Група за биоинформатику

Семинар
Рачунарство и информатика
10. фебруар 2011.

Универзитет у Београду
Математички факултет

Биоинформатика

-

Проблеми

-
  -
  -

Проблеми (2)

-
  -
  -
  -
  -
  -

Блиске области

-
  -
  -
  -
  -
  -
-
  -
  -

Основне карактеристике

-
-
-

Група за биоинформатику

-
  -

Почетак

-
  -
  -
    -
-
  -
    -
    -

Организација рада

-
  -
  -
  -
-
  -
  -
  -

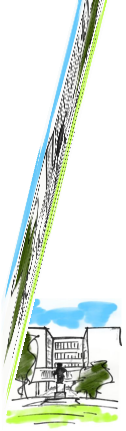
Састав групе

-
  -
  -
  -
  -
  -
  -
  -

Састав групе (2)

Сарадња

- *Center for information science and technology, Temple University, Philadlphia*



Области

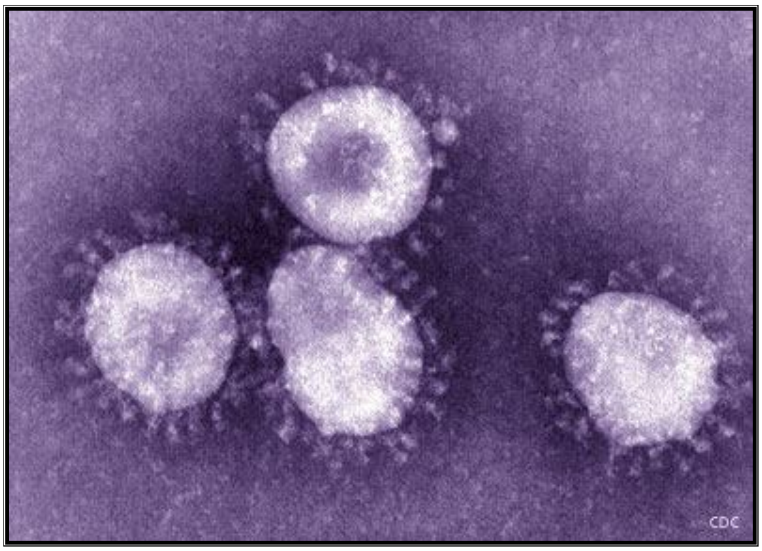
-
  -
  -

1. БИ анализа виралних генома

-
  - *SARS CoV Severe acute respiratory syndrome*
  -
  -

Биоинф. анализа виралних генома (2)

-
  -
  -

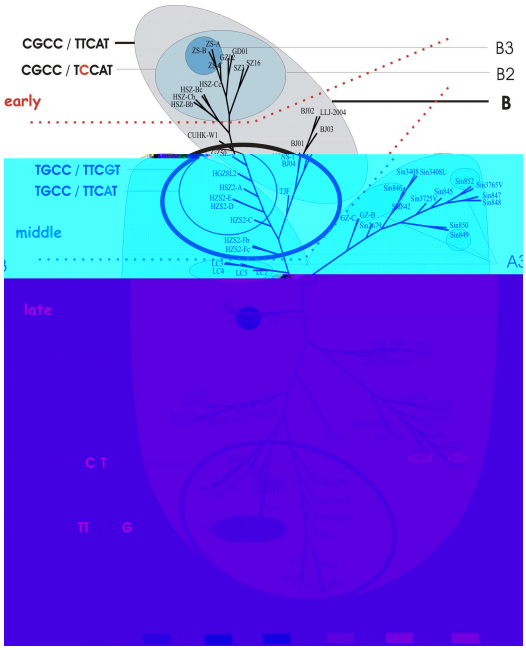


SARS

- *S*

Биоинф. анализа виралних генома (4)

-
  -
  -
  - *N*
  -
  - *S*



BMC Bioinformatics

BioMed Central

Research article

Open Access

Bioinformatics analysis of SARS coronavirus genome polymorphism

Gordana M Pavlović-Lažetić[1], Nenad S Mitić*[1] and Miloš V Beljanski[2]

Address: [1]Faculty of Mathematics, University of Belgrade, P.O.B. 550, Studentski trg 16, 11001 Belgrade, Serbia and Montenegro and [2]Institute of General and Physical Chemistry, P.O.B. 551, Studentski trg 16, 11001 Belgrade, Serbia and Montenegro

Email: Gordana M Pavlović-Lažetić - gordana@matf.bg.ac.yu; Nenad S Mitić* - nenad@matf.bg.ac.yu; Miloš V Beljanski - mbel@matf.bg.ac.yu

* Corresponding author

Published: 25 May 2004

BMC Bioinformatics 2004, 5:65

Received: 24 December 2003

Accepted: 25 May 2004

This article is available from: http://www.biomedcentral.com/1471-2105/5/65

Abstract

Background: We have compared 38 isolates of the SARS-CoV complete genome. The main goal was twofold: first, to analyze and compare nucleotide sequences and to identify positions of single nucleotide polymorphism (SNP), insertions and deletions, and second, to group them according to

Article

SARS-CoV Genome Polymorphism: A Bioinformatics Study

Gordana M. Pavlović-Lažetić[1]*, Nenad S. Mitić[1], Andrija M. Tomović[2], Mirjana D. Pavlović[3], and Miloš V. Beljanski[3]

[1] *Faculty of Mathematics, University of Belgrade, 11001 Belgrade, Serbia and Montenegro;* [2] *Friedrich Miescher Institute for Biomedical Research, CH-4058 Basel, Switzerland;* [3] *Institute of General and Physical Chemistry, 11001 Belgrade, Serbia and Montenegro.*

A dataset of 103 SARS-CoV isolates (101 human patients and 2 palm civets) was investigated on different aspects of genome polymorphism and isolate classification. The number and the distribution of single nucleotide variations (SNVs) and insertions and deletions, with respect to a "profile", were determined and discussed ("profile" being a sequence containing the most represented letter per position). Distribution of substitution categories per codon positi

and non-synonymous substitutions in coding regions of annotated isolates, was determined, along with amino acid (a.a.) property changes. Similar analysis was performed for the spike (S) protein in all the isolates (55 of them being predicted for the first time). The ratio Ka/Ks confirmed that the S gene was subjected to the Darwinian selection during virus transmission from animals to humans. Isolates from the dataset were classified according to genome polymorphism and

types. Genome polymorphism yields to two groups, one with a small number of SNVs and another with a large number of SNVs, with up to four subgroups with respect to insertions and deletions. We identified three basic nine-locus genotypes: TTTT/TTCGG, CGCC/TTCAT, and TGCC/TTCGT, with four subgenotypes. Both classifications proposed are in accordance with the new insights into possible epidemiological spread, both in space and time.

2. Секундарна структура протеина

-
-
  -
  -
-
  -
  -
  -

Секундарна структура протеина (2)

Секундарна структура протеина (3)

MIODRAG ŽIVKOVIĆ[1]
SAŠA MALKOV[1*]
SNEŽANA ZARIĆ[2]
MILENA VUJOŠEVIĆ-JANIČIĆ[1]
JELENA TOMAŠEVIĆ[1]
GORAN PREDOVIĆ[1]
NOVICA BLAŽIĆ[1]
MILOŠ V. BELJANSKI[3]

[1]Department of Mathematics, University of Belgrade, Beograd, Serbia and Montenegro
[2]Department of Chemistry, University of Belgrade, Beograd, Serbia and Montenegro
[3]Institute of General and Physical Chemistry, Beograd, Serbia and Montenegro

SCIENTIFIC PAPER

547.96:547.466:519.23

STATISTICAL DEPENDENCE OF PROTEIN SECONDARY STRUCTURE ON AMINO ACID BIGRAMS

The statistical dependence of protein secondary structure on amino acid bigram frequencies was studied.
Proteins in the PDBSELECT subset of the Protein Data Bank database were investigated. Protein secondary structures were determined using DSSP software. The conditional probabilities of protein secondary structures were calculated and presented. The results on bigrams show the frequencies of all the possible bigrams in all secondary structure types. These results elucidate some factors important for the prediction of the secondary structures of proteins based on the amino acid sequence.

Key words: Amino–acid pairs, Protein secondary structure, Bigram, Bigram frequencies.

Pharmacology and biotechnologies using protein engineering strongly depend on the prediction of the generic protein function and structure, since the protein function is related to its structure. Because detailed protein 3D structure determination is a very costly process, there is extensive development of methods for the prediction of protein structure based on the amino acid sequence. In many of these methods the first step is the prediction of the protein secondary structure.

bigrams can be used as a tool for efficient protein comparison and classification [14].

The occurrence of amino acids and amino acid pairs in different positions in α–helices has been studied by Gobel et al. [15] and Engel et al. [16]. Exceptional amino acid pairs are identified as pairs, the frequency of which substantially differ from the expected value. Statistical data on the amino acid pair compatibility both spatially nearest neighbors and adjacent resi

Protein Journal, 2009

Protein J
DOI 10.1007/s10930-009-9166-3

A Reexamination of Correlations of Amino Acids with Particular Secondary Structures

Saša N. Malkov · Miodrag V. Živković · Miloš V. Beljanski · Srđan Đ. Stojanović · Snežana D. Zarić

Abstract Using the data from Protein Data Bank the correlations of primary and secondary structures of proteins [illegible] on α-helices and strands at some distance in the sequence. The diagrams corresponding to polar amino acids are [illegible] with similar diagrams have similar local [illegible] with the particular secondary structure differ at [illegible] requirements [illegible] is are significant at distances between −9 and 10. [illegible] show that the substituents on Cβ or Cγ atoms of [illegible] play major role in their preference for particular structure at the same position in the sequence. [illegible] polarity of amino acid has significant influence [illegible] sequence [illegible] most 25. [illegible] correlation [illegible] The results [illegible] amino acid [illegible] secondary [illegible] while the

Keywords Protein · Amino acid · Protein secondary structure · Statistical correlation

Abbreviations
DSSP Define secondary structure of proteins the

3. БИ анализа бактеријских генома

- *Genomic Islands GI*
  -
  - *HGT*
  -
- *GI*
  -
  -
  -
  -

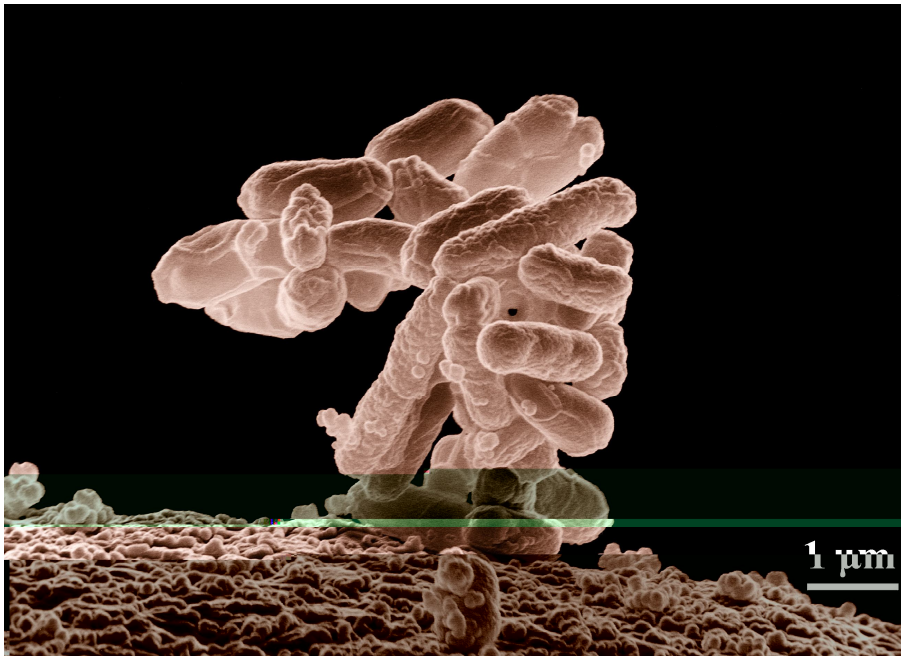
БИ анализа бактеријских генома (2)

- *pathogenicity islands, PAIs*
  - *GI*
  -
- *PAI*
  - ***композициона*** *GC*
  - ***функционална***

БИ анализа бактеријских генома (3)

-
  - *GI*
  - *GI*
-
  - *n-*
  -
-
  - *Escherichia coli O157:H7 EDL933*
  -

Колонија бактерија *E-coli*



БИ анализа бактеријских генома (5)

- *E-coli O157:H7 EDL933*
- *Enterobacteriaceae Escherichia Shigella*
- *N*
- *GI*
- *N*

БИ анализа бактеријских генома (6)

-
  -
    - *GI*
  -
  - *PAI-HGT*

Journal of Biomedical Informatics, 2008

Volume 93, Issue 3, Pages 241-256 (March 2009)

n-Gram characterization of genomic islands in bacterial genomes

Gordana M. Pavlović-Lažetić[a], Nenad S. Mitić[a], Miloš V. Beljanski[b]

Received 20 April 2008; received in revised form 10 September 2008; accepted 21 October 2008.

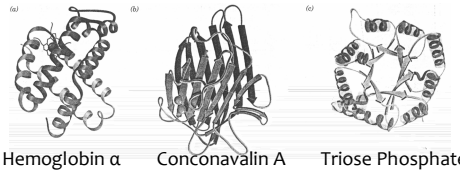
Abstract

The paper presents a novel, *n*-gram-based method for analysis of bacterial genome segments known as *genomic islands* (GIs). Identification of GIs in bacterial genomes is an important task since many of them represent inserts that may contribute to bacterial evolution and pathogenesis. In order to characterize and distinguish GIs from rest of the genome, binary classification of islands based on *n*-gram frequency distribution have been performed. It consists of testing the agreement of islands *n*-gram frequency distributions with the complete genome and backbone sequence. In addition, a statistic based on the maximal order Markov model is used to identify significantly overrepresented and underrepresented *n*-grams in islands. The results may be used as a basis for Zipf-like analysis suggesting that some of the *n*-grams are overrepresented in a subset of islands and underrepresented in the backbone, or vice versa, thus complementing the binary classification. The method is applied to strain-specific regions in the *Escherichia coli* O157:H7 EDL933 genome (O-islands), resulting in two groups of O-islands with different *n*-gram characteristics. It refines a characterization based on other compositional features such as G + C content and codon usage, and may help in identification of GIs, and also in research and development of adequate drugs targeting virulence genes in them.

Abbreviations: HGT, horizontal gene transfer, GI, genomic island, OI, O-island, PAI, pathogenesis isla
CU, codon usage

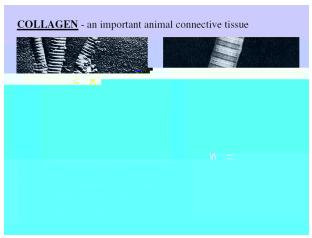
4. БИ анализа неуређености протеина код прокариота

-
  -



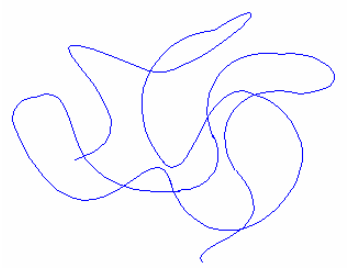
Hemoglobin α Conconavalin A Triose Phosphate isomerase

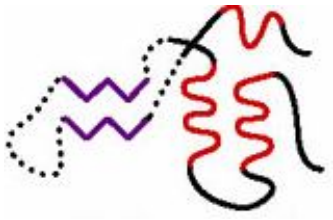
  -



БИ анализа неуређености протеина код прокариота (2)

-
  -

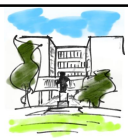




БИ анализа неуређености протеина код прокариота (3)

-
  - *експериментално или*
  - *биоинформатичком анализом*
  -
  -
  -
  -

БИ анализа неуређености протеина код прокариота (4)



-
  -
    -

Clusters of Orthologous Groups - COG

БИ анализа неуређености протеина код прокариота (5)

-
  - **Bioinformatics analysis of disordered proteins in prokaryotes**
  -
-

5. БИ истраживање података

- *Data Mining*
  -
  -
-
  -
  -

БИ истраживање података (2)

-
  - *PDB*
    -
  -
    -

БИ истраживање података (3)

-
  -
    -
  -
    -
  -
    -

БИ истраживање података (4)

-

6. Поравнавање геномских секвенци

-
  -
  -
-
-
-

Поравнавање геномских секвенци (2)

-
  -

контиге

-
  -
  -
  -
    -

Поравнавање геномских секвенци (3)

Планови за наредни период

Докторске дисертације у изради

-
-
-
-
-
-

Хвала на пажњи!

Група за биоинформатику
Математички факултет
Универзитет у Београду