

Plan upravljanja istraživačkim podacima - AltRecPath

Zahradka, Ksenija

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2025**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:241:643337>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-02-05**



Repository / Repozitorij:

[Fulir DATA - Ruđer Bošković Institute Research Data Repository](#)

PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)

Opće informacije		
	Ime i prezime predlagatelja	Ksenija Zahradka
	Matična organizacija	Institut Ruđer Bošković
	Naziv projekta	Alternativni putevi rekombinacije i popravka DNA kod bakterija <i>Escherichia coli</i> i <i>Deinococcus radiodurans</i>
	Upravitelj podacima	Jelena Repar, Davor Zahradka, Ksenija Zahradka
1. Prikupljanje podataka i dokumentacija		
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite formate, vrste i opseg svih podataka s kojima ćete raditi, a ne samo krajnji skup podataka koji će biti rezultat istraživanja)	<p>Podaci iz osnovnih analiza bakterijskih sojeva:</p> <ul style="list-style-type: none"> - brzina rasta i vijabilnost - preživljenje kod raznih doza oštećivača DNA, npr. UV i gama zračenja - transdukcijaska rekombinacija <p>Vrsta podataka: numerički (očitanja sposobnosti stvaranja kolonija (CFA) i optičke gustoće (OD)) Format: podaci će biti sakupljeni u xlsx tablice Volumen podataka: ~10 Mb</p> <p>Podaci iz fluorescencijske i konfokalne mikroskopije:</p> <p>Vrsta podataka: Fotografije Format: jpg, tiff Volumen podataka: ~20 Gb</p> <p>Podaci iz elektroforeze u pulsirajućem polju (PFGE):</p> <p>Vrsta podataka: Fotografije Format: tiff Volumen podataka: ~1 Gb</p> <p>Podaci iz sekvenciranja cjelog bakterijskog genoma:</p> <p>Vrsta podataka: tekst (očitanja DNA sekvenci) Format: fastq ili bam Volumen podataka: ~100 Gb</p>
	Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete, načine organiziranja podataka te alate i instrumente kojima ćete se koristiti za	<p>Podaci iz osnovnih analiza bakterijskih sojeva:</p> <p>Podaci će biti sakupljeni kao očitavanja CFA i OD tijekom normalnog rasta i nakon različitih tretmana (npr. UV i gama zračenje). Konstruirat ćemo krivulje preživljenja za različite doze oštećivača DNA. Protokol i metodologija sakupljanja podataka je već opisana i objavljena u znanstvenim radovima (npr. Buljubašić et al 2019, DNA repair, DOI:10.1016/j.dnarep.2019.102670). Bakterijski sojevi divljeg tipa bit će korišteni kao kontrole prilikom karakterizacije mutantnih sojeva. Ekvivalentno procesirani, ne-tretirani uzorci bit će korišteni kao kontrole gdje bude potrebno. Barem tri neovisna mjerenja po dozi i po soju će biti sakupljena. Tablice će biti organizirane tako da se tretmani, njihove doze i sojevi bakterija mogu jasno razlikovati.</p> <p>Podaci iz fluorescencijske i konfokalne mikroskopije:</p> <p>Podaci će biti prikupljeni fluorescentnim invertnim mikroskopom Zeiss Axiovert 35 i konfokalnim laserskim skenirajućim mikroskopom Leica TCS SP8 X. Protokol i metodologija prikupljanja podataka prethodno su opisani (npr.</p>

	prikupljanje i obradu)	<p>Buljubašić et al 2019, DNA repair, DOI:10.1016/j.dnarep.2019.102670). Uzorci divljeg tipa koristit će se kao kontrole pri analizi mutantnih sojeva. Snimat će se slike koje obuhvaćaju najmanje 1000 stanica po eksperimentalnom uvjetu i bakterijskom soju kako bi se omogućila kvantifikacija različitih fenotipova, kao što su stanice bez DNA i filamentozne stanice. Snimljene slike će biti odvojene u mape imenovane prema bakterijskom soju i eksperimentalnom uvjetu. <i>In silico</i> kvantifikacija fenotipova izvršit će se prema potrebi, korištenjem programa ImageJ.</p> <p>Podaci iz elektroforeze u pulsirajućem polju (PFGE): Podaci će biti prikupljeni pomoću sustava elektroforeze CHEF-DR III (Bio-Rad). Protokol i metodologija prikupljanja podataka prethodno su opisani (Repar et al 2010, DNA repair, DOI:10.1016/j.dnarep.2010.08.003). Uzorci divljeg tipa koristit će se kao kontrole pri analizi mutantnih sojeva. Neozračeni uzorci koristit će se kao kontrola pri praćenju kinetike popravka DNA nakon gama zračenja. Snimljene slike će biti imenovane prema bakterijskom soju i eksperimentalnom uvjetu.</p> <p>Podaci iz sekvenciranja cjelog bakterijskog genoma: Očitavanja DNA sekvenci genoma dobit će se korištenjem vanjske usluge sekvenciranja (npr. Macrogen) tehnologijom sekvenciranja Illumina ili PacBio. Očitavanja DNA će biti sravnjena s referentnim genomima E. coli K-12 MG1655 i D. radiodurans R1 preuzetima iz baze podataka NCBI Refseq. Za to ćemo upotrijebiti program za sravnjivanje kao što je BWA-MEM ili Minimap2. Sravnjenja će se koristiti za (i) pronalaženje nepodudarnosti između sekvenciranih sojeva i referentnog genoma, tj. otkrivanje mutacija i preuređivanja genoma, pomoću programa DNAdiff iz paketa Mummer ili sličnog, i (ii) za istraživanje diferencijalne replikacije kromosomskih regija koristeći programski paket samtools (opcija depth) ili sličan. Kako bi se otkrila diferencijalna replikacija kromosomskih regija i izbjegli lažno pozitivni/lažno negativni rezultati koji bi mogli proizaći iz diferencijalnog sekvenciranja, stanice stacionarne faze također će se sekvencirati i koristiti kao kontrola. Duboko sekvenciranje koristit će se kako bi se izbjegle pogreške uzrokovane stohastičkim varijacijama u pokrivenosti genoma očitavanjima DNA sekvenci. Ispitat će se i usporediti višestruki pristupi otkrivanju genomskih rearanžmana. Prilikom analize genomskih rearanžmana u (hipotetiziramo) pogrešno popravljenim genomima ozračenih, "mrtvih" stanica soja D. radiodurans recA, očekuje se da će stohastička priroda takvih pogrešaka proizvesti mnoštvo različitih promjena u različitim stanicama. Kako bismo razlikovali pogreške prilikom popravka genoma od pogrešaka u sekvenciranju, kao kontrolu ćemo koristiti sekvence divljeg tipa dobivenih u ekvivalentnim uvjetima (npr. mjerenjem broja rearanžmana po bp). Nazivi datoteka i njihovo odvajanje u mape prema vrsti eksperimenta i vrsti analize osigurati će točno praćenje podataka.</p>
	Koji ćete dokumentaciju i metapodatke izraditi osim podataka? (dokumentacija mora sadržavati informacije i standarde potrebne korisnicima kako bi mogli samostalno čitati i interpretirati podatke u budućnosti, primjerice, kodne knjige, <i>ReadMe</i> datoteke i sl.)	<p>Za osnovne analize bakterijskih sojeva, PFGE i mikroskopiju slijedit ćemo naše standardne laboratorijske protokole koji su prethodno razvijeni i objavljeni (npr. vidi Buljubašić et al 2019, DNA repair, DOI:10.1016/j.dnarep.2019.102670, Repar et al 2010, DNA popravak, DOI:10.1016/j.dnarep.2010.08.003). Ove reference bit će navedene prema potrebi u tekstualnom dokumentu koji prati podatke.</p> <p>Za neobrađene podatke bit će osiguran tekstualni dokument s datumima i podacima o bakterijskim sojevima i eksperimentalnim uvjetima, kao i opisi strukture podataka i sadržaja datoteka.</p> <p>Za analize sekvenciranja izradit će se tekstualni dokument s detaljnim protokolom. U slučaju pisanja vlastitih programskih skripti (u programskom jeziku java ili R), osim u slučaju kad služe za osnovno rukovanje ili promjenu formata podataka, skripte će biti deponirane na github (github.com) nakon objavljivanja odgovarajućeg znanstvenog rada, a poveznice će biti navedene prema potrebi u tekstualnom popratnom dokumentu uz podatke.</p>
2.	Pravna i sigurnosna pitanja	

<p>Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci obrađuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka, navesti metode anonimizacije podataka)?</p>	<p>Nema ograničenja jer radimo s reprezentativnim bakterijskim sojevima koji se mogu nabaviti iz javnih i komercijalnih repozitorija bakterijskih sojeva kao što je DHMZ (https://www.dsmz.de/collection/catalogue/microorganisms/catalogue).</p>
<p>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?</p>	<p>Podaci će biti pohranjeni lokalno u više kopija, na tvrdim diskovima računala istraživača i vanjskim tvrdim diskovima. Do objave rezultata u znanstvenim radovima pristup podacima imat će samo istraživači koji rade na projektu. Kod objave znanstvenih radova, standardna praksa u struci je da se podaci putem objave učine dostupnima kao prilozi radu (eng. „supplementary data“). Podaci nastali dubokim sekvenciranjem bakterijskih sojeva bit će pohranjeni u bazi podataka NCBI Sequence Read Archive (SRA), te će postati javno dostupni nakon objavljivanja odgovarajućeg znanstvenog rada.</p>
<p>Kako ćete upravljati zaštitom autorskih prava i drugog intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka?</p>	<p>Podaci i analize bit će objavljeni u znanstvenim radovima i njihovim priložima. Gdje je to financijski moguće, birat će se znanstveni časopisi s otvorenim pristupom.</p>
<p>3. Pohrana i čuvanje podataka</p>	
<p>Kako će radne verzije podataka biti pohranjene tijekom projekta? Kako će se napraviti sigurnosne kopije tih podataka (<i>backup</i>)? Koja je očekivana količina</p>	<p>Radne verzije podataka bit će spremljene u različite mape ili datoteke imenovane prema verziji. Sigurnosne kopije svih podataka radit će se pomoću vanjskih tvrdih diskova. Očekivana količina podataka koja će se prikupiti i čuvati tijekom projekta (sirovi podaci + analize) je ~1 Tb.</p>

	podataka koja će se prikupiti i čuvati tijekom projekta (izraženo u MB/GB/TB)?	
	Kako će se završne verzije podataka dugotrajno pohraniti i čuvati (i nakon završetka projekta)? U kojim će se formatima čuvati podaci? Koja je očekivana količina podataka koja će se trajno pohraniti (izraženo u MB/GB/TB)?	Podaci i njihove analize postat će javno dostupni putem objavljenih znanstvenih radova. Lokalno, unutar laboratorija, pohranit ćemo podatke u najmanje dvije kopije: podatke dobivene osnovnim analizama bakterijskih sojeva u .xlsx formatu, te mikroskopske i PFGE slike u originalnim (jpg, tiff) formatima. Podaci iz sekvenciranja bit će poslani i pohranjeni u bazi podataka NCBI Sequence Read Archive (SRA), kako bi bili javno dostupni nakon objavljivanja odgovarajućeg znanstvenog rada. NCBI SRA bazu podataka sigurnosno je kopirana i putem Europskog instituta za bioinformatiku (EBI) i Japanske banke podataka DNA (DDBJ). Nakon projekta ,oko 130 Gb podataka bit će trajno pohranjeno u našem laboratoriju, dostupno na zahtjev.
4.	Dijeljenje i ponovna uporaba podataka	
	Kako i gdje će se podaci dijeliti? Koji repozitorij će se koristiti za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke?	Podaci će se dijeliti kroz znanstvene radove i njihove priloge. Podaci iz sekvenciranja dijelit će se putem NCBI repozitorija. Potencijalni korisnici će o podacima saznati putem znanstvenih publikacija ili pretraživanjem odgovarajućih repozitorija.
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	Nije primjenjivo.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Da.
	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.