

Analiza genoma primjenom bioinformatičkih alata (269074)

Nositelj predmeta

[doc. dr. sc. Vladimir Brajković](#)

Opis predmeta

Kolegij Analiza genoma primjenom bioinformatičkih alata nudi praktičan uvod u tehnologiju nove generacije sekvenciranja i naprednu upotrebu računala za obradu i analizu cijele sekvence genoma te interpretaciju osnovnih analiza. Ishodi učenja fokusirani su na razvoju iskustva u primjeni bioinformatičkih alata na operativnom sustavu Linux i poznavanju formata u procesima obrade i analize sekvence genoma. Također, ne manje važan je i tehnički aspekt rukovanja velikim podatkovnim datotekama, kao što je sekvenca genoma koja po uzorku/jedinci može iznositi i 60 GB, te posljedično rad na udaljenim računalima (klasterima) s dovoljno resursa za obradu i analizu istog. Isto tako, kolegij je predložen i kao odgovor na zahtjeve tzv. genomske ere tj. razvoj i sve veću primjenu tehnologije nove generacije sekvenciranja u prirodnim znanostima, a posljedično i sve više dostupnih velikih podataka. Kolegij je od važnosti za bolje razumijevanje i implementaciju nadolazećih tehnologija sekvenciranja i analiza genoma. Spomenute metode su općenito primjenjive u području biotehnologije, neovisno o vrsti organizma koji se proučava kako bi se moglo odgovoriti na specifična istraživačka pitanja.

ECTS: **4.00**

Engleski jezik: **R1**

E-učenje: **R1**

Sati nastave: 40

Predavanja: 20

Vježbe u praktikumu: 20

Izvođač predavanja

- [doc. dr. sc. Vladimir Brajković](#)
- [Prof. dr. sc. dr. h. c. Ino Čurik](#)

Izvođač vježbi

- [doc. dr. sc. Vladimir Brajković](#)

Ocjenjivanje

Dovoljan (2): 60-70%

Dobar (3): 71-80%

Vrlo dobar (4): 81-90%

Izvrstan (5): 91-100%

Uvjeti za dobivanje potpisa

Redovito pohađanje nastave. Izostanak s predavanja ili vježbi potrebno je nadoknaditi zadatkom iz odgovarajuće nastavne jedinice.

Vrsta predmeta

- Diplomski studij / [Genetika i oplemenjivanje životinja](#) (Izborni predmet, 2. semestar, 1. godina)

Opće kompetencije

Cilj ovog kolegija je upoznati studente s primjenom tehnologije sekvenciranja nove generacije kroz razumijevanje procesa i formata u obradi i analizi sekvence genoma. Studenti će naučiti kako instalirati Linux operativni sustav na osobno računalo te osnovne naredbe u komandnoj liniji kako bi mogli instalirati i primjenjivati bioinformatičke alate u obradi i analizu sekvence genoma te interpretaciji osnovnih analiza. Također, studenti će biti upoznati s uslugom naprednog računanja na udaljenim računalima/klasterima Sveučilišnog računskog centra. Za potrebe korištenja klastera studentima će biti kreirani korisnički računi te će naučiti kako se spajati na iste te primijeniti bioinformatičke alate i naučiti kako definirati potrebne resurse za izvođenje reda poslova kroz obradu i analizu sekvenci genoma.

Oblici nastave

- Predavanja
- Vježbe

Ishodi učenja i način provjere

Ishod učenja	Način provjere
Razumijevanje tehnologije sekvenciranja nove generacije.	
Razumijevanje procesa i formata u obradi i analizi sekvence genoma.	
Primjena Linux operativnog sustava u obradi i analizi sekvence genoma.	
Primjena bioinformatičkih alata u obradi i analizi sekvence genoma.	
Razumijevanje i primjena usluge naprednog računanja Sveučilišnog računskog centra – SRCE.	
Interpretacija osnovnih analiza seta obrađenih sekvenci genoma na primjeru specifične regije ili gena na genomu.	

Način rada

Obveze studenta

Pohađati predavanja i vježbe, samostalno rješavati zadatke na vježbama, čitati odabrane znanstvene radove i prezentirati ih.

Tjedni plan nastave

1. Opis i primjena sekvenciranja nove generacije (engl. Next generation sequencing - NGS) iliti tehnologije masivnog paralelnog sekvenciranja.
2. Opis ključnih procesa u NGS obradi i analizi sekvenci cijelog genoma: a) provjera kvalitete odsekvencirane sekvence genoma (.fastq format), b) čišćenje sekvence prema parametrima kvalitete sekvenciranja, c) mapiranje i vizualizacija sekvence genoma jedinke na referentni genom, d) određivanje dubine i duljine pokrivenosti mapiranog genoma, e) filtriranje mapiranih odsječaka prema parametrima kvalitete, f) utvrđivanje i vizualizacija varijabilnih/polimornih mjesta (.vcf format - engl. variant call format), g) anotacija varijabilnih/polimornih mjesta u svrhu identifikacije varijabilnih mjesta koja su relevantna za studije bolesti, važna fenotipska svojstva, populacijsku genetiku i evoluciju.
3. Opis i vizualizacija formata u obradi sekvence genoma (.fastq, .sam, .sai, .bam, .bai, .vcf, .bed, .fasta, .fai, .dict, SRA).
4. Postavljanje računala za NGS analizu podataka: a) Opis Linux operacijskog sustava u bioinformatičkim analizama, b) instalacija Windows podsistema za Linux (engl. Windows Subsystem for Linux - WSL) na osobna računala s Windows 10 ili Windows 11 operativnim sustavima.
5. Osnovne naredbe u komandnoj liniji Linux operativnog sustava: kreiranje novog direktorija, ispis sadržaja direktorija, promjena pozicije u datotečnom sustavu, provjera trenutne putanje direktorija u kojem se nalazimo, kopiranje datoteka iz jedne lokacije u drugu, brisanje datoteka, korištenje NANO - linijski orijentiranog uređivača teksta, primjer petlje za automatizaciju ponavljajućih zadataka, pregled sadržaja datoteke i slično.
6. Instalacija i upoznavanje s bioinformatičkim alatima za obradu i analizu sekvence genoma (Conda, FastQC, IGV, BWA, BMap, Samtools, Picard, VCFtools, BCFTools, GATK4, Bedtools, SnpEff) (1. dio).
7. Instalacija i upoznavanje s bioinformatičkim alatima za obradu i analizu sekvence genoma (Conda, FastQC, IGV, BWA, BMap, Samtools, Picard, VCFtools, BCFTools, GATK4, Bedtools, SnpEff) (2. dio).
8. Primjena obrade i analize sekvenci genoma na windows podsistemu za Linux na osobnom računalu (1. dio).
9. Primjena obrade i analize sekvenci genoma na windows podsistemu za Linux na osobnom računalu (2. dio).
10. Predstavljanje usluge naprednog računanja uz opis računalnih sustava i okruženja za rješavanje resursno zahtjevnih izazova na Sveučilišnom računskom centru SRCU kao što su računalni klaster Supek, Vrančić, Padobran i Galaxy platforma.
11. Stvaranje korisničkog računa studentima za pristup računalnom klasteru Padobran i Galaxy platformu u svrhu obrade i analize sekvenci. Poduka osnovama spajanja na Padobran putem ssh protokola komandne linije, kopiranja/prijenosa sekvenci uporabom protokola SCP ili FTP te pokretanje reda poslova uz definiranje resursa potrebnih za analize.
12. Primjena obrade i analize sekvence na Galaxy platformi (1. dio).
13. Primjena obrade i analize sekvence na Galaxy platformi (2. dio).
14. Interpretacija osnovnih analiza seta obrađenih sekvenci na primjeru regije ili gena na određenom kromosomu genoma ili na mitogenomu kroz: a) anotacija varijabilnih/polimornih mjesta, b) određivanje nukleotidne raznolikosti, c) procjene udaljenosti populacije FST statistikom, d) određivanje raznolikosti haplotipova. (1. dio)
15. Interpretacija osnovnih analiza seta obrađenih sekvenci na primjeru regije ili gena na određenom kromosomu genoma ili na mitogenomu kroz: a) anotacija varijabilnih/polimornih mjesta, b) određivanje nukleotidne raznolikosti, c) procjene udaljenosti populacije FST statistikom, d) određivanje raznolikosti haplotipova. (2. dio)

Obvezna literatura

1. Prezentacije predavanja doc. dr. sc. Vladimira Brajković, prof. dr. sc. dr. h. c. Ine Čurika, dr. sc. Dalibora Hršaka, znanstvenog suradnika
2. Izbor preglednih znanstvenih radova (do 5 aktualnih radova iz područja) prema odabiru doc. dr. sc. Vladimira Brajkovića
3. Ismail H. D. (2023). Bioinformatics: A Practical Guide to Next Generation Sequencing Data Analysis. CRC Press.

Preporučena literatura

1. Kappelmann-Fenzl M. (2021). Next generation sequencing and data analysis. Springer International Publishing.
2. Elkins, K. M., Zeller, C. B. (2021). Next generation sequencing in forensic science: a primer. CRC Press.