

Bioinformatička analiza genomskih sekvenci

Gordana Pavlović-Lažetić

Matematički fakultet,
Beogradski univerzitet
email: gordana@matf.bg.ac.yu

Apstrakt. Biće predstavljeni rezultati istraživanja, razvoja i primene informatičkih metoda u analizi i predikciji kompozicije i strukture genomskih sekvenci. Reč je o nekim rezultatima grupe za bioinformatiku projekta "Automatsko rezonovanje i napredne obrade velikih količina podataka i teksta" (MNTR 144030) odnosno prethodnog projekta "Bioinformatika" (MNTR 1856), Matematičkog fakulteta. Prvo će biti prikazana komparativna studija izolata SARS-CoV genoma sa ciljem da se ustanovi genomski polimorfizam i evolucija varijanti. Zatim će biti predstavljena na n-gramima zasnovana metoda za analizu i predikciju segmenata u bakterijskim genomima poznatih kao genomski ostrava (GI). Ona su značajna jer mnoga od njih predstavljaju ubačene delove (inserte) koji doprinose bakterijskoj evoluciji i patogenezi. Karakterizacija i predikcija GI je izvedena na osnovu raspodele n-gramskih frekvencija i zifovske (Zipf) analize prezastupljenih i podzastupljenih n-grama, identifikovanih statistikom baziranom na Markovljevom modelu maksimalnog reda. Korišćenjem najboljih kriterijuma, dobijenih obučavanjem na soju *Escherichia coli* O157:H7 EDL933, izvršeno je predviđanje GI u 14 genoma familije Enterobacteriaceae i u 21 genomu slučajno odabranih bakterija. Predviđanja su upoređena sa rezultatima dobijenim iz dve posojeće baza genomskih i patogenih ostrva. Rezultati pokazuju da primena n-grama poboljšava i preciznost i odziv predviđanja. Biće ukratko predstavljena i neka tekuća istraživanja.

References

1. G.M. Pavlović-Lažetić, N.S. Mitić, M.V. Beljanski: n-Gram characterization of genomic islands in bacterial genomes, *Computer methods and programs in biomedicine*, 93 (2009) 241-256
2. N.S. Mitić, G.M. Pavlović-Lažetić, M.V. Beljanski: Could n-gram analysis contribute to genomic island determination?, *Journal of Biomedical Informatics*, 41(6) (2008) 936-943
3. G.M. Pavlović-Lažetić, N.S. Mitić, M.V. Beljanski: Bioinformatics analysis of SARS coronavirus genome polymorphism, *BMC Bioinformatics*. 2004; 5: 65
4. G.M. Pavlović-Lažetić, N.S. Mitić, A.M. Tomović, M.D. Pavlović, M.V. Beljanski: SARS-CoV Genome Polymorphism: A Bioinformatics Study, *Geno. Prot. Bioinfo.* Vol. 3 No. 1 February 2005, 18-35