Sekvenciranje tehnologijom nanopora i sklapanje genoma ogulinske špiljske spužvice Eunapius subterraneus

Glavaš, Dunja

Master's thesis / Diplomski rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:530982

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

Sveučilište u Zagrebu
Prirodoslovno - matematički fakultet
Biološki odsjek

Dunja Glavaš

Sekvenciranje tehnologijom nanopora i sklapanje genoma ogulinske špiljske spužvice *Eunapius subterraneus*

Diplomski rad

Zagreb, 2018.

Ovaj rad je izrađen u Grupi za bioinformatiku na Zavodu za molekularnu biologiju Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu i Laboratoriju za molekularnu genetiku na Zavodu za molekularnu biologiju Instituta Ruđer Bošković, pod vodstvom prof. dr.sc. Kristiana Vlahovičeka i dr.sc. Helene Ćetković. Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistra molekularne biologije.

Zahvaljujem se obitelji na podršci, Maji Kuzman na pomoći, i mentorima prof. dr.sc. Kristianu Vlahovičeku i dr.sc. Heleni Ćetković na strpljenju i vodstvu.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Diplomski rad

SEKVENCIRANJE TEHNOLOGIJOM NANOPORA I SKLAPANJE GENOMA OGULINSKE ŠPILJSKE SPUŽVICE *EUNAPIUS SUBTERRANEUS*

Dunja Glavaš
Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Spužve su zbog ranog odvajanja od ostalih skupina dobar model za proučavanje rane evolucije i razvoja životinja, ali do sada je sekvenciran genom samo jedne vrste. Osim iz evolucijskog i razvojnog gledišta, spužve su zanimljive i kao modelni organizmi za proučavanje prilagodbe okolišu. Poseban interes predstavlja vrsta *Eunapius subterraneus*, endem krškog područja Hrvatske i jedina poznata slatkovodna stigobiontna spužva na svijetu. Sekvenciranje i sklapanje genoma nije jednostavno zbog prisutnosti brojnih mikroorganizama u i na tijelu spužvi. Sekvenciranjem tehnologijom nanopora dobivaju se dugački očitani sljedovi s većom razinom pogreške nego kod tehnologija sekvenciranja druge generacije. U hibridnom sklapanju genoma dugački sljedovi treće generacije se koriste za spajanje i zatvaranje prekida između neprekinutih sljedova dobivenih sklapanjem kratkih i točnih sljedova druge generacije. Sekvencirala sam genom spužve *E. subterraneus* pomoću uređaja Oxford Nanopore Technologies MinION. Provela sam hibridno sklapanje genoma pomoću sljedova dobivenih sekvenciranjem tehnologijama Illumina i Oxford Nanopore Technologies koristeći programe temeljene na pohlepnom algoritmu i metodi de Bruijnovog grafa. Procijenila sam kvalitetu sklopljenih genoma. Veći broj dugačkih očitanih sljedova korišten u sklapanju rezultira duljim neprekinutim sljedovima. U sklopljenom genomu su prisutne brojne kontaminacije i potrebno je daljnje sekvenciranje kako bi se genom spužve *E. subterraneus* sklopio u cijelosti.

(42 stranice, 12 slika, 14 tablica, 55 literaturnih navoda, jezik izvornika: hrvatski)
Rad je pohranjen u Središnjoj biološkoj knjižnici.
Ključne riječi: tehnologije sekvenciranja treće generacije, hibridno sklapanje genoma, de Bruijnov graf, pohlepni algoritam
Voditelji: prof. dr. sc. Kristian Vlahoviček
dr.sc. Helena Ćetković, znanstvena savjetnica, IRB
Ocjenitelji: prof. dr. sc. Kristian Vlahoviček
izv. prof. dr. sc. Damjan Franjević
prof. dr. sc. Zlatko Liber
Zamjena: doc.dr.sc. Rosa Karlić
Rad prihvaćen:

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Graduation Thesis

SEQUENCING BY NANOPORE TECHNOLOGY AND GENOME ASSEMBLY OF ENDEMIC CAVE SPONGE *EUNAPIUS SUBTERRANEUS*

Dunja Glavaš
Rooseveltov trg 6, 10000 Zagreb, Croatia

Sponges are good model organisms for early Metazoan evolution and development because of their early separation from other groups of animals. However, genomic research on sponges is slow because of the scarce availability of genomic information - only one sponge species has had its genome completely sequenced so far. Sponges are also good models for environmental adaptation studies, in which species Eunapius subterraneus is of special interest. *E. subterraneus* is the only known freshwater stygobiont sponge in the world and it is endemic to karst region of Croatia. Sequencing and assembly of Poriferan genomes is not a trivial task due to the presence of numerous endo- and exosymbiontic microorganisms. Sequencing by Nanopore technology yields long reads with higher error rates in comparison to sequencing technologies of second generation. Hybrid approach to genome assembly uses long third-generation reads for closing gaps and joining contigs assembled from accurate but relatively short second-generation reads. I sequenced the genome of *E. subterraneus* using Oxford Nanopore Technologies MinION sequencer. I assembled reads sequenced on Illumina and Oxford Nanopore Technologies platforms using hybrid approach and then assesed the quality of assembled genomes. Programs I used for the assembly are based on greedy algorithm and de Bruijn graph method. Using more long third-generation reads resulted in longer contigs. Although I obtained long contigs, my assembly contains many contaminations. Further sequencing is needed in order to complete the genome of *E. subterraneus.*

(42 pages, 12 figures, 14 tables, 55 references, original in: Croatian)
Thesis deposited in the Central Biological library
Key words: third generation sequencing, hybrid genome assembly, de Bruijn graph, greedy algorithm
Supervisors: Professor Kristian Vlahoviček, PhD
Helena Ćetković, PhD, Institute Rudjer Boskovic
Reviewers: Professor Kristian Vlahoviček, PhD
Assoc. Professor Damjan Franjević, PhD
Professor Zlatko Liber, PhD
Substitution: Asst. Professor Rosa Karlić, PhD
Thesis accepted:

Sadržaj

1. UVOD 1

1.1. Koljeno Porifera (spužve) 1

1.1.1. Građa tijela i razmnožavanje 1

1.1.2. Taksonomska podjela 2

1.1.3. Istraživanja genoma 3

1.1.4. Vrsta Eunapius subterraneus 4

1.2. Sekvenciranje 6

1.2.1. Sekvenciranje tehnologijom Illumina 7

1.2.2. Sekvenciranje tehnologijom nanopora 8

1.3. Sklapanje genoma 10

1.3.1. Predobrada očitanih sljedova 11

1.3.2. Sklapanje genoma de novo metodom de Bruijnovog grafa 11

1.3.3. Hibridno sklapanje genoma 13

1.4. Procjena kvalitete sklopljenih sljedova 14

2. CILJ ISTRAŽIVANJA 15

3. MATERIJALI I METODE 16

3.1. Izolacija genomske DNA 16

3.1.1. Agarozna gel-elektroforeza 16

3.1.2. Određivanje koncentracije DNA 16

3.2. Potvrda vrste pomoću genetičkog biljega ITS2 16

3.2.1. Lančana reakcija polimerazom 17

3.2.2. Izolacija PCR produkta iz gela 18

3.2.3. Sekvenciranje Sangerovom metodom 18

3.3. Priprema knjižnica za sekvenciranje tehnologijom nanopora 19

3.4. Sekvenciranje tehnologijom nanopora na uređaju ONT MinION 20

3.5. Korištene knjižnice 20

3.6. Predobrada nukleotidnih sljedova 21

3.6.1. Predobrada sljedova dobivenih sekvenciranjem tehnologijom Illumina 21

3.6.2. Predobrada sljedova dobivenih sekvenciranjem tehnologijom nanopora 22

3.7. Sklapanje genoma 22
3.8. Analiza sklopljenih sljedova 22
4. REZULTATI 23
4.1. Izolacija genomske DNA 23
4.2. Potvrda vrste pomoću genetičkog biljega ITS2 23
4.3. Priprema knjižnica za sekvenciranje tehnologijom nanopora 24
4.4. Sekvenciranje tehnologijom nanopora na uređaju ONT MinION 25
4.5. Predobrada nukleotidnih sljedova 25
4.5.1. Predobrada sljedova dobivenih sekvenciranjem tehnologijom Illumina 25
4.5.2. Predobrada sljedova dobivenih sekvenciranjem tehnologijom nanopora 25
4.6. Sklapanje genoma 27
4.7. Analiza sklopljenih sljedova 28
4.7.1. Procjena dovršenosti genoma programom BUSCO 28
4.7.2. Usporedba sklopljenih sljedova s bazom podataka proteinskih sljedova 30
4.7.3. Usporedba sklopljenih sljedova s genomom vrste Amphimedon queenslandica 32
5. RASPRAVA 34
6. ZAKLJUČCI 37
7. LITERATURA 38
8. PRILOZI 42

Kratice

DNA	deoksiribonukleinska kiselina
dNTP	deoksiribonukleotid-trifosfat
kb	kilobaza, 10^3 baza
Mb	megabaza, 10^6 baza
NFW	voda bez nukleaza
nt	nukleotid
ONT	Oxford Nanopore Technologies
pb	par baza
PCR	lančana reakcija polimerazom

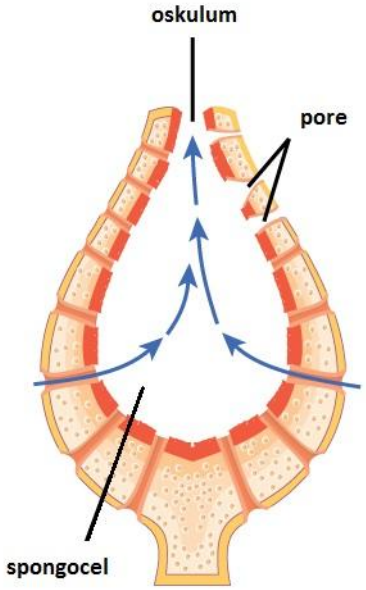
1. UVOD

1.1. Koljeno Porifera (spužve)

Koljeno Porifera vjerojatno predstavlja evolucijski najstariju skupinu višestaničnih životinja, odnosno prvu skupinu koja se odvojila od zajedničkog pretka svih višestaničnih životinja (Metazoa) (Pick i sur. 2010.; Pisani i sur. 2015.). Spužve su građene od mnoštva pora i kanala koji čine sustav za filtriranje i prehranu, što je u skladu s njihovim sjedilačkim načinom života. Gotovo sve su morske, izuzev jedne porodice Spongillidae, čiji predstavnici žive u kopnenim vodama. Stanice koje izgrađuju tijelo spužvi su specijalizirane ali nema udruživanja u prava tkiva. Zbog te jedinstvene karakteristike, vrste iz ove skupine dobar su model za proučavanje evolucijske pojave viših razina stanične organizacije.

1.1.1. Građa tijela i razmnožavanje

Veličina spužvi varira ovisno o vrsti od nekoliko milimetara do više metara u promjeru. Većina ima asimetrična tijela, samo one jednostavnije građe imaju radijalnu simetriju. Brojni manji otvori na površini (pore ili ostije) omogućavaju ulazak vode u šuplju unutrašnjost spužve (spongocel) gdje se odvija apsorpcija kisika i hranjivih čestica. Iskorištena voda i gamete izbacuju se kroz veće otvore (oskulume) (Slika 1).



SLIKA 1. Shema građe tijela jednostavne spužve. Plave strelice označavaju smjer protoka vode. Preuzeto i prerađeno sa https://archive.cnx.org/contents/c985edf6-66b3-40b4-aabe-bec43c407362@5/sponges-and-cnidarians.

Na histološkoj razini u tijelu spužve razlikuju se tri sloja. Vanjski epidermalni sloj sastoji se od plosnatih poligonalnih stanica pinakocita koje mogu imati sposobnost kontrakcije. Pore su otvori u pinakocitnom sloju, a kod jednostavnijih spužvi mogu biti građene od specijaliziranih cjevastih stanica porocita. Pinakocite nalazimo i u endodermalnom sloju koji je okrenut prema unutarnjim šupljinama. Najznačajnije stanice u endodermalnom sloju su hoanociti, vrčaste stanice s bičem čije pokretanje osigurava protok vode kroz organizam. Između epidermalnog i endodermalnog sloja nalazi se rahliji unutarnji sloj mezenhima. On sadrži skelet, želatinozni matriks mezogleju i putujuće stanice amebocite. Različiti tipovi amebocita imaju uloge u izgradnji sincicijalne mreže, pohrani pigmenata ili pričuvnih tvari, izgradnji skeleta (skleroblasti i spongoblasti) i razmnožavanju (arheociti). Tipični skelet spužve sastavljen je od iglica mineralnog porijekla i vlakana proteina spongina. Iglice se nazivaju spikule ili sklere i izgrađene su od silicija ili kalcijevog karbonata. Neke spužve uopće nemaju spikule već im se skelet sastoji samo od spongina (Matoničkin 1990.).

Spolne stanice se razvijaju iz hoanocita ili amebocita. Velik broj vrsta spužvi su hermafroditi s vremenski odvojenom proizvodnjom spermija i jajnih stanica. Zreli spermiji se otpuštaju u vodu i dolaze u doticaj s površinom „ženske“ spužve gdje ih preuzimaju hoanociti i potom transportiraju do jajne stanice u unutrašnjosti. Zigota se razvija u trepetljikavu ličinku koja izlazi iz tijela, te pliva ili puže dok ne pronađe prikladnu podlogu na koju će se pričvrstiti.

Nespolno, spužve se razmnožavaju pupanjem i gemulama. Kod pupanja se dio tijela odvaja i nastavlja život kao zasebna jedinka u blizini (tako nastaju kolonije). Gemule su posebno formirana malena okruglasta tijela koja sadrže veći broj arheocita unutar vanjskog zaštitnog sloja. Zaštitni sloj omogućava preživljavanje nepovoljnih uvjeta, a u povoljnim uvjetima se iz svake gemule može razviti odrasla jedinka. Brojne slatkovodne spužve na taj način preživljavaju zimu (Matoničkin 1990.).

1.1.2. Taksonomska podjela

Koljeno Porifera dijeli se na četiri razreda: Demospongiae, Homoscleromorpha, Calcarea i Hexactinellida (Van Soest i sur. 2012.). Najviše današnjih vrsta pripada razredu **Demospongiae** (kremenorožnjače). Skelet se može sastojati od silicijskih spikula, organskih vlakana, kombinacije toga dvoga ili ne mora uopće biti prisutan. Žive u gotovo svim morskim staništima, od intertidalne do abisalne zone. Sve poznate vrste slatkovodnih spužvi pripadaju ovom razredu. **Homoscleromorpha** su mala grupa isključivo morskih spužvi. Zbog morfoloških sličnosti dugo se smatralo da čine podrazred unutar razreda Demospongiae, no nedavna molekularna istraživanja pokazala su da ih treba klasificirati kao zaseban razred

(Gazave i sur. 2011.). Skelet izostaje ili je sastavljen od silicijskih spikula. Većina poznatih vrsta obitava u mračnim staništima u plitkim vodama (primjerice špilje). Skelet **Calcarea** (vapnenjača) izgrađuju karbonatne iglice. Jedinke najčešće veličinom ne prelaze 10 centimetara. Žive isključivo u slanoj vodi, a najviše ih se nalazi u plitkim područjima tropskih mora. **Hexactinellida** (staklače) imaju karakteristične silicijske šesterozrakaste sklere. Stanice su često spojene u sincicij, zbog čega spužva nema sposobnost kontrakcije ali se signali relativno brzo provode kroz cijelo tijelo. Isključivo su morski organizmi i uglavnom žive u dubokim vodama (od 200 do preko 6000 metara dubine), a mogu formirati i grebene (Van Soest i sur. 2012.).

Klasifikacija na nižim taksonomskim razinama podložna je čestim promjenama. Jednostavna građa spužvi znači manje očitih morfoloških razlika između vrsta i teže određivanje srodstva. U novije vrijeme, brojni rezultati genetičkih istraživanja dokazuju ili barem sugeriraju da je potrebna promjena dosadašnje klasifikacije pojedinih vrsta (primjerice Harcet i sur. 2010.a).

1.1.3. Istraživanja genoma

Genetika spužvi relativno je slabo istraženo područje. U potpunosti je sekvenciran genom samo jedne vrste, *Amphimedon queenslandica* (Srivastava i sur. 2010.). Mitohondrijski genomi dovršeni su za 56 vrsta (NCBI Organelle Genome Resources 2018.). Napredak u ovom području usporavaju mala količina otprije dostupnih podataka i specifične poteškoće kod istraživanja genetičkog materijala spužvi. Naime spužve često žive u simbiozi s algama, gljivama, bakterijama i drugim organizmima (Taylor i sur. 2007.; Webster i sur. 2009.), a analiza genoma *A. queenslandica* dokazala je i horizontalni transfer gena između spužve i prokariota (Conaco i sur. 2016.). Zbog toga je ponekad teško razdvojiti DNA domaćina i simbiontskog organizma, bilo u laboratoriju ili *in silico*. S druge strane, duga koegzistencija s brojnim simbiontima čini spužve dobrim modelom za proučavanje molekularne osnove simbioze.

Sekvenciranje je koristan alat za identifikaciju vrsta. Od 2012. godine u tijeku je Sponge Barcoding Project - pothvat kojem je krajnji cilj kreirati bazu podataka genetičkih markera svih poznatih vrsta spužvi (Vargas i sur. 2012.). Opširnije i dublje sekvenciranje genoma pojedinih vrsta pomoglo je riješiti neke od problema u sistematici (Gazave i sur. 2011.). Genetika spužvi zanimljiva je i s evolucijskog gledišta. Spužve su prva skupina koja se odvojila od zajedničkog pretka svih višestaničnih životinja (Pick i sur. 2010.), pa su današnje vrste dobar model za proučavanje genetičke osnove najranije evolucije Metazoa. Komparativnom analizom genoma *A. queenslandica* pronađeni su geni odgovorni za ključna

obilježja višestaničnosti (Srivastava i sur. 2010.) i složeniju građu tijela viših životinja (Adamska i sur. 2007.). Riesgo i suradnici (2014.) pokazali su da predstavnici svih razreda spužvi dijele velik broj gena s drugim višestaničnim životinjama, između ostalog i gene uključene u biološke funkcije koje spužve nemaju (npr. imunološko prepoznavanje). Analiza transkriptoma vrsta *Suberites domuncula* i *Lubomirskia baicalensis* otkrila je neočekivanu kompleksnost genoma s obzirom na izostanak pravih tkiva (Harcet i sur. 2010.b). Rezultati upućuju na to da je zajednički predak svih višestaničnih životinja bio genetski kompleksniji nego što se mislilo, i da je do usložnjavanja genoma došlo prije pojave bilateralne simetrije. Podatci dobiveni daljnjim sekvenciranjem genoma predstavnika koljena Porifera mogli bi pomoći u dobivanju odgovora na brojna otvorena pitanja u evolucijskoj biologiji.

1.1.4. Vrsta Eunapius subterraneus

Vrstu *Eunapius subterraneus* (hrv. ogulinska špiljska spužvica) opisali su Sket i Velikonja 1984. godine. Jedina je poznata stigobiontska slatkovodna spužva na svijetu. Obitava isključivo u podzemnim vodama krškog područja oko Ogulina i na Velikoj Kapeli (Bedek i sur. 2008.). Živi pričvršćena na stijene u potpunom mraku ali tolerira i malu količinu svjetlosti. Pronađena je na rasponu dubina od 0 do 23 metra, pri temperaturama između 7 i 11 stupnjeva (Bilandžija i sur. 2007.).

Oblik tijela može biti jajolik s izbrazdanom površinom (Slika 2) ili tanjurast bez brazdi. Jedinke jajolike morfologije velike su između 1 i 8 centimetara. One tanjurastog oblika imaju široku bazu kojom se prihvaćaju na podlogu i izbočinu na sredini s oskulumom na vrhu. Tijelo spužve je rahlo, mekano i bez pigmenata. Skelet je sastavljen od pojedinačnih vrlo velikih i lagano zakrivljenih spikula, međusobno povezanih s malo spongina (Bedek i sur. 2008.). Gemule su žućkasto-smeđe, okruglaste i nalaze se pričvršćene na stijenu blizu baze spužve (Bilandžija i sur. 2007.).

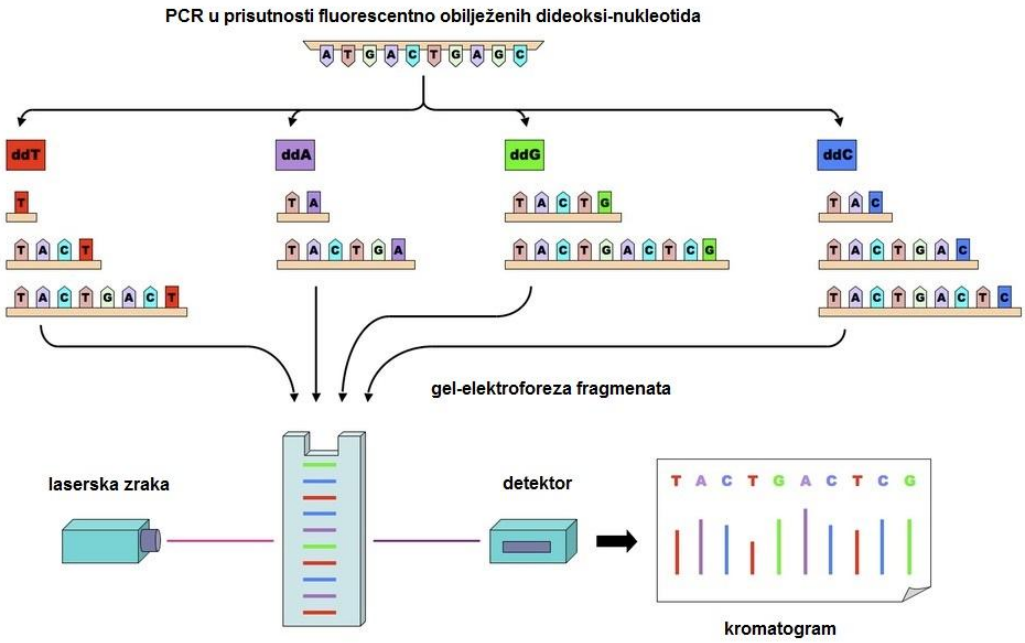


SLIKA 2. *E. subterraneus* jajolikog oblika u prirodnom staništu na lokalitetu Tounjčica. Preuzeto sa https://www.hbsd.hr/SkupineZ_spuzvica_eng.html.

Prema podatcima dostupnim na stranici World Porifera Database (2018.), vrsta *E. subterraneus* pripada razredu Demospongiae, podrazredu Heteroscleromorpha, redu Spongillida, obitelji Spongillidae. Međutim, pripadnost rodu *Eunapius* dovedena je u pitanje istraživanjem koje su proveli Harcet i suradnici (2010.a). Rezultati filogenetičke analize temeljene na tri genetička biljega (18S rDNA, ITS2 i COI) pokazuju da je *E. subterraneus* bliži srodnik slatkovodnim spužvama *Ephydatia muelleri* i *Lubomirska baikalenskiis* nego ostalim vrstama iz roda Eunapius. Točna klasifikacija ostaje neriješeno pitanje. Dovršena sekvenca mitohondrijskog genoma potvrđuje smještaj vrste u monofiletsku skupinu s ostalim slatkovodnim spužvama, a visok stupanj homologije unutar skupine sugerira da je do odvajanja slatkovodnih spužvi došlo nedavno na evolucijskoj skali (Pleše i sur. 2011.).

1.2. Sekvenciranje

Sekvenciranje je postupak određivanja slijeda dušičnih baza u molekuli DNA. Do sada razvijene tehnologije možemo podijeliti u tri generacije. Prva generacija je počela razvojem Sangerove metode (Sanger i sur. 1977.). Ona se temelji na umnožavanju DNA u prisutnosti ireverzibilnih terminatora sinteze, te kasnijem razdvajanju dobivenih fragmenata po veličini. Provode se četiri odvojene reakcije elongacije DNA u prisutnosti sva četiri nukleotida i po jednog dideoksi-nukleotida koji onemogućava nastavak sinteze. Po završetku se reakcijske smjese pročišćavaju i podvrgavaju elektroforezi, a s dobivenog gela se iščitava redoslijed baza u molekuli DNA. Pojava kapilarne elektroforeze i fluorescentno obilježenih terminatora sinteze DNA omogućila je spajanje četiri reakcije u jednu, strojnu detekciju signala i time automatizaciju procesa (Slika 3) (Ansorge i sur. 1986.; Swerdlow i Gesteland 1990.). Prosječna duljina očitanih sljedova (engl. *reads*) nešto je manja od 1 kb (Heather i Chain 2016.).



SLIKA 3. Shematski prikaz sekvenciranja automatiziranom Sangerovom metodom. Preuzeto i prerađeno sa http://www.vce.bioninja.com.au/aos-3-heredity/molecular-biology-technique/sequencing.html.

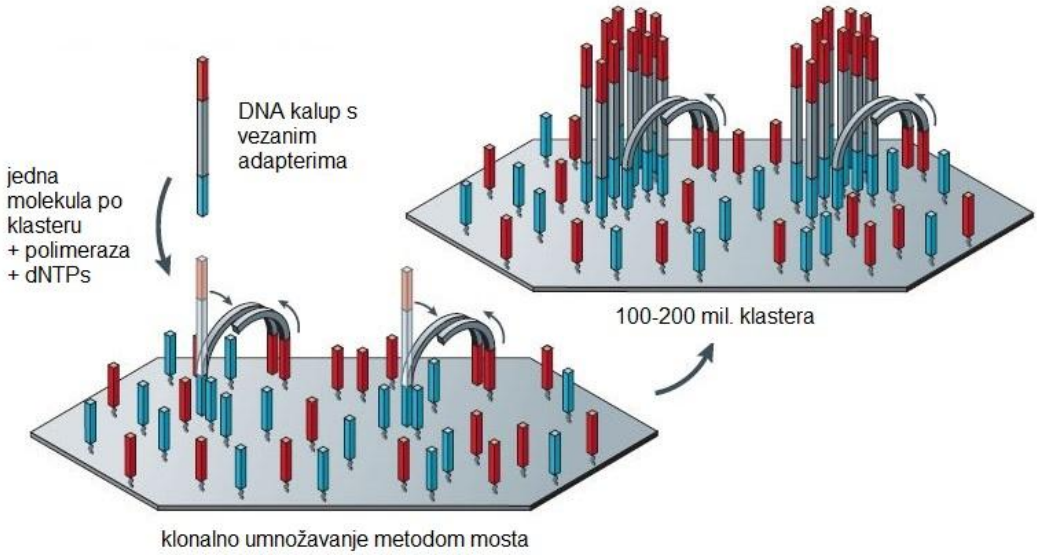
Tehnologije druge generacije omogućile su paralelno sekvenciranje velikog broja molekula odjednom. Najznačajnije platforme koriste jedan od tri principa rada: tehnologiju reverzibilnih terminatora (Illumina/Solexa), pirosekvenciranje (454/Roche) ili sekvenciranje ligacijom (ABI SOLiD) (Kchouk i sur. 2017.). Zajedničke značajke su umnožavanje uzorka lančanom reakcijom polimeraze (engl. *polymerase chain reaction*, PCR) prije reakcije sekvenciranja, fiksiranje molekula uzorka na čvrstu podlogu (kuglice ili pločicu) tijekom

reakcije, i direktna detekcija signala bez potrebe za elektroforezom. U odnosu na prvu generaciju metode su puno brže i jeftinije ali produciraju kraće očitane sljedove i više pogrešaka (Verma i sur. 2017.).

Treću generaciju karakterizira očitanje signala u realnom vremenu bez zaustavljanja sinteze (engl. *real-time sequencing*) i mogućnost sekvenciranja bez prethodnog umnožavanja uzorka (engl. *single-molecule sequencing*). Time se povećava brzina reakcije i smanjuje vrijeme potrebno za pripremu uzorka (Kchouk i sur. 2017.). Izostanak umnožavanja znači da se u uzorak ne uvode pogreške tipične za PCR, a predstavlja prednost i kod kvantitativnih metoda zbog očuvanja originalnih odnosa količine pojedinih fragmenata. Helicos Biosciences, Pacific Biosciences i Oxford Nanopore Technologies razvili su platforme treće generacije (Heather i Chain 2016.). Glavna prednost ovih tehnologija je velika duljina očitanih sljedova, a glavni nedostatak za sada još uvijek mala pouzdanost odnosno relativno velik udio pogrešno pročitanih baza.

1.2.1. Sekvenciranje tehnologijom Illumina

Sve platforme koje je Illumina razvila temelje se na sekvenciranju sintezom uz cikličku reverzibilnu terminaciju. Na fragmente željene DNA ligiraju se adapteri koji su komplementarni oligonukleotidima fiksiranim na čvrstoj površini protočne ćelije (engl. *flow cell*). Tako pripremljene molekule se vežu na protočnu ćeliju u određenoj gustoći. Slijedi klonalno umnožavanje metodom mosta (engl. *bridge amplification*), pri čemu nastaju lokalizirani „otočići“ odnosno klasteri sastavljeni od puno kopija istog fragmenta (Slika 4). Taj korak je potreban kako bi kasnije intenzitet signala bio dovoljan za detekciju. Sekvenciranje se provodi DNA polimerazom u prisutnosti univerzalnih početnica i modificiranih deoksiribonukleotida (dNTP). Četiri različita dNTP obilježeni su svaki sa svojom fluorescentnom bojom, a fluorofor ujedno služi i kao 3'-kapa koja zaustavlja daljnju sintezu DNA. Prvi korak u svakom ciklusu sekvenciranja je dodavanje sva četiri dNTP odjednom. Nakon ugradnje jednog dNTP po molekuli sinteza je zaustavljena, nevezani dNTP se ispiru i očitava se signal (boja) na svakom klasteru. Slijedi cijepanje fluorofora i ponovno ispiranje, nakon čega se ciklus ponavlja kako bi se ustanovio identitet iduće baze (Bentley i sur. 2008.).



SLIKA 4. Shema klonalnog umnožavanja metodom mosta. Preuzeto i prerađeno iz Metzker 2010.

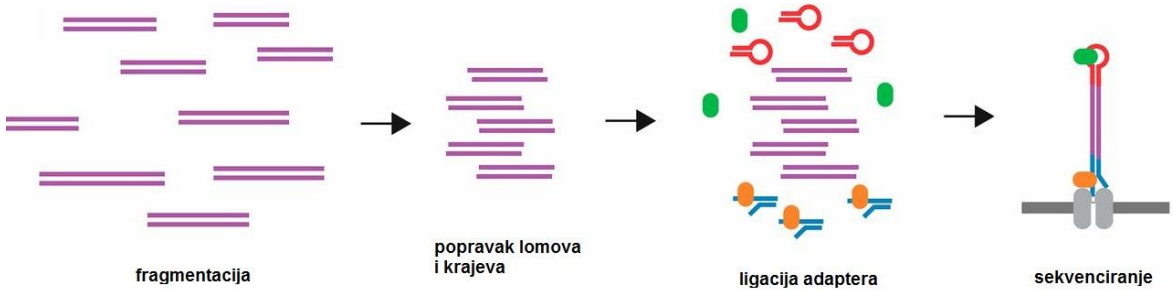
Procjenjuje se da je Illumina najčešće korištena tehnologija sekvenciranja (Hodzic i sur. 2017.), vjerojatno zbog relativno niske cijene i velike brzine. Kao i kod svih drugih metoda u kojima se nukleotidi ugrađuju jedan po jedan, duljina očitanih sljedova ograničena je pojavom tzv. *dephasinga* - pošto ugradnja dNTP ima uspješnost manju od 100%, nakon određenog broja ciklusa pozadinski šum postane prejak i više nije moguće očitati stvarni signal. Illumininom tehnologijom moguće je dobiti očitane sljedove prosječne duljine od 35 do 200 pb, ovisno o platformi. Ukupna pogreška je otprilike 1%, pri čemu su najčešći tip pogreške supstitucije (Kchouk i sur. 2017.).

1.2.2. Sekvenciranje tehnologijom nanopora

Tehnologija sekvenciranja nanoporama počiva na činjenici da prolazak molekule DNA kroz poru u membrani utječe na strujanje iona kroz tu poru. Ukoliko je membrana pod naponom, promjena ionske struje odražava se u mjerljivoj promjeni potencijala na membrani. Oligomer nukleotida koji prolazi kroz poru uzrokuje niz promjena potencijala, a uzorak tih promjena je specifičan ovisno o sastavu i redoslijedu baza u oligomeru. Konstantnim praćenjem promjena potencijala dobiva se zapis u realnom vremenu, kojeg specijalizirani program (tzv. *basecaller*) „prevodi“ u redoslijed baza (Wang i sur. 2014.).

Priprema uzorka ne zahtijeva obilježavanje ni amplifikaciju DNA. Izolirana DNA se fragmentira da se osigura željena veličina molekula, no sam korak fragmentacije nije nužan. Zatim se popravljaju lomovi, fosforiliraju 5'-krajevi i adeniliraju 3'-krajevi. Sljedeći i ključni korak u pripremi je ligacija adaptera na dobivene dvolančane molekule. Na jedan kraj molekule ligira se adapter u obliku ukosnice (HP adapter) koji kovalentno spaja dva

komplementarna lanca, a na drugi kraj dvolančani adapter (Y adapter) čiji 5'-kraj prvi ulazi u poru. Oba adaptera imaju vezna mjesta za motorne proteine. Protein E5 veže se na Y adapter, razdvaja komplementarne lance i pomaže ulazak u poru. Protein E3 na HP adapteru usporava brzinu prolaska kroz poru kad sekvenciranje dođe do ukosnice (Wei i Williams 2016.) (Slika 5). Sekvenciranje oba lanca DNA korisno je jer smanjuje razinu pogreške u očitanim sljedovima, koje onda zovemo 2D sljedovi. 1D sljedovi su niže kvalitete i dobiju se kad je pročitan samo jedan lanac iz komplementarnog para.



SLIKA 5. Priprema uzorka i sekvenciranje tehnologijom nanopora. Preuzeto i prerađeno iz Wei i Williams 2016.

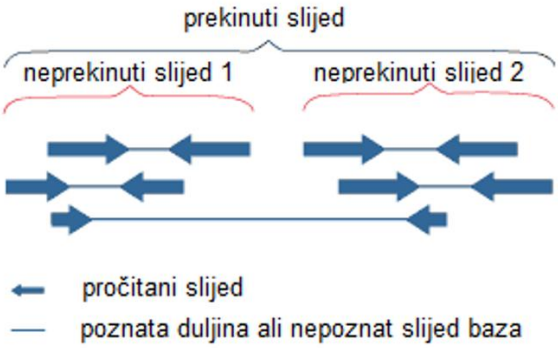
Oxford Nanopore Technologies (ONT) je za sad jedina kompanija koja je razvila ovakvu platformu. 2014. godine je predstavljen uređaj ONT MinION, vrlo mali sekvenator koji teoretski može generirati očitane sljedove duljine preko 150 kb. Protočna ćelija uređaja sadrži polimernu membranu velikog električnog otpora u koju su umetnute biološke nanopore. U prvoj generaciji kao nanopora je služio modificirani porin MspA iz mikobakterija. Novije verzije koriste modificirani CsgG, bakterijski kanalni protein za sekreciju amiloida, koji osigurava bolju protočnost (de Lannoy i sur. 2017.). Prednosti ove tehnologije su niska cijena, vrlo dugački očitani sljedovi, te mogućnost pristupanja podatcima u realnom vremenu dok se sekvenciranje odvija. ONT MinION je prikladan i za terenski rad zbog veličine uređaja i relativno jednostavne pripreme uzorka. Glavni nedostatak je vrlo visoka razina pogreške od barem 12% (3% pogrešno pročitane baze, 4% insercije, 5% delecije) (Kchouk i sur. 2017.) za starije verzije, zbog čega je potrebno ispravljanje sljedova prije sklapanja genoma. Nova generacija (verzije od R9 na dalje) razvijena je s ciljem povećanja protočnosti i smanjenja šuma u podatcima, kako bi se mogao zaobići računalno skup korak ispravljanja. Unaprijeđen je i proces prevođenja električnog signala u redoslijed baza. Prijašnji programi za prevođenje (primjerice Metrichor i DeepNano (Boža i sur.2017.)) radili su pod pretpostavkom da se u nanopori u bilo kojem trenutku nalazi pentamer nukleotida. Nakon što je ustanovljeno da određeni sljedovi baza i neke sekundarne strukture molekule DNA drugačije utječu na protok iona kroz poru, razvijen je novi program Albacore koji kod prevođenja razmatra varijabilni broj nukleotida u pori (de Lannoy i sur. 2017.).

1.3. Sklapanje genoma

Sklapanje genoma je rekonstrukcija točne sekvence DNA iz zbira nasumično uzorkovanih fragmenata (Narzisi i Mishra 2011.) - drugim riječima, postupak kojim se sekvencirani fragmenti slažu ispravnim redoslijedom i u točnom broju kopija kako bi dobiveni slijed odgovarao onom u stvarnoj molekuli DNA. Već složeni genom srodne vrste može se iskoristiti kao pomoć kod sklapanja i znatno olakšati proces, što se naziva komparativnim slaganjem genoma. U slučajevima kad to nije moguće, novi genom se mora sklapati samo na osnovi očitanih sljedova (*de novo*).

Sklapanje genoma *de novo* je matematički problem sa razinom kompleksnosti NP. To znači da nije moguće pronaći učinkovito računalno rješenje (Pop 2009.). Stoga se programi za sklapanje genoma oslanjaju na aproksimacije i heurističke metode. Postoje dva osnovna pristupa: algoritmi ili provode analizu nizova (engl. *string-based*) ili prikazuju i analiziraju podatke u obliku grafa (engl. *graph-based*) (Narzisi i Mishra 2011.). Najznačajniji predstavnik prve grupe je tzv. pohlepni (engl. *greedy*) algoritam, a programi iz druge grupe koriste OLC metodu (od engl. *overlap-layout-consensus*) ili de Bruijnov graf (Miller i sur. 2010.; Ye i sur. 2012.).

Za dobivanje potpunog i neprekinutog slijeda cijelog eukariotskog genoma u pravilu je potrebno više postupaka sekvenciranja i nekoliko iteracija sklapanja. Rezultat jednog sklapanja genoma obično je skup dužih neprekinutih sljedova (engl. *contigs*) koji ne pokrivaju cijeli genom već se između njih nalaze prekidi (engl. *gaps*). Postupak određivanja relativnih pozicija i međusobnih udaljenosti neprekinutih sljedova naziva se sklapanje okvira (engl. *scaffolding*), a njime se dobivaju tzv. prekinuti sljedovi (engl. *scaffolds*) (Slika 6). Za sklapanje okvira mogu se koristiti informacije o uparenosti sljedova dobivene sekvenciranjem uparenih krajeva s većom duljinom inserta (engl. *mate-pair*), optičko mapiranje, ili dugački sljedovi dobiveni sekvenciranjem treće generacije.



SLIKA 6. Sklapanje okvira. Preuzeto i prerađeno sa https://genome.jgi.doe.gov/portal/help/scaffolds.jsf.

1.3.1. Predobrada očitanih sljedova

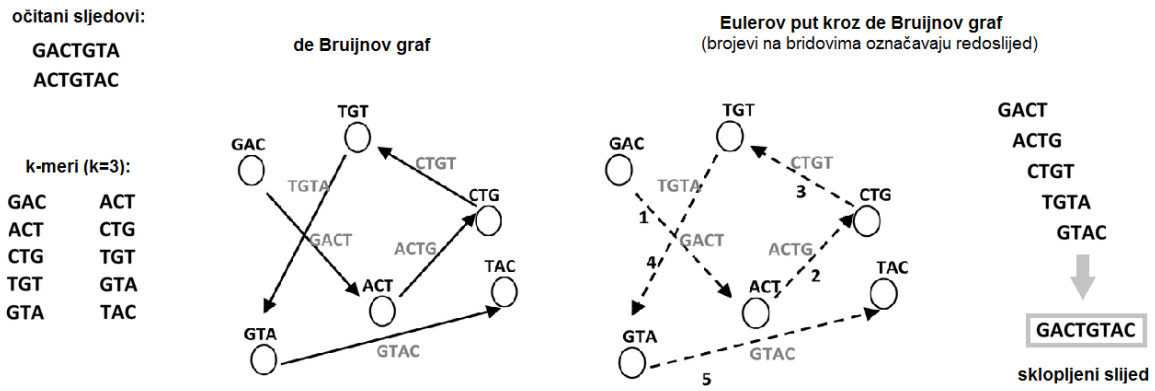
Prije sklapanja potrebno je obraditi sirove podatke dobivene sekvenciranjem. Što se tiče sljedova dobivenih sekvenciranjem tehnologijama druge generacije, očiti primjer je uklanjanje adapterskih sekvenci. Ukoliko je proveden postupak uparenog sekvenciranja krajeva (engl. *paired-end sequencing*), dobivaju su dva seta podataka koji se prije sklapanja mogu spojiti u jedan. Predobradom se može smanjiti i ukupna pogreška, na način da se uklone dijelovi slijeda s malom pouzdanošću jer je za njih statistički vjerojatno da predstavljaju pogrešno očitane nukleotide. Još jedan način detekcije pogrešaka je normalizacija na temelju dubine pokrivenosti (engl. *sequencing depth*), gdje se istražuju susjedni sljedovi s velikom razlikom u učestalosti. Ako se nakon promjene sumnjivih baza u jednom slijedu učestalost tog slijeda poveća tako da bude sličnija onoj susjednog slijeda, smatra se da su te baze bile krivo očitane.

Predobrada sljedova dobivenih tehnologijama treće generacije je u odnosu na drugu generaciju puno zahtjevnija, zbog veće duljine sljedova i veće ukupne pogreške. Čest pristup je da se kratki točni sljedovi druge generacije sravnjuju s dugačkim sljedovima treće generacije i koriste kao referenca za ispravljanje pogrešaka u dugačkim sljedovima. Program Nanocorr provodi takvo sravnjenje između sljedova dobivenih sekvenciranjem tehnologijama Illumina i ONT MinION koristeći alat BLAST (od engl. *Basic Local Alignment Search Tool*) (Altschul i sur. 1990.). Nepouzdana i prekratka sravnjenja te nesravnjeni ONT sljedovi se odbacuju, nakon čega algoritam LIS (engl. *longest increasing subsequence*) pronalazi najbolji skup sravnjenih kratkih sljedova za svaki dugački slijed. Konsenzusna sekvenca se izračunava pomoću algoritma pbdagcon (Goodwin i sur. 2015.). Program LoRDEC namijenjen je ispravljanju dugačkih sljedova treće generacije s većim brojem insercija i delecija. On sastavlja de Bruijnov graf od kratkih sljedova (detaljnije opisano u sljedećem poglavlju) i ispravlja netočne regije u dugačkim sljedovima traženjem prikladnog puta kroz graf za svaki dugački slijed pojedinačno (Salmela i Rivals 2014.).

1.3.2. Sklapanje genoma de novo metodom de Bruijnovog grafa

Osnovni algoritam metode de Bruijnovog grafa započinje sastavljanjem liste svih sljedova duljine k (tzv. k-meri) koji su prisutni u ulaznim podatcima. Odabrani k je uvijek manji od duljine pojedinačnog očitanog slijeda. Time se dobiva skup podataka u kojem su svi fragmenti jednake duljine i savršeno uzorkuju genom (svaka baza u sekvenci odgovara početku točno jednog k-mera). Zatim se konstruira graf u kojem su k-meri predstavljeni čvorovima, a čvorovi koji dijele sekvencu duljine (k-1) su međusobno povezani bridovima

(Slika 7). Problem sklapanja genoma se time svodi na problem pronalaženja Eulerovog puta kroz graf (put koji prolazi jednom svakim bridom) (Compeau i sur. 2011.; Pop 2009.).



SLIKA 7. Konstrukcija de Bruijnovog grafa i Eulerov put. Preuzeto i prerađeno iz El-Metwally i sur. 2013.

Ovaj pristup štedi memoriju jer se svaki različiti *k*-mer u listu sprema samo jednom, uz informaciju o tome koliko puta se pojavljuje u očitanim sljedovima. Cijela operacija sklapanja se provodi transformacijama grafa, čime se izbjegava eksplicitni izračun preklapanja između fragmenata što je računalno najskuplji korak u drugim algoritmima. S druge strane, razlamanjem očitanih sljedova na manje dijelove gube se informacije o povezanosti međusobno udaljenijih područja. Također, svaka pogreška u sekvenciranju stvara veći broj „lažnih“ *k*-mera, a broj mogućih putova kroz graf raste eksponencijalno s brojem čvorova. Sličan problem se javlja kod gotovo identičnih ponavljanja. Odabir veličine *k* jako utječe na konačni izgled grafa: manje vrijednosti čine graf zapetljanijim jer se veći broj ponavljanja prikazuje istim čvorovima, dok se kod većih detektira manje preklapanja pa graf postaje izlomljen (Bankevich i sur. 2012.).

Osnovni algoritam se može modificirati kako bi se riješili gore navedeni problemi. Informacije izgubljene razlamanjem vraćaju se na način da program prvo zabilježi dijelove puta koji odgovaraju originalnim očitanim sljedovima i potom traži „superput“ kroz cijeli graf koji sadržava te dijelove. Upareni de Bruijnov graf (PDBG) (Medvedev i sur. 2011.) uključuje i informacije o parovima očitanih sljedova dobivenim uparenim sekvenciranjem krajeva. Problemi koji se javljaju zbog pogrešaka u sekvenciranju mogu se djelomično riješiti korekcijom sljedova prije sklapanja, a daljnji napredak u tom području postignut je uvođenjem *A*-Bruijnovog grafa (Pevzner i sur. 2004.). Riječ je o generalizaciji de Bruijnovog grafa koja prepoznaje gotovo identične sljedove i „lijepi“ ih zajedno, čime se smanjuje broj čvorova i mogućih putova kroz graf. *Multisized* de Bruijnovi grafovi omogućavaju variranje vrijednosti *k* ovisno o pokrivenosti određene regije, pri čemu niži *k* daju bolje rezultate u regijama s manjom pokrivenošću.

Program SPAdes (Bankevich i sur. 2012.) implementira sve navedene modifikacije i jednu dodatnu. Za razliku od većine ostalih programa, njegov PDBG modul može iskoristiti informacije o uparenosti i kad udaljenost između članova para nije konstantna u cijelom skupu podataka - drugim riječima, kad se kod različitih parova upareni očitani sljedovi nalaze na različitoj međusobnoj udaljenosti.

1.3.3. Hibridno sklapanje genoma

Hibridno sklapanje genoma je postupak sklapanja u kojem se koriste očitani sljedovi dobiveni sekvenciranjem na više različitih platformi. Različite metode sekvenciranja imaju različite prednosti i nedostatke. Kombiniranje podataka iz više izvora omogućava kvalitetnije sklapanje jer se relativni nedostatci jedne platforme (primjerice kratka očitanja ili visok stupanj pogreške) mogu nadoknaditi ili ispraviti podatcima s druge.

Česta kombinacija kod hibridnog sklapanja su kratki sljedovi dobiveni tehnologijom Illumina i dugački sljedovi dobiveni sekvenciranjem treće generacije. Sekvenciranje tehnologijom Illumina daje točnija očitanja i bolju pokrivenost (veći broj očitanih sljedova po jedinici sekvence), pa se ti sljedovi koriste za ispravljanje baza koje su na drugoj platformi pogrešno identificirane. Veća duljina očitanih sljedova treće generacije pak omogućava sklapanje neprekinutih sljedova koji su duži nego što bi to bio slučaj kad bi se u sklapanju koristili samo kratki sljedovi (Heather i Chain 2016.; Lu i sur. 2016.). Kad se suprotni krajevi jednog dugačkog slijeda mogu sravniti s dva neprekinuta sklopljena slijeda (engl. *contigs*), može se odrediti veličina prekida između tih sljedova. Ukoliko je neki prekid pokriven s više dugačkih sljedova može se pronaći konsenzusna sekvenca i zatvoriti prekid. Dugački sljedovi korisni su i za rješavanje ponavljajućih regija u genomu.

Programi za hibridno sklapanje genoma (primjerice MaSuRCA, SPAdes i ALPACA) počinju sklapanjem kratkih sljedova i potom dodaju dugačke kako bi produljili neprekinute sljedove, zatvorili prekide i riješili ponavljanja. Razvijeni su i programi koji kao ulazne podatke uzimaju dugačke očitane sljedove treće generacije i neprekinute sljedove dobivene prethodnim sklapanjem genoma od kratkih sljedova. Potonji se često temelje na pohlepnom algoritmu: prvo se uzimaju u obzir najpouzdaniji podatci, a ostatak se iterativno uklapa dok ne dođe do konflikta s prethodno konstruiranim slijedom. Takav algoritam je brz i jednostavan ali podložan (lažnim) lokalnim maksimumima kod sklapanja genoma od kraćih sljedova. Pri sklapanju dugačkih očitanih sljedova treće generacije i prethodno sklopljenih neprekinutih sljedova taj efekt nije toliko izražen zbog manjeg broja i veće duljine sljedova s kojima se radi. Jedan od takvih programa je npScarf (Cao i sur. 2017.). npScarf prvo detektira

jedinstvene neprekinute sljedove, odnosno one koji ne predstavljaju repetitivne sekvence u genomu. Zatim sravnjuje dugačke ONT sljedove s jedinstvenim neprekinutim sljedovima koristeći program BWA-MEM (*Burrow-Wheeler Aligner*). Sravnjenje jednog ONT slijeda s dva jedinstvena neprekinuta slijeda određuje njihovu povezanost i međusobnu udaljenost. Pohlepni algoritam spaja jedinstvene neprekinute sljedove u veće neprekinute sljedove (prvo parovi s najvećom kvalitetom sravnjenja, ostatak se dodaje iterativno), a ponavljajući neprekinuti sljedovi se koriste za zatvaranje prekida.

1.4. Procjena kvalitete sklopljenih sljedova

Kvaliteta sklapanja nekog genoma određuje se na temelju duljine i točnosti dobivenih neprekinutih i prekinutih sljedova. Najčešće korištene statističke mjere su veličina najkraćeg i najduljeg neprekinutog slijeda, i N50. N50 je duljina najmanjeg neprekinutog slijeda koji zajedno sa svim jednako dugačkim i duljim neprekinutim sljedovima čini barem 50% ukupne sklopljene sekvence. Programom QUAST (Gurevich i sur. 2013.) mogu se izračunati ove i druge statističke vrijednosti.

Za procjenu točnosti sklopljenih sljedova mogu se koristiti informacije o genomu istog ili srodnih organizama. BLAST je alat koji kao ulazni podatak uzima neki slijed nukleotida ili aminokiselina i potom u bazi podataka traži slične ili identične sljedove. Slijed koji se ispituje se uspoređuje sa sljedovima iz baze podataka metodom lokalnog maksimuma, sravnjenja se boduju, a kao rezultat se vraćaju sravnjenja koja imaju veći broj bodova od granične vrijednosti koju određuje korisnik. Program DIAMOND također uspoređuje sljedove na sličan način, ali je posebno razvijen za sravnjivanje velikih setova podataka sa proteinskim sekvencama i u tom slučaju puno brži od BLAST-a (Buchfink i sur. 2015.). Jedna od mjera dovršenosti sklopljenog genoma je udio sklopljenih gena. OrthoDB (Waterhouse i sur. 2013.) je baza podataka ortologa koja, osim samih sljedova gena, uključuje i informacije o evoluciji i srodstvenim odnosima. Program BUSCO (od engl. *Benchmarking Universal Single-Copy Orthologs*) (Simão i sur. 2015.) u sklopljenom genomu traži sljedove gena za koje se, s obzirom na tijek evolucije, može očekivati da su prisutni u ispitivanom genomu u jednoj kopiji.

2. CILJ ISTRAŽIVANJA

Cilj istraživanja je sekvencirati genom ogulinske špiljske spužvice *Eunapius subterraneus* tehnologijom Oxford Nanopore Technologies MinION, dobivene podatke iskoristiti u sklapanju genoma (uz već dostupne genomske knjižnice dobivene tehnologijama Illumina i ONT MinION) te usporediti kvalitetu nekoliko sklapanja napravljenih uz upotrebu različitih programa za sklapanje, prevođenje baza i ispravljanje pogrešaka.

3. MATERIJALI I METODE

3.1. Izolacija genomske DNA

Koristila sam uzorak spužve *E. subterraneus* prikupljen na lokalitetu špilja Tounjčica (blizu Ogulina) i konzerviran u etilnom alkoholu. Odvojila sam manji komad tijela spužve i tri puta ga isprala s destiliranom vodom po par minuta na sobnoj temperaturi. Pomoću pincete sam iscijedila vodu i potom odvagala 90 mg uzorka. Usitnila sam ga skalpelom i sterilnim štapićem.

Za izolaciju visokomolarne genomske DNA koristila sam komplet QIAGEN Blood & Cell Culture DNA koji sadrži set pufera Qiagen Genomic DNA Buffer Set i kolonice Qiagen Genomic-tip 20/G. Stanice se liziraju, centrifugiraju i supernatant u kojem se nalazi DNA se nanosi na kolonu. Silikatna kolona djeluje kao anionski izmjenjivač, što znači da uvjeti visoke ionske jakosti omogućuju vezanje DNA a ispiranje proteina i drugih nečistoća. Na kraju se pročišćena DNA eluira s kolone. Izolirala sam DNA prema uputama proizvođača za izolaciju iz uzorka tkiva.

3.1.1. Agarozna gel-elektroforeza

Dva puta po 5 µL izolata podvrgnula sam elektroforezi u 0,8 postotnom agaroznom gelu pripremljenom s TE puferom i etidij-bromidom pri naponu od 60 volti. Ukupan naboj molekule DNA proporcionalan je njenoj duljini pa se primjenom električnog polja na molekule koje se nalaze u poroznom gelu one razdvajaju po veličini. Etidij-bromid se interkalira u DNA i omogućava vizualizaciju jer fluorescira narančasto pod UV svjetlom. Kao referencu za veličinu fragmenata na gelu koristila sam DNA biljeg GeneRuler DNA Ladder Mix proizvođača Thermo Scientific.

3.1.2. Određivanje koncentracije DNA

Koncentraciju DNA u izolatu izmjerila sam pomoću uređaja BioSpec-nano Micro-volume UV-VIS Spectrophotometer. Izoliranu DNA iskoristila sam za potvrdu vrste pomoću genetičkog biljega ITS2 i pripremu knjižnice za sekvenciranje tehnologijom nanopora.

3.2. Potvrda vrste pomoću genetičkog biljega ITS2

U svrhu potvrde vrste jedinke iz koje je izolirana genomska DNA provedena je lančana reakcija polimerazom (PCR) kojom je umnožen jezgrin ITS2 (eng. *internal transcribed spacer*), molekularni biljeg za određivanje nižih taksonomskih kategorija. Za umnažanje fragmenta ITS2 duljine 361 pb lančanom reakcijom polimerazom korištene su početnice

ITS2F (5' CGGCTCGTGCGTCGATGAAGAAC 3') i ITS2R (5' CGCCGTTACTGGGGGAATCCCTGTTG 3') iz Harcet i suradnici (2010.a).

Metodom PCR se postiže umnažanje ciljnog fragmenta DNA kalupa (određenog slijedom baza u početnicama koje se vežu za kalup i započinju sintezu) pomoću termostabilne DNA polimeraze. Periodičke promjene temperature uzrokuju ponavljanje ciklusa denaturacije kalupa, prianjanja početnica i elongacije - što u konačnici rezultira velikim brojem kopija istog fragmenta (Mullis i sur. 1986.). ITS2 (od engl. *internal transcribed spacer*) je nekodirajuća regija koja se nalazi između visoko konzerviranih ribosomalnih gena kod svih eukariota. U genomu je prisutan u velikom broju kopija, a zbog malog evolucijskog pritiska na nekodirajuće sekvence pokazuje visok stupanj varijabilnosti između vrsta. Zbog tih svojstava ITS2 se često koristi kao genetički biljeg pri određivanju vrsta gotovo svih eukariota, pa tako i spužvi (Itskovich i sur. 2008.).

3.2.1. Lančana reakcija polimerazom

Kao kalup sam koristila genomsku DNA čija izolacija je opisana u poglavlju 3.1. Pripremila sam smjesu za PCR reakciju sljedećeg sastava: oko 30 ng kalupa (konačna koncentracija otprilike 1,2 ng/μL), 12,5 μL reagensa Taq PCR Reaction Mix With $MgCl_2$ proizvođača Sigma, 1 μL početnice ITS2F (konačna koncentracija 0,4 μM), 1 μL početnice ITS2R (konačna koncentracija 0,4 μM), voda bez nukleaza do ukupnog volumena 25 μL. Uvjeti provođenja reakcije opisani su u Tablici 1.

TABLICA 1. Uvjeti provođenja PCR reakcije za umnožavanje ITS2 regije u izolatu genomske DNA.

temperatura (°C)	**vrijeme**	**broj ponavljanja**
95	3 min	1
94	30 s	35
55	45 s	
72	1 min 30 s	
72	10 min	1
4	∞	1

Po završetku PCR reakcije, prebacila sam ukupni volumen reakcijske smjese u jažice pripremljenog agaroznog gela. Provela sam elektroforezu na način opisan u poglavlju 3.1.1.. Koristila sam i DNA biljeg (isti kao ranije), kako bih mogla odrediti veličinu fragmenta umnoženog PCR reakcijom.

3.2.2. Izolacija PCR produkta iz gela

Nakon što sam potvrdila da je PCR reakcijom umnožen fragment očekivane veličine (410 pb), izrezala sam PCR produkt iz gela. Iz toga sam izolirala DNA prema uputama proizvođača koristeći komplet QIAquick Gel Extraction kit, koji se zasniva na ionskoj izmjeni u koloni. Dobiveni izolat koristila sam kao kalup u reakciji sekvenciranja Sangerovom dideoksi metodom.

3.2.3. Sekvenciranje Sangerovom metodom

Za sekvenciranje sam pripremila dvije PCR reakcije. Sastav reakcijske smjese bio je sljedeći: oko 200 ng DNA kalupa, 1 μL reakcijske smjese za sekvenciranje i 1 μL pufera za sekvenciranje iz kompleta ABI PRISM BigDye Terminator v3.1 Ready Reaction Cycle Sequencing Kit (Applied Biosystems), te 1 μL početnice (konačna koncentracija 1 μM) u ukupnom volumenu reakcijske smjese 10 μL. Za jednu reakciju sam koristila početnicu ITS2F a za drugu ITS2R. Uvjeti reakcije opisani su u Tablici 2.

TABLICA 2. Uvjeti provođenja PCR reakcije u sekvenciranju Sangerovom metodom.

temperatura (°C)	**vrijeme**	**broj ponavljanja**
96	1 min	1
96	1 min	25
50	5 s	
60	4 min	
4	10 min	1
4	∞	1

Za pročišćavanje DNA fragmenta po završetku reakcije dodala sam u svaku tubicu 2 μL EDTA (konc. 125 mM), 2 μL natrijevog acetata (konc. 3 M) i 50 μL apsolutnog etanola pa inkubirala 15 minuta na sobnoj temperaturi u mraku. Nakon centrifugiranja (15 minuta pri maksimalnoj brzini i 4°C) odsisala sam supernatant i dodala 70 μL prethlađenog 70-postotnog etanola. Ponovno sam centrifugirala (5 minuta pri maksimalnoj brzini i 4°C), odsisala supernatant i osušila talog 5 minuta u vakuum-centrifugi. Resuspendirala sam talog dodavanjem13 μL formamida. Uzorke sam stavila na denaturaciju (3 minute pri 96°C), a zatim na led najmanje 2 minute. Programirala sam uređaj ABI PRISM 3100 Avant Genetic Analyser (proizvođač Applied Biosystems) i njime provela kapilarnu elektroforezu pripremljenih uzoraka.

Rezultate sekvenciranja analizirala sam pomoću programa ChromasPro2. Uklonila sam nepouzdano očitane baze dok prosječna mjera kvalitete u pomičućem prozoru duljine 10 nt nije bila veća od 15. Napravila sam reverzni komplement slijeda dobivenog u reakciji s početnicom ITS2R, sravnila ga sa slijedom dobivenim reakcijom s ITS2F, i izračunala konsenzusni slijed (Prilog 1). Potražila sam slične sljedove u bazi podataka NCBI Nucleotide (https://www.ncbi.nlm.nih.gov/nuccore/) koristeći alat BLAST.

3.3. Priprema knjižnica za sekvenciranje tehnologijom nanopora

Osnovni koraci u pripremi izolirane DNA za sekvenciranje na uređaju ONT MinION su:

1. cijepanje na fragmente
2. FFPE popravak lomova
3. popravak i pripremanje krajeva molekule
4. ligacija adaptera
5. pročišćavanje pomoću MyOne C1 kuglica
6. elucija s kuglica.

Zbog duljine protokola koraci su detaljnije opisani u Prilogu 2. Provedbom ovog postupka dobiva se mješavina za sekvenciranje (engl. *pre-sequencing mix*) spremna za nanošenje na protočnu ćeliju uređaja ONT MinION.

Pripremila sam tri knjižnice za sekvenciranje koristeći Ultra II End Prep kit, FFPE DNA Repair kit i Blunt/TA Ligase Master Mix proizvođača New England Biolabs. Za dvije knjižnice koristila sam još i Nanopore sequencig kit SQK-MAP006 (odgovara R7 verziji kemije na protočnoj ćeliji), a za treću Nanopore sequencig kit SQK-NSK007 (za verziju R9) proizvođača Oxford Nanopore Technologies. Protokoli za pripremu knjižnice pomoću kompleta SQK-MAP006 i SQK-NSK007 su isti.

Provedena je optimizacija protokola i napravljene su modifikacije u službenom protokolu proizvođača navedenom u Prilogu 2. U koraku 3. (popravak i pripremanje krajeva molekule) produljila sam vremena inkubacije s enzimima za popravak na 20°C i 65°C sa 5 minuta na 30 minuta kako bih dobila veći udio molekula s krajevima pogodnim za ligaciju adaptera u sljedećem koraku. U koracima 2. (FFPE popravak lomova) i 3. koristila sam 0,4 volumena kuglica za pročišćavanje u odnosu na originalni protokol proizvođača jer manja količina kuglica pogoduje vezanju duljih fragmenata i smanjuje vezanje kratkih. U istim koracima sam produljila vremena elucije DNA s kuglica s 2 minute na 15 minuta, kako bih smanjila gubitke u pročišćavanju i povećala prinos.

Koncentraciju DNA u manjim volumenima odvojenim tijekom različitih koraka pripreme (alikvoti 1-7) izmjerila sam uređajem Invitrogen Qubit 4 Fluorometer (proizvođač Thermo Fisher Scientific).

3.4. Sekvenciranje tehnologijom nanopora na uređaju ONT MinION

Isprala sam protočnu ćeliju ONT MinION R7 FAD08562 otopinama iz kompleta za ispiranje Oxford Nanopore Technologies Wash Kit. Kontrolom kvalitete ustanovljeno je da ima 269 aktivnih pora. Ćeliju sam pripremila za rad prema uputama proizvođača. Smjesu za sekvenciranje pripremila sam miješanjem 75 μL pufera RNB, 23 μL vode bez nukleaza, 4 μL reagensa FMX, i po 24 μL obje knjižnice za sekvenciranje pripremljene kompletom SQK-MAP006 (verzija R7). Smjesu sam nanijela na protočnu ćeliju. Pokrenuta je reakcija sekvenciranja kojom je dobivena knjižnica ONT 1.

Protočnu ćeliju ONT MinION R9 FAD21998 isprala sam i pripremila za rad prema uputama proizvođača. Ustanovljeno je 216 aktivnih pora. Pomiješala sam 75 μL reagensa RNB, 63 μL vode bez nukleaza i 12 μL knjižnice za sekvenciranje pripremljene kompletom SQK-NSK007 (u verziji R9 reagens RNB osim pufera sadrži i reagens FMX, što nije slučaj u verziji R7). Tako dobivenu smjesu nanijela sam na protočnu ćeliju, nakon čega je pokrenuto sekvenciranje. Rezultat je knjižnica ONT 2.

3.5. Korištene knjižnice

Koristila sam pet otprije dostupnih knjižnica sekvenciranih tehnologijom Illumina, dvije na platformi Illumina MiSeq i tri na platformi Illumina HiSeq (uključujući knjižnicu Illumina Macrogen). Obje knjižnice Illumina MiSeq i knjižnica Illumina HiSeq 1 dobivene su postupkom uparenog sekvenciranja krajeva (engl. *paired-end sequencing*). Knjižnica Illumina HiSeq 2 dobivena je uparenim sekvenciranjem uz veću veličinu fragmenta (engl. *mate-pair sequencing*). Podatci o broju i duljini sljedova u pojedinim knjižnicama nalaze se u Tablici 3.

TABLICA 3. Broj i duljina sljedova i ukupan broj baza u korištenim Illumina knjižnicama.

knjižnica	**broj sljedova**	**duljina slijeda (pb)**	**ukupna duljina (pb)**
Illumina MiSeq 1	30 871 686	251	7 748 793 186
Illumina MiSeq 2	31 456 208	251	7 895 508 208
Illumina HiSeq 1	41 894 314	94	3 938 065 516
Illumina HiSeq 2	37 859 628	100	3 785 962 800
Illumina Macrogen	381 488 100	151	57 604 703 100

Koristila sam uspješne 2D sljedove iz dviju knjižnica ONT MinION koje sam sekvencirala u sklopu ovog rada (ONT 1 i ONT 2) i dviju otprije dostupnih knjižnica (ONT 3 i ONT 4). Električni signal je tijekom sekvenciranja knjižnica ONT 1-4 prevođen u redoslijed baza programom Metrichor. Knjižnica ONT 5 predstavlja skup svih uspješnih sljedova sekvenciranih u većem broju reakcija na ONT protočnim ćelijama nove generacije, uz korištenje programa za prevođenje Albacore. Podatci o broju korištenih sljedova, ukupnoj duljini, N50, prosječnoj i maksimalnoj duljini sljedova za pojedine knjižnice nalaze se u Tablici 4.

TABLICA 4. Statistički podatci o korištenim sljedovima u ONT knjižnicama.

knjižnica	**broj sljedova**	**ukupna duljina (pb)**	**prosječna duljina slijeda (pb)**	**N50 (pb)**	**najdulji slijed (pb)**
ONT 1	1 548	3 988 041	2 517	5 497	15 004
ONT 2	96	710 316	7 399	9 957	31 493
ONT 3	70 293	88 766 199	1 263	3 411	27 822
ONT 4	25 773	117 600 907	4 562	6 360	17 629
ONT 5	441 055	2 128 982 552	4 827	9 986	94 485

3.6. Predobrada nukleotidnih sljedova

3.6.1. Predobrada sljedova dobivenih sekvenciranjem tehnologijom Illumina

Za predobradu sljedova dobivenih sekvenciranjem tehnologijom Illumina koristila sam programe iz programskog paketa BBTools.

Knjižnicama Illumina MiSeq 1, MiSeq 2 i HiSeq 1 uklonila sam adapterske sljedove minimalne duljine 27 nt pomoću programa BBDuk. Unutar svake knjižnice sam provela ispravljanje preklapajućih dijelova uparenih sljedova programom BBMerge, te spojila uparene sljedove u jednu datoteku. Zatim sam BBDukom uklonila dijelove sljedova s lijeve i desne strane koji su imali prosječnu mjeru kvalitete manju od 20. Na kraju sam uz pomoć programa BBNorm s parametrom k=31 provela normalizaciju dubine pokrivenosti knjižnica i ispravljanje pogrešaka.

Illumina HiSeq 2 je knjižnica uparenih sljedova s velikim razmakom (engl. *mate-pair*) pa na njoj nisam provodila korekciju preklapajućih dijelova jer takvi ne postoje. Pomoću BBDuka sam uklonila adaptere i dijelove s lijeve i desne strane sljedova s prosječnom mjerom kvalitete manjom od 20, te BBNormom normalizirala dubinu pokrivenosti uz parametar k=31 i ispravljanje pogrešaka.

Knjižnici Illumina Macrogen uklonila sam adaptere minimalne duljine 19 nt i regije s prosječnom kvalitetom nižom od 20 pomoću programa BBDuk2. Kod kraćenja s obzirom na kvalitetu, oba slijeda u jednom paru skraćena su na istu duljinu i odbačeni su svi parovi u kojima je barem jedan član bio kraći od 100 nt.

3.6.2. Predobrada sljedova dobivenih sekvenciranjem tehnologijom nanopora

Sljedove sam prebacila iz hijerarhijskog formata .fast5 u format .fastq koji prihvaća većina programa za analizu. Za to sam koristila programski paket poretools. Zbog lakšeg baratanja sve sljedove iz knjižnica ONT 1-4 sam spojila u jednu datoteku. Ispravljanje sljedova u knjižnicama ONT 1-4 pomoću programa Nanocorr i LoRDEC provela sam uz pomoć podataka iz predobrađenih knjižnica Illumina MiSeq 1 i 2 i HiSeq 1 i 2. Sljedove u knjižnici ONT 5 nisam ispravljala.

3.7. Sklapanje genoma

Illuminine knjižnice MiSeq 1 i 2 i HiSeq 1 i 2 sklopila sam programom SPAdes uz vrijednosti parametra k 33, 55 i 77. Knjižnicu Illumina Macrogen sklopila sam koristeći isti program, uz vrijednosti parametra k 21, 33, 55 i 77. Dobivene neprekinute sljedove sklopila sam s ONT sljedovima pomoću programa npScarf.

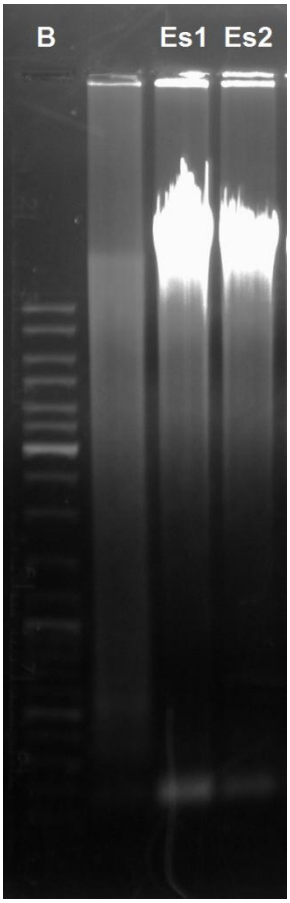
3.8. Analiza sklopljenih sljedova

Za procjenu kvalitete sklopljenih sljedova koristila sam statistike N50, broj i ukupnu duljinu sklopljenih sljedova te duljinu najduljeg neprekinutog slijeda, koje sam izračunala pomoću programa QUAST. Detekciju 978 visoko očuvanih gena prisutnih kod svih životinja (Metazoa) u sklopljenim sljedovima provela sam programom BUSCO. Sklopljene nukleotidne sljedove sam prevela u aminokiselinske i usporedila ih s bazom podataka proteinskih sekvenci dostupnom preko servisa NCBI (NCBI Resource Coordinators 2018.) služeći se programom DIAMOND. Iz rezultata sam izvukla taksonomske podatke pomoću paketa funkcija taxize za programski jezik R (Chamberlain i Szöcs 2013.) i baze podataka NCBI Taxonomy (NCBI Resource Coordinators 2018.). Sklopljene sljedove sam također usporedila s genomom spužve *Amphimedon queenslandica* koristeći program BLAST.

4. REZULTATI

4.1. Izolacija genomske DNA

Gel nakon provedene elektroforeze izolata genomske DNA prikazan je na Slici 8. U liniji označenoj s B nalazi se DNA biljeg, a u linijama Es1 i Es2 produkt izolacije genomske DNA opisane u poglavlju 2.1. Materijala i metoda. U gornjoj trećini linija Es1 i Es2 vidljiva je visokomolekularna genomska DNA spužve a pri dnu su vidljivi mali fragmenti.

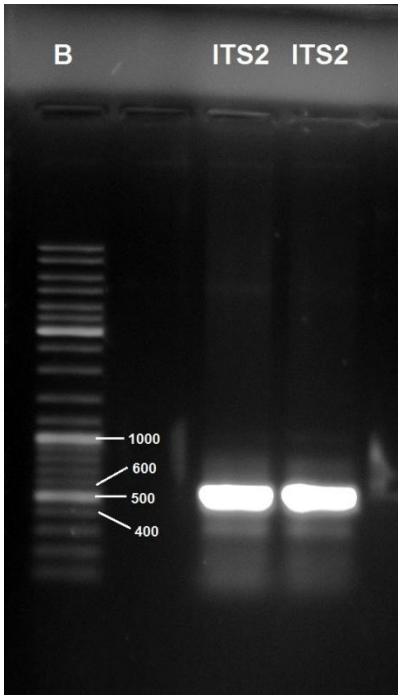


SLIKA 8. 0,8%-tni agarozni gel nakon elektroforeze genomske DNA *E. subterraneus* i DNA biljega. B - DNA biljeg, Es1 i Es2 - uzorci izolata iz poglavlja 2.1. Materijala i metoda.

Uređajem BioSpec-nano Micro-volume UV-VIS Spectrophotometer izmjerena je koncentracija DNA od 200 ng/µL, uz prihvatljive vrijednosti OD260/280 (1,88) i OD260/230 (2,1).

4.2. Potvrda vrste pomoću genetičkog biljega ITS2

Umnažanje fragmenta jezgrinog ITS2 lančanom reakcijom polimerazom bilo je uspješno. Na slici 9. prikazan je agarozni gel sa vidljivim PCR fragmentom (ITS2) očekivane veličine.



SLIKA 9. 0,8%-tni agarozni gel nakon elektroforeze umnoženog PCR produkta (jezgrinog ITS2 spužve *E. subterraneus*) i DNA-biljega. Brojevi označavaju broj parova baza u pojedinim fragmentima DNA biljega. B - DNA biljeg, ITS2 - umnoženi PCR produkt.

Dobiveni fragment izrezan je iz gela i pročišćen pomoću kompleta QIAquick Gel Extraction Kit. Sljedovi dobiveni sekvenciranjem automatiziranom Sangerovom metodom obrađeni su programom ChromasPro2. Pregledom baze podataka nukleotidnih sekvenci dostupnom preko stranice NCBI pomoću alata BLAST potvrđeno je da se uistinu radi o vrsti *Eunapius subterraneus* (Eunapius subterraneus internal transcribed spacer 2, partial sequence (identifikacijski kod FJ715436.1) uz postotak identičnosti 89% i E vrijednost $2*10^{-118}$).

4.3. Priprema knjižnica za sekvenciranje tehnologijom nanopora

Izmjerene su koncentracije DNA u tri pripremljene knjižnice za sekvenciranje. U knjižnicama pripremljenim pomoću kompleta reagensa SQK-MAP006 (verzija R7) koncentracije DNA su bile 1,70 ng/μL i 1,30 ng/μL. U knjižnici pripremljenoj pomoću kompleta SQK-NSK007 (verzija R9) izmjerena je koncentracija DNA 1,85 ng/μL. Preporuka proizvođača za optimalan rezultat je oko 250 ng DNA u 25 μL gotove knjižnice, odnosno koncentracija od otprilike 10 ng/μL. Koncentracije DNA izmjerene u alikvotima prikazane su tablicom u Prilogu 3.

4.4. Sekvenciranje tehnologijom nanopora na uređaju ONT MinION

Podatci o broju i duljini dobivenih sljedova navedeni su ranije u Tablici 4. Obje knjižnice sadrže mali broj uspješno očitanih sljedova (u slučaju ONT 2 iznimno mali) ali je vrijednost N50 prilično visoka. Knjižnica ONT 2 sadrži i najdulji od svih korištenih sljedova.

4.5. Predobrada nukleotidnih sljedova

4.5.1. Predobrada sljedova dobivenih sekvenciranjem tehnologijom Illumina

Podatci o broju i duljini sljedova u knjižnicama Illumina MiSeq 1 i 2 i HiSeq 1 i 2 nakon uklanjanja adaptera i dijelova niske kvalitete, ispravljanja pogrešaka na preklapajućim dijelovima i normalizacije dubine nalaze se u Tablici 5. U Tablici se nalaze i podatci o broju i duljini sljedova za knjižnicu Illumina Macrogen nakon uklanjanja adaptera i dijelova s niskon kvalitetom.

TABLICA 5. Broj i prosječna duljina sljedova te ukupan broj baza u Illumina knjižnicama nakon predobrade.

knjižnica	**broj sljedova**	**prosječna duljina slijeda (pb)**	**ukupna duljina (pb)**
Illumina MiSeq 1	21 973 554	217	4 759 103 260
Illumina MiSeq 2	18 638 612	213	3 972 025 696
Illumina HiSeq 1	33 404 712	88	2 927 673 822
Illumina HiSeq 2	28 187 268	90	2 536 650 533
Illumina Macrogen	241 264 840	140	33 892 263 448

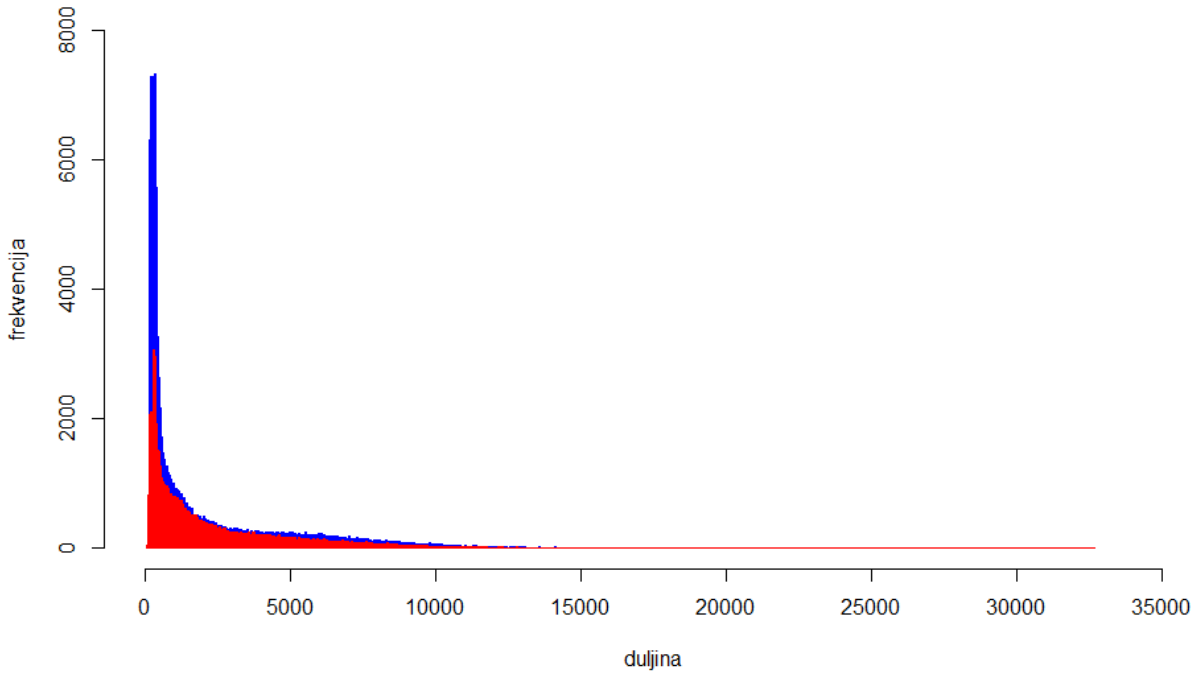
4.5.2. Predobrada sljedova dobivenih sekvenciranjem tehnologijom nanopora

Podatci o broju sljedova, ukupnoj duljini, N50 i prosječnoj duljini za neispravljene i ispravljene sljedove iz knjižnica ONT 1-4 nalaze se u Tablici 6.

TABLICA 6. Statistički podatci o ONT sljedovima prije i nakon različitih metoda ispravljanja.

program za ispravljanje	**broj sljedova**	**ukupna duljina (pb)**	**prosječna duljina slijeda (pb)**	**N50 (pb)**	**najdulji slijed (pb)**
-	97 746	211 065 463	2 159	5 386	31 493
Nanocorr	60 104	139 946 106	2 328	4 675	32 621
LoRDEC	97 746	211 438 983	2 163	5 408	32 493

Na Slici 10 prikazana je distribucija duljine neispravljenih sljedova i sljedova ispravljenih programom Nanocorr iz knjižnica ONT 1-4. Distribucija duljina sljedova ispravljenih pomoću programa LoRDEC gotovo je identična distribuciji neispravljenih sljedova pa nije prikazana.



SLIKA 10. Histogram duljina neispravljenih sljedova iz knjižnica ONT 1-4 (plavo) i istih sljedova ispravljenih pomoću programa Nanocorr (crveno).

Ispravljeni sljedovi iz knjižnica ONT 1-4 uspoređeni su s neispravljenim pomoću programa BLAST, pri čemu je zadržan samo najbolji pogodak za svaki ispravljeni slijed. Postotak pokrivenosti sravnjenjima za sljedove ispravljene programom Nanocorr iznosi 55,24 %, a za sljedove ispravljene programom LoRDEC 91,58 %. Podatci o broju i duljini ostvarenih sravnjenja, te prosječnom broju pogrešno uparenih baza (engl. *mismatch*) i prekida (engl. *gaps*) po sravnjenju navedeni su u Tablici 7.

TABLICA 7. Statistički podatci o sravnjenju ispravljenih i neispravljenih sljedova iz knjižnica ONT 1-4.

program za ispravljanje	**broj sravnjenja**	**prosječna duljina (pb)**	**prosječan broj pogrešno uparenih baza**	**prosječan broj prekida**
Nanocorr	57 842	1336,39	86,87	90,51
LoRDEC	97 672	1983,55	43,07	48,01

Ispravljeni i neispravljeni sljedovi iz knjižnica ONT 1-5 su sravnjeni s genomom spužve *A. queenslandica* pomoću programa BLAST, a kao pogoci su zadržana samo identična

sravnjenja. Broj i ukupna duljina sravnjenja te broj pogodaka prosječno ostvaren po 100 000 pb navedeni su u Tablici 8.

TABLICA 8. Statistički podatci nakon usporedbe ispravljenih i neispravljenih knjižnica ONT 1-5 s genomom *A. queenslandica*.

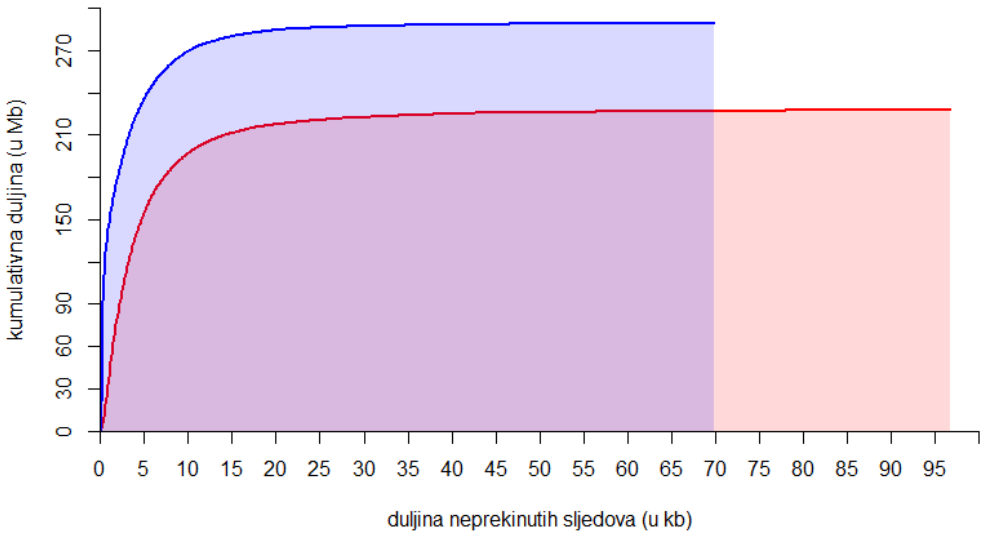
ONT knjižnice	**broj pogodaka**	**ukupna duljina sravnjenja (pb)**	**broj pogodaka po 100 000 baza**
ONT 1-4	154	4 460	0,07
ONT 1-4 (Nanocorr)	1 211	35 874	0,86
ONT 1-4 (LoRDEC)	333	10 039	0,16
ONT 5	28 659	826 687	1,35

4.6. Sklapanje genoma

U Tablici 9 prikazane su statistike za neprekinute sljedove dobivene sklapanjem pomoću programa SPAdes. Knjižnice korištene u pojedinim sklapanjima navedene su u prvom stupcu. Slika 11 prikazuje kumulativnu duljinu neprekinutih sljedova dobivenih u dva sklapanja.

TABLICA 9. Statistički podatci o neprekinutim sljedovima sklopljenim programom SPAdes.

knjižnice	**broj neprekinutih sljedova**	**ukupna duljina (pb)**	**N50 (pb)**	**najdulji neprekinuti slijed (pb)**
MiSeq 1 i 2 HiSeq 1 i 2	159 448	227 586 314	2 993	96 768
Macrogen	803 833	289 462 365	931	69 828



SLIKA 11. Distribucija kumulativne duljine neprekinutih sljedova dobivenih sklapanjem knjižnice Illumina Macrogen (plavo) i knjižnica Illumina MiSeq 1 i 2 i HiSeq 1 i 2 (crveno).

Podatci o broju i ukupnoj duljini neprekinutih sljedova, statistici N50 i veličini najduljeg neprekinutog slijeda dobivenog sklapanjem pomoću programa npScarf prikazani su u Tablici 10. Skupovi neprekinutih sljedova na koje su sklapani ONT sljedovi navedeni su u prvom stupcu, a ONT knjižnice korištene u pojedinim sklapanjima u drugom stupcu. Program za ispravljanje sljedova, ukoliko je korišten, naveden je u zagradi.

TABLICA 10. Statistički podatci o neprekinutim sljedovima dobivenim sklapanjem okvira programom npScarf.

neprekinuti sljedovi	**ONT knjižnice**	**broj neprekinutih sljedova**	**ukupna duljina neprekinutih sljedova (pb)**	**N50 (pb)**	**najdulji neprekinuti slijed (pb)**
MiSeq + HiSeq	ONT 1-4 (Nanocorr)	80 467	140 581 155	4 086	96 768
	ONT 1-4 (LoRDEC)	73 508	140 779 500	5 341	96 768
	ONT 5	46 781	177 566 399	41 636	393 798
	ONT 1-4 (Nanocorr) + ONT 5	46 655	176 754 425	42 063	314 450
	ONT 1-4 (LoRDEC) + ONT 5	46 623	176 600 587	42 205	323 568
Macrogen	ONT 5	395 242	190 579 017	10 635	305 409
	ONT 1-4 (Nanocorr) + ONT 5	395 195	190 285 384	10 727	230 405
	ONT 1-4 (LoRDEC) + ONT 5	395 190	189 506 336	10 451	301 259

4.7. Analiza sklopljenih sljedova

4.7.1. Procjena dovršenosti genoma programom BUSCO

Broj gena koji su pronađeni u cijelosti, pronađeni parcijalno ili nedostaju u pretrazi sklopljenih sljedova programom BUSCO prikazan je u Tablici 11. U zagradi je naveden udio u ukupnom broju gena za koje je izvršena pretraga (978). Neprekinuti sljedovi i ONT knjižnice korišteni u pojedinim sklapanjima navedeni su u prva dva stupca.

TABLICA 11. Broj i udio pronađenih cijelih, pronađenih djelomičnih i izostajućih od 978 visoko očuvanih životinjskih gena u sklopljenim sljedovima.

neprekinuti sljedovi	**ONT knjižnice**	**cijeli**	**parcijalni**	**nisu pronađeni**
MiSeq + HiSeq	-	535 (54,7 %)	189 (19,3 %)	254 (26,0 %)
	ONT 1-4 (Nanocorr)	529 (54,1 %)	190 (19,4 %)	259 (26,5 %)
	ONT 1-4 (LoRDEC)	562 (57,5 %)	172 (17,6 %)	244 (24,9 %)
	ONT 5	684 (69,9 %)	91 (9,3 %)	203 (20,8 %)
	ONT 1-4 (Nanocorr) + ONT 5	694 (71,0 %)	92 (9,4 %)	192 (19,6 %)
	ONT 1-4 (LoRDEC) + ONT 5	704 (72,0 %)	86 (8,8 %)	188 (19,2 %)
Macrogen	-	511 (52,2 %)	210 (21,5 %)	257 (26,3 %)
	ONT 5	553 (56,6 %)	137 (14,0 %)	288 (29,4 %)
	ONT 1-4 (Nanocorr) + ONT 5	558 (57,1 %)	136 (13,9 %)	284 (29,0 %)
	ONT 1-4 (LoRDEC) + ONT 5	557 (57,0 %)	140 (14,3 %)	281 (28,7 %)

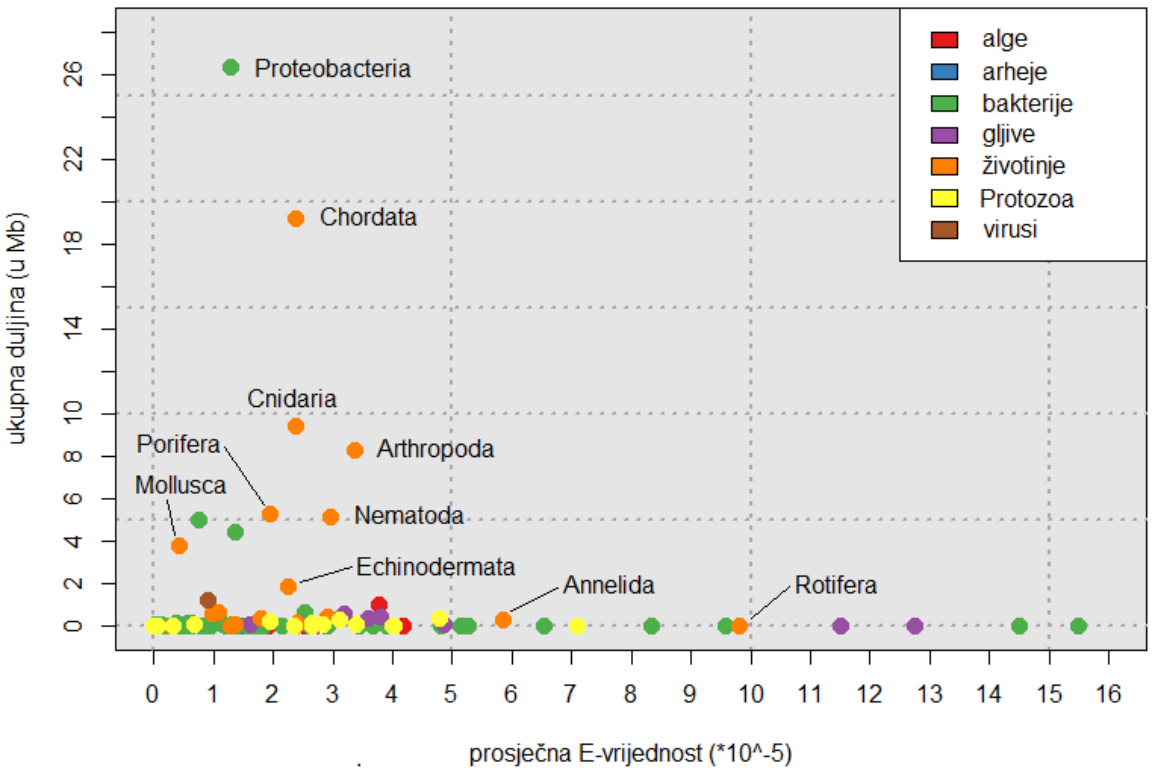
Rezultati su podijeljeni u tri grupe:

1. sklapanja provedena na neprekinutim sljedovima dobivenim od knjižnica Illumina HiSeq i MiSeq bez sljedova iz knjižnice ONT 5 (prva tri reda u Tablici 10)
2. sklapanja provedena na neprekinutim sljedovima dobivenim od knjižnica Illumina HiSeq i MiSeq uz sljedove iz knjižnice ONT 5 (druga tri reda u Tablici 10)
3. sklapanja provedena na neprekinutim sljedovima dobivenim od knjižnice Illumina Macrogen (zadnja četiri reda Tablice 10).

Srednje vrijednosti broja ukupno pronađenih gena (cijelih i parcijalnih) unutar grupa iznose redom 725,67, 783,67 i 700,50, a srednja vrijednost za sva sklapanja iznosi 732,60. Standardne devijacije udjela ukupno pronađenih gena unutar grupa iznose redom 0,78 %, 0,79 % i 1,43 %.

4.7.2. Usporedba sklopljenih sljedova s bazom podataka proteinskih sljedova

Skup sljedova dobiven zajedničkim sklapanjem knjižnica Illumina MiSeq i HiSeq, ONT 1-4 ispravljenih programom LoRDEC i ONT 5 je uspoređen s proteinskim sljedovima iz baze podataka NCBI Protein, uz mogućnost pomaka okvira pri sravnjivanju. Za ovaj korak korišten je program DIAMOND. Za svaki ispitivani slijed zabilježeno je 25 najboljih sravnjenja, pri čemu je kao mjerilo kvalitete korišten tzv. *bit score* (bodovanje sravnjenja transformirano na način da ne ovisi o veličini prostora pretrage). Svi sravnjeni sljedovi iz baze podataka su grupirani po pripadnosti pojedinom koljenu i statistički obrađeni. Na Slici 12. prikazan je odnos prosječne ostvarene E-vrijednosti i ukupne duljine sravnjenja za svako koljeno, a različite boje se odnose na pripadnost koljena većim skupinama. Označeni su i nazivi istaknutijih koljena. U Tablici 12 nalaze se podatci o broju, prosječnoj i ukupnoj duljini sravnjenja te prosječnoj ostvarenoj E-vrijednosti za više taksonomske kategorije.



SLIKA 12. Odnos prosječne E-vrijednosti i ukupne duljine sravnjenja po koljenu.

TABLICA 12. Statistički podatci nakon usporedbe neprekinutih sljedova s proteinskom bazom podataka NCBI Protein.

skupina	**broj sravnjenja**	**prosječna duljina (pb)**	**ukupna duljina (pb)**	**prosječna E-vrijednost (*10^{-5})**
životinje	179 378	172,69	55 525 160	2,096
bakterije	102 023	295,24	38 338 723	2,144
gljive	5 579	173,79	1 423 257	7,826
virusi	3 756	286,14	1 189 767	0,894
Protozoa	3 663	205,18	1 161 617	2,662
alge	3 839	218,62	1 033 512	2,855
arheje	458	299,32	130 638	0,283

Ista usporedba je zatim provedena uz korištenje algoritma LCA (engl. *last common ancestor*) u sklopu programa DIAMOND. Algoritam za svaki ispitivani neprekinuti slijed odabire 10% najboljih sravnjenja, za svako određuje taksonomsku klasifikaciju organizama iz kojih sekvence potječu, i traži najbližeg zajedničkog pretka. Od 10 249 neprekinutih sljedova za koje je bilo moguće odrediti najbližeg zajedničkog pretka na razini koljena ili niže, izdvaja se 2 011 slijed za kojeg zajednički predak pripada koljenu Porifera (od toga 1728 *Amphimedon queenslandica*, 78 *Ephydatia fluviatilis* i 75 *Lubomirskia baicalensis*). Deset koljena s najvećim brojem neprekinutih sljedova za koje su određena kao najbliži zajednički predak prikazana su u Tablici 13, uz podatak o prosječnoj E-vrijednosti sravnjenja.

TABLICA 13. Podatci nakon usporedbe neprekinutih sljedova s proteinskom bazom podataka NCBI Protein i određivanja poslijednjeg zajedničkog pretka po neprekinutom slijedu.

koljeno	carstvo	broj neprekinutih sljedova	prosječna E-vrijednost sravnjenja (*10^{-5})
Porifera	Metazoa	2011	0,575
Proteobacteria	Bacteria	1616	2,462
Cnidaria	Metazoa	426	11,658
Chordata	Metazoa	321	12,944
Arthropoda	Metazoa	307	13,924
Bacteroidetes	Bacteria	272	3,993
Firmicutes	Bacteria	156	2,655
Nematoda	Metazoa	106	1,998
Echinodermata	Metazoa	102	4,240
Mollusca	Metazoa	86	11,548

4.7.3. Usporedba sklopljenih sljedova s genomom vrste Amphimedon queenslandica

Svi sklopljeni sljedovi su uspoređeni s genomom spužve *A. queenslandica* pomoću programa BLAST. U Tablici 14 navedeni su broj ostvarenih sravnjenja, ukupna i prosječna duljina sravnjenja, prosječan broj pogrešno uparenih baza (engl. *mismatch*) i prekida (engl. *gaps*) po sravnjenom slijedu, te postotak pokrivenosti ispitivanih sljedova.

TABLICA 14. Statistički podatci o sravnjenju sklopljenih sljedova s genomom *A. queenslandica*.

neprekinuti sljedovi	**ONT knjižnice**	**broj sravnjenja**	**prosječna duljina (pb)**	**ukupna duljina (pb)**	**prosječni broj pogrešno uparenih baza**	**prosječni broj prekida**	**pokrivenost ispitivanih sljedova (%)**
MiSeq + HiSeq	-	6787	46,74	317225	2,77	0,55	0,1394
	ONT 1-4 (Nanocorr)	6241	41,35	258088	1,71	0,41	0,1836
	ONT 1-4 (LoRDEC)	6214	41,48	257734	1,73	0,42	0,1831
	ONT 5	6195	42,72	264675	1,95	0,46	0,1491
	ONT 1-4 (Nanocorr) + ONT 5	6274	42,54	266866	1,96	0,45	0,1510
	ONT 1-4 (LoRDEC) + ONT 5	6214	42,72	265452	1,97	0,45	0,1503
Macrogen	-	23346	35,47	827982	0,98	0,30	0,2860
	ONT 5	17193	34,77	597873	0,90	0,24	0,3137
	ONT 1-4 (Nanocorr) + ONT 5	17005	34,82	592054	0,90	0,24	0,3111
	ONT 1-4 (LoRDEC) + ONT 5	17216	34,72	597791	0,89	0,24	0,3154

5. RASPRAVA

Izolacijom DNA dobivena je visokomolekularna genomska DNA s vrlo malo kratkih fragmenata koja je bila prikladna za pripremu knjižnica za sekvenciranje. Umnažanjem fragmenta jezgrinog ITS2 lančanom reakcijom polimerazom (PCR) s ciljem potvrde vrste dobila sam očekivani rezultat. Dobiveni fragment se nalazi na poziciji koja odgovara molekuli DNA duljine između 400 i 500 parova baza, što je i očekivana veličina fragmenta umnoženog korištenim početnicama. Rezultat pretrage programom BLAST potvrđuje da uzorak spužve s kojim sam radila zbilja potječe iz jedinke vrste *Eunapius subterraneus*.

Koncentracije u pripremljenim knjižnicama za sekvenciranje bile su pet do sedam puta manje od preporučenih pa sam zato za jednu reakciju sekvenciranja objedinila dvije knjižnice i nanijela ih zajedno na protočnu ćeliju. Preniska koncentracija DNA u knjižnicama i mali broj aktivnih pora na korištenim protočnim ćelijama su razlog zašto je očitano malo uspješnih sljedova. Ipak, dobiveno je više jako dugačkih sljedova koji su korisni u hibridnom sklapanju genoma (Tablica 4).

Obrađene knjižnice Illumina HiSeq 1 i 2 i Illumina Miseq 1 i 2 sadrže 71,9 % sljedova prisutnih u neobrađenim podatcima, a ukupna duljina smanjena je na 60,7 %. U slučaju knjižnice Illumina Macrogen ti postotci iznose 63,24 % i 58,84 %. Neprekinuti sljedovi dobiveni sklapanjem knjižnice Illumina Macrogen dosta su fragmentiraniji, što se može iščitati iz podataka u Tablici 8 i prikaza kumulativne duljine na Slici 11.

Ispravljanje programom Nanocorr znatno smanjuje ukupni broj i duljinu sljedova (Tablica 6 i Slika 10), ali dodatno produljuje duže sljedove (što je vidljivo iz prosječne duljine i duljine najduljeg slijeda). Program LoRDEC puno je brži, jednostavniji za korištenje i ne dovodi do gubitka sljedova. Usporedbom ispravljenih i neispravljenih sljedova ustanovila sam da program Nanocorr uvodi puno više promjena u sljedove koje ispravlja od programa LoRDEC (Tablica 7). Usporedba sa završenim genomom *E. subterraneus* bila bi idealna za procjenu kvalitete programa za ispravljanje, ali pošto to nije moguće, provela sam usporedbu s genomom srodne vrste. Neispravljeni ONT sljedovi starije generacije očekivano ostvaruju vrlo mali broj pogodaka po jedinici slijeda pri sravnjenju s genomom *A. queenslandica*, dok neispravljeni sljedovi iz knjižnice ONT 5 ostvaruju najviše. Ipak, treba uzeti u obzir da je knjižnica ONT 5 višestruko veća od knjižnica ONT 1-4 što samo po sebi uzrokuje veći broj pogodaka. Sudeći po broju i duljini pogodaka, ispravljanje sljedova programom Nanocorr bilo je uspješnije od ispravljanja programom LoRDEC.

Sklapanje ispravljenih sljedova iz knjižnica ONT 1-4 na neprekinute sljedove dobivene sklapanjem knjižnica Illumina HiSeq i MiSeq smanjilo je fragmentiranost (što je vidljivo iz broja neprekinutih sljedova i vrijednosti N50), ali nije proizvelo dulji slijed od onog već prisutnog (Tablica 10). Dulji sljedovi ispravljeni programom LoRDEC rezultirali su većom prosječnom duljinom neprekinutih sljedova i s najviše pronađenih kompletnih visoko očuvanih gena unutar grupe (Tablica 11). Iz relativno male standardne devijacije udjela ukupno pronađenih gena unutar grupe (0,78 %) može se zaključiti da knjižnice ONT 1-4 ne pridonose značajno točnosti sklopljenih sljedova.

Sklapanja provedena na neprekinutim sljedovima dobivenim od knjižnica Illumina HiSeq i MiSeq uz sljedove iz knjižnice ONT 5 dala su daleko najbolje rezultate što se tiče kvalitete sklapanja i dovršenosti genoma (Tablice 10 i 11). Velik broj sljedova sa smanjenim šumom iz knjižnice ONT 5 uz kvalitetne neprekinute sljedove knjižnica Illumina HiSeq i MiSeq u ovom su slučaju bili ključ za uspješno sklapanje genoma. Zanimljivo je da je dodatak ispravljenih sljedova iz knjižnica ONT 1-4 povećao statistiku N50, ali smanjio duljinu najduljeg neprekinutog slijeda. U sklapanjima iz ove grupe pronađen je i najveći udio visoko očuvanih gena.

Sklapanja s neprekinutim sljedovima dobivenim iz knjižnice Illumina Macrogen imaju prihvatljive vrijednosti N50 i duljine najduljih sljedova, ali i iznimno velik ukupan broj neprekinutih sljedova (Tablica 10). Drugim riječima, stupanj fragmentiranosti sklopljenog genoma ostao je visok i uz prisutnost brojnih dugačkih sljedova iz knjižnice ONT 5. Usprkos većoj ukupnoj duljini, broj pronađenih visoko očuvanih gena ne razlikuje se puno u odnosu na broj pronađen u sklapanjima knjižnica Illumina HiSeq i MiSeq bez knjižnice ONT 5 (Tablica 11). Unutar ove grupe postoji primjetna razlika između sklapanja u slučajevima kad je korištena samo knjižnica Illumina Macrogen i kad su dodani sljedovi iz knjižnice ONT 5 (to je uzrok veće standardne devijacije udjela pronađenih gena unutar grupe).

Sljedovi pronađeni usporedbom rezultata najboljeg sklapanja s bazom podataka NCBI Protein većinom su porijeklom iz životinja (56,20% ukupne sravnjene duljine) i bakterija (38,81% ukupne duljine). Brojnost bakterijskih sljedova može se objasniti kontaminacijom genetičkog materijala spužve pri sekvenciranju zbog prisutnosti simbiontskih i drugih bakterija. Bakterije i kralješnjaci su skupine za koje je dostupno najviše sekvenciranih genoma, što također utječe na rezultate. Koljeno Proteobacteria izdvaja se s najboljim omjerom ukupne duljine i prosječne E-vrijednosti (Slika 12), ali vidljivo je da više koljena iz skupine Metazoa ostvaruje dobar rezultat. Skupine Chordata, Cnidaria i Arthropoda zastupljenije su od koljena Porifera, što nije iznenađujuće s obzirom na broj dostupnih sekvenci za pojedinu skupinu.

Rezultati analize pomoću algoritma LCA sugeriraju da je barem dio dobivenih neprekinutih sljedova porijeklom iz genetičkog materijala spužve i nekontaminiran sljedovima iz drugih organizama. Neprekinuti sljedovi za koje nije bilo moguće odrediti najbližeg zajedničkog pretka predstavljaju ili hibridne sekvence ili (vjerojatnije) regije genoma koje ne kodiraju proteine. Druga koljena osim Porifera prisutna u većem broju u rezultatima u skladu su s onim što se može očekivati kao kontaminacija u organizmu spužve u vodenom okolišu.

Što se tiče sravnjenja s genomom *A. queenslandica*, u svim slučajevima ostvareno je relativno malo prilično kratkih pogodaka (Tablica 14). Sklapanja provedena na neprekinutim sljedovima knjižnice Illumina Macrogen imaju veći broj pogodaka, manje prekida i pogrešno uparenih baza od sklapanja provedenih na neprekinutim sljedovima dobivenim iz knjižnica Illumina HiSeq i MiSeq. Moguće je da su sljedovi knjižnice Illumina Macrogen točniji, ali zbog visokog stupnja fragmentiranosti sklapanja provedena na njima nisu dala najbolje rezultate.

6. ZAKLJUČCI

- Sekvenciranje tehnologijom nanopora na uređaju ONT MinION daje dugačke očitane sljedove čak i u suboptimalnim uvjetima.
- Prednost programa za ispravljanje dugačkih sljedova treće generacije Nanocorr je točnost, a prednosti programa LoRDEC brzina, jednostavnost korištenja i duljina dobivenih sljedova.
- Velik broj očitanih sljedova kod novije generacije sekvenciranja tehnologijom nanopora ONT MinION koristan je kod sklapanja genoma, ali prisutnost čak i manje količine dugačkih sljedova značajno smanjuje fragmentiranost sklopljenog genoma.
- Program npScarf je brz i uspješan u hibridnom sklapanju neprekinutih sljedova. Kvaliteta sklapanja ovisi o fragmentiranosti složenih sljedova na kojima se sklapanje provodi, i količini dugačkih sljedova treće generacije.
- Sklopljeni neprekinuti sljedovi sadrže relativno dobar postotak visoko očuvanih gena, ali dijele dosta mali udio sekvence s genomom *A. queenslandica*. U sklopljenom genomu prisutan je velik broj kontaminacija - sljedova porijeklom iz drugih organizama, najviše bakterija.
- Bolje pročišćavanje spužve prije sekvenciranja da se smanji kontaminacija i dostupnost više poznatih sljedova iz srodnih organizama pomoći će u dovršetku sklapanja genoma *Eunapius subterraneus*.

7. LITERATURA

Adamska M, Degnan SM, Green KM, Adamski M, Craigie A, Larroux C, *i sur* (2007). Wnt and TGF-β Expression in the Sponge Amphimedon queenslandica and the Origin of Metazoan Embryonic Patterning. *PLoS One* **2**: e1031.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Ansorge W, Sproat BS, Stegemann J, Schwager C (1986). A non-radioactive automated method for DNA sequence determination. *J Biochem Biophys Methods* **13**: 315–23.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *i sur* (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **19**: 455–477.

Bedek J, Bilandžija H, Jalžić B (2008). Ogulinska špiljska spužvica Eunapius subterraneus Sket et Velikonja, 1984, rasprostranjenost i ekologija vrste i staništa. *Modruški Zb* **2**: 103–130.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, *i sur* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–9.

Bilandžija H, Bedek J, Jalžić B, Gottstein S (2007). The morphological variability, distribution patterns and endangerment in the Ogulin cave sponge Eunapius subterraneus Sket & Velikonja, 1984 (Demospongiae, Spongillidae). *Nat Croat* **16**: 1–17.

Boža V, Brejova B, Vinar T (2017). DeepNano: Deep recurrent neural networks for base calling in MinION nanopore reads. *PLoS One* **12**: e0178751.

Buchfink B, Xie C, Huson DH (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60.

Cao MD, Nguyen SH, Ganesamoorthy D, Elliott AG, Cooper MA, Coin LJM (2017). Scaffolding and completing genome assemblies in real-time with nanopore sequencing. *Nat Commun* **8**: 14515.

Chamberlain SA, Szöcs E (2013). taxize: taxonomic search and retrieval in R. *F1000Research* **2**: 191.

Compeau PEC, Pevzner PA, Tesler G (2011). How to apply de Bruijn graphs to genome assembly. *Nat Biotechnol* **29**: 987–991.

Conaco C, Tsoulfas P, Sakarya O, Dolan A, Werren J, Kosik KS (2016). Detection of Prokaryotic Genes in the Amphimedon queenslandica Genome. *PLoS One* **11**: e0151092.

El-Metwally S, Hamza T, Zakaria M, Helmy M (2013). Next-Generation Sequence Assembly: Four Stages of Data Processing and Computational Challenges. *PLoS Comput Biol* **9**: e1003345.

Gazave E, Lapébie P, Ereskovsky A V., Vacelet J, Renard E, Cárdenas P, *i sur* (2011). No longer Demospongiae: Homoscleromorpha formal nomination as a fourth class of Porifera. *Anc Anim New Challenges* 3–10doi:10.1007/978-94-007-4688-6_2.

Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res* **25**: 1750–6.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075.

Harcet M, Bilandžija H, Bruvo-Mađarić B, Ćetković H (2010a). Taxonomic position of Eunapius subterraneus (Porifera, Spongillidae) inferred from molecular data – A revised classification needed? *Mol Phylogenet Evol* **54**: .

Harcet M, Roller M, Cetkovic H, Perina D, Wiens M, Muller WEG, *i sur* (2010b). Demosponge EST Sequencing Reveals a Complex Genetic Toolkit of the Simplest

Metazoans. *Mol Biol Evol* **27**: 2747–2756.
Heather JM, Chain B (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**: 1–8.
Hodzic J, Gurbeta L, OmanovicMiklicanin E, Badnjevic A (2017). Overview of Next-generation Sequencing Platforms Used in Published Draft Plant Genomes in Light of Genotypization of Immortelle Plant (Helichrysium Arenarium). *Med Arch* **71**: 288.
Itskovich V, Gontcharov A, Masuda Y, Nohno T, Belikov S, Efremova S, *i sur* (2008). Ribosomal ITS Sequences Allow Resolution of Freshwater Sponge Phylogeny with Alignments Guided by Secondary Structure Prediction. *J Mol Evol* **67**: 608–620.
Kchouk M, Gibrat JF, Elloumi M (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biol Med* **09**: .
Lannoy C de, Ridder D de, Risse J (2017). The long reads ahead: de novo genome assembly using the MinION. *F1000Research* **6**: 1083.
Lu H, Giordano F, Ning Z (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* **14**: 265–279.
Matoničkin I (Školska knjiga: Zagreb, 1990). *Beskralješnjaci: biologija nižih avertebrata*. .
Medvedev P, Pham S, Chaisson M, Tesler G, Pevzner P (2011). Paired de Bruijn Graphs: A Novel Approach for Incorporating Mate Pair Information into Genome Assemblers. *J Comput Biol* **18**: 1625–1634.
Metzker ML (2010). Sequencing technologies — the next generation. *Nat Rev Genet* **11**: 31–46.
Miller JR, Koren S, Sutton G (2010). Assembly algorithms for next-generation sequencing data. *Genomics* **95**: 315–327.
Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**: 263–73.
Narzisi G, Mishra B (2011). Comparing De Novo Genome Assembly: The Long and Short of It. *PLoS One* **6**: e19175.
NCBI Organelle Genome Resources (2018). Porifera, organelle- and plasmid-only records. Dostupno na https://www.ncbi.nlm.nih.gov/genome?term=porifera[orgn] NOT genome[PROP] AND non_genome[filter]. Pristupljeno 19.7.2018.
NCBI Resource Coordinators NR Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bourexis D, Brister JR, Bryant SH, Canese K, Cavanaugh M, Charowhas C, Clark K, Dondoshansky I, Feolo M, Fitzpatrick L, Funk K, Geer LY, Gorelenkov V, Graeff A, Hlavina W, Holmes B, Johnson M, Kattman B, Khotomlianski V, Kimchi A, Kimelman M, Kimura M, Kitts P, Klimke W, Kotliarov A, Krasnov S, Kuznetsov A, Landrum MJ, Landsman D, Lathrop S, Lee JM, Leubsdorf C, Lu Z, Madden TL, Marchler-Bauer A, Malheiro A, Meric P, Karsch-Mizrachi I, Mnev A, Murphy T, Orris R, Ostell J, O'Sullivan C, Palanigobu V, Panchenko AR, Phan L, Pierov B, Pruitt KD, Rodarmer K, Sayers EW, Schneider V, Schoch CL, Schuler GD, Sherry ST, Siyan K, Soboleva A, Soussov V, Starchenko G, Tatusova TA, Thibaud-Nissen F, Todorov K, Trawick BW, Vakatov D, Ward M, Yaschenko E, Zasypkin A, Zbicz K. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **46**: D8–D13.
Pevzner PA, Tang H, Tesler G (2004). De novo repeat classification and fragment assembly. *Genome Res* **14**: 1786–96.
Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, *i sur* (2010). Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol Biol Evol* **27**: 1983–7.
Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, *i sur* (2015). Genomic

data do not support comb jellies as the sister group to all other animals. *Proc Natl Acad Sci U S A* **112**: 15402–7.

Pleše B, Lukić-Bilela L, Bruvo-Mađarić B, Harcet M, Imešek M, Bilandžija H, *i sur* (2011). The mitochondrial genome of stygobitic sponge Eunapius subterraneus: mtDNA is highly conserved in freshwater sponges. *Anc Anim New Challenges* 49–59doi:10.1007/978-94-007-4688-6_6.

Pop M (2009). Genome assembly reborn: recent computational challenges. *Brief Bioinform* **10**: 354–366.

Riesgo A, Farrar N, Windsor PJ, Giribet G, Leys SP (2014). The Analysis of Eight Transcriptomes from All Poriferan Classes Reveals Surprising Genetic Complexity in Sponges. *Mol Biol Evol* **31**: 1102–1120.

Salmela L, Rivals E (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* **30**: 3506–3514.

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**: 5463–7.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–2.

Sket B, Velikonja M (1984). Prethodni isvjestaj o nalazima slatkovodnih spuzvi (Porifera, Spongillidae) u spiljama Jugoslavije. *Deveti Jugoslavenski Speleoloski Kongr* 553–557.

Soest RWM Van, Boury-Esnault N, Vacelet J, Dohrmann M, Erpenbeck D, Voogd NJ De, *i sur* (2012). Global Diversity of Sponges (Porifera). *PLoS One* **7**: e35105.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, *i sur* (2010). The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* **466**: 720–726.

Swerdlow H, Gesteland R (1990). Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res* **18**: 1415–9.

Taylor MW, Radax R, Steger D, Wagner M (2007). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **71**: 295–347.

Vargas S, Schuster A, Sacher K, Büttner G, Schätzle S, Läuchli B, *i sur* (2012). Barcoding Sponges: An Overview Based on Comprehensive Sampling. *PLoS One* **7**: e39345.

Verma M, Kulshrestha S, Puri A (2017). Genome Sequencing. *Methods Mol Biol* **1525**: 3–33.

Wang Y, Yang Q, Wang Z (2014). The evolution of nanopore sequencing. *Front Genet* **5**: 449.

Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva E V. (2013). OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res* **41**: D358–D365.

Webster NS, Taylor MW, Behnam F, Lücker S, Rattei T, Whalan S, *i sur* (2009). Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* **12**: 2070–2082.

Wei S, Williams Z (2016). Rapid Short-Read Sequencing and Aneuploidy Detection Using MinION Nanopore Technology. *Genetics* **202**: 37–44.

World Porifera Database (2018). Species - Eunapius subterraneus Sket & Velikonja, 1984. Dostupno na http://www.marinespecies.org/porifera/porifera.php?p=taxdetails&id=167168. Pristupljeno 19.7.2018.

Ye C, Ma ZS, Cannon CH, Pop M, Yu DW (2012). Exploiting sparseness in de novo genome assembly. *BMC Bioinformatics* **13 Suppl 6**: S1.

Korišteni programi

Albacore	https://github.com/dvera/albacore
BBTools	https://jgi.doe.gov/data-and-tools/bbtools/
BLAST	https://www.ncbi.nlm.nih.gov/BLAST/
BUSCO	https://busco.ezlab.org/
BWA-MEM	http://bio-bwa.sourceforge.net/
ChromasPro2	https://technelysium.com.au/wp/chromaspro/
DIAMOND	https://github.com/bbuchfink/diamond
LoRDEC	www.atgc-montpellier.fr/lordec/
Metrichor	https://metrichor.com/
Nanocorr	https://github.com/jgurtowski/nanocorr
npScarf	https://github.com/mdcao/npScarf
poretools	https://github.com/arq5x/poretools
QUAST	http://quast.sourceforge.net/
R	https://www.r-project.org/
SPAdes	http://cab.spbu.ru/software/spades/

8. PRILOZI

PRILOG 1. Konsenzusni slijed nukleotida dobiven sekvenciranjem automatiziranom Sangerovom metodom uz korištenje početnica ITS2F i ITS2R.

AGTGTGATTGCAGATTCCGTGATCTCGAGTCTTTGACGCAATTGCGCCCTCGGTTT
GACGCCGGGGGCCGTCTGTCTGAGCGTCCGTTTCGTTTGNGNCTCCCCGCGCGGG
CGANCGTTTNNTTTGNTTAACGTTCGNTTAACGCGGAGGNGCGNTGAGGCGTCGT
CCGGAAACGGGCGTCCCTTCAAGTGCGAANCGCTCCGGTTCGGAAGGNCTCGCTC
GCGGCTCGAGNGNCCCTCCTTTNCCACCTTGCGCGTCGGGAACTCGACGANGACA
AGGAGGGAGAGGAGGTTNACCTCGNTCGCGAGGGGNCCGTCGNTCCGAAGAACN
AANACAAGTCGGTTNAATACGCGAGGANCCNTTCTCCCCTTNAAAACGGNGNAA
GNTCTTNCATCCTGGNCCTCAGCTCAGGCGGNCTCCCNGCTANNTNAGATAGAAA
AAAAGAGGGTTTCCTTCCCGGGAAGG

PRILOG 2. Protokol za pripremu mješavine za sekvenciranje na uređaju ONT MinION za verzije kemije MAP006 (R7) i NSK007 (R9). Preuzeto sa službene mrežne stranice proizvođača Oxford Nanopore Technologies (https://community.nanoporetech.com/protocols/) u srpnju i rujnu 2016.

Cijepanje DNA na fragmente veličine 8 kb: Otprilike 1- 1,5 μg izolirane DNA dopuniti s NFW do ukupnog volumena 46 μL i nježno promiješati uvlačenjem u pipetu. Prebaciti u tubicu za fragmentaciju g-TUBE proizvođača Covaris. Centrifugirati 1 minutu na 6000 okretaja po minuti u stolnoj centrifugi Eppendorf 5424, potom okrenuti tubicu na drugu stranu i ponoviti centrifugu.

FFPE popravak lomova: Fragmentiranoj DNA dodati NFW do ukupnog volumena 47 μL pa promiješati inverzijom. Odvojiti 2 μL (alikvot 1). Na ostatak dodati 8,5 μL NFW, 6,5 μL FFPE Repair pufera i 2 μL FFPE Repair Enzyme Mixa, promiješati inverzijom, spustiti tekućinu i inkubirati 15 minuta na 20°C. Dodati 62 μL prethodno vorteksiranih AMPure XP kuglica i nježno promiješati uvlačenjem u pipetu. Inkubirati 5 minuta na sobnoj temperaturi pa peletirati kuglice na magnetskom nosaču. Od supernatanta odvojiti 10 μL (alikvot 2) pa ostatak ukloniti. Na pelet nježno dodati 200 μL 70 postotnog etanola pa ukloniti pipetom nakon 30 sekundi. Ponoviti ispiranje etanolom na isti način pa kratko pustiti da se osuši. Nježno resuspendirati kuglice pipetom u 47 μL NFW. Inkubirati 2 minute na sobnoj temperaturi pa peletirati kuglice na magnetskom nosaču. Prebaciti supernatant u LoBind tubicu i odvojiti 2 μL (alikvot 3).

Popravak i pripremanje krajeva molekule: U supernatant iz prethodnog koraka dodati 7 µL End-prep pufera, 3 µL end-prep Enzyme Mixa i 5 µL NFW, promiješati inverzijom pa spustiti tekućinu. Inkubirati 5 minuta pri 20°C i 5 minuta pri 65°C. Dodati 60 µL prethodno vorteksiranih AMPure XP kuglica i nježno promiješati uvlačenjem u pipetu. Inkubirati 5 minuta na sobnoj temperaturi pa peletirati kuglice na magnetskom nosaču. Od supernatanta odvojiti 10 µL (alikvot 4) a ostatak ukloniti. Dva puta isprati sa 200 µL 70 postotnog etanola. Spustiti tekućinu, peletirati kuglice na magnetskom nosaču i izvući rezidualni etanol pipetom. Kratko sušiti pa nježno resuspendirati kuglice pipetom u 32 µL NFW. Inkubirati 2 minute na sobnoj temperaturi. Peletirati kuglice na magnetskom nosaču i prebaciti supernatant u LoBind tubicu. Odvojiti 2 µL (alikvot 5).

Ligacija adaptera: U supernatant iz prethodnog koraka dodati redom reagense 8 µL NFW, 10 µL Adapter Mix, 2 µL HPA, 50 µL Blunt/TA ligase MM pa nježno promiješati inverzijom. Spustiti tekućinu i inkubirati 10 minuta na sobnoj temperaturi. Dodati 1 µL sidra HPT, promiješati inverzijom pa spustiti tekućinu. Inkubirati 10 minuta na sobnoj temperaturi.

Pročišćavanje: Odvojiti 50 µL u MyOne C1 kuglica u DNA LoBind tubicu, dva puta ih isprati sa po 100 µL pufera BBB uz peletiranje na magnetskom nosaču i na kraju resuspendirati u 100 µL BBB. Dodati kuglice u smjesu dobivenu na kraju ligacije adaptera i nježno resuspendirati pipetom. Inkubirati 5 minuta na sobnoj temperaturi pa peletirati kuglice na magnetskom nosaču. Odvojiti 10 µL supernatanta (alikvot 6) i ostatak ukloniti pipetom. Kuglice resuspendirati pipetom u 150 µL BBB, peletirati na magnetskom nosaču i ukloniti supernatant. Ponoviti ispiranje s BBB na isti način. Spustiti tekućinu, peletirati na magnetskom nosaču i ukloniti rezidualni BBB.

Elucija: Pomoću pipete resuspendirati pelet kuglica iz prethodnog koraka u 26 µL pufera ELB. Inkubirati 10 minuta pri 37°C pa peletirati na magnetskom nosaču. Supernatant je pripremna mješavina za sekvenciranje. Odvojiti 2 µL (alikvot 7), ostatak prebaciti u DNA LoBind tubicu i odmah pohraniti na -80°C. Alikvote također pohraniti na -80°C.

PRILOG 3. Koncentracije DNA (u ng/μL) izmjerene u alikvotima odvojenim tijekom pripreme knjižnica za sekvenciranje tehnologijom nanopora.

knjižnica	**alikvot**					
	1	**2**	**3**	**4**	**5**	**6**
R7 1	-	-	20,2	-	14,0	-
R7 2	14,2	-	8,7	-	7,3	-
R9	67,2	1,83	13,9	prenisko za detekciju	15,1	1,35

Životopis

OSOBNE INFORMACIJE

Dunja Glavaš
Horvaćanska 21, 10000 Zagreb
datum rođenja: 3.4.1992.
spol: ženski
državljanstvo: hrvatsko

OBRAZOVANJE I OSPOSOBLJAVANJE

10/2014 - 09/2018	Diplomski studij molekularne biologije Biološki odsjek, Prirodoslovno-matematički fakultet, Sveučilište u Zagrebu znanje i vještine vezane uz laboratorijski rad, programiranje, statistiku i analizu podataka; znanje iz područja molekularne biologije
2010 - 2014	Preddiplomski studij molekularne biologije Biološki odsjek, Prirodoslovno-matematički fakultet, Sveučilište u Zagrebu znanje i vještine vezane uz laboratorijski rad; znanje iz područja molekularne biologije

TEČAJEVI

13. - 17.6.2016.	Exaltum Bioinformatics and Statistics for Next Generation Sequencing, Zagreb, Hrvatska