

## Chapter 9

# DNA Fragment Assembly Using Quantum-Inspired Genetic Algorithm

**Manisha Rathee**

*Jawaharlal Nehru University, India*

**Kumar Dilip**

*Jawaharlal Nehru University, India*

**Ritu Rathee**

*Indira Gandhi Delhi Technical University for Women, India*

### ABSTRACT

*DNA fragment assembly (DFA) is one of the most important and challenging problems in computational biology. DFA problem involves reconstruction of target DNA from several hundred (or thousands) of sequenced fragments by identifying the proper orientation and order of fragments. DFA problem is proved to be a NP-Hard combinatorial optimization problem. Metaheuristic techniques have the capability to handle large search spaces and therefore are well suited to deal with such problems. In this chapter, quantum-inspired genetic algorithm-based DNA fragment assembly (QGFA) approach has been proposed to perform the de novo assembly of DNA fragments using overlap-layout-consensus approach. To assess the efficacy of QGFA, it has been compared genetic algorithm, particle swarm optimization, and ant colony optimization-based metaheuristic approaches for solving DFA problem. Experimental results show that QGFA performs comparatively better (in terms of overlap score obtained and number of contigs produced) than other approaches considered herein.*

## **1. INTRODUCTION**

Understanding the functioning (as well as malfunctioning) of living beings require determination and interpretation of their genome sequences (Kikuchi & Chakraborty, 2006). The genome of an organism is made up of deoxyribonucleic acid (DNA) strands which encode its hereditary information and determine its body structure, functions and protein formation (Watson & Berry, 2003). DNA strands consist of two types of nitrogenous bases namely purines (adenine (A) and guanine (G)) and pyrimidines (cytosine (C) and thymine (T)). DNA has a double helical structure consisting of two strands running anti-parallel to each other and having complementary bases where a purine on one strand is paired with pyrimidine on the other and vice-versa in such a way that A is always paired with T and G is always paired with C (Watson & Berry, 2003; Watson & Crick, 1953). The process of determining the complete sequence of bases in all the strands of DNA (i.e. the genome) is termed as DNA sequencing. The genome sequences are generally very large ranging from few thousand base pairs for small viruses to  $3 \times 10^9$  base pairs for humans,  $1.7 \times 10^{10}$  base pairs for wheat and  $1.2 \times 10^{11}$  base pairs for lily (Kikuchi & Chakraborty, 2006). A number of techniques are available for sequencing but none of the available techniques is capable of reading more than 1000 bases at a time, let alone reading an entire genome at once. This limitation of DNA sequencing methods is overcome by shotgun sequencing where the target DNA is replicated to generate multiple copies which are then randomly broken into a number of smaller fragments so that the fragments are short enough to be sequenced by any of the available sequencing techniques (Dorransoro, *et al.*, 2008). After sequencing, the sequenced fragments need to be combined back to obtain the original sequence as the whole genome sequence is required for phylogenetic and genomic research activities. But, as the fragments were generated randomly, the information about ordering of the fragments on the parent strand or the strand to which a particular fragment belongs is lost thereby resulting in the DNA fragment assembly problem (Meksangsouy & Chaiyaratana, 2003). DFA problem involves reconstruction of target DNA from several hundred (or thousands) of sequenced fragments by identifying the proper orientation and order of the sequenced fragments (Meksangsouy & Chaiyaratana, 2003). The sequenced fragments are called as reads and are provided as input to the assembly procedure.

DFA is proved to a NP-Hard combinatorial optimization problem (Medvedev, *et al.*, 2007). The complexity arises due to a very large search space as there are  $2^k \times k!$  possible solutions in worst case for a set containing  $k$  fragments (Kubalik, *et al.*, 2010). Therefore obtaining exact solutions using traditional optimization techniques is not possible. Metaheuristic techniques have the capability to handle large search spaces and therefore well suited to solve the hard optimization problems. In this chapter, Quantum inspired genetic algorithm (QIGA) has been adapted for performing the de novo assembly of DNA fragments using overlap-layout-consensus approach. QIGA is a relatively recent metaheuristic technique which blends the principals of quantum computing with the concepts of genetic algorithm (Han & Kim, 2002). Due to the parallel processing capabilities of QIGA, it is capable of providing better solutions (in terms of diversity, quality and convergence) with a smaller population size. Also, QIGA has an edge over other metaheuristic techniques as very less number of parameters needs to be adjusted in case of QIGA.

Quantum inspired Genetic algorithm-based DNA Fragment Assembly (QGFA) has been proposed in this paper with the objective of maximizing the sum of overlaps between adjacent fragments in a layout. The main contribution of the proposed work is mentioned below:

- QIGA has been adapted for solving the DFA problem. To the best of our knowledge, none of the works existing in literature has used QIGA for solving the DFA problem.

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/dna-fragment-assembly-using-quantum-inspired-genetic-algorithm/277775](http://www.igi-global.com/chapter/dna-fragment-assembly-using-quantum-inspired-genetic-algorithm/277775)

## Related Content

---

### A Review on Quantum Deep Machine Learning Model for Predicting Rice Husk Ash Compressive Strength

Dorothy Blessing Agboola, Micheal Olaolu Arowolo and Amit Kumar Tyagi (2023). *Handbook of Research on Quantum Computing for Smart Environments* (pp. 103-120).

[www.irma-international.org/chapter/a-review-on-quantum-deep-machine-learning-model-for-predicting-rice-husk-ash-compressive-strength/319864](http://www.irma-international.org/chapter/a-review-on-quantum-deep-machine-learning-model-for-predicting-rice-husk-ash-compressive-strength/319864)

### Complex Action Methodology for Enterprise Systems (CAMES): A System to Contextualize the Behavioral Management Issue as Quantum Mechanical Variable

Olaf Cames and Meghann L. Drury-Grogan (2021). *Research Anthology on Advancements in Quantum Technology* (pp. 387-399).

[www.irma-international.org/chapter/complex-action-methodology-for-enterprise-systems-cames/277786](http://www.irma-international.org/chapter/complex-action-methodology-for-enterprise-systems-cames/277786)

### An Improved Generalized Quantum-Inspired Evolutionary Algorithm for Multiple Knapsack Problem

Sulabh Bansal and C. Patvardhan (2021). *Research Anthology on Advancements in Quantum Technology* (pp. 22-50).

[www.irma-international.org/chapter/an-improved-generalized-quantum-inspired-evolutionary-algorithm-for-multiple-knapsack-problem/277768](http://www.irma-international.org/chapter/an-improved-generalized-quantum-inspired-evolutionary-algorithm-for-multiple-knapsack-problem/277768)

### Introduction and Beginners Guide to Quantum Computing

Poornima Nedunchezian and Rajkumar Rajasekaran (2022). *Technology Road Mapping for Quantum Computing and Engineering* (pp. 1-10).

[www.irma-international.org/chapter/introduction-and-beginners-guide-to-quantum-computing/300513](http://www.irma-international.org/chapter/introduction-and-beginners-guide-to-quantum-computing/300513)

### Multi-Process Analysis and Portfolio Optimization Based on Quantum Mechanics (QM) Under Risk Management in ASEAN Exchanges: A Case Study of Answering to the E-Commerce and E-Business Direction

Chukiat Chaiboonsri and Satawat Wannapan (2021). *Research Anthology on Advancements in Quantum Technology* (pp. 400-415).

[www.irma-international.org/chapter/multi-process-analysis-and-portfolio-optimization-based-on-quantum-mechanics-qm-under-risk-management-in-asean-exchanges/277787](http://www.irma-international.org/chapter/multi-process-analysis-and-portfolio-optimization-based-on-quantum-mechanics-qm-under-risk-management-in-asean-exchanges/277787)