

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 2350

Učinkovito sažimanje genoma korištenjem referentnog genoma

Juraj Radanović

Zagreb, travanj 2021.

DIPLOMSKI ZADATAK br. 2530

Pristupnik: **Juraj Radanović (0036503660)**
Studij: Računarstvo
Profil: Računarska znanost
Mentor: izv. prof. dr. sc. Mirjana Domazet-Lošo

Zadatak: **Učinkovito sažimanje genoma korištenjem referentnog genoma**

Opis zadatka:

U okviru ovoga diplomskog rada potrebno je proučiti problem sažimanja skupa genoma korištenjem referentnog genoma te postojeće metode koje se koriste za taj problem. U sklopu rada potrebno je proučiti i implementirati algoritam SCCG te vlastitu i originalnu implementaciju primijeniti na kompresiju i dekompresiju skupa genoma.

Rok za predaju rada: 28. lipnja 2021.

*Zahvaljujem se mentorici doc. dr. sc. Mirjani Domazet-Lošo na vremenu i pomoći
u izradi rada*

SADRŽAJ

1. Uvod	1
2. Zaključak	2
Literatura	3

1. Uvod

Određivanje DNK organizma, tj. sekvenciranje je postalo jeftinije i brže nego prije, zahvaljujući novim tehnologijama sekvenciranja. Jedna od svrha sekvenciranja bi bila detekcija abnormalnih DNK u organizmu koja mogu upućivati na bolest. Sekvenciranje je jedan proces, koji dovodi do pohranjivanja DNK sekvence, prijenosa i procesiranja iste. Svaki od ovih procesa zahtjeva specifične implementacije kako bi svaki proces bio što više optimiziran, memorijski i vremenski. U ovome radu se bavim pohranjivanjem genoma, konkretnije njihovom kompresijom. Što se više genoma sekvencira, to je više podataka za pohranu i treba ih moći pohraniti na efikasne načine u smislu memorije i smanjenja gubitaka podataka tijekom dekompresije. Zapis ljudskog genoma je u redovima 10^9 okteta i držanje takvih datoteka je nemoguće. Stoga je potrebno napraviti algoritam koji će efikasno pohraniti taj genom. Postoje općeniti algoritmi za kompresiju, kao npr. 7zip, ali oni su neučinkoviti za pohranu genoma iz razloga što su genomi srodnih vrsta veoma slični. Konkretno, ljudski DNK su 99.9% slični, što znači da postoje puno redundantnih podataka. U ovom radu ću implementirati algoritam koji koristi referentni genom uz ciljani genom koji se komprimira, kako bi se iskoristila neka inherentna svojstva genoma i time ostvario veći stupanj kompresije.

2. Zaključak

Zaključak.

LITERATURA

Tsachy Weissman Dmitir Pavlichin. The Desperate Quest for Genomic Compression Algorithms, 2018.

Mao Luo Min Chen Wei Shi, Jianhua Chen. High efficiency referential genome compression algorithm. 2018.

Učinkovito sažimanje genoma korištenjem referentnog genoma

Sažetak

Sažetak na hrvatskom jeziku.

Ključne riječi: Ključne riječi, odvojene zarezima.

Title

Abstract

Abstract.

Keywords: Keywords.