New Algorithm Transgenic with Homology to solve the problem of OneMax

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

Christian José Correa Villalón[1], Felipe Padilla Díaz[1], Eugenio Pérez Molphe[1], Eunice Ponce de Leon, Fátima Sayuri Quezada Aguilera[1] and Julio César Ponce Gallegos[1]

[1]Universidad Autonoma de Aguascalienes (Departamento de Inteligencia Artificial); UAA
crisjc_2000@yahoo.com.mx, fpadilla@correo.uaa.mx, eponce@correo.uaa.mx,
sayu_172002@yahoo.com.mx, jcponce@correo.uaa.mx, apadilla@correo.uaa.mx

Article received on July 07, 2009; accepted on October 31, 2009

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" [SCDB, 1992]. 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 [CNICE, 2001]. 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

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 vivientes, o derivados de estos, para hacer o modificar productos o procesos para un uso específico"* [SCDB, 1992]. 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 especifico, generalmente *para mejorar ha dicho individuo o resalar algunas de sus* propias características [CNICE, 2001]. 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 positiva y negativa.

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" [SCDB, 1992].

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 [CNICE, 2001]. Individuals who have had some changes in their genes through biotechnology techniques are known as transgenic, they can be animals or plants. The changes applied to these individuals are made to pursue a specific objective, usually said to improve individual resale or some of its characteristics. In this way there are a lot of examples can be mentioned of genetically modified organisms (GMOs) may range from at least imaginable as the alteration of a plant snuff, which added to the plant genes of firefly, specifically the characteristic of light in contact with a substance can cause some radiation [David, 2005], until things are as common as the cereal we eat in the morning or cotton which are produced with the ticket that commonly pay.

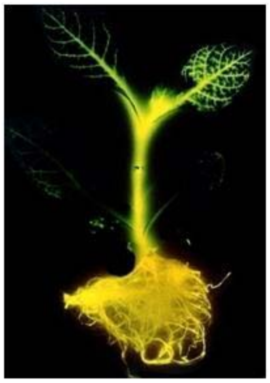


Fig. 1. Snuff of a plant that has been genetically engineered to express a gene taken from fireflies (specifically: the Photinus of pyralis) which produces Lucifero.

1.2 New Algorithm Transgenic

For the first transgenic algorithm [Correa C. J., 2008], 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.

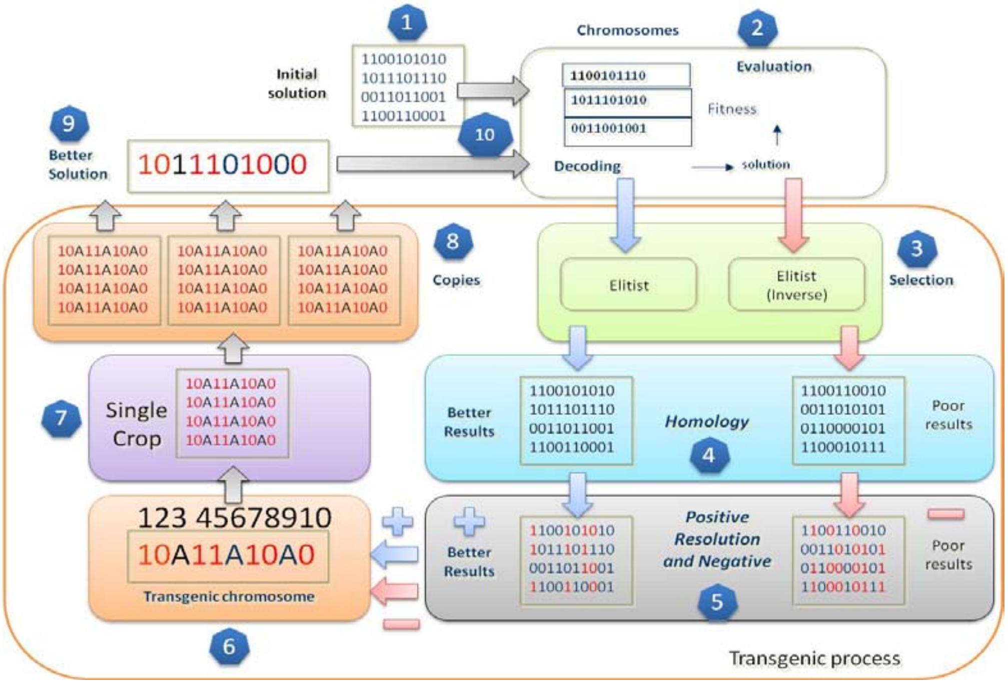


Fig. 2. New Algorithm Transgenic

This new algorithm incorporates new features such as **Homology, Alignment, and Positive and negative resolution.** 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 used to find similar patterns in the best and worst individuals, using an ***Alignment*** of chromosomes, this technique is used in biotechnology to identify the gene to be isolated and manipulated to create new individuals.
3. With the ***alignment of chromosomes*** can be found gene patterns in more fit individuals, this is known as ***positive resolution***, when we found gene patterns in less fit individuals, is known as ***negative resolution***. .
4. We built an individual, respecting the genes of the resolution positive and negative, to produce a chromosome with good gen, but excluding the gen 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 and 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 as seen in Fig 3.

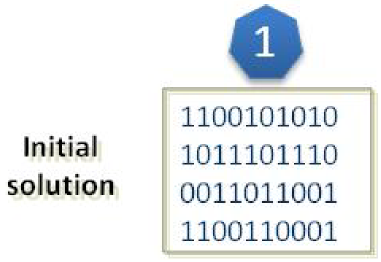


Fig. 3. Initial solution

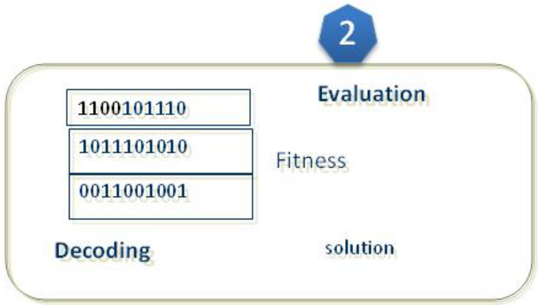


Fig. 4. Evaluation of the objective function

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 is 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 Fig 4.

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 gen 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 ***inverse selection***, the elitist selection to identify the best qualified individuals, the ***inverse elitist*** selection , seeks to identify individuals least able, thus the one hand we have the fittest individuals, and on the other hand we have individuals who are less fit, fig 5.

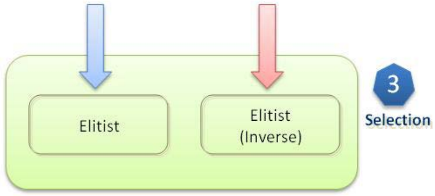


Fig. 5. Elitist selection and reverse selection

```
AAB24882      TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881      --------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
                                  ****: .***:  * *:** * :****.:* *******..

AAB24882      PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881      HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
               **** *:***********:***:**.: .***************    :  *.: :
```

Fig. 6. Alignment of sequences generated by ClustalW between two zinc finger proteins identified by the GenBank access number [De Luna, 1996]

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 [De Luna, 1996]. This method is used in biotechnology in the creation of transgenic organisms, to identify patterns of gen contained in the body that will receive the new gen in the body that is going to donate, Thus the homology identified the gene that can be modified and thus lead to an improvement in an organism, but also identifies that gene should not be altered as this would lead to an unwanted mutation in an organism. 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, fig. 6 [De Luna, 1996]. The homology is used in the algorithm transgenic for the same purpose of finding similar patterns exist that tell us we can manipulate gene and that gene cannot manipulate, and create more fit individuals Fig 7.

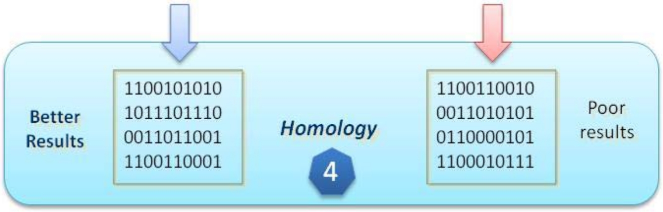


Fig. 7. Homology used in the best and poor results

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).

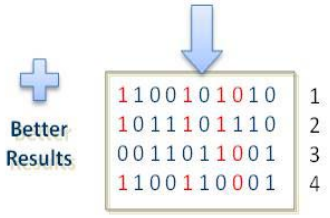


Fig. 8. Alignment of only the best individuals

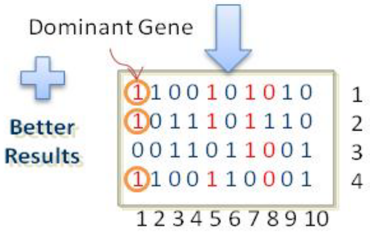


Fig. 9. Positive dominant gene

In Fig. 8 one can observe an alignment of the best individuals, the way that you look for patterns from column to column, we see the ***dominant gene*** in red, the domain of a gene is determined by the number of times this appears in a column, for this example we can see that in column 1 of Fig 9 that there is a ***dominant gene*** that is number 1 in red, so therefore we assume that this is the gene that prevail in the next harvest in that position. Column No. 2 Fig.10 we see that there is no absolute rule for any gene and appears the same number of times both 1 and 0, therefore there is no domain of a gene, so in this position next harvest will be a random value, this process is repeated for each of the columns.

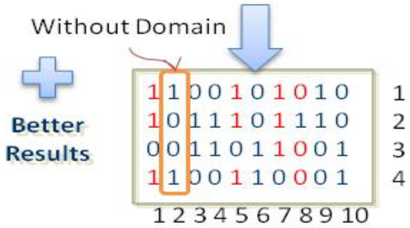


Fig. 10. Column without a dominant gene

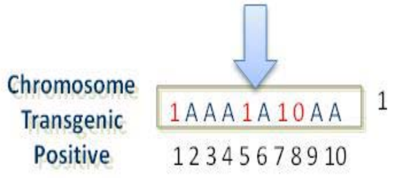


Fig. 11. Results Positive Alignment

At the end of the alignment of individual gen, we obtain a resultant vector, with which we can build the next harvest. In Fig. 11 shows the result of the positive line, the result is a chromosome which starts at position 1 to 10 and also marks 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 a 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 [Prescott, 1993]. The antisense process (negative alignment) is similar to positive alignment, the difference is that the alignment process is applied to individuals resulting from the inverse elitist selection gets the least skilled, with the goal of finding patterns in the genes that generate mutations in chromosomes and generate bad results, only shows the resultant vector of Antisense Fig. 12.

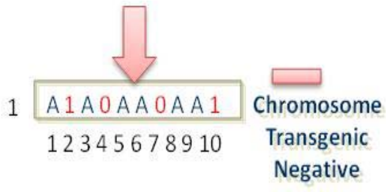


Fig. 12. Result of the antisense

Fig. 13. Resolution positive and negative

In fig. 12 is observed resulting from the antisense chromosome, is a vector that starts at position 1 to 10, please note that in positions 2, 4, 7, 10 there is a dominant gene in the other positions are generated at random chromosomes absence of a dominant gene.

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** Fig 13. Resolution positive and negative, are alienated from the results of the two resulting vectors, the vector positive and negative vector, to obtain a single vector called "***transgenic chromosome***" Fig 14, this vector will indicate how to build the next harvest fig 15. As mentioned before the resulting chromosome will emerge to confront the **positive alignment, and antisense (negative alignment).**

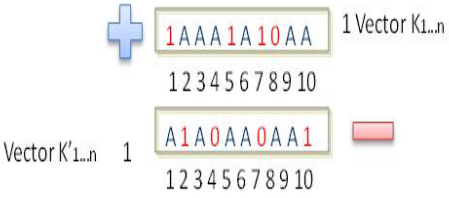


Fig. 14. Positive and Negative Vector

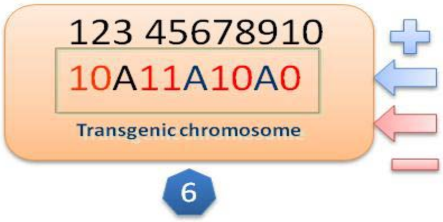


Fig. 15. Transgenic chromosome

The way you get the vector containing the chromosome transgenic R, is comparing the positive vector k, and the negative vector k ', remember that the positive vector contains the genes that we want to preserve and negative (antisense) genes that we want to keep fig 14. The conditions for generating the resultant vector R are:

```
For j = 1 to n = size of chromosome
 {
  If $_{kj}$ = $k'_j$ Then
      $R_j$ = Random
  Else
      $R_j$ is equal to $k_j$ is but that is different from $k'_j$
 }
```

Where, j indicates the position of the gene being evaluated, n is the size of chromosome, so if the gene at position j, the positive vector k_j is equal to gene j, the negative vector (antisense) k'_j, j gene in the transgenic vector R_j, is created randomly A, otherwise the gene j in the transgenic vector R_j have the value of the gene j of the positive vector k_j, but never the gene j, the negative vector k'_j , so the chromosome transgenic R will have only the patterns found in the genes of best individuals and excluding those masters of the worst individuals. The vector resulting can be seen in Fig. 15. The resulting chromosome R will create the basis for seeding for the new population.

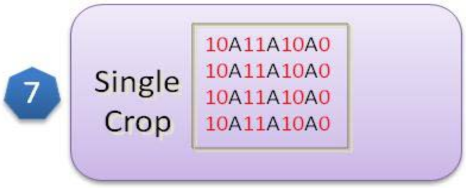


Fig. 16. Vector R

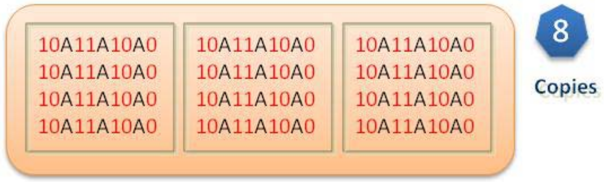
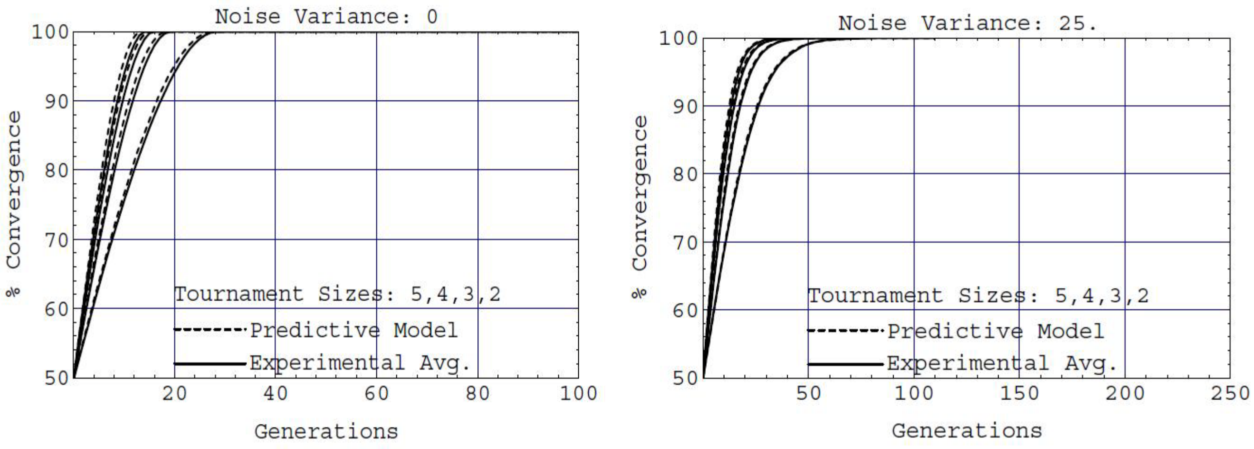


Fig. 17. Copies of the Sowing

Repetition. Using the same principle to generate a transgenic population creates a seed that can be contained only fig. 16, or copies of the resulting population fig. 17, to observe the same behavior when creating copies obtained satisfaction objective function of each, and take the best cases for each of the copies, as well as cases in which the objective function was not executed very good shape. As shown in Figure 19 are generated replicas while respecting the unique seed, where the genes are defined which are unchanged and are not defined supplementing the gene with a random number. Finally repeat the process until you find a stop condition or who have met all the plantings.

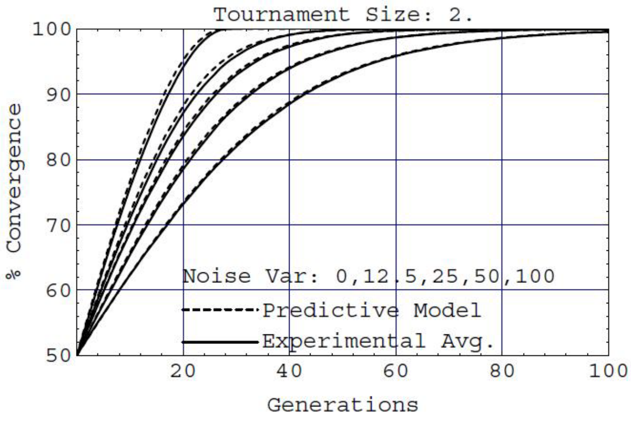
3 Results

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 [Miller, 1995], 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.



(a) Variance: 0

(b) Variance: 25



(c) Tournament Size: 2

Fig. 18. Onemax experiments with different tournament sizes and noise levels

In the graphic a, b, c 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.

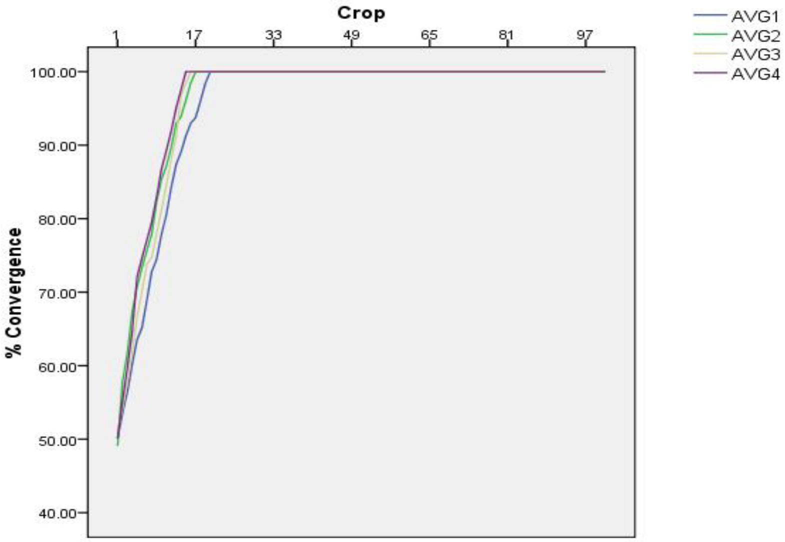


Fig. 19. Algorithm Transgenic experiments, with averages

In Figure 19 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 [Llorá, 2005]. 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.

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.

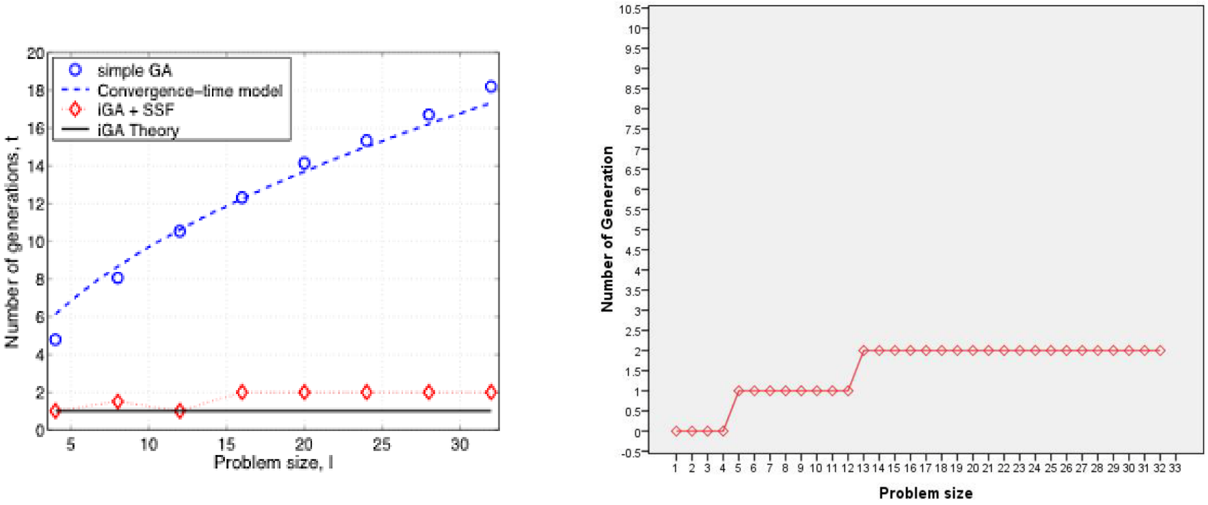


Fig. 20. Comparison of Genetic, Genetic iterative and Transgenic (AT) for OneMax

4 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.

5 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 [Helsgaun, 2000] [Melian B., 2003] 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

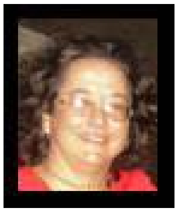
1. **United Nations. (1993).** Convenio sobre la Diversidad Biológica. United Nations. Treaty Series. Vol. 1760 I-30619. Retrieved from http://www.cbd.int/doc/legal/cbd-un-es.pdf
2. **Correa, C. J., Padilla, F., Pérez, E. & Quezada, F. (2007).** Algoritmo Transgénico para resolver el problema del TSP. *Tercer Congreso Mexicano de Computación Evolutiva COMCEV' 07*, Aguascalientes, México, 75-78.
3. **De Luna, E. & Mishler, B.D. (1996)**. El concepto de homología filogenética y la selección de caracteres taxonómicos. *Boletín de la Sociedad Botánica de México*, 59(1), 131-146.
4. **Dorigo, M. & Stützle, T. (2004).** *Ant Colony Optimization*. Cambridge, MA: MIT Press.
5. **Goldberg, D.E. (1989).** *Genetic Algorithms in Search, Optimization and Machine Learning.* Reading MA: Addison-Wesley.

6. **Helsgaun, K. (2000).** An Effective implementation of the Lin-Kernighan traveling sales heuristic. *European Journal of Operation Research*, 126(1), 106-130.
7. **Ignacimuthu, S. (2008).** *Biotechnology: an introduction*. Oxford, UK: Alpha Science International.
8. **Llorà, X., Sastry, K, Goldberg, D. E., Gupta, A., & Lakshmi, L. (2005).** Combating User Fatigue in iGAs: Partial Ordering, Support Vector Machines, and Synthetic Fitness. *Genetic and Evolutionary Computation Conference GECCO'05,* Washington, USA, 1363-1370.
9. **Melián, B., Moreno, J.A. & Moreno, J.M. (2003).** Metaheurísticas: Una visión global. *Revista Iberoamericana de Inteligencia Artificial*, 7(19), 7-28.
10. **Miller, B. L. & Goldberg, D. E. (1995).** Genetic Algorithms, Tournament Selection, and the Effects of Noise. *Complex Systems*, 9(3), 193-212.
11. **Ow, D. W., de Wet, J. R., Helinski, D. R., Howell, S. H., Wood, K. V. & de Luca, M. (1986).** Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science,* 234(4778), 856-859.
12. **Papadimitriou, C.H. (1998).** *Combinatorial Optimization: Algorithms and Complexity*. Mineola, NY: Dover Publications.
13. **Prescott, L., Harley, J., & Klein, D.** (**1999).** *Microbiología* (4). Madrid: McGraw-Hill Interamericana.
14. **Villalobos, V. (2007).** *Los Transgénicos: oportunidades y amenazas* (1). Madrid: Mundi-Prensa.

Christian José Correa Villalón *received the B.S. degree in computer system engineering from the Instituto Tecnológico de Aguascalientes in 2003, and the M.S. degree in computer sciences from the Universidad Autónoma de Aguascalientes in 2007. He is currently working towards the PhD. Degree In the creation of a new optimization algorithm based on transgenic techniques. from the UAA.*

Felipe Padilla Díaz *received the B.S. degree in computer science engineering from the Instituto Tecnológico de Queretaro in 1985, received the M.S. degree in enterprise management from the Universidad Autónoma de Aguascalientes in 1999, and the PhD. in engineering (Artificial Intelligence). His research interests include Evolutionary Computation, Data Mining and learning classifier systems.*

Eugenio Pérez Molphe *received the MSc degree in Plant Biology (1991), and the Doctor degree in Plant Biotechnology (1997), both in the Plant Genetic Engineering Department of the CINVESTAV-IPN. He works as professor and researcher in the Department of Chemistry of the Universidad Autónoma de Aguascalientes. His research interest includes the development of transgenic plants.*



Eunice Esther Ponce de León Sentí *received her PhD degree in Mathematics Science from Cybernetic, Mathematics and Physics Institute from Ministry of Science, Technology and Environment, Cuba in 1998. She works as full time professor-researcher at Computer Science Department of Autonomous University of Aguascalientes, Mexico. Her current research interests include, Combinatorial Optimization, Computational Complexity and Metaheuristics***.**

Fatima Sayuri Quezada Aguilera. *Received her B.S. in Computer Systems from the Autonomous University of Aguascalientes (UAA) in 2003. Actually she is studying the Doctorate in Computer Science at the UAA. Her research interests include Evolutionary Computation and Artificial Immune Systems.*

Julio Cesar Ponce Gallegos *received the B.S. degree in computer system engineering from the Universidad Autónoma de Aguascalientes in 2003, and the M.S. degree in computer sciences from the Universidad Autónoma de Aguascalientes in 2007. He is currently working towards the PhD. Degree in Ant Colony Optimization from the UAA. He is currently an professor in the Universidad Autónoma de Aguascalientes. His research interests include Evolutionary Computation, and Data Mining.*