

# Stringology Combats Microbiological Threats

Michal Ziv-Ukelson

Ben Gurion University of the Negev, Israel

<https://www.cs.bgu.ac.il/~negevcb/>

michaluz@cs.bgu.ac.il

---

## Abstract

A major concern worldwide is the acquisition of antibiotic resistance by pathogenic bacteria. Genomic elements carrying resistance and virulence function can be acquired through horizontal gene transfer, yielding a broad spread of evolutionary successful elements, both within and in between species, with devastating effect. Recent advances in pyrosequencing techniques, combined with global efforts to study microbial adaptation to a wide range of ecological niches (and in particular to life in host tissues that we perceive as pathogenesis), yield huge and rapidly-growing databases of microbial genomes.

This big new data statistically empowers genomic-context based approaches to functional analysis: the idea is that groups of genes that are clustered locally together across many genomes usually express protein products that interact in the same biological pathway, and thus the function of a new, uncharacterized gene can be deciphered based on the previously characterized genes that are co-localized with it in the same gene cluster. Identifying and interpreting microbial gene context in huge genomic data requires efficient string-based data mining algorithms. Additionally, new computational challenges are raised by the need to study the grammar and evolutionary spreading patterns of microbial gene context.

In this talk, we will review some classical combinatorial pattern matching and data mining problems, previously inspired by this application domain. We will re-examine the biological assumptions behind the previously proposed models in light of some new biological observations. We will consider the computational challenges arising in accomodating the new biological observations, and in exploiting them to scale up the algorithmic solutions to the huge new data. Our goal is to inspire interesting new problems that harness Stringology to the study of microbial adaptation and to the fight against microbiological threats . . .

**2012 ACM Subject Classification** Applied computing → Computational biology

**Keywords and phrases** comparative genomics, syntenic blocks, gene clusters, reconciliation of gene and species trees

**Digital Object Identifier** 10.4230/LIPIcs.CPM.2019.3

**Category** Invited Talk



© Michal Ziv-Ukelson;

licensed under Creative Commons License CC-BY

30th Annual Symposium on Combinatorial Pattern Matching (CPM 2019).

Editors: Nadia Pisanti and Solon P. Pissis; Article No. 3; pp. 3:1–3:1

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany