

Chapter 12

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.

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

DOI: 10.4018/978-1-5225-0140-4.ch012

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 however, human metagenomics interactions are very essential to utilize the meaningful result due to its impact on mankind need as drugs, foods and diseases identifications (Qin et al., 2012; Behre C.J, et al., 2013; Greenblum et al., 2011).

In the age of information superhighway, numerous metagenomic analyses are based on gene-centric processes and as well as new aim to cluster the complete genomic shape of the association by virtue of shotgun metagenomics. Mathematical analysis based on shotgun sequencing impact helped to mapped database to a orthologous gene groups such as KEGG (Kanehisa et al., 2012), COG (Tatusov et al., 2003), EggNOG (Powell et al., 2012), M5NR (wike et al., 2012) to investigate similarities to genes or proteins with real and annotated activities. It is important to know that the biological activities among similar genes are work together in contrast, with taxonomic identify resulted in the experimental analysis. BLAST is frequently used to mapped to obtain greater score values on sequence alignment and annotations (Dalevi et al., 2008). This annotation mechanism and the measurements of the data set alignment that mapped individual activity is a an important part in computational metagenomic research and it helps to identify the functional shifts related with disease, as it is the basis of general genomic of analyses.

An important hypothesis afterwards, this annotation process and the key to properly saving the absolute structures of the metagenome alignment where small genome sequencing input holds abundant information to precise map to the exact activity. By considering different annotation rules and assumptions that each input will be directed to either its gene of or to an orthologous gene from an efficient distant genome annotated with the same function. These rules for future assumption that input initiating from intergenic regions or from orthology collection with unknown activity will not be properly mapped to any of the active annotated genes in our database. These considerations are vital for both estimation as well as identifying the presence or absence of gene groups. Now-a-days, but, no details and systematic studies has not checked this analysis on a broad scale, nor have rules for this annotation system been regulated.

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/149395

Related Content

Microscopic Biopsy Image Segmentation Using Hybrid Color K-Means Approach

Rajesh Kumar, Rajeev Srivastava and Subodh Srivastava (2017). *International Journal of Computer Vision and Image Processing* (pp. 79-90).

www.irma-international.org/article/microscopic-biopsy-image-segmentation-using-hybrid-color-k-means-approach/177202

Deep Neural Network With Feature Optimization Technique for Classification of Coronary Artery Disease

Pratibha Verma, Sanat Kumar Sahu and Vineet Kumar Awasthi (2023). *Handbook of Research on Computer Vision and Image Processing in the Deep Learning Era* (pp. 257-269).

www.irma-international.org/chapter/deep-neural-network-with-feature-optimization-technique-for-classification-of-coronary-artery-disease/314001

A Review of Services-Based Architectures for Video Surveillance

Henry Duque-Gomez and Jorge Azorin-Lopez (2021). *International Journal of Computer Vision and Image Processing* (pp. 39-46).

www.irma-international.org/article/a-review-of-services-based-architectures-for-video-surveillance/270875

Reporting Clinical Gait Analysis Data

Raymond White and Robert Noble (2008). *User Centered Design for Medical Visualization* (pp. 58-72).

www.irma-international.org/chapter/reporting-clinical-gait-analysis-data/30624

Object Recognition via Contour Points Reconstruction Using Hurwitz - Radon Matrices

Dariusz Jakóbczak (2013). *Image Processing: Concepts, Methodologies, Tools, and Applications* (pp. 998-1018).

www.irma-international.org/chapter/object-recognition-via-contour-points/77584