Plan upravljanja istraživačkim podacima

|  |
| --- |
| Opće informacije |
|  | Ime i prezime predlagatelja  | Martina Šeruga Musić |
|  | Matična organizacija | Prirodoslovno-matematički fakultet |
|  | Naziv projekta | IPS-2019-04, Strategije patogenosti fitoplazmi: efektori, faktori virulencije i pokretni genetički elementi |
|  | Upravitelj podacima | Prof. dr. sc. Martina Šeruga Musić (martina.seruga.music@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) | Tijekom projekta generirat ćemo 6 kategorija podataka:1. Fotografije
2. Fotografije elektroforetskih gelova, SDS-PAGE gelova, Western i Southern blot membrana

Fotografije elektroforetskih agaroznih gelova, SDS-PAGE gelova, Western i Southern blot membrana bit će generirane na uređaju Amersham ImageQuant 800 ili uređaju DigiGeneious te pohranjene u *.jpg* ili .tiff formatu. Ove će fotografije zauzeti otprilike 10 GB memorije. 1. Fotografije biljaka vrsta *Nicotiana benthamiana* i *Arabidopsis thaliana* – zdravih te tretiranih i uzgojenih molekularno-biološkim metodama

U svrhu fenotipske evaluacije utjecaja fitoplazmatskih efektora, biljke će biti transformirane metodom „floral dip“ te će se sakupljati i selektirati transformirano sjeme. Ovako uzgojene biljke će se fotografirati u svrhu mjerenja morfoloških parametara. Ove fotografije zauzeti će otprilike 1 GB memorije.1. Fotografije snimljene na konfokalnom fluorescencijskom mikroskopu

U svrhu određivanja interakcijskih partnera između fitoplazmatskih efektora i biljnih proteina, snimati će se slike na konfokalnom fluorescencijskom mikroskopu Leica na Institutu Ruđer Bošković. Ove fotografije zauzeti će otprilike 35 GB memorije.1. Numeričke tablice
2. Popisi uzoraka zajedno s pisanim bilješkama s terena i izmjerenim koncentracijama ekstrahirane DNA, RNA i plazmida bit će spremljeni u obliku *excel* pregledne tablice. Također, podaci o koncentraciji nukleinskih kiselina mjereni tijekom analitičkih protokola (Nanodrop, Qubit) za Nanopore sekvenciranje bit će zabilježeni u obliku excel tablica te dugoročno pohranjeni u csv formatu. Tablice će zauzeti otprilike 5 GB memorije.
3. Analize genomskih sekvenci biti će djelomično obrađene i pohranjene u *excel* pregledne tablice. Tablice će zauzeti otprilike 5 GB memorije.
4. Sirove i obrađene sekvence, te popratni podaci

Podaci dobiveni Nanopore ili Sanger sekvenciranjem bit će pohranjeni u *fasta* ili *fastq* formatima. Prilikom bioinformatičke obrade sirovih sekvenci generirat ćemo veliki broj popratnih dokumenata u različitim formatima *(sam, bam, txt, gz, csv*, ...). Draft sekvence genoma te ostale sekvence bit će pohranjene u bazi podataka GenBank. Ovaj set podataka zahtjeva najveću količinu memorije procijenjenu na minimalno 1TB.1. Laboratorijski dnevnici i registratori (bilješke i sl.)

Sve bilješke istraživača nastale tokom izvedbi laboratorijskih protokola i bioinformatičke obrade podataka bit će zapisane u laboratorijskim dnevnicima i u „*Word“* dokumentima. Protokoli unaprijed pripremljeni ili stvarani tijekom laboratorijskog rada bit će pohranjeni u registratoru projekta. 1. Ostatak biljnog materijala

Ostaci biljnog materijala čuvat će se također u laboratoriju kao zamrznute porcije biljnog materijala te i izolirani genetički materijal u ledenici (-20 ili -80°C) laboratorija. Izolirano i osušeno sjeme čuvati će se u laboratorijskoj hladnjači na 4°C.1. Bakterijske kulture

Bakterijske kulture pohraniti će se u glicerolu te sustavu za pohranjivanje Microbank™ u laboratorijskoj ledenici (-80°C). |
| b | Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka) | Svakom uzorku bit će dodijeljena određena oznaka te će dodijeljenjim oznakama biti obilježeni pohranjeni sirovi uzorci, fotografije s terena, izolirana DNA, RNA i plazmidi kao i sekvence.Ostale tablice, popisi i fotografije bit će imenovani na način da sadrže kratak opis sadržaja, akronim projekta (PathoPhyto), datum stvaranja ili zadnje izmjene te inicijale istraživača koji dokument kreira.Budući da jedna od aktivnosti projekta i sekvenciranje, sastavljanje i anotacija genoma fitoplazmi, konačne sekvence bit će detaljno bioinformatički obrađene te procesuirane kroz razne dostupne i besplatne bioinformatičke programe kao i program Geneious software v10.2.6 (Biomatters Ltd., Auckland, New Zealand) za koji je kupljena licenca u sklopu projekta. Dio osiguranja kvalitete je provjera sekvenci od strane znanstvenika koji uređuju NCBI-bazu podataka sekvenci GenBank jer se za svaku sekvencu koja se obrađuje predviđa unošenje u GenBank zbog dobivanja pristupnog broja (Accession Number) pri čemu ona prolazi eksterne i objektivne postupke kontrole kvalitete. |
|  | 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) | Primarni način na koji ćemo ustupiti podatke javno biti će u znanstvenim publikacijama. Izvori i način prikupljanja i obrade podataka biti će detaljno objašnjeni u dijelu o materijalima i metodama, a rezultati će biti prikazani slikama, tablicama, plotovima te pojašnjeni tekstualno i legendama. Sekvence genoma bit će pohranjene online u Nacionalni centar za biotehnološke informacije (NCBI) s detaljima prikupljanja, obrade i s anotacijama. Jedinstveni, dodijeljeni broj sekvence (tzv. accession number) bit će objavljen u sklopu publikacije te na taj način javno dostupan za pregledavanje i korištenje. Tijekom projekta bit će u izradi i diplomski rad studenta te doktorski rad poslijediplomanda iz teme projekta na hrvatskom jeziku uz još detaljnije objašnjene postupke dolaska do podataka te uz objavljene dodatne rezultate koji nisu postali dio objave u znanstvenom radu. |
| 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)? | Budući da ne radimo s ljudskim ispitanicima ili životinjskim uzorcima, nemamo osjetljive podatke. Podaci će biti dostupni samo istraživačima uključenima u projekt do trenutka objave ili znanstvenog skupa.Prilikom pisanja sažetka ili cjelokupnog rada, a prije same objave sve istraživače uključene u rad obavijestit ćemo o planovima o koautorstvu. Oni će moći sudjelovati u analizi podataka i stvaranju rukopisa, te će moći pregledati, dopuniti ili izmijeniti tekst prije no što se slože s konačnom verzijom tj. objavljenim materijalom. |
|  | 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?  | Budući da nemamo osjetljivih podataka, ne postoje drugi rizici pri regulaciji pristupa podacima osim da netko objavi rezultate prije nas. S ciljem izbjegavanja takvog rizika sirovi podaci će se obrađivati, a potom i pohranjivati isključivo tako da su dostupni unutar projektnog tima do trenutka objave. Po objavi znanstvenog rada, kako bi naši rezultati bili ponovljivi i provjerljivi poslati ćemo potrebne informacije svakom zainteresiranom znanstveniku. |
|  | 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? | Do trenutka javne objave rada kada su osigurana autorska prava svih osoba koje su doprinijele radu, podaci pripadaju nositeljici projekta, izv. prof. dr. sc. Martini Šeruga Musić te je potrebno njeno dopuštenje za uporabu podataka. Sve sekvence koje objavimo biti će javno dostupne na korištenje šire znanstvene javnosti u svrhu usporedbi i dodatnih analiza. |
| 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 će biti podijeljeni s obzirom na istraživačke ciljeve projekta te prema tome grupirani u 4 cjeline: (i) Istraživanja raznolikosti i polimorfizma efektora u genomu fitoplazme '*Candidatus* Phytoplasma solani', (ii) Ispitivanje ekspresije potencijalnih efektora i njihove interakcije s biljnim domaćinom (iii) Ispitivanje interakcije fitoplazmatskih proteina Stamp s proteinima kukaca-vektora, (iv) Istraživanje raznolikost potencijalnih mobilnih jedinica (PMUs) u genomu fitoplazme '*Candidatus* Phytoplasma solani'. Podaci će istovremeno biti pohranjeni na računalima poslijedoktoranda dr. sc. Brune Polaka, doktorandice mag. biol. exp. Marine Drčelić i izv. dr. sc. prof. Martine Šeruge Musić, a sigurnosne kopije podataka biti će čuvane u zajedničkoj oblak pohrani Dropbox servisa i na vanjskom tvrdom disku (18TB).  |
|  | Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati? | Podaci će se čuvati na računalima poslijedoktoranda dr. sc. Brune Polaka, doktorandice mag. biol. exp. Marine Drčelić i izv. dr. sc. prof. Martine Šeruge Musić , u zajedničkoj oblak pohrani Dropbox servisa i na vanjskom tvrdom disku (18TB). na U repozitoriju SRCA (https://puh.srce.hr) pohranit ćemo konačne rezultate projekta. Formati dugoročno pohranjenih podataka podrazumijevaju .pdf, .csv, ..xlsx, fasta, .fastq formate podataka. |
| 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 prije objave rezultata ograničeno je na članove projektnog tima. Po objavi podataka, projektni rezultati postat će dostupni, na zahtjev. Dobiveni podaci o rezultatima će biti predstavljeni na znanstvenim i stručnim skupovima te u relevantnim znanstvenim publikacijama peer-review časopisima (Q1 ili Q2 prema WoS-u). Podaci o sekvencama (što su ujedno i ključni konačni rezultati ovog projekta) dostupni su brzo nakon pohrane u NCBI-u, pri čemu trenutak kada postaju dostupni ovisi isključivo o administratorima NCBI-a. |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Nemamo takvih podataka.  |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | Potvrđujemo da ćemo po završetku projekta finalne verzije ključnih rezultata pohraniti na repozitoriju PUH. |
|  | 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). | Potvrđujemo da ćemo po završetku projekta finalne verzije ključnih rezultata pohraniti na repozitoriju PUH. |

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%3A327)