

New Algorithm Transgenic with Homology to solve the problem of OneMax

Christian José Correa Villalón¹, Felipe Padilla Díaz¹, Eugenio Pérez Morphe¹,
Fátima Sayuri Quezada Aguilera¹, Ochoa Ortiz-Zezzatti.

¹Universidad Autónoma de Aguascalientes (Departamento de Inteligencia Artificial); UAA
crisjc_2000@yahoo.com.mx, sayu_172002@yahoo.com.mx, megamax8@hotmail.com

Abstract. Already proposed a new vision, a heuristic based on a modern branch of biology called Biotechnology. "This is any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use"[3]. For individuals who have undergone any change in their genes through biotechnology techniques are known as **Transgenic**, they can be animals or plants. The changes applied to these individuals are made for a specific purpose; usually to improve an individual has highlighted some of its own characteristics [4]. Now incorporates new elements to the first algorithm for transgenic improvement. Homology has been incorporated in the search for the fittest individuals. The Homology has been incorporated in the search for the fittest individuals. With the use of positive and negative homology.

Keywords: Transgenic Organisms, Biotechnology, New Algorithms Transgenic, GMO, Directed Crossing, Homology, Anti sense, Resolution

Nuevo Algoritmo Transgénico con Homología, para resolver el problema del OneMax.

Resumen. Ya se propuso una nueva visión, una heurística basada en una rama de la biología moderna llamada Biotecnología. "Es cualquier aplicación tecnológica que usa sistemas biológicos, organismos vivos, o derivados de estos, para hacer o modificar productos o procesos para un uso específico" [3]. A los individuos que han sufrido alguna alteración en sus genes a través de técnicas de biotecnología reciben el nombre de **Transgénicos**, estos pueden ser animales o plantas. Los cambios aplicados a dichos individuos son hechos persiguiendo un fin o propósito específico, generalmente *para mejorar ha dicho individuo o resaltar algunas de sus propias características* [4]. Ahora se incorporan nuevos elementos al primer algoritmo transgénico. La homología se ha incorporado en la búsqueda de los individuos más aptos. Con el uso de la homología positivos y negativos.

Palabras clave: Organismos Transgénicos, Biotecnología, Nuevo Algoritmo Transgénico, OMG, Cruza Dirigida, Homología, Anti sentido, Resolución.

1 Introduction

The inspiration for what is proposed is a heuristic-based biotechnology. What is biotechnology, there are several definitions but we will stay with pronounced Convention on Biological Diversity, organized by the United Nations: "It is any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for a specific use "[8].

To better understand the concept of biotechnology we go back to their history and some of its uses more knowledge, the advantages and disadvantages, as well as their implementation, as seen from a biological point of view. For example, the **genetic manipulation of plants** for the benefit of humans is part of the Biotechnology [9]. Individuals who have had some changes in their genes through biotechnology techniques are known as *transgenic*, they can be animals or plants.

1.2 New Algorithm Transgenic

For the first transgenic algorithm, the basis is the process used in genetic improvement of plants, where the process was really simple. To continue the study of transgenic techniques is found that there are other factors involved in the production of transgenic individuals who appear to be analyzed has a significant impact on the actual technical development, considering these factors, we create an algorithm Transgenic improved.

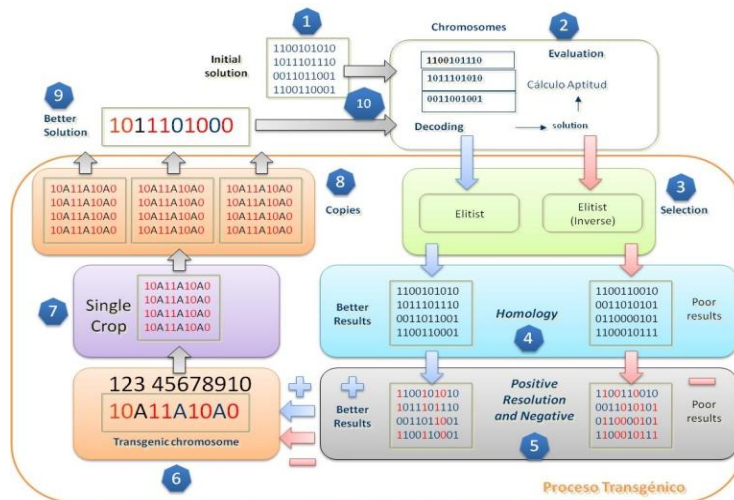


Fig. 1. New Algorithm Transgenic

This new algorithm incorporates new features such as **Alignment, Homology, Order positive and negative**. Explaining each process and see the difference with respect to the first algorithm created. The points numbered in blue indicate the steps that must be followed to understand the diagram.

1. Create an initial population. We get the satisfaction of each of the individuals according to the objective function. Select a percentage of the fittest individuals (who best meets the objective function) and select a percentage of individuals less able "Reverse Elitism".
 2. Homology is applied to detect patterns in the genes of each individual.
 3. When patterns are found in genes of the fittest individuals, such as those that are not, this is known as negative and positive resolution.
 4. We built an individual, respecting the genes of the resolution positive and negative, to produce a chromosome with good genes, but excluding the genes that could generate good individuals, which is not positive or negative sense genes are created at random.
 5. Generates the seed of a new population with the specifications and Resolution Passive negative, always remembering that each generation is unique and does not necessarily have to relate with each other.
 6. As each generation is unique, repeated planting a small number of times, not to lose the properties found in the resolution positive and negative.
 7. It takes the best solution found between the results of each planting to ensure a leap in evolution, so it requires fewer generations to find a good result.
- The process is repeated until a stop function or the algorithm converges.

2 Related Work

This section is an implementation of transgenic algorithm to the problem of OneMax to show the operation thereof, will explain each item as shown in Fig 2. The parameters that were used to solve the problem were three harvests per generation and with a positive and negative homology of 25% of the population size and with a hardness of 70% for individuals who meet the objective function.

Creating the initial population. The first step is to create a completely random initial population; this process is the same as in the old transgenic algorithm that is also similar to other evolutionary techniques like genetic algorithms.

Assessment and decoding. The next step is to evaluate the initial population were obtained from each individual level of satisfaction of the objective function. Decoding is the interpretation given to the data contained in each individual, such as in this case we are testing the maximum amount of 1 that are contained in each chromosome. Until this point the operation is almost similar to other techniques, from third to observe the two most important changes and define the new algorithm transgenic.

Selection. As is known, there are many types of selection for each of the heuristics and each use that is most convenient for the transgenic algorithm, using a totally elitist selection. The purpose of selection is to identify the best qualified individuals, if we were talking about a genetic algorithm these individuals would be selected for a crossing and generate new individuals to preserve the genes of the fittest parents in transgenic algorithm, the elitist selection plays the same role, it identifies the fittest individuals of the population, the difference in the new transgenic algorithm is also used a selection that we call **reverse selection**, the elitist selection to identify the best qualified individuals, the reverse elitist selection , seeks to identify individuals least

able in this way not only to identify individuals who meet the objective function in a more accurate, we also identified individuals who do not comply.

Homology. On the epistemological "**homology**" is a correspondence between two or more traits of organisms. It is not a directly observable, but the homology is detected as a result of an inferential method of analysis. This theoretical framework allows recognizing the assumptions of homology and as the epistemological basis for the comparative study of the similarities between organisms [15]. This method is used in the creation of transgenic organisms to identify patterns of genes contained in the body that will receive the new gene in the body that is going to donate, the homology with transgenic processes can identify genes that can be modified and which should not be modified. This led to computing the homology is used to detect patterns in the different chromosomes that make up the population, these chromosomes can be positive or negative. To perform the homology is required to apply a technique called alignment, a sequence alignment in bioinformatics, is a way to represent and compare two or more sequences or strings of DNA, RNA, or protein primary structures to highlight their areas of similarity that could indicate functional or evolutionary relationships between genes or proteins consulted. The aligned sequences are written with the letters (representing amino acids or nucleotides) in rows of a matrix in which, if necessary, insert spaces for areas with similar or identical structure to align [15]. The homology is used in Transgenic algorithm to find similar patterns that indicate the existence of good results.

Resolution positive and negative. The alignment will enable us to see if similar patterns exist in the chromosomes, also remember that it had the satisfaction of each of the individuals in the objective function, such as individuals who met at least the objective function, to obtain positive patterns (which we want to preserve) and individuals least able to obtain a negative pattern (which we do not want to preserve).

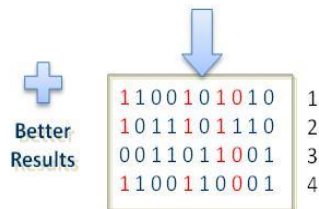


Fig. 2. Alignment of only the best individuals.

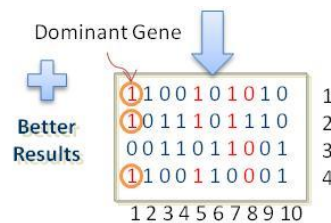


Fig. 3. Positive dominant gene.

In Fig. 2 shows an alignment of the best individuals, seeking patterns in each column, for this example shows that in column 1 there is a dominant gene that is the number 1, so therefore we assume that this is the gene prevail in the next generation. In column 2 shows that there is no dominance by any chromosome, so in this position in the next generation will have a random value, this process is repeated for each column.

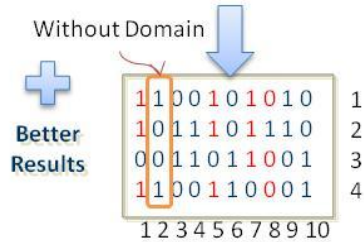


Fig. 4. Column without a dominant gene.

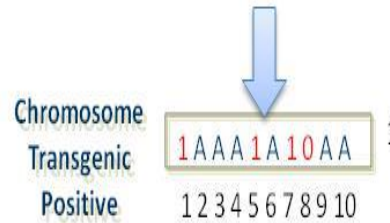


Fig. 5. Results Positive Alignment

At the end of the alignment of individual genes, we obtain a resultant vector, with which we can build the next harvest. In Fig. 5 shows the result of the positive line, the result is a chromosome which starts at position 1 to 10 and also mark that position 1, 5, 7, the dominant gene must be 1, and the position 8 is placed a zero, in other places in the absence of dominant patterns were chosen by generating random again. The process of negative alignment is performed in the same way, and is known as antisense, an antisense mRNA is an mRNA transcript that is complementary to an endogenous mRNA, in other words, the complementary strand (non-coding) to a sequence of mRNA (coding). The introduction of a transgenic coding for antisense mRNA is a technique used to block the expression of a gene of interest. An antisense mRNA of radioactively tagging can be used to show the level of gene transcription in several cell types [16]. The alignment process is similar to positive, thus only shows the resulting vector of Antinsense.

It is important to mention that until this point the Transgenic algorithm can work with each of the separate techniques, but also can work combining the two techniques, this process is called **resolution positive and negative**. In resolution positive and negative, combine the results of the two resulting vectors to obtain a single vector, this vector will tell us how to build the next harvest. As mentioned before the resulting chromosome will emerge to confront the **positive alignment, and antisense**.

The way in which the vector is obtained by comparing the resulting vector k positive, with the vector k ' negative, remember that the positive vector has the gene that we want to retain the negative genes that do not want to keep. The conditions for generating the resultant vector R are:

```

For j = 1 to n = size of chromosome
{
  If kj = k'j Then
    Rj = Random
  Else
    Rj is equal to kj is but that is different from
k'j
}

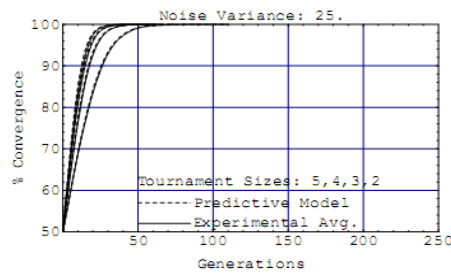
```

The resulting chromosome will create the basis for seeding for the new population.

Repetition. Using the same principle to generate a transgenic population, generate copies of the resulting population to observe the same behavior when creating replicas obtained the degree of satisfaction of the objective function of each, and takes the best cases for each of the replicas, also cases in which the objective function was not performed very well. As shown in Fig. 1 are generated replicas respecting sowing only, where the positions are already defined are not modified, but which indicates that the number that must be generated is random. Finally the process is repeated until a stop condition or that all the plantings.

3 Results

Shows a comparison of a genetic algorithm, and a Transgenic algorithm, the results show that the genetic algorithm is a publication of Brad L. Miller and David E. Goldberg [17], which shows the impact of tournament selection and the noise is generated, these tests were implemented to solve the problem of Onemax, a bit size of 100's. The algorithm was tested with the same size Transgenic Bit's, as demonstrated, is the number of generation which generates convergence.



(b) Variance: 25

Fig. 19. Onemax experiments with different tournament sizes and noise levels

In the graphic we can see different results for different values of the selection tournament for different noise levels of a genetic algorithm, this experiment it is of interest is the number of generation where the algorithm converges, the results observed in the graph are averages of several runs.

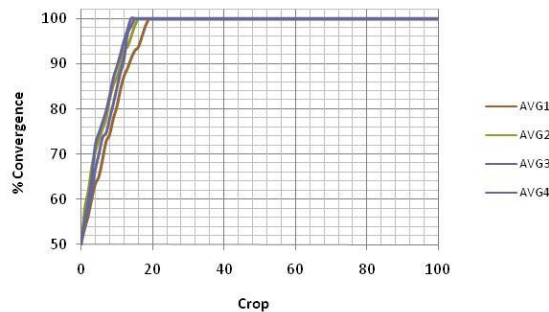


Fig. 20. Algorithm Transgenic experiments, with averages.

In Figure 20 show the averages for each crop and shows that the convergence begins harvesting 14 for the four replications, surpassing the Genetic Algorithm, the Genetic convergence starts from the generation 20, to remember that cases are 100 bits Onemax.

3.1 Transgenic Algorithm Vs Interactive Genetic Algorithm

Another comparison of the performance of the AT (Transgenic Algorithm), emerges with the publication of Xavier Llorà, Kumara Sastry, David E. Goldberg, Abhimanyu Gupta, Lalitha Lakshmi [18]. Which compares the performance of an Interactive Genetic Algorithm with a Genetic algorithm, also applied to the problem of different sizes Onemax chromosome, (4, 8, 12, 16, 20, 24, 28, 32). The GM algorithm was also tested with these sizes chromosome.

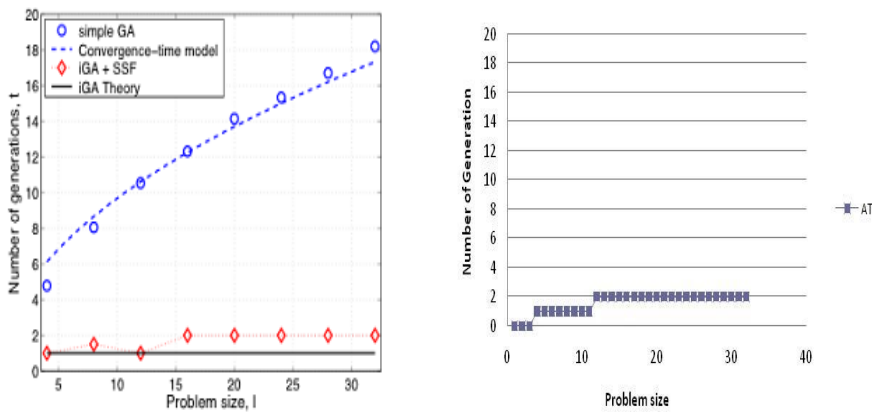


Fig. 21. Comparison of Genetic, Genetic iterative and Transgenic (AT) for OneMax

In the graph shows that there is a large difference of conventional AG and an iterative AG; almost AGI, requires only 2 generations to solve the instance where the chromosome is equal to 32. Note that the Transgenic can be compared with a virtually AGI and compared with genetic, is above the results of these.

5 Conclusions

An important point of departure for any heuristic is always compared to genetic algorithms as the basis for many of the existing heuristics, and should always be one of the heuristics to overcome. It can be concluded that with the new changes that have suffered the GM algorithm, its performance has raised, the comparisons that have been shown to be enhanced rather than a genetic algorithm. The new mechanisms were added, are an important part in the creation of transgenic organisms, the homology, positive and negative resolution, and consider each as a single crop.

4 Future Work.

Desired, as it happened with the first transgenic algorithm solving the traveling salesman problem, since the initial design was to solve the problem. Where there is an adjustment to the TSP problem, which aims to use the algorithm Transgenic, a BechMark where we know the optimal route for the case of TSP. In recent years, the colony of ants has shown a great acceptance for their performance in various applications, it will be one of the contenders for the AT (Transgenic Algorithm). Lin-Kernighan[19] is an algorithm that has proven its great functionality to the problem of TSP with excellent results, so it is a suitable candidate for the competition, remember that is the problem that will be the basis for testing the AT.

References

- [1] "Algoritmo Transgénico" TERCER CONGRESO MEXICANO DE COMPUTACIÓN EVOLUTIVA COMCEV' 07, 2008
- [2] G. Polya, "How to Solve It", 2nd ed., Princeton University Press, 1957, ISBN 0-691-08097-6
- [3] ANSI/IEEE Std 100-1984
- [4] Centro Nacional de Información y Comunicación Educativa (C.N.I.C.E.) Laboratorio de Genética
- [5] Enciclopedia Autodidáctica Millenium, Editorial Norma, S.A., Edición 2001.
- [6] The University of Utah, Genetic Science Learning Center 2006
- [7] Access Excellence The National Health Museum 2006
- [8] UN Economic and Social Development
- [9] Centro Nacional de Información y Comunicación Educativa Departamento de genética España 2001
- [10] David W. Ow, Keith V. Wood, Marlene DeLuca, Jeffrey R. de Wet, Donald R. Helinski and Stephen H. Howell. US Dept. of Agriculture and the National Science Foundation.
- [11] Department of Soil and Crop Sciences at Colorado State University, 1999-2004
- [12] Juan J. Merelo Guervós Depto. Arquitectura y Tecnología de Computadores Escuela Técnica Superior de Ingeniería Informática Campus Aynadamar Daniel Saucedo Aranda, s/n 18071 Granada (Spain)
- [13] Metaheurísticas: una visión global* Belén Melian, José A. Moreno Pérez, J. Marcos Moreno Vega DEIOC. Universidad de La Laguna 38271 La Laguna 2003
- [14] J.D. Schaffer and L.J. Eshelman. "On crossover as an evolutionary viable strategy". In R.K. Belew and L.B. Booker, editors. *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 61-68, Morgan Kaufmann, 1991.
- [15] De Luna E. & B.D. Mishler. 1996 El concepto de homología filogenética y la selección de caracteres taxonómicos. BOLETIN DE LA SOCIEDAD BOTANICA DE MEXICO 59: 131-146
- [16] Prescott, L. (1993). Microbiology, Wm. C. Brown Publishers, [ISBN 0-697-01372-3](https://doi.org/10.1080/00912249308983923)
- [17] Brad L. Miller & David E. Goldberg. 1995. Genetic Algorithms, Tournament Selection, and the Effects of Noise, Illigal Report No. 95006, July 1995.
- [18] Xavier Llorà, Kumara Sastry, David E. Goldberg, Abhimanyu Gupta, and Lalitha Lakshmi. Combating User Fatigue in iGAs: Partial Ordering, Support Vector Machines, and Synthetic Fitness, 2005.
- [19] K. Helsgaun. An Effective implementation of the Lin-Kernighan traveling sales heuristic. European Journal of Operation Research, 126:106-130, 2000. Source code currently available from the author's <http://www.dat.ruc.dk/~keld/>.