Plan upravljanja istraživačkim podacima - Livadni procjepak (Chouardia litardierei, Hyacinthaceae) kao istraživački sustav ekološke divergencije (IP-2020-02-8099)

|  |  |  |
| --- | --- | --- |
| Opće informacije | | |
|  | Ime i prezime predlagatelja | Ivan Radosavljević |
|  | Matična organizacija | Prirodoslovno-matematički fakultet Sveučilišta u Zagrebu |
|  | Naziv projekta | Livadni procjepak (Chouardia litardierei, Hyacinthaceae) kao istraživački sustav ekološke divergencije |
|  | Upravitelj podacima | Ivan Radosavljević; ivanrad@biol.pmf.hr |
| 1. | Prikupljanje podataka i dokumentacija | |
|  | Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka) | Prikupljani podaci odnose se na sekvence/regije DNA prikazane kao primarna struktura DNA. Kroz dvije vrste sekvenciranja DNA (PacBio i HiC\_Illumina) dobivaju se informacije o redoslijedu nukleotida unutar obrađenih DNA sekvenci. Tako dobiveni primarni podaci koriste se za sklapanje genoma istraživane jedinke. Dobiveni podaci se tipično pohranjuju u fastq. formatu. S obzirom da se u dotičnom istraživanju analize provode na vrlo velikom biljnom genomu, i dobiveni podaci s kojima se barata su veliki. Veličina podataka iznosi:  1. 180 GB za PacBio sekvenciranje  2. 400 GB za HiC\_Illumina sekvenciranje  3. 4 GB za sklopljeni genom |
|  | Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka) | Svi podaci su generirani od strane međunarodno-priznatih komercijalnih servisa kod kojih se uslužno vrši sekvenciranje. Informacije o kvaliteti primarno dobivenih podataka se automatski generiraju od strane uređaja za sekvenciranje, te u tom smislu ne postoji dodatna vanjska kontrola inicijalno dobivenih podataka. Korištene metode sekvenciranja imaju standardnu točnost od >99%. U kasnijem tijeku analize i bioinformatičke obrade podataka provedena je dodatna kontrola dobivenih sekvenci upotrebom računalnih, za tu svrhu specifično dizajniranih programa, kao npr. fastQC (Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk /projects/fastqc/).  S obzirom da se u slučaju PacBio i HiC\_Illumina sekvenciranja radi o analizi na razini jedne jedinke, organizacija podataka se provela po kriteriju tipa sekvenciranja. |
|  | Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se standardi koristiti pri tumačenju podataka) | Iako su dobiveni podaci obimni, njihova je kategorizacija istovremeno iznimno jednostavna. S obzirom da se radi o podacima dobivenima sekvenciranjem ukupne DNA samo jedne jedinke (odabrane biljke), podatke je moguće podijeliti jedino u dvije grupe, prema tipu sekvenciranja (PacBio, odnosno HiC\_Illumina). Posljedično, bilo kakva popratna dokumentacija je vrlo oskudna, te se sastoji tek od osnovnih informacija: geografska lokacija populacije kojoj istraživana jedinka pripada, tip tkiva iz kojeg je DNA za potrebe sekvenciranja izolirana, te pripadnost sekvenci s obzirom na tip sekvenciranja.  Na posljetku, sklopljeni genom istraživane jedinke se sastoji od većeg broja vrlo dugačkih sklopljenih sekvenci koje su organizirane u jedinstvenu datoteku. |
| 2. | Pravna i sigurnosna pitanja | |
|  | Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci pohranjuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka)?Pri | Pri provedbi ovog projekta ne krše se etička načela. |
|  | Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka? | Podaci će se obrađivati i njima upravljati u zaštićenom nemrežnom okruženju. Do objave završnih znanstvenih publikacija kojima će biti prezentirani dobiveni rezultati, pristup podacima biti će omogućen samo najužem krugu suradnika ovog projekta. |
|  | Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka? | S obzirom da se radi o istraživanju evolucijskih procesa kod odabrane biljne svojte, ne očekuje se da će rezultat istraživanja dovesti do patenta. Do završetka istraživanja i očekivane objave znanstvenih publikacija kojima će se znanstvenoj zajednici prezentirati dobiveni rezultati, svi dobiveni podaci su pohranjeni u NCBI bazi podataka (National Center for Biotechnology Information) kao najpriznatijoj svjetskoj bazi u dotičnom znanstvenom području. U NCBI bazi podaci nisu dostupni javnosti do isteka embarga definiranog od strane unositelja podataka. Po isteku embarga na dostupnost, podaci će biti slobodno dostupni prema CC BY licenci. |
| 3. | Pohrana i čuvanje podataka | |
|  | Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (*backup*) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolažete? Kojim se procedurama koristite za sigurnosnu kopiju (*backup*)? | Podaci su pohranjeni na tri lokacije: dva vanjska tvrda diska, te navedenoj međunarodnoj bazi podataka. |
|  | Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati? | Podaci će se trajno čuvati u trenutnom formatu, kao fastq. dokumenti. |
| 4. | Dijeljenje i ponovna uporaba podataka | |
|  | Kako i gdje će se podaci dijeliti? Na kojem repozitoriju planirate dijeliti podatke? Kako će potencijalni korisnici doznati za podatke? | Dijeljenje podataka izvan kruga projektnog tima biti će omogućeno po završetku istraživanja, tj. po objavi znanstvenih publikacija proizašlih iz istraživanja. Podaci će biti dostupni u NCBI bazi gdje se i sada nalaze. S obzirom da je NCBI glavna svjetska baza za ovaj tip podataka, očekivano, svi zainteresirani su u to dobro upućeni, te će provoditi pretraživanje upravo ove baze podataka. U već objavljenoj znanstvenoj publikaciji gdje je predstavljen sklopljeni genom istraživane svojte, navedena je šifra dotičnog seta podataka pri NCBI, te će ona svima zainteresiranima služiti za jednostavno pronalaženje željenih podataka. |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Dok istraživanje ne završi i iz njega proizašli znanstveni radovi ne budu publicirani, svi dobiveni podaci će se nalaziti pod embargom zbog prioriteta uporabe. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | Da, potvrđujem. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije komercijalan). | Da, potvrđujem. |

Ref:

[1] Celjak, D., Dorotić Malič, I., Matijević, M., Poljak, Lj., Posavec K. i Turk, I.: „Istraživački podaci - što s njima?“ [Istraživački podaci - što s njima? : priručnik o upravljanju istraživačkim podacima | Digitalni repozitorij Srca (unizg.hr)](https://repozitorij.srce.unizg.hr/islandora/object/srce:327)