu}@nuim.ie

Abstract. Recent advances in genetics controversially suggest that the model plant *Arabidopsis thaliana* performs genetic repair using genetic information that originates in the individual's grandparent generation. We apply this ancestral genetic repair strategy within an Evolutionary Algorithm (EA) to solve a constraint based optimisation problem. Results indicate that the grandparent based *genetic repair* strategy outperforms the parent alternative. Within this framework, we investigate the impact of storing only the fittest ancestors for use as a repair template. The influence of performing repair in a fixed direction is compared to randomly varying the direction in which error detection proceeds. Finally we explore the impact of varying the direction of repair on the results produced. All results seem to support the non-Mendelian inheritance process suggested by Lolle *et al.*

1 Introduction

This paper attempts to strengthen the parallels between evolutionary algorithms and recent advances in genetics, within the framework of naturally inspired computation. We explore a recent controversy in biology that centres on a non-Mendelian inheritance mechanism that has been supposedly identified in the model plant *Arabidopsis thaliana* [1]. The mechanism that has been controversially [2] suggested by Lolle *et al* involves the transfer of genetic information from a grand-parent's genes into an offspring - apparently by-passing the parents' genes. This non-Mendelian inheritance process was suggested not as part of the plants normal inheritance process, but in response to specific genetic errors that were apparently corrected by a grandparent-based genetic repair process. Specific mechanisms for carrying out this non-Mendelian repair process include an RNA based archive [1] and a form of "archival" DNA [3].

In this paper we examine the proposed non-Mendelian inheritance mechanism from the perspective of an evolutionary algorithm. This repair strategy is adapted to operate with a constrained optimisation process. This approach allows us to examine the performance of Lolle *et al*'s proposed non-Mendelian inheritance based genetic repair process. In particular, this strategy will be compared against simpler alternative approaches.

The structure of the paper is as follows. First we compare the performance of the ancestral based, template driven, GeneRepair process against the alternative

penalty function system. Next, we compare the performance of the parent and grandparent repair templates. In the third set of experiments, we manipulate GeneRepair by using fitness as a selection criterion on the stored repair templates. In the fourth set of experiments, we manipulate the direction in which the error detection phase of GeneRepair proceeds, examining its impact on the fitness of the population. Finally we compare the earlier results against the use of great-grandparent based GeneRepair, examining the impact of using even more ancient genetic data to guide the repair process.

2 Constraint-based Optimisation with Evolutionary Optimisation

Evolutionary algorithms (EA) has been successful in exploring complex solution spaces, but EA are ill-suited to supporting constraints on these search spaces. For the purposes of this paper, we see both natural and simulated evolution focusing on a highly constrained optimisation problem, where a minimal change to a feasible solution is very likely to generate an unfeasible solution [4]. Evolutionary optimisation (EO) is the term given to the subset of EA that addresses constraints. We investigate if these EO can be improved so that they can handle constraints in a reasonable manner by incorporating a genetic repair process that reflects recent biological findings.

Four distinct approaches have been adopted to enforce constraints on evolutionary searches [5]. Firstly, crossover and mutation operators are modified to ensure that all solutions obey the constraints. This can also take the form of problem specific representations in which invalid solutions cannot be generated. This is not only a biologically implausible approach, it also limits the evolutionary algorithm to solving one (or just a few) problem types. The second approach to dealing with constraints is to use a penalty function (including the death penalty). This method ensures that valid genotypes have the greatest influence on subsequent generations by restricting the reproduction of invalid individuals. (The "death penalty" completely eliminates the influence of invalid individuals on subsequent generations). The biological plausibility of this approach rests in the fact that invalid individuals do not form viable phenotypes and thus can not influence subsequent generations. The penalty inflicted can range from a light penalty for constraint violation to the death penalty [6] being placed on erroneous individuals. The third constraint enforcement method is to adopt a pareto-optimal approach [5]. The final method of enforcing constraints is through a genetic repair operator and it is this 'GeneRepair'[7], [8] approach that is explored in this paper.

A number of techniques for genetic repair have been explored, including: use of heuristics [9], the template approach [10] and the harmonisation operator [11]. This paper explores biologically inspired variants of the GeneRepair approach, that overcomes the problem dependence associated with many constraint enforcement techniques [12]. The GeneRepair approach offers the advantage of using an unmodified evolutionary engine to explore the problem space. Signifi-

cantly the genetic repair process presented in this paper uses ancestral genomic data which can always be made available for any problem domain.

The results in this paper were produced using a simple experimental setup described in section 5. For this paper we have used the TSPLIB eil51 51 city Travelling Salesman Problem (TSP) to evaluate our hypothesis and compare results. This is a combinatorial optimisation problem where each individual in the EO represents a tour of all of the cities in the TSP, and each element represents a city. The number of possible solutions for the TSP is $n!$, where n is the number of cities on the tour. However, the total space that can be explored by an un-constrained EO is n^n where $n^n \gg n!$ for large values of n .

Thus, an un-constrained EO will explore not just the search space required by the TSP but will also generate a large number of invalid solutions. The genetic repair operator modifies invalid solutions that lie outside the valid solution space and converts them back into valid members of the solution space. The biological process that inspired our approach shall now be described.

3 Genetic Repair in *Arabidopsis thaliana*

Arabidopsis thaliana (thale cress) is a model plant used for a wide variety of detailed studies and it was the first plant genome to be fully sequenced. Lolle *et al* (2005) investigated *A. thaliana* plants with an organ fusion mutation on the Hothead gene (*hth*) [13], which resulted in an abnormal formation of the plant's flower. The studies of Lolle *et al* revealed that two plants with the *hth* mutation can produce offspring without this abnormality. The resultant offspring have the normal form of the hothead gene (HTH), even though this information was present in neither of the parent genomes. Surprisingly, approximately 10% of the offspring of two mutant plants were found to revert to the normal form of the hothead gene [14]. This rate of reversion is far higher than can be explained by random mutation 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 flax.

It was found that these revertant genomes all appeared to inherit genetic information from their grand-parents genomes, which had 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 *et al*'s (2005) controversial explanation relies on a cache of RNA inherited from previous generations, our approach is more similar to the explanation offered by Ray (2005) that is compatible with Lolle *et al*'s findings. Ray's 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.

In this paper we examine the efficiency of this “parallel path of inheritance” using the TSP as our benchmark problem for generating and comparing results. There were two main objectives behind these experiments: first to evaluate the efficiency of grand-parent based genetic repair, comparing it to a variety of alternative strategies. Secondly, these experiments attempt to assess if there is any reason to believe that Lolle *et al*’s controversial findings might, in fact, find some support in the analogous domains of evolutionary algorithms.

4 Ancestor based Genetic Repair Strategies

The GeneRepair operator used in this paper is modelled on that of the model plant *Arabidopsis thaliana*, ensuring that each individual maintains its own archive of ancestral genetic information. Thus, each individual in the population has its own repair template based on its ancestors genetic information. These repair templates are called upon when an invalid individual is produced by the evolutionary process.

Each allele in our EO encodes a single city and each city is uniquely encoded within the fixed length representation. Therefore there is a 1-to-1 association between cities of the TSP problem and the city’s representation within the EO. Solutions to the TSP (tours) are formed as an ordered list of cities and the entire population is composed of a fixed number of individual tours (see Figure 2). So, the relative order of cities determines their position within a tour.

The constraints of the TSP problem require that each city must be visited exactly once. Our fixed length genetic representation resulting in two types of genomic error, which are found in pairs. Firstly, duplication errors occur when a city is repeated within a tour (individual). Secondly, omission errors occur when a city is absent from a candidate solution. Thus, an omission error always has a corresponding duplicate error. As can be seen in Figure 1 duplication of the “1” causes omission of “6” from the genetic sequence. The GeneRepair operator identifies all duplicate alleles and replaces each one with the corresponding element in the repair template.

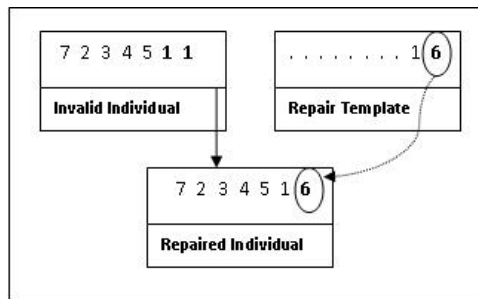


Fig. 1. GeneRepair using a Fixed Template

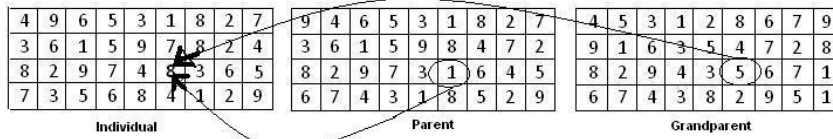


Fig. 2. GeneRepair using an Inherited Template

Ancestral repair templates are stored as independent archived populations. For efficiency only the required template type is stored (See Section 5.3)

In the next section we investigate the relative efficiency of the *parent* and the *grandparent* templates.

5 Results

All of the following experiments use the same experimental setup. Our EO was written in Java and incorporates the MersenneTwister PRNG [16]. This random number generator differs to that used in our previous work [8]. The EO uses the following experimental setup; single point crossover, swap mutation at 2%, rank selection, a population size of 500 and each experiment was run for 500,000 generations. The data below was generated for the *eil51* problem set from TSPLIB [17]. Due to the stochastic nature of EO, each experimental condition was tested 26 times.

In this section we report on the effectiveness of the grandparent based genetic repair strategy, as applied to the TSP. First, grandparent based repair is compared with the alternative “death penalty” strategy. In this strategy, invalid individuals are subject to the “death penalty” and so are removed from the population and do not influence subsequent generations.

5.1 Grandparent GeneRepair Template and the Death Penalty

The results detailed in Table 1 indicate that grandparent based genetic repair produces significantly better results than the alternative “death penalty” strategy, for our sample problem of the TSP. Thus the grandparent strategy produces far fitter individuals for this minimisation task.

5.2 Parent and GrandParent Repair Template Results

In our second experiment we compare the performance of grandparent based repair against the simpler parent based strategy. Parent based repair uses the genetic information from each individual’s parent as a template to repair genetic errors in the child generation, while grandparent based repair uses the individual’s grandparent as the template.

Table 1. Template Inheritance and the Penalty Function

| | Min | Mean |
|----------------------|--------|----------|
| Penalty Function | 1486.4 | 1584.094 |
| Grandparent Template | 483 | 519.81 |

The following table summarises the results across three directions of repair (right to left, left to right and random) and two inheritance selections (fittest and random). All of these experimental conditions are investigated further in the following sections. Grandparent based repair generates the fittest individual

Table 2. Effect of Ancestry on GeneRepair

| Template | Min | Mean |
|-------------|-----|--------|
| Parent | 463 | 519.49 |
| Grandparent | 453 | 516.60 |

(minimum tour length) and also generates the lowest mean result. As grandparent repair produced the lowest average and standard deviation this suggests that the grandparent is the superior repair template. Repeated trials have shown that the best overall result generally originates from the grandparent repair strategy (Mann-Whitney $P=0.1423$). These findings combined seem to suggest that the grandparent strategy introduces greater diversity into the population and that this diversity can yield the best result across repeated trials. These results support the controversial findings of Lolle *et al* [1].

5.3 A Comparison of Template Fitness

While the most plausible interpretation of genetic repair in *A. thaliana* involves recording randomly selected parent or grandparent information, we next explore the impact that fitness might have on the results produced. The next results compare the performance of storing the fittest parent or grandparent against a randomly selected template of the appropriate ancestry.

The results detailed in Table 3 show that using a randomly selected template produces better results than selecting the fittest of the available templates (MANN-WHITNEY $P=0.1894$) The results detailed in Table 3 show that using a randomly selected template produces better results than selecting the fittest of the available templates (Mann-Whitney $P=0.0764$). This was a relatively surprising result, as it was felt that the fitter template would produce fitter results. However, this result also seems to lend support to Lolle’s theory as the random selection process appears to produce the best results (and no mechanism for favouring the fittest ancestor would appear to be naturally available).

Table 3. Template Fitness

| Template | Min | Mean | Standard Deviation |
|---------------------|-----|--------|--------------------|
| Fittest Parent | 469 | 524.5 | 34.72 |
| Random Parent | 463 | 516.15 | 27.87 |
| Fittest Grandparent | 458 | 519.62 | 33.72 |
| Random Grandparent | 455 | 517.04 | 29.75 |

5.4 The Effect of Repair Direction

In this section we explore the impact of varying the direction in which GeneRepair operates. GeneRepair has two stages; the first **error detection** phase is followed by a distinct **error correction** phase. Errors are the elements of the individual that violate the problem constraints. When multiple errors are detected, the order in which errors are corrected can impact on the resultant individual. In this paper, the order in which the errors detection operates is dictated by the repair template.

Error correction is then carried out in one of three different directional strategies: randomly changing direction, fixed Left-to-Right and fixed Right-to-Left. Duplicates are identified using a fixed replacement strategy. Error detection always begins at one end of the genome and proceeds along until a genetic error is detected. For the TSP, an error is detected when a duplicate city is found. This erroneous genetic item is then dispatched to the error correction phase. Error correction examines the ancestral genomic data to identify the missing data, using a fixed replacement strategy. Thus, the repaired information originates in the ancestral genome, as proped in *A. thaliana*.

We compare these three repair directions to further investigate the use of GeneRepair. Table 4 outlines the results created when a random grandparent template was used as a repair template and repair was carried out in the three directions listed. These results show that repair acting in a randomly changing direction produces the best minimum result and this is upheld by the fact that it also produces the lowest mean. This can be explained by the fact that repairing in a random direction increases the diversity in a solution as opposed to repairing in a fixed direction.

These early results in Table 4 also show that repairing in a right to left direction outperforms repairing in a left to right direction. This suggests that our fixed replacement strategy favours right to left repair and opens the door to further experiments where the error correction phase is also driven by the ancestral template. First, the random-varying direction was compared to the fixed Left-to-Right detection strategy ($P = 0.1762$) indicating that the random direction reliably produced better results. Similarly, a comparison between the random and fixed Right-to-Left directions $P = 0.0384$, again indicating that the random-varying direction produced the better results. These results are compounded further when we compare all four experimental sets, that is fittest and random parent and grandparent. We see that overall, repair carried out in a

randomly changing direction produces the minimum lowest result and this is supported by the fact that it also has the lowest mean (Table 5).

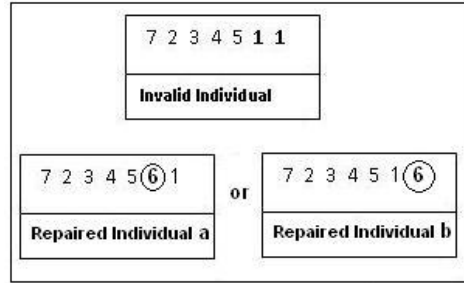


Fig. 3. Effect of Different Repair Directions

Table 4. Repair Direction - Random Grandparent and Random Great-grandparent

| Template | Direction | Min | Mean | Standard Deviation |
|--------------------|---------------|-----|--------|--------------------|
| Random Grandparent | Left to Right | 483 | 514.15 | 22.95 |
| Random Grandparent | Right to Left | 464 | 525.12 | 35.82 |
| Random Grandparent | Random | 453 | 505.54 | 26.37 |

Table 5. Repair Direction - Parent and Grandparent

| Direction | Min | Mean |
|---------------|-----|--------|
| Left to Right | 473 | 519.32 |
| Right to Left | 458 | 520.96 |
| Random | 453 | 515.30 |

5.5 Great-grandparent Repair Template

We expanded our investigation to look at whether extending the ancestry of our repair templates produces better results. We ran experiments to look at the results produced the great-grandparent template. For this experiment we used the superior algorithmic parameters found in the results above, that is random inheritance acting in a random direction.

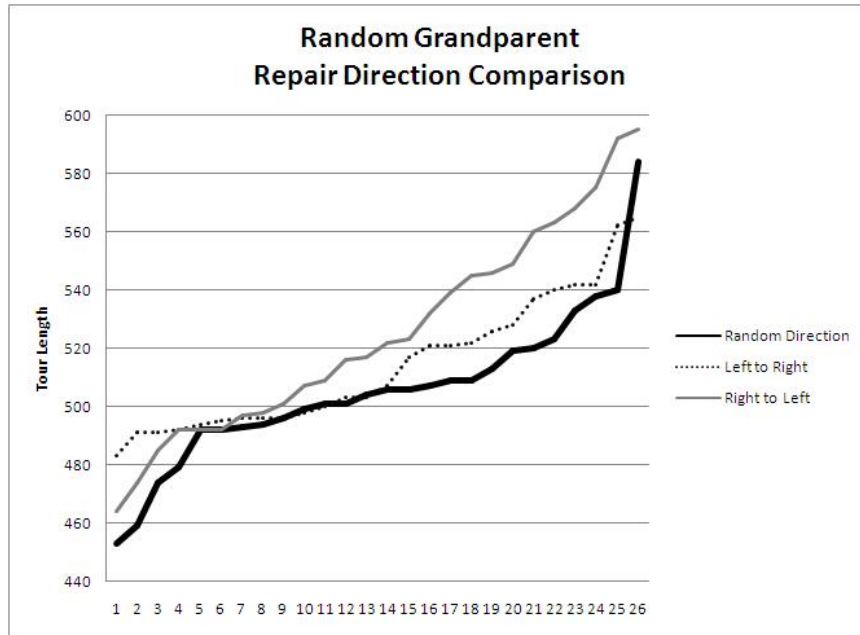


Fig. 4. Grandparent Repair Direction Comparison

Table 6. Random Grandparent and Random Great-grandparent

| Template | Direction | Min | Mean |
|--------------------------|-----------|-----|--------|
| Random Parent | random | 470 | 514.76 |
| Random Grandparent | random | 453 | 509.96 |
| Random Great-grandparent | random | 461 | 518.33 |

The minimum result produced by the great-grandparent template is lower than that produced by the parent template but the mean of the great-grandparent results is highest. The grandparent appears to produce results superior to that of the great-grandparent which leaves the door open to further investigation into the benefit of retreating further in generations versus the cost on memory. This result seems to suggest that grandparent repair introduces just the right amount of diversity, while great-grandparent produces too much diversity in the repaired populations(at least for the algorithmic parameters discussed above). This result may suggest that *A thaliana* (or other organisms) may be unlikely to access ancestral information that extends back more than two generations.

6 Conclusion

Recent advances in the study of the model organism *Arabidopsis thaliana* suggest that this plant uses a novel strategy for repairing errors to its own DNA. Lolle *et al* [1] controversially suggest that *A. thaliana* uses genetic information from the grandparent generation to repair errors that occur in the grandchild generation.

In this paper we present a new (and largely domain independent) strategy for enforcing constraints on the search space of an Evolutionary Optimisation process. The new evolutionary optimisation is inspired by the phenomenon of non-Mendelian inheritance contentiously witnessed in nature [1] by using stored genetic information from previous generations. This is a broadly applicable strategy, unlike many approaches to enforcing constraints on evolutionary searches. This strategy repairs genetic errors using ancestral data to repair constraint violations, so that all problem constraints are obeyed. We explored a total of twelve different methods for GeneRepair in the experiments outlined above.

We began by comparing grandparent based genetic repair to the death penalty system to enforce constraints. The grandparent GeneRepair showed itself to produce significantly better results than the death penalty. In fact the result produced by the death penalty was almost four times the size of the result produced using the grandparent template GeneRepair. We went on to compare the relative performance of the following repair templates: parent, grandparent and great-grandparent. The results show that parent is the weaker template, which is in support of Lolle *et al* [1].

Next, we explored the influence of fitness on the results produced, based on the intuition that storing the fittest parent or grandparent would produce better results. Results showed that using a randomly selected template constantly outperforms selecting the fittest template from the appropriate ancestry. This surprising result seems to suggest that increasing the diversity of the population might play a role in grandparent based repair. We feel that this results again provides indirect support for Lolle *et al*.

Up to this point experiments had been carried out using repair in a constant left to right direction, so we investigated other possible directions of repair. The

experiments showed that repairing in a randomly changing direction produced superior results to repairing from right to left or from left to right.

In summary, our results show that it is possible to adapt an evolutionary strategy to handle constraints in a biologically plausible manner, by amending the algorithmic parameters to repair mechanisms found in biology. This paper shows that using a grandparent or great-grandparent repair template to enforce constraints can outperform the use of a parent template (for our selected problem domain). This supports the findings of Lolle *et al* that non-Mendelian inheritance of genomic information can lead to stronger individuals and populations.

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. Peng, Chan, Shah, Jacobsen: Increased outcrossing in hothead mutants. *Nature* **443**(E8) (2006) 28
3. Ray, A.: Plant genetics: Rna cache or genome trash? *Nature* **437**(E1-E2) (2005)
4. Colorni, Dorigo, Maniezzo: Genetic algorithms and highly constrained problems : The time-table case. Proceedings of the First International Workshop on Parallel Problem Solving from Nature, Dortmund, Germany, Lecture Notes in Computer Science 496 (1990)
5. 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
6. Kalanmoy, D.: An efficient constraint handling method for genetic algorithms. *Computer Methods in Applied Mechanics and Engineering* **2-4**(186) (2000) 311–338
7. Mitchell, G., O'Donoghue, D.: Generepair-repair operator for genetic algorithms. GECCO Conference Illinois (2003)
8. FitzGerald, A., O'Donoghue, D.P.: Genetic repair for optimization under constraints inspired by Arabidopsis thaliana. *Parallel Problem Solving From Nature* (2008) 399–408
9. Arroyo, C.: A parallel repair genetic algorithm to solve the unit commitment problem. *IEEE Transactions on Power Systems* **17**(4) (2002)
10. Mitchell, G.: Evolutionary computation applied to combinatorial optimisation problems. Dublin City University (2007)
11. Nakano, R., Yamado, T.: Conventional genetic algorithm for job shop problems. *Proc. 4th International Conference on Genetic Algorithms* (1991) 474–479
12. Michalewicz, Z., Fogel, D.B.: How to solve it: modern heuristics. 1 edn. Springer-Verlag (2000)
13. Frugoli, J.: *Medicago truncatula handbook*. Clemson University (November 2006)
14. Coghlan, A.: Rogue weeds defy rules of genetics. *New Scientist* **2492** (26 March 2005) 8
15. Weigel, Jurgens: Hothead healer. *Nature* **434** (2005) 443
16. Matsumoto, M., Nishimura, T.: Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation* **8 No. 1** (1998) 3–30
17. Reinalt, G.: Tsplib - a travelling salesman problem library. *OSRA Journal of Computing* **3** (1991) 376 – 384