

# Chapter 59

## An Optimized Graph– Based Metagenomic Gene Classification Approach: Metagenomic Gene Analysis

**Md Sarwar Kamal**

*East West University, Bangladesh*

**Mohammad Ibrahim Khan**

*Chittagong University of Engineering and Technology, Bangladesh*

**Kaushik Dev**

*Chittagong University of Engineering and Technology, Bangladesh*

**Linkon Chowdhury**

*Chittagong University of Engineering and Technology, Bangladesh*

**Nilanjan Dey**

*Techno India College of Technology, India*

### ABSTRACT

*Biological interaction mainly depends on the interactions of various genes and genomes. To identify actual meaning of interactions we have to find out the facts and reasons for these interactions. Gene analysis allows to verify such environment. Gene annotation means to identify the exon regions in metagenomic samples. The de Bruijn graph plays significant role in gene prediction and next generation sequencing (NGS). Apart from that, Euler Path of de Bruijn graph introduced generalized gene annotation for translational and splicing signals, exon introns separation and coding regions. set of graph reduction rules have used to build a de Bruijn graph. Accurate solution for large scale sequencing, trims space complexity and generates optimal gene annotation have tested.*

DOI: 10.4018/978-1-7998-1204-3.ch059

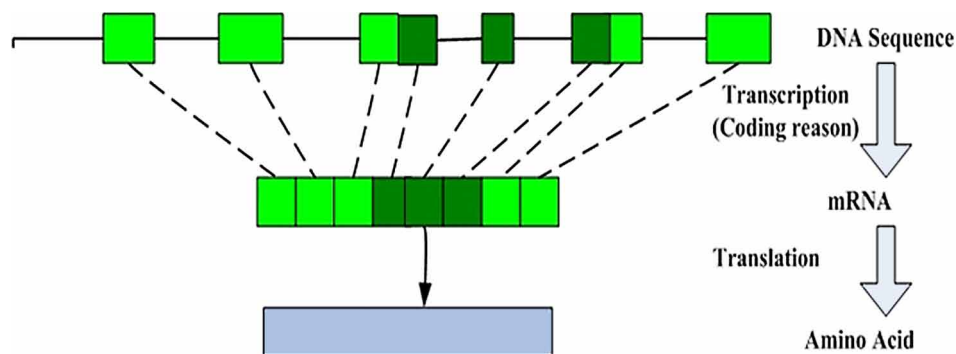
## INTRODUCTION

The approach that specifies or categorized the genetic substance from environmental collections is called metagenomics. This process is important due to its uniqueness on data clustering facilities. I.e. the system permits to organize huge molecular data into a small area. By imposing metagenomics concept, various vital and unique molecules with pivotal activities and applications have been determined. The populations and demand of foods and daily necessary goods are increasing day by day. To fulfill the demand of daily commodities as well as medical support, metagenomics will play very important role. Lot of companies are willing and investing in this sector to clear the unknown genetic information. So metagenomics gene analyses will dominant in enzymes processes, medicines, agriculture, ecology and new generic products. Metagenomics ensure to supply new molecules and novel enzymes with dynamic activities and provide factors from set of enzymes from the cultivable microorganisms. Apart from that, the uses of metagenomics are used to uncovering novel biocatalysts from nature, bioremediation and xenobiotic metabolism.

DNA sequence consists of four nucleotides A, G, C and T. These sequencing efforts generate a large amount of raw data as the DNA sequence of a eukaryote is often longer than a hundred million base pairs. The genome of human has 3.2 billion (appx) base pairs. The annotation of these sequences of biological data is needed as a computation tool for gene finding. Locating the genes is not only helpful but also a prerequisite for further analysis, such as the characterization of the function of the gene product, determining the phylogeny of different species or understanding gene regulation. But the problem of finding genes in a genomic DNA sequence is difficult and high space complexity.

Metagenomics, the investigation of crude bacterium under the study of genomic substances gathered precisely from natural data sets, has been applied to analyze the shape, activities, and mechanical interactions of small communities. Various analysis permit us to know about different microbes in multiple areas (Ley et al., 2008; Rusch et al., 2007; Wu et al., 2011), their composite architectures of the communities they built (Huttenhower et al., 2012), and the pivotal encounter they maintain on the associated environment (Tyson et al., 2004). It important to analyze the various bacterium interactions

*Figure 1. The central dogma is a framework for understanding the flow of genetic information. It states that DNA makes RNA, and RNA makes protein (amino acid.) In transcription process DNA convert into RNA specially mRNA(messenger RNA) and in translation RNA convert into protein(amino acid).*



23 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/an-optimized-graph-based-metagenomic-gene-classification-approach/243162](http://www.igi-global.com/chapter/an-optimized-graph-based-metagenomic-gene-classification-approach/243162)

## Related Content

---

### Medical Image Segmentation and Analysis

Ramgopal Kashyap (2019). *Advanced Classification Techniques for Healthcare Analysis* (pp. 132-160).  
[www.irma-international.org/chapter/medical-image-segmentation-and-analysis/222144](http://www.irma-international.org/chapter/medical-image-segmentation-and-analysis/222144)

### An Experimental Data of Lithium-Ion Battery Time Series Analysis: ARIMA and SECTRAL Analysis

Liming Xie (2021). *International Journal of Data Analytics* (pp. 1-26).  
[www.irma-international.org/article/an-experimental-data-of-lithium-ion-battery-time-series-analysis/285465](http://www.irma-international.org/article/an-experimental-data-of-lithium-ion-battery-time-series-analysis/285465)

### A Predictive Analytics Framework for Blood Donor Classification

Kavita Pabrejaand Akanksha Bhasin (2021). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-14).  
[www.irma-international.org/article/a-predictive-analytics-framework-for-blood-donor-classification/277644](http://www.irma-international.org/article/a-predictive-analytics-framework-for-blood-donor-classification/277644)

### Identifying the Factors Associated With Inpatient Admissions for Non-COVID-19 Illnesses: Application of Regression Analysis and NFL Theorem

Chamila K. Dissanayakeand Dinesh R. Pai (2022). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-24).  
[www.irma-international.org/article/identifying-the-factors-associated-with-inpatient-admissions-for-non-covid-19-illnesses/312576](http://www.irma-international.org/article/identifying-the-factors-associated-with-inpatient-admissions-for-non-covid-19-illnesses/312576)

### Fuzzy Logic-Based Predictive Model for the Risk of Sexually Transmitted Diseases (STD) in Nigeria

Jeremiah A. Balogun, Florence Alaba Oladeji, Olajide Blessing Olajide, Adanze O. Asinobi, Olayinka Olufunmilayo Olusanyaand Peter Adebayo Idowu (2020). *International Journal of Big Data and Analytics in Healthcare* (pp. 38-57).  
[www.irma-international.org/article/fuzzy-logic-based-predictive-model-for-the-risk-of-sexually-transmitted-diseases-std-in-nigeria/259987](http://www.irma-international.org/article/fuzzy-logic-based-predictive-model-for-the-risk-of-sexually-transmitted-diseases-std-in-nigeria/259987)