

NEW HYPER-HEURISTIC ALGORITHM FOR GENE FRAGMENT ASSEMBLY

MURNIYANTI BINTI MALIK

UNIVERSITI TEKNOLOGI MALAYSIA

NEW HYPER-HEURISTIC ALGORITHM FOR GENE FRAGMENT ASSEMBLY

MURNIYANTI BINTI MALIK

A thesis submitted in fulfillment of the  
requirements for the award of the degree of  
Master of Philosophy

Faculty of Computing  
Universiti Teknologi Malaysia

FEBRUARY 2017

To my beloved family

## ACKNOWLEDGEMENT

First of all, I would like to acknowledge my supervisor, Dr. Afnizanfaizal Abdullah for his guidance and advice throughout the making of this thesis and for his patience while dealing with a student like me. I would like to thanks the panel of examiners involved during my VIVA presentation for evaluating my thesis. I am very grateful for your help. Secondly, I would like to thanks my friends and seniors for their helps, supports, opinions and continuous information shared throughout the semester to help me finish my thesis. Without them, I would be a lost cause. Thirdly, I would like to give a thousand thanks to the staff of Faculty of Computing, Universiti Teknologi Malaysia, Mrs. Nizamra, Ms. Norshahida, Dr Norafida, Dr Zalmiyah, and Deputy Dean (Academic) for helping me throughout the process of submitting my thesis. Last but not least, I would like to express my deepest gratitude to my parent and family for their support and encouragement during my master study.

## ABSTRAK

Perhimpunan gen adalah teknik untuk mengenalpasti jujukan gen berdasarkan serpihan gen yang dijana oleh mesin penjujukan. Serpihan gen tersebut adalah pendek dan banyak. Sekiranya bilangan serpihan gen meningkat, kerumitan masalah meningkat, dan situasi ini menjadikan ruang penyelesaian menjadi semakin luas. Untuk menyelesaikan masalah ini, serpihan gen perlu disusun di dalam susunan yang betul. Namun, disebabkan kerumitan dan ruang penyelesaian yang besar, penyelesaian sukar didapati. Melihat dari perspektif pengkomputeran, masalah perhimpunan serpihan gen dianggap sebagai masalah polinomial tidak berketentuan (NP), dimana masalah ini boleh diselesaikan dengan menggunakan algoritma metaheuristik. Algoritma metaheuristik mengoptimumkan masalah dengan mencari penyelesaian yang hampir optimal. Dalam penyelidikan ini, satu algoritma hiper-heuristik dicadangkan untuk menyelesaikan masalah perhimpunan serpihan gen. Penyelidikan ini dibina berdasarkan tiga objektif. Pertama, untuk menganalisa dua algoritma metaheuristik, iaitu Pengoptimuman Reaksi Kimia (CRO) dan Algoritma Kuantum yang diinspirasikan daripada Algoritma Evolusi (QIEA). Kedua, algoritma hiper-heuristik yang baru dibangunkan berdasarkan CRO dan QIEA. Ketiga, penyelesaian yang didapati daripada ketiga-tiga algoritma dinilai menggunakan analisis statistik. Prestasi algoritma-algoritma dinilai dengan menggunakan analisis penumpuan. Persamaan draf gen yang dijana oleh algoritma dianalisis dengan menggunakan Alat Pencarian Penjajaran Tempatan (BLAST). Hasil kajian menunjukkan bahawa QCRO boleh mencari susunan serpihan-serpihan gen dengan betul dan dapat menyelesaikan masalah perhimpunan serpihan gen. Kesimpulannya, penyelidikan ini membentangkan algoritma hiper-heuristik baru untuk menyelesaikan masalah perhimpunan serpihan gen yang dibuat berdasarkan dua algoritma metaheuristik. Algoritma ini boleh mencari susunan serpihan gen yang betul dan menyelesaikan masalah perhimpunan gen.

## ABSTRACT

Gene assembly is a technique to construct a gene sequence by referring to gene fragments generated by sequencing machine. The gene fragments are often short and come in large number. As the number of gene fragments increases, the complexity of the problem increases, and this situation produces a wider solution space. To solve the gene assembly problem, the gene fragments need to be arranged in the right order. However, due to the complexity and wide solution space, the accurate solution to this problem is difficult to be found. By looking from the computational perspective, gene assembly problem is considered as nondeterministic-polynomial (NP) problem, where the gene assembly problem can be solved by using metaheuristic algorithms. Metaheuristic algorithms optimize the problem by searching for almost optimal solution. In this research, a hyper-heuristic algorithm is proposed to solve gene assembly problem due to its advantages that overcome the metaheuristic algorithms. This research is conducted based on three objectives. First, to analyze two metaheuristic algorithms, Chemical Reaction Optimization (CRO) and Quantum Inspired Evolutionary Algorithm (QIEA), to solve the problem. Second, a new hyper-heuristic algorithm (QCRO) is developed based on CRO and QIEA. Third, the solutions generated from all three algorithms are evaluated by using statistical analysis. The performance of the algorithms is evaluated by convergence analysis. The similarities of the draft gene sequence generated by the algorithms are analyzed by using Basic Local Alignment Search Tool (BLAST). The findings show that QCRO is competent in finding the right order of the fragments and solving the gene assembly problem. In conclusion, this research presented a new hyper-heuristic algorithm to solve gene fragment assembly problem that is derived from two metaheuristic algorithms. This algorithm is capable of finding the right order of the gene fragments and thus solves the gene assembly problem.

## TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	<b>DECLARATION</b>	<b>ii</b>
	<b>ACKNOWLEDGEMENT</b>	<b>iv</b>
	<b>ABSTRAK</b>	<b>v</b>
	<b>ABSTRACT</b>	<b>vi</b>
	<b>TABLE OF CONTENTS</b>	<b>vii</b>
	<b>LIST OF TABLES</b>	<b>x</b>
	<b>LIST OF FIGURES</b>	<b>xi</b>
	<b>LIST OF ABBREVIATIONS</b>	<b>xii</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Problem Background	1
	1.2 Problem Statement	3
	1.3 Research Goal and Objectives	3
	1.4 Research Scope and Significance	4
	1.5 Thesis Outline	4
	1.6 Summary	5
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>6</b>
	2.1 Gene Sequence	6
	2.2 Overview on Gene Fragment Assembly	8
	2.3 Sequence Alignment Analysis	9
	2.4 Metaheuristic Algorithm	11

	2.4.1	Genetic Algorithm (GA)	12
	2.4.2	Particle Swarm Optimization (PSO)	13
	2.4.3	Firefly Algorithm (FA)	14
	2.4.4	Ant Colony Optimization (ACO)	15
	2.4.5	Quantum-inspired Evolutionary Algorithm (QIEA)	15
	2.4.6	Chemical Reaction Optimization Algorithm (CRO)	18
	2.5	Overview on Several Available Gene Assembly Software	22
	2.6	Hyper-heuristic Algorithm	26
	2.7	Research Trends and Directions	28
	2.8	Summary	30
<b>3</b>		<b>RESEARCH METHODOLOGY</b>	<b>31</b>
	3.1	Introduction	31
	3.2	Research Framework	31
	3.3	Dataset for Experiment	34
	3.4	Problem Formulation	35
	3.5	Performance Measurement	36
	3.5.1.	Statistical Analysis	36
	3.5.2.	Fitness Convergence	37
	3.5.3.	Validation with BLAST Alignment	37
	3.6	Summary	38
<b>4</b>		<b>DEVELOPMENT OF CRO AND QIEA FOR GENE FRAGMENT ASSEMBLY</b>	<b>39</b>
	4.1	Overview	39
	4.2	CRO for Gene Fragment Assembly	39
	4.3	QIEA for Gene Fragment Assembly	41
	4.4	Experimental Result	43
	4.5	Summary	48
<b>5</b>		<b>DEVELOPMENT OF HYPER-HEURISTIC</b>	<b>49</b>
	5.1	Overview	49
	5.2	Metaheuristic-based Hyper-heuristic Algorithm	51



5.3	Convergence Analysis of CRO, QIEA, and QCRO	53
5.4	Statistical Analysis	59
5.5	Validation with BLAST Analysis	61
5.6	Summary	62
<b>6</b>	<b>CONCLUSION AND FUTURE WORKS</b>	<b>63</b>
6.1	Conclusion	63
6.2	Future Works	64
	<b>REFERENCES</b>	<b>65</b>

## LIST OF TABLES

TABLE NO.	TITLE	PAGE
<b>Table 2.1</b>	Elementary reactions and its suggested operators (Lam and Li, 2010a, 2012)	22
<b>Table 2.2</b>	Gap analysis table of available software for gene assembly.	24
<b>Table 3.1</b>	Details of datasets.	35
<b>Table 4.1</b>	Experimental result with CRO. The experiment is tested on various controlled and manipulated variable. The bolded value is the maximum fitness score for each instance.	45
<b>Table 4.2</b>	Experimental result with QIEA. The experiment is tested on various controlled and manipulated variable. The bolded value is the maximum fitness score for each instance.	46
<b>Table 4.3</b>	CRO analysis. The CRO algorithm was run for 10 times. Variable $b$ is the best fitness score, $w$ is the worst fitness score, $m$ is the mean of the maximum fitness scores, and $s$ is the standard deviation of fitness scores. Parameters: iteration=1000, population size=1000.	47
<b>Table 4.4</b>	QIEA analysis. The QIEA algorithm was run for 10 times. Variable $b$ is the best fitness score, $w$ is the worst fitness score, $m$ is the mean of the maximum fitness scores, and $s$ is the standard deviation of fitness scores. Parameters: iteration=1000, population size=1000.	47
<b>Table 5.1</b>	Statistical comparison of CRO, QIEA, and QCRO. The best maximum fitness score of each algorithm is labelled with *.	60
<b>Table 5.2</b>	Best result alignment from CRO, QIEA, and QCRO compared for the similarity with the original sequence.	61

## LIST OF FIGURES

<b>FIGURE</b>	<b>TITLE</b>	<b>PAGE</b>
<b>Figure 2.1</b>	A double helix of DNA strands (Zephyris, 2011)	7
<b>Figure 2.2</b>	Global alignment (Mount, 2001).	10
<b>Figure 2.3</b>	Local alignment (Mount, 2001).	11
<b>Figure 2.4</b>	Process flow of QIEA (Han and Kim, 2002)	18
<b>Figure 2.5</b>	Framework of CRO	19
<b>Figure 2.6</b>	Framework of hyper-heuristic (Burke <i>et al.</i> , 2003)	27
<b>Figure 2.7</b>	Several framework of hyper-heuristics	29
<b>Figure 3.1</b>	Resarch framework	33
<b>Figure 3.2</b>	Example of BLAST output.	38
<b>Figure 4.1</b>	Pseudocode of CRO used to solve gene fragment assembly. The parameter involved is population size, iteration, kinetic energy loss rate, initial kinetic energy, alpha, beta, and molecule collision rate, and the value of the parameters is set as 1000, 1000, 0.2, 500, 10, 0.2, respectively.	41
<b>Figure 4.2</b>	Pseudocode of repair algorithm.	43
<b>Figure 5.1</b>	Hyper-heuristic framework modified from $F_A$ in Figure 2.7 to fit with algorithm QIEA and CRO criteria	50
<b>Figure 5.2</b>	Framework of hyper-heuristic QCRO	52
<b>Figure 5.3</b>	Convergence analysis graph for instance A	54
<b>Figure 5.4</b>	Convergence analysis graph for instance B	54
<b>Figure 5.5</b>	Convergence analysis graph for instance C	55
<b>Figure 5.6</b>	Convergence analysis graph for instance D	56
<b>Figure 5.7</b>	Convergence analysis graph for instance E	57
<b>Figure 5.8</b>	Convergence analysis graph for instance F	58
<b>Figure 5.9</b>	Convergence analysis graph for instance G	58

**LIST OF ABBREVIATIONS**

A	-	Adenine
ACO	-	Ant colony optimization
C	-	Cytosine
CRO	-	Chemical reaction optimization
DNA	-	Deoxyribonucleic acid
FA	-	Firefly algorithm
G	-	Guanine
GA	-	Genetic algorithm
mRNA	-	Messenger RNA
NP	-	Nondeterministic polynomial
NWA	-	Needleman Wunsch algorithm
OS	-	Operating system
PSO	-	Particle swarm optimization
QIEA	-	Quantum-inspired evolutionary algorithm
RNA	-	Ribonucleic acid
SA	-	Simulated annealing
SWA	-	Smith-Waterman algorithm
T	-	Thymine
TSP	-	Travelling salesman problem
U	-	Uracil

# CHAPTER 1

## INTRODUCTION

### 1.1 Problem Background

Metabolic engineering is a study of biochemical reactions in biochemical reaction pathways. The fundamental of the metabolic engineering is to analyze cells as the integral units. This study involves pathway synthesis, thermodynamic feasibility, and pathway flux and flux control. One important application of metabolic engineering is to manipulate the yield and productivity of products synthesized by microorganisms (Keasling, 2010). To date, various researches that involved the manipulation of metabolic engineering have been carried out extensively.

As the synthetic biology and metabolic engineering area of study partially overlap, a number of studies have presented the used of synthetic biology to facilitate metabolic engineering especially in the applications of industrial biotechnology (Keasling, 2012) and industrial microbiology (Zhang and Nielsen, 2014). Synthetic biology aims to create new biologically functional parts, modules and systems by utilizing various molecular biology and synthetic deoxyribonucleic acid (DNA) tools together with the use of mathematical methodologies to perform new tasks (Chandran *et al.*, 2011; Copeland *et al.*, 2012). The concepts of synthetic biology follow the concepts of computer engineering hierarchy (Andrianantoandro *et al.*, 2006; Chandran *et al.*, 2011; Kronberger, 2012).

The hierarchy consists of few layers: physical layer, device layer, and module layer. In this hierarchy, in physical layer, the transistor, capacitors, and resistors represent

the DNA, ribonucleic acid (RNA), proteins and metabolites. In the device layer, the biochemical reactions in microorganisms represent the engineered logic gates that perform computations in a computer. In the module layer, a library of biological devices to assemble pathways represented integrated circuits. The modules and their integration into host cells can be modified in a programmatic fashion. Biological devices and modules are often dependable to each other. Hence, when the devices or modules are engineered, it will modify the whole cells itself (Andrianantoandro et al., 2006; Chandran et al., 2011; Kronberger, 2012).

One major focus in synthetic biology is to engineer a complex cellular behavior by assembling and expressing genes that will encodes well – characterized biological components. Each cellular function is carried out by ‘modules’ made up of numerous species of interacting molecules. The modules are separable by function. Insulation of modules allow cell to have many diverse reactions without bring any harm to the cell. Connectivity of modules allows one function to influence another (Ajikumar et al., 2010). Some challenges of metabolic engineering require the use of synthetic biology. For example, metabolic engineering is about designing, engineering and optimizing pathways to produce variety of products. Synthetic biology provides synthetic DNA for the constructed metabolic pathway.

Heuristic refers to experience-based techniques to find or to discover by trial and error of a problem. Metaheuristic method means to find or to discover the problem by using higher level heuristic method and perform better than simple heuristics. Hyper-heuristic method is a search method that includes the integration of machine learning techniques to automate the process of selecting, combining, generating or adapting few heuristics or its components to solve computational search problems.

## 1.2 Problem Statement

In genome sequencing, a set of gene fragments is generated from the sequencing machine. The gene fragments itself have several problems. The gene fragments are not in order. The gene fragments order is decided base on the computed overlap score.

Several challenges are identified. The fragments have unknown orientation. The sequence can be read as 5' to 3' or 3' to 5'. If the algorithm is not able to assemble a set of fragments into single contig, the solution is said to have incomplete coverage. The gene fragments may have repeated regions.

The total amount of fragments generated is proportional to the size of original gene. The fragments are not in order, wide solution space, and it is time consuming. Hence, the running time increase with the number of fragments. This problem can be solved by using optimization algorithm.

## 1.3 Research Goal and Objectives

The goal of this research is to apply metaheuristic algorithm to solve assembly problem. The following are the objectives of research.

- a. To analyze two metaheuristic algorithms, CRO and QIEA to solve gene fragment assembly problem.
- b. To develop a new hyper-heuristic algorithm based on CRO and QIEA, called QCRO to solve gene fragment assembly problem.
- c. To validate the result from CRO, QIEA, and QCRO using existing gene sequence database.

## 1.4 Research Scope and Significance

Several DNA fragment benchmark datasets were used to study the performance of the proposed algorithm for gene assembly problem. The dataset is provided in <http://chac.sis.uia.mx/fragbench/> website by Mallén-Fullerton *et al.* (2013). The file format of the dataset is in the form of FASTA and Comma-Separated Values (CSV) for its tabular data. The score of each overlap fragments are presented in matrices form. The method used to optimize the problem is CRO and QIEA. The algorithms were programmed by using programming language Python. Several open source dependencies libraries for Python such as numPy (Van der Walt *et al.*, 2011), sciPy, Matplotlib (Hunter, 2007), is used for multidimensional array, numerical routines, and graph plotting purposes, respectively. The experimental result is generated *in silico*.

The significance of the research is addressed as follows. First, new hyper-heuristic algorithm is designed and developed in terms of the computational contribution. The algorithm employs the heuristic operations of CRO and QIEA to increase the search ability in order to find a good quality of overlap order without the being dependent of the problem and parameters tuning. The outcomes of this research may benefit the biotechnology industry. This is due to the contribution of a new approach in solving gene assembly problem. Since the hyper-heuristic algorithm is problem-independent, the algorithm can be reused and applied for another computational problem such as TSP (Burke *et al.*, 2010; Burke *et al.*, 2003; Burke *et al.*, 2013).

## 1.5 Thesis Outline

This thesis is composed of few chapters.

- Chapter 1: This chapter provides the introduction of the research which includes the research background, problem statement, goal and objectives, and scope and significance of the study.



- Chapter 2: This chapter provides the literature review of the research. It starts with the overview of gene sequence. Then, it proceeds with the discussion of gene fragment assembly problem. Next, the chapter continues on the overview of sequence alignment analysis and discussed the potential metaheuristic algorithms should be used on the research. The research trends and directions on the related issues discussed.
- Chapter 3: This chapter presents the overview of the research operational framework, the flow of the thesis development, and the metaheuristic algorithms, programming language and software used in this study.
- Chapter 4: This chapter discusses the design and development of Chemical reaction optimization algorithm and Quantum-inspired evolutionary algorithm to solve gene fragment assembly problem.
- Chapter 5: This chapter discusses the design and development of hyper-heuristic algorithm to solve gene fragment assembly problem.
- Chapter 6: Conclusion of the research. This chapter also includes contribution of the works and future plans

## 1.6 Summary

In conclusion, synthetic biology area of study follows the concepts and principle of engineering to produce new biological parts, systems, or devices with improved functionality. The use of computational tools in this area of study provides necessary knowledges to solve problem encounters. One problem can have numerous possible solutions available. To find the best solution out of the possible solution, metaheuristic algorithm is used. The metaheuristic algorithm can come in many forms and the result is varied. The metaheuristic algorithm is also problem dependent. Hence, selecting the correct algorithm is necessary to solve specific problem.

## REFERENCES

- Abdullah, A., Deris, S., Mohamad, M. S., and Anwar, S. (2013). An Improved Swarm Optimization for Parameter Estimation and Biological Model Selection. *PLoS ONE*. 8 (4): e61258.
- Andrianantoandro, E., Basu, S., Karig, D. K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol*. 2 2006.0028.
- Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and information science*. 3 (1): 180-184.
- Bastide, M., and McCombie, W. R. (2007). *Assembling genomic DNA sequences with PHRAP*. *Current Protocols in Bioinformatics*. (11.14. 11-11.14. 15).
- Burke, E., Hyde, M., Kendall, G., Ochoa, G., Özcan, E., and Woodward, J. (2010). *A Classification of Hyper-heuristic Approaches*. M. Gendreau and J.-Y. Potvin. *Handbook of Metaheuristics*. (449-468). Springer US.
- Burke, E., Kendall, G., Newall, J., Hart, E., Ross, P., and Schulenburg, S. (2003). *Hyper-heuristics: An emerging direction in modern search technology*. F. Glover and G. A. Kochenberger. *International series in operations research and management science*. (457-474). Springer US.
- Burke, E. K., Gendreau, M., Hyde, M., Kendall, G., Ochoa, G., Oezcan, E., *et al.* (2013). Hyper-heuristics: a survey of the state of the art. *Journal of the Operational Research Society*. 64 (12): 1695-1724.
- Chandran, D., Bergmann, F., Sauro, H., and Densmore, D. (2011). *Computer-Aided Design for Synthetic Biology*. H. Koepl, G. Setti, M. di Bernardo and D. Densmore. *Design and Analysis of Biomolecular Circuits*. (203-224). Springer New York.
- Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M., Kim, K. H., Sleight, S. C., *et al.* (2012). Computational tools for metabolic engineering. *Metabolic Engineering*. 14 (3): 270-280.

- Dorigo, M., and Birattari, M. (2010). *Ant colony optimization. Encyclopedia of Machine Learning.* (36-39). Springer US.
- Eberhart, R. C., and Shi, Y. (2001). Particle swarm optimization: developments, applications and resources. *Evolutionary Computation, 2001.* COEX, Seoul, Korea, 81-86.
- Engle, M. L., and Burks, C. (1993). Artificially Generated Data Sets for Testing DNA Sequence Assembly Algorithms. *Genomics.* 16 (1): 286-288.
- Engle, M. L., and Burks, C. (1994). GenFrag 2.1: new features for more robust fragment assembly benchmarks. *Comput Appl Biosci.* 10 (5): 567-568.
- Feng, X. Y., Wang, Y., Ge, H. W., Zhou, C. G., and Liang, Y. C. (2006). *Quantum-inspired evolutionary algorithm for Travelling salesman problem.* G. R. Liu, V. B. C. Tan and X. Han. *Computational Methods.* (1363-1367). Dordrecht: Springer Netherlands.
- Gutjahr, W. J. (2010). *Convergence Analysis of Metaheuristics.* V. Maniezzo, T. Stützle and S. Voß. *Matheuristics: Hybridizing Metaheuristics and Mathematical Programming.* (159-187). Boston, MA: Springer US.
- Han, K.-H., and Kim, J.-H. (2002). Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. *Evolutionary Computation, IEEE Transactions on.* 6 (6): 580-593.
- Han, K. H. (2003). *Quantum-inspired evolutionary algorithm.* Korea Advanced Institute of Science and Technology, KAIST.
- Huang, K.-W., Chen, J.-L., Yang, C.-S., and Tsai, C.-W. (2014). A memetic particle swarm optimization algorithm for solving the DNA fragment assembly problem. *Neural Computing and Applications.* 26 (3): 465-506.
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering.* 9 (3): 90-95.
- Keasling, J. D. (2010). Manufacturing molecules through metabolic engineering. *Science.* 330 (6009): 1355-1358.
- Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering.* 14 (3): 189-195.
- Kronberger, N. (2012). Synthetic biology: taking a look at a field in the making. *Public Understanding of Science.* 21 (2): 130-133.

- Lam, A. Y., and Li, V. O. (2010a). Chemical-reaction-inspired metaheuristic for optimization. *Evolutionary Computation, IEEE Transactions on*. 14 (3): 381-399.
- Lam, A. Y., and Li, V. O. (2010b). Chemical reaction optimization for cognitive radio spectrum allocation. *2010 IEEE Global Telecommunications Conference GLOBECOM 2010* 6-10 December 2010. Miami, Florida, USA, 1-5.
- Lam, A. Y., and Li, V. O. (2012). Chemical Reaction Optimization: a tutorial. *Memetic Computing*. 4 (1): 3-17.
- Li, Z., Li, Z., Nguyen, T. T., and Chen, S. (2015). Orthogonal chemical reaction optimization algorithm for global numerical optimization problems. *Expert Systems with Applications*. 42 (6): 3242-3252.
- Luque, G., and Alba, E. (2005). Metaheuristics for the DNA fragment assembly problem. *International Journal of Computational Intelligence Research*. 1 (1): 98-108.
- Mallén-Fullerton, G. M., and Fernandez-Anaya, G. (2013). DNA fragment assembly using optimization. *2013 IEEE Congress on Evolutionary Computation*. Cancun, 1570-1577.
- Mallén-Fullerton, G. M., Hughes, J. A., Houghten, S., and Fernández-Anaya, G. (2013). Benchmark datasets for the DNA fragment assembly problem. *International Journal of Bio-Inspired Computation*. 5 (6): 384-394.
- Matsumoto, K.-i., Arai, M., Ishihara, N., Ando, A., Inoko, H., and Ikemura, T. (1992). Cluster of fibronectin type III repeats found in the human major histocompatibility complex class III region shows the highest homology with the repeats in an extracellular matrix protein, tenascin. *Genomics*. 12 (3): 485-491.
- Mitchell, M. (1998). *An introduction to genetic algorithms*. Cambridge, MA, USA: MIT press.
- Mount, D. W. (2001). *Bioinformatics: sequence and genome analysis*. Cold spring harbor laboratory press New York.
- Mullikin, J. C., and Ning, Z. (2003). The Phusion Assembler. *Genome Research*. 13 (1): 81-90.
- Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., et al. (2000). A whole-genome assembly of Drosophila. *Science*. 287 (5461): 2196-2204.

- Ozcan, E., Bilgin, B., and Korkmaz, E. E. (2008). A comprehensive analysis of hyper-heuristics. *Intelligent Data Analysis*. 12 (1): 3.
- Parsons, R. J., Forrest, S., and Burks, C. (1995). Genetic algorithms, operators, and DNA fragment assembly. *Machine Learning*. 21 (1-2): 11-33.
- Pevzner, P. (2000). *Computational molecular biology: an algorithmic approach*. Cambridge, Massachusetts: MIT press.
- Poli, R., and Graff, M. (2009). *There is a free lunch for hyper-heuristics, genetic programming and computer scientists*. L. Vanneschi, S. Gustafson, A. Moraglio, I. D. Falco and M. Ebner. *Genetic Programming*. (195-207). Berlin, Heidelberg: Springer.
- Radenbaugh, A. J. (2008). *Applications of genetic algorithms in bioinformatics*. Master thesis. San Jose University.
- Shaj, V., Akhil, P., and Asharaf, S. (2016). Edge-PSO: A recombination operator based PSO algorithm for solving TSP. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. 21-24 September. Jaipur, India, 35-41.
- Sommer, D., Delcher, A., Salzberg, S., and Pop, M. (2007). Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics*. 8 (1): 64.
- Sutton, G. G., White, O., Adams, M. D., and Kerlavage, A. R. (1995). TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Science and Technology*. 1 (1): 9-19.
- Van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*. 13 (2): 22-30.
- Venter, G., and Sobieszczanski-Sobieski, J. (2003). Particle swarm optimization. *AIAA journal*. 41 (8): 1583-1589.
- Vidal, P., and Olivera, A. C. (2014). *A Parallel Discrete Firefly Algorithm on GPU for Permutation Combinatorial Optimization Problems*. G. Hernández, C. J. Barrios Hernández, G. Díaz, C. García Garino, S. Nesmachnow, T. Pérez-Acle, M. Storti and M. Vázquez. *High Performance Computing: First HPCLATAM - CLCAR Latin American Joint Conference, CARLA 2014, Valparaiso, Chile, October 20-22, 2014. Proceedings*. (191-205). Berlin, Heidelberg: Springer Berlin Heidelberg.

- Wetcharaporn, W., Chaiyaratana, N., and Tongsimma, S. (2006a). *DNA Fragment Assembly by Ant Colony and Nearest Neighbour Heuristics*. L. Rutkowski, R. Tadeusiewicz, L. Zadeh and J. Żurada. *Artificial Intelligence and Soft Computing – ICAISC 2006*. (1008-1017). Springer Berlin Heidelberg.
- Wetcharaporn, W., Chaiyaratana, N., and Tongsimma, S. (2006b). *DNA Fragment Assembly: An Ant Colony System Approach*. F. Rothlauf, J. Branke, S. Cagnoni, E. Costa, C. Cotta, R. Drechsler, E. Lutton, P. Machado, J. Moore, J. Romero, G. Smith, G. Squillero and H. Takagi. *Applications of Evolutionary Computing*. (231-242). Springer Berlin Heidelberg.
- Wolpert, D. H., and Macready, W. G. (1997). No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on*. 1 (1): 67-82.
- Yang, X.-S. (2010). *Firefly Algorithm. Engineering Optimization*. (221-230). John Wiley & Sons, Inc.
- Yu, J. J., Lam, A., and Li, V. O. (2015). Adaptive chemical reaction optimization for global numerical optimization. *2015 IEEE Congress on Evolutionary Computation (CEC)*. 25-28 May 2015. Sendai, Japan, 3192-3199.
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 18 (5): 821-829.
- Zhang, G. (2011). Quantum-inspired evolutionary algorithms: a survey and empirical study. *Journal of Heuristics*. 17 (3): 303-351.
- Zhang, W., and Nielsen, D. R. (2014). Synthetic biology applications in industrial microbiology. *Frontiers in Microbiology*, 5 (1), 451-452. Frontiers Media S.A.
- Zhou, W., Zhou, C., Huang, Y., and Wang, Y. (2005). *Analysis of gene expression data: application of quantum-inspired evolutionary algorithm to minimum sum-of-squares clustering*. D. Ślęzak, J. Yao, J. F. Peters, W. Ziarko and X. Hu. *Rough sets, fuzzy sets, data mining, and granular computing*. (383-391). Berlin, Heidelberg: Springer.