Extra-Intracellular Transgenetic Algorithm applied to the Graph Coloring Problem

*Marco César Goldbarg** *Elizabeth Gouvêa** *Lívia M. M. Silva**

* Umiversidade Federal do Rio Grande do Norte
Campus Universitário Lagoa Nova - Natal - RN
Email: {gold, bgouvea, livia}@ufrnet.br

1 Introduction

This paper presents the Extra-Intracellular Transgenetic Algorithm, *EITA*, an evolutionary algorithm based on the Computational Transgenetics metaphor. Computational Transgenetics aims at including information into an evolutionary process. The paradigms of Computational Transgenetics are briefly introduced. EITA is applied to the Graph Coloring Problem.

Darwinian evolution has been the most usual metaphor to design evolutionary algorithms. Nevertheless, some researchers criticize the imperfection of the darwinian metaphor in solving several microbiological phenomena and in modelling evolution of species[1]. Genetic Algorithms, GA, introduced by Holland[12], are based on the darwinian paradigm and are successful in finding approximate solutions for many NP-hard problems. However, there are also many cases for which GAs fail. Genetic Algorithms use operators which imitate sexual reproduction to share genes between individuals, generating offspring. Since it implies on interactions between sets of cells, the sexual reproduction is said to be extracellular [4]. In fact, the extracellular paradigm is the main mechanism used by GAs to explore the search space. Though extracellular interactions modify DNA, definitive changes in the DNA may also result from intracellular interactions and from epigenetic rules. Recent researches sho! w ! that genes and culture are inherently linked [14]. Individuals evolve both by anatomical and behavioral selection. Rules that cause anatomical and behavioral elements to come together are called *epigenetic*. Genes prescribe epigenetic rules, which are the regularities of sensorial perception and mental development that motivate and direct culture acquisition. Culture helps to determine which prescriptive genes survive and multiply from one generation to the next. New well succeeded genes modify the epigenetic rules of the population. The modified epigenetic rules alter the direction and efficiency of the cultural acquisition channels [22]. A meme is an element of a culture that may be considered to be passed on by non-genetic means [16]. Thus, memes are elements of cultural concepts. Richard Dawkins used the term *meme* to refers to cultural transmission unities in analogy to the role of genes in biological evolution [3]. A meme and a gene have both the dual capacity: carriers of information and creators of behavior. Considering memes and epigenetic rules in evolutionary computation is still a challenge [2]. In the computational context memes were initially associated to exogenous information - information out of the extracellular process - such as local search in GAs [17]. One of the main contributions of the transgenetic paradigm is that it has means to explore both environmental and cultural dimensions of the evolutionary process. In this paper, a meme is a piece of information, with origins in a specific context, used to rearrange a set or block of genes. To express epigenetic rules in transgenetic algorithms, Computational Transgenetics uses intracellular flow agents called transgenetic agents.

In order to build on proteins, DNA uses a mechanism distinct to the one used to create new individuals. Protein folding depends on the interactions that occur into the intracellular context, involving several agents. Interactions in the intracellular context can affect the reproductive cells and be inherited by offspring. One of the main characteristics of the intracellular flow is that it represents an intense and significative way to carry on physical-chemical interferences on the DNA. Microbiological processes are carried on by means of the intracellular flow. Several adaptation and evolution tasks correspond exclusively to the intracellular context [18]. Genetic Engineering, GE, studies the intracellular mechanisms for genetic interactions and their impact over the organisms. By means of GE tools [19], it is possible to remove DNA segments from an organism and insert them in other or in the same organism. Genetic researchers have investigated gene structure for many years and found out some means to carry on genetic manipulations. To insert a gene into the nucleus of a cell, the genetic engineers use a vector. The vectors, among others, may be viruses, bacterias or mobile genetic particles. The transgenetic agents are the computational tools that play the role of such vectors. The proposals to rearrange genes are loaded on the transgenetic agents that manipulate the chromosomes of a given population. The intracellular flow is the operational medium of Genetic Engineering. It is multiagent and associative.

Computational Transgenetic algorithms are designed to consider the intracellular and epigenetic paradigms into an evolutionary process. The next section presents CT and the transgenetic agent Virus. Computational results of applying EITA to the Graph Coloring Problem are reported on Section 3. Finally, Section 4 presents some conclusions.

2 Computational Transgenetics

Computational Transgenetics (CT) [8] brings the following ideas to the evolutionary context: To use exogenous and endogenous information to interfere on the processes of formation and modification of individuals of a given population; to use the intracellular flow as an operational way to carry the required manipulations on the individuals; to explore new processes of population improvement using transgenetic agents and competition between agents and individuals; to guide the evolutionary process allowing the occurrence of evolutionary jumps. The transgenetic agent used to express the memes into the computational context, in this paper, is the *virus* [11]. A meme is a piece of information, such as heuristic information, previous knowledge about the problem or information originated within the evolutionary process itself - immune process. Some memes are originated from epigenetic rules and compose a database or, simply, a meme base. A meme is loaded on a transgenetic agent which encodes it as a DNA segment - originating a building block - or as a genetic manipulation procedure. The transgenetic agents manipulate the chromosomes of a given population. The population evolves and may return information to generate new memes wich will form new agents. The evolution of chromosomes and agents characterizes a process of geneXmeme coevolution. Thus, it can be said that CT is a coevolutionary informed search technique where genes and memes are the players. CT is an approach that makes possible to build evolution through an association of individuals and agents - simbiotic assimilation and coevolution - and may, inclusively, make no use of sexual reproduction, consequently making no use of classical genetic operators.

A transgenetic agent is composed by one or more memes and an operation method. A virus is a transgenetic agent which has the ability to manipulate a chromosome, consequently altering its fitness, and to keep the information into the chromosome for a certain period of time, called *virus lifetime*. Let P be a population of size N, P = $\{ S_1, S_2, ..., S_N \}$, and f a fitness function. Each chromosome S_i, i=1,.., N is an integer-valued string of length n representing a problem solution, such that f: $S_i \rightarrow \mathcal{R}^+$, i = 1,...,N, where f returns a real value. A virus, from the computational transgenetics point of view, is an information string, an *Infection* and a *Desinfection* method. The virus infection consists of three stages: attack, transcription and lifetime.

Attack (A) - This is a decision stage where a value *true* or *false* is assigned to a chromosome to establish whether the chromosome is sensitive to the manipulation induced by a given virus. That is,$A : S_i \rightarrow$ {true, false}, i = 1,...,N. The sensitiveness of a chromosome depends on the virus and on its own features. If the virus manipulation improves the fitness of a chromosome, say S_i, then A(S_i)

= "true". Otherwise, the virus manipulation decreases S_i fitness. In this case the value of A(S_i) is function of a pre-defined immune scheme.

Transcription (Γ) - If A(S_j) = "true", then genetic information transported by the virus will be transferred to S_j, according to a specific Transcription Operator (Γ_i) which belongs to the virus transcription operators set. S'_j is the resulting chromosome. Chromosome manipulation by the action of a virus is a transcription where: $\Gamma = (\Gamma_1, \Gamma_2, ..., \Gamma_m), \Gamma_i : S_j \rightarrow S'_j$, $j = 1,...,N$.

Lifetime (Ψ) - Ψ is a function that gives the number of generations to which the transcribed information must be propagated to offspring.

There are several types of viruses [8]. The type of virus used in the implementations of this work is the *beagle virus*. A beagle virus is composed by a meme encoded as a genetic manipulation procedure. Considering the Graph Coloring Problem, by definition, the colors assigned to adjacent vertices must be distinct. Thus, the colors assigned to the vertices of a clique must be all distinct. An epigenetic rule derived from this result may be "assign distinct colors to the vertices of a clique". In this case, a meme is a clique plus a statement to color its vertices with distinct colors.

CT provides two main algorithmic compositions: Extra-Intracellular Transgenetic Algorithms (EITAs)[9] and Proto-Gene Algorithm (ProtoG) [11]. This work presents the former. The EITA algorithm is based on the three paradigms previously considered: extracellular, intracellular and epigenetic. In general, EITAs can be described as below. The arrows ($\rightarrow$) mark the statements where intracellular manipulations occur, the puzzle signs (#) mark the epigenetic steps and the asterisks (*) mark the steps of the underlying genetic algorithm.

Extra-Intracellular Transgenetic Algorithm (EITA)

```
Begin
    # Load a Meme Base
    # Design a set of Viruses according to the Meme Base
    (*) Generate and evaluate an initial population
    Repeat
        Begin
            → Load a subset of Viruses
            (*) Select a set of parents to generate offspring
            → Manipulate sensitive parents according to epigenetic rules
            (*) Crossover
            (*) Mutation
            → Liberate chromosomes that can be set free (end of Virus lifetime)
            (*) Evaluate population fitness
        End
    Until the stop criterion be satisfied
End
```

3 Computational Experiments

In order to check the transgenetic potential, the approach was applied to the to the Graph Coloring Problem (GCP). The objective of the computational experiment was to test whether the transgenetic planning is able to accelerate a classical genetic algorithm.

Coloring the vertices of a graph $G = (V,E)$ such that no adjacent vertices receive the same color is a known NP-complete problem [7]. Some evolutionary approaches were presented for the GCP by Fleurent and Ferland [6], Eiben et al. [5] and Lorena [13]. This work presents the results of an EITA implementation which used the algorithm presented in [13] as its underlying GA. The implemented version used a beagle virus as its transgenetic agent. Memes loaded on the viruses are cliques plus

Instance	Vertices	Edges	Colors	Virus
Anna	138	493	11	10
David	87	406	11	10
Huck	74	301	11	11
Jean	80	254	10	21
Games120	120	638	9	9
Miles250	128	387	8	15
Miles500	128	1170	20	19
Queen55	25	160	5	6

Table 1: Experiment data

Instance	Average EITA	Average GA	Median EITA	Median GA
Anna	0.53	2.07	0.0	2.0
David	0.3	1.37	0.0	2.0
Huck	0.03	0.37	0.0	0.0
Jean	0.47	0.87	0.0	1.0
Games120	0.27	0.57	0.0	0.5
Miles250	0.5	2.8	0.0	3.0
Miles500	2.5	54.8	2.5	55.0
Queen55	6.38	6.82	6.0	7.0

Table 2: E_c comparison

a statement to color the vertices with distinct colors. The cliques were generated by the algorithm No-Choice presented in [21]. Chromosome fitness is computed as a function of the number of edges with both end vertices colored with the same color, E_c. Thus, the objective is minimize E_c. If a transcription did not improve chromosome fitness (increase E_c), the infection chance was computed by the following expression: $P = E_c\beta + E_{cc}$, where E_c and E_{cc} denote chromosome fitness before and after the virus infection, respectively and β is a control parameter, $\beta > 0$. Virus lifetime was set to three iterations. Instances were obtained from Trick [20]. Table 1 summarizes instances data and the number of viruses used on the experiments. Table 2 shows the average and the median for E_c values on 30 runs of both algorithms. EITA found the best solution for all instances. The GA failed to find the best solution for Miles500. The final E_c value found by the GA for Miles500 was 42. The experimental data summarized on Table 2 show that EITA obtained the best average and median results for all instances.

4 Conclusion and Remarks

Computational experiments showed that the chromosome manipulation by means of transgenetic agents was a powerful tool to guide the search in the solution space for the considered instances. EITA obtained the best average and median results when compared to the Genetic Algorithm.

The transgenetic approach was successfully applied to the Quadratic Assignment Problem [11] and to the Piston Pump Mobile Unit Tour [10, 15], a real world problem of the petroleum industry.The Piston Pump Mobile Unit Tour is a special case of the vehicle routing problem.

During the computational experiments it was noticed that EITA is sensitive to information quality. That is, the better the information the faster the algorithm converges to the optimum. Transgenetic metaphor enlarges the evolutionary algorithms context, yielding to consider evolutionary jumps that complement the gradual character of Darwin's approach.

5 Acknowledgments

This research was funded by FINEP/CTPetro.

6 References

[1] Behe, M., 1996, *Darwin's Black Box: The Biochemical to Evolution*, The Free Press, New York.

[2] Chattoe, E., 1998, Just How (Un)realistic are Evolutionary Algorithms as Representations of Social Processes? *Journal of Artificial Societies and Social Simulation* vol. 1, no. 3. http://www.soc.surrey.ac.uk/JASSS/1/3/2.html

[3] Dawkins, R., 1976, *The Selfish Gene*, Oxford University Press.

[4] Dudley, W, 1990, *Genetic Engineering*, San Diego: Greenhave Press, Inc.

[5] Eiben, A. E., Hauw, J. K. and J. I. Hemert, 1998, Graph Coloring with Adaptive Evolutionary Algorithms, to appear in J*Journal of Heuristics* 4:1.

[6] Fleurent, C., and J. A. Ferland, 1996, Genetic and Hybrid Algorithms for Graph Coloring, In Laporte, G., Hammer, P. L. (Eds), *Annals of Operations Research* 63, 437-461.

[7] Garey, M. R. and D. S. Johnson, 1978, *Computers and Intractability: a Guide to the Theory of NP-Completeness*, San Francisco, W. H. Freemann.

[8] Goldbarg, M. C., and E. F. Gouvêa, 2000, Computational Transgenetics, *X CLAIO*, September, Mexico City, Mexico.

[9] Goldbarg, M. C., and E. F. Gouvêa, 2001, Extra-Intracellular Transgenetic Algorithm, In *Proceedings of GECCO 2001 Genetic and Evolutionary Computation Conference*, July, San Francisco, CA, USA.

[10] Goldbarg, M. C., Gouvêa, E. F., and A. C. C. Medeiros, 2001, A Transgenetic Algorithm Applied to the Piston Pump Mobile Unity Tour Problem, *EURO 2001 - The European Operational Research Conference*, July 9-11, Rotterdam.

[11] Gouvêa, E., 2001, *Computational Transgenetics, an Algorithmic Investigation*, D. Sc. Thesis, Program of Production Engineering, COPPE/UFRJ. In Portuguese.

[12] Holland, J. H., 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.

[13] Lorena, L. A. N., and Filho, G. R., 1998, *Constructive Genetic Algorithm for Graph Coloring*, Technical Report, LAC/INPE, Brazil.

[14] Lumsden, C. J., and E. O. Wilson, 1981, *Genes, Mind and Culture*, Harvard University Press, Cambridge.

[15] Medeiros, A. C. C., 2001, *Transgenetic Algorithms applied to the Piston Pump Mobile Unity Tour Problem*, M. Sc. Monograph, Universidade Federal do Rio Grande do Norte. In Portuguese.

[16] Plotkin, H. C., 1996, Non-genetic transmission of information: candidate cognitive processes and the evolution of culture. *Behavioural Processes*, 35, 207-213.

[17] Radcliffe, N. J., and P. D. Surry, 1994, *Formal Memetic Algorithms*, Edinburgh Parallel Computing Centre, King's Buildings, University of Edinburgh.

[18] Selander, R. K., Clark, A. G., and T. S. Whittam, 1991, *Evolution at the Molecular Level*, Sinauer

Associates, Sunderland, MA.

[19] Steinbrecher R., 1999, What is Genetic Engineering ?, *Synthesis/Regeneration: A Magazine of Green Social Thought*, Vol 18, 9-12.

[20] Trick, M. A., 1996, *Clique and Coloring Problems, a Brief Introduction with Project Ideas*, http://mat.gsia.cmu.edu/COLOR/general/ccreview/ccreview.html.

[21] Turner, J. S., 1988, Almost all k-Colorable Graphs are Easy to Color, *Journal of Algorithms* 9, 63-82.

[22] Wilson, E. O., 1998, *Consilience*, A. A. Knopfs (Ed.), New York, USA.