

# GENOMIKA I OSOBITOSTI PONAVLJAJUĆIH DNA ŠKOLJKAŠA IZ PORODICE OSTREIDAE

---

Tunjić Cvitanić, Monika

Doctoral thesis / Doktorski rad

2024

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://um.nsk.hr/um:nbn:hr:217:270005>

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

Download date / Datum preuzimanja: **2024-08-16**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)





University of Zagreb

FACULTY OF SCIENCE  
DEPARTMENT OF BIOLOGY

Monika Tunjić Cvitanić

**GENOMICS AND PECULIARITIES  
OF REPETITIVE DNAs OF THE OYSTERS  
IN THE FAMILY OSTREIDAE**

DOCTORAL THESIS

Zagreb, 2024



Ovaj rad izrađen je u „Laboratoriju za strukturu i funkciju heterokromatina“ u Zavodu za molekularnu biologiju Instituta Ruđer Bošković u Zagrebu, pod mentorstvom nasl. prof. dr. sc. Miroslava Plohla i nasl. doc. dr. sc. Eve Šatović Vukšić u sklopu Sveučilišnog poslijediplomskog doktorskog studija Biologije pri Biološkom odsjeku Prirodoslovno-matematičkog fakulteta, Sveučilišta u Zagrebu.



Sveučilište u Zagrebu

Doktorski rad

Prirodoslovno-matematički fakultet

Biološki odsjek

GENOMIKA I OSOBITOSTI PONAVLJAJUĆIH DNA ŠKOLJKAŠA IZ PORODICE  
OSTREIDAE

Monika Tunjić Cvitanić

Institut Ruđer Bošković

Ponavljajuća DNA u mnogim organizmima čini velik, ali neistražen dio genoma. Za vrstu *Crassostrea gigas* bioinformatički je utvrđeno da repeatom dominantno grade satDNA i PGE, a najbrojnija satDNA je Cg170. Bioinformatički su utvrđeni satelitomi srodnih vrsta kamenica *C. gigas*, *Crassostrea angulata*, *Crassostrea virginica* i *Ostrea edulis*. U sklopu centromere *C. gigas* najbrojnije DNA su transpozoni i satDNA, a u heterokromatinu najbrojniji su DNA transpozoni. Više od 75 % satelitoma te vrste pokazalo je sličnost s elementima Helitron. Analizom postranih regija monomera ili nizova satDNA utvrđeno je da se samo jedna satDNA nalazi isključivo u sklopu elementa Helitron, dvjema je većina monomera/nizova monomera samostalna, a osam satDNA u podjednakim je omjerima prisutno samostalno i u sklopu elementa. Niz satDNA zajednički je svim ispitanim vrstama, a *C. gigas* i *C. angulata* dijele najveći broj zajedničkih satDNA. U *C. gigas* nije otkriveno tipično nakupljanje satDNA na određenim kromosomskim pozicijama, čime je u ovoj vrsti ustanovljen novi (visokoraspršeni) obrazac organizacije satDNA na razini genoma.

(239 stranica, 31 slika, 26 tablica, 4 grafikona, 302 literaturna navoda, jezik izvornika na hrvatskom)

**Ključne riječi:** školjkaši, satelitom, repeatom, centromere, heterokromatin

Mentori: emeritus prof. dr. sc. Miroslav Plohl, IRB, Zagreb; doc. dr. sc. Eva Šatović Vukšić, znanstvena suradnica, IRB, Zagreb

Ocjenjivači: prof. dr. sc. Višnja Besendorfer, redovita profesorica u trajnom zvanju, PMF, Zagreb; prof. dr. sc. Đurđica Ugarković, znanstvena savjetnica u trajnom zvanju, IRB, Zagreb; prof. dr. sc. Petra Korać, redovita profesorica, PMF, Zagreb

Rad prihvaćen:



University of Zagreb

Doctoral thesis

Faculty of Science

Department of Biology

GENOMICS AND PECULIARITIES OF THE REPETITIVE DNAs OF THE OYSTERS IN  
THE FAMILY OF OSTREIDAE

Monika Tunjić Cvitanić

Ruđer Bošković Institute

Repetitive DNA constitutes big, but still uninvestigated genome part. For the oyster *Crassostrea gigas* it was bioinformatically determined that the repeatome is mainly constituted of satDNAs and mobile elements, and the most abundant satDNA is Cg170. The satellitomes of four closely related species, *C. gigas*, *Crassostrea angulata*, *Crassostrea virginica* and *Ostrea edulis*, were bioinformatically determined. The most abundant sequences in the centromere of *C. gigas* are DNA transposons and satDNAs, while in heterochromatin the most abundant are DNA transposons. More than 75% of satellitome of this species showed similarity to Helitrons. Analysis of conserved boxes adjacent to monomers or tandem repeats of satDNAs showed that only one satDNA exists integrated into Helitrons, two have majority of their monomers/tandem repeats as standalone organizational form, and eight satDNAs have equally present monomers/tandem repeats in both forms. All analyzed species have several satDNAs in common, but *C. gigas* and *C. angulata* share the highest number of common satDNAs. SatDNAs in *C. gigas* did not show typical clustering at certain chromosomal positions but new organizational pattern: highly scattered across the genome.

(239 pages, 31 figures, 26 tables, 4 graphs, 302 references, original in Croatian)

**Key words:** bivalves, satellitome, repeatome, centromere, heterochromatin

Mentors: Emeritus Prof. Dr. Miroslav Plohl, IRB, Zagreb; Assoc. Prof. Dr. Eva Šatović Vukšić, Scientific Associate, IRB, Zagreb

Reviewers: Prof. Dr. Višnja Besendorfer, Full Professor, PMF, Zagreb

Prof. Dr. Đurđica Ugarković, Scientific Advisor, IRB, Zagreb

Prof. Dr. Petra Korać, Full Professor, PMF, Zagreb

Thesis accepted:





## Sadržaj

1. UVOD.....	11
2. LITERATURNI PREGLED.....	15
2.1. Satelitna DNA .....	17
2.1.1. Strukturna obilježja satDNA.....	17
2.2. Evolucija i strukturna dinamika satDNA .....	19
2.3. Otkriće PGE .....	22
2.4. Tipovi pokretnih genetičkih elemenata .....	22
2.4.1. Elementi LINE .....	24
2.4.2. Elementi SINE (od engl. short interspersed nuclear elements, SINEs).....	25
2.4.3. DNA transpozoni .....	26
2.4.4. Elementi Helitron.....	26
2.5. Povezanost sekvencija satDNA i PGE.....	30
2.6. Obilježja heterokromatina i njegova povezanost s ponavljajućim DNA .....	33
2.7. Centromere .....	35
2.8. Obilježja školjkaša i školjkaških genoma .....	39
2.8.1. Obilježja školjkaša .....	39
2.8.2. Kamenice (Ostreidae) .....	46
2.8.3. Heterokromatin školjkaša .....	47
2.9. Ponavljajuće sekvencije DNA u genomima školjkaša .....	48
2.9.1. SatDNA u porodici Ostreidae .....	48
2.9.2. Pokretni genetički elementi u genomima školjkaša.....	49

2.10.	Analize genoma i detekcija ponavljajuće DNA bioinformatičkim metodama.....	52
3.	MATERIJALI I METODE.....	55
3.1.	Biološki materijal .....	56
3.2.	Izolacija genomske DNA školjkaša za potvrdu vrste.....	56
3.3.	Izolacija nativnog kromatina i razgradnja mikrokokalnom nukleazom.....	58
3.4.	Testiranje antitijela za protein Cg-CenH3 metodom <i>western blotting</i> -a.....	58
3.5.	Testiranje specifičnosti antitijela za protein Cg-CenH3 .....	60
3.6.	Nativna kromatinska imunoprecipitacija .....	61
3.7.	Sekvenciranje DNA.....	63
3.8.	Obrada jedinki vrste <i>C. gigas</i> korištenih u citogenetičkim pokusima.....	63
3.9.	Priprema citogenetičkih preparata s mitotskim kromosomima.....	64
3.10.	Priprema citogenetičkih preparata s gonadnim stanicama .....	64
3.11.	Označavanje sonde za pokuse imunofluorescencije .....	65
3.11.1.	Označavanje sonde za FISH .....	65
3.11.2.	Označavanje sonde DNA iz frakcija kromatina u asocijaciji sa centromernim histonom H3 i histonom H3 trimetiliranim na lizinu 9 .....	70
3.12.	Fluorescencijska hibridizacija <i>in situ</i> .....	70
3.13.	Pokusi imunofluorescencijske detekcije (IF).....	72
3.13.1.	Imunofluorescencijska detekcija .....	72
3.13.2.	Dvostruka imunofluorescencijska detekcija.....	72
3.13.3.	IF-FISH .....	72
3.14.	Računalne analize .....	73

3.14.1.	Analize dostupne unutar paketa <i>RepeatExplorer</i> .....	73
3.14.2.	Analiza <i>RepeatExplorer</i> .....	73
3.14.3.	Analiza <i>ChIP-Seq Mapper</i> .....	74
3.14.4.	Modificirana analiza <i>ChIP-Seq Mapper</i> .....	75
3.14.5.	Analiza <i>TAREAN</i> .....	75
3.15.	Usporedba s bazom podataka <i>Rebase</i> .....	76
3.16.	Analize napravljene u programu <i>Geneious</i> .....	76
3.16.1.	Usporedba sekvencija satDNA dobivenih analizama različitih setova očitavanja u analizi <i>TAREAN</i> .....	76
3.16.2.	Lokalizacija detektiranih satDNA na kromosomima vrsta <i>C. gigas</i> .....	77
3.16.3.	Određivanje postranih regija satDNA vrste <i>C. gigas</i> .....	77
3.16.4.	Lokaliziranje regija sličnosti s elementima Helitron u izdvojenim postranim regijama	77
4.	REZULTATI .....	79
4.1.	Izolacija kromatina i razgradnja mikrokokalnom nukleazom.....	80
4.2.	Testiranje antitijela anti-Cg-CenH3 .....	81
4.3.	Imunoprecipitacija kromatina asociiranog s proteinom CenH3.....	82
4.4.	Lokalizacija proteina Cg-CenH3.....	83
4.5.	Cjelokupni sastav ponavljajućih sekvencija DNA u genomu vrste <i>C. gigas</i> .....	86
4.6.	Sekvencije koje grade centromere u genomu vrste <i>C. gigas</i> .....	87
4.7.	Karakterizacija najzastupljenijih centromerno obogaćenih sekvencija u genomu vrste <i>C. gigas</i>	98

4.7.1.	Kolokalizacija ukupne imunoprecipitirane DNA i proteina Cg-CenH3 na jezgrama gonadnih stanica vrste <i>C. gigas</i> .....	102
4.7.2.	Lokalizacija centromerne satDNA Cg170 na metafaznim kromosomima vrste <i>C. gigas</i> i njezina kolokalizacija s proteinom Cg-CenH3 na jezgrama spolnih stanica.....	103
4.7.3.	Lokalizacija obogaćenih ponavljajućih sekvencija DNA detektiranih u centromernoj frakciji imunoprecipitiranog kromatina vrste <i>C. gigas</i> .....	106
4.8.	Kolokalizacija ponavljajućih sekvencija DNA obogaćenih u centromernoj frakciji vrste <i>C. gigas</i> s proteinom Cg-CenH3 .....	109
4.9.	Sekvence DNA koje izgrađuju heterokromatin vrste <i>C. gigas</i> .....	111
4.9.1.	Lokalizacija sekvencija iz imunoprecipitiranog heterokromatina na metafaznim kromosomima i jezgrama gonadnih stanica vrste <i>C. gigas</i> te kolokalizacija proteina Cg-CenH3 i H3K9me3 na gonadnim stanicama vrste <i>C. gigas</i> .....	115
4.10.	Analiza satelitoma pacifičke kamenice <i>C. gigas</i> .....	118
4.10.1.	Sastav satDNA u genomu pacifičke kamenice <i>C. gigas</i> .....	118
4.10.2.	Anotacija <i>in silico</i> satDNA u genomu <i>C. gigas</i> .....	122
4.11.	Određivanje dominantnog organizacijskog oblika ponavljajuće DNA povezane s PGE Helitron u genomu <i>C. gigas</i> .....	129
4.12.	Fluorescencijska hibridizacija <i>in situ</i> najistaknutijih satDNA.....	136
4.13.	Satelitomi triju srodnih školjkaških vrsta .....	138
4.14.	Satelitom vrste <i>C. angulata</i> .....	138
4.15.	Satelitom vrste <i>C. vriginica</i> .....	140
4.16.	Satelitom vrste <i>O. edulis</i> .....	142
4.17.	Komparativna analiza satelitoma četiriju vrsta kamenica .....	144
5.	RASPRAVA.....	165

6. ZAKLJUČAK.....	182
7. REFERENCE .....	184
8. PRILOZI.....	219
9. ŽIVOTOPIS.....	238



## **1. UVOD**





Ponavljajuće DNA vrlo često zauzimaju značajne udjele eukariotskih genoma. U nekih biljaka one čine i preko 90 % genoma, u genomu čovjeka udio im premašuje 50 % (Lander i sur., 2001., Feschotte i sur., 2002., Messing i sur., 2004.). Ponavljajuću DNA čine uzastopno ponovljene sekvencije, satelitna DNA (satDNA), i raspršeno ponovljene sekvencije, pokretni genetički elementi (PGE). Iako ih se dugo smatralo *otpadom* genoma, tzv. *junk DNA* (Ohno, 1972.), strukturni i funkcionalni značaj ponavljajućih DNA potvrđuje sve veći broj istraživanja (Ugarković, 2005., Pezer i Ugarković, 2012., Feliciello i sur., 2015., Garrido-Ramos, 2017., Louzada i sur., 2020.). SatDNA uključene su u funkcionalnost, građu i evoluciju kromosoma, imaju ulogu u strukturnoj stabilnosti cjelokupnog genoma, ali satDNA mogu biti i regulatori ekspresije gena (Feliciello i sur., 2015.).

Centromere su kromosomske strukture koje imaju ključnu ulogu u pravilnom razdvajanju kromosoma tijekom stanične diobe (McKinley and Cheeseman, 2016.). Kod većine biljnih i životinjskih vrsta centromere su bogate sekvencijama satDNA (Talbert i sur., 2018.). Epigenetički su centromere označene prisutnošću centromernog histona H3. Heterokromatin predstavlja kromosomske domene koje su u pravilu transkripcijski utišane, a epigenetička oznaka transkripcijski utišanog heterokromatina je trimetilacija Lys9 na histonu H3. U području centromere nalazi se centrokromatin koji je okružen područjima pericentromernog heterokromatina kojega karakterizira prisutnost dugih nizova satDNA isprekidanih s umetnutim PGE (Almouzni i Probst, 2007.). U heterokromatinu, osobito u (peri)centromernom kromatinu, satDNA i PGE mogu interagirati i biti povezani na različite načine: preferencijalna mjesta ugradnje PGE mogu biti regije bogate satDNA, satDNA se po genomu mogu širiti uz pomoć PGE, satDNA mogu nastati iz njih, ili biti sastavni dio pokretnih elemenata (pregled u Meštrović i sur., 2015.). Regije genoma bogate ponavljajućom DNA vrlo često su prepreka u slaganju sekvenciranih genoma te su podzastupljene u složenim genomima (Plohl i sur., 2012., Sedlazeck i sur., 2018.). Problem predstavlja nemogućnost utvrđivanja točnog redoslijeda sekvencija i procjena duljina regija koje tvore ponavljajuće DNA velike sličnosti, ponovljene velik broj puta (Šatović Vukšić i sur., 2020.).

Školjkaši predstavljaju gospodarski i ekološki značajnu skupinu beskralješnjaka. Važnost im se prije svega temelji na uzgoju u akvakulturi, a zbog načina na koji se hrane, filtriranjem, imaju

veliki učinak na čistoću okoliša u kojem žive te se zbog toga koriste kao živi sustavi za biomonitoring okoliša (Vaughn i Hoellein, 2018.). Nadalje, u zadnjih nekoliko godina značaj u istraživanjima im je porastao zbog toga što se veliki broj saznanja o nastanku i mehanizmima širenja neoplazija i nastanku ostalih bolesti kod ljudi temelji upravo na proučavanju biologije školjkaša (Robledo i sur., 2018.), a peptidi izolirani iz školjkaša potencijalni su novi antibiotici (Gerdol i Venier, 2015.). U istraživanju ponavljajućih DNA, repeatoma i satelitoma, osobito je značajna vrsta *Crassostrea gigas* zbog nekoliko istaknutih obilježja svog genoma. Ta vrsta ima velik broj neklasificiranih ponavljajućih sekvencija (Zhang i sur., 2012., Peñaloza i sur., 2021.), velik broj elemenata Helitron (Peñaloza i sur., 2021.) te nisku količinu heterokromatina (Bouilly i sur., 2008.). Važna osobitost genoma ove vrste je i prisutnost kratkih nizova satDNA uklopljenih u elemente Helitron/Helentron (Gaffney i sur., 2003., Vojvoda Zeljko i sur., 2020.).

U zadnjih nekoliko godina, s razvojem nove generacije sekvenciranja, broj sekvenciranih genoma školjkaša značajno je narastao. U svega tri godine, sekvencirano je 60-ak genoma (prema bazi *National Center for Biotechnology Information*, NCBI, podaci iz kolovoza 2023.). Analizom velikog broj kratkih sekvencija koje daju nisku pokrivenost genoma novim bioinformatičkim alatima može se dobiti potpuniji uvid u ponavljajuću DNA genoma školjkaša, odnosno repeatom i satelitom (Weiss-Schneeweiss i sur., 2015., Ruiz-Ruano i sur., 2016., Pita i sur., 2017.).

Svrha ovog rada produbljuje znanja o raspodjeli, položaju, udjelu, evoluciji, i mogućoj ulozi pojedinih tipova ponavljajuće DNA u različitim segmentima genoma školjkaša, koje je još uvijek nepotpuno. Razumijevanje strukture i evolucije ponavljajućih DNA važno je za razumijevanje strukturne dinamike i evolucije eukariotskih genoma u cjelini. Ciljevi ovog istraživanja su utvrditi i okarakterizirati ponavljajuće sekvencije specifične za centromerni i heterokromatinski dio genoma sekvenciranjem DNA dobivene kromatinskom imunoprecipitacijom antitijelima za epigenetske oznake centromernog kromatina i heterokromatina te definirati cjelokupni repeatom i satelitom vrste *C. gigas* sekvenciranjem ukupne genomske DNA i korištenjem bioinformatičkih alata za analizu DNA iz imunoprecipitiranog kromatina i ukupne genomske DNA. Pretpostavljeno je da će udio ponovljenih sekvencija biti nizak te će se razlikovati u heterokromatinskim i centromernim

segmentima u odnosu na ostatak genoma. Nadalje, cilj je usporediti satelitom vrste *C. gigas* s tri srodne vrste (*Crassostrea angulata*, Lamarck, 1819., *Crassostrea virginica*, Gmelin, 1719. i *Ostrea edulis*, Linnaeus, 1759.). Pretpostavljeno je da će uspoređene vrste imati set satelitnih sekvencija koji će biti specifičan za svaku od njih, a najrodnije vrste dijelit će pojedine satDNA.

Sljedeći je cilj istražiti vezu satDNA i PGE u vrsti *C. gigas*. Pretpostavljeno je da će većina satDNA u genomu vrste *C. gigas* biti uklopljena u PGE (npr. Helitrone), a udio klasičnih satelitnih nizova bit će mali. Konačno, neke od ponavljajućih sekvencija detektiranih u centromernom kromatinu i heterokromatinu te repeatomu i satelitomu vrste *C. gigas* lokalizirati eksperimentalnim metodama na metafaznim kromosomima i bionformatičkim metodama, *in silico*, na sastavljenim kromosomima.



## **2. LITERATURNI PREGLED**



Genomi eukariotskih organizama sastoje se od kodirajućih i nekodirajućih sekvencija DNA. Nekodirajuća DNA u promjenjivim je udjelima sačinjena od ponavljajućih sekvencija DNA (ponavljajuće DNA) koje su u genomu ponovljene od nekoliko puta do nekoliko tisuća puta. One predstavljaju glavni faktor o kojem ovisi veličina genoma nekog organizma (Wright, 2017.) te su neposredni uzrok tzv. enigme C-vrijednosti (Elliot i Gregory, 2015.). Cjelokupnu ponavljajuću DNA prisutnu u nekom genomu nazivamo repeatom (Kim i sur., 2014.). S obzirom na organizaciju u genomu, ponavljajuću DNA dijelimo na uzastopno ponavljajuću (satDNA, ali i uzastopno ponovljeni geni) i raspršeno ponavljajuću DNA (pokretni genetički elementi, PGE, primjerice DNA transpozoni i retrotranspozoni). U genomima, ponovljene mogu biti i retrotransponirane sekvencije, poput retrogena i retropseudogena, kao i genskih porodica. Također, i dijelovi DNA veći od 1 kb mogu biti ponovljeni uzastopno ili raspršeno, a nazivaju se segmentalne duplikacije (od engl. segmental duplications, SDs) (Ruiz-Ruano i sur., 2016.). Šezdesetih godina prošlog stoljeća izvedeni su prvi pokusi kojima je pokazano da udio ponavljajuće DNA u genomu može značajno varirati od organizma do organizma (Britten i Kohne, 1968.). Autori su pokusima reasocijacije disociranih lanaca izolirane genomske DNA utvrdili da brzina reasocijacije ovisi o udjelu i načinu ponavljanja u genomu. Na taj je način utvrđeno da DNA viših eukariota reasocira kao tri glavne komponente: sporo reasocirajuća DNA (geni prisutni u jednoj kopiji u genomu), umjereno brzo reasocirajuća DNA (raspršeno ponovljene sekvencije) i brzo reasocirajuća DNA (uzastopno ponovljene sekvencije). Također, udjeli različitih tipova ponavljajućih DNA značajno se razlikuju od organizma do organizma. Primjerice, udio satDNA u nekim biljnim genomima varira između 0,1 i 36 %, u nekim životinjskim vrstama može iznositi manje od 0,5 % i više od 50 %. Udio PGE u genomu čovjeku iznosi preko 50 % dok u nekim biljnim genomima taj udio doseže 90 % genoma (Lander i sur., 2001., Feschotte i sur., 2002., Messing i sur., 2004.). U nekim vrstama postoje velike razlike u udjelu satDNA i PGE u genomu. Primjerice, u zobi, udio ponavljajuće DNA iznosi oko 70 % od čega satDNA čine svega 2 % (Liu, 2019.). Iako se dugo smatralo da je ponavljajuća DNA svojevrsno *smeće* u genomu (od engl. junk DNA; Ohno 1972.) ili dio genoma kojemu je svrha samo vlastito umnažanje pa je zbog toga nazvana *sebična DNA* (Orgel i Crick, 1980.), sve veći broj istraživanja pokazuje njezinu veliku funkcionalnu važnost (npr. Ugarković, 2005.).



## 2.1. Satelitna DNA

### 2.1.1. Strukturna obilježja satDNA

Sekvencije DNA koje su u genomu uzastopno ponovljene veliki broj puta u orijentaciji glava-rep te izgrađuju duge nizove svojstvene heterokromatinskim genomskim odjeljcima tradicionalno se nazivaju satelitne DNA, prema dodatnoj *satelitnoj* vrpici, vidljivoj pri centrifugiranju genomske DNA u gradijentu gustoće (Kit, 1961.). Prema veličini osnovne jedinice ponavljanja (monomera), duljini niza kojeg tvori, stupnju ponovljenosti, broju lokusa na kojima se pojavljuje te lokaciji na kromosomu, satDNA mogu se dodatno podijeliti na mikrosatelitne, minisatelitne, satelitne i makrosatelitne DNA. Konsenzus o točnim duljinama monomernih sekvencija za svaku od navedenih skupina još nije donešen (Tautz, 1993.; Richard i sur., 2008.), ali se satDNA od, primjerice, mikrosatelitne i minisatelitne DNA može razlikovati po tome što tvori puno dulje nizove (od desetaka kilobaznih parova do megabaznih parova nukleotida) koji često tvore blokove konstitutivnog heterokromatina u pericentromernim i telomernim (Eymery i sur., 2009.; Schueler i Sullivan, 2006.), ali i interkalarnim regijama kromosoma sa specifičnim strukturnim i epigenetičkim obilježjima (Garrido-Ramos, 2017.). Monomerne sekvencije satDNA najčešće su duge 150-180 pb ili 300-360 pb što odgovara veličini nukleotidne sekvencije potrebne za omatanje DNA oko jednog ili dva nukleosoma te se smatra evolucijski favoriziranom veličinom monomera (Henikoff i sur., 2001., Heslop-Harrison i Schwarzacher, 2013.). No, duljina im može iznositi od nekoliko pb (Lohe i Roberts, 1988.) do nekoliko kpb (Benedum i sur., 1986.). SatDNA u genomima nekih vrsta zauzimaju i preko 50 % (Elder and Turner, 1995.; Schmidt i Heslop-Harrison, 1998.).

SatDNA obično ima promijenjen udio A i T nukleotida u odnosu na ostatak genoma. Kada se izolirana visokomolekularna genomska DNA centrifugira u gradijentu cezijeveg klorida, vidljive su tri vrpce: glavni dio genomske DNA, dio bogat bazama A i T koji se nadsvođuje nad glavnim dijelom genomske DNA, te dio bogat bazama G i C koji se smješta ispod glavnog dijela genomske DNA (Kit, 1961.). Iako se satDNA od ostatka genoma u pravilu razlikuje po većem udjelu A i T nukleotida, satDNA može karakterizirati i veći udio G i C nukleotida u odnosu na ostatak genoma (npr. Petrović i Plohl, 2005.)

Uzastopna ponovljenost satelitnih monomera u dugim nizovima potvrđena je istraživanjem satDNA razgradnjom restriktivnim enzimima (Bachmann i sur., 1991.). Kada se cjelokupna genomska DNA razgradi restriktivnim enzimom čije mjesto prepoznavanja postoji u monomerima uzastopno ponovljene sekvencije i dobiveni produkt razgradnje elektroforetski razdvoji na agaroznom ili poliakrilamidnom gelu dobiva se karakteristični tzv. ljestvičasti obrazac elektroforetičkih vrpca. Ovaj obrazac posljedica je cijepanja dugih nizova monomera i, zbog varijabilnosti u sekvenciji DNA koju prepoznaje taj restriktivni enzim u pojedinim monomerima, veličine koji su višekratnici monomerne duljine satDNA. Što je više vrpca prisutno, to je više mutiranih restriktivnih mjesta unutar niza neke satDNA.

SatDNA u genomu neke vrste razlikuju se po nukleotidnom slijedu, duljini monomera, složenosti sekvencija, i zastupljenosti u genomu, a monomeri mogu imati različita strukturalna svojstva i evolucijsko porijeklo (Plohl i sur., 2012.). Iznimno velike razlike u veličini monomera satDNA možemo pronaći čak i u istoj vrsti, primjerice u mušici *Drosophila melanogaster* u kojem može iznositi od 4 pb do 359 pb (pregled u Sneideman i Meller, 2021.)

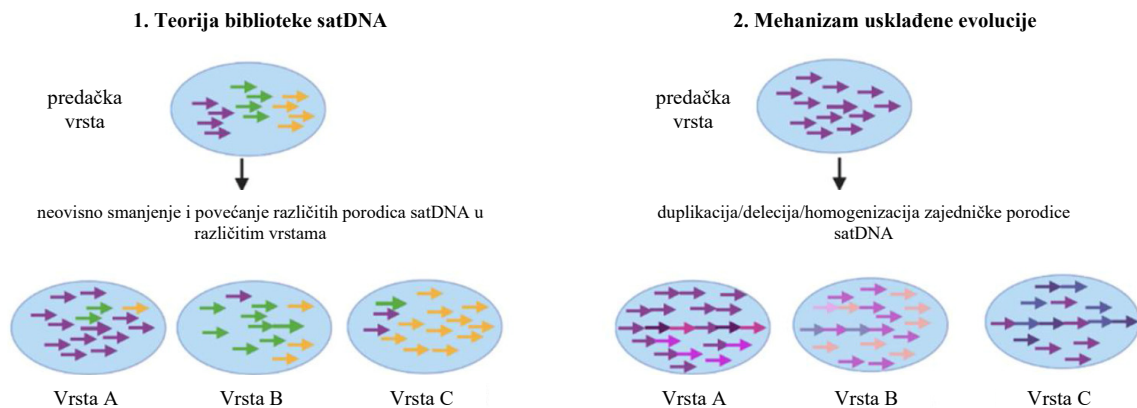
Monomeri satDNA mogu tvoriti tzv. ponavljanja višeg reda (od engl. higher-order repeats, HOR) u kojima nekoliko monomera (dva ili više) čini novu i veću osnovnu jedinicu uzastopnih ponavljanja. Ponavljanja višeg reda pronalazimo npr. u centromernom području kromosoma pripadnika nadporodice Hominoidea, uključujući rod Homo (Sujiwattanarat i sur., 2015.). U genomu čovjeka, takvu organizaciju ima alfa-satelitna DNA kojoj je osnovna jedinica ponavljanja duga ~170 pb te je sastavni dio centromere (Koga i sur., 2014.). U sklopu monomera alfa-satDNA čovjeka pronalazimo i karakterističan konzervirani *CENP-B* motiv, sekvenciju dugu 17 pb, koji predstavlja mjesto interakcije proteina CENP-B sa centromernom DNA tijekom odvajanja kromosoma u diobi (Romanova i sur., 1996.).

Osim u heterokromatinskim regijama genoma, satDNA pronalazimo i u eukromatinskim regijama u kojima se *heterokromatinske satDNA* nalaze u obliku kratkih ponavljanja, samostalnih monomera ili njihovih fragmenata (Kuhn, 2015.; Ruiz-Ruano i sur., 2016.; de Lima i sur., 2017.; Chaves i sur., 2017.) te mogu biti smještene u blizini kodirajućih regija (Brajković i sur., 2018.). No, u eukromatinskim regijama pronalazimo i dugačke nizove satDNA, karakteristične samo za eukromatinske regije (Pavlek i sur., 2015.; Pita i sur., 2017.). Novija istraživanja potvrđuju postojanje fenomena transkripcije satDNA, uključivo i u

heterokromatinskim područjima. Primjerice, transkribirane satDNA u mušici *D. melanogaster* mogu sudjelovati u stvaranju heterokromatina kao male interferirajuće molekule RNA (od engl. small-interfering RNA, siRNA) u koje se procesiraju nakon transkripcije (Usakin i sur., 2007.).

## 2.2. Evolucija i strukturna dinamika satDNA

Sedamdesetih godina prošlog stoljeća, teorijom satelitne biblioteke pretpostavljeno je da blisko srodne vrste dijele određeni skup satDNA, porijeklom od zajedničkog pretka tih vrsta, unutar kojeg pojedine satDNA variraju u broju kopija kod svake vrste, a u određenoj mjeri i nukleotidnoj sekvenciji (Slika 1., Fry i Salser, 1977.). Vrsno specifični profili satDNA (Ugarković i Plohl, 2002.; Plohl, 2008.) nastaju kada se promijene odnosi u broju kopija satDNA u nekoj vrsti koja dijeli tu biblioteku (Fry i Salser, 1977.). Prvi je put to eksperimentalno potvrđeno na vrstama kukaca iz roda *Palorus* (Meštrović i sur., 1998.) u kojima su prisutne četiri različite satDNA, a u svakoj od ispitanih vrsta jedna se satDNA izdigla kao dominantna. Do sad je ovaj model, kao i pretpostavka da se u genomima najčešće nalazi veliki broj različitih satDNA, potvrđen za velik broj životinjskih i biljnih vrsta (npr. Macas i sur., 2015.; Silva i sur., 2017.; Pita i sur., 2017.; Palacios-Gimenez i sur., 2018.). Tako genom graška sadrži 16 različitih satDNA (Macas i sur., 2007.), u genomu europskog skakavca nalazimo 62 satDNA (Ruiz-Ruano i sur., 2016.), u genomu skakavca *Vandiemenella viatica* 129 (Palacios-Gimenez i sur., 2020.), a u genomu ribe *Megaleporinus microcephalus* okarakterizirane su 164 satDNA (Utsunomiya i sur., 2019.).



Slika 1. Grafički prikaz teorije biblioteke satDNA i mehanizma usklađene evolucije (slika prilagođena iz Thakur i sur., 2021.).

Sekvencije monomera neke porodice satDNA često su međusobno puno sličnije unutar jedne jedinice te drugih jedinki iste vrste nego između jedinki dalje srodnosti. SatDNA unutar vrste evoluiraju međusobno ovisno, odnosno usklađeno, a nova varijanta neke sekvencije simultano se širi porodicom neke satDNA (homogenizacija) i među jedinkama određene populacije iste vrste (fiksacija). Time se varijabilnost sekvencije na razini genoma održava niskom te fiksiranom na razini vrste (obično <5 %) (Elder i turner 1995.; Dover 1982., 1986.; Plohl i sur. 2008., 2012.). Proces kojim se objašnjava usklađena evolucija (Zimmer i sur., 1980.) monomera satDNA naziva se molekularni *drive* (Dover, 1982.) te je on neovisan o prirodnoj selekciji i molekularnom *drift*-u (Dover, 1986.). Homogenizacija i fiksacija satDNA unutar vrste događa se puno brže nego što bi to bilo za očekivati kad bi one bile podložne isključivo procesu molekularnog *drift*-a i prirodne selekcije (Dover, 1982.). Rekombinacijski procesi kojima se objašnjava brzo širenje novih varijanti monomera satDNA ili dugo održavanje postojećih (Strachan i sur., 1985.) su: nejednoliko prekrivanje, genska konverzija, replikacijsko proklizavanje, mehanizam kotrljajućeg kruga te mehanizmi zasnovani na transpoziciji (Dover, 1986., Feliciello i sur., 2006., Birky i Skavaril, 1976., Levinson i Gutman, 1987., Plohl i sur., 2012., Vondrak i sur., 2021.).

Usprkos čestoj pojavi vrsno specifičnih varijanti satDNA koje naseljavaju genome srodnih vrsta, nukleotidne sekvencije nekih satDNA ostaju nepromijenjene tijekom dugih evolucijskih razdoblja i kao takve prisutne u vrstama taksonomske skupine. Primjerice, monomerne sekvencije satDNA PRAT stare 50-60 milijuna godina prisutne u nekoliko taksonomski udaljenih vrsta u porodicama Tenebrionidae i Chrysomelidae pokazuju izrazito visok stupanj međusobne sličnosti, čak 98,7 % (Mravinac i sur., 2002.). Ovakve spore promjene u nukleotidnoj sekvenciji satDNA vjerojatno su posljedica *pristranosti* u mehanizmima obrtaja u procesu molekularnog *drive-a* (Strachan i sur., 1985.). U genomima postoje mnoge podporodice i nepovezane porodice satDNA te se one veoma razlikuju u zastupljenosti u genomu, od < 0,5 % do > 50 %, što je moguća posljedica brzih promjena broja kopija u kratkim evolucijskim periodima (Ugarković i Plohl, 2002.). S obzirom na navedeno, čak i ako su nukleotidne sekvencije satDNA dobro očuvane, vrsno specifični profili satDNA mogu nastati zbog razlika u broju kopija pojedinih satDNA čak i između vrlo blisko srodnih vrsta (Fry i Salser, 1977., Meštrović i sur., 1998., Ugarković i Plohl, 2002.).

S pojavom visokoprotočnih metodologija sekvenciranja te slaganjem i analiziranjem sastavljenih genoma, ali i genomskih odjeljaka bogatih satDNA, poput heterokromatina, produbilo se znanje o transkripciji i mogućim funkcijama satDNA. Poznato je da su satDNA uključene u funkcionalnost, građu i evoluciju kromosoma, kromatinsko moduliranje, reproduktivnu izolaciju, genomsku stabilnost i evoluciju (Henikoff i sur., 2001., Slamovits i Rossi 2002., Pezer i Ugarković, 2008., Adegas i sur., 2009., Ferree i Barbash, 2009., Garrido-Ramos 2017., Lower i sur., 2018., Louzada i sur., 2020., Shatskikh i sur., 2020.). U sekvencijama alfa satDNA u centromernim područjima ljudskih kromosoma, u satDNA drugih sisavaca, ali i kod različitih beskralješnjaka pronađene su očuvane sekvencije motiva CENP-B (Kipling i Warburton, 1997., Lorite i sur., 2002., Mravinac i sur., 2002.), regije koje služe kao mjesto vezanja proteina CENP-B (Masumoto i sur., 1989.), te se smatra da imaju ključnu ulogu u slaganju centromernog kromatina (Ohzeki i sur., 2002.). Funkcionalna istraživanja elementa homolognog najzastupljenijoj satDNA brašnara *Tribolium castaneum*, TCAST1, koja je većinom smještena u pericentromernom heterokromatinu, ali je prisutna i u eukromatinu u blizini gena, pokazala su da je odgovorna za supresiju ekspresije nekoliko gena, a supresija je to jača što je veći broj kopija monomera te satDNA u elementu (Felicciello i sur., 2015.). U istoj je vrsti pokazano da okolišni

stres (temperaturni šok) utječe na ekspresiju satDNA i utječe na strukturu heterokromatina (Pezer i Ugarković, 2012.). Nadalje, velik broj istraživanja pokazao je da lomovi DNA u području centromera i posljedične translokacije te nepravilna segregacija kromosoma dovode do mnogih tipova tumora (Barra i Fachinetti, 2018. i literaturni navodi u tom radu) i drugih bolesti ljudi (Miga, 2019., Louzada i sur., 2020.).

### 2.3. Otkriće PGE

Pokretni genetički elementi otkriveni su sredinom prošlog stoljeća, a opisala ih je Barbara McClintock u svom radu u kojem je objašnjen mehanizam naslijeđivanja boje zrna kukuruza (McClintock, 1950.). Prva je dala eksperimentalne dokaze da je struktura genoma dinamična. Distribucija PGE u genomu nije nasumična, a njihova propagacija i transpozicija u genomu događa se u interakciji PGE međusobno, ali i s ostalim dijelovima stanice (Venner i sur., 2009.), kroz procese parazitizma, kooperacije i kompeticije (Robillard i sur., 2016.). Sultana i sur. (2017.) u svom preglednom radu predlažu da su specifična mjesta u genomu na koja se PGE premještaju takva da što manje štete nanose domaćinu, ali što više koristi PGE-u u smislu mobilnosti. Primjerice, jedno takvo mjesto u genomu je i promotorska regija gena kojeg transkribira RNA polimeraza III, a koju ciljaju pokretni elementi u kvascu i plijesni prilikom svoje transpozicije (Guo i sur., 2014.; Spaller i sur., 2016.; Cheung i sur., 2018.). Nadalje, evolucijski proces selekcije utječe i na fiksaciju transponiranih elemenata u određenim regijama. Element LINE L1 u genomu čovjeka gotovo nikad nije fiksiran u kodirajućim regijama gena (Gotea i Makalowski, 2006.), iako se oni insertiraju i u područja genskih egzona (Hancks i Kazazian, 2016.).

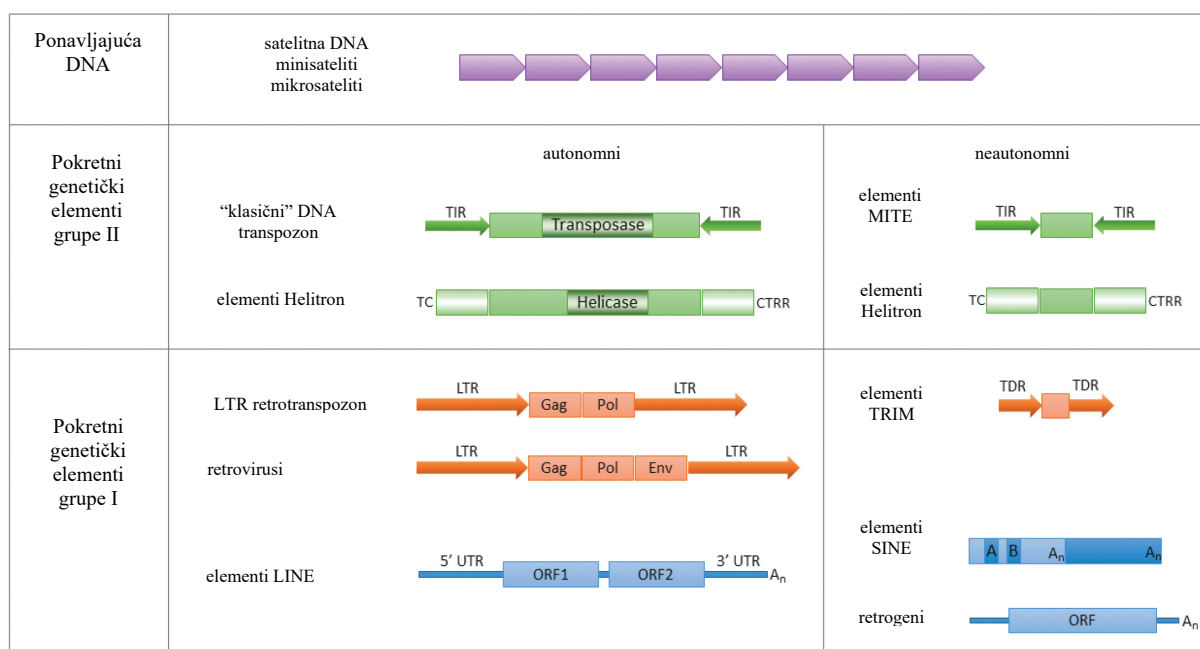
### 2.4. Tipovi pokretnih genetičkih elemenata

Pokretni genetički elementi, PGE, predstavljaju raznoliku skupinu segmenata DNA koji mogu mijenjati položaj na kromosomu i u genomu kroz proces transpozicije. Osim unutar istog genoma i naslijeđivanjem putem potomstva (tj. vertikalno), mogu se širiti i između različitih genoma tzv. horizontalnim prijenosom među različitim taksonomskim skupinama (Arhipova i Yushenova, 2019.). Prema mehanizmu kojim mijenjaju mjesto u genomu, PGE se klasificiraju kao grupa I (transpozicija ovisna o RNA) i grupa II (transpozicija ovisna o DNA) (Finnegan, 1989.). Obje grupe sadrže i autonomne i neautonomne elemente. Autonomni retrotranspozoni

kodiraju vlastitu reverznu transkriptazu (RT) kao i ostale enzime potrebne za retrotranspoziciju, a neautonomni koriste enzime autonomnih. Glavne skupine PGE-a pokazane su na Slici 2.

Prvu grupu čine retrotranspozoni koji u genomu stvaraju dodatne kopije uz nastanak RNA intermedijera. Prilikom transpozicije, njihova se sekvencija najprije prepisuje i zatim se njena reverzno transkribirana kopija ugrađuje na novo mjesto u genomu (Boeke i sur., 1985.). Zbog toga se njihov mehanizam transpozicije naziva replikativna transpozicija ili *copy-and-paste* mehanizam. Grupa I mobilnih genetičkih elemenata može se prema prisutnosti LTR regija (od engl. long-terminal repeats) podijeliti na LTR i *non-LTR* retrotranspozone. LTR retrotranspozoni dijele se u četiri glavne grupe: Copia, Gypsy, BEL i endogeni retrovirusi (ERV). Na krajevima imaju LTR regije, a u središnjem dijelu nalazi se regija s genima koji kodiraju za nekoliko enzima (proteaza, RT, ribonukleaza H i integraza) kao jedan otvoreni okvir čitanja (od engl. open-reading frame, ORF) te regija s drugim otvorenim okvirom čitanja koja kodira strukturni protein GAG. Endogeni retrovirusi posebna su skupina retrotranspozona koja uz navedene proteine ima i gen env koji kodira glikoproteine kapside koji obično omogućuju širenje među jedinkama i infekciju stanica. Ostali retrotranspozoni ili uopće nemaju gen env ili im je on nefunkcionalan.

*Non-LTR* retrotranspozoni ne sadrže LTR regije i na 3' kraju obično završavaju poli-A repom ili jednostavnim ponavljanjima. Prema filogenetskoj karakterizaciji dijele se u 3 nadporodice: LINE (od engl. long interspersed nuclear element), SINE (od engl. short interspersed element) i elementi Penelope. Uz LTR i *non-LTR* retrotranspozone, prema novijim klasifikacijama po mehanizmu integracije i/ili filogeniji reverzne transkriptaze, postoje još dvije podgrupe retrotranspozona: retrotranspozoni DIRS (od engl. Dyctiostelium intermediate repeat sequence; Gloeckner i sur., 2001., Poulter i Goodwin, 2005.) i retrotranspozoni nalik elementu Penelope (od engl. Penelope-like elements; Arkhipova i sur., 2003.).



Slika 2. Glavni tipovi ponavljajuće DNA u eukariotskim genomima (slika prilagođena iz Šatović i sur., 2020.).

#### 2.4.1. Elementi LINE

Elementi LINE dominantni su retrotranspozoni u kralješnjaka, posebice sisavaca u čijim genomima čine oko 20% (Kapitonov i Jurka, 2009.). U svom punom obliku dugi su oko 6 kb. U 5' netranslatiranoj regiji sadrže interni promotor, dva otvorena okvira čitanja (jedan gen kodira RNA-vezujući protein, a drugi kodira protein s endonukleaznom i reverznotranskriptaznom funkcijom, a u 3' netranslatiranoj regiji poli-A rep). Većina elemenata LINE transponira složenim procesom koji se naziva reverzna transkripcija uvjetovana ciljnom DNA (od engl. target-primed reverse transcription; Luan i sur., 1993., Wiener 2002.), iako se neki mogu integrirati u postojeće lomove u molekuli DNA procesom ovisnim o endonukleazi (Morrish i sur., 2002.).

Retrotranspozicija elemenata LINE nije proces velike vjerodostojnosti pa se reinsertirani elementi vrlo često značajno razlikuju od ishodišnih elemenata, a reinsertije na nova mjesta u genomu, nerijetko uvelike mijenjaju genomske okoliš. Jedan od razloga za velike razlike koje nastaju je slabi intrinzični signal kojeg posjeduje poli-A rep. Zbog nedovoljne jačine signala za kraj transkripcije elementa, ona se nastavlja i u regiju iza poli-A repa koji se nalazi iza završetka



elementa. Takav se *nadopunjeni* element zatim ugrađuje na novo mjesto u genomu gdje zbog novog, dodatnog dijela potencijalno nastaju genske duplikacije (Moran i sur., 1999.). Nadalje, translacija elementa vrlo često ne dođe do kraja (5' netranslatirane regije) prije insercije, pa se u genom na novo mjesto može ugraditi i nefunkcionalni, skraćeni element koji je *mrtav na dolasku* (od engl. dead-on-arrival, Lynch, 2007.). Reinsercije elemenata popraćene su velikim lokalnim rearanžmanima, za koje uzroci nisu do kraja razjašnjeni. Ti rearanžmani uključuju delecije velikih regija (do nekoliko kb), inverzije, te insercije kimeričkih elemenata koji potječu od dviju različitih molekula mRNA (Gilbert i sur., 2002., Symer i sur., 2002.) što dovodi do razlika u krajnjem insertiranom elementu u odnosu na ishodišni. Retrotranspozicijom funkcionalnih elemenata LINE istovremeno se mogu transponirati i oni elementi LINE koji sami nisu kompetentni za transponiranje (Wei i sur., 2001.), kao i ostali neautonomni elementi.

#### 2.4.2. Elementi SINE (od engl. short interspersed nuclear elements, SINEs)

Elementi SINE neautonomni su retrotranspozoni značajno kraći od elemenata LINE (duljine im iznose 100-600 pb, obično oko 300 pb), ali brojem kopija nadmašuju elemente LINE, s više od 1 500 000 kopija u genomu čovjeka. Više od 70 % tih kopija pripada elementu *Alu*, specifičnom za primata (Lynch, 2007.). Budući da ne kodiraju vlastite enzime za transpoziciju, reverznu transkriptazu i endonukleazu koriste od elemenata LINE. RNA-polimeraza II ili III transkribira elemente SINE, nakon čega se reverznom transkriptazom prepisuju u DNA i nakon toga se endonukleaznom aktivnošću smještaju u različite dijelove genoma (Eickbush i sur., 1992., Ohshima i sur. 1996., Kramerov i Vassetzky, 2011.). Većina elemenata SINE potječu od tRNA (Okada i Ohshima, 1995.), a element *Alu* potječe od 7SL RNA (Ullu i Tschudi, 1984.). Element *Alu* dugačak je 300 pb i ima dimernu strukturu. Može se pronaći u velikom broju gena i njihovih transkriptata, a na gensku ekspresiju utječe na različite načine te može djelovati na poliadenilaciju (Roy-Engel i sur., 2005.; Chen i sur., 2009.), izrezivanje egzona (Shen i sur., 2011.) i editiranje RNA (Chen i sur., 2008.). Elementi *Alu* uzrok su brojnih ljudskih bolesti i genetičkih poremećaja jer mijenjaju ljudski genom insercijama *de novo* ili međusobnom rekombinacijom (pregled u Kim i sur., 2016.).

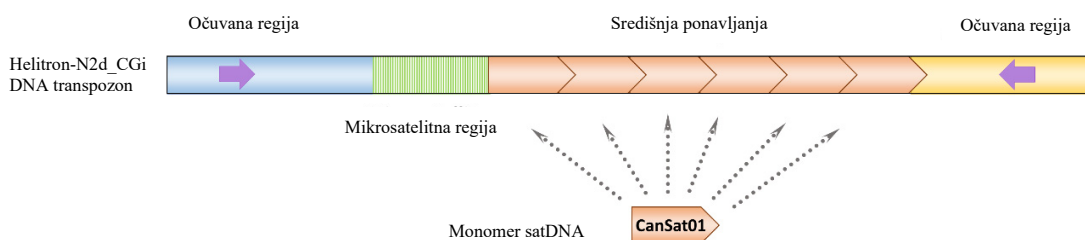
### 2.4.3. DNA transpozoni

DNA transpozoni druga su skupina PGE. Za razliku od retrotranspozona, u procesu transpozicije ne mogu stvarati DNA, a za replikaciju koriste enzime iz stanice domaćina. Također, prilikom transpozicije, svoju kopiju ne ostavljaju na dotadašnjem mjestu, već se u potpunosti premještaju na novu lokaciju. Prema mehanizmu transpozicije dijele se na četiri skupine (Bourque i sur., 2018.), a klasificiraju se prema enzimima koje elementi koriste u procesu transpozicije, strukturnim sličnostima i vezama između sekvencija (Kapitonov i Jurka, 2008.). Prva skupina transponira prema tzv. mehanizmu *izreži i zalijepi* (od engl. cut-and-paste). U njoj se nalaze TE poput hAT i Tc1/mariner. Tu skupinu transpozona karakteriziraju inverzna ponavljanja, TIR-ovi (od engl. terminal inverted repeats), između kojih se nalazi gen za transpozazu, a nakon insercije u novo mjesto u genomu obično stvaraju kratke (2-10 pb) duplikacije ciljnog mjesta, TSD-ove (od engl. target-site duplication) (Yuan i Wessler, 2011.). Postoje i TE u toj skupini koji prilikom insercije na novo mjesto u genomu ne stvaraju duplikacije ciljnog mjesta. To su elementi *Spy* otkriveni u genomu dudovog svilca (Han i sur., 2014.). U drugoj se skupini nalaze elementi Helitron koji transponiraju tzv. *mehanizmom kotrljajućeg kruga* (od engl. rolling-circle replication) za koji koriste HUH endonukleazu (Kapitonov i Jurka, 2001.). Treća skupina obuhvaća iznimno dugačke pokretne elemente Polinton/Maverick koji kodiraju sve enzime potrebne i za vlastitu replikaciju i za transponiranje zbog čega se nazivaju *samosintetizirajući elementi* (od engl. self-synthesizing elements) (Kapitonov i Jurka, 2006.; Pritham i sur., 2007.). Elementi Crypton čine četvrtu skupinu DNA transpozona koji za transpoziciju koriste tirozin rekombinazu (Goodwin i sur., 2003.).

### 2.4.4. Elementi Helitron

Elementi Helitron otkriveni su najprije u riži, uročnjaku i obliću *Caenorhabditis elegans* (Kapitonov i Jurka, 2001.), a danas je poznato da su Helitroni prisutni u brojnim eukariotskim genomima u različitim udjelima. U genomima vrsta iz roda *Drosophila* udio Helitrona varira između 1 i 5 %, 0-3 % u sisavcima, > 0,5 % u genomu žabe, oko 3 % u vrsti *Nematostella vectensis* te < 0,1 % u plijesni *Aspergillus nidulans* (Kapitonov i Jurka 2007. i literaturni navodi u tom radu). Glavna su obilježja elemenata Helitron postojanje očuvanih segmenata na 5' i 3'

krajevima koji su važni za transpoziciju te nedostatak TIR i TSD. Temeljna struktura elementa Helitron prikazana je na Slici 3.



Slika 3. Temeljna struktura elementa Helitrona. Na 5' kraju nalazi se tzv. očuvana regija (od engl. conserved box) (ljubičasta strelica) nakon kojih slijedi regija mikrosatelitnih ponavljanja, Iza nje nalazi se regija središnjih ponavljanja koju čine monomeri satDNA. Nakon te regije, nalazi se još jedna očuvana regija u suprotnoj orijentaciji. Slika prilagođena prema Tunjić Cvitanić i sur., u pripremi.

Autonomni elementi Helitroni sadrže ORF koji kodira transpozazu dugu oko 1800 aminokiselina, a sadrži DNA vezujuću domenu na N-kraju iza koje slijedi endonukleaza Rep, koja je tipa HUH (engl. histidine-hydrophobic-hystidin), s helikazom na C-kraju koja odmata DNA u smjeru 5' — 3'. U procesu transpozicije u katalitičkoj domeni endonukleaze HUH ključnu ulogu igraju dva Tyr, a tijekom transpozicije stvara jednolančani kružni intermedijer (del Pilar Garcillán-Barcia i sur., 2001.). Na 5' kraju nalazi se specifična regija duljine 150 pb koja započinje očuvanim dinukleotidom, a regija iste duljine na 3' završava očuvanim tetranukleotidom kojemu prethodi struktura ukosnice koja ima ulogu zaustavljanja transpozicije. Grabundžija i sur. (2018.) pokazali su da su za transpoziciju Helitrona *Helraiser*, okarakteriziranog u genomu malog smeđeg šišmiša (*Myotis lucifugus*), nužne lijeva terminalna regija i struktura ukosnice u desnoj terminalnoj regiji te da nepostojanje strukture ukosnice povećava učestalost skraćivanja i transdukcija sekvencija u blizini desne terminalne regije *ex vivo*.

U svom središnjem dijelu elementi Helitron sadrže uzastopna ponavljanja koja mogu činiti do 50 % duljine samog elementa (Thomas i sur., 2014.). Uzastopna su ponavljanja regije podložne visokoj stopi mutacija pa zbog toga blisko srodne kopije elemenata mogu pokazivati značajne razlike u sekvenciji (Gemayel i sur., 2010.). Neke porodice elemenata Helitron mogu sadržavati mikrosatelitne i minisatelitne regije (Thomas i Pritham, 2015. i lit. navodi u tom radu).

Uočeno je da Helitroni mogu postojati i u organizaciji jedan do drugoga tvoreći uzastopna ponavljanja u orijentaciji glava-rep (Du i sur., 2008.). Primjer takve organizacije su tandemizirani elementi Helitron 27 biljnih vrsta (Xiong i sur., 2016.) te dva Helitrona detektirana u genomu malog smeđeg šišmiša, s genomskim udjelom od 2,5 % (Pritham i Feschotte, 2007.).

Elementi Helitron pokazuju širok spektar ciljnih mjesta. Primjerice, i za biljne i za životinjske elemente Helitron pokazano je da se preferencijalno ugrađuju na AT-bogata mjesta u genomu (Yang i Bennetzen, 2009.a; Yang i Bennetzen, 2009.b). Nadalje, elementi Helitron preferencijalno premještaju u heterokromatinska područja siromašna genima, na primjer, elementi Helitron u genomu uročnjaka u pericentromerna (Yang i Bennetzen, 2009.a). Elementi Helitron u genomu kukuruza specifični su po tome što preferencijalno premještaju u heterokromatinska područja u blizini gena, no ne i u gene (Yang i Bennetzen, 2009.b). Također, elementi Helitron gravitiraju blizini drugih elemenata iste ili srodne porodice (Yang i Bennetzen, 2009.a), a mogu se naći i u centromerama (Wang i sur., 2021.).

Elementi Helitron značajni su faktori u evoluciji genoma zbog svoje sposobnosti da u svoju strukturu inkorporiraju dijelove gena koje prilikom svoje translokacije sele po genomu, i mogućnosti insercije u promotorske i intronske dijelove gena. Osim sposobnosti akumulacije određenih dijelova gena, elementi Helitron prilikom transpozicije mogu izgubiti interne dijelove svojih sekvencija (Dong i sur., 2011.).

Po sposobnosti ugradnje dijelova gena u svoju strukturu posebno se ističu elementi Helitron kukuruza (Barbaglia i sur., 2012.). Autori smatraju da je otprilike 11 000 transkripata u genomu kukuruza porijeklom od Helitrona te da ti elementi mogu inkorporirati dijelove više različitih gena te da transkripti gena inkorporiranih u Helitronu mogu biti alternativno prekrajani. Svojom insercijom u intronske regije elementi Helitron mogu utjecati na fenotipske karakteristike organizama. Primjerice, insercijom u gen *DFR-B* odgovoran za pigmentaciju antocijaninom osiguravaju stabilnu bijelu boju cvjetova vrste *Ipomoea tricolor*. Ovojnica sjemenki u vrsti

*Brassica nappa* žute je boje zbog insercije Helitrona u intron gena *BrTT8*. Nadalje, insercija Helitrona u promotorsku regiju gena *BnSP11-1* uzrokuje autoinkompatibilnost u vrsti *Brassica napus* (Hu i sur., 2019. i literaturni navodi u tom radu). Svojom insercijom u promotorske regije elementi Helitron mogu utjecati na raznolikost odgovora organizma u stresnim uvjetima toplinskog šoka, kao što je pokazano za vrstu *Caenorhabditis elegans* (Garrigues i sur., 2019.). Huang i sur. (2016.) pokazali su da sekvencije porijeklom od elemenata Helitron mogu utjecati na gensku ekspresiju djelujući kao cis-regulatorni elementi.

Ubrzo nakon otkrića elemenata Helitron, otkriveni su srodni DNA transpozoni koji uz replikazu i helikazu u istom okviru čitanja kodiraju apurinsko-apirimidinsku endonukleazu (Poulter i sur., 2003.). Te varijante Helitrona nazvane su Helentroni. Uz razlike u kodirajućem dijelu, razlike između Helitrona i Helentrona postoje i u strukturi. Helentroni u subterminalnim regijama imaju invertna ponavljanja (12-15 pb), tzv. subTIR-ove (od engl. subterminal inverted repeats). Ponavljanje na 5' kraju palindrom je sekvencija koja se nalazi u njegovoj blizini, obično udaljenoj 3-22 pb. U nekim porodicama Helentrona sami subTIR-ovi su palindromske sekvencije. Helentroni se integriraju na nova mjesta u genomu u TT dinukleotid koji se ponekad nalazi u dužem segmentu T nukleotida (Thomas i sur., 2014.). U Helentrone je svrstan i element DINE-1, najzastupljenija ponavljajuća sekvencija u genomu vrste *D. melanogaster*, čija je klasifikacija dugo bila nerazriješena zbog nedostatka sposobnosti kodiranja (Thomas i sur., 2014.).

Još jednu varijantu Helitrona, elemente Helitron2, opisali su Bao i Jurka (2013.) u genomu alge *Chlamydomonas reinhardtii*. Strukturno se ti elementi od Helitrona razlikuju po postojanju kratkih asimetričnih invertnih ponavljanja na 5' i 3' kraju elementa, te postojanju strukture ukosnice na oba kraja. Pretpostavlja se da su obje strukturne razlike važne za proces transpozicije. Uz te, Helitron2 elementi sadrže i TSD ponavljanja različitih duljina (7-19 pb) na krajevima (Bao i Jurka, 2013.).

Smatra se da PGE izbjegavaju *izumiranje* uzrokovano vertikalnim prijenosom tako da se šire horizontalnim prijenosom između različitih vrsta. Do sada je za četiri porodice Helitrona utvrđen horizontalni prijenos (Thomas i Pritham, 2015.). Jedna takva porodica je Heli-gloria koja je opisana u četiri različite vrste (šišmiš *M. lucifugus*, kukac *Rhodnius prolixus*, gušter *Anolis*

*carolinensis*, riba *Petromyzon marinus*), a njihovo je vrijeme divergencije više od 750 milijuna godina (Thomas i sur., 2010.).

## 2.5. Povezanost sekvencija satDNA i PGE

SatDNA i PGE povezani su na mnogo različitih načina (pregled u Meštrović i sur., 2015.). Na primjer, mogu se pronaći višetruke insercije PGE u nizove satDNA, ponavljanja satDNA mogu se formirati umnažanjem cjelovitog PGE ili bilo kojeg njegovog dijela ili ekspanzijom iz kratkih nizova koji se nalaze unutar PGE. U nastavku su navedeni primjeri svih nabrojanih tipova povezanosti.

Najučestalije *mjesto susreta* satDNA i PGE jest heterokromatin, a interakcije tih sekvencija osobito su česte u području (peri)centromera (Wong i Choo, 2004.; Heslop-Harrison i Schwarzacher, 2011.). U biljnim genomima dobro je proučena porodica CR LTR retroelementa koji se nalaze u centromerama riže (CRR, Cheng i sur., 2002.), kukuruza (CRM, Zhong i sur., 2002.), pšenice (CRW, Liu i sur., 2008.) i ječma (*cereba*, Houben i sur., 2007.). Ti se elementi u centromernom heterokromatinu nalaze umetnuti u ponavljanja satDNA ili u druge retrotranspozone, poput elementa CRR iz genoma riže koji u području centromera može biti umetnut u satDNA CentO ili druge elemente CRR te i sam tvoriti duga tandemska ponavljanja (Cheng i sur., 2002.). Primjer povezanosti satDNA i PGE u animalnim genomima jest element L1, visoko zastupljen u pericentromerama kromosoma čovjeka (Lander i sur., 2001.). Pokazano je da je preferencijalno mjesto za ugradnju elementa L1 regija bogata polinukleotidom TA (Feng i sur., 1996., Jurka, 1997.) što bi moglo objasniti njihovu visoku zastupljenost u pericentromernim regijama ljudskih kromosoma kojima dominiraju satDNA bogate nukleotidima A i T (Feng i sur., 1996., Jurka, 1997.). Centromere vinske mušice formiraju se u područjima složenih sekvencija DNA koje su bogate retroelementima koje okružuju dugi nizovi ponavljajućih satDNA (Chang i sur., 2019.).

Osim u (peri)centromernim regijama, satDNA i PGE u međusobne interakcije mogu doći i u drugim regijama genoma. U genomu brašnara *Tribolium castaneum* pronađeno je nekoliko različitih organizacijskih oblika satDNA TCAST. Prva varijanta nalikuje pravim satDNA, a druga varijanta uklopljena je u element kompleksne strukture koji nalikuje (retro)transpozonomima. Sastavljen je od gotovo cjelovitog monomera TCAST i 121 pb dugog monomernog segmenta u

invertnoj orijentaciji. Druga varijanta pronađena je i u 5' netranslatiranoj regiji CR1-3\_TCa retrotranspozona te je moguće da je proces retrotranspozicije uključen u disperziju ove sekvencije po genomu. Obje varijante pronađene su u blizini gena te se nagađa da mogu sudjelovati u kontroli genske ekspresije (Brajković i sur., 2012.). U genomima daždevnjaka iz roda *Hydromantes* pronađeni su kompozitni elementi koji su nastali insertiranjem uzastopno ponovljene DNA u element SINE, no njihova lokalizacija nije ograničena na specifične kromosomske regije poput centromera, telomera i C-vrpce već su se smjestili duž krakova kromosoma (Batistoni i sur., 1994.).

SatDNA pvB370 nađena u genomu vrste *Drosophila virilis* posjeduje značajnu sličnost u sekvenciji s LTR-om transpozona pDv (Heikkinen i sur., 1995.), a sličan odnos primijećen je i kod satDNA Sobo u krumpiru (Tek i sur. 2005). SatDNA HinfI iz genoma pilića sadrži retrotranspozon CR1 u sklopu svojih ponavljanja (Li i Leung, 2006.). Kod pericentromerno locirane glavne satDNA kitova reda Cetacea primijećeno je 63 % sličnosti s 3' krajem elementa L1 (Kapitonov i sur. 1998.). Slično tome, satDNA koja se nalazi u pericentromernom području vrste *Arabidopsis thaliana* pokazuje sličnost s 5' krajem DNA transpozona En/Spm (Kapitonov i Jurka, 1999.).

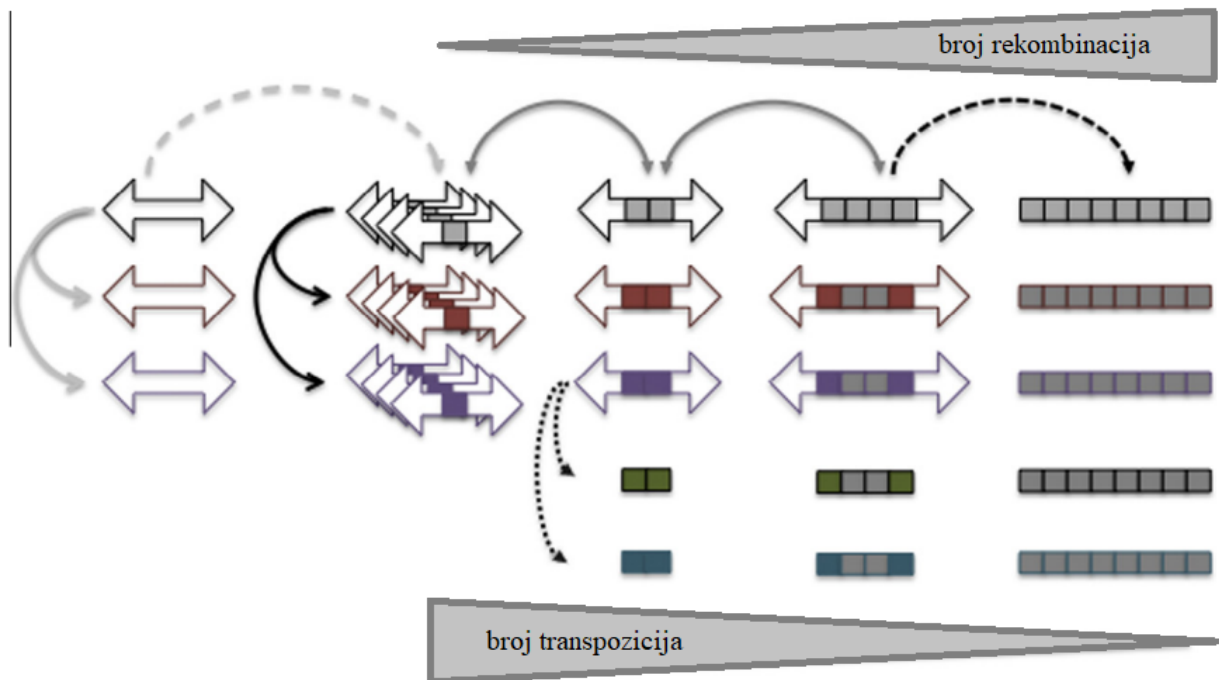
U genomima vrsta roda *Xenopus* nađene su dvije uzastopno ponovljene sekvencije, Xstir (*X. leavis* i *X. borealis*) i tropXstir (*X. tropicalis*) pri čemu su obje slične različitim dijelovima MITE elementa Xmix. Zaključak je autora da su te dvije satDNA nastale nezavisno iz MITE elementa Xmix i to nakon razdvajanja linija kojima pripadaju te vrste (Hikosaka i Kawahara, 2004.).

Hikosaka i Kawahara (2004.) predlažu mehanizam nastanka uzastopnih ponavljanja u genomu iz središnjih ponavljanja elementa MITE na temelju sličnosti uzastopnih ponavljanja i elementa MITE iz spomenutih triju vrsta žaba. Budući da nejednoliko prekrivanje i replikacijsko proklizavanje mogu tvoriti nova ponavljanja samo iz već postojećih ponavljanja, autori predlažu dva mehanizma tandemizacije početnih ponavljanja u elementu MITE, a oba se temelje na djelomičnoj deleciji sekvencija dvaju susjednih elemenata MITE (Hikosaka i Kawahara, 2004.).

Scalvenzi i Pollet (2014.) također su istraživali elemente MITE te predložili mogući mehanizam nastanka satDNA iz tih PGE. U njihovom modelu, predačka sekvencija elementa MITE, prije nego što je u sebi sadržavala recentnu sekvenciju satDNA, mogla je već sadržavati ishodišnu sekvenciju sa satDNA ili je ona mogla nastati nakupljanjem supstitucija. Sekvencija satDNA u

tom elementu (miDNA4) okružena je sekvencijama bogatima A i T nukleotidima koje tvore kratka direktna ponavljanja (4-5 pb), tzv. područja mikrohomologije. Područja mikrohomologije mogu rezultirati internim delecijama ili integracijama tijekom replikacije ili popravka DNA (Negoua i sur., 2013.). U ovom kontekstu, mikrohomologije između sekvencija bogatih A i T nukleotidima mogući su dokaz integracije sekvencije satDNA nakon popravka izrezane miDNA. Budući da nema ostataka miDNA4 bez sekvencije satDNA, pretpostavlja se da sekvencija satDNA potječe od predačke sekvencije miDNA4 unutar elementa MITE. Na taj način predačka sekvencija miDNA4 potencijalno je doprinijela nastanku nekoliko različitih sekvencija satDNA. Nastala sekvencija satDNA mogla je evoluirati u dva smjera. Jedan smjer uključuje klasičnu amplifikaciju i evoluciju satDNA, a drugi uključuje transpoziciju uz pomoć elementa MITE. Spomenuta satDNA u elementu miDNA4 pokazuje visok stupanj konzerviranosti na razini genoma zbog usklađene evolucije te veću raznolikost u nekoliko prvih i zadnjih monomera u nizu uzastopno ponovljenih monomera što dovodi do zaključka da slijedi klasični put evolucije satDNA. U subtelomernim regijama kromosoma nalaze se dugi nizovi sekvencije satDNA iz miDNA4 što se može objasniti većim brojem rekombinacijskih događaja u tim regijama. Prisutnost sekvencija satDNA u tim regijama može se objasniti početnom translokacijom elementa miDNA4 u to područje, a produljivanje niza satDNA većim brojem rekombinacija u tom području. Kako se niz produljivao, tako se povećavala i duljina regije između invertiranih terminalnih ponavljanja što je uzrokovalo smanjenje broja transpozicija samog elementa MITE. Nakupljanjem mutacija tijekom vremena nastali su nizovi satDNA bez okolnih regija porijeklom od elementa MITE. Nastanak satDNA iz elementa MITE prikazan je shematski na Slici 4. Nadalje, i za područje primarnih konstrikcija (centromera) kromosoma pokazano je da su pogodna za stvaranje dugih nizova satDNA. U genomu vrste *Lathyrus sativus* devet je satDNA pronađeno u blizini LTR retrotranspozona te je za neke od njih utvrđeno da su nastali ekspanzijom subterminalnih ponavljanja u 3' netranslatiranoj regiji elementa Oge. Svih devet satDNA dominantno je lokalizirano u području primarnih konstrikcija (Vondrak i sur., 2021.).





Slika 4. Slika prikazuje potencijalni model nastanka satDNA iz elementa MITE. Rubni dijelovi elementa MITE predstavljeni su dvostranim strelicama, a uzastopna ponavljanja satDNA pravokutnicima. Slika s desna na lijevo prikazuje predloženi model nastanka nizova satDNA unutar elementa MITE te produljivanje niza s povećanjem broja rekombinacija. Od vrha prema dnu, slika prikazuje kako element translocira, počevši od ishodišnog elementa predstavljenog crnom strelicom. Element može translocirati neovisno o duljini niza satDNA. No, s povećanjem niza satDNA i samog elementa MITE, opada broj transpozicija elementa. Uslijed nakupljenih mutacija, nizovi satDNA više nemaju postrane regije podrijetlom od elementa MITE (pravokutnici bez strelica). Slika je modificirana prema Scalvenzi i Pollet, 2014.

## 2.6. Obilježja heterokromatina i njegova povezanost s ponavljajućim DNA

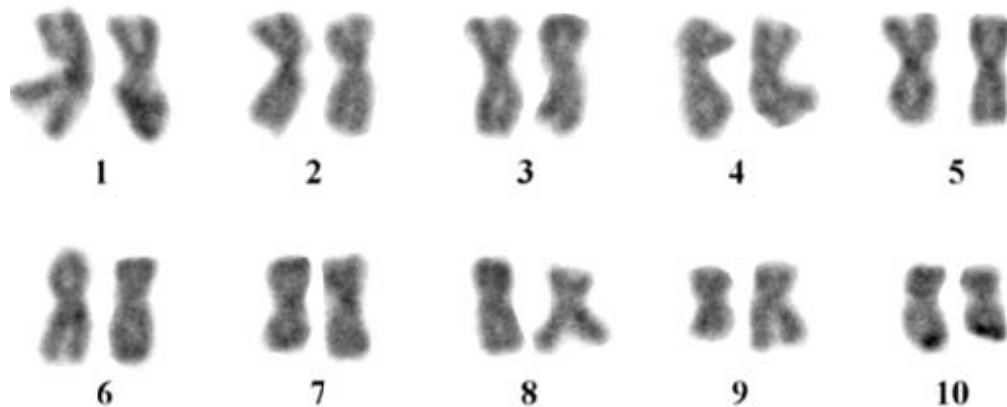
Kromatin je kompleksna struktura koju formiraju DNA, histoni i ostali proteini u eukariotskim stanicama, a podjela na heterokromatin i eukromatin uspostavljena je na temelju longitudinalne diferencijacije kromosoma nakon bojanja bazičnim bojama (Heitz, 1928.). Prvi stupanj u namatanju kromatina jest nukleosomska struktura. Njega čine dva namotaja molekule DNA

(otprilike 360 pb), tj. kromatinske niti omotane oko četiri para histona (H2A, H2B, H3 i H4). Između dva nukleosoma otprilike je 20-80 pb razmaka, regija na koju se veže histon H1 te pomaže u povezivanju dva susjedna nukleosoma, a kromatin se dalje kompaktno sabija sve do strukture kromosoma. Heterokromatin predstavljaju one regije na kromosomima koje tijekom staničnog ciklusa ostaju citološki kondenzirane, a prilikom tehnike tzv. C-pruganja one su intenzivno obojane (Slika 5.). Svjetlije obojane pruge na kromosomima su eukromatin, tip kromatina koji je bogat genima i često transkripcijski aktivan, a tijekom interfaze je dekondeziran (Maison i Almouzni, 2004.; Slika 5.). Heterokromatin se dalje može podijeliti na fakultativni i konstitutivni heterokromatin. Fakultativni heterokromatin predstavlja kondenzirani heterokromatin koji je privremeno transkripcijski utišan, ali zadržava potencijal konverzije u eukromatin te se transkribira pod određenim vremenskim i prostornim uvjetima. Primjerice, premještanje neke kromatinske niti iz periferije u centar jezgre ili obrnuto može potaknuti ekspresiju gena u fakultativnom heterokromatinu, ili ona može biti potaknuta tijekom specifične faze staničnog ciklusa ili razvoja (Trojer i Reinberg, 2007.). Konstitutivni heterokromatin čini specifične regije, lokalizirane najčešće uz centromere i telomere kromosoma, područja ključna za integritet i stabilnost genoma (Almouzni i Probst, 2011.).

Histonske modifikacije mogu kromatin *označiti* kao aktivan ili kao utišan. Histoni H3 i H4 u području eukromatina acetilirani su, a Lys4 u histonu H3 (H3K4me3) metiliran je stoga se te modifikacije smatraju biljezima eukromatina i genske aktivnosti prema histonskom kodu (Jenuwein i Allis, 2001.). H3K9me oznaka je transkripcijskog utišanog kromatina. Različiti stupnjevi metiliranosti iste aminokiseline mogu biti dodatni načini regulacije aktivnosti kromatina. Primjerice, H3K4me2 oznaka je *permisivnog* kromatina koji je potencijalno transkripcijski aktivan, dok je trimetilacija tog istog lizina oznaka transkripcijske aktivnosti. H3K9me2 i H3K9me3 oznake su fakultativnog i konstitutivnog heterokromatina (Peters i sur., 2003.).

Kromatin smješten u području centromera naziva se centrokromatin, a oko njega se nalazi pericentromerni heterokromatin. Centrokromatin je direktno uključen u stvaranje kinetohora. Regije pericentromernog heterokromatina na jednom kromosomu mogu stupati u interakcije s

istim regijama na drugim kromosomima te zajedno tijekom interfaze mogu tvoriti strukture koje se nazivaju kromocentri (Maison i Almouzni, 2004.).



Slika 5. Na slici je prikazan kariogram pacifičke kamenice *C. gigas* obojan tehnikom C-pruganja. Na 5. paru kromosoma vidljiv je tamniji heterokromatinski signal u centromernom području, a na 10. paru kromosoma vidljiv je tamni signal u telomernom području. Svjetlije regije su eukromatinska područja (slika preuzeta iz Bouilly i sur., 2008.).

Pericentromerni heterokromatin u pravilu se sastoji se od dugih nizova satDNA (često su bogate nukleotidima A i T), isprekidanih s umetnutim PGE (Probst i Almouzni, 2007.). U odnosu na ostatak heterokromatinskih regija, pericentromerni je heterokromatin hipoacetiliran na histonima H3 i H4 (Jeppesen i sur., 1992.). U biljkama ključnu ulogu u održavanju konstitutivnog heterokromatina imaju male molekule RNA koje se transkribiraju iz satDNA (May i sur., 2005.) i PGE u pericentromernom području (Almouzni i Probst, 2011.). Najveći uvid u molekularna svojstva heterokromatina danas se dobivaju kroz tehnike poput kromatinske imunoprecipitacije (ChIP, od engl. chromatin immunoprecipitation) (Penagos-Puig i Furlan-Magaril, 2020. i literaturni navodi u tom radu).

## 2.7. Centromere

Centromere su kromosomske domene nužne za pravilnu segregaciju kromosoma tijekom stanične diobe. Za pravilnu segregaciju nužne su kinetohore, kompleksne proteinske strukture

koje se stvaraju na centromerama tijekom diobe te se vežu za mikrotubule (McKinley i Cheeseman, 2016.). Protein CenH3 važan je epigenetski biljeg koji je nužan za vezanje kinetohora za centromere. CenH3 je homolog kanonskog histona H3 te je specifičan za centromerne nukleosome, a poznat je još i kao CENP-A (životinje i neke gljive), Cse4 (kvasci) te CID (*D. melanogaster*) (Talbert i Henikoff, 2020.). Taj je protein ujedno i glavna značajka funkcionalnih centromera, dok DNA u osnovi kromatina u tom području može biti vrlo različita po svojoj nukleotidnoj sekvenciji, ali i strukturnim i organizacijskim osobinama (Plohl i sur., 2014.). Kod ljudi i vinskih mušica, centrokromatin je označen modifikacijom H3K4me2, dok su histoni pericentromernog područja označeni modifikacijom H3K9me3 (Plohl i sur., 2014.). Pri karakterizaciji molekularnih značajki centromere, važno je voditi se njezinom funkcijom, a ne formom jer ona jako varira od vrste do vrste (McKinley i Cheeseman, 2016.). Primjerice, centromere kvasca regije su duge otprilike 125 pb koje se sastoje od tri funkcionalne domene bogate nukleotidima A i T i nazivaju se točkaste centromere. One su vjerojatno nastale od kratkih regionalnih centromera koje su karakteristične za mnoge eukariote, poput praživotinja i algi (Talbert i Henikoff, 2020.). Regionalne centromere puno su dulje od točkastih centromera te njihova duljina iznosi od nekoliko kb (gljive) do nekoliko Mb (biljke i životinje). U velikom broju do sada istraženih vrsta s regionalnim centromerama (npr. Lynch i sur., 2010.; Kapoor i sur., 2015.; Sanyal i sur., 2004.) utvrđeno je da su takve centromere bogate nukleotidima A i T te najčešće sadrže ponavljajuću DNA (pregled u Plohl i sur., 2014.). Metoda koja se koristi za određivanje tipa sekvencija DNA koje se nalaze u području centromera kromatinska je imunoprecipitacija (Lee i sur., 2005.).

Centromere koje u svojoj genomskoj osnovi imaju satDNA smještene su u znatno širim regijama uzastopnih ponavljanja koje mogu biti dugačke i do 1 Mb, ali samo dijelovi tih regija epigenetički su označeni proteinom CenH3, odnosno imaju funkciju centromere (Talbert i Henikoff, 2020. i literaturni navodi u tom radu). Većina biljnih i životinjskih vrsta sadrži centromere koje u svom sastavu imaju satDNA (Talbert i sur., 2018.). Dugi nizovi satDNA omogućuju povoljno periodičko smještanje nukleosoma koji sadrže protein CenH3 što stabilizira napetost u centromerama koja je posljedica gibanja kromosoma u anafazi (Prytkova i sur., 2011.). Nizovi satDNA koje grade centromere mogu se skraćivati i produljivati, najčešće kao posljedica nejednolikog prekriženja.

U amebi *Dictyostelium discoideum*, 86 % DNA komponente centromere čine PGE, od kojih gotovo polovinu čini LTR retrotranspozon *Dictyostelium Intermediate Repeat Sequence* (DIRS) (Gloeckner i Heidel, 2009.). U plijesni *Neurospora crassa* centromeru čine elementi LINE, retrotranspozoni i DNA transpozoni te nukleosomi označeni trimetilacijom histona H3 na Lys9 i nukleosomi s proteinom CenH3. Nukleosomi označeni trimetilacijom H3 na Lys9 prostiru se u područje oko centromera, a u istom je organizmu pokazano da i heterokromatin ima ulogu u stvaranju centromere (Smith i sur., 2012.). Nekim organizmima, poput alge *Phytophthora sojae*, centromeru grade PGE („Copia-like transposon“, CoLT) koji su centromerno specifični (Fang i sur., 2020.). Pokretni genetički elementi nakupljaju se u centromerama jer je u području centromera otežana rekombinacija prekrizanjem, a heterokromatinizacija centromernog područja potiče njihovo zadržavanje u tom području te omogućuje zaštitu centromera i vezivanje kohezina (Talbert i Henikoff, 2020. i literaturni navodi u tom radu).

Razlog zašto su satDNA najbrojnije centromerne sekvencije u tako velikom broju organizama vjerojatno je posljedica dvaju svojstava samih satDNA: zadržavaju homogenost u jako dugačkoj regiji, ali su podložne i jako brzom evolucijskoj promjeni (pregled u Plohl i sur., 2008.). SatDNA u centromernom području neprestano evoluiraju sa ciljem što jačeg privlačenja proteina kinetohore te posljedične propagacije na potomstvo u organizmima s asimetričnom mejozom (Talbert i Henikoff, 2020. i literaturni navodi u tom radu). Kako bi centromere zadržale svoju funkciju, centromerni proteini, poput histona CenH3, koevoluiraju sa satelitim sekvencijama na način da kompenziraju promjene u satDNA (Dawe i Henikoff, 2006.). Ta pojava koevolucije proteina CenH3 sa satDNA prisutnima u centromerama naziva se centromerni *drive* (Henikoff i sur., 2001.). Pretpostavlja se da se centromerni *drive* događa u vrstama koje imaju asimetričnu mejozu ženki pri kojoj se homologni kromosomi natječu za prisutnost u jajnoj stanici. Iako ne postoji univerzalna sekvencija koja bi određivala identitet centromere, prema hipotezi centromernog *drive*-a proteini CenH3 i ostali proteini kinetohore u interakciji su sa satDNA u području centromere i to na način ovisan o specifičnosti sekvencije. Prema teoriji, alelno produljivanje niza satDNA u jednom od homologa rezultira nastankom jače centromere koja može privući više proteina kinetohore i na taj način povećati šansu da taj homolog bude prisutan u jajnoj stanici. Takva asimetrija dovodi do mejotskih defekata i smanjene fertilnosti. Kako bi se smanjile negativne posljedice centromernog *drive*-a, događaju se promjene u sekvenciji proteina

CenH3 koje utječu na njegove preferencije u vezivanju za DNA čime se ponovno uspostavlja mejotska *ravnopravnost*. Ovaj proces uspostavlja svojevrsnu evolucijsku utrku između *sebičnih* satDNA i proteina kinetohore. Rezultat je adaptivna koevolucija centromernih ponavljajućih DNA među vrstama i proteina CenH3 ili drugih proteina kinetohore koji su u direktnom kontaktu sa centromernim sekvencijama (Henikoff i sur., 2001., Plohl i sur., 2008., Malik, 2009.)

Za evoluciju centromera koje u svom sastavu imaju satDNA zaslužno je nekoliko mehanizama. Na razini sekvencije monomera djeluju mutacije pojedinačnih nukleotida, a na razini cjelovitih monomera djeluju duplikacije i delecije (Plohl i sur., 2008.). Centromere koje u svom sastavu imaju satDNA podložne su i većim rearanžmanima, poput nejednolikog prekriženja, segmentalnih duplikacija, delecija PGE, hemicentričnih inverzija i insercija (Talbert i Henikoff, 2020. i literaturni navodi u tom radu).

Neocentromere mogu nastati na mjestima na kromosomu gdje centromere prethodno nisu bile prisutne, a jedna od uočenih pojava na tim lokacijama je moguća tandemizacija nakupljenih PGE koji s vremenom postaju nizovi satDNA. Novi nizovi satDNA na centromerama mogu nastati i replikacijom kotrljajućeg kruga nekih ekstrakromosomskih DNA (Cohen i sur., 2010.) ili iz već postojećih PGE (Talbert i Henikoff, 2020. i literaturni navodi u tom radu).

Postoje još i centromere koje su prisutne cijelom duljinom kromosoma osim na krajevima, a nazivaju se holocentromere. One su vjerojatno nastale od monocentromera, a u biljnim i životinjskim genomima pojavile su se neovisno nekoliko puta tijekom evolucije. Smatra se da su meta-policentrični kromosomi, prisutni u grašku, prijelazni oblik iz metacentričnih u holocentrične kromosome (Neumann i sur., 2012.). Holocentromere su bogate ponavljajućom DNA, primjerice, centromere graška sadrže velik broj različitih porodica satDNA i jedan retrotranspozon iz porodice CR (Neumann i sur., 2012.), a biljke iz roda *Luzula* sadrže na centromerama više od 30 različitih porodica satDNA i jedan LTR retrotranspozon (Heckmann i sur., 2013.). No, postoje i biljke s holocentromerama koje sadrže primjerice samo dvije porodice satDNA, poput biljke *Rynchospora pubera*, gdje dvije glavne satDNA tvore HOR-ove koji su ispresijecani retrotranspozonom (Marques i sur., 2015.).

## 2.8. Obilježja školjkaša i školjkaških genoma

### 2.8.1. Obilježja školjkaša

Školjkaši (latinski Bivalvia) su skupina morskih i slatkovodnih organizama iz koljena mekušaca čije je tijelo građeno od dvije međusobno spojene ljuštore koje se pokreću mišićem aduktorom, a broje preko 20 000 vrsta. Budući da su većinom sesilni organizmi, hrane se filtriranjem što pozitivno utječe na čistoću okoliša u kojem žive. Uz to, školjkaši mogu utjecati i na biogeokemijska svojstva okoliša u kojem žive te sudjelovati u prijenosu minerala i organskih tvari (npr. kalcija) (Saavedra i Bachere, 2006.). Nadalje, školjkaši se već dugo koriste kao živi sustavi za biomonitoring u ekotoksikologiji te pomažu u ranom otkrivanju štetnih učinaka ksenobiotika na organizme i vodena staništa (Vaughn i Hoellein, 2018.). No, akumulacija štetnih tvari iz okoliša u tkiva školjkaša koji se primjenjuju za ljudsku kozumaciju može imati štetne posljedice za ljude. Nadalje, školjkaši imaju važnu ulogu u bioremedijaciji, kruženju i pohranjivanju hranjivih tvari, poticanju primarne i sekundarne proizvodnje, u stvaranju i mijenjanju staništa te biogeokemijskoj transformaciji (Vaughn i Hoellein, 2018.).

Osim po ekološkom značaju, školjkaši se ističu i kao ekonomski vrlo važna skupina organizama. Neke vrste školjkaša prednjače u uzgoju mekušaca u svjetskoj akvakulturi (npr. kamenice). Pacifička kamenica predstavlja tzv. ciljnu vrstu (od engl. target species) u akvakulturi koja je ciljano uvedena u europska staništa iz Pacifika zbog nekoliko odlika, primjerice prilagodljivosti na meka staništa tamo gdje nema tvrdih stjenovitih (Ruesink i sur., 2005.), otporna je na bolesti i druge nepovoljne ekološke čimbenike, daje brz prinos zbog brzog rasta te su za nju smanjeni troškovi uzgoja (Hewitt i sur., 2006.). Smatra se da je uvođenje vrste *C. gigas* u europska staništa najveći pothvat uvođenja neke vrste u nova staništa zbog akvakulture (Wasson i sur., 2001, Ruesink i sur., 2005.). Istovremeno, u određenim regijama ova se vrsta značajno širi i predstavlja prijetnju autohtonim vrstama, izraženu kroz kompeticiju za prostor i hranu, popraćenu larvifagijom. Takvo širenje invazivnih vrsta može imati negativan učinak na ekosustav, predstavljajući prijetnju njegovoj stabilnosti (Vaughn i Hoellein, 2018., Ezgeta Balić i sur., 2019.).

Zbog svoje velike ekološke i ekonomske značajnosti, morski su školjkaši jedna od najviše istraženih skupina organizama (Ruesink i sur., 2005.). Prva su se istraživanja stoga bazirala na

proučavanju i unaprijeđivanju genetičkih svojstava povezanih s uzgojem (otpornost na bolesti i okolišne čimbenike, rast, metabolizam) (Saveedra i Bachère, 2006.).

Posljednjih godina, zanimanje za školjkaše proširilo se i u druge domene. Primjerice, istraživanja urođene imunosti školjkaša te njihovih mikrobioma pomažu u razumijevanju istih kod čovjeka (Robledo i sur., 2018.), a peptidi izolirani iz školjkaša potencijalni su novi antibiotici (Gerdol i Venier, 2015.). Mehanizmi neoplazija školjkaša predstavljaju vrijedne modele u istraživanju molekularnih mehanizama tumora kod ljudi, školjkaši predstavljaju i značajne modelne organizme u istraživanju matičnih stanica (Robledo i sur., 2018. i lit. navodi u tom radu).

Recentna istraživanja usmjerena su na istraživanje genoma školjkaša u cjelini (Gomes-dos-Santos i sur., 2020.) čemu možda najbolje svjedoči *eksplozija* sekvenciranih školjkaških genoma u zadnjih nekoliko godina. Prema podacima u bazi *PubMed*, do kolovoza 2023. sekvencirani su genomi ukupno 97 školjkaša (Tablica 1.), s tim da je u zadnje tri godine sekvencirano njih 60.



Tablica 1. Vrste školjkaša čiji su genomi sekvencirani do razine kontiga, *scaffold*-a ili kromosoma do sada (kolovoz 2023., baza NCBI).

Carstvo	Koljeno	Razred	Infrarazred	Red	Superporodica	Porodica	Rod	Vrsta
Animalia	Mollusca	Bivalvia	Heteroconchia	Adapedonta	Hiatelloidea	Hiatellidae	<i>Panopea</i>	<i>Panopea generosa</i>
				Adapedonta	Solenioidea	Pharidae	<i>Sinonovacula</i>	<i>Sinonovacula constricta</i>
				Adapedonta	Solenioidea	Solenidae	<i>Solen</i>	<i>Solen grandis</i>
				Cardiida	Cardioidea	Cardiidae	<i>Cerastoderma</i>	<i>Cerastoderma edule</i>
							<i>Clinocardium</i>	<i>Clinocardium nuttallii</i>
							<i>Fragum</i>	<i>Fragum fragum</i>
							<i>Hippopus</i>	<i>Hippopus hippopus</i>
							<i>Tridacna</i>	<i>Tridacna crocea</i>
								<i>Tridacna gigas</i>
					Tellinoidea	Donacidae	<i>Donax</i>	<i>Donax variabilis</i>
						Psammobiidae	<i>Gari</i>	<i>Gari tellinella</i>
							<i>Nuttallia</i>	<i>Nuttallia obscurata</i>
						Tellinidae	<i>Limecola</i>	<i>Limecola balthica</i>
							<i>Macoma</i>	<i>Macoma nasuta</i>
				Dreissenoidea	Dreisseninae	Dreissenidae	<i>Dreissena</i>	<i>Dreissena rostriformis</i>
				Lucinida	Cardiidae	Fraginae	<i>Fragum</i>	<i>Fragum whitleyi</i>
					Thyasioidea	Thyasiridae	<i>Conchocele</i>	<i>Conchocele bisecta</i>
				Myida	Dreissenoidea	Dreissenidae	<i>Congeria</i>	<i>Congeria kusceri</i>
							<i>Dreissena</i>	<i>Dreissena polymorpha</i>
					Myoidea	Myidae	<i>Mya</i>	<i>Mya arenaria</i>
					Pholadoidea	Teredinidae	<i>Bankia</i>	<i>Bankia setacea</i>

Carstvo	Koljeno	Razred	Infracarstvo	Red	Superporodica	Porodica	Rod	Vrsta
				Unionida	Unionoidea	Margaritiferidae	<i>Margaritifera</i>	<i>Margaritifera hembeli</i>
						Unionidae	<i>Beringiana</i>	<i>Beringiana fukuharai</i>
							<i>Cristaria</i>	<i>Cristaria clessini</i>
							<i>Hyriopsis</i>	<i>Hyriopsis cumingii</i>
							<i>Inversidens</i>	<i>Inversidens brandtii</i>
							<i>Inversiunio</i>	<i>Inversiunio reinianus</i>
							<i>Lanceolaria</i>	<i>Lanceolaria oxyrhyncha</i>
							<i>Megaloniaias</i>	<i>Megaloniaias nervosa</i>
							<i>Nodularia</i>	<i>Nodularia douglasiae</i>
							<i>Nodularia</i>	<i>Nodularia nipponensis</i>
							<i>Obovalis</i>	<i>Obovalis omiensis</i>
							<i>Potamilus</i>	<i>Potamilus alatus</i>
							<i>Potamilus</i>	<i>Potamilus streckersoni</i>
							<i>Pronodularia</i>	<i>Pronodularia japonensis</i>
							<i>Sinanodonta</i>	<i>Sinanodonta tumens</i>
							<i>Unio</i>	<i>Unio delphinus</i>
							<i>Unio</i>	<i>Unio pictorum</i>
							<i>Venustaconcha</i>	<i>Venustaconcha ellipsiformis</i>
				Venerida	Cyrenoidea	Cyrenidae	<i>Corbicula</i>	<i>Corbicula fluminea</i>
					Mactroidea	Mactridae	<i>Lutraria</i>	<i>Lutraria rhynchaena</i>
							<i>Mactra</i>	<i>Mactra quadrangularis</i>
							<i>Spisula</i>	<i>Spisula solida</i>
							<i>Spisula</i>	<i>Spisula solidissima</i>

<b>Crastvo</b>	<b>Koljeno</b>	<b>Razred</b>	<b>Infrarazred</b>	<b>Red</b>	<b>Superporodica</b>	<b>Porodica</b>	<b>Rod</b>	<b>Vrsta</b>
					Veneroidea	Veneridae	<i>Cyclina</i>	<i>Cyclina sinensis</i>
							<i>Mercenaria</i>	<i>Mercenaria mercenaria</i>
							<i>Ruditapes</i>	<i>Ruditapes philippinarum</i>
							<i>Saxidomus</i>	<i>Saxidomus gigantea</i>
							<i>Saxidomus</i>	<i>Saxidomus purpurata</i>
			Pteriomorphia	Arcida	Arcoidea	Arcidae	<i>Anadara</i>	<i>Anadara broughtonii</i>
							<i>Anadara</i>	<i>Anadara kagoshimensis</i>
							<i>Tegillarca</i>	<i>Tegillarca granosa</i>
				Mytilida	Mytilioidea	Lithophaginae	<i>Botula</i>	<i>Botula fusca</i>
						Mytilidae	<i>Gigantidas</i>	<i>Gigantidas platifrons</i>
							<i>Lithophaga</i>	<i>Lithophaga antillarum</i>
							<i>Lithophaga</i>	<i>Lithophaga nigra</i>
							<i>Modiolus</i>	<i>Modiolus philippinarum</i>
							<i>Mytilisepta</i>	<i>Mytilisepta virgata</i>
							<i>Brachidontes</i>	<i>Brachidontes variabilis</i>
							<i>Geukensia</i>	<i>Geukensia demissa</i>
							<i>Limnoperna</i>	<i>Limnoperna fortunei</i>
				Mytilida	Mytilioidea	Mytilidae	<i>Modiolus</i>	<i>Modiolus philippinarum</i>
							<i>Mytilus</i>	<i>Mytilus californianus</i>
								<i>Mytilus coruscus</i>
								<i>Mytilus edulis</i>
								<i>Mytilus galloprovincialis</i>
								<i>Mytilus unguiculatus</i>
							<i>Perna</i>	<i>Perna viridis</i>
							<i>Vignadula</i>	<i>Vignadula atrata</i>

Carstvo	Koljeno	Razred	Infrarazred	Red	Superporodica	Porodica	Rod	Vrsta
				Ostreida	Ostreoidea	Ostreidae	<i>Crassostrea</i>	<i>Crassostrea angulata</i>
								<i>Crassostrea ariakensis</i>
								<i>Crassostrea gigas</i>
								<i>Crassostrea hongkongensis</i>
								<i>Crassostrea virginica</i>
							<i>Ostrea</i>	<i>Ostrea denselamellosa</i>
								<i>Ostrea edulis</i>
								<i>Ostrea lurida</i>
								<i>Ostrea stentina</i>
							<i>Saccostrea</i>	<i>Saccostrea cucullata</i>
								<i>Saccostrea glomerata</i>
					Pinnoidea	Pinnidae	<i>Atrina</i>	<i>Atrina japonica</i>
							<i>Pinna</i>	<i>Pinna nobilis</i>
					Pterioidea	Margaritidae	<i>Pinctada</i>	<i>Pinctada fucata</i>
								<i>Pinctada imbricata</i>
						Pteriidae	<i>Margaritifera</i>	<i>Margaritifera margaritifera</i>
				Pectinida	Anomioidea	Anomiidae	<i>Pododesmus</i>	<i>Pododesmus macrochisma</i>
					Pectinoidea	Pectinidae	<i>Amusium</i>	<i>Amusium pleuronectes</i>
							<i>Argopecten</i>	<i>Argopecten irradians</i>
								<i>Argopecten purpuratus</i>
							<i>Azumapecten</i>	<i>Azumapecten farreri</i>
							<i>Chlamys</i>	<i>Chlamys rubida</i>
							<i>Crassadoma</i>	<i>Crassadoma gigantea</i>
							<i>Mimachlamys</i>	<i>Mimachlamys varia</i>
							<i>Mizuhopecten</i>	<i>Mizuhopecten yessoensis</i>

<b>Carstvo</b>	<b>Koljeno</b>	<b>Razred</b>	<b>Infrarazred</b>	<b>Red</b>	<b>Superporodica</b>	<b>Porodica</b>	<b>Rod</b>	<b>Vrsta</b>
							<i>Pecten</i>	<i>Pecten maximus</i>
							<i>Ylistrum</i>	<i>Ylistrum balloti</i>
						Propaemussidae	<i>Archivessica</i>	<i>Archivesica marissinica</i>

### 2.8.2. Kamenice (Ostreidae)

Kamenice imaju široku geografsku rasprostranjenost i naseljavaju obale gotovo svih kontinenata (osim Antarktike i nekih otoka) (Gunter, 1950.). Obično se nalaze pričvršćene za kamenu podlogu u području izmjene plime i oseke te u plitkim i zaštićenim estuarijima (Gosling, 2003.). Porodica kamenica broji 16 rodova, od kojih je rod *Crassostrea* vrsno najbogatiji, a vrste tog roda četvrte su po udjelu u svjetskoj proizvodnji morskih organizama (FAO, 2020.). Kamenice pokazuju visok stupanj fenotipske plastičnosti koja je pod utjecajem okolišnih čimbenika, poput tipa supstrata na kojem žive i režima izmjene plime i oseke koji utječe na njihovu okolinu te je zbog toga njihova klasifikacija na temelju morfoloških značajki nepouzdana (Seilacher i sur., 1985.). Za klasifikaciju kamenica koriste se genetički biljezi: fragmenti ribosomalnih gena (podjedinice rRNA 16S i 28S), konvencionalni *gen za barkodiranje*-podjedinica I mitohondrijske citokrom oksidaze (COI) te regija ITS2 (od engl. internal transcribed spacer) (Salvi i sur., 2014.).

U ovom istraživanju korištene su četiri kamenice iz porodice Ostreidae: pacifička kamenica, *Crassostrea gigas* (Thunberg, 1753.), portugalska kamenica, *Crassostrea angulata* (Lamarck, 1819.), američka kamenica, *Crassostrea virginica* (Gmelin, 1791.) i europska kamenica, *Ostrea edulis* (Linnaeus, 1758.).

Pacifička kamenica (Slika 6.A) kao nativna vrsta naseljava obale Japana i Južne Koreje te sjeverne obale Kine, a široko je rasprostranjena u području zapadnog Pacifika od otoka Sakhalin u Rusiji do rijeke Yangtze u Kini. Na hrvatskoj obali Jadranskoga mora pacifička kamenica se kao invazivna vrsta pojavila sedamdesetih godina prošlog stoljeća, ali vjerojatno i ranije. Pacifička kamenica ima negativan ekološki utjecaj na nativnu vrstu školjkaša (europsku kamenicu) te na druge native vrste morskih organizama (Martinez-Garcia i sur., 2021.).

Američka kamenica (Slika 6.B) raste u slanim plitkim uvalama, lagunama i estuarijima uz obale Atlantika i Meksičkog zaljeva. Glavni svjetski proizvođači te kamenice su SAD, Kanada i Meksiko (FAO, 2020.).

Portugalska kamenica (Slika 6.C) nativna je vrsta sjeverozapadnog Pacifika koja u 16. stoljeću unesena u Europu. Danas se njeno stanište u Europi prostire jugozapadnim obalama Iberijskog polutoka. Živi pričvršćena na kamenju i drugim tvrdim površinama u riječnim ušćima u oceane.

Europska kamenica (Slika 6.D) nativna je europska vrsta rasprostranjena uz zapadne obale europskog



Slika 6. Vrste kamenice korištene u ovom istraživanju. A- *Crassostrea gigas*, B- *Crassostrea virginica*, C- *Crassostrea angulata*, D- *Ostrea edulis*. (Slike A, C i D preuzete s <https://en.wikipedia.org/>, a slika B preuzeta s <https://www.jaxshells.org/crassost.htm>.)

kontinenta od Norveške do Gibraltara, obale Mediteranskog mora, obale Maroka, sjeverne obale SAD-a te obale Južnoafričke Republike. Europska kamenica raste u visokoproduktivnim estuarijima i plitkim vodama uz obalu na čvrstim kamenim podlogama, ali i na čvrstom blatu i mulju te ostacima ljuštura drugih školjkaša. Hrvatska se ističe kao jedan od glavnih europskih proizvođača europske kamenice (FAO, 2018.).

### 2.8.3. Heterokromatin školjkaša

Glavno obilježje školjkaškog heterokromatina jest visoka varijabilnost u njegovom sastavu, udjelu i raspodjeli u genomu čak i kod blisko srodnih vrsta. Saznanja o obilježjima heterokromatina školjkaša dobivena su citogenetičkim metodama poput C-pruganja (Leitao i Chaves, 2008.), fluorescencijske hibridizacije *in situ* (FISH) i bojanja fluorokromima (Petrović i sur., 2008.; Garcia-Souto i sur., 2017.c). Portugalska kamenica *C. angulata* u genomu ima visok udio heterokromatina smještenog pericentromerno, telomerno i interkalarno na većini kromosoma (Cross i sur. 2005.). Njoj bliskosrodna vrsta s kojom može stvarati hibride, pacifička kamenica *C. gigas*, u genomu ima iznimno malo heterokromatina smještenog centromerno na jednom paru kromosoma i telomerno na drugom paru kromosoma (Bouilly i sur., 2008.). Martinez-Lage i sur. (1995.) proučavali su tri blisko srodne vrste roda *Mytilus*. Vrste *M. edulis* i *M. trossolus* pokazivale su ujednačene obrasce rasporeda heterokromatina dominantno na telomerama kromosoma dok je *M. galloprovincialis* pokazivala i interkalarni smještaj heterokromatina. U vrstama roda *Sphaerium*, heterokromatin je smješten samo u

pericentromernom području (Petkevičiute i sur., 2018.), dok je kod vrste *Donax trunculus*, heterokromatin potpuno odsutan iz pericentromernog područja (Petrović i sur., 2008.). Heterokromatin vrste *D. trunculus* pokazuje još jedno neuobičajeno svojstvo, a to je da je iznimno GC-bogat (Garcia-Souto i sur., 2016.; Petrović i sur., 2008.).

## 2.9. Ponavljajuće sekvencije DNA u genomima školjkaša

Većina dosadašnjih istraživanja ponavljajućih sekvencija u genomima školjkaša temelji se na konvencionalnim metodama istraživanja ponavljajućih sekvencija, primjerice razgradnji ukupne genomske DNA restrikcijским enzimima te elektroforetsko razdvajanje i kloniranje monomera odnosno multimera ili generiranje parcijalnih genomskih biblioteka nakon čega slijedi *colony-lift* te hibridizacija s fragmentiranom cjelokupnom genomskom DNA (pregled u Šatović i sur., 2018.). Do sada su analizirane satDNA 48 vrsta školjkaša iz osam porodica (pregled u Šatović i sur., 2018.).

### 2.9.1. SatDNA u porodici Ostreidae

Prva otkrivena satDNA u porodici Ostreidae bila je satDNA Cg170 (Clabby i sur., 1996.) lokalizirana u području centromera na četiri do sedam kromosomskih parova vrste *C. gigas* (Wang i sur., 2001.). Duljina monomera satDNA Cg170 iznosi otprilike 166 pb, a udio u genomu iznosi 1-4 % (Clabby i sur., 1996.). Monomer je bogat nukleotidima A i T (60 %), a homologija monomernih sekvencija iznosi 79-94 % (Clabby i sur., 1996.). Ista satDNA detektirana je i nakon razgradnje genomske DNA enzimom HindIII u sedam drugih vrsta školjkaša, uključujući vrste *C. virginica*, *C. angulata* i *O. edulis* (Lopez-Flores i sur., 2004.). Dvije sekvencije u bazi podataka Repbase, SAT-1\_CGi (Jurka, 2012.) i SATREP\_CGi (Clabby i sur., 1996.) također odgovaraju satDNA Cg170/HindIII. Sekvencije te satDNA detektirane su i kao središnja ponavljanja elemenata Helitron u kamenici *C. gigas* (Šatović i sur., 2016.). U vrsti *O. edulis* nakon razgradnje enzimom BclI detektirana je satDNA duljine monomera 150 pb i slična postranim regijama mikrosatelita CT u genomu vrsta *C. gigas* i *O. edulis*. Ta je satDNA detektirana i u pet drugih vrsta školjkaša među kojima su i *C. gigas* i *C. angulata* (Lopez-Flores i sur., 2010.).

Vojvoda Zeljko i sur. (2020.) u genomu vrste *C. gigas* detektirali su ponavljanja nalik pravim satDNA raspršena po genomu, a dio su neautonomnih elemenata Helitron koji pripadaju DNA transpozonomima. Autori su opisali 13 najzastupljenijih ponavljanja koja su prema sličnosti grupirana u 11 elemenata koje su nazvali elementi *Cg\_HINE*. Svaki element *Cg\_HINE* sastavljen je od središnjeg dijela koji sadrži



ponavljanja nalik satDNA, te od dviju postranih regija. Neka od središnjih ponavljanja pokazuju sličnost sa satDNA Cg170 što ukazuje na povezanost PGE i satDNA.

Lopez-Flores i sur. (2004.) okarakterizirali su satDNA dobivenu cijepanjem genomske DNA enzimom HindIII koja je detektirana u genomima nekoliko vrsta kamenica. Utvrđeno je da je ta satDNA zapravo centromerni satelit iz genoma vrste *C. gigas*, a pokazala je i sličnost sa satDNA pDTE24-6 detektiranom u genomu vrste *D. trunculus*. SatDNA HindIII pokazala je sličnost i s elementom MITE iz genoma kukuruza i s elementom CvA iz genoma kamenice *C. virginica*. Zbog svih navedenih sličnosti, autori pretpostavljaju postojanje PGE koji bi bio zajednički predak elemenata s kojima je detektirana satDNA pokazala sličnost, iz kojeg su zatim i nastale detektirane satDNA u nekoliko vrsta kamenica.

Postoji i niz satDNA prisutnih kod velikog broja školjkaških vrsta i izvanredno duge starosti. Primjer satDNA prisutne u nekoliko vrsta mekušaca iz čak četiri različita razreda jest satDNA PjHhaI prvotno otkrivena u genomu jakobove kapice (*Pecten jacobaeus*, Linnaeus, 1758.) (Petraccioli i sur., 2015.). Još dvije školjkaške satDNA detektirane su u nekoliko vrsta te istovremeno pokazuju iznimnu evolucijsku starost. Prvi primjer je satDNA DTHS3 detektirana u 12 vrsta iz podrazreda Heterodonta i Pteriomorphia (Šatović i Plohl, 2018.). Minimalna starost te satDNA procjenjuje se na 516 milijuna godina (Šatović i Plohl, 2018.) prema evolucijskom razdvajanju tih dviju skupina (Bieler i sur., 2014.). Druga satDNA, BIV160, otkrivena je u devet vrsta iz triju glavnih podrazreda školjkaša: Protobranchia, Pteriomorphia i Heteroconchia. Prema procjenama temeljenima na takvoj raspodjeli, stara je barem 540 milijuna godina (Plohl i sur., 2010.).

### 2.9.2. Pokretni genetički elementi u genomima školjkaša

Prvi identificirani neautonomni DNA transpozoni u školjkašima bio je element CvA. Gaffney i sur. (2003.) detektirali su element CvA u genomu vrste *C. virginica* prisutan u 70 000 kopija. Isti je element otkriven i u vrsti *Anadara trapezia* (Deshayes, 1839.) u čijem se genomu nalazi u intronu 2 globinskog gena.

Element CvA sadrži nekoliko značajnih strukturnih obilježja: središnja uzastopna ponavljanja, invertna ponavljanja koja mogu tvoriti strukture ukosnice, terminalna direktna ponavljanja, konzerviranu regiju koja pripada promotorskom motivu *Box A* gena za RNA polimerazu III koji se nalazi u monomeru središnjeg dijela elementa te mikrosatelitna ponavljanja. Također, dijelovi elementa CvA slični su ponavljanjima koja se nalaze u genomu vrste *C. gigas* u pretpostavljenom genu za transglutaminazu (Gaffney i sur., 2003.). Element Ana-1 vrste *A. trapezia* sadrži strukturna obilježja poput elementa CvA.

Mikrosatelitne regije oba elementa slična su mikrosatelitnim regijama elementa ISTU4 koji pripada porodici transpozona Tsp (elementi MITE) detektiranom u vrsti *Strongylocentrotus purpuratus*. Također, elementi su pokazali sličnost sa satDNA nekoliko vrsti iz grupe vrsta *Drosophila obscura* (Gaffney i sur., 2003.)

U genomu vrste *C. virginica* isti su autori detektirali još tri elementa slična elementu CvA, prisutne u puno manjem broju kopija. Elementu CvB okarakterizirana je samo 5' regija i regija središnjih ponavljanja. Regija na 5' kraju slična je onoj kraju elementa Ana-1. Elementi CvE i CvG okarakterizirani su u cijeloj duljini te dijele strukturne značajke s elementom CvA, ali ne i nukleotidnu sekvenciju. Elementima CvA i CvB monomer središnjeg ponavljajućeg dijela dug je 150-180 pb te pokazuje sličnost sa satDNA Cg170 u genomu vrste *C. gigas* i DTE1 vrste *D. trunculus*. Autori su sve detektirane elemente koji strukturom nalikuju elementima MITE, u vrsti *C. virginica*, svrstali u posebnu skupinu koju su nazvali *pearl*.

U genomu vrste *Mytilus galloprovincialis* nalazi se element MgE prvotno detektiran u intronima gena *hsp70* i *hsp71*. Element MgE čini satelit Mg1 iz iste vrste sa postranim regijama koje pokazuju sličnost s elementom CvE detektiranim u vrsti *C. virginica* (Kourtidis i sur., 2006.).

U školjkašu *D. trunculus* otkriven je element DTC84, izvorno svrstan kao MITE (Šatović i Plohl, 2013.), koji u središnjem dijelu sadrži do otprilike šest ponavljajućih monomera sa točno određenim redosljedom. Monomeri središnjeg dijela se prema položaju u nizu mogu međusobno razlikovati po dijagnostičkim promjenama u sekvenciji. Duljina monomera iznosi 156 pb. Ovaj element u genomu vrste *D. trunculus* služi kao svojevrsna *kazeta* za unutarnja ponavljanja koja se pojavljuju u određenim kombinacijama. Svaki monomer sadrži palindromsku sekvenciju na spojnom mjestu sa sljedećim monomerom u ponavljanju koja je potencijalno ključno mjesto za rekombinacije koje uzrokuju razlike u broju monomera središnjeg dijela elementa. Uz jedan kraj monomernog niza nalaze se mikrosatelitna ponavljanja, a element počinje i završava sačuvanim regijama u kojima se nalaze kratka invertna ponavljanja. Zbog razlika u broju monomera u središnjem dijelu elementa te razlikama u postranim regijama, svaki element DTC84 u genomu vrste *D. trunculus* jedinstven je (Šatović i Plohl, 2013.). Kasnije su ovaj element, kao i oni iste strukturne građe, elementi *pearl* i element MgE, svrstani kao elementi nadporodice Helentron (Thomas i Pritham, 2015.).

Isti su autori u vrsti *D. trunculus* otkrili još dva elementa MITE, DTC M1 i DTC M2, duge 261 i 286 pb. Osim međusobne strukturne sličnosti, elementi ne dijele sličnost u sekvencijama niti međusobno niti s drugim poznatim ponavljajućim sekvencijama osim kratkog segmenta homologije u području kratkih

inverznih ponavljanja. U toj je regiji element DTC M1 sličan jednom duljem elementu u genomu vrste *C. gigas*. Elementi DTC M1 i DTC M2 u genomu vrste *D. trunculus* prisutni su kao cjeloviti i kao fragmentirani elementi (Šatović i Plohl, 2017.).

Šatović i sur. (2018.) po prvi su puta u školjkašima detektirali elemente TRIM (od engl. terminal repeat retrotransposons in miniature). To su kratki, neautonomni LTR retrotranspozoni koji se sastoje od središnje domene koju okružuju dva kratka direktna ponavljanja, a do tad su bili opisani većinom u biljaka. Elementi biv-TRIM detektirani su u 11 vrsta školjkaša, među kojima je i kamenica *C. gigas*. Udio u genomu u svim vrstama manji je od 1 %. Osim kanonski strukturiranih elemenata, u genomima istraženih vrsta nalaze se i pojedinačna direktna ponavljanja iz elementa raspršena po genomu, kao i dio elementa uzastopno ponavljen te nizovi raspršeni po genomu.

Puzakov i sur. (2018.) u genomima sedam vrsta školjkaša, među kojima su i *C. gigas* i *C. angulata*, detektirali su DNA transpozon koji pripada Tc-1/mariner skupini DNA transpozona. Taj je element nisko zastupljen u genomima analiziranih školjkaša, karakterizira ga nedostatak terminalnih invertiranih ponavljanja, katalitička domena s *potpisom* DD36E te postojanje introna s mogućnošću izrezivanja u regiji koja kodira za transpozazu. Autori su ga nazvali element TLEWI te ga smatraju *udomaćenim* u genomima istraženih školjkaša.

U velikom broju školjkaških vrsta detektirani su elementi SINE koji se mogu rasporediti u pet nadporodica elemenata SINE (V-SINE, Deu-SINE, CORE-SINE, Ceph-SINE i Meta-SINE). U kamenicama *C. gigas* i *O. edulis* detektiran je element BivaV-SINE2 koji pripada nadporodici V-SINE, a u kamenici *C. virginica* detektiran je element BivaMD-SINE1 koji pripada superporodici Meta-SINE (Nishihara i sur., 2016.). Elementi u superporodici Meta-SINE otkriveni su u nekoliko vrsta životinja (od žarnjaka do riba), a svima im je zajednički središnji dio od 66 pb za koji se smatra da je zajedničkog porijekla u svim vrstama. Promotori otkrivenih elemenata Meta-SINE potječu od molekula tRNA (Nishihara i sur., 2016.).

Luchetti i sur. (2016.) otkrili su još jedan element V-SINE prisutan u 16 školjkaških vrsta, ali i u nekoliko vrsta mekušaca te vrstama iz koljena Polyplacophora i Solenogastres. Autori su element RUDI SINE detektirali kao dva klona sekvencije elementa u vrsti *Ruditapes decussatus* (D12 i D25) i kao jedan klon u vrsti *Ruditapes philippinarum* (P14F). Ponavljanje D25 sadrži poli- A rep na jednom kraju iza kojeg slijede savršena direktna ponavljanja. Na drugom kraju ponavljanja nalazi se potencijalna sekundarna struktura koja je tipična za seleneocisteinsku tRNA s dobro očuvanim promotorima za RNA polimerazu III.

## 2.10. Analize genoma i detekcija ponavljajuće DNA bioinformatičkim metodama

U novije vrijeme analiziranje ponavljajućih DNA temelji se na analizama sekvencija dobivenih sekvenciranjem nove generacije, NGS (od engl. next-generation sequencing) na platformama poput one *Illumina*. Prije NGS-a, informacije o ponavljajućoj DNA u nekom genomu temeljile su se na analizama setova monomera nakon digestije restriktivnim enzimima i konstrukcije biblioteka sekvencija u plazmidima i njihovom sekvenciranju (primjerice Ugarković i sur., 1996., Gaffney i sur., 2003.). No, sve te eksperimentalne analize bile su veoma niske osjetljivosti, što je rezultiralo u uspješnoj identifikaciji samo najzastupljenijih satDNA kojima su sekvencije imale određena specifična svojstva. Malo se moglo saznati o satDNA koje su u genomu bile niskozastupljene ili nisu imale mjesta prepoznavanja restriktivnog enzima. S razvojem tehnologija sekvenciranja nove generacije i porastom broja istraživanja koja ih koriste, generirana je ogromna količina sekvencijskih podataka i metoda za slaganje genoma. Međutim, genomske regije koje su obogaćene ponavljajućom DNA uzrokuju značajne tehničke poteškoće u sekvenciranju i slaganju genoma. Problem je u nemogućnosti rekonstrukcije točnog redoslijeda sekvencija i procjeni duljina regija koje tvore ponavljajuće DNA velike sličnosti i ponovljene velik broj puta (Šatović Vukšić i sur., 2020.). Primjerice, udio jedne satDNA u genomu brašnara *Tribolium castaneum* procijenjen je na 17 % (Ugarković i sur., 1996.) u neposloženim genomskim *scaffold*-ima, dok u posloženom genomu ona čini tek 0,3 % genoma (Wang i sur., 2008.). Čak ni kombinacija nekoliko različitih metoda, objedinjavanja fozmida, NGS-a i hijerarhijskog slaganja, nije učinkovita u dobivanju precizno posloženih nizova satDNA niti u genomima s malim udjelom ponavljajuće DNA i heterokromatina, poput vrste *C. gigas* (Zhang i sur., 2012.). Precizan redoslijed uzastopno ponovljenih sekvencija u dugim nizovima ponavljajuće DNA moguće je dobiti sekvenciranjem *de novo* korištenjem tzv. tehnologije *SMRT* (od engl. single molecule real-time sequencing).

Korištenje visokoprotočnih tehnika koje uključuju kombinaciju sekvenciranja DNA niske pokrivenosti kratkim očitajima i specijaliziranih bioinformatičkih alata, pokazalo se vrlo učinkovito u identificiranju ponavljajućih DNA u genomima, odnosno dobivanju potpunih setova ponavljajućih DNA i satDNA, tzv. repeatoma i satelitoma (Kim i sur., 2014., Ruiz-Ruano i sur., 2016.). Takve metode posebno su korisne u istraživanju nemodelnih organizama za koje genomi još nisu sastavljeni. Kod organizama koji imaju sastavljene genome igraju ključnu ulogu u popunjavanju praznina koje nastaju prilikom slaganja genoma zbog posljedica prisutnosti ponavljajućih DNA. Stečeno znanje o sastavu i genomskoj

lokalizaciji ponavljajućih DNA važan su korak prema boljem razumijevanju cjelokupne genomske arhitekture, izgledu regija s ponavljajućom DNA i funkcioniranju genoma u cjelini (Belyayev i sur., 2019.).

Iako dobivanje jako velikog broja kraćih sekvencija tehnologijom sekvenciranja nove generacije nije dovoljno za precizno slaganje čitavih genoma, ipak je dovoljno za dobivanje reprezentativnog uzorka očitavanja za analizu dijelova genoma koji sadrže ponavljajuću DNA (Novák i sur., 2010.). Danas postoji velik broj alata kojima je moguće detektirati ukupnu ponavljajuću DNA u nekom genomu ili pak specifično satDNA ili zasebne skupine PGE (pregled u Šatović i sur., 2020.). Većina analiza postojećim alatima temelji se ili na traženju sličnosti sa sekvencijama u postojećim bazama repetitivnih sekvencija, ili je za njihov rad potrebno postojanje referentnog genoma vrste čija se ponavljajuća DNA istražuje.

Jedan od bioinformatičkih alata kojim je moguće detektirati širok spektar ponavljajućih DNA jest programski paket *RepeatExplorer v.2*, RE (Novák i sur., 2010., Novák i sur., 2020.), a bazira se na velikom broju kratkih očitavanja dobivenih sekvenciranjem nove generacije te za detekciju ponavljajućih DNA ne treba postojeće posložene genome. *RepeatExplorer v. 2* predstavlja paket analiza kojima je moguće određivati sastav ponavljajućih DNA u genomu jedne vrste (analiza RE), uspoređivati ponavljajuće DNA istovremeno u nekoliko različitih vrsta (komparativna analiza RE), identificirati centromernu ponavljajuću DNA (analiza *Chip-Seq Mapper*) te određivati sastav satDNA u genomu jedne vrste (analiza TAREAN). Iako je detekcija satDNA uključena i u analizu *RepeatExplorer*, autori programa predlažu i pokretanje analize TAREAN za isti genom jer se takvom analizom dobiva detaljniji prikaz satDNA prisutnih u genomu istraživane vrste s dodatnim mogućnostima obrade detektiranih satDNA, primjerice prijedlog konstrukcije sonda za analizu FISH (Novák i sur., 2020.). Programski paket mrežno je dostupan (<https://repeatexplorer-elixir.cerit-sc.cz/galaxy>), a osim spomenutih analiza sadrži i alate kojima se neobrađena očitavanja dobivena sekvenciranjem nove generacije mogu obraditi i urediti kako bi bila pogodan ulazni materijal za navedene analize (Novák i sur., 2020.). Algoritam programa u paketu RE temelji se na *usporedbi svih očitavanja sa svim očitanjima* (od engl. all-to-all comparison). Broj sličnosti u takvim usporedbama proporcionalan je genomskom broju kopija sekvencija koje se uspoređuju: najviši za one sekvencije prisutne u velikom broju kopija u genomu (ponavljajuća DNA), a najniži za one sekvencije koje se pojavljuju kao pojedinačne sekvencije (geni). Zbog toga je važno da se za ulazni dokument analize odabere skup očitavanja iz sekvenciranja koji će predstavljati nisku pokrivenost genoma koji se istražuje. Na taj način se osigurava primarna detekcija sličnosti očitavanja porijeklom iz ponavljajućih sekvencija (Novák i sur., 2020.).

komponenta analize temelji se na *algoritmu klasteriranja temeljenog na grafovima* (od engl. graph-based clustering algorithm). U tom algoritmu, klasteri se identificiraju analiziranjem izgleda virtualnih grafova u kojima je svako očitavanje/sekvencija predstavljeno linijom, krugovi koji spajaju dvije linije predstavljaju preklapanje dvaju očitavanja, a veličina kruga predstavlja međusobnu sličnost tih preklapajućih očitavanja (Novák i sur., 2020.). Ovakvo klasteriranje rezultira tisućama klastera koji mogu biti sačinjeni od samo dva očitavanja pa sve do 2000 očitavanja, ovisno o njihovoj zastupljenosti u genomu. Samo određeni klasteri koriste se za daljnje procesiranje u analizi *RepeatExplorer*, a granica za veličinu klastera koji će se dalje analizirati može se proizvoljno postaviti prije pokretanja analize. Procesni tijek analize dalje koristi ugrađene, lokalno izrađene, baze nukleotida za određivanje sličnosti detektiranih klastera i tipa ponavljajuće sekvencije kojega klaster predstavlja. Za određivanje tipa pokretnog genetičkog elementa kojega klaster predstavlja koristi se ugrađena baza konzerviranih proteinskih domena REXdb (Neumann i sur., 2019., Novák i sur., 2020.).



### **3. MATERIJALI I METODE**





### 3.1. Biološki materijal

U istraživanju su korištene četiri vrste kamenica: pacifička kamenica *Crassostrea gigas* prikupljena u Jadranskom moru u Istri i Aveiru u Portugalu, američka kamenica *Crassostrea virginica* (Gmelin, 1791.) prikupljena u Atlantskom oceanu u Floridi u SAD-u, portugalska kamenica *Crassostrea angulata* (Lamarck, 1819.) prikupljena na obalama Atlantskog oceana u Portugalu te europska kamenica *Ostrea edulis* (Linnaeus, 1758.) prikupljena u Jadranskom moru.

### 3.2. Izolacija genomske DNA školjkaša za potvrdu vrste

Zbog visokog stupnja fenotipske plastičnosti kojeg kamenice pokazuju, za potvrdu vrste potrebno je provesti DNA barkodiranje. Genomska DNA školjkaša izolirana je prema uputama u komercijalnom setu za izolaciju DNA *Dneasy Blood and Tissue Kit* (Qiagen) iz komadića mišića aduktora (do 25 mg tkiva). Koncentracija izolirane DNA izmjerena je na uređaju *Qubit Fluorometer* (ThermoFisher Scientific, Waltham, MA, SAD). U lančanoj reakciji polimerazom (PCR, od engl. *polymerase chain reaction*) korištenjem specifičnih početnica za podjedinicu I gena citokrom C oksidaze te za veliku podjedinicu ribosoma (16S) (Tablica 2.) molekularno su identificirane sve kamenice korištene u ovom radu.

Tablica 2. sekvencije početnica za gen za podjedinicu I citokrom C oksidaze i veliku podjedinicu (16S) ribosoma kojima su molekularno identificirane sve kamenice korištene u ovom radu.

LCO-1490	5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3'
HCO-2198	5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3'
16S-forward	5'-CGC CTG TTT ATC AAA AAC AT-3'
16S-reverse	5'-CCG GTC TGA ACT CAG ATC ACG T-3'

Sadržaji reakcijskih smjesa PCR-a prikazani su Tablici 3.

Tablica 3. Sadržaj reakcijske smjese za reakciju PCR-a kojom su molekularno identificirane kamenice korištene u ovom radu.

Reagens	Volumen ( $\mu\text{L}$ )
5 x Go Taq Flexi pufer	2
MgCl <sub>2</sub> (25 mM)	1
dNTP mix (10 mM svaki)	0,2
početnica <i>forward</i> (10 pmol/ul)	0,4
početnica <i>reverse</i> (10 pmol/ul)	0,4
DNA polimeraza Taq (5 UI)	0,1
H <sub>2</sub> O	5,4
kalup DNA	0,5
UKUPNI VOLUMEN	10

Volumen dodanog kalupa DNA varirao je ovisno o koncentraciji pojedine DNA, tako da je u reakciju dodano otprilike 100 ng DNA.

Uvjeti pri kojima se PCR odvijao prikazani su u Tablici 4. za dva različita para početnica.

Tablica 4. Prikaz uvjeta PCR-a korištenih za barkodiranje vrsta korištenih u ovom radu.

	COI	16S	
denaturacija	94 °C, 30 s	94 °C, 30 s	30 x
hibridizacija	52 °C, 30 s	55 °C, 30 s	
produljivanje	72 °C, 30 s	72 °C, 30 s	

### 3.3. Izolacija nativnog kromatina i razgradnja mikrokokalnom nukleazom

Nativni je kromatin izoliran u nekoliko ponavljanja iz prethodno naglo smrznutog mišića aduktora vrste *C. gigas*. Kromatin je izoliran homogeniziranjem mišićnog tkiva korištenjem tkivnog homogenizatora (tip *Dounce*) u puferu CIB (15 mM Tris HCl pH 7,5, 60 mM KCl, 15 mM NaCl, 0,34 M saharoza, 0,15 mM spermin, 0,5 mM spermidin) s dodatkom tablete proteaznih inhibitora (PI; Roche, Basel, Švicarska; 1 tableta za 10 mL pufera), PMSF-a (2 mM) i Tritona X-100 (0,5 %). Mješavina je centrifugirana 5 min na 2500 x g na 4 °C, talog je ispran u CIB puferu bez mješavine PI-a i deterdženata te ponovno centrifugiran. Talog je resuspendiran u puferu CIB (bez deterdženta ili PI-a), u suspenziju je dodan CaCl<sub>2</sub> konačne koncentracije 1 mM i grijan 1 min na 37 °C. Razgradnja mikrokokalnom nukleazom (MNaza; Thermo Scientific 88216, LOT SF251245) napravljena je na 37 °C u trajanju od 5 min, korištenjem 1,5 U enzima po mg tkiva. Reakcija je zaustavljena dodavanjem 0,5 M EDTA do konačne koncentracije od 20 mM, a produkti razgradnje centrifugirani su 5 min na 8000 x g na 4 °C. Talog je resuspendiran u puferu PBS (137 mM NaCl, 2,7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1,8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7,4) s dodatkom NaCl-a do koncentracije od 500 mM i inkubiran na 4 °C u trajanju od 2 sata na rotatoru. Nakon inkubacije, uzorci su centrifugirani 15 min na 15000 x g na 4 °C te je sakupljen supernatant u kojem se nalazio izolirani kromatin. Stupanj razgradnje nativnog kromatina provjeren je gel-elektroforezom na 1 %-tnom agaroznom gelu.

### 3.4. Testiranje antitijela za protein Cg-CenH3 metodom *western blotting-a*

Poliklonalno antitijelo za protein Cg-CenH3 napravljeno je za potrebe Laboratorija prema prvotnim analizama dr. sc. Vojvoda Zeljko u vanjskoj instituciji Piñeda Antikoerper-Service u Berlinu. Plazma imuniziranih zečeva vađena je 60., 90. i 135. dan nakon imunizacije. Najbolja reaktivnost pokazana je za antitijela u plazmi vađenoj 135. dan nakon imunizacije te su se antitijela izolirana iz te plazme koristila u citogenetičkim imunofluorescencijskim pokusima. Antitijela su dobivena iz 3 različita zeca s gotovo jednakom imunološkom reaktivnošću, a za pokuse su korištena antitijela iz plazme jednog zeca.

Imunogeničnost antijela na protein Cg-CenH3 testirana je metodom *western blotting-a*. Za tu metodu korišten je ukupni proteinski izolat vrste *C. gigas* (samu izolaciju proteina napravila dr. sc. Mravinac) koji odgovara masi od približno 30 µg proteina po uzorku. Dodan je odgovarajući

volumen pufera Laemmli (62,5 mM Tris, 70 mM SDS, 10 % glicerol, 100 mM DTT). Smjesa je denaturirana 10 minuta na 65 °C te ohlađena na sobnu temperaturu i centrifugirana 1 minutu pri brzini 10 000 rpm. Proteinski uzorci razdvojeni su na 4-20 %-tnom gradijntnom poliakrilamidnom gelu SDS PAGE u puferu 1 x SDS PAGE (25 mM Tris, 192 mM glicin, 0,1 % SDS) pri konstantnoj vrijednosti električnog napona od 200 V u trajanju 30 min. Prije transfera proteina s gela na membranu, gel je ekvilibriran u ledeno hladnom puferu 1 x Transfer (25 mM Tris, 192 Mm Gly, 10 % metanol) 15 min. Nitrocelulozna membrana s porama veličine 0,2 µm namočena je 20 sekundi u reH<sub>2</sub>O te potom tijekom 5 min ekvilibrirana u hladnom puferu 1 x Transfer. Prijenos proteina proveden je pri konstantnoj jakosti struje od 200 mA, 40 minuta pri hladnim uvjetima na magnetskoj mješalici. Odmah nakon završetka prijenesa, uspješnost prijenesa proteina provjerena je otopinom za detekciju proteina *Ponceau S* i potom ispirana nekoliko puta u destiliranoj vodi u trajanju od 1 min. Membrana je zatim isprana puferom TBS/T (20 mM Tris, 150 mM NaCl, 0,1 % Tween) 3 puta po 5 min na tresilici pri sobnoj temperaturi te je potom inkubirana u blokirajućem puferu (5 %-tna otopina BSA u puferu TBS/T), 1 sat na sobnoj temperaturi. Membrana je razrezana te su dijelovi membrane inkubirani zasebno s preimuniziranom plazmom kunića i pročišćenim i ugušćenim monospecifičnim IgG-ima razrijeđenima u 5 %-tnoj blokirajućoj otopini BSA u puferu TBS/T u omjeru 1:500 te s jednim uzorkom nerazrijeđenog monoklonalnog IgG-a. Inkubacija je provedena preko noći pri temperaturi od 4 °C uz laganu trešnju membrana. Sljedeći dan, nevezana antitijela isprana su puferom TBS/T, 5 puta po 5 min na sobnoj temperaturi pri laganom trešnji, a zatim je membrana inkubirana sa sekundarnim anti-zečjim antitijelom konjugiranim s enzimom HRP (od engl. horseradish peroxidase), koji je razrijeđen 1:2000 u 5 %-tnoj BSA/TBS/T, 1 sat pri sobnoj temperaturi uz laganu trešnju. Nakon što su membrane isprane od suviška sekundarnog antitijela 5 puta po 5 min u puferu TBS/T pri sobnoj temperaturi uz laganu trešnju, inkubirane su 1 min u smjesi reagensa za detekciju (*Pierce ECL Western Blotting Detection Reagents*, Thermo Fisher Scientific, Waltham, Massachusetts, SAD). Signali su vizualizirani ručnim razvijanjem rentgenskih filmova u razvijaču i fiksiru (oba Kodak, Rochester, New York, SAD) s ekspozicijom u trajanju 3 min.

### 3.5. Testiranje specifičnosti antitijela za protein Cg-CenH3

Specifičnost vezivanja antitijela na protein CenH3 pokušali smo dokazati na mitotskim kromosomima vrste *C. gigas* napravljenih prema nekoliko različitih protokola. U različitim protokolima varirani su korišteni citostatici, kemikalije za hipotonični šok i fiksativi te su varirani načini na koji su kamenice predtretirane citostaticima (cijele kamenice ili samo škrge). Različite varijacije protokola prikazane su u Tablici 5.

Tablica 5. Varijacije protokola koji su se koristili pri izradi preparata s metafaznim kromosomima vrste *C. gigas*.

	tkivo kamenice tretirano citostatikom	citostatik	hipotonični šok	fiksativ
Protokol 1	cijela kamenica	kolhicin (0,005 %)	50 %-tna morska voda, 25 %-tna morska voda, 0,56 %-tni KCl i 0,9 %-tni Na-acetat	Carnoyev fiksativ (etanol i octena u omjeru 3:1)
Protokol 2	škrge	kolhicin (0,005 %)	50 %-tna morska voda, 25 %-tna morska voda, 0,56 %-tni KCl i 0,9 %-tni Na-acetat	Carnoyev fiksativ (etanol i octena u omjeru 3:1)
Protokol 3	škrge	kolhicin (0,005 %)	0,56 %-tni KCl	Carnoyev fiksativ (etanol i octena u omjeru 3:1)
Protokol 4	škrge	kolhicin (0,005 %)	0,9 %-tni Na-acetat	Carnoyjev fiksativ (etanol i octena u omjeru 3:1)
Protokol 5	škrge	kolhicin (0,005 %)	0,9 %-tni Na-acetat	2 %-tni paraformaldehid

Tkivo kamenice najprije je podvrgnuto hipotoničnom šoku u trajanju od 25 min u jednom od reagensa za hipotonični šok (ovisno o protokolu iz Tablice 4.). Nakon toga su tkiva, ili cijela kamenica, prebačeni u otopinu citostatika preko noći. Nakon tretiranja citostatikom, škrge su

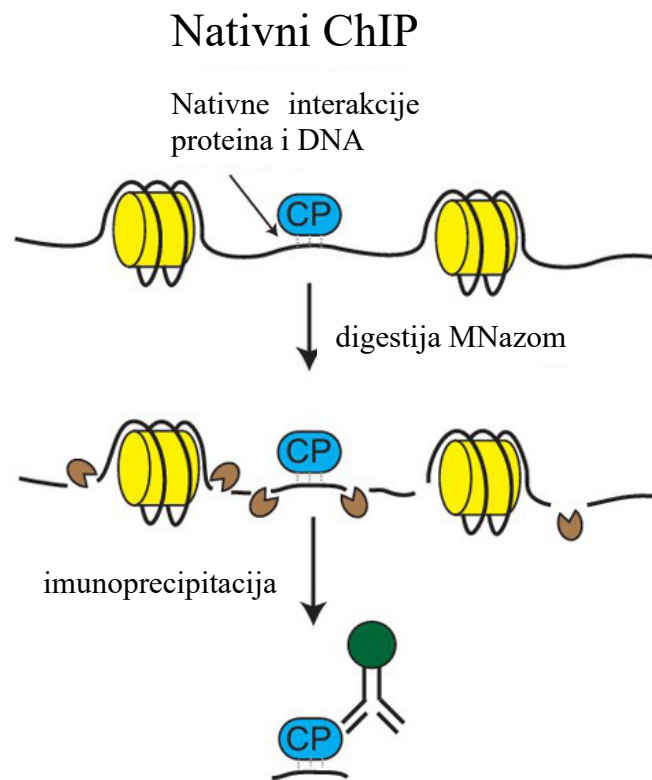
macerirane kapalicom i žiletom u otopini 60 %-tne octene kiseline (u svakom od protokola) te nakapane na predmetnice zagrijane na 50 °C i očišćene octenom kiselinom.

Kao pozitivne kontrole na dobivenim preparatima korištena su dva komercijalna monoklonalna antitijela, antitijelo za H3K9me3 (Cell Signaling 13969 and Merck 05-1242), razrijeđeno 1:800 i 1:75 te antitijelo za H3 (Cell Signaling, 4499), razrijeđeno 1:400. Niti antitijelo razvijeno za potrebe laboratorija niti tri komercijalna antitijela nisu dala signale na dobivenim preparatima mitotskih kromosoma. Zaključili smo da je razlog za to korištenje visokopostotne octene kiseline (60 %-tne) za maceriranje škržnog tkiva prilikom izrade preparata. Visokopostotna octena kiselina koristi se u pripremi preparata mitotskih kromosoma zbog svog svojstva da dobro razgrađuje zaostalo tkivo pa se zbog toga mogu dobiti čiste metafazne ploče, bez zaostalog tkiva koje na sebe nespecifično veže antitijela. Nažalost, korištenje visokopostotne octene kiseline ujedno je uništilo ciljne proteinske epitope što je rezultiralo izostankom signala, dok protokoli u kojima nije korištena visokopostotna octena kiselina nisu dali dovoljno čiste preparate. Stoga su eksperimenti imunolokalizacije provedeni na gonadnim preparatima koji imaju drugačiji način pripreme, kao što će biti objašnjeno kasnije.

### 3.6. Nativna kromatinska imunoprecipitacija

Mononukleosomalna frakcija kromatina izolirana iz stanica mišića aduktora iz vrste *C. gigas* razgrađenog MNazom korištena je u pokusima native kromatinske imunoprecipitacije, CHIP. Shematski prikaz imunoprecipitacije pokazan je na Slici 7. Imunoprecipitacija je napravljena prema protokolu dostupnom u komercijalnom setu *PIERCE magnetic CHIP KIT* (Thermo Fisher Scientific, 26157) uz nekoliko modifikacija. A/G magnetne kuglice blokirane su 1 h u 1 x PBS + 1 % BSA na 4 °C na rotatoru. Kromatin u imunoprecipitacijskom puferu (0,01 % SDS, 1,1 % Triton X-100, 1,2 mM EDTA, 16,7 mM Tris-HCl pH 8,1 i 167 mM NaCl) s dodatkom mješavine proteaznih inhibitora (PI) bio je pročišćen na kuglicama *Dynabeads* u trajanju od 1 sata na 4 °C na rotatoru. Za kontrolu imunoprecipitacijske reakcije, sačuvano je 10 % ukupne mase ulaznog kromatina (*input*) koji je pohranjen na 4 °C do zadnjeg koraka elucije. Antitijela su na magnetne kuglice vezane 2 h na 4 °C u 100 µL dilucijskog pufera *IP* koji je sastavni dio komercijalnog seta. Na kuglice su vezana dva antitijela, anti-Cg-CenH3 antitijelo koje je za potrebe laboratorija sintetizirano u vanjskoj instituciji (Piñeda Antikoerper-Service, Berlin, Njemačka) te

komercijalno monoklonalno anti-H3K9me3 antitijelo (Cell Signaling, 13969). Monoklonalno anti-H3K4me3 antitijelo (Cell Signaling 9751) i zečji IgG (Cell Signaling, 2729) korišteni su kao pozitivna i negativna kontrola. Sva antitijela dodana su u količini predloženoj od proizvođača (1  $\mu$ L na  $\mu$ g kromatina). Za svaku od reakcija kromatinske imunoprecipitacije korišteno je otprilike 5  $\mu$ g kromatina vrste *C. gigas*. ChIP s antitjelom na Cg-CenH3 napravljen je u dvije biološke replike, a ChIP s antitjelom na H3K9me3 u jednoj. Magnetne kuglice s vezanim antitjelom pomiješane su s pročišćenim kromatinom razrijeđenim 1:2 s dilucijskim puferom, a mješavina je inkubirana preko noći na 4 °C na rotatoru. Precipitirani imunokompleksi isprani su, eluirani i dalje obrađivani prema uputama iz protokola dostupnog u komercijalnom setu *PIERCE magnetic ChIP KIT*.



Slika 7. Na slici je prikazana shema imunoprecipitacije nativnog kromatina izoliranog iz stanica mišića aduktora vrste *C. gigas*. Nakon izolacije, kromatin se razgradi mikrokokalnom nukleazom (MNazom) te imunoprecipitira antitjelom na protein od interesa (na shemi označen plavo). U



ovom radu, imunoprecipitiran je kromatin vezan na protein Cg-CenH3 i histonsku modifikaciju K9me3 histona H3. Slika je prilagođena prema He i Bonasio (2017.).

### 3.7. Sekvenciranje DNA

U vanjsku instituciju (*Admera Health, SAD*) na sekvenciranje su poslani sljedeći uzorci: genomske DNA vrste *C. gigas*, uzorci kromatina imunoprecipitirani dvama antitijelima, anti-Cg-CenH3 (dvije replike) i anti-H3K9me3, za vrstu *C. gigas* te kontrolni uzorci *input*. Za potrebe komparativne satelitomike sekvencirane su i genomske DNA vrsta *C. virginica*, *C. angulata* i *O. edulis*. U vanjskoj instituciji pripremljene su biblioteke i DNA je sekvencirana na platformi *Illumina HiSeqX*. Tablica 6. prikazuje broj očitavanja nakon sekvenciranja genomske DNA za svaku od četiriju vrsta i svih tipova imunoprecipitiranog i kontrolnog kromatina.

Tablica 6. Prikaz broja i duljine očitavanja dobivenih nakon sekvenciranja za svaku vrstu i sve tipove imunoprecipitiranog i kontrolnog kromatina.

	Vrsta	Broj očitavanja	Duljina pojedinog očitavanja (pb)
genomska DNA	<i>C. gigas</i>	2 x 2 768 912	151
	<i>C. virginica</i>	2 x 18 121 416	151
	<i>C. angulata</i>	2 x 16 587 089	151
	<i>O. edulis</i>	2 x 16 954 299	151
imunoprec. kromatin	Cg-CenH3, 1. replika	2 x 2 419 817	151
	Cg-CenH3, 2. replika	2 x 6 463 282	151
	H3K9me3	2 x 5 516 696	151
	<i>Input</i> (1. replika Cg-CenH3)	2 x 2 518 230	151
	<i>Input</i> (2. replika Cg-CenH3 i H3K9me3)	2 x 5 570 873	151

### 3.8. Obrada jedinki vrste *C. gigas* korištenih u citogenetičkim pokusima

Mlade jedinke vrste *C. gigas* prikupljene su u svom prirodnom staništu u Aveiru u Portugalu. Kamenice su tjedan dana intenzivno hranjene mikroalgama u prozračivanim i filtriranim akvarijima na 18 °C kako bi se potaknuo njihov somatski rast i sazrijevanje gonada. Nakon

perioda intenzivnog hranjenja, jedinke su stavljene u otopinu citostatika kako bi se što više stanica iz tkiva škruga zaustavilo u metafazi stanične diobe.

### 3.9. Priprema citogenetičkih preparata s mitotskim kromosomima

Preparati s mitotskim kromosomima pripremljeni su na način opisan u Martinez-Exposito i sur. (1994.), uz nekoliko manjih modifikacija. Nakon 12-osatnog tretmana u 0,005 %-tnoj otopini kolhicina, kamenice su otvorene, tkivo škruga izrezano te stavljeno u hipotoničnu otopinu morske vode, najprije u 50 %-tnu 25 min, a zatim u 25 %-tnu 25 min. Nakon toga, tkivo je fiksirano u Carnoyevom fiksativu (etanol i octena kiselina u omjeru 3:1). Tkivo je svakih pola sata prebačeno u svježi fiksativ, a nakon treće izmjene tkivo je do daljnje upotrebe pohranjeno u fiksativu na -20 °C. Stanična suspenzija dobivena je sjeckanjem škržnog tkiva u 60 %-tnoj octenoj kiselini i nakapavana na predmetna stakalca predgrijana na 56 °C. Pod svjetlosnim mikroskopom odabrana su stakalca sa zadovoljavajućim brojem prisutnih metafaza. Stakalca su isušena u seriji alkohola (najprije 70 %-tni, zatim 90 %-tni i na kraju 100 %-tni) i osušena preko noći na sobnoj temperaturi prije pohrane na -80 °C.

Kao što je spomenuto, priprema preparata mitotskih kromosoma s fiksativima koji sadrže visok udio octene kiseline (> 5 %) nisu se pokazali pogodnima za pokuse fluorescencijske imunodetekcije ni komercijalnim monoklonalnim antitijelima ni poliklonalnim antitijelom proizvedenim za potrebe laboratorija (anti-Cg-CenH3). Naime, visoki udjeli octene kiseline uništavaju proteinske epitope koji su ključni za detekciju proteina od interesa metodom imunofluorescencijske detekcije (IF). Zbog toga su se za pokuse IF-a, uključujući kombiniranu metodu IF-a i fluorescencijske hibridizacije *in situ* (FISH), koristili preparati s muškim gonadnim stanicama u različitim razvojnim fazama pripremljeni s fiksativima koji ne sadrže octenu kiselinu ili sadrže njen mali postotak.

### 3.10. Priprema citogenetičkih preparata s gonadnim stanicama

Za potrebe lokalizacije centromernog histona H3, napravljeni su citogenetički preparati koji sadrže muške spolne stanice različitih razvojnih faza. Za pripremu tih preparata isprobano je nekoliko različitih protokola. U Tablici 7. vidljivi su parametri koji su mijenjani u svakom pojedinom protokolu (citostatici, fiksativi).

Tablica 7. Prikaz parametara koji su mijenjani u svakom od fiksativa sa ciljem dobivanja što kvalitetnijih preparata.

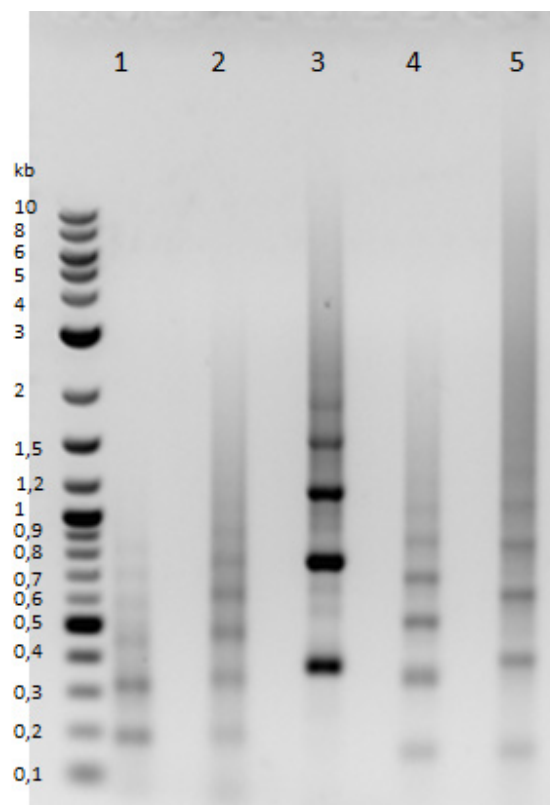
	Citostatik	Fiksativ
Protokol 1	-	metanol:aceton=1:1
Protokol 2	kolcemid (0,01 %)	metanol:aceton=1:1
Protokol 3	kolcemid (0,01 %)	paraformaldehid (3,7 %) + octena kiselina (5 %) + NaCl (0,9 %)
Protokol 4	kolcemid (0,01 %)	etanol:octena=3:1
Protokol 5	kolcemid (0,01 %)	metanol:octena=3:1
Protokol 6	kolcemid (100 %)	metanol:aceton=1:1
Protokol 7	kolhicin (0,1 %)	metanol:aceton=1:1
Protokol 8	kolhicin (0,1 %)	etanol:octena=3:1

Nakon izolacije gonada iz jedinki vrste *C. gigas*, stavljene su u 0,4 M NaCl te su pod svjetlosnim mikroskopom izabrane gonade muških jedinki koje su sadržavale stanice u različitim razvojnim fazama. Od gonada izabranih jedinki napravljene su stanične suspenzije. Najprije su gonade isprane tri puta u PBS-u, nakon toga fiksirane u mješavini metanola i acetona u omjeru 1:1. Fiksirane gonadne stanice nakapavane su na stakalca predtretirana poly-D-lizinom. Stakalca su zatim prekrivena pokrovnicom i nakratko uronjena u tekući dušik. Zatim su izvađena te je uklonjena pokrovnica, a stakalca pohranjena na -80 °C do upotrebe. Faza sazrijevanja gonadnih stanica određena je prema Franco i sur. (2008.).

### 3.11. Označavanje sondi za pokuse imunofluorescencije

#### 3.11.1. Označavanje sondi za FISH

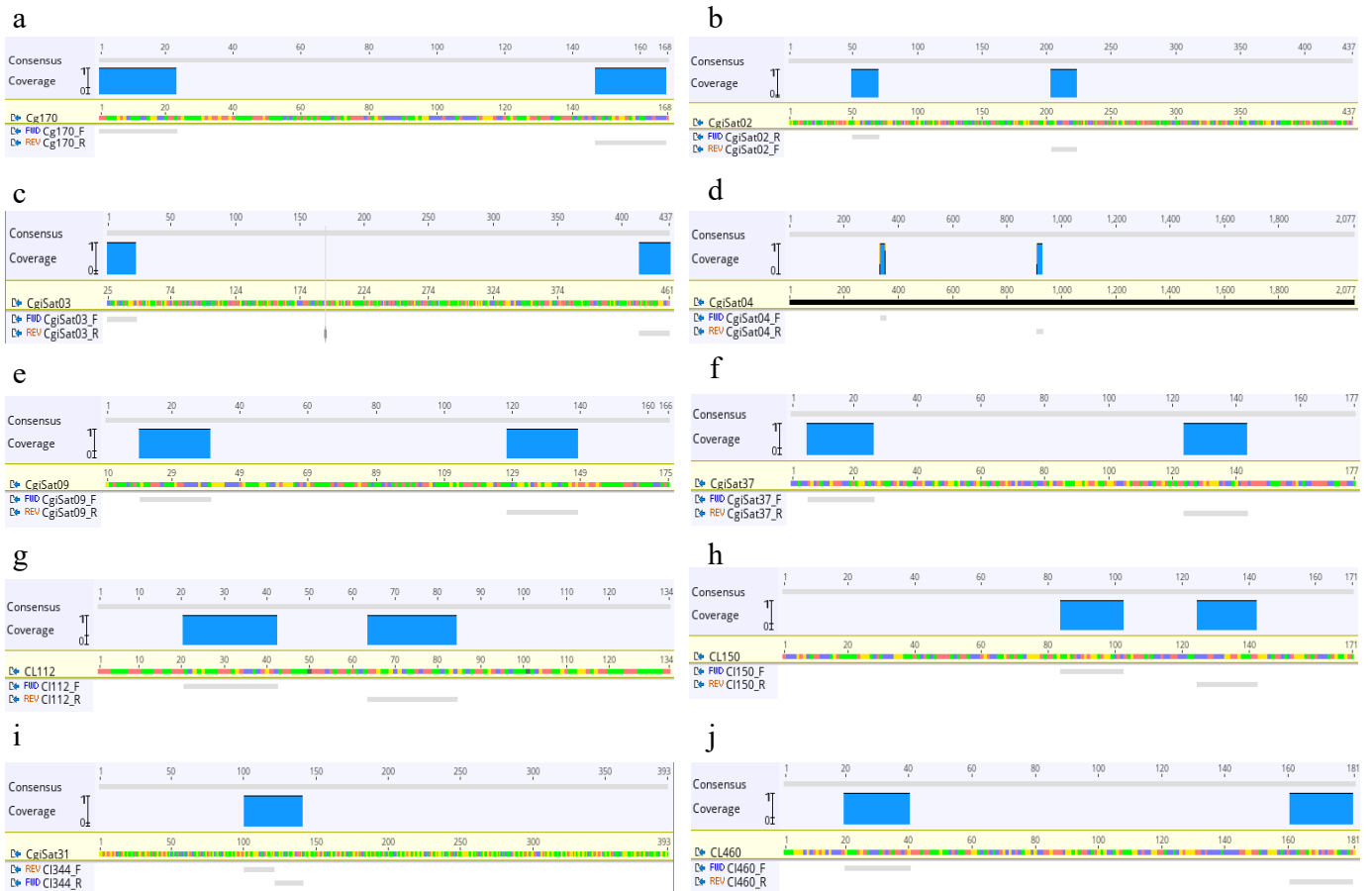
Početnice za sonde napravljene su u programu *Geneious* te sintetizirane u ustanovi Macrogen. Sonde za FISH sintetizirane su i označene lančanom reakcijom polimerazom. Za sonde: CgiSat01 (Cg170), C1112, C1150, C1344, C1460 i C1485 reakcijska smjesa sadržavala je 250 ng DNA, 50 μM dATP, dGTP i dCTP, 32 μM dTTP i 18 μM biotin-16-dUTP, početnice (1 μM svaka; Jena BioScience), 2,5 mM MgCl<sub>2</sub>, 2,5 U Go Taq DNA polymerase i 1 x Go Taq Flexi Reaction Buffer (sve Promega osim ako nije već navedeno) u volumenu od 50 μL. Dobiveni karakteristični ljestvičasti obrazac razdvajanja na agaroznom gelu za navedene satDNA prikazan je na Slici 8.



Slika 8. Slika prikazuje karakterističan ljestvičasti obrazac razdvajanja satDNA na agaroznom gelu. Kolone sadrže satDNA: C1112 (1), C1150 (2), C1344 (3), C1460 (4), C1485. Takav obrazac nastaje zbog umnažanja satDNA u višekratnicima duljine osnovnog monomera te satDNA.

Za sonde: CgiSat02, 03, 04, 05, 09, 17, 28, 37, 46 i 5S rDNA, reakcijska je smjesa sadržavala 50 ng DNA, 100  $\mu$ M dATP, dGTP i dCTP, 65  $\mu$ M dTTP, 2,5 mM MgCl<sub>2</sub>, 2,5 U GoTaq G2 Flexi Taq DNA polymerase, 1 x GoTaq Flexi Reaction Buffer (sve Promega), početnice (svaka 1  $\mu$ M, Macrogen) i 35  $\mu$ M biotin-16-dUTP (sondee satDNA; Jena Bioscience) ili 35  $\mu$ M digoksigenin-11-dUTP (za 5S rDNA, Roche) u volumenu od 50  $\mu$ L. Početna denaturacija trajala je 5 min na 94 °C, a zatim je slijedilo 30 ciklusa umnažanja pod uvjetima navedenima za svaki set početnica navedenih u Tablici 8. Završna ekstenzija trajala je 7 min na 72 °C. Sekvencije svih početnica i uvjeti svih reakcija prikazani su u Tablici 7. Sekvencija početnica za kontrolnu sondu (5S rDNA) sintetizirane su prema Pèrez-Garcia (2011.). Sonde su pročišćene korištenjem komercijalnog kompleta *QIAquick PCR Purification Kit* (Qiagen) prema protokolu dostupnom u kompletu. Sonde su provjerene na 1 %-tnom agaroznom gelu, a koncentracija pročišćenih sondi mjerena je

uređajem *Qubit Fluorometer*. Za svaku hibridizacijsku reakciju korišteno je otprilike 30 ng sonde po preparatu. Prikaz mjesta u sekvenciji monomera satDNA CgiSat01, 02, 03, 04, 09, 37, 46, C1112, 150, 344 i 460 na koja se vežu početnice prikazan je na Slici 9. Ponovljene sekvencije koje su u sklopu analize ChIP-Seq Mapper izdvojene kao relevantne (C1112, 150, 344 i 460), nakon analize satelitoma klasificirane su kao CgiSat17, CgiSat46, CgiSat31 i CgiSat28, stoga se pojavljuju pod oba imena.



Slika 9. Slika prikazuje na koja se mjesta u sekvenciji monomera pojedine satDNA vežu sintetizirane početnice. Slika prikazuje redom satDNA: CgiSat01 (Cg170) (a), CgiSat02 (b), CgiSat03 (c), CgiSat09 (d), CgiSat37 (e), CL112 (CgiSat17) (f), CL150 (CgiSat46) (g), CL344 (CgiSat31) (h), CL460 (CgiSat28) (i).

Tablica 8. Sekvencije početnica i uvjeti reakcija kojima su sintetizirane sonde za FISH.

Sekvencije početnica	Uvjeti PCR-a		
	Početno razdvajanje	Umnažanje (30 x)	Završno produljivanje
CgiSat01 (Cg170)_F: AATTTGAGCTTTCCTGGCCTAAT CgiSat01 (Cg170)_R: GGGTGGAAGCATCCTCAGATA	94 °C/5 min	94 °C/30 s 51 °C/30 s 72 °C/30 s	72 °C/7 min
Cl112_F: TCGTGTTTAAATGACCTCAGAC Cl112_R: GCAGAATTGTAAGCGCAACTT		94 °C/30 s 62 °C/30 s 72 °C/30 s	
Cl150_F: TTGTACAGAGGTAGGTCAC Cl150_R: TTAAACTTCCGGGTTGGG		94 °C/30 s 47 °C/30 s 72 °C/30 s	
Cl344_F: CACAAGTGAAGCCCCAACTAC Cl344_R: CTTGTGCTCGTTGGCAAAGC		94 °C/30 s 60 °C/30 s 72 °C/30 s	
Cl460_F: GTTTAGATCCCTACATAACAG Cl460_R: TAGGACTTGTGCTCATTAGG		94 °C/30 s 60 °C/30 s 72 °C/30 s	
Cl485_F: CTGAAGCAAGTACCACGAGC Cl485_R: AGACTCCGTGCTGCCTAC		94 °C/30 s 62 °C/30 s 72 °C/30 s	
CgiSat02_F: TAGGGGTTTTTAGGGGCCAA CgiSat02_R: TAGGGGCCGATCCCTTATCTC		94 °C/30 s 56 °C/30 s 72 °C/30 s	
CgiSat03_F: CCTCTAGATCCTTAATGGGGACA CgiSat03_R: GCTGGCCCCTAAAATATTCAATCA		94 °C/30 s 58 °C/30 s 72 °C/30 s	
CgiSat04_F: TGCAGCCATTTTGCTGTCTG CgiSat04_R: CAGCTCGCCCTACATGTGAT		94 °C/30 s 55 °C/30 s 72 °C/30 s	
CgiSat05_F: GGAAGTAAAGCTGGAAGTTCCTGTA CgiSat05_R: TGTCACTTCCGGTCGAAACC		94 °C/30 s 58 °C/30 s 72 °C/30 s	
CgiSat09_F: GTTAAACTTTGAACCCCGCCT CgiSat09_R: CTCAAGAACCACTGCACCAGA		94 °C/30 s 58 °C/30 s 72 °C/30 s	
CgiSat37_F: AGCATCCGAAACCTATGGCTC CgiSat37_R: TTGCAGGTGCCCTCTGATGA		94 °C/30 s 56 °C/30 s 72 °C/30 s	

5S rDNA_F: GAAGTTAAGCAACGTAGAG 5S rDNA_R: TCAATCTTTCTATCTGTG		94 °C/30 s 43 °C/30 s 72 °C/30 s	
---	--	--	--

### 3.11.2. Označavanje sonde DNA iz frakcija kromatina u asocijaciji sa centromernim histonom H3 i histonom H3 trimetiliranim na lizinu 9

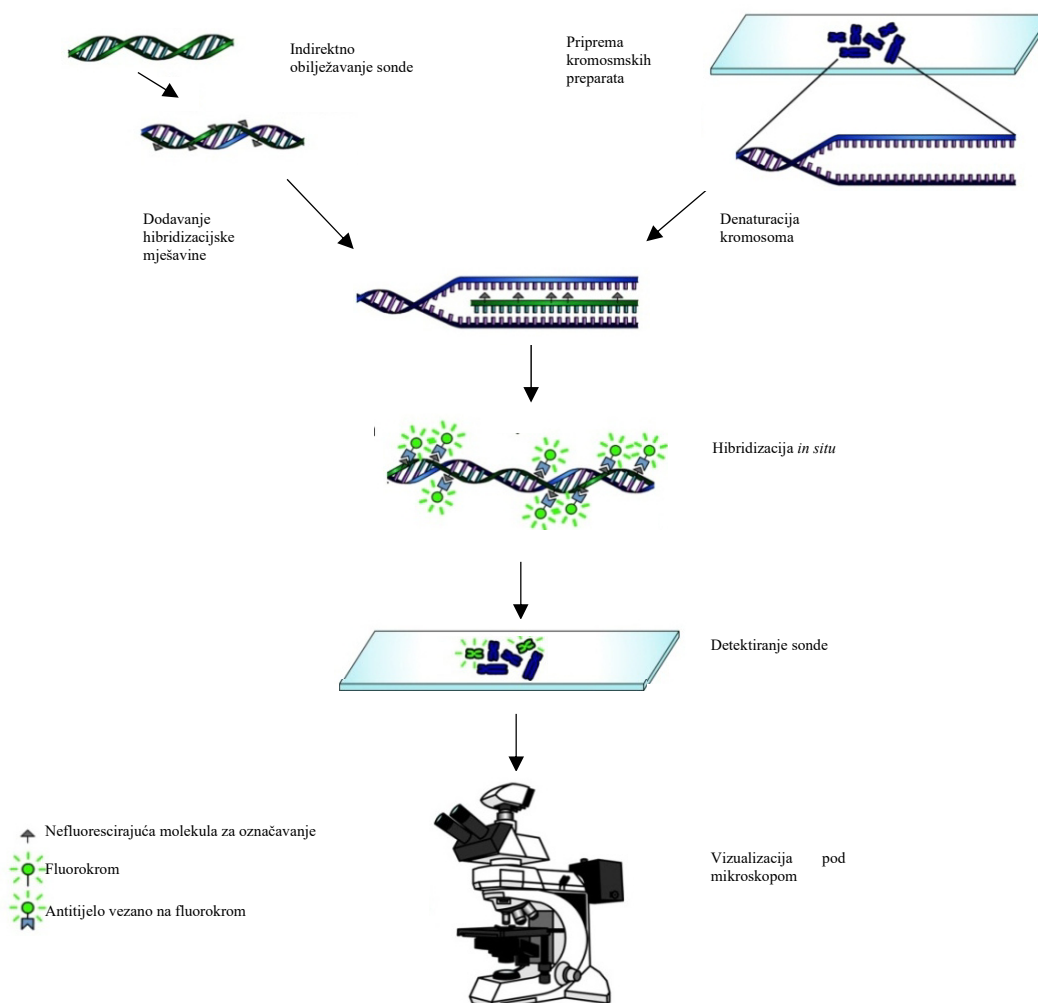
Sonde za dvije kromatinske frakcije obilježene su metodom nasumičnog vezanja heksanukleotidnih početnica i umnažanjem Klenowljevim fragmentom prema protokolu iz komercijalnog seta Random Primed DNA labeling Kit (Roche, 1100476001, Basel, Švicarska). Sonde su označene biotinom obilježenim d-UTP-om. Reakcije obilježavanja sadržavale su: 50  $\mu$ M nukleotide dATP, dGTP, dCTP te 45  $\mu$ M dTTP i 5  $\mu$ M biotinom označeni dUTP u svakoj od reakcijskih smjesa, 100 ng DNA imunoprecipitirane antitijelom na Cg-CenH3 u jednoj ili 50 ng DNA imunoprecipitirane antitijelom na H3K9me3 u drugoj reakcijskoj smjesi te pufer NEB 2 (dostupan u komercijalnom *kit-u*) i 7,5 U Klenowljevog fragmenta u svakoj od reakcija. Uspješnost obilježavanja sonde i kolorimetrijski test za mjerenje koncentracija dobivenih sonde napravljeni su prema protokolu iz istog komercijalnog seta. Koncentracije dobivenih sonde izmjerene su uređajem *Qubit Fluorometer*.

### 3.12. Fluorescencijska hibridizacija *in situ*

Pokusi FISH napravljeni su prema protokolu opisanom u Perez-Garcia i sur. (2011.), s promjenom u koraku pepsinske razgradnje (5 min na 37 °C). Shematski je metoda prikazana na Slici 10. Prije hibridizacije, DNA sonde denaturirane su 8 min na 80 °C i odmah nakon toga stavljene 2 min na led. Hibridizacija se odvijala preko noći u vlažnoj komori na 37 °C. Nakon toga su uslijedila tri ispiranja u trajanju od 5 minuta u puferu PBS uz dodatak 0,2 % deterdženta Tween-20. Avidin-D obilježen fluorescein izotiocijanatom (FITC) i biotinizirani anti-avidin D (oba Vectashield, Newark, Kalifornija, SAD) korišteni su za detekciju sonde obilježenih biotiniziranim nukleotidom. Anti-digoksinin Fab fragmenti obilježeni rodaminom korišteni su za detekciju sonde obilježenih digoksinin-11-dUTP-om. Za detekciju biotinom označenih nukleotida korištena tzv. metoda sendviča kako bi se pojačalo vezivanje antitijela na biotin.



Preparati su najprije inkubirani s antitijelom na biotin (razrjeđenje 1:500) u trajanju 45 min na 37 °C u vlažnoj komori. Druga inkubacija je bila s biotiniziranim antitijelom anti-avidin u razrjeđenju 1:100, također u trajanju od 45 min na 37 °C u vlažnoj komori. Treća inkubacija ponovno je bila s antitijelom anti-biotin, ovaj put u razrjeđenju 1:1000 uz iste uvjete hibridizacije. Između svake inkubacije uslijedila je serija od tri ispiranja u puferu PBS s dodatkom 0,2 % deterdženta Tween. U pokusima FISH-a u kojima su istovremeno detektirani i satDNA i sonde za 5S rDNA, u prvu i treću inkubacijsku smjesu s detekcijskim antitijelom za biotin dodani su i rodaminom označeni antidigoksinogeninski Fab fragmenti iz ovce (Roche, Basel, Švicarska). U prvoj inkubaciji razrjeđenje je bilo 1:200, a u trećoj inkubaciji, razrjeđenje je bilo 1:400. Kromosomi su obojani otopinom (100 ng/mL) 4',6'-diamidino-2-fenilindola (DAPI; Sigma-Aldrich) te zaštićeni sredstvom protiv izbjeljivanja fluorescencije Mowiol 4-88 (Sigma-Aldrich). Za vizualizaciju preparata korišteni su Nikon Eclipse-800 fluorescencijski mikroskop i Leica TCS SP8 X laserski pretražni mikroskop.



Slika 10. Shematski prikaz metode fluorescencijske hibridizacije *in situ*. U ovom radu, korištena je metoda indirektnog označavanja sonde za hibridizaciju koja se kasnije detektira antitijelima vezanima za fluorofore. Slika prilagođena prema Brammer i sur., 2013.

### 3.13. Pokusi imunofluorescencijske detekcije (IF)

#### 3.13.1. Imunofluorescencijska detekcija

Nakon pripreme opisane u odlomku 8., preparati su isprani 2 min u 1 x PBS-u i 2 x 5 min u 1 x PBS/0,2 % Tween 20. Nespecifična vezna mjesta blokirana su 2,5 %-tnom otopinom govedeg serumskog albumina (BSA, od engl. bovine serum albumin). Nakon toga, gonadne stanice inkubirane su sa zečjim anti-Cg-CenH3 primarnim antitijelom, razrijeđenim u omjeru 1:400 s 1 x PBS/0,2 %-tni Tween20/2,5 %-tni BSA preko noći na 37 °C u vlažnoj komori. Nakon prekonoćne inkubacije, preparati su 3 puta isprani u 1 x PBS/0,2 % Tween 20 i potom inkubirani 1 h na 37 °C u vlažnoj komori sa sekundarnim antitijelom (kozji anti-zečji IgG označen s fluroforom Alexa 594; Abcam ab 150080) razrijeđenim u omjeru 1:1000 s 2,5 %-tnim BSA u 1 x PBS/0,2 %-tni Tween20. Potom su uslijedila dva ispiranja po 5 min u 1 x PBS/0,2 %-tni Tween20 i jedno ispiranje od 2 min u 1 x PBS. Jezgre su obojane otopinom DAPI (100 ng/mL) i zaštićene sredstvom Mowiol 4-88.

#### 3.13.2. Dvostruka imunofluorescencijska detekcija

Pokusi dvostruke IF napravljeni su prema istom protokolu. Zečji anti-Cg-CenH3 IgG u razrjeđenju 1:400 i mišji anti-H3K9me3 IgG (Merck 05-1242, Darmstadt, Njemačka) razrijeđen 1:75 korišteni su kao primarna antitijela. Sekundarna antitijela bila su kozji anti-zečji IgG označen Alexa-594 fluoroforom (Abcam ab150090, Cambridge, Ujedinjeno Kraljevstvo; 1:1000) i magareći anti-mišji IgG označen fluoroforom Alexa – 488 (Sigma-Aldrich, SAB4600035, St. Louis, Missouri, SAD; 1:500). Preparati su istovremeno inkubirani s oba primarna antitijela tijekom prve inkubacije i s oba sekundarna antitijela tijekom druge inkubacije.

#### 3.13.3. IF-FISH

Kombinacija dviju tehnika, IF-a i FISH-a, napravljena je prema opisanim protokolima. Nakon inkubacije sa sekundarnim antitijelom i ispiranja s PBS/Tween-om i PBS-om u pokusu IF-a,

preparati nisu bili obojani otopinom DAPI-ja nego se odmah započelo s pokusom FISH-a. Razgradnja RNazom (koncentracije 10 µg/mL) trajala je 45 min, a razgradnja pepsinom nije primijenjena. Ostatak pokusa proveden je prema ranije opisanom protokolu.

### 3.14. Računalne analize

#### 3.14.1. Analize dostupne unutar paketa *RepeatExplorer*

Za analize detekcije ponavljajućih sekvencija DNA korišten je paket računalnih analiza *RepeatExplorer* besplatno dostupan u sklopu *Galaxy* servera, na poveznici: <https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>. Sve analize temelje se na *de novo* identifikaciji ponavljajućih sekvencija uz pomoć klasteriranja temeljenog na grafikonima (Novák i sur., 2013.). U ovom radu korištene su sljedeće analize: identifikacija ponavljajućih sekvencija *de novo* u cjelokupnom genomu pojedine vrste uz pomoć analize *RepeatExplorer*, ponavljajuće sekvencije DNA obogaćene u uzorcima centromernog kromatina i kromatina označenog trimetilacijom Lys9 na histonu H3 detektirane su analizom *ChIP-Seq Mapper*, u nekoliko krugova analiza *TAREAN* određen je cjelokupni satelitom svake od vrsta analiziranih u ovom radu, komparativnom *RepeatExplorer* analizom uspoređeni su satelitomi četiriju vrsta analiziranih u ovom radu. Pojedina ponavljanja analiza *TAREAN* razlikovali su se u broju očitavanja uzetih za pojedinu analizu te u drugačijem setu odabranih očitavanja za svaku analizu u pojedinoj vrsti.

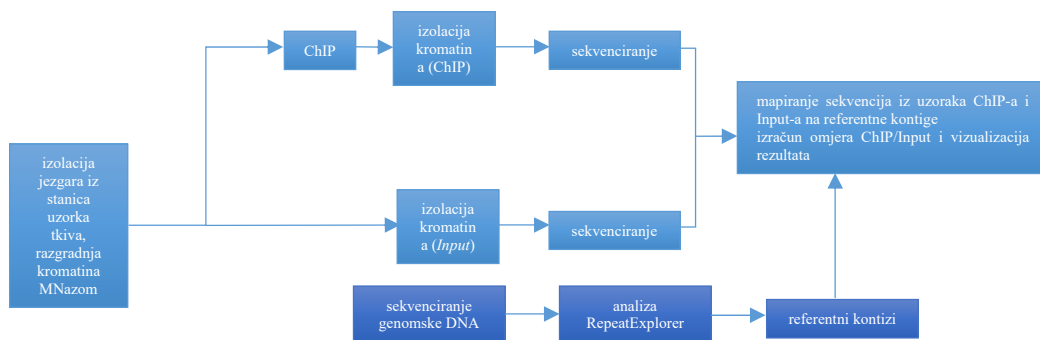
#### 3.14.2. Analiza *RepeatExplorer*

Identifikacija ponavljajućih sekvencija u cjelokupnoj genomske DNA određena je analizom *RepeatExplorer* temeljenoj na klasteriranju prema sličnosti očitavanja napravljenoj prema zadanim vrijednostima programa. Genomska očitavanja su prije analize probrana prema kvaliteti, skraćena na krajevima kako bi se uklonili adapteri i nedovoljno precizna očitavanja, spareni te su izabrani oni parovi očitavanja koji nisu imali preklop u nukleotidnoj sekvenciji. Procjena kvalitete očitavanja napravljena korištenjem posebnog alata u sklopu servera *Galaxy* i paketa analiza *RepeatExplorer* (*Preprocessing of FASTQ paired-end reads*). Kvaliteta očitavanja (engl. *quality score*) mora biti  $\geq 10$  za minimalno 95 % baza i bez neidentificiranih nukleotida. Očitavanja su skraćena na duljinu 115 nt (preporučena duljina očitavanja za ove računalne analize je 100-200 nt). Adapteri se tijekom sekvenciranja nove generacije vezuju na molekule DNA jer se tijekom jednog sekvenciranja

istovremeno sekvencira DNA nekoliko vrsta te adapteri služe za razlikovanje svake pojedine vrste. Na taj su način obrađena očitavanja za svaku od analiza napravljenih u paketu *RepeatExplorer*.

### 3.14.3. Analiza ChIP-Seq Mapper

Obogaćene ponavljajuće sekvencije DNA u uzorcima kromatina imunoprecipitiranog antitijelima na protein Cg-CenH3 i protein H3K9me3, određeni su analizom *ChIP-Seq Mapper*. Shematski prikaz analize vidljiv je na Slici 11. Tom se analizom detektira postoje li obogaćenja pojedinih sekvencija unutar sekvencija dobivenih sekvenciranjem nove generacije uzoraka dobivenih pokusima kromatinske imunoprecipitacije, pri čemu se kao referentna baza ponavljajućih sekvencija koriste ponavljanja identificirana nakon klasteriranja genomske DNA u analizi *RepeatExplorer*. Prema protokolu analize, BLASTN pretragom određuje se sličnost imunoprecipitiranih sekvencija s referentnom bazom. Očitavanja se dodjeljuju onim klasterima koja su pokazala najveću sličnost s određenim očitanjem i svako očitavanje dodjeljuje se samo jednom klasteru. Nakon raspodjele očitavanja, evaluira se ukupni broj očitavanja iz imunoprecipitiranog i *input* uzorka dodijeljenih svakom od klastera te se oni međusobno usporede prema normaliziranom broju očitavanja i utvrdi se koje sekvencije pokazuju obogaćenje (Neumann i sur., 2012.).



Slika 11. Shematski prikaz postupaka prilikom analize *ChIP-Seq Mapper*. Kao ulazni podaci za analizu koriste se sekvencije dobivene sekvenciranjem uzoraka kromatina (*input*-a te antitijelima precipitiranih uzoraka) i referentne sekvencije, kontizi dobiveni analizom *RepeatExplorer* sekvencija dobivenih sekvenciranjem potpune genomske DNA.

#### 3.14.4. Modificirana analiza *ChIP-Seq Mapper*

Centromerno specifične sekvencije u genomu *C. gigas* utvrđene su analizom *ChIP-Seq Mapper*. Prema rezultatima prve analize, obogaćenja sekvencija prisutnih u većem broju očitavanja nisu visoke značajnosti (između 1 i 2 prema omjeru imunoprecipitiranog i *input* uzorka). Značajnija obogaćenja pokazale su sekvencije koje su bile niže zastupljene. Primjerice, određeni klasteri koji su pokazali obogaćenje od 4 x, u imunoprecipitiranom su uzorku bili prisutni sa samo osam očitavanja, a u uzorku *input* s dva očitavanja što je dalo pogrešnu percepciju o obogaćenju takvih klastera. Također, uočeno je da se unutar pojedinog klastera može nalaziti više različitih kontiga, u kojim se dotična sekvencija ne nalazi u istom organizacijskom obliku, niti svi sadrže identičnu nukleotidnu sekvenciju. Detaljnijim analizama utvrđeno je da je obogaćenje klastera usrednjenje obogaćenja pojedinih kontiga koji sačinjavaju klaster te da, sukladno tome, nema svaki kontig isto obogaćenje. Da bi se utvrdilo koji točno kontizi sadrže sekvencije koje pokazuju najveće obogaćenje, kontizi su razdvojeni te je u ponovljenoj analizi *ChIP-Seq Mapper* svaki kontig tretiran zasebno, tj. kao zasebni klaster.

S takvim razdvojenim kontizima uspoređena su očitavanja imunoprecipitiranih kromatina te kontrolnih reakcija (*input*). Očitavanja koja su imala omjer ChIP/Input > 1,7 te ukupni broj pojavljivanja (*hit*-ova) > 500, smatrana su obogaćenim očitavanjima u ChIP uzorku. Odabrani kriteriji temeljili su se na prosjeku obogaćenja koja su korištena u istraživanjima gdje je za protein CenH3 pokazano da je povezan s mnogim različitim sekvencijama (primjerice Kowar i sur., 2016. i Zhang i sur., 2017.). Očitavanja dobivena iz obje biološke replike kromatinske imunoprecipitacije s antitijelom na CenH3 za vrstu *C. gigas* analizirana su modificiranom analizom *ChIP-Seq Mapper*. Centromernim kandidatima smatrana su ona očitavanja koja su u obje biološke replike zadovoljavala navedene kriterije.

#### 3.14.5. Analiza *TAREAN*

Ukupni sastav satDNA u svakoj od četiri analizirane vrste određen je analizom *TAREAN*. Kao ulazni podaci za analizu vrste *C. gigas* uzeti su setovi očitavanja korišteni u analizi *RepeatExplorer*, a kao ulazni podaci za analize *TAREAN* preostalih triju vrsta kamenica korištena su četiri seta podataka za četiri analize za svaku vrstu (Tablica 9.).

Tablica 9. Brojevi očitavanja korištenih kao ulazni dokument za analizu TAREAN za vrste *C. virginica*, *C. angulata* i *O. edulis*.

vrsta	br. očitavanja	pokrivenost genoma
<i>C. gigas</i>	2 x 1000000	0,25 x
	2 x 2000000	0,5 x
<i>C. virginica</i>	2 x 1221425	0,25 x
	2 x 2442858	0,5 x
<i>C. angulata</i>	2 x 1053572	0,25 x
	2 x 2107143	0,5 x
<i>O. edulis</i>	2 x 2035715	0,25 x
	2 x 4071429	0,5 x

Jedna od grupa očitavanja koji pripada pokrivenosti genoma od 0,25 x za svaku od vrsta izabrana je po klici 10 u alatu za nasumični odabir grupa očitavanja za svaku od vrsta, a druga grupa podataka koja je dala pokrivenost od 0,25 x za svaku od vrsta izabrana je po klici 13. Na jednak način izabrane su grupe koje su dale pokrivenost od 0,5 x.

### 3.15. Usporedba s bazom podataka *Repbase*

Konsenzusi ponavljajućih sekvencija obogaćenih u analizama *Chip-Seq Mapper* i dobivenih analizama satelitoma kamenica uspoređeni su s bazom podataka *Repbase* (Jurka, 1998.) korištenjem programa CENSOR. Baza sadrži dostupne ponavljajuće sekvencije iz različitih vrsta. Navedeni program unutar baze podataka *Repbase* dostupan je besplatno na poveznici: <https://www.girinst.org/censor/index.php>.

### 3.16. Analize napravljene u programu *Geneious*

#### 3.16.1. Usporedba sekvencija satDNA dobivenih analizama različitih setova očitavanja u analizi *TAREAN*

U programu *Geneious* napravljene su međusobne usporedbe dobivenih sekvencija satDNA iz svake pojedine analize *TAREAN* za svaku od vrsta. Za svaku detektiranu satDNA određena je konsenzus sekvencija koja je korištena u daljnim analizama (lokalizacija na kromosomima za vrste *C. gigas* i *C. virginica* i usporedba s bazom *Repbase* za sve četiri analizirane vrste).

### 3.16.2. Lokalizacija detektiranih satDNA na kromosomima vrsta *C. gigas*

Konsenzus sekvencije satDNA detektirane analizom TAREAN anotirane su na kromosome dviju vrsta, *C. gigas*. Za svako anotiranje dopuštena je razlika od 30 % u odnosu na konsenzus sekvenciju kako bi se obuhvatile različite varijante pojedine satDNA. Kromosomi obje vrste povučeni su iz baze NCBI. Za vrstu *C. gigas* korištene su dvije javno dostupne verzije genoma, sastavljene do razine kromosoma. Prvi se set nalazi pod oznakom GCA\_902806645.1. i sadrži 10 kromosoma i 200 nesvrstanih segmenata (od engl. unplaced scaffolds), a drugi po oznakom GCA\_011032805.1 i sadrži 10 kromosoma.

### 3.16.3. Određivanje postranih regija satDNA vrste *C. gigas*

Sa ciljem utvrđivanja dominantnog organizacijskog oblika satDNA u genomu vrste *C. gigas*, za neke od satDNA detektiranih u genomu vrste *C. gigas* određene su 2000 pb duge regije koje prethode i slijede svakom satDNA monomeru ili nizu lokaliziranom na sastavljenim kromosomima te vrste. Detektirane postrane regije izrezane su, zajedno s pripadajućim satelitnim nizom ili monomerom, uz pomoć interne skripte *Python* (Dodatak 1.) napisane za potrebe izrade ovog rada (autor dr. sc. T. Cvitanić). Izlazni dokument skripte *Python* bio je u formatu *fasta*. Skripta *Python* parser\_scaffolds.py (Dodatak 2.) napisana je za potrebe izrezivanja okolnih regija i satelitnih monomera ili nizova prisutnih u nesvrstanim segmentima.

### 3.16.4. Lokaliziranje regija sličnosti s elementima Helitron u izdvojenim postranim regijama

Nukleotidne sekvencije elemenata Helitron koji su nakon pretrage u bazi *Repbases* pokazale visok stupanj sličnosti s detektiranim satDNA vrste *C. gigas*, povučene su iz baze. Nakon sravnjivanja nukleotidnih sekvencija elemenata Helitron u programu *Geneious*, utvrđene su očuvane regije na krajevima elemenata Helitron koji pripadaju istoj skupini. Te su regije izdvojene i zatim anotirane na izdvojene sekvencije koje sadrže monomere ili nizove satDNA i njihove pripadajuće postrane regije, uz dopuštenu razliku od 30 % u odnosu na konsenzusnu sekvenciju. Za utvrđivanje prisutnosti očuvanih regija u postranim regijama, korištena je još jedna skripta *Python* (Dodatak 3.). Izlazni dokument te skripte *Python* bio je .csv dokument u kojem je prisutnost neke od konzerviranih regija u okolnoj regiji satelitnog niza ili monomera bila označena s 1, a odsutnost s 0. Na jednak način utvrđena je prisutnost konzerviranih regija u

postranim regijama izreanima iz *scaffolda* uz pomoć skripte *Python* naziva `boxer_scaffolds.py` (Dodatak 4.).



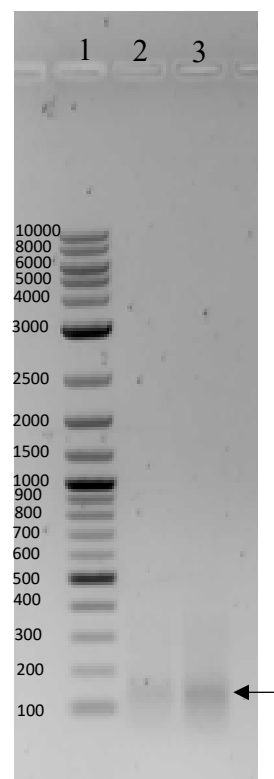


## **4. REZULTATI**



#### 4.1. Izolacija kromatina i razgradnja mikrokokalnom nukleazom

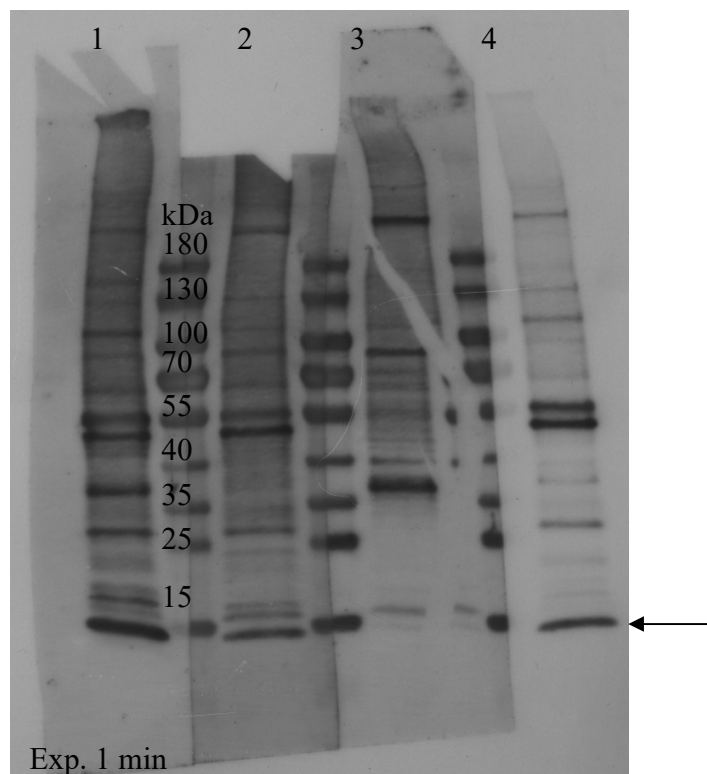
U pokusima izolacije i razgradnje kromatina nastojalo se dobiti kromatin razgrađen na veličinu mononukleosoma (~180 pb) jer je ta veličina optimalna za ovaj tip istraživanja. Kromatin pacifičke kamenice izoliran je u 3 serije izolacija, iz ukupno devet jedinki, a ukupna masa izoliranog kromatina iznosila je 22,99 µg. Kromatin izoliran iz različitih serija izolacija najprije je razgrađen mikrokokalnom nukleazom do veličine mononukleosoma (Slika 12.), a zatim su sve frakcije objedinjene u jedan uzorak koji je korišten za daljnje pokuse. Duljina fragmenata DNA u razgrađenom kromatinu provjerena je gel-elektroforezom (Slika 12.).



Slika 12. DNA detektirana etidij-bromidom nakon obrade kromatina proteinazom i razdvajanja agaroznom gel-elektroforezom. Na gelu su vidljive vrpce (označene strelicom) u kojima se nalaze fragmenti DNA veličine od oko 160 pb što odgovara veličini mononukleosoma. U prvu jažicu nanesen je marker veličina od 100 bp do 10 kbp, u drugu uzorak 10 ng DNA, a u treću uzorak 20 ng DNA.

#### 4.2. Testiranje antitijela anti-Cg-CenH3

Antitijela anti-Cg-CenH3, dobivena imunizacijom triju kunića, testirana su na ukupnom proteinskom izolatu tkiva škrge vrste *C. gigas*. Testirani su alikvoti seruma triju kunića uzetih prije imunizacije, te 60, 90 i 135 dana nakon početka imunizacije. Testirani su i afinitetno pročišćeni serumi, odnosno monospecifične frakcije antitijela IgG za protein Cg-CenH3. Na Slici 13. prikazani su rezultati testiranja seruma koji sadrži anti-Cg-CenH3 i njima pripadajućeg preimunoseruma uzetog iz kunića prije postupka imunizacije te frakcije IgG-a iz afinitetno pročišćenog imuniziranog seruma 135 dana nakon imunizacije. U daljnjim pokusima korišten je imunoserum kunića br. 2 te su na Slici 13. prikazani rezultati njegovog testiranja. U uzorku preimunoseruma nije detektirano antitijelo na Cg-CenH3 (uzorak 3, Slika 13.). U uzorcima seruma imuniziranih životinja detektirano je antitijelo za Cg-CenH3 (uzorci 1, 2 i 4).



Slika 13. Slika prikazuje rezultate testiranja seruma kunića metodom *western blotting*-a. Uzorak prikazuje koncentriranu frakciju monospecifičnog IgG-a izoliranog iz imuniziranog seruma kunića, uzorak 2 prikazuje nepročišćeni imunizirani serum kunića, uzorak 3 preimuni serum kunića, a uzorak 4 neugušćenu frakciju monospecifičnog IgG-a. Strelica pokazuje protein očekivane veličine 17 kDa.

#### 4.3. Imunoprecipitacija kromatina asociranog s proteinom CenH3

Kromatinskom imunoprecipitacijom izdvojeni su odsječci kromatina vrste *C. gigas* koji u svom sastavu sadrže protein CenH3. Postupak je detaljno opisan u Materijalima i metodama, potpoglavlje 3.6. Izmjerena je koncentracija DNA u imunoprecipitiranim eluatima uređajem *Qubit 3 Fluorometer* te su rezultati prikazani u Tablici 10. Kao pozitivna kontrola u eksperimentima kromatinske imunoprecipitacije korišteno je antijelo anti-H3K4me3, a kao negativna kontrola zečji IgG. Za svaku od reakcija imunoprecipitacije korišteno je ~ 4600 ng kromatina. U krajnjem desnom stupcu Tablice 10. prikazan je postotak eluiranog kromatina u

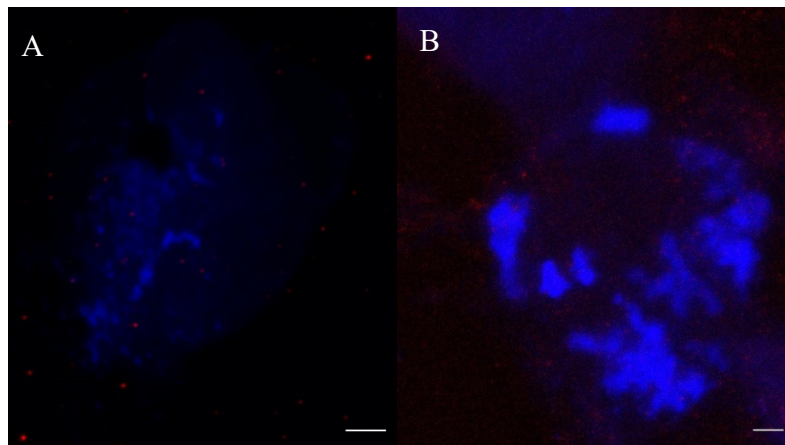
odnosu na početnu masu kromatina. U uzorku negativne kontrole nije detektirana DNA (uz granicu osjetljivosti mjernog uređaja od 0,05 ng/μl) zbog čega je zaključeno da nije došlo do nespecifičnog precipitiranja DNA u pokusu kromatinske imunoprecipitacije.

Tablica 10. U Tablici su prikazane izmjerene koncentracije DNA početnog ulaznog uzorka kromatina za imunoprecipitaciju (*input*), uzoraka kromatina imunoprecipitiranog antitijelom na protein Cg-CenH3 iz dvije biološke replike te pozitivne i negativne kontrole. Krajnji desni stupac prikazuje udio precipitirane DNA u odnosu na početnu masu kromatina.

uzorak	koncentracija	udio
<i>Input</i> (10 %)	30,1 ng/μl	32,72 %
CenH3	2,76 ng/μl	3,00 %
	2,45 ng/μl	2,66 %
H3K4me3	2,55 ng/μl	2,77 %
IgG	nemjerljivo niska koncentracija (< 0.010 ng/μl)	

#### 4.4. Lokalizacija proteina Cg-CenH3

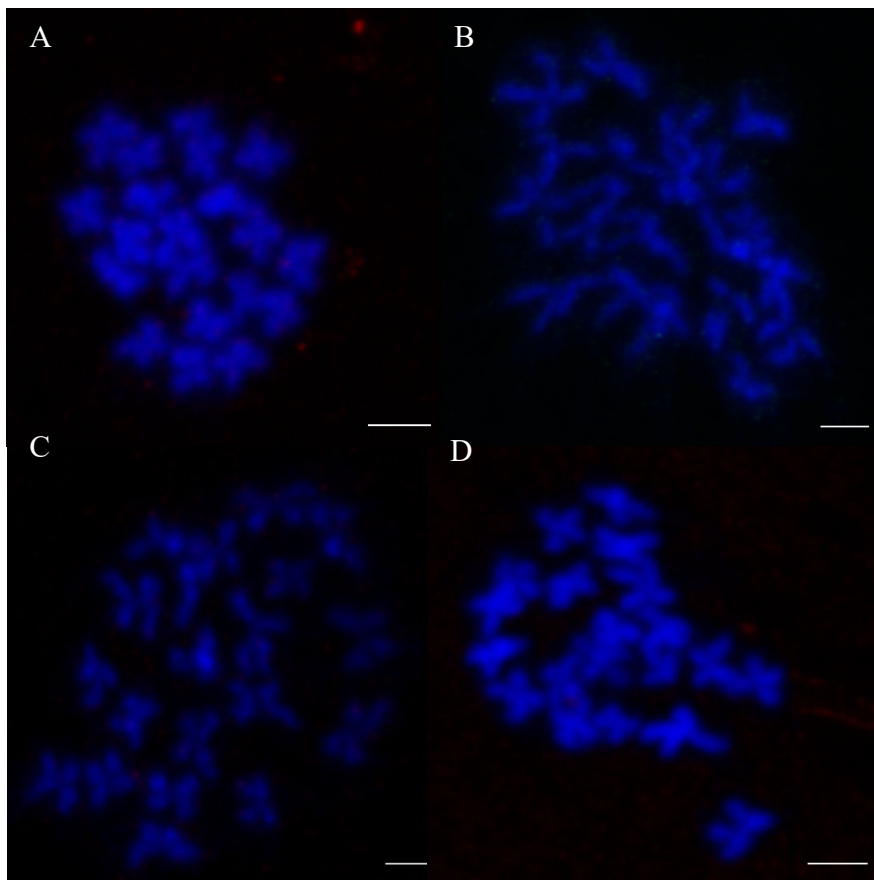
Sa ciljem lokalizacije proteina Cg-CenH3 pripremani su preparati metafaznih kromosoma na nekoliko različitih načina i variranjem različitih fiksativa (poglavlja 3.9. i 3.10. Materijala i metoda). Ovdje će biti prikazano samo nekoliko primjera. Na preparatima prilikom čije izrade je kao fiksativ korištena octena kiselina u niskom postotku zaostalo je dosta komadića tkiva i nije bilo prisutnih metafaznih ploča kromosoma kada su preparati bili pripremljeni iz tkiva škrge kao ni kada su bili pripremljeni iz tkiva gonada (Slika 14.).



Slika 14. Slika preparata pripremljenih s malim udjelom octene kiseline u fiksativu korištenom u izradi preparata. A) Tkivo škrge pripremljeno fiksativom sastavljenim od etanola i octene kiseline u omjeru 3:1. B) Preparati s tkivom gonada pripremljeni fiksativom sastavljenim od metanola i octene kiseline u omjeru 3:1. Skala predstavlja 5  $\mu\text{m}$ .

U protokolima u kojima je za izradu preparata metafaznih kromosoma korištena visokopostotna octena kiselina (60 %), nužna za dobivanje preparata očišćenih od ostataka tkiva, vezanje antitijela na protein CenH3 je izostalo (Slika 15A.). Poznato je da korištenje visokopostotne octene kiseline može uništiti proteinske epitope, što je i potvrđeno nakon što je na preparatima pripremljenima na jednak način, testirano i nekoliko komercijalno dostupnih antitijela (Slika 15. B, C i D).



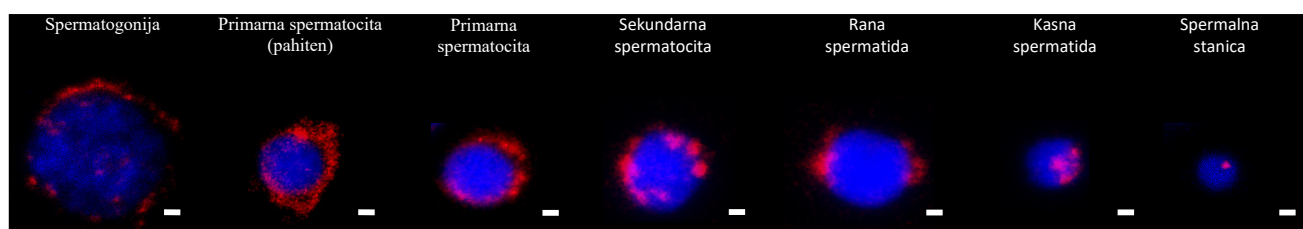


Slika 15. Slika preparata metafaznih kromosoma koji su pripremljeni fiksativom s visokim udjelom octene kiseline (60 %) koja je onemogućila vezanje komercijalnih antitijela i antitijela proizvedenog za potrebe laboratorija. Na svim preparatima vidljiv je slabi signal oko metafaznih ploča, bez jakih specifičnih signala po samim kromosomima. Korištena su sljedeća antitijela: A) Antitijelo na protein Cg-CenH3 detektirano antitijelom obilježenim fluoroforom *Alexa-594*. B) Antitijelo na protein H3K4me3 (Merck) detektirano antitijelom obilježenim fluorescein-izotiocijanatom. C) Antitijelo na protein H3K4me3 (Cell Signaling) detektirano antitijelom obilježenim fluoroforom *Alexa-594*. D) Antitijelo na protein H3 (Cell Signaling) detektirano antitijelom obilježenim fluoroforom *Alexa-594*. Skale predstavljaju 5  $\mu\text{m}$ .

Budući da visokopostotna octena kiselina prilikom pripreme preparata metafaznih kromosoma uništava proteinske epitope na koje se antitijela mogu vezati, u daljnjim pokusima

imunofluorescencijske detekcije koršiteni su preparati gonadnih stanica za čiju se pripremu ne koristi octena kiselina.

Vežanje antitijela na protein Cg-CenH3 pokazano je pokusima imunofluorescencijske detekcije na gonadnim stanicama *C. gigas* u različitim razvojnim fazama. Na Slici 16. vidljiva je lokalizacija proteina Cg-CenH3 na jezgrama sedam razvojnih faza muških spolnih stanica *C. gigas*. U stadiju spermatogonije protein Cg-CenH3 prisutan je po periferiji jezgre stanice. Moguće je primijetiti da se dio jezgre na kojem protein Cg-CenH3 lokalizira, mijenja sa sazrijevanjem stanice.



Slika 16. Imunofluorescencijska detekcija proteina Cg-CenH3 na različitim razvojnim fazama muških spolnih stanica vrste *C. gigas*. Skala predstavlja 1  $\mu\text{m}$ . Izvor slike: Tunjić Cvitanić i sur., 2020.

#### 4.5. Cjelokupni sastav ponavljajućih sekvencija DNA u genomu vrste *C. gigas*

Sa ciljem dobivanja referentne baze koja se može koristiti za usporedbu sastava sekvencija dobivenih imunoprecipitacijom kromatina koji sadrže protein CenH3 i modifikaciju H3K9me3, najprije je utvrđen cjelokupni sastav ponavljajućih sekvencija DNA u genomu vrste *C. gigas*. Iz uzorka izolirane cjelokupne genomske DNA dobivena su očitavanja sekvenciranjem NGS na platformi *Illumina*. Milijun nasumično odabranih i obrađenih sekvencija korišten je kao ulazni dokument za analizu *RepeatExplorer*.

Detektirane ponavljajuće DNA u genomu vrste *C. gigas* čine 20,69 % genoma (Tablica 11.), pri čemu su detektirane ponavljajuće DNA s udjelom u genomu  $> 0,01$  % zbog ograničenja metode. Analizom *RepeatExplorer* one su raspoređene u 83 856 klastera. Prema rezultatima analize najzastupljenije ponavljajuće sekvencije su satDNA s 2,56 %, od čega 1,76 % iznosi udio satDNA Cg170 koji predstavlja 68,75 % svih detektiranih satDNA. Zatim slijede PGE s 2,07 %

od čega su najbrojniji Ty3/Gypsy retrotranspozoni s 0,09 %. U genomu *C. gigas*, udio 45S rDNA iznosi 0,4 %. Najniži udio među klasificiranim ponavljajućim DNA zauzimaju mitochondrijalne ponavljajuće DNA, 0,02 %. Neklasificirano je ostalo 15,64 % detektiranih ponavljajućih sekvencija. Ovi su rezultati sumarno prikazani na Grafikonu 1.

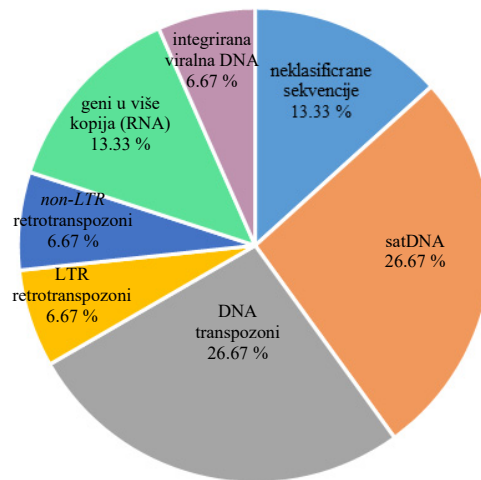
Tablica 11. Udio pojedinih tipova detektiranih ponavljajućih DNA u genomu vrste *C. gigas*.

Tip ponavljajuće DNA	Udio u genomu
Satelitne DNA	2.56
Cg170	1.76
Pokretni genetički elementi	2.07
Ty3 gypsy	0.09
DIRS	0.04
LINE	0.11
Penelope	0.04
45S rDNA	0.40
Mitochondrijska DNA	0.02
Neklasificirane	15.64

Razdvojeni kontizi koji su sačinjavali klastere detektiranih ponavljajućih sekvencija korišteni su kao referentne sekvencije za analizu *ChIP-Seq Mapper*.

#### 4.6. Sekvencije koje grade centromere u genomu vrste *C. gigas*

Sekvencije koje grade centromere u genomu vrste *C. gigas* utvrđene su modificiranom analizom *ChIP-Seq Mapper* koja je napravljena za obje biološke replike imunoprecipitacije kromatina vezanog s proteinom Cg-CenH3. Iz već navedenih razloga (Materijali i metode, sekcija 3.14.1.4.), kontizi koji sačinjavaju svaki od detektiranih klastera u *ChIP-Seq Mapper* analizi razdvojeni su na zasebne kontige te je u ponovljenoj analizi *ChIP-Seq Mapper* svaki kontig tretiran kao zasebni klaster. Rezultati za prvu biološku repliku prikazani su Tablici 12., a za drugu u Tablici 13. Grafikon 1. prikazuje kojim tipovima ponavljajućih sekvencija pripadaju one koje grade centromeru vrste *C. gigas*.



Grafikon 1. Udio pojedinih tipova ponavljajućih sekvencija koje grade centromeru vrste *C. gigas*.

Tablica 12. Prikaz dobivenih klastera u biološkoj replici 1 nakon analize očitavanja dobivenih iz sekvenciranja DNA iz kromatina imunoprecipitiranog antitijelom na protein Cg-CenH3. Prikazani su br. očitavanja u uzorku imunoprecipitiranog kromatina i ulaznog kromatina (*input*) te omjer brojeva očitavanja ChIP/Input, odnosno obogaćenje pojedinog klastera.

Ime klastera	Br. očitavanja (ChIP)	Br. očitavanja (Input)	Omjer Chip/Input
>CL1 7	617	361	1.7091
>CL1 48	555	454	1.2225
>CL1 108	9068	7317	1.2393
>CL3 95	17100	12360	1.3835
>CL4 33	1537	1245	1.2345
>CL4 100	841	652	1.2899
>CL5 20	1626	1353	1.2018
>CL9 15	1847	1452	1.2720
>CL10 10	801	594	1.3485
>CL12 44	501	341	1.4692
>CL14 18	1039	850	1.2224
>CL14 60	1002	671	1.4933
>CL14 51	708	482	1.4689
>CL14 75	570	355	1.6056
>CL15 23	1440	1070	1.3458
>CL16 69	755	434	1.7396
>CL22 54	5184	3803	1.3631
>CL22 68	906	645	1.4047
>CL22 9	949	790	1.2013
>CL31 15	1937	1590	1.2182
>CL33 1	15473	6780	2.2822
>CL35 39	1040	847	1.2279
>CL40 32	2560	1724	1.4849
>CL43 3	2080	1179	1.7642
>CL43 6	2616	1535	1.7042
>CL46 6	10275	7114	1.4443
>CL46 15	3105	2003	1.5502
>CL53 4	2486	1878	1.3237
>CL59 17	534	387	1.3798
>CL59 11	3503	2786	1.2574
>CL59 18	5792	4350	1.3315
>CL68 9	694	568	1.2218
>CL69 1	669	362	1.8481
>CL69 3	5587	2012	2.7768
>CL69 4	546	272	2.0074

>CL69 5	1334	635	2.1008
>CL69 6	8955	4119	2.1741
>CL69 9	567	256	2.2148
>CL69 10	3029	1307	2.3175
>CL72 21	2937	2201	1.3344
>CL73 3	889	601	1.4792
>CL73 22	589	444	1.3266
>CL77 6	1553	1183	1.3128
>CL77 22	516	376	1.3723
>CL83 9	959	642	1.4938
>CL86 2	671	462	1.4524
>CL90 17	839	660	1.2712
>CL94 25	3716	2408	1.5432
>CL102 1	762	628	1.2134
>CL103 8	999	793	1.2598
>CL103 15	659	472	1.3962
>CL106 17	2777	2233	1.2436
>CL109 11	723	478	1.5126
>CL109 12	723	580	1.2466
>CL111 27 24 sc 0.493648 1 1665	573	430	1.3326
>CL112 7	687	417	1.6475
>CL112 11	1136	662	1.7160
>CL112 4	850	652	1.3037
>CL112 12	542	373	1.4531
>CL114 10	918	706	1.3003
>CL116 10	704	425	1.6565
>CL118 3	558	378	1.4762
>CL118 8	51438	36383	1.4138
>CL121 7	936	596	1.5705
>CL121 11	1327	807	1.6444
>CL134 3	681	539	1.2635
>CL141 14	531	402	1.3209
>CL147 6	1582	1127	1.4037
>CL150 3	1147	897	1.2787
>CL150 11	2467	1196	2.0627
>CL152 27 1 sc 0.704409 1 743	1755	1260	1.3929
>CL173 4	2887	2280	1.2662
>CL178 14	1247	935	1.3337
>CL179 11	562	372	1.5108
>CL182 12	1544	1134	1.3616
>CL182 20	834	584	1.4281
>CL184 5	1173	677	1.7326
>CL195 13	734	493	1.4888

>CL198 6	523	405	1.2914
>CL202 1	683	567	1.2046
>CL204 7	3322	1987	1.6719
>CL204 10	597	452	1.3208
>CL231 3	1641	1116	1.4704
>CL232 9	722	586	1.2321
>CL245 5	1014	690	1.4696
>CL246 15 1 sc 0.000709167 1 3	30317	19925	1.5216
>CL246 11 8 sc 0.0124728 1 18	949	697	1.3615
>CL253 5	902	557	1.6194
>CL262 4	1884	1149	1.6397
>CL264 2	687	538	1.2770
>CL282 3	1848	732	2.5246
>CL282 5	1102	498	2.2129
>CL290 27 2 sc 0.414801 1 96	1042	525	1.9848
>CL296 8	885	731	1.2107
>CL298 4	4270	3555	1.2011
>CL308 7	2817	2320	1.2142
>CL310 2	700	473	1.4799
>CL310 3	620	492	1.2602
>CL341 27 3 sc 0.448146 1 45	1292	706	1.8300
>CL341 27 6 sc 0.228821 1 90	508	264	1.9242
>CL344 1	1218	653	1.8652
>CL347 4	860	624	1.3782
>CL347 12	1344	920	1.4609
>CL353 1	594	402	1.4776
>CL353 2	545	443	1.2302
>CL366 2	2298	1885	1.2191
>CL436 1	1878	899	2.0890
>CL442 1	1802	710	2.5380
>CL447 4	574	428	1.3411
>CL455 1	1292	614	2.1042
>CL458 1	694	334	2.0778
>CL460 1	785	332	2.3645
>CL485 1	1151	476	2.4181
>CL490 2	909	563	1.6146
>CL499 2	1484	694	2.1383
>CL504 4	1792	1188	1.5084
>CL548 1	811	486	1.6687
>CL561 1	570	432	1.3194
>CL562 1	1977	795	2.4868
>CL605 2	575	265	2.1698
>CL653 1	705	313	2.2524

>CL698 1	570	347	1.6427
>CL712 3	658	444	1.4820
>CL721 3	616	430	1.4326
>CL783 1	533	366	1.4563
>CL779 1	590	308	1.9156
>CL908 2	1629	1257	1.2959
>CL930 1	620	264	2.3485
>CL961 1	1237	614	2.0147
>CL1070 1	1097	703	1.5605
>CL1142 1	547	327	1.6728
>CL1191 1	5435	3436	1.5818



Tablica 13. Prikaz dobivenih klastera u biološkoj replici 2 nakon analize očitavanja dobivenih sekvenciranjem DNA iz kromatina imunoprecipitiranog antitijelom na protein CenH3. Prikazani su br. očitavanja u uzorku imunoprecipitiranog kromatina i ulzanog kromatina (*Input*) te omjer tih očitavanja, odnosno obogaćenje pojedinog klastera.

Ime klastera	Br. očitavanja (ChIP)	Br. očitavanja ( <i>Input</i> )	Omjer ChIP/ <i>Input</i>
>CL1 98	1183	955	1.23874
>CL3 95	9664	7342	1.31626
>CL4 33	1160	863	1.34415
>CL4 100	789	555	1.42162
>CL5 16	924	691	1.33719
>CL7 67	1159	912	1.27083
>CL10 56	718	544	1.31985
>CL11 7	598	439	1.36219
>CL11 11	511	417	1.22542
>CL14 18	552	390	1.41538
>CL14 60	521	370	1.40811
>CL15 23	937	664	1.41114
>CL16 69	665	319	2.08464
>CL22 54	2925	2185	1.33867
>CL22 68	906	523	1.73231
>CL33 1	12499	3585	3.48647
>CL39 27 2 sc 0.506491 1 892	1572	1199	1.31109
>CL40 32	3772	2922	1.29090
>CL41 28	1294	853	1.51700
>CL43 3	1876	776	2.41753
>CL43 6	2569	800	3.21125
>CL46 6	4032	2721	1.48181
>CL46 15	2497	1278	1.95383
>CL58 18	1026	780	1.31538
>CL59 1	1402	840	1.66905
>CL59 17	615	335	1.83582
>CL69 1	1018	275	3.70182
>CL69 3	4351	1172	3.71246
>CL69 4	536	190	2.82105
>CL69 5	949	310	3.06129
>CL69 6	12328	4561	2.70292
>CL69 9	1806	700	2.58000
>CL69 10	2969	1190	2.49496
>CL72 21	1975	1523	1.29678
>CL73 3	595	446	1.33408

>CL73 22	619	497	1.24547
>CL77 6	907	717	1.26499
>CL86 2	678	389	1.74293
>CL90 17	824	650	1.26769
>CL103 8	629	447	1.40716
>CL106 17	2024	1464	1.38251
>CL109 11	531	309	1.71845
>CL109 12	771	441	1.74830
>CL111 27 24 sc 0.493648 1 1665	595	470	1.26596
>CL112 7	539	181	2.97790
>CL112 10	539	333	1.61862
>CL112 11	657	331	1.98489
>CL114 10	932	709	1.31453
>CL116 10	519	299	1.73579
>CL118 3	518	421	1.23040
>CL118 4	598	454	1.31718
>CL118 8	63474	40375	1.57211
>CL121 7	558	373	1.49598
>CL121 11	1144	630	1.81587
>CL144 1	539	398	1.35427
>CL150 3	1511	1158	1.30484
>CL150 11	1216	460	2.64348
>CL152 27 1 sc 0.704409 1 743	1293	744	1.73790
>CL184 5	799	382	2.09162
>CL202 1	959	754	1.27188
>CL204 7	3434	2680	1.28134
>CL231 3	1291	681	1.89574
>CL246 15 1 sc 0.000709167 1 3	23145	13523	1.71153
>CL253 5	667	324	2.05864
>CL262 4	1740	1286	1.35303
>CL264 2	555	421	1.31829
>CL282 3	1469	180	8.16111
>CL282 5	960	235	4.08511
>CL290 27 2 sc 0.414801 1 96	1965	1081	1.81776
>CL344 1	1063	446	2.38341
>CL353 1	507	354	1.43220
>CL436 1	1623	949	1.71022
>CL455 1	1489	316	4.71203
>CL460 1	1396	300	4.65333
>CL485 1	722	258	2.79845
>CL490 2	881	305	2.88852
>CL504 4	1710	1164	1.46907
>CL553 1	617	355	1.73803

>CL562 1	671	323	2.07740
>CL605 2	824	407	2.02457
>CL698 1	827	367	2.25341
>CL783 1	651	363	1.79339
>CL798 1	603	276	2.18478
>CL971 1	502	277	1.81227
>CL1070 1	876	463	1.89201
>CL1201 1	556	255	2.18039

Nakon provedene analize *ChIP-Seq Mapper* s promijenjenim postavkama, odnosno razdvojenim kontizima, kontizi s obogaćenjem > 1,7 i brojem pojavljivanja u imunoprecipitiranom uzorku > 500 smatrali su se centromerno obogaćenim sekvencijama. Kriteriji su primijenjeni za obje biološke replike reakcije imunoprecipitacije s antitijelom na Cg-CenH3. Sekvencije koje su zadovoljile kriterije u obje replike i njihove karakteristike prikazane su u Tablici 14. Dominantno su zastupljene ponavljajuće sekvencije klasificirane kao DNA transpozoni i satDNA (svaka skupina sa po 26,67 %). Zatim slijede retrotranspozoni s 13,34 % (LTR i *non-LTR* retrotranspozoni s podjednakim udjelom) te ribosomalni geni s 13,33 % i integrirane viralne sekvencije DNA sa 6,67 %. Neidentificirano je ostalo 13,33 % sekvencija (Grafikon 1.).

Tablica 14. Sekvencije obogaćene u obje biološke replike kromatina imunoprecipitiranog antitijelom na protein Cg-CenH3. Navedena su obogaćenja u obje biološke replike, klasifikacija prema bazi *Repbse* te klasifikacija prema analizi *RepeatExplorer*.

Kontig	ChIP 1 obogaćenje	ChIP 2 obogaćenje	Klasifikacija prema bazi <i>Repbse</i>		Klasifikacija prema analizi <i>RepeatExplorer</i>
>CL16_69	2,08	1,74	Helitron-N55 CGi	DNA transposon	ostalo
>CL33_1	3,49	2,28	hAT-N1B_Crp	DNA transpozon	-
>CL43_3	2,42	1,76	LSU-rRNA_Mfr	gen u više kopija (RNA)	-
>CL43_6	3,21	1,70			
>CL69_1	3,70	1,85	LSU-rRNA_Hsa	gen u više kopija (RNA)	-
>CL69_3	3,71	2,78			
>CL69_4	2,82	2,01	LTR-18B_OS-LTR	LTR retrotranspozon	
>CL69_5	3,06	2,10	Cre-1_MB	<i>non-LTR</i> retrotranspozon	
>CL69_6	2,70	2,17	LSU-	gen u više	

>CL69_9	2,58	2,21	rRNA_Hsa	kopija (RNA)	
>CL69_10	2,49	2,32			
>CL112_7	2,98	1,65	Helitron-N35_CGi	DNA transpozon	ostalo
>CL112_11	1,98	1,72			
>CL150_11	2,64	2,06	<i>root-related</i> ponavljanje	-	-
>CL184_5	2,09	1,73	<i>root-related</i> ponavljanje	-	ostalo
>CL282_3	8,16	2,52	<i>root-related</i> ponavljanje	-	satDNA (niska vjerojatnost)
>CL282_5	4,09	2,21			
>CL290_27_2_sc_0.414801_1_96	1,82	1,98	<i>root-related</i> ponavljanje	satDNA	satDNA (visoka vjerojatnost)
>CL344_1	2,38	1,87	DNAV-1e_LVa	integrirani virus	satDNA (niska vjerojatnost)
>CL436_1	1,71	2,09	SAT2_CM	satDNA	-
>CL455_1	4,71	2,10	<i>root-related</i> ponavljanje	-	satDNA (niska vjerojatnost)
>CL460_1	4,65	2,36	<i>root-related</i> ponavljanje	-	satDNA (niska vjerojatnost)
>CL485_1	2,80	2,42	DNAV-1e_LVa	integrirani virus	-
>CL562_1	2,08	2,49	Polinton-1_XT	DNA transpozon	-

Za ponavljanje okarakterizirano kao *root-related* te onu označenu -, izostala je klasifikacija.

#### 4.7. Karakterizacija najzastupljenijih centromerno obogaćenih sekvencija u genomu vrste *C. gigas*

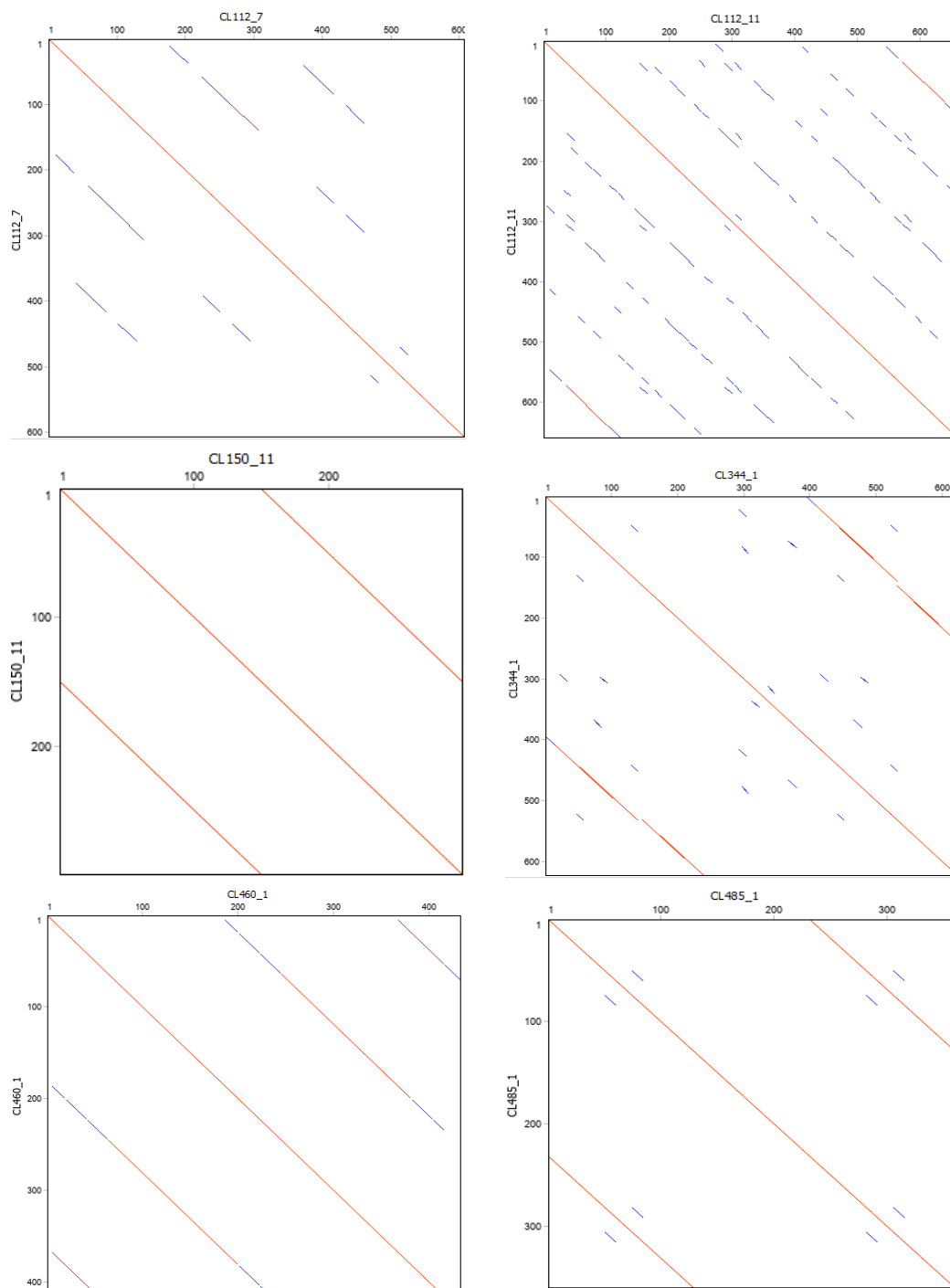
Centromerno obogaćenje pokazala su 24 kontiga raspodijeljenih u 15 različitih klastera. Svaki od kontiga analiziran je programom *Geneious* za regije autohomologije (Slika 17.) koje su naznaka uzastopno ponavljajuće DNA. Autohomologija prisutna je u svim klasterima, osim u njih šest (CL43\_3, CL43\_6, CL69\_1, CL69\_3, CL69\_4, CL69\_5). Analizom BLASTN u bazi NCBI provjerena je prisutnost kontiga na kromosomima/vezanim grupama (od engl. linkage groups) i nesmjještenim regijama (od engl. unplaced scaffolds), te postojanje regija uzastopne ponovljenosti u genomu. Na svih 10 kromosoma/vezanih grupa smješteno je ukupno devet klastera (CL16\_69, 33\_1, 112\_7, 112\_1, 184\_5, 282\_3, 282\_5, 455\_1, 460\_1), među kojima se CL16\_69 nalazi i na 40 nesmjještenih regija, CL282\_3 i 282\_5 na njih sedam, CL184\_5 na pet, CL112\_7 na četiri, CL33\_1 na jednom, a CL112\_1, 455\_1 i 460\_1 ni na jednoj regiji. CL344\_1 nalazi se na sedam vezanih grupa i 14 nesmjještenih regija. Na šest vezanih grupa nalaze se CL43\_3 i 43\_6, a 43\_3 se nalazi i na jednoj regiji. CL69\_3 nalazi se na pet vezanih grupa i nijednoj regiji. Na četiri vezane grupe nalaze se CL69\_1, 69\_4, 69\_6 i 69\_9, a među njima se 69\_1, 69\_6 i 69\_1 nalaze i na po jednoj nesmjještenoj regiji. CL150\_1 nalazi se na tri vezane grupe i nijednoj nesmjještenoj regiji. CL69\_5, 69\_10 i 562\_1 nalaze se na po dvije vezane grupe i nijednoj nesmjještenoj regiji. Na po jednoj vezanoj grupi nalaze se CL290\_27\_2\_sc\_0.414801\_1\_96, 436\_1 i 485\_1. Ti se klasteri ne nalaze ni na jednoj nesmjještenoj grupi. Tablica 15. pokazuje sumirane rezultate analize za svaki od analiziranih kontiga.

Tablica 15. Tablica prikazuje karakteristike utvrđene za svaki od analiziranih klastera iz obje replike imunoprecipitacije kromatina s antitijelom za Cg-CenH3. U Tablici je naznačena prisutnost autohomologije u sekvenciji klastera (da/ne), prisutnost na kromosomima/vezanim grupama ili nesmjještenim regijama te postojanje uzastopno ponovljenih sekvencija klastera u genomu (da/ne).

IME KONTIGA	DULJINA/pb	AUTOHOMOLOGIJA	PRISUTNOST NA KROMOSOMIMA/NESMJEŠTENIM REGIJAMA	UZASTOPNA PONOVLJENOST U GENOMU
>CL16_69 (kontig 69 klastera 16)	400	da	da (10/10; 40)	ne
>CL33_1	5455	da	da (10/10; 1)	ne
>CL43_3	1860	ne	da (6/10; 1)	ne
>CL43_6	2250	ne	da (6/10)	ne
>CL69_1	809	ne	da (4/10; 1)	ne
>CL69_3	921	ne	da (5/10)	da
>CL69_4	635	ne	da (4/10)	ne
>CL69_5	473	ne	da (2/10)	ne
>CL69_6	1291	da	da (4/10; 1)	ne
>CL69_9	349	da	da (4/10, 1)	ne
>CL69_10	532	da	da (2/10)	ne
>CL112_7	604	da	da (10/10; 4)	da
>CL112_11	656	da	da (10,10)	da
>CL150_11	297	da	da (3/10)	da
>CL184_5	772	da	da (10/10; 5)	da
>CL282_3	590	da	da (10/10; 7)	da
>CL282_5	357	da	da (10/10; 7)	da
>CL290_27_2 sc 0.414801_1_96	384	da	da (1/10)	ne
>CL344_1	620	da	da (7/10; 14)	da
>CL436_1	426	da	da (1/10)	da
>CL455_1	327	da	da (10/10)	da

>CL460 1	431	da	da (10/10)	da
>CL485 1	358	da	da (1/10)	da
>CL562 1	550	da	da (2/10)	da

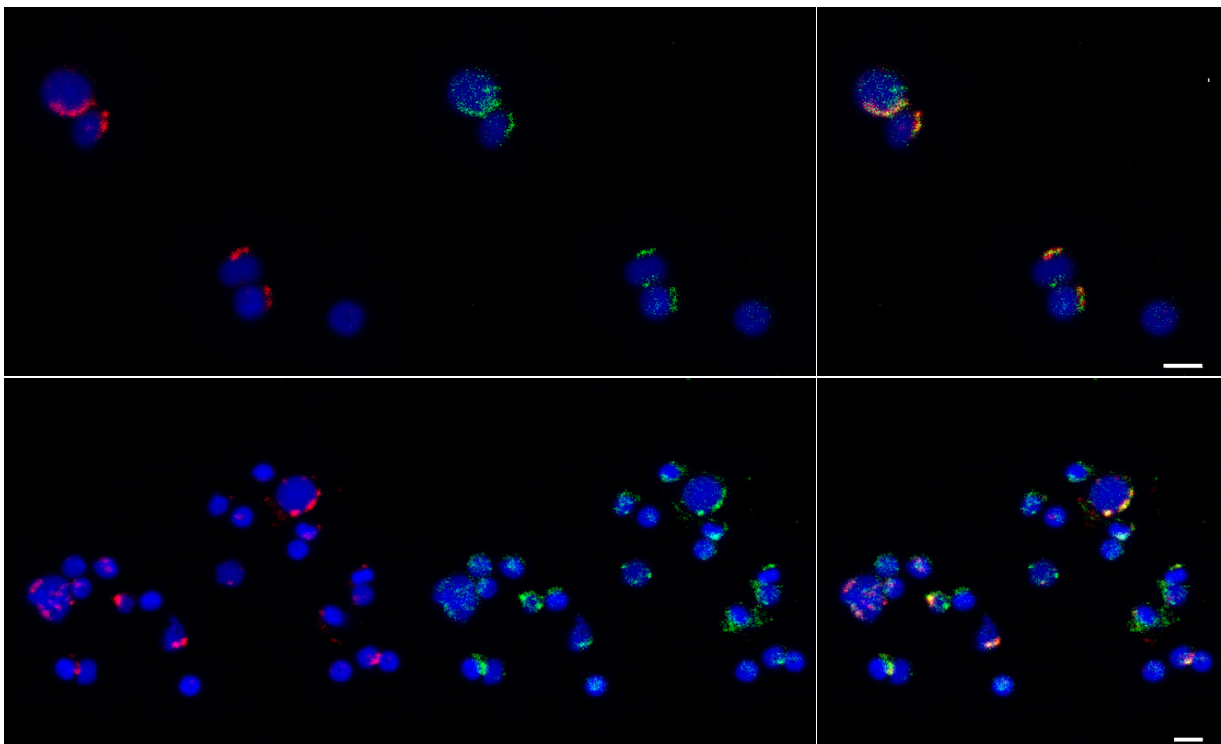




Slika 17. Slika prikazuje grafove autohomologije, tzv. *Dot-plot* grafove, za pet klastera koji su odabrani za pokuse lokalizacije na metafaznim kromosomima pacifičke kamenice. Grafovi prikazuju redom: CL112\_7, CL112\_11, CL150\_11, CL344\_1, CL460\_1, CL485\_1.

#### 4.7.1. Kolokalizacija ukupne imunoprecipitirane DNA i proteina Cg-CenH3 na jezgrama gonadnih stanica vrste *C. gigas*

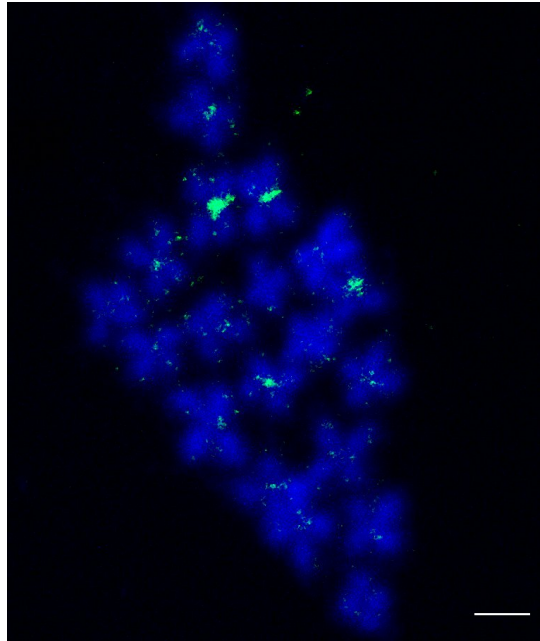
Zbog nemogućnosti izvedbe pokusa imunofluorescencijske detekcije na metafaznim kromosomima iz ranije objašnjenih razloga, kombinacija metoda FISH i IF napravljena je na preparatima muških spolnih stanica u različitim razvojnim fazama. Time se indirektno potvrdila kolokalizacija proteina Cg-CenH3 i sekvencija imunoprecipitiranih s antitijelom na protein Cg-CenH3. Na Slici 18. vidljivo je djelomično preklapanje signala proteina Cg-CenH3 i signala sekvencija DNA kromatina imunoprecipitiranog antitijelom na protein Cg-CenH3. To je moguća posljedica nepostojanja sekvencija koje su centromerno specifične te su one lokalizirane i po ostatku kromosoma.



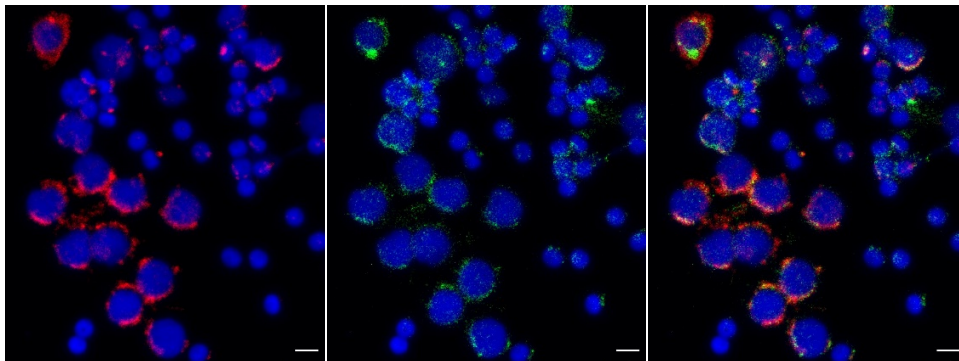
Slika 18. Kolokalizacija proteina Cg-CenH3 i sekvencija DNA imunoprecipitiranih antitijelom na protein Cg-CenH3 na jezgrama muških spolnih stanica u različitim razvojnim fazama. Crveno (fluorokrom Alexa 594) je detektiran protein Cg-CenH3, a zeleno (fluorokrom Alexa 488) imunoprecipitirane sekvencije DNA. Na jezgrama muških spolnih stanica vidljiva su područja

preklapanja crvenog i zelenog signala, što odgovara centromernim područjima na kromosomima, ali i područja gdje je vidljiv samo zeleni signal što odgovara područjima raspršeno raspoređenih imunoprecipitiranih sekvencija i izvan centromernog područja. Skala predstavlja 3  $\mu$ m. Izvor slike: Tunjić Cvitanić i sur., 2020.

4.7.2. Lokalizacija centromerne satDNA Cg170 na metafaznim kromosomima vrste *C. gigas* i njezina kolokalizacija s proteinom Cg-CenH3 na jezgrama spolnih stanica Wang i sur. (2001.) su na kromosomima vrste *C. gigas* prema morfologiji kromosoma i signalima FISH-a zaključili da satDNA Cg170 sjeda na područja primarne konstrikcije nekoliko parova kromosoma. U ovom je radu potvrđen zaključak Wanga i sur. (2001.) što je vidljivo na Slici 19. Signali dobiveni FISH-em nalaze se u centromernim područjima nekoliko kromosoma, ali i raspršeni duž krakova kromosoma. Nadalje, u spolnim stanicama *C. gigas* kolokalizirana je satDNA Cg170 s proteinom Cg-CenH3. Na Slici 20. vidljivo je da postoje mjesta gdje se signali porijeklom od proteina Cg-CenH3 i satDNA Cg170 preklapaju, što bi odgovaralo područjima centrokromatina. Na spolnim stanicama vidljiva su i područja gdje nema preklapanja satDNA i proteina. Ta područja odgovarala bi raspršeno raspoređenim sekvencijama satDNA Cg170 duž krakova kromosoma.



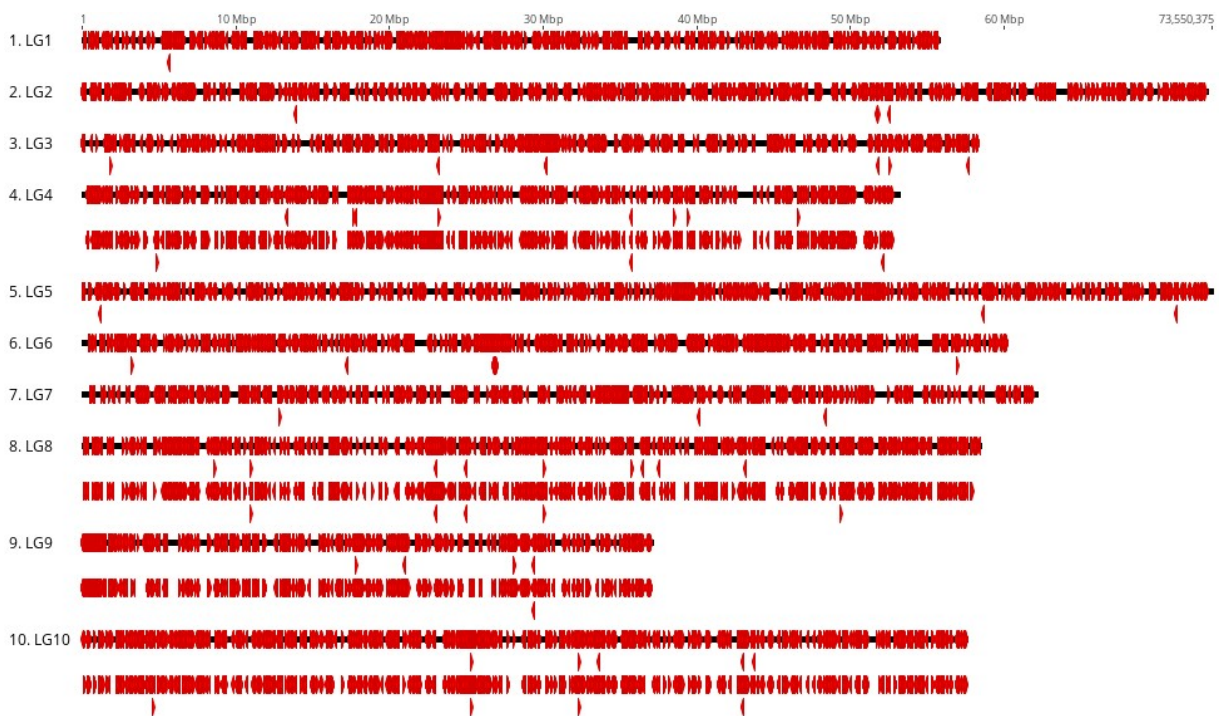
Slika 19. Lokalizacija sekvencija satDNA Cg170 na metafaznim kromosomima vrste *C. gigas*. Skala predstavlja 5  $\mu$ m. Izvor slike: Tunjić Cvitanić i sur., 2020.



Slika 20. Kombinacijom IF-a i FISH-a pokazano je da se Cg-CenH3 i Cg170 u velikim područjima jezgara muških spolnih stanica *C. gigas* međusobno preklapaju, ali postoje i područja gdje je Cg170 lokaliziran samostalno, bez preklapanja s proteinom Cg-CenH3. Skala predstavlja 5  $\mu$ m. Izvor slike: Tunjić Cvitanić i sur., 2020.

Značajan raspršeni smještaj satDNA Cg170 potvrđen je i lokalizacijom te satDNA na kromosomima *C. gigas in silico* (Slika 21.). Pretragom baze Repbase utvrđeno je da satDNA Cg170 odgovara središnjim ponavljanjima elemenata Helitron. Uzastopna ponavljanja satDNA Cg170 čine središnji dio čak 12 elemenata Helitron (Tablica 16.). Značajna raspšerenost te satDNA po kromosomima vjerojatno je posljedica njezine povezanosti s PGE Helitronom.

Među obogaćenim sekvencijama imunoprecipitiranim antitijelom na protein Cg-CenH3 nalazi se i sekvencija klastera CL16\_69 (Tablica 15.; opširnija karakterizacija u dijelu 3.2.1.). Ta sekvencija pripada dijelu sekvencije koja omeđuje središnja ponavljanja u elementu Helitron-N55\_CGi.



Slika 21. Na Slici je vidljiv smještaj satDNA Cg170 na kromosomima (vezanim grupama) vrste *C. gigas in silico*.

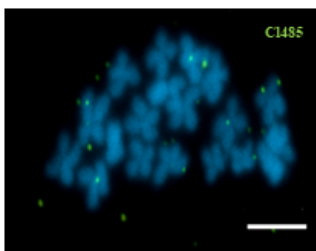
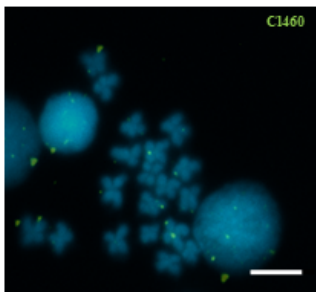
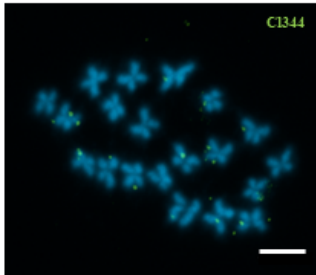
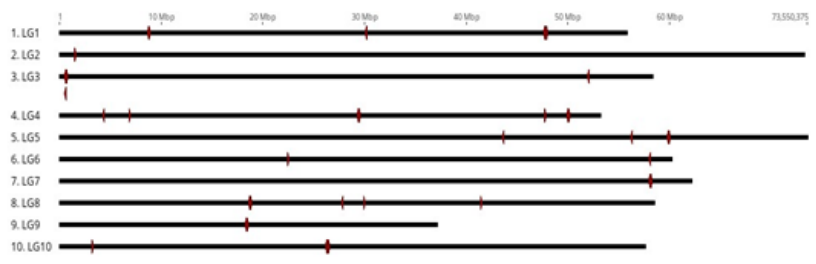
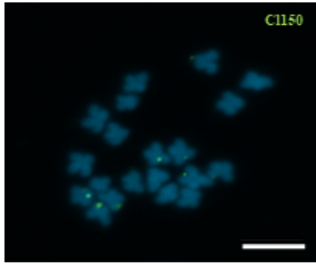
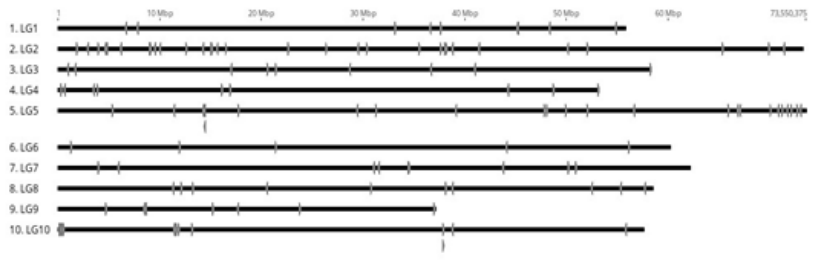
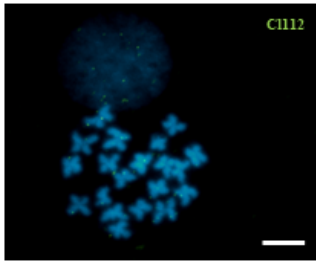
Tablica 16. Elementi s kojima sekvencija satDNA Cg170 pokazuje sličnost te podaci o unosu u bazu *Repbase* za svaki pojedini element.

Element	Unos u bazu <i>Repbase</i>
Helitron-N1B CGi	Repbase Reports 13(1), 525-525 (2013)
Helitron-N1C CGi	Repbase Reports 13(1), 526-526 (2013)
Helitron-N2 CGi	Repbase Reports 13(1), 530-530 (2013)
Helitron-N2c CGi	Repbase Reports 13(1), 529-529 (2013)
Helitron-N2d CGi	Repbase Reports 13(1), 531-531 (2013)
Helitron-N2e CGi	Repbase Reports 13(1), 532-532 (2013)
Helitron-N2f CGi	Repbase Reports 13(1), 533-533 (2013)
Helitron-N2g CGi	Repbase Reports 13(1), 534-534 (2013)
Helitron-N14 CGi	Repbase Reports 13(1), 552-552 (2013)
Helitron-N37 CGi	Repbase Reports 13(1), 575-575 (2013)
Helitron-N39 CGi	Repbase Reports 13(3), 1377-1377 (2013)
Helitron-N55 CGi	Repbase Reports 13(3), 1234-1234 (2013)

#### 4.7.3. Lokalizacija obogaćenih ponavljajućih sekvencija DNA detektiranih u centromernoj frakciji imunoprecipitiranog kromatina vrste *C. gigas*

Velik dio sekvencija obogaćenih u kromatinu imunoprecipitiranom antitijelom na Cg-CenH3 pripada različitim skupinama raspršeno ponavljajućih DNA čiji je smještaj potencijalno i izvan centromernih regija kromosoma *C. gigas*. Nekoliko kandidata iz Tablice 15. izabrano je za pokus lokalizacije na metafaznim kromosomima *C. gigas*. Pri odabiru kandidata, u obzir su uzeti njihovo obogaćenje i/ili klasifikacija, s prednošću danoj satDNA i PGE Helitron koji sadrže uzastopna središnja ponavljanja zbog toga što je te sekvencije, zbog duljine niza, potencijalno lakše vizualizirati na kromosomima metodom FISH. Prema navedenim kriterijima sekvencije klastera C1112, C1150, C1344, C1460 i C1485 izabrane su za pokuse vizualizacije na kromosomima. C1460 pokazao je visok stupanj sličnosti sa C1282 i C1455 (otprilike 80 % sličnosti) te su zbog toga ta tri klastera tretirana kao jedan (C1460). Početnice za izradu sonda sintetizirane su prema konsenzusu svih monomernih ponavljanja unutar pojedinog kontiga (Tablica 15.). Položaj početnica na konsenzusnoj sekvenciji monomera vidljiv je na Slici 12. Hibridizacijski signal bio je prisutan u centromeri nekoliko kromosoma svih pet odabranih kandidata uz povremeni raspršeni smještaj duž krakova kromosoma (Slika 22.). Na Slici 22. prikazan je i smještaj sekvencija monomera svakog od odabranih klastera na kromosomima *C.*

*gigas in silico*. Sekvencije klastera CL112 i CL150 pokazuju smještaj na svih deset kromosoma, CL344 i CL485 na po jednom kromosomu, a CL460 na tri kromosoma prema analizi *in silico*. Takav smještaj razlikuje se od onog vidljivog na metafaznim kromosomima detektiranog metodom FISH. Vjerojatno je ta razlika posljedica nepotpune posloženosti genoma, odnosno kromosoma/vezanih grupa na kojima je rađena analiza *in silico*. Genom pod oznakom GCA\_902806645.1 u bazi NCBI trenutačno je referentni genom za vrstu *C. gigas* i sadrži velik broj nesmjještenih *scaffold*-a na kojima se potencijalno nalazi još sekvencija klastera CL344 i CL485.

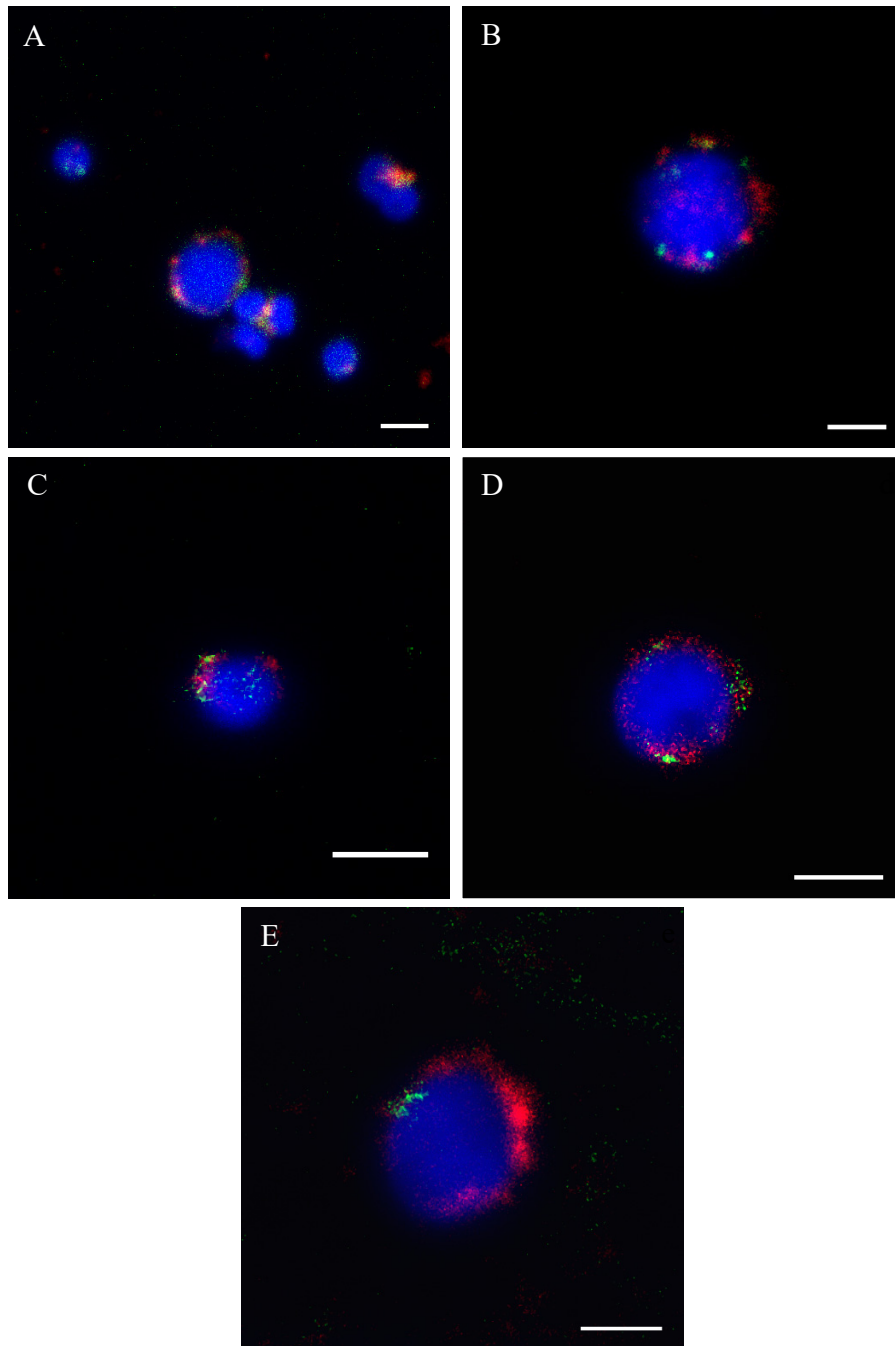




Slika 22. Slika prikazuje lokalizaciju sekvencija centromerno obogaćenih klastera na metafaznim kromosomima *C. gigas* (lijevi dio slike) i na kromosomima (vezanim grupama) *in silico* (desni dio slike). Izvor slike: Tunjić Cvitanić i sur., 2020.

#### 4.8. Kolokalizacija ponavljajućih sekvencija DNA obogaćenih u centromernoj frakciji vrste *C. gigas* s proteinom Cg-CenH3

Budući da centromerna lokalizacija odabranih pet kandidata klastera nije mogla biti potvrđena na preparatima metafaznih kromosoma, napravljen je pokus njihove kolokalizacije s proteinom Cg-CenH3 kombinacijom IF-a i FISH-a na preparatima spolnih stanica. Kako je vidljivo sa Slike 23., svaka od odabranih sekvencija DNA u određenim regijama jezgre spolne stanice kolokalizira s proteinom Cg-CenH3 što potvrđuje njihov smještaj i u području centromere kromosoma.



Slika 23. Kolokalizacija proteina Cg-CenH3 (crveno) i DNA sekvencija (zeleno) centromerno obogaćenih klastera kombinacijom IF-a i FISH-a na gonadnim stanicama *C. gigas*: C1112 (A), C1150 (B), C1344 (C), C1460 (D) i C1485 (E). Skala predstavlja 3  $\mu$ m. Izvor slike: Tunjić Cvitanić i sur., 2020.

#### 4.9. Sekvencije DNA koje izgrađuju heterokromatin vrste *C. gigas*

Za potrebe utvrđivanja sastava heterokromatinske frakcije genoma *C. gigas*, napravljena je kromatinska imunoprecipitacija izoliranog kromatina *C. gigas* monoklonalnim komercijalnim antitijelom na H3K9me3. Ta je epigenetička modifikacija oznaka transkripcijski utišanog heterokromatina. Za određivanje sekvencija obogaćenih u kromatinu s oznakom H3K9me3, primijenjeni su isti kriteriji kao i za odabir sekvencija obogaćenih u centromernom heterokromatinu (omjer ChIP/Input > 1,7; broj pojavljivanja u imunoprecipitiranom uzorku > 500). Kontizi obogaćeni u kromatinu imunoprecipitiranom antitijelom na protein H3K9me3 te njihova klasifikacija i obilježja prikazani su u Tablici 17.

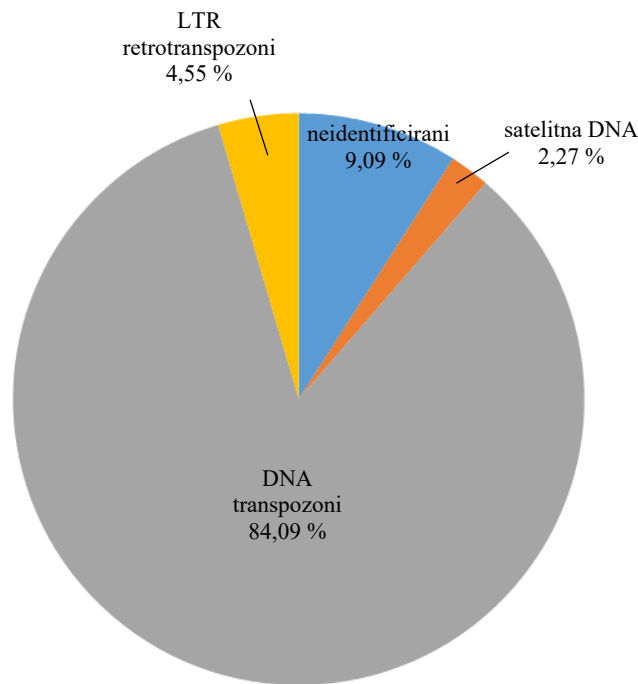
Tablica 17. U Tablici su prikazani kontizi obogaćeni u frakciji kromatina *C. gigas* imunoprecipitiranom antitijelom na H3K9me3. Uz ime kontiga i stupanj obogaćenja, navedene su njihove klasifikacije prema bazi *Repbases* i analizi *RepeatExplorer*.

Klaster/kontig	Obogaćenje (ChIP/Input)	Klasifikacija prema <i>Repbases</i> -u		Karakterizacija prema <i>RepeatExplorer</i> -u
>CL5_51	1,92	root-related repeat	DNA transpozon	satelit (niska vjerojatnost)
>CL5_77	1,87	Helitron-N29_Cgi		
>CL16_57	1,85	Helitron-N1B_CGi	DNA transpozon	ostalo
>CL18_24	2,00	DNA2-8_CGi	DNA transpozon	ostalo
>CL18_71	1,90	Helitron-N7C_CGi	DNA transpozon	ostalo
>CL20_45	1,74	Helitron-N17_CGi	DNA transpozon	ostalo
>CL24_19	1,95	Helitron-N51_Cgi	DNA transpozon	ostalo
>CL24_32	1,71	root-related repeat		
>CL25_58	1,80	IS3EU-2_SK	DNA transpozon	ostalo
>CL28_38	1,91	DNA3-1_CGi	DNA transpozon	ostalo
>CL29_4	2,04	DNA8-10_Cgi Helitron-N6	DNA transpozon	ostalo
>CL29_7	2,34			
>CL29_20	2,28			
>CL29_28	1,73			
>CL29_52	1,88			
>CL37_13	1,79	DNA4-5_CGi	DNA transpozon	ostalo
>CL37_23	2,63			
>CL42_31	1,70	Helitron-N6_CGi	DNA transpozon	ostalo
>CL48_13	1,77	Kolobok-N3_CGi	DNA transpozon	ostalo
>CL59_10	1,95	DNA6-9	DNA transpozon	ostalo

>CL59_13	2,44			
>CL67_13	1,72	DNA4-8_CGi	DNA transpozon	DNA transpozon
>CL80_1	1,95	Kolobok-N4_CGi	DNA transpozon	ostalo
>CL94_22	1,72	TE-5_CGi	DNA transpozon	ostalo
>CL105_23	2,31	Helitron-N15	DNA transpozon	ostalo
>CL120_7	2,08	root-related repeat	N/A	ostalo
>CL122_3	2,02	Gypsy-3 PAb-I	LTR transpozon	ostalo
>CL148_19	1,78	DNA4-32_Cgi	DNA transpozon	ostalo
>CL149_10	2,06	Helitron-N2d_CGi	DNA transpozon	0.23 % organel/plastid
>CL158_1	1,78	DNA6-10_CGi	DNA transpozon	ostalo
>CL159_7	2,22	DNA6-10_Cgi	DNA transpozon	ostalo
>CL159_9	1,84			
>CL188_5	1,80	Helitron-N6_Cgi	DNA transpozon	ostalo
>CL190_5	2,08	DNA6-5_CGi	DNA transpozon	ostalo
>CL195_10	2,21	root-related repeat	N/A	ostalo
>CL199_18	1,74	DNA6-11_CGi	DNA transpozon	ostalo
>CL199_26	1,91			
>CL236_2	2,06	DNA4-39_CGi	DNA transpozon	ostalo
>CL239_5	2,13	DNA6-13_CGi	DNA transpozon	ostalo
>CL246_11_3_sc_0.00184917_1_7	1,80	root-related repeat	satelitna DNA	satelit (visoka vjerojatnost)
>CL246_11_6_sc_0.00286439_1_11	1,92			
>CL246_11_7_sc_0.0108049_1_15	2,48			
>CL246_11_8_sc_0.0124728_1_18	1,82			

>CL246_15_2_sc_0.000783816_1_8	3,21			
>CL246_15_4_sc_0.00216483_1_11	2,67			
>CL246_15_5_sc_0.00261272_1_13	2,76			
>CL246_15_6_sc_0.00694237_1_16	2,04			
>CL246_15_8_sc_0.0135488_1_47	1,97			
>CL246_19_1_sc_0.00415321_1_20	2,01			
>CL257_3	1,91	DNA8-2B_CGi	DNA transpozón	ostalo
>CL294_5	1,85	ISL2EU-7	DNA transpozón	ostalo
>CL358_1	1,99	Sola3-1_CGi	DNA transpozón	ostalo
>CL367_2	1,75	DNA3-12_CGi	DNA transpozón	N/A
>CL385_1	1,74	DNA6-5_CGi	DNA transpozón	N/A
>CL385_3	2,52			
>CL385_5	2,10			
>CL438_1	2,34	DNA4-4C_CGi	DNA transpozón	N/A
>CL462_1	2,03	DNA3-11_CGi	DNA transpozón	N/A
>CL579_1	2,38	Gypsy-13_ALY-I	LTR retrotranspozón	N/A
>CL611_4	3,60	DNA6-13B_CGi	DNA transpozón	N/A
>CL680_1	2,84	Harbinger-5_DR	DNA transpozón	N/A
>CL700_3	3,35	IS3EU-N2_CGi	DNA transpozón	N/A
>CL772_1	3,22	root-related repeat	N/A	N/A
>CL1282_1	2,98	root-related repeat	N/A	N/A

Najbrojnija grupa ponavljajućih sekvencija DNA u heterokromatinu s oznakom H3K9me3 jesu DNA transpozoni s 84,09 %, dok satDNA čine 2,27 %. LTR retrotranspozoni čine 4,55 % ukupnih sekvencija, a 9 % sekvencija ostalo je neidentificirano. Sastav ponavljajućih sekvencija DNA heterokromatina *C. gigas* grafički je prikazan na Grafu 2.

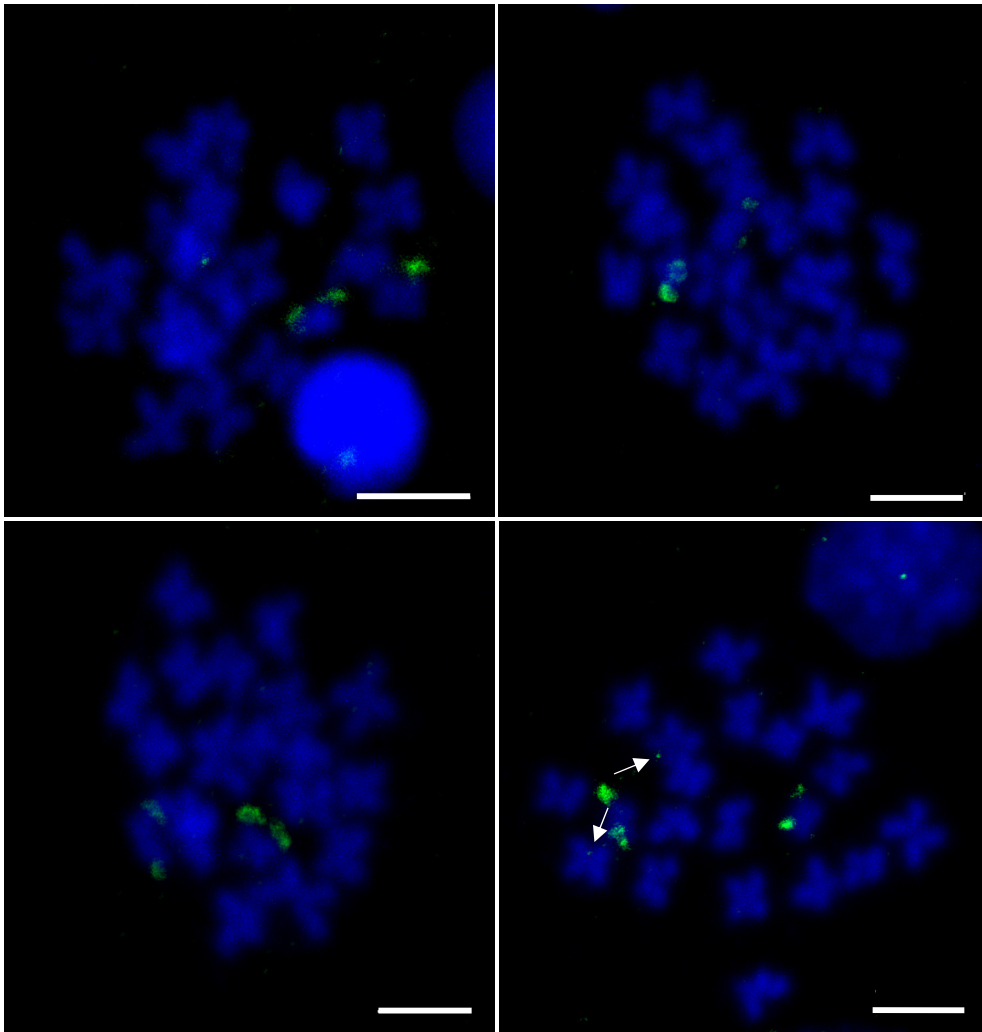


Grafikon 2. Grafički prikaz sastava ponavljajućih sekvencija DNA u heterokromatinu vrste *C. gigas* imunoprecipitiranom komercijalnim monoklinalnim antitijelom na H3K9me3.

#### 4.9.1. Lokalizacija sekvencija iz imunoprecipitiranog heterokromatina na metafaznim kromosomima i jezgrama gonadnih stanica vrste *C. gigas* te kolokalizacija proteina Cg-CenH3 i H3K9me3 na gonadnim stanicama vrste *C. gigas*

Za lokalizaciju DNA sekvencija iz kromatina imunoprecipitiranog antitijelom na H3K9me3, napravljena je hibridizacijska sonda DNA koristeći sveukupnu DNA iz imunoprecipitiranog kromatina. Sonda je označena biotinom označenim dUTP-om i detektirana antitijelom označenim fluoroforom fluorescein izotiocijanatom. Lokalizacijom DNA iz kromatina imunoprecipitiranog antitijelom za protein H3K9me3 detektirani su jaki signali na telomernim regijama najmanjeg para kromosoma i slabi (peri)centromerni signali na još jednom paru

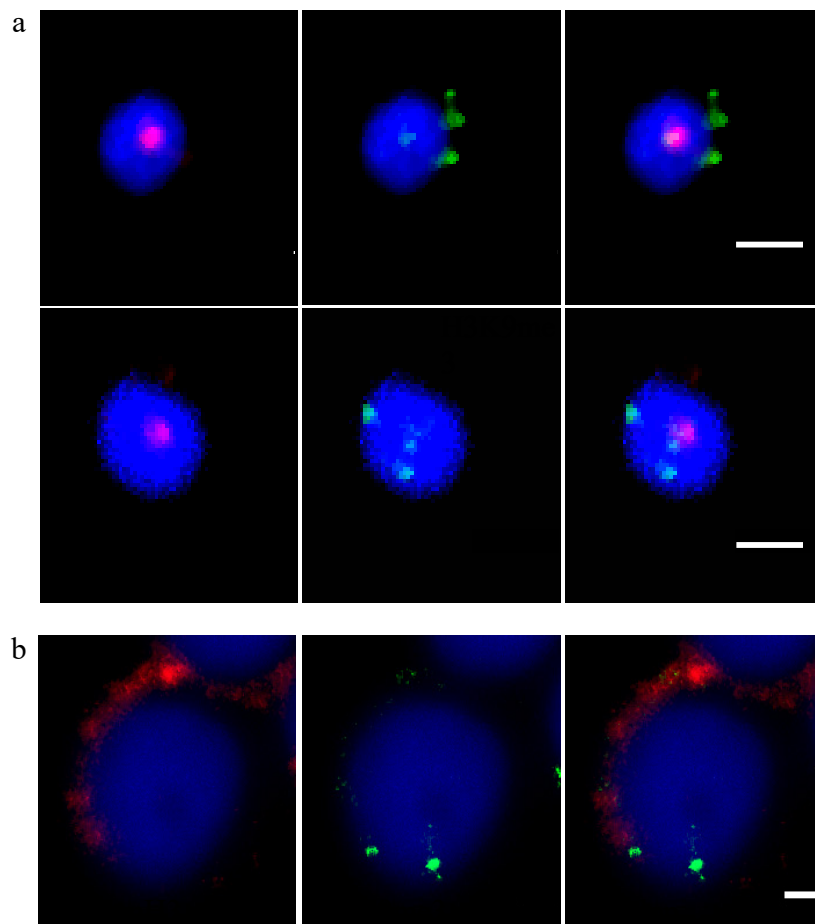
kromosoma (Slika 24.). Ovim pokusom potvrđeno je da razmještaj heterokromatina odgovara onom utvrđenom od strane Bouilly i sur. (2008.) metodom C-pruganja. Autori su utvrdili smještaj heterokromatina u (peri)centromernoj regiji 5. para kromosoma i telomernoj regiji 10. para kromosoma (Slika 5.).



Slika 24. Lokalizacija DNA iz kromatina imunoprecipitiranog antitijelom na H3K9me3 na metafaznim kromosomima vrste *C. gigas* metodom FISH-a. Strelice pokazuju lokalizaciju heterokromatina u pericentromernom području jednog kromosomskog para. Na još jednom paru kromosoma vidi se snažan signal u telomernim regijama. Skala predstavlja 5  $\mu$ m. Izvor slike: Tunjić Cvitanić i sur., 2020.



Nadalje, ispitana je međusobna pozicija centrokromatina i heterokromatina Cg-CenH3 i H3K9me3. Pokusom dvostruke imunofluorescencijske detekcije na jezgrama muških gonadnih stanica, kolokalizirani protein Cg-CenH3 (crveno) i histonska modifikacija H3K9me3 (zeleno). Na Slici 25. vidljiva je kolokalizacija u području slabog (peri)centromernog signala H3K9me3 koji odgovara slabom (peri)centromernom signalu uočenom na metafaznim kromosomima (Slika 24.), dok preklapanje signala u području jakog telomernog signala H3K9me3 očekivano izostaje.



Slika 25. Kolokalizacija kromatina obilježenog histonom Cg-CenH3 i heterokromatina obilježenog modificiranim histonom H3K9me3 na zrelim spermijima (a) i spermatociti (b) vrste *C. gigas* metodom dvostruke IF. Skala predstavlja 1  $\mu$ m. Izvor slike: Tunjić Cvitanić i sur., 2020.

#### 4.10. Analiza satelitoma pacifičke kamenice *C. gigas*


##### 4.10.1. Sastav satDNA u genomu pacifičke kamenice *C. gigas*

Nakon analize pojedinih genomskih frakcija koje sadrže sekvencije ponavljajuće DNA, istražili smo ukupni sastav satDNA u genomu *C. gigas*, satelitom. On je određen analizom sekvencija dobivenih sekvenciranjem NGS, a sekvencije su korištene kao ulazni podaci za analizu TAREAN. Rezultati analize TAREAN monomerne su konsenzus sekvencije satDNA s podacima o udjelu klastera u genomu, grafičkim izgledom klastera te određenom najsličnijom ponavljajućom DNA (primjer rezultata vidljiv je na Slici 26.). Napravljene su ukupno četiri analize TAREAN s različitim parametrima (klica po kojoj su birane ishodišne NGS sekvencije i broj sekvencija za analizu). Monomeri satelitnih DNA dobiveni četirima različitim analizama TAREAN međusobno su uspoređeni programom Geneious kako bi se odredilo koji monomeri porijeklom iz različitih analiza TAREAN zapravo pripadaju istoj satDNA. U genomu *C. gigas* na taj je način uvtrđeno postojanje seta od 52 satDNA (Tablica 18.) s ukupnim udjelom u genomu od 6,33 %. SatDNA detektirana u svakoj od četiri analize TAREAN navedena je u tablici u Dodatku 5.

Tablica 18. Set od 52 satDNA koji je analizama TAREAN detektiran u genomu *C. gigas*. U Tablici su navedena glavna obilježja detektiranih satDNA.

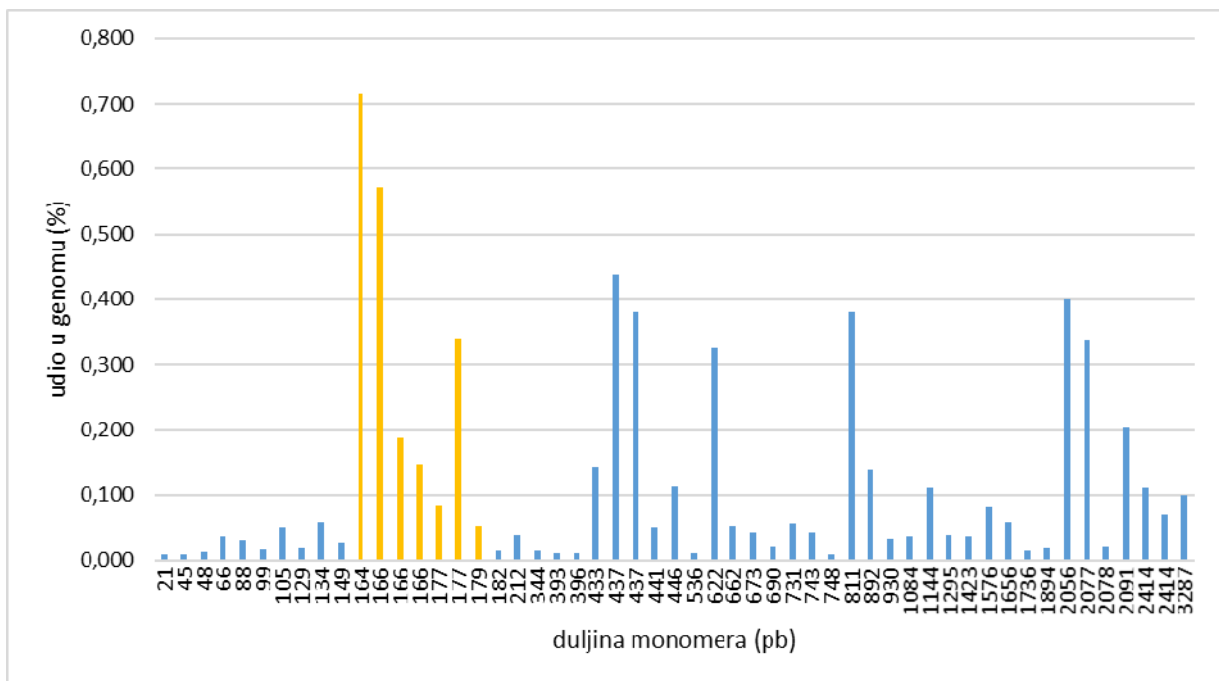
satDNA	duljina monomera (pb)	prosječni udio u genomu (%)	udio u satelitemu (%)	u udio nukleotida A i T (%)	sličnost s ponavljajućim sekvencijama <i>Rebase</i> baze
<b>CgiSat01a</b>	164	0,715	11,294	59,8	DNA/Helitron
<b>CgiSat01b</b>	166	0,573	9,043	57,2	DNA/Helitron
<b>CgiSat02</b>	437	0,438	6,911	65,7	DNA/Helitron
<b>CgiSat03</b>	437	0,380	6,002	67,5	DNA/Helitron
<b>CgiSat04</b>	2077	0,337	5,318	67,3	DNA/Helitron
<b>CgiSat05</b>	811	0,380	6,002	67,6	DNA/Helitron
<b>CgiSat06</b>	622	0,325	5,134	67,4	DNA/Helitron
<b>CgiSat07</b>	2091	0,205	3,238	67	DNA/Helitron
<b>CgiSat08</b>	166	0,188	2,962	67,5	DNA/Helitron
<b>CgiSat09</b>	166	0,148	2,330	63,3	DNA/Helitron
<b>CgiSat10</b>	433	0,143	2,264	63	DNA/Helitron
<b>CgiSat11</b>	892	0,140	2,211	64,6	LTR/BEL
<b>CgiSat12</b>	446	0,113	1,777	64,8	-
<b>CgiSat13</b>	2414	0,110	1,738	67,4	DNA/Kolobok
<b>CgiSat14</b>	1144	0,110	1,738	68,9	DNA/Helitron
<b>CgiSat15</b>	177	0,084	1,319	68,9	-
<b>CgiSat16</b>	2414	0,071	1,122	66,8	DNA/Kolobok
<b>CgiSat17</b>	134	0,058	0,921	76,1	DNA/Helitron
<b>CgiSat18</b>	731	0,055	0,864	63,9	Interspersed Repeat
<b>CgiSat19</b>	212	0,038	0,596	72,6	-
<b>CgiSat20</b>	662	0,051	0,810	65,4	-
<b>CgiSat21</b>	441	0,049	0,774	69,2	DNA/Helitron
<b>CgiSat22</b>	743	0,042	0,659	53,3	LTR/Gypsy
<b>CgiSat23</b>	1295	0,038	0,600	64,9	DNA/MuDR
<b>CgiSat24</b>	66	0,037	0,577	57,6	-
<b>CgiSat25</b>	88	0,031	0,495	61,4	NonLTR/R1
<b>CgiSat26</b>	1894	0,019	0,300	60,4	LTR/Gypsy
<b>CgiSat27</b>	99	0,017	0,261	60,6	
<b>CgiSat28</b>	182	0,016	0,253	44	DNA/MuDR
<b>CgiSat29</b>	344	0,015	0,229	52,9	DNA/DNA4-44
<b>CgiSat30</b>	48	0,014	0,221	68,7	-
<b>CgiSat31</b>	393	0,011	0,179	52,2	IntegratedVirus/DNAV

<b>CgiSat32</b>	536	0,011	0,166	66,6	DNA/Mariner
<b>CgiSat33</b>	105	0,051	0,798	46,7	-
<b>CgiSat34</b>	1423	0,036	0,569	67,3	DNA/DNA3-8
<b>CgiSat35</b>	690	0,021	0,324	68,8	DNA/IS3EU
<b>CgiSat36</b>	45	0,010	0,163	60	-
<b>CgiSat37</b>	177	0,340	5,371	56,5	DNA/Helitron
<b>CgiSat38</b>	3287	0,100	1,580	68	DNA/DNA2-7
<b>CgiSat39</b>	179	0,052	0,821	52	-
<b>CgiSat40</b>	673	0,041	0,648	62,9	DNA/Crypton
<b>CgiSat41</b>	1736	0,015	0,237	66,5	DNA/Ginger1
<b>CgiSat42</b>	930	0,033	0,521	66,9	DNA/DNA4-31
<b>CgiSat43</b>	21	0,010	0,158	66,7	-
<b>CgiSat44</b>	748	0,010	0,158	68,6	LTR/DIRS
<b>CgiSat45</b>	1576	0,083	1,311	64,1	DNA/DNA4-2
<b>CgiSat46</b>	149	0,026	0,411	54,4	-
<b>CgiSat47</b>	129	0,019	0,300	69,8	DNA/IS3EU
<b>CgiSat48</b>	2056	0,400	6,318	64,8	DNA/Helitron
<b>CgiSat49</b>	1084	0,036	0,569	69,4	DNA/DNA2-12
<b>CgiSat50</b>	2078	0,021	0,332	64,1	-
<b>CgiSat51</b>	396	0,012	0,190	63,1	LTR/Gypsy
<b>CgiSat52</b>	1656	0,058	0,916	61,5	DNA/Polinton

klaster	udio u genomu				duljina monomera	monomerna sekvencija	grafički prikaz					anotirana sličnost				
30	<a href="#">30</a>	0.180	0.180	1832	0.02030	166	<pre> TAGCCAAAATTTTCAAACAAATTTTCGTCAAGATTTCTCAGCAACTATTTATCGCAGATGETTGAAAATTTTACACAGT ATTTGTTTAGGCATGCCATATCGTGGGATATATTTTGTACCAATCGGACGTCACCTTCTGTTAAATGACGACTTTGTT TATTTT </pre>		<a href="#">report</a>	0.718	0.618	0.478	1832	329000	0.0000	0.05% Class_I/LTR/Bel-Pao:BEL-INT 0.05% Class_II/Subclass_1/TIR/Academ:Academ-TPase

Slika 26. Primjer rezultata analize TAREAN. Za svaki detektirani klaster u analizi u izlaznom dokumentu, među ostalima, generiraju se podaci o njegovom udjelu u genomu, duljina konsenzusa monomera, monomerna sekvencija, grafički izgled klastera te se klasteru odredi tip najbližnje ponavljajuće DNA.

Udio pojedine satDNA varira između 0,01 % (donja granica za detekciju ponavljajuće DNA u analizama programa *RepeatExplorer*) i 0,715 %. Udio nukleotida A i T u satDNA iznosi od 44 do 76,1 %. Duljina najkraćeg monomera satDNA iznosi 21 pb (CgiSat43), a duljina najdulje iznosi 3287 pb (CgiSat38). SatDNA s duljinom monomera od 160 do 180 pb (narančasti stupci na Grafikonu 3.) dominiraju te im je ukupni udio u satelitemu 33 %.

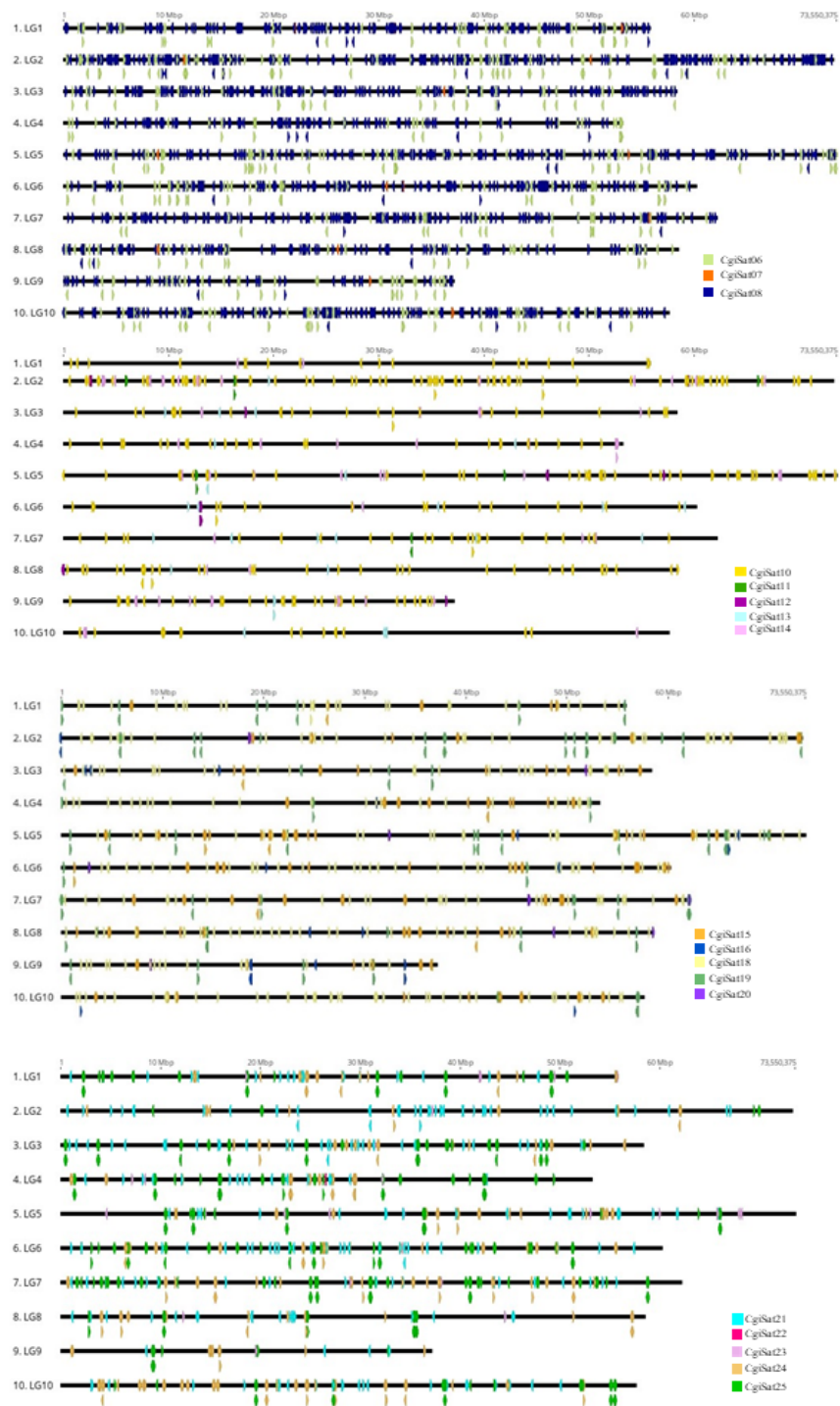


Grafikon 3. Grafički prikaz udjela satDNA u odnosu na duljinu monomerne sekvencije u genomu *C. gigas*. Monomeri s duljinom od 160 pb do 180 pb (narančasto označeni na grafu) dominiraju s ukupnim udjelom u genomu od 33 %.

#### 4.10.2. Anotacija *in silico* satDNA u genomu *C. gigas*

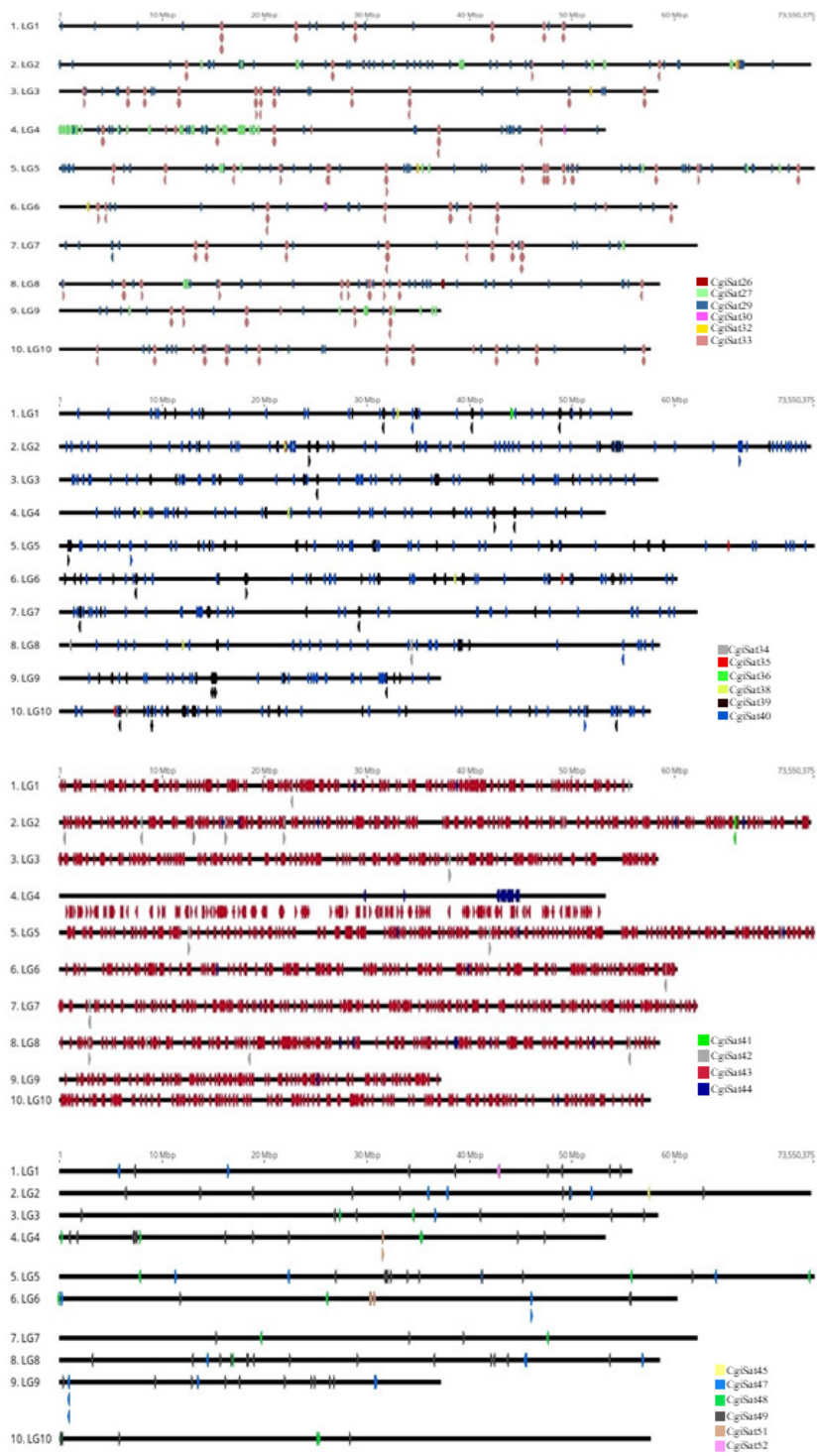
Ukupni satelitem vrste *C. gigas* lokaliziran je na dva posložena genoma te vrste, dostupna u bazi NCBI. Prvi genom, pod oznakom GCA\_902806645.1 (Peñaloza i sur., 2021.), složen je u 10 vezanih grupa (LG1-LG10) i 226 nesmjještenih regija te se smatra trenutno referentnim genomom za *C. gigas*. Drugi genom, pod oznakom GCA\_011032805.1, složen je u 10 cjelovitih kromosoma.

Detektirane satDNA anotirane na kromosome *in silico*, pokazuju međusobnu različitost u kromosomskom smještaju, rasprostranjenosti i monomernoj organizaciji po kromosomima. Na Slici 27. vidljiv je kromosomski smještaj, rasprostranjenost i monomerna organizacija za sve detektirane satDNA osim za: CgiSat01 koja je prikazana na Slici 21. te CgiSat17, 28, 31 i 46 prikazane na Slici 22. koje odgovaraju redom klasterima C1112, C1460, C1344 i C1485 te C1150. SatDNA CgiSat02, 03, 04, 05, 09 i 37 bit će prikazane kasnije na Slici 31., zajedno sa smještajem na metafaznim kromosomima.



Slika 27. Prikaz anotacije *in silico* satDNA detektiranih u genomu vrste *C. gigas* na kromosome referentnog genoma GCA 902806645.1.





Slika 27. Nastavak.

Za svaku satDNA, utvrđena je pojavnost na svakom pojedinom kromosomu te ukupan broj monomera na svim kromosomima i svim *scaffold*-ima (za genom GCA\_902806645.1). Podaci su prikazani u Tablici 19.

Na referentnom genomu vrste *C. gigas* od ukupno 52 satDNA, njih 25 smješteno je na svih 10 kromosoma, sedam ih je smješteno na samo jednom kromosomu, po jedna satDNA je smještena na šest, pet i tri kromosoma, pet satDNA je smješteno na sedam kromosoma, tri su smještene na četiri kromosoma, a dvije na dva kromosoma. Za usporedbu, na drugom posloženom genomu iz baze NCBI, od ukupno 52 satDNA, njih 27 je smješteno na svih 10 kromosoma, a na samo jednom kromosomu šest satDNA. Na devet i sedam kromosoma smještene su dvije satDNA, na osam i šest i dva kromosoma smještene su tri satDNA, a na četiri kromosoma jedna satDNA. SatDNA CgiSat50 nije detektirana ni na jednom kromosomu ni u jednom dostupnom genomu niti na *scaffold*-ima. Analizom te satDNA provjerom sličnosti sa sekvencijama u bazi NCBI utvrđeno je da se radi o sekvenciji mitohondrijske rDNA, čija se prisutnost na kromosomima niti ne očekuje.

Tablica 19. Prikaz pojavnosti monomera satDNA na pojedinoj vezanoj grupi/kromosomu te broj monomera na svakoj vezanoj grupi/kromosomu i nesmjještenim *scaffold*-ima na dva posložena genoma dostupna u bazi NCBI.

GENOM: GCA\_902806645.1 (Peñaloza i sur., 2021.)

GENOM: GCA\_011032805.1

satDNA	kromosom										br. monomera na kromosomima	br. monomera na nesmjještenim <i>scaffold</i> -ima	kromosom										br. monomera na kromosomima
	1	2	3	4	5	6	7	8	9	10			1	2	3	4	5	6	7	8	9	10	
CgiSat01	+	+	+	+	+	+	+	+	+	+	28120	3946	+	+	+	+	+	+	+	+	+	28486	
CgiSat02	+	+	+	+	+	+	+	+	+	+	6499	1249	+	+	+	+	+	+	+	+	+	5563	
CgiSat03	+	+	+	+	+	+	+	+	+	+	2502	411	+	+	+	+	+	+	+	+	+	2116	
CgiSat04	+	+	+	+	+	+	+	+	+	+	223	32	+	+	+	+	+	+	+	+	+	198	
CgiSat05	+	+	+	+	+	+	+	+	+	+	1811	312	+	+	+	+	+	+	+	+	+	1553	
CgiSat06	+	+	+	+	+	+	+	+	+	+	1628	202	+	+	+	+	+	+	+	+	+	1574	
CgiSat07	+	+	+	-	+	+	+	+	+	+	55	11	+	+	+	+	+	+	+	+	+	48	
CgiSat08	+	+	+	+	+	+	+	+	+	+	5171	1030	+	+	+	+	+	+	+	+	+	4949	
CgiSat09	+	+	+	+	+	+	+	+	+	+	4411	406	+	+	+	+	+	+	+	+	+	4609	
CgiSat10	+	+	+	+	+	+	+	+	+	+	1082	328	+	+	+	+	+	+	+	+	+	1463	
CgiSat11	-	+	-	-	+	-	+	-	-	-	83	50	+	-	+	+	+	-	+	-	-	150	
CgiSat12	-	+	+	-	+	+	-	+	+	-	244	21	+	+	+	+	+	-	+	+	-	672	
CgiSat13	-	+	+	+	+	+	+	+	+	+	39	2	+	+	-	+	+	-	+	+	+	39	
CgiSat14	+	+	+	+	+	+	+	+	+	+	441	119	+	+	+	+	+	+	+	+	+	257	
CgiSat15	+	+	+	+	+	+	+	+	+	+	1768	163	+	+	+	+	+	+	+	+	+	1897	
CgiSat16	-	+	+	+	+	+	+	+	+	+	66	1	+	+	+	+	-	+	-	-	+	34	
CgiSat17	+	+	+	+	+	+	+	+	+	+	1619	309	+	+	+	+	+	+	+	+	+	1153	
CgiSat18	+	+	+	+	+	+	+	+	+	+	363	25	+	+	+	+	+	+	+	+	+	372	
CgiSat19	+	+	+	+	+	+	+	+	+	+	947	352	+	+	+	+	+	+	+	+	+	1304	
CgiSat20	-	+	+	-	+	+	+	+	+	-	148	54	+	-	-	-	+	+	+	+	-	245	
CgiSat21	+	+	+	+	+	+	+	+	+	+	542	109	+	+	+	+	+	+	+	+	+	505	
CgiSat22	-	-	-	+	-	-	-	-	-	-	45	0	-	-	-	-	-	-	+	-	-	31	
CgiSat23	+	+	+	+	+	+	+	+	+	-	78	14	+	+	+	+	+	+	+	+	+	148	

CgiSat24	+	+	+	+	+	+	+	+	+	+	2740	128	+	+	+	+	+	+	+	+	+	2829	
CgiSat25	+	+	+	+	+	+	+	+	+	+	2468	67	+	+	+	+	+	+	+	+	+	2983	
CgiSat26	-	-	-	-	-	-	-	+	-	-	13	0	-	-	-	+	-	-	-	-	-	15	
CgiSat27	-	+	+	+	+	-	+	+	+	-	2030	57	+	-	+	+	+	-	+	+	-	+	1554
CgiSat28	+	+	+	+	+	+	+	+	+	+	487	167	+	+	+	+	+	+	+	+	+	548	
CgiSat29	+	+	+	+	+	+	+	+	+	+	277	19	+	+	+	+	+	+	+	+	+	267	
CgiSat30	-	-	-	+	-	+	-	-	-	-	2	1	-	-	-	-	+	-	-	+	-	4	
CgiSat31	+	-	-	-	-	-	-	-	-	-	52	0	-	-	-	-	-	+	-	-	-	61	
CgiSat32	-	+	+	-	+	+	-	-	-	-	147	0	+	-	-	-	-	-	+	-	-	75	
CgiSat33	-	+	+	+	+	+	+	+	+	+	3191	254	-	+	-	+	+	+	+	+	+	2789	
CgiSat34	-	+	-	-	-	-	-	+	+	+	12	2	+	+	-	+	+	+	+	+	+	43	
CgiSat35	-	+	-	-	+	+	-	-	-	+	6	0	-	-	+	-	-	-	-	-	-	4	
CgiSat36	+	-	-	-	-	-	-	-	-	-	2	57	-	-	-	-	+	+	+	-	-	+	96
CgiSat37	+	+	+	+	+	+	+	+	+	+	6367	516	+	+	+	+	+	+	+	+	+	6242	
CgiSat38	+	+	-	+	-	+	-	+	-	-	6	0	+	-	+	-	-	+	+	+	-	10	
CgiSat39	+	+	+	+	+	+	+	+	+	+	1617	64	+	+	+	+	+	+	+	+	+	1235	
CgiSat40	+	+	+	+	+	+	+	+	+	+	402	31	+	+	+	+	+	+	+	+	+	402	
CgiSat41	-	+	-	-	-	-	-	-	-	-	11	0	+	-	-	-	-	-	-	-	-	7	
CgiSat42	+	+	+	-	+	+	+	+	-	-	124	31	+	-	+	+	+	-	+	-	-	104	
CgiSat43	+	+	+	+	+	+	+	+	+	+	3145	380	+	+	+	+	+	+	+	+	+	3416	
CgiSat44	+	+	+	+	+	+	+	+	+	+	54	2	+	+	+	+	+	+	+	+	+	80	
CgiSat45	-	+	-	-	-	-	-	-	-	-	1	11	+	-	-	-	-	-	-	-	-	4	
CgiSat46	+	-	-	-	+	-	-	-	-	+	25	0	-	+	-	+	+	+	-	+	+	359	
CgiSat47	+	+	+	-	+	+	-	+	+	-	251	79	+	-	+	+	+	+	+	+	+	340	
CgiSat48	-	-	+	+	+	+	+	+	-	+	20	4	+	+	+	+	-	+	-	-	-	10	
CgiSat49	+	+	+	+	+	+	+	+	+	+	77	4	+	+	+	+	+	+	+	+	+	79	
CgiSat50	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	0	
CgiSat51	-	-	-	+	-	+	-	-	-	-	122	0	-	+	-	-	-	-	+	-	-	87	
CgiSat52	+	-	-	-	-	-	-	-	-	-	17	0	-	-	-	-	-	+	-	-	-	6	

Prema satDNA koje se nalaze smještene na samo jednom kromosomu, moguće je utvrditi koje vezane grupe iz jednog posloženog genoma (GCA\_902806645.1) odgovaraju kromosomima iz drugog posloženog genoma (GCA\_011032805.1.). LG1 i chr7, LG2 i chr1, LG4 i chr8 te LG8 i chr4 dijele isti satelitomski profil (Tablica 19.).

#### 4.11. Određivanje dominantnog organizacijskog oblika ponavljajuće DNA povezane s PGE Helitron u genomu *C. gigas*

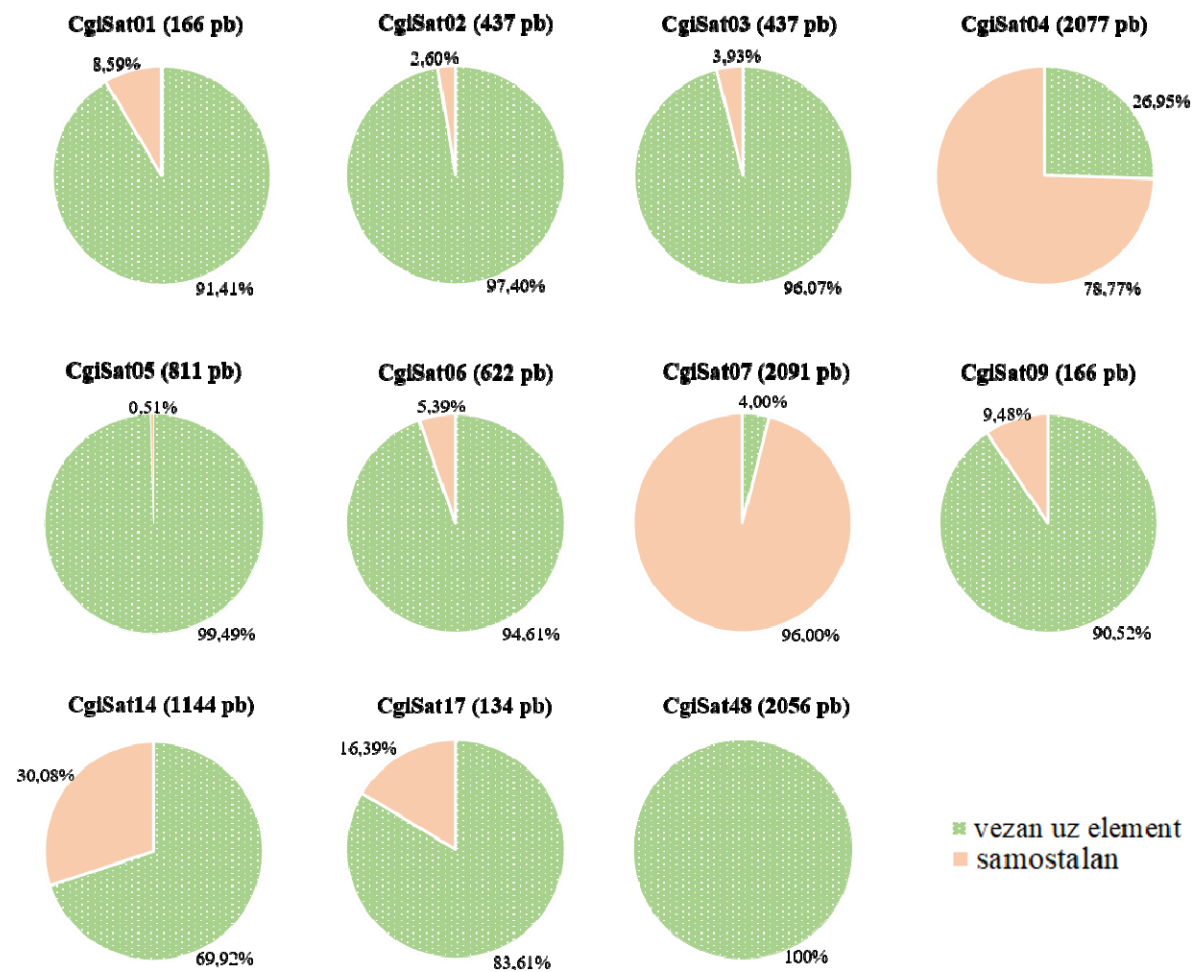
Više od 75 % satDNA u genomu vrste *C. gigas* pokazuje sličnost s elementima Helitron, a to je otprilike 5 % genoma *C. gigas*. Zbog tako značajnog udjela, analizirano je u kakvom se organizacijskom obliku te sekvencije najčešće nalaze: kao samostalni nizovi, povezane s elementom Helitron, ili oboje.

Pretragom baze *Repbase* utvrđeno je da 16 satDNA pokazuje sličnost s 15 elemenata Helitron (Tablica 10.). Helitroni su elementi s dobro definiranom strukturom (Šatović i sur., 2013., Dias i sur., 2014., Thomas i Pritham, 2015., Xiong i sur., 2016., Peñaloza i sur., 2021.): na krajevima elementa nalaze se lijeva i desna očuvana postrana sekvencija (konzervirani *box*-evi), a u središnjem dijelu mikrosatelit iza kojeg slijedi dio s različitim brojem uzastopnih ponavljanja nalik satDNA. Međusobnom usporedbom tih 15 elemenata Helitron utvrđeno je da su njihove sekvencije najbližije u područjima postranih regija na krajevima elemenata, u području tzv. konzerviranih *box*-eva. Elementi Helitron\_N2C-Cgi i Helitron\_N2-Cgi pokazuju međusobnu sličnost u regiji dugoj 44 pb na početku elementa (*Box 1*), regiji dugoj 49 pb na kraju elementa (*Box 2*) te 156 pb dugoj regiji koja se nalazi između mikrosatelitne regije elementa i regije središnjih ponavljanja (*Box 3*). Elementi Helitron-N3\_Cgi, Helitron-N4\_Cgi, Helitron-N28\_Cgi, Helitron-N29\_Cgi, Helitron-N31\_Cgi, Helitron-N32\_Cgi, Helitron-N35\_Cgi i Helitron-N40\_Cgi pokazuju međusobnu sličnost u regiji dugoj 53 pb na početku elementa (*Box 4*) i regiji dugoj 49 pb na kraju elementa (*Box 5*). Helitroni koji pokazuju međusobnu sličnost te glavna obilježja i sekvencije regija sličnosti prikazane su sumarno u Tablici 20. Za elemente Helitron-1\_DEu, Helitron-N25\_Cgi, Helitron-N62B\_Cgi i Helitron-N12\_Cgi nije pronađena nikakva međusobna sličnost niti sličnost s ostalim elementima.

Tablica 20. Prikaz elemenata Helitron i regija njihove međusobne sličnosti te glavna obilježja tih regija.

SatDNA	Ime Helitrona	Duljina regije sličnosti	Pozicija u elementu	Ime regije sličnosti	Sekvencija regije sličnosti
CgiSat01a CgiSat01b	Helitron-N2_Cgi Helitron-N2C_Cgi	44 pb	početak	Box 1	TAGCTCACCTGAGCCGAAGGCTCAAGTGAGCTTTTCTGATCACA
		49 pb	kraj	Box 2	TGTTGCTCAGGTGAGCGATGTGGCCCATGGGCCTCTTGT
		156 pb	između mikrostatitne regije i središnjih ponavljanja	Box 3	AAACTTTTCACATTTTCAACTTCTTCTCAAGAACCACTGGGCCAA TTTCAACCAAATTTGGCACAAAGCATCCTTAGGBDAAGGGGATT HAAATTTGTTAAAAATDAAGGGCCACGCCCTTTTBAAAAGGGAGA TAATTGVAATTADTGAAAATT
CgiSat04 i CgiSat07 CgiSat05 CgiSat14 CgiSat48	Helitron-N3_Cgi Helitron-N4_Cgi Helitron-N28_Cgi Helitron-N29_Cgi	53 pb	početak	Box 4	TAGGGTCTTCCGTTTTCCAACGGAAGACCCATTGTTATTGTTCT GTTTCTT
CgiSat03 CgiSat06 CgiSat17 CgiSat02	Helitron-N31_Cgi Helitron-N32_Cgi Helitron-N35_Cgi Helitron-N40_Cgi				

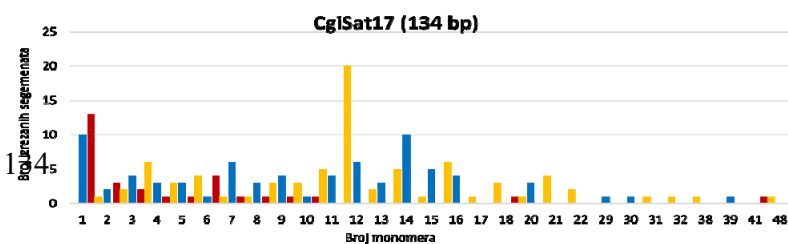
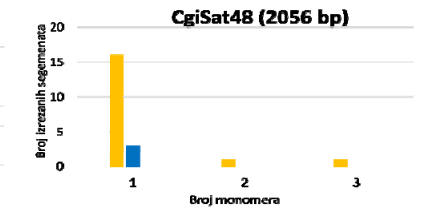
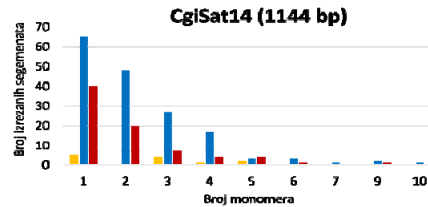
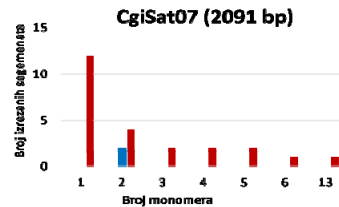
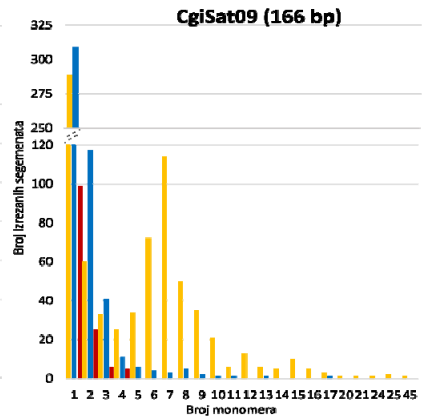
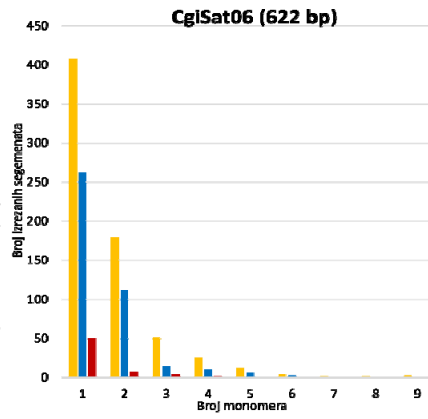
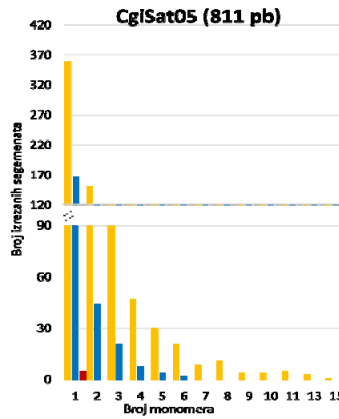
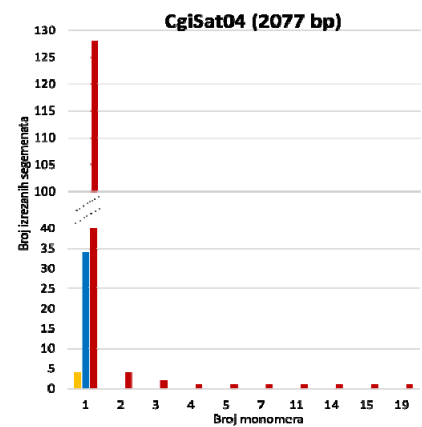
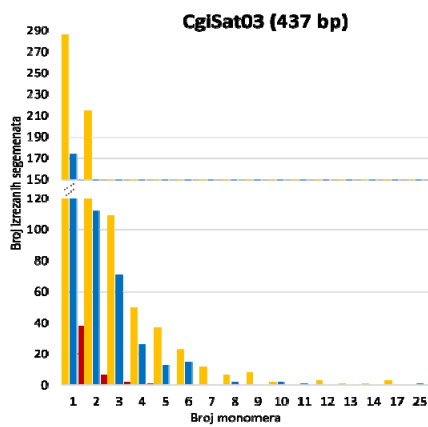
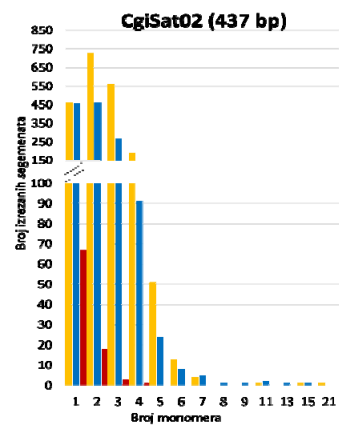
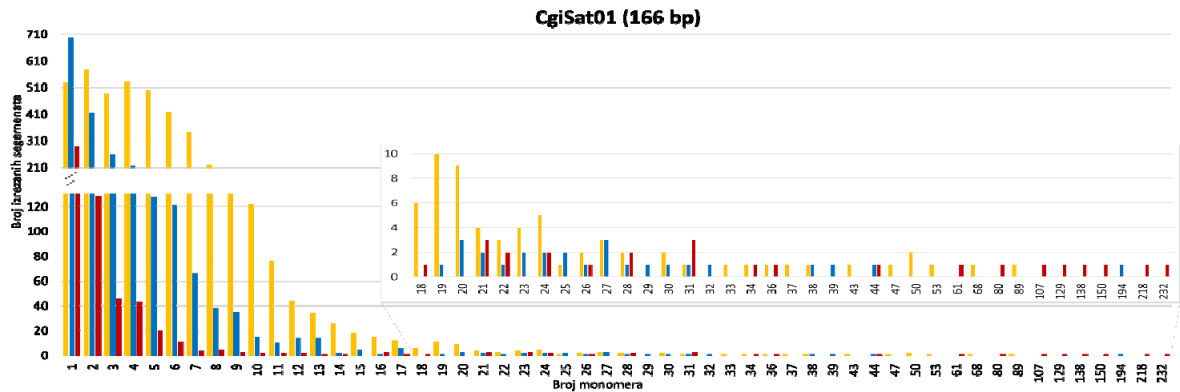
Konzervirane regije sličnosti navedene u Tablici 20. korištene su u daljnjim analizama određivanja glavnih organizacijskih oblika za 11 satDNA, s ukupnim udjelom u satelitemu od 64,25 %, koje su pokazale sličnost s elementima Helitron: CgiSat01, 02, 03, 04, 05, 06, 07, 09, 14, 17 i 48. Za određivanje organizacijskih oblika korišteni su konzervirani *box*-evi *Box 1* i *Box 2* te *Box 4* i *Box 5*. Sekvencije navedenih satDNA najprije su anotirane na kromosome (vezane grupe) referentnog genoma *C. gigas* GCA\_902806645.1. Zatim je za svaki niz monomera (ili pojedinačni monomer ukoliko je nađen u samostalnom obliku izvan niza) koji se pojavljuje na kromosomima (vezanim grupama), uz pomoć skripte napisane programom *Python*, izrezan segment koji uključuje 2000 pb regije koja prethodi monomeru/nizu, sam monomer/niz te 2000 pb regije koja se nalazi iza njega. Tako izrezanim segmentima kromosoma (vezanih grupa) programom *Geneious* anotirane su sekvencije konzerviranih regija određenih u elementima Helitron *Box 1-4*. Nastavno je drugom skriptom napisanom u programu *Python* određeno koji se *box*-evi nalaze u postranim regijama monomera/nizova satDNA (detaljnije o analizi u segmentu 14.5. Materijala i metoda). Prilikom utvrđivanja povezanosti satDNA s pokretnim elementima Helitron u obzir je uzimana prisutnost *box*-a s barem jedne strane monomera/niza monomera satDNA (lijeva ili desna postrana regija). Krnji oblici elemenata Helitron vrlo su česta pojava (Cross i sur., 2005., Xiong i sur., 2016.) te su zbog toga monomeri/nizovi monomera satDNA koji i samo s jedne strane imaju prisutan dio elementa Helitron smatrani povezani s elementom. Oni monomeri/nizovi monomera satDNA koji nisu imali prisutan konzervirani dio elementa Helitron niti s jedne strane, smatrani su samostalnim. Na Slici 28. vidljiv je dominantni organizacijski oblik svake od analiziranih satDNA.



Slika 28. Grafički prikazi udjela monomera/niza monomera u jednom od dva organizacijska oblika za analizirane satDNA. U zagradama uz ime satDNA navedena je duljina monomera.



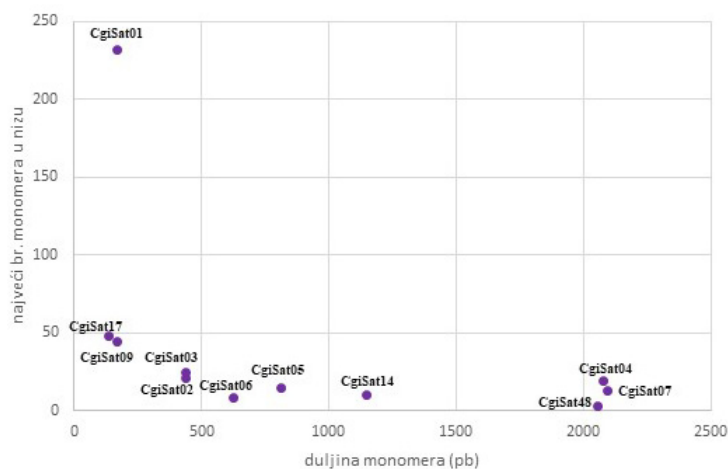
Za svaku satDNA utvrđeno je koliko se monomera nalazi u svakom pojedinom izrezanom segmentu te nalazi li se taj monomer/niz monomera dominantno vezan uz PGE ili samostalno. Rezultati su prikazani na Slici 29. Svim analiziranim satDNA, bez obzira nalaze li se u samostalnom obliku ili vezani uz element, najzastupljenija kategorija ekstrakcija su monomerne sekvencije.



■ oba *box*-a
 ■ jedan *box*
■ nijedan *box*

Slika 29. Broj izrezanih segemenata i broj monomera po izrezanom segmentu za analizirane satDNA. Prisutnost *box*-eva u izrezanim segmentima označena je stupcima: žutim (oba *box*-a), plavim (jedan *box*) i crvenim (nijedan *box*).

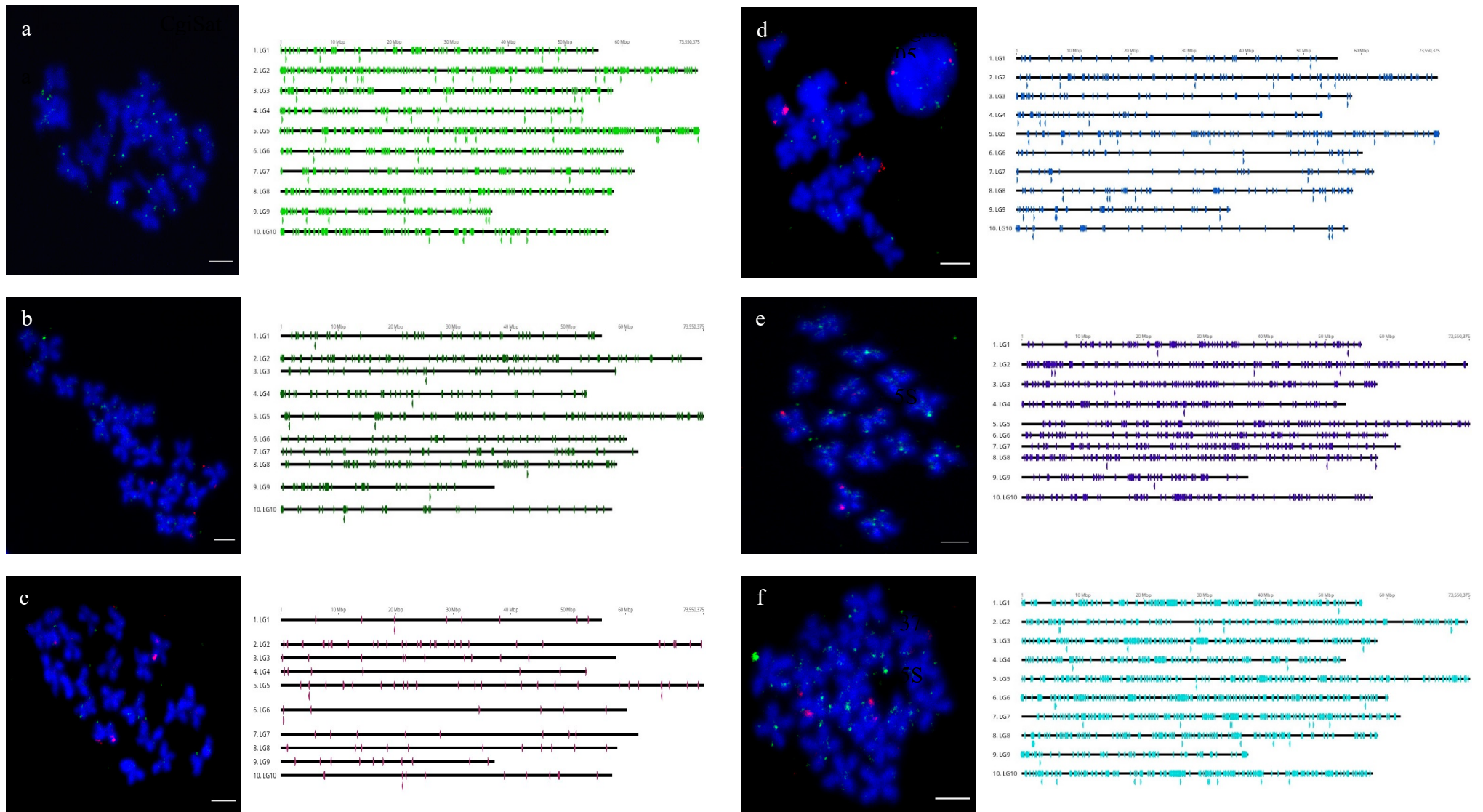
Najzastupljenija satDNA u genomu *C. gigas*, CgiSat01 (Cg170/HindIII) ima najviše izrezanih segemenata te najdulji niz uzastopno ponovljenih monomera. Duljina niza uzastopno ponovljenih monomera CgiSat01 vezanog uz pokretni element iznosi 89 monomera, a duljina najduljeg samostalnog niza uzastopno ponovljenih monomera CgiSat01 iznosi 232 monomera (otprilike 40 kb). U segmentu samostalnog najduljeg niza CgiSat01 detektirana je i praznina u sekvenciji prilikom slaganja genoma tako da je taj najdulji niz potencijalno i dulji. Nalazi se smješten na vezanoj grupi LG3. CgiSat48 u analizi TAREAN okarakterizirana je kao satDNA, međutim, ona se pojavljuje samo kao monomer, dimer i trimer pa bi se zbog toga mogla nazvati ponavljajućom sekvencijom nalik satDNA. Dominantno je prisutna u obliku monomera s dijelovima elementa Helitron prisutnima na obje strane monomera. Ta je sekvencija ujedno i najveće duljine monomera, 2056 pb, te se dodatnim analizama nije mogla razložiti u jednostavnije podsekvencije. Za ostale satDNA najveći broj monomera u nizu iznosi od devet (CgiSat06) do 48 (CgiSat17). Nije uočena međuovisnost najveće duljine niza i povezanosti s elementom. Za sve satDNA osim za CgiSat01, najdulji niz nalazi se u obliku koji je dominantan za tu satDNA, bez obzira je li to samostalno ili povezano s elementom. Također, najveći broj monomera koji se može pojaviti u nizu za neku satDNA nije ovisan o duljini monomerne sekvencije te satDNA (Grafikon 4.).



Grafikon 4. Prikaz odnosa duljine monomera te najvećeg broja monomera koji se može pronaći uzastopno ponovljenih u genomu za svaku od 11 analiziranih satDNA. Iz grafikona je vidljivo da najveći broj monomera koji se može pronaći u nizu za neku satDNA ne ovisi o duljini monomerne sekvencije te satDNA.

#### 4.12. Fluorescencijska hibridizacija *in situ* najistaknutijih satDNA

Sa ciljem utvrđivanja lokalizacije nekih od analiziranih satDNA na metafaznim kromosomima vrste *C. gigas*, napravljena je fluorescencijska hibridizacija *in situ*. Rezultati hibridizacije najzastupljenije satDNA u genomu CgiSat01 (Cg170/HindIII) prikazani su u odlomku 3.2.3. gdje je na Slici 19. vidljivo da ta satDNA lokalizira na centromerama nekoliko kromosoma te intersperzno po krakovima kromosomima. Nekoliko je satDNA dobivenih analizom satelitoma *C. gigas* odabrano za fluorescencijsku hibridizaciju *in situ* preovisno prema udjelu u genomu, ali i prema duljinama nizova u kojima se pojavljuju u genomu jer ih je zbog toga potencijalno lakše detektirati metodom FISH. Na Slici 30. vidljiva je lokalizacija CgiSat02 (a), CgiSat03 (b), CgiSat04 (c), CgiSat05 (d), CgiSat09 (e), CgiSat37 (f) na metafaznim kromosomima *C. gigas* te lokalizacija *in silico* napravljena programom *Geneious* na kromosomima *C. gigas* (vezane grupe gena GCA\_902806645.1). Dio gena 5S rDNA, za koji je poznato da u vrsti *C. gigas* lokalizira na kromosomima 4 i 5 pokazujući jak hibridizacijski signal na subtelomernoj regiji jednog para i slabe signale na drugom paru kromosoma (Cross i sur., 2005.), služio je kao pozitivna kontrola i kolokaliziran je s nekoliko satDNA (CgiSat03, CgiSat04, CgiSat05, CgiSat09 i CgiSat37).



Slika 30. Lokalizacija nekoliko najistaknutijih satDNA u genomu vrste *C. gigas* fluorescencijskom hibridizacijom *in situ* te *in silico* korištenjem programa Geneious: CgiSat02 (a), CgiSat03 (b), CgiSat04 (c), CgiSat05 (d), CgiSat09 (e) i CgiSat37 (f). Uz CgiSat03, 04, 05, 09 i 37 (zeleno) lokalizirana je sonda za 5S rDNA kao pozitivna kontrola (crveno). Skala predstavlja 3  $\mu$ m.

#### 4.13. Satelitomi triju srodnih školjkaških vrsta

Osim za vrstu *C. gigas*, u ovom je radu određen satelitom za još tri vrste kamenica (*C. angulata*, *C. virginica* i *O.edulis*). Odredili smo ukupne setove satDNA u genomima korištenjem očitavanja dobivenih na platformi *Illumina* sekvenciranjem NGS. Tako dobivena očitavanja korištena su u nekoliko krugova analiza dostupnih na serveru *Galaxy*, u paketu *RepeatExplorer*. Za svaki krug analiza, manji setovi očitavanja koji odgovaraju željenoj pokrivenosti genoma nasumično su odabrani, a sumarni rezultati detektiranih satDNA za svaku vrstu prikazani su Tablicama 21.-23.

#### 4.14. Satelitom vrste *C. angulata*

U genomu vrste *C. angulata* detektirana je ukupno 61 satDNA, a ukupno čine 7,70 % genoma te vrste (Tablica 21.). Najbrojnija satDNA u genomu je CanSat01 koja je detektirana i u genomu vrste *C. gigas* (rezultati ovog rada; Clabby i sur., 1996.; López-Flores i sur., 2004.). Ova satDNA pripada obitelji satDNA Cg170/HindIII, a u genomu vrste *C. angulata*, kao i u genomu vrste *C. gigas*, nalazi se u dvije varijante, CanSat01a (164 pb) i CanSat01b (166 pb), te čini 1,32 % genoma što je 17,11 % satelitoma te vrste. Sekvencije detektiranih satDNA uspoređene su s bazom ponavljajućih DNA *Repbase*. Ukupno 51 satDNA, koje su 7,05 % genoma i 91,53 % satelitoma, pokazala je sličnost s nekom od ponavljajućih DNA u bazi *Repbase*, a među njima 17 satDNA koje čine 5,01 % genoma i 64,19 % satelitoma pokazalo je sličnost s pokretnim genetičkim elementima Helitron (Tablica 21.). Duljina monomera najkraće satDNA, CanSat61, iznosi 28 pb, a duljina monomera najdulje satDNA, CanSat15, iznosi 8637 pb. Udio baza A i T u monomernim sekvencijama iznosi 46,8 % za satDNA s najmanjim udjelom (CanSat57) te 76,2 % za satDNA (CanSat59) s najvećim udjelom tih nukleotida u monomernoj sekvenciji.

Tablica 21. Glavna obilježja 61 satDNA detektirane u genomu vrste *C. angulata*.

ime satelita	duljina monomera (pb)	udio u genomu	udio u satelitomu	% AT	Sličnost s ponavljajućim sekvencijama baze <i>Repbase</i>
CanSat01a	164	0,60	7,79	60,4	DNA/Helitron
CanSat01b	166	0,72	9,32	57,2	DNA/Helitron
CanSat02	437	0,49	6,40	65,7	DNA/Helitron
CanSat03	829	0,45	5,81	66,5	DNA/Helitron
CanSat04	805	0,42	5,49	67,4	DNA/Helitron

<b>CanSat05</b>	559	0,37	4,81	66,2	DNA/Helitron
<b>CanSat06</b>	296	0,36	4,71	67,2	DNA/Helitron
<b>CanSat07</b>	177	0,35	4,55	57,1	DNA/Helitron
<b>CanSat08</b>	2564	0,33	4,29	64	DNA/Helitron
<b>CanSat09</b>	662	0,30	3,93	65,3	-
<b>CanSat10</b>	2163	0,30	3,90	67,5	DNA/Helitron
<b>CanSat11</b>	2090	0,23	2,92	66,8	DNA/ISL2EU
<b>CanSat12</b>	446	0,22	2,86	65	LTR/Copia
<b>CanSat13</b>	892	0,19	2,40	64,8	LTR/BEL
<b>CanSat14</b>	166	0,18	2,34	67,5	DNA/Helitron
<b>CanSat15</b>	8637	0,15	1,95	61,8	LTR/BEL
<b>CanSat16</b>	165	0,13	1,69	63	DNA/Helitron
<b>CanSat17</b>	2425	0,12	1,56	67,5	DNA/Kolobok
<b>CanSat18</b>	2532	0,12	1,56	66,3	DNA
<b>CanSat19</b>	675	0,11	1,36	64	DNA/Helitron
<b>CanSat20</b>	1153	0,09	1,22	68,8	DNA/Helitron
<b>CanSat21</b>	933	0,09	1,18	61,4	DNA/Crypton
<b>CanSat22</b>	1295	0,08	1,05	64,9	LTR/Gypsy
<b>CanSat23</b>	730	0,08	0,99	63,7	Interspersed_Repeat
<b>CanSat24</b>	1262	0,07	0,94	64,6	Interspersed_Repeat
<b>CanSat25</b>	555	0,07	0,87	66,3	LTR/Gypsy
<b>CanSat26</b>	6179	0,07	0,84	64,9	DNA/Harbinger
<b>CanSat27</b>	1450	0,06	0,84	67,3	NonLTR/L2
<b>CanSat28</b>	177	0,06	0,76	68,9	DNA/MuDR
<b>CanSat29</b>	2128	0,06	0,72	61	Multicopy_gene/rRNA
<b>CanSat30</b>	2414	0,05	0,65	66,7	NonLTR/Outcast
<b>CanSat31</b>	674	0,05	0,64	63,1	DNA/Crypton
<b>CanSat32</b>	181	0,05	0,61	69,6	-
<b>CanSat33</b>	441	0,05	0,60	69,2	DNA/Helitron
<b>CanSat34</b>	743	0,05	0,59	53,2	LTR/Gypsy
<b>CanSat35</b>	105	0,04	0,57	47,6	-
<b>CanSat36</b>	179	0,04	0,55	51,4	-
<b>CanSat37</b>	181	0,04	0,51	73,5	-
<b>CanSat38</b>	182	0,04	0,49	50,5	DNA/MuDR
<b>CanSat39</b>	300	0,03	0,41	63,7	DNA/Helitron
<b>CanSat40</b>	363	0,03	0,39	62,5	-
<b>CanSat41</b>	182	0,03	0,36	69,2	-
<b>CanSat42</b>	1510	0,03	0,35	69,7	DNA/Helitron
<b>CanSat43</b>	401	0,02	0,31	67,3	DNA/IS3EU
<b>CanSat44</b>	393	0,02	0,30	52,2	IntegratedVirus/DNAV
<b>CanSat45</b>	348	0,02	0,30	53,4	DNA/MuDR

<b>CanSat46</b>	2079	0,02	0,30	64,3	DNA/MuDR
<b>CanSat47</b>	1424	0,02	0,29	69,4	NonLTR/Penelope/Athena
<b>CanSat48</b>	66	0,02	0,29	57,6	-
<b>CanSat49</b>	1602	0,02	0,27	61	LTR/Gypsy
<b>CanSat50</b>	1748	0,02	0,27	66,5	LTR/Gypsy
<b>CanSat51</b>	2042	0,02	0,24	60,3	Interspersed_Repeat
<b>CanSat52</b>	615	0,02	0,21	57,4	-
<b>CanSat53</b>	99	0,02	0,19	60,6	-
<b>CanSat54</b>	357	0,01	0,18	69,7	DNA/hAT
<b>CanSat55</b>	747	0,01	0,18	52,1	DNA/Crypton/CryptonV
<b>CanSat56</b>	997	0,01	0,17	59,9	LTR/Gypsy
<b>CanSat57</b>	231	0,01	0,16	46,8	LTR
<b>CanSat58</b>	1506	0,01	0,16	62,9	LTR/Gypsy
<b>CanSat59</b>	42	0,01	0,14	76,2	DNA
<b>CanSat60</b>	642	0,01	0,14	62,6	DNA/EnSpm/CACTA
<b>CanSat61</b>	28	0,01	0,13	53,6	-

#### 4.15. Satelitom vrste *C. virginica*

U genomu vrste *C. virginica* detektirane su ukupno 33 satDNA koje čine 2,92 % genoma te vrste (Tablica 22.). Najzastupljenija satDNA je CviSat01 duljine 436 pb s udjelom u genomu 0,58 % što je gotovo petina satelitoma te vrste. Duljine monomera iznose od 39 pb za najkraću satDNA (CviSat29) do 2748 pb za najdulju satDNA (CviSat31). Najmanji udio nukleotida A i T iznosi 28 % za satDNA CviSat09, a najviši udio 70,3 % za satDNA CviSat28. Od 33 detektirane satDNA, njih 26 (2 % genoma, 68,53 % satelitoma) pokazalo je sličnost s nekim od elemenata u bazi *Repbase*. Pet satDNA pokazalo je sličnost s elementima Helitron, a udio u genomu iznosi im 0,24 % što je 8,26 % satelitoma.



Tablica 22. Glavna obilježja 33 satDNA detektirane u genomu vrste *C. virginica*.

ime satelita	duljina monomera (pb)	udio u genomu	udio u satelitomu	% AT	sličnost s ponavljajućim sekvencijama baze <i>Rebase</i>
CviSat01	436	0,58	19,78	67,70	DNA/Helitron
CviSat02	1135	0,35	11,80	62,70	DNA/Helitron
CviSat03	180	0,32	10,94	63,30	DNA
CviSat04	1098	0,18	5,98	63,30	DNA/Helitron
CviSat05	161	0,17	5,81	61,50	DNA
CviSat06	178	0,17	5,81	56,20	DNA
CviSat07	578	0,14	4,79	63,10	DNA/Helitron
CviSat08	520	0,11	3,76	68,10	LTR/Gypsy
CviSat09	93	0,08	2,74	28,00	-
CviSat10	1971	0,08	2,56	63,40	DNA/Harbinger
CviSat11	173	0,07	2,43	56,60	DNA/Sola/Sola3
CviSat12	177	0,07	2,37	65,50	-
CviSat13	170	0,06	2,11	52,40	DNA
CviSat14	964	0,06	1,96	68,90	DNA/EnSpm/CACTA
CviSat15	1023	0,05	1,81	65,70	DNA/Crypton
CviSat16	2415	0,05	1,74	59,40	LTR/BEL
CviSat17	1430	0,04	1,39	60,70	Multicopy_gene/rRNA
CviSat18	42	0,04	1,37	59,50	-
CviSat19	181	0,04	1,33	45,30	DNA/MuDR
CviSat20	1111	0,04	1,24	62,40	DNA
CviSat21	743	0,03	1,10	64,10	DNA/hAT
CviSat22	1760	0,03	0,97	68,90	LTR/Gypsy
CviSat23	381	0,03	0,88	64,30	-
CviSat24	719	0,02	0,81	53,80	LTR/Copia
CviSat25	2031	0,02	0,78	64,00	NonLTR/Penelope
CviSat26	1806	0,02	0,58	64,50	DNA
CviSat27	678	0,02	0,57	61,10	DNA/Helitron
CviSat28	263	0,02	0,52	70,30	-
CviSat29	39	0,01	0,47	53,80	-
CviSat30	900	0,01	0,47	65,70	ERV/ERV1
CviSat31	2748	0,01	0,43	65,80	LTR/Copia
CviSat32	181	0,01	0,40	66,30	-
CviSat33	892	0,01	0,30	63,50	DNA/Helitron

#### 4.16. Satelitom vrste *O. edulis*

U genomu vrste *O. edulis* analizom *TAREAN* detektirane su 52 satDNA (Tablica 23.) koje čine 6,67 % genoma te vrste. Najzastupljenija satDNA u genomu vrste *O. edulis* jest OedSat01 koja pripada porodici Cg170/HindIII ponavljajućih DNA. Za razliku od vrsta *C. gigas* i *C. angulata* u čijim se genomima nalazi prisutna u dvije varijante od 164 pb i 166 pb, u vrsti *O. edulis*, ova satDNA nalazi se prisutna u jednoj, onoj od 166 pb. Iako je prisutna samo u jednoj varijanti, njezin udio iznosi 1,3 % genoma što je gotovo jednako ukupnom udjelu dviju varijanti u vrstama *C. gigas* i *C. angulata*. Ova satDNA čini gotovo petinu ukupnog satelitoma ove vrste. Najkraći monomer u genomu je onaj satDNA OedSat43 duljine 23 pb, a najdulji monomer pripada satDNA OedSat26 čija duljina iznosi čak 16 346 pb. Udio nukleotida A i T varira od 50 % (OedSat25) do 74,40 % (OedSat50). Više od polovine detektiranih satDNA, njih 30 koje čine 5,05 % genoma i 75,74 % satelitoma, pokazalo je sličnost s nekom od ponavljajućih DNA iz baze *Repbse*. Njih 13, s ukupnim udjelom 3,23 % u genomu i 48,36 % u satelitomu, pokazalo je sličnost s PGE Helitron.






Tablica 23. Glavna obilježja 52 satDNA detektirane u genomu vrste *O. edulis*.

ime satelita	duljina monomera (pb)	udio u genomu	udio u satelitomu	% AT	sličnost s ponavljajućim sekvencijama baze <i>Repbse</i>
OedSat01	166	1,30	19,48	60,20	DNA/Helitron
OedSat02	142	0,47	6,97	57,00	DNA/Helitron
OedSat03	146	0,41	6,14	62,30	DNA/IS3EU
OedSat04	2408	0,38	5,62	56,60	DNA/hAT
OedSat05	1283	0,36	5,40	63,40	DNA/Helitron
OedSat06	190	0,30	4,42	65,30	-
OedSat07	1080	0,23	3,48	63,80	Interspersed_Repeat
OedSat08	1655	0,23	3,37	65,60	DNA/Helitron
OedSat09	140	0,20	3,00	62,10	DNA/Helitron
OedSat10	2048	0,20	2,92	64,70	DNA/Helitron
OedSat11	379	0,16	2,32	56,50	-
OedSat12	1590	0,15	2,30	64,80	DNA/Helitron
OedSat13	1692	0,15	2,29	64,40	DNA/Helitron
OedSat14	151	0,15	2,25	59,60	-

<b>OedSat15</b>	1628	0,13	1,95	67,70	DNA/MuDR
<b>OedSat16</b>	177	0,13	1,95	67,80	-
<b>OedSat17</b>	601	0,11	1,57	60,70	DNA/Academ
<b>OedSat18</b>	789	0,10	1,47	64,90	Interspersed_Repeat
<b>OedSat19</b>	29	0,09	1,42	58,60	-
<b>OedSat20</b>	334	0,09	1,38	60,20	-
<b>OedSat21</b>	360	0,08	1,16	74,20	DNA/Harbinger
<b>OedSat22</b>	29	0,08	1,12	58,60	-
<b>OedSat23</b>	202	0,07	1,10	65,80	-
<b>OedSat24</b>	29	0,07	1,08	58,60	-
<b>OedSat25</b>	124	0,07	1,07	50,00	-
<b>OedSat26</b>	16346	0,07	1,06	64,80	LTR/Gypsy
<b>OedSat27</b>	431	0,06	0,95	66,60	DNA/Helitron
<b>OedSat28</b>	1995	0,06	0,91	72,20	Interspersed_Repeat
<b>OedSat29</b>	1932	0,06	0,87	59,70	NonLTR/L1
<b>OedSat30</b>	443	0,06	0,86	64,10	DNA/Helitron
<b>OedSat31</b>	41	0,05	0,82	58,50	-
<b>OedSat32</b>	41	0,05	0,79	58,50	-
<b>OedSat33</b>	240	0,05	0,73	59,60	-
<b>OedSat34</b>	36	0,04	0,66	66,70	-
<b>OedSat35</b>	56	0,04	0,63	67,90	-
<b>OedSat36</b>	36	0,04	0,63	66,70	-
<b>OedSat37</b>	170	0,04	0,56	54,10	DNA
<b>OedSat38</b>	989	0,04	0,55	59,00	-
<b>OedSat39</b>	2938	0,04	0,54	61,50	DNA/Mariner
<b>OedSat40</b>	1875	0,04	0,53	66,30	DNA
<b>OedSat41</b>	1023	0,03	0,45	67,30	Multicopy_gene/rRNA
<b>OedSat42</b>	181	0,03	0,43	50,30	DNA/MuDR
<b>OedSat43</b>	23	0,03	0,41	60,90	-
<b>OedSat44</b>	428	0,03	0,37	62,60	DNA/Helitron
<b>OedSat45</b>	2409	0,02	0,36	56,30	LTR/DIRS
<b>OedSat46</b>	32	0,02	0,30	68,70	-
<b>OedSat47</b>	597	0,02	0,26	54,80	DNA/Merlin
<b>OedSat48</b>	3640	0,02	0,25	57,10	DNA/Helitron
<b>OedSat49</b>	63	0,02	0,23	52,40	-
<b>OedSat50</b>	43	0,01	0,21	74,40	-
<b>OedSat51</b>	1477	0,01	0,19	67,60	DNA/Helitron
<b>OedSat52</b>	190	0,01	0,18	61,60	-

#### 4.17. Komparativna analiza satelitoma četiriju vrsta kamenica

Programom *RepeatExplorer* napravljena je komparativna analiza četiriju satelitoma. Istodobno je u toj analizi kao referentni dokument korištena lokalno izrađena baza koja je sadržavala sekvencije svih setova satDNA detektiranih u zasebnim analizama satelitoma četiriju vrsta kamenica. Svi klasteri iz komparativne analize u programu *RepeatExplorer* koji su pokazali sličnost veću od 0,1 % s nekim od zasebno detektiranih satDNA u analizama *TAREAN* izdvojeni su, te je napravljena nova lokalna baza koja je sadržavala sve kontige tih klastera. Tako formirana baza kontiga uspoređena je programom *Geneious* s lokalnom bazom koja je sadržavala sve setove satDNA iz zasebnih analiza *TAREAN*. Ukupno 223 klastera iz komparativne analize (RE) uspoređena su s lokalnom bazom setova satDNA iz zasebnih analiza *TAREAN*. Za detaljniju analizu korišteni su oni klasteri iz komparativne analize kojima je u barem jednoj od četiriju analiziranih vrsta broj očitavanja veći od 1000. Rezultati te usporedbe sumarno su prikazani u Tablici 24.

Tablica 24. Tablica prikazuje 223 klastera dobivena komparativnom analizom u programu *RepeatExplorer* te njihovu prisutnost iskazanu brojem očitavanja detektiranih za svaku vrstu (izraženu numerički te nastavno prikazano naranačastim stupcima lijevo od vertikalne crne linije) te njihovu usporedbu napravljenu programom *Geneious* sa satDNA koje su detektirane u zasebnim analizama *TAREAN* za svaku pojedinu vrstu (stupci desno od vertikalne crne linije). Prve dvije kolone nijansama sive boje pokazuju prisutnost/sličnost svakog pojedinog klastera po vrstama detektiranu u dvije različite analize, *RepeatExplorer* i *Geneious*. Legenda:  -klaster prisutan u samo jednoj vrsti,  -prisutan u dvije vrste,  -prisutan u tri vrste,  -prisutan u četiri vrste,  -nije pokazao sličnost ni sa jednom vrstom (slučajevi samo za analizu programom *Geneious*).





















RE	Geneious	komparativna analiza programom RepeatExplorer								usporedba programom Geneious				
		Klaster	broj očitavanja				broj očitavanja				sličnost			
			<i>C.gigas</i>	<i>C.angulata</i>	<i>C.virginica</i>	<i>O.edulis</i>	<i>C.gigas</i>	<i>C.angulata</i>	<i>C.virginica</i>	<i>O.edulis</i>	<i>C.gigas</i>	<i>C.angulata</i>	<i>C.virginica</i>	<i>O.edulis</i>
		CL1	4345	5203	40194	7613					CgiSat01a	CanSat01a		
										CgiSat01b	CanSat01b			
										CgiSat37	CanSat07			
											CanSat16		OedSat01	
										CgiSat38	CanSat18			
										CgiSat09				
												CviSat14		
												CviSat31		
													OedSat14	
		CL2	19374	20923	0	52								
										CgiSat01a	CanSat01a			
										CgiSat01b	CanSat01b			
		CL5	13	15	10	29303							OedSat01	
		CL6	2354	2547	23372	226				CgiSat03	CanSat03	CviSat08		
											CanSat08		OedSat05	
		CL7	13128	13805	0	3								
		CL8	9409	10303	0	0								
										CgiSat01a	CanSat01a			
										CgiSat01b	CanSat01b			
											CanSat43			
		CL9	12	5	18704	158				CgiSat16	CanSat30	CviSat22		
												CviSat02		
												CviSat11		
												CviSat23		
													OedSat03	
		CL12	7858	9694	0	0				CgiSat03	CanSat03	CviSat08		
		CL13	8012	9131	0	1				CgiSat35	CanSat43			
										CgiSat48	CanSat08			
													OedSat45	
		CL14	0	0	8021	9131							OedSat01	
													OedSat03	
													OedSat04	
		CL15	7470	8896	1	0				Cgisat05	CanSat04			
		CL16	8549	7738	0	0				CgiSat02	CanSat02	CviSat08		

		<b>CL18</b>	8185	7922	6	3					CgiSat04 CgiSat08 CanSat05 CanSat14 CanSat21
		<b>CL19</b>	7649	8029	0	144					CgiSat04 CanSat10
		<b>CL20</b>	7358	7551	0	0					CgiSat01b CanSat01b CgiSat37 CanSat07
		<b>CL21</b>	6998	7600	4	90					CgiSat06 CanSat06
		<b>CL23</b>	5902	6946	0	1130					CgiSat48 CgiSat04 CanSat10 CgiSat05 CanSat04 OedSat05
		<b>CL25</b>	84	58	12303	18					CviSat09
		<b>CL26</b>	5659	6695	26	45					CgiSat01b CanSat01b CgiSat01a CanSat01a CgiSat38 OedSat01
		<b>CL28</b>	6415	5677	3	173					CgiSat48 CanSat08 CgiSat23 CanSat22 CgiSat02 CanSat03 CgiSat03 CgiSat05 CanSat19 CanSat02 CviSat08 CviSat09 OedSat05 OedSat08 OedSat48

	<b>CL29</b>	3762	4520	964	2298		CgiSat05 CgiSat02 CgiSat03 CgiSat48 CanSat03 CanSat08 CanSat19 CviSat08 CviSat10 OedSat05
	<b>CL30</b>	5635	5834	6	0		CviSat29
	<b>CL33</b>	2734	2650	3538	2076		CgiSat16 CanSat30 CviSat22
	<b>CL34</b>	4560	6280	9	4		Cgisat10
	<b>CL35</b>	5318	5422	0	4		CgiSat48 CgiSat05 CgiSat17 CanSat08 CanSat19 CanSat03 CanSat06 CviSat12 OedSat05 OedSat08
	<b>CL36</b>	2111	3112	4910	601		CanSat15
	<b>CL37</b>	223	285	9075	1123		CviSat02 CviSat16 CviSat32
	<b>CL39</b>	0	0	3305	6969		OedSat05
	<b>CL40</b>	4049	3992	1870	331		CanSat39
	<b>CL42</b>	0	0	10	9945		OedSat14



	<b>CL44</b>	5387	4460	0	4		CgiSat09 CanSat16 CanSat01a CanSat01b CanSat27 CgiSat01b	OedSat01
	<b>CL46</b>	0	0	2	9613		CgiSat40 CanSat31	OedSat02
	<b>CL47</b>	4503	4942	0	18		CgiSat42 CanSat05 CanSat08 CanSat25	
	<b>CL48</b>	4748	4631	14	42		CgiSat45	
	<b>CL51</b>	4233	4616	0	0		CgiSat07 CanSat11	
	<b>CL54</b>	1	2	7903	725			CviSat11 CviSat30 CviSat31
	<b>CL55</b>	109	493	334	7636			OedSat04
	<b>CL56</b>	3926	4296	12	204		CgiSat05 CgiSat04 CgiSat14	CanSat04 CanSat10 CanSat20
	<b>CL58</b>	990	1078	3675	2481			OedSat15
	<b>CL59</b>	3799	3905	489	5		CgiSat04	CanSat03 CanSat10 CanSat19 OedSat05
	<b>CL60</b>	3906	4278	0	0		CanSat02 CgiSat02	CviSat08
	<b>CL61</b>	4243	3919	0	0		CgiSat08 CanSat14	
	<b>CL66</b>	4009	3884	0	0		CgiSat04 CgiSat49	CanSat10 CanSat24 CanSat42

	<b>CL68</b>	3676	4128	2	3					CgiSat03	CanSat03 CanSat05 CanSat08 CanSat19	OedSat05
	<b>CL71</b>	3355	4087	15	0					CanSat03 CanSat08 CanSat49	OedSat05	
	<b>CL72</b>	2714	3247	611	702					CgiSat38	CanSat18 CviSat29	OedSat40
	<b>CL75</b>	2883	4081	0	0					CgiSat11	CanSat13	
	<b>CL76</b>	2273	2466	2072	37						CanSat21	
	<b>CL77</b>	0	0	6707	0						CviSat27	
	<b>CL81</b>	0	3	6612	3						OedSat05 OedSat09	
	<b>CL83</b>	954	657	830	4022					CgiSat01a CgiSat01b CgiSat09 CgiSat33	CanSat01a CanSat01b CanSat16 CanSat32 CanSat35	OedSat01 CviSat16 CviSat27 CviSat31 OedSat01
	<b>CL84</b>	3251	3113	0	0					CgiSat05 CgiSat34	CanSat04 CanSat21 CanSat24 CanSat27	



	<b>CL119</b>	0	0	4874	7							CanSat03 CanSat08 CanSat19 CgiSat48 CviSat08 CviSat09 CviSat10 OedSat05 OedSat08 OedSat09
	<b>CL121</b>	0	0	0	4738							OedSat04
	<b>CL122</b>	0	0	10	4723							OedSat07 OedSat12
	<b>CL123</b>	0	0	4508	181							CviSat28 OedSat02 OedSat03
	<b>CL125</b>	0	0	4605	15							OedSat14
	<b>CL126</b>	0	0	0	4610							Cgisat48 CanSat19 CviSat12 OedSat08
	<b>CL127</b>	1258	947	2062	313							CanSat19 CgiSat23 CanSat22 CviSat08 CviSat17 OedSat05
	<b>CL128</b>	1254	1226	955	1145							CgiSat52 CanSat29 CviSat03 OedSat29
	<b>CL131</b>	1898	2000	347	285							CanSat42 CanSat58 CgiSat49
	<b>CL132</b>	55	22	598	3729							CgiSat16 CanSat30 OedSat10
	<b>CL133</b>	0	0	4	4393							OedSat01
	<b>CL135</b>	0	0	4361	7							OedSat05

	<b>CL138</b>	0	1	4329	0				CviSat16 CviSat30 CviSat32 OedSat14 OedSat42
	<b>CL140</b>	2329	1985	0	1		CgiSat14	CanSat20	
	<b>CL142</b>	0	0	4273	0				CviSat30 CviSat31
	<b>CL145</b>	2219	2024	0	0		CgiSat02 CgiSat03	CanSat02 CanSat03 CanSat08 CanSat19	CviSat08 CviSat10 OedSat05
	<b>CL146</b>	1	2	4031	187				
	<b>CL147</b>	0	0	4213	0				CviSat01
	<b>CL148</b>	0	1	4178	0				CviSat11 CviSat28 CviSat30
	<b>CL150</b>	0	0	1	4150				OedSat09
	<b>CL152</b>	1909	2187	0	0				CanSat10 CanSat42 CgiSat04 CgiSat45
	<b>CL155</b>	1831	2195	3	8		CgiSat42 CgiSat01a CgiSat16 CgiSat49	CanSat25 CanSat01a CanSat30	CviSat16 CviSat20 CviSat22 OedSat10 OedSat37

	<b>CL157</b>	2137	1887	0	0						CgiSat37	CanSat07
	<b>CL159</b>	1733	2226	14	1							CanSat21 CanSat23 CanSat24 CgiSat18
	<b>CL160</b>	1893	2070	0	10							CanSat17 CanSat42
	<b>CL161</b>	1960	1998	1	7							CanSat42
	<b>CL165</b>	2013	1920	0	0							CgiSat38
	<b>CL167</b>	139	139	1740	1901							CanSat10 CanSat31 CgiSat04 CgiSat40 OedSat13
	<b>CL169</b>	2119	1798	0	0							CanSat05
	<b>CL170</b>	1397	1402	1081	0							CanSat21
	<b>CL172</b>	1841	1843	25	55							CgiSat04 CanSat05
	<b>CL181</b>											
	<b>CL184</b>	1713	1900	0	0							CgiSat19 CanSat37 CanSat10 CanSat32 CanSat41 CgiSat04
	<b>CL190</b>	0	0	0	3520							OedSat11
	<b>CL191</b>	2377	1035	29	37							
	<b>CL194</b>	1545	1912	0	0							CgiSat04 CgiSat23 CanSat22
	<b>CL197</b>	0	0	0	3351							
	<b>CL199</b>	0	0	0	3287							CanSat05
	<b>CL200</b>	940	781	1552	0							OedSat28
	<b>CL206</b>	1902	1307	0	0							CgiSat15 CanSat28

	<b>CL210</b>	0	0	0	3189						CanSat23
	<b>CL211</b>	263	297	2625	0						CanSat03 CanSat08 CanSat19 CviSat10 OedSat05
	<b>CL213</b>	0	0	0	3152						OedSat12 OedSat07
	<b>CL214</b>	0	0	0	3152						OedSat13
	<b>CL216</b>	0	0	0	3146						OedSat13
	<b>CL217</b>	0	0	1	3141						CanSat39 CviSat28
	<b>CL218</b>	0	0	3137	0						
	<b>CL219</b>	1296	1838	0	0						CanSat39 CanSat42
	<b>CL220</b>	1511	1588	14	11						CgiSat10
	<b>CL222</b>	1290	1359	0	449						CanSat23 CgiSat18 CgiSat49
	<b>CL223</b>	261	272	1565	998						CgiSat08 CviSat33
	<b>CL224</b>	1497	1576	0	0						
	<b>CL227</b>	1711	1332	0	0						CanSat26

		CL228	1491	1353	0	191		CgiSat40 CgiSat08 CgiSat16 CgiSat40 CgiSat42 Cgisat49	CanSat31 CanSat14 CanSat24 CanSat25 CanSat30 CanSat31	CviSat20 CviSat22	OedSat10
		CL231	4	1	1493	1502		CgiSat13	CanSat17		
		CL233	10	12	2955	3				CviSat12	OedSat08
		CL234	1337	1630	0	0			CanSat42		
		CL236	13	6	0	2917			CanSat05		
		CL240	1109	1785	0	0		CgiSat20	CanSat09		
		CL241	1346	1542	0	0					
		CL244	694	738	47	1405			CanSat05 CanSat41		
		CL246	1469	1405	0	0			CanSat32		
		CL248	1421	1443	0	0			CanSat16		
		CL249	1357	1505	0	0					OedSat01
		CL250	1	2	2733	116				CviSat19	
		CL254	1477	1306	0	28					
		CL255	1224	1584	0	1		CgiSat04	CanSat10		
		CL256	1298	1484	0	0		CgiSat01a			
		CL258	1278	1497	0	0			CanSat24		





	CL304	0	0	0	2339							OedSat17
	CL311	422	422	1108	314							CgiSat26 CanSat51 CviSat20
	CL315	1068	1184	0	0							Cgisat29 CanSat45 CgiSat41 CanSat50
	CL319	1164	1061	0	0							CgiSat39 CanSat36 CanSat11 CgiSat07
	CL320	789	1434	0	0							CgiSat34 CanSat27
	CL323	881	1121	189	0							Cgisat35 CanSat43
	CL325	0	0	0	2187							CgiSat18 CanSat23 OedSat18
	CL327	0	0	950	1230							CviSat18 OedSat28
	CL329	1223	945	0	0							CanSat03 CanSat19 CgiSat17 CgiSat48 OedSat05
	CL336	1225	897	0	8							CgiSat19 CanSat37 CgiSat47 CanSat54
	CL338	1037	1072	0	0							CgiSat45
	CL340	347	358	1394	5							OedSat03
	CL341	56	53	1994	0							CviSat19
	CL344	0	0	2063	5							OedSat14
	CL345	996	1066	0	0							
	CL347	0	0	2049	0							CviSat23
	CL353	0	0	0	2009							CanSat24 OedSat18
	CL358	939	1057	0	1							CgiSat40 CanSat40 CanSat23 CanSat31 CgiSat18
	CL359	0	0	0	1966							

	CL367	850	1078	0	0				CanSat49
									CgiSat45
	CL370	1	0	0	1913				CgiSat16 CanSat30 CviSat22 OedSat10 OedSat39
	CL371	1077	824	0	0				CgiSat04 CanSat10
	CL372	1076	823	0	0				CgiSat38 CanSat18
	CL373	1018	877	0	0				
	CL376	587	1022	5	0				
	CL383	0	0	0	1844				
	CL390	0	0	1786	1				CanSat03 CanSat08 CanSat19  OedSat05 OedSat07 OedSat10 OedSat12
	CL392	0	0	0	1777				
	CL393	270	255	0	1251				OedSat03
	CL400	0	0	0	1737				OedSat21
	CL410	0	0	0	1687				
	CL412	0	0	0	1681				OedSat18
	CL413	0	0	0	1679				OedSat03
	CL416	0	0	0	1669				OedSat25
	CL418	0	0	51	1611				OedSat02
	CL434	0	0	368	1222				CgiSat16 CanSat30  Cvisat22 OedSat10
	CL437	0	0	1571	0				CviSat15
	CL445	0	0	1536	0				CviSat16
	CL447	0	0	0	1509				OedSat27
	CL454	0	0	0	1463				OedSat23
	CL465	0	0	1	1433				
	CL469	0	0	0	1417				OedSat39

	CL474	0	0	0	1392							OedSat20
	CL480	69	17	1245	40							CviSat11
	CL491	0	0	0	1328							CanSat03 CanSat04 CanSat06 CanSat08 CanSat19 CgiSat05 OedSat05 OedSat08 OedSat30
	CL496	0	0	1278	2							CviSat10
	CL497	0	0	33	1242							OedSat31 OedSat32
	CL498	0	0	1190	84							CanSat38 CviSat32 OedSat13 OedSat42
	CL501	75	66	0	1114							CanSat40 OedSat01
	CL504	0	0	1250	0							CviSat17
	CL508	93	99	9	1022							
	CL510	0	0	0	1214							
	CL512	34	25	16	1134							
	CL514	0	0	0	1190							OedSat18
	CL516	0	0	1186	0							
	CL522	0	0	0	1175							
	CL523	9	11	1063	86							OedSat02
	CL529	0	0	0	1151							
	CL531	0	0	1144	0							CviSat15
	CL532	0	0	0	1121							OedSat10
	CL555	0	0	1011	9							CgiSat07 CanSat11

Broj očitavanja za pojedini klaster detektiran u komparativnoj analizi u programu RE prikazan je brojčano i grafički u stupcima lijevo od vertikalne crne linije u Tablici 24. Prema komparativnoj analizi u programu RE, očitavanja sekvencija satDNA iz svih četiriju vrsta prisutna su u 57 klastera, iz tri vrste u 39 klastera, iz dvije vrste u njih 79, a iz samo jedne vrste u njih 48. Iz vrsta *C. gigas* i *C. angulata* očitavanja su prisutna u 52 klastera što je najveći broj zajedničkih klastera nekim dvjema vrstama u ovoj analizi. Očitavanja tih klastera podjednaka su, odnosno ti su klasteri/satDNA prisutni u podjednakim udjelima u genomima tih dviju vrsta. Nakon njih slijede *C. virginica* i *O. edulis* s 22 zajednička klastera. Očitavanja iz vrsta *C. angulata* i *C. virginica* prisutna su u dva zajednička klastera, kao i vrsta *C. angulata* i *O. edulis*. *C. gigas* i *O. edulis* imaju jedan zajednički klaster, a *C. gigas* i *C. virginica* nemaju zajedničkih klastera koji bi bili dijeljeni i s drugim vrstama.

Očitavanja iz 24 klastera detektirana su istovremeno u *C. gigas*, *C. angulata* i *O. edulis*, iz 12 klastera u *C. gigas*, *C. angulata* i *C. virginica*, iz dva klastera u *C. angulata*, *C. virginica* i *O. edulis* te iz jednog klastera u *C. gigas*, *C. virginica* i *O. edulis*.

Očitavanja iz 35 klastera detektirana su samo u vrsti *O. edulis*, a očitavanja iz 13 klastera detektirana su samo u vrsti *C. virginica*. Prema komparativnoj analizi RE nema očitavanja koja su prisutna samo u vrsti *C. gigas* ili samo u vrsti *C. angulata*. Ovi su rezultati sumarno prikazani u Tablici 25.

Tablica 25. Tablica prikazuje u kojim su klasterima detektirane sekvencije satDNA iz pojedinih vrsta školjkaša u analizi programom *RepeatExplorer*. Cgi = *C. gigas*, Cvi = *C. vriginica*, Can = *C. angulata*, Oed = *O. edulis*.

VRSTE	KLASTERI
Cgi + Can + Cvi + Oed	1, 5, 6, 9, 18, 21, 25, 26, 28, 29, 33, 34, 36, 37, 40, 48, 54, 55, 56, 58, 59, 68, 72, 76, 83, 92, 93, 104, 107, 112, 127, 128, 131, 132, 146, 155, 159, 161, 167, 172, 181, 191, 220, 223, 231, 233, 244, 250, 263, 273, 298, 311, 340, 480, 508, 512, 523
Cgi + Can + Cvi	15, 30, 71, 109, 170, 200, 211, 277, 290, 323, 341, 376
Cgi + Can + Oed	2, 7, 13, 19, 23, 35, 44, 47, 117, 118, 140, 160, 222, 228, 236, 254, 255, 264, 274, 295, 336, 358, 393, 501
Can + Cvi + Oed	81, 100
Cgi + Cvi + Oed	270
Cgi + Can	8, 12, 16, 20, 51, 60, 61, 66, 75, 84, 90, 96, 98, 101, 102, 106, 110, 145, 152, 157, 165, 169, 184, 194, 206, 219, 224, 227, 234, 240, 241, 246, 248, 249, 256, 258, 269, 272, 276, 291, 294, 302, 315, 319, 320, 329, 338, 345, 367, 371, 372, 373
Cvi + Oed	14, 39, 42, 46, 119, 122, 123, 125, 133, 135, 150, 217, 327, 344, 390, 418, 434, 465, 496, 497, 498, 555
Can + Cvi	138, 148
Can + Oed	106, 300
Cgi + Oed	370
Cgi + Cvi	/
Oed	91, 95, 116, 121, 126, 190, 197, 199, 210, 213, 214, 216, 259, 275, 304, 325, 353, 359, 383, 392, 400, 410, 412, 413, 416, 447, 454, 469, 474, 491, 510, 514, 522, 529, 532
Cvi	77, 142, 147, 218, 268, 283, 285, 347, 437, 445, 504, 516, 531
Cgi	/
Can	/

Stupci desno od vertikalne crne linije u Tablici 24. prikazuju rezultate usporedbe klastera dobivenih komparativnom analizom RE i setova ukupnih satDNA dobivenih analizama TAREAN za svaku pojedinu vrstu. Navedena usporedba napravljena je programom *Geneious*. Sličnost sa satDNA sve 4 vrste iz analiza TAREAN utvrđena je za 21 klaster. SatDNA vrsta *C. gigas* i *C. angulata* pokazale su sličnost sa 47 klastera iz analize RE, satDNA vrsta *C. gigas* i *C. virginica* s tri klastera, vrsta *C. virginica* i *O. edulis* sa šest klastera, vrsta *C. gigas* i *O. edulis* s jednim klasterom, a vrsta *C. angulata* i *O. edulis* sa četiri klastera. SatDNA vrsta *C. gigas*, *C. angulata* i *C. virginica* zajedno su pokazale sličnost s pet klastera, satDNA vrsta *C. gigas*, *C. angulata* i *O. edulis* s njih 19, a satDNA vrsta *C. gigas*, *C. virginica* i *O. edulis* te vrsta *C. angulata*, *C. virginica* i *O. edulis* s po jednim klasterom. Potencijalne vrsno specifične satDNA za pojedine vrste su one satDNA koje su pokazale sličnost s osam klastera za vrstu *C. gigas*, s 15 klastera za vrstu *C. angulata*, s 19 klastera za vrstu *C. virginica* i sa 41 klasterom za vrstu *O. edulis*. Sumarno su ovi rezultati prikazani u Tablici 26. Ukupno 29 klastera nije pokazalo sličnost ni sa jednom satDNA iz lokalne baze. Sa satDNA iz sve 4 vrste sličnost je pokazao ukupno 21 klaster.

Tablica 26. Tablica prikazuje vrsnu distribuciju klastera. Usporedba je napravljena programom *Geneious*. Cgi = *C. gigas*, Cvi = *C. v. virginica*, Can = *C. angulata*, Oed = *O. edulis*.

VRSTE	KLASTERI
Cgi + Can+ Cvi + Oed	1, 6, 9, 28, 29, 35, 72, 83, 90, 107, 119, 126, 127, 128, 145, 155, 211, 228, 276, 370, 434
Cgi + Can + Cvi	12, 16, 33, 60, 311
Cgi + Can + Oed	5, 13, 23, 26, 44, 46, 59, 68, 93, 96, 116, 132, 167, 248, 295, 298, 325, 329, 491
Can + Cvi + Oed	498
Cgi + Cvi + Oed	263
Cgi + Can	2, 7, 8, 15, 18, 19, 20, 21, 47, 51, 56, 61, 66, 75, 84, 98, 101, 102, 109, 111, 131, 140, 152, 157, 159, 172, 184, 194, 206, 222, 231, 236, 240, 244, 255, 269, 272, 277, 291, 315, 319, 320, 323, 336, 358, 367, 371, 372, 555
Cvi + Oed	123, 138, 233, 259, 285, 327, 555
Can + Cvi	/
Can + Oed	71, 353, 390, 501
Cgi + Oed	100
Cgi + Cvi	112, 118, 223

Oed	5,14, 39, 42, 55, 58, 81, 91, 106, 121, 122, 125, 133, 135, 150, 190, 200, 213, 214, 216, 264, 273, 275, 300, 304, 340, 344, 393, 400, 412, 413, 416, 418, 447, 454, 469, 474, 497, 514, 523, 532
Cvi	25, 30, 37, 54, 142, 147, 148, 250, 268, 283, 290, 341, 347, 437, 445, 480, 496, 504, 531
Cgi	34, 48, 165, 220, 256, 274, 294, 338
Can	36, 40, 75, 77, 160, 169, 170, 199, 217, 219, 227, 234, 246, 258, 302

Za većinu klastera dvije su analize pokazala različitu prisutnost po vrstama, odnosno, klaster koji je po usporedbi programom *Geneious* pokazao sličnost sa satDNA iz samo jedne vrste, po analizi RE pronađena su očitavanja svih vrsta u tom klasteru u svim vrstama. Primjerice, CL55 je usporedbom programom *Geneious* pokazao sličnost samo sa satDNA OedSat04, dok su prema analizi RE u tom klasteru pronađena očitavanja svih vrsta, a najveći broj u vrsti *O. edulis*. Ponavljajuća DNA prisutna u CL104 u podjednakom broju očitavanja (udjelu u genomu) prisutna je u vrstama *C. virginica* i *O. edulis* dok je prema usporedbi programom *Geneious* ta sekvencija pokazala sličnost samo sa CanSat15 iz vrste *C. angulata*. Prema usporedbi programom *Geneious*, u toj, kao i u *C. gigas*, prisutna su očitavanja u podjednakom broju, iako ih je otprilike 2,5 puta manje nego u vrstama *C. virginica* i *O. edulis*. Sekvencija klastera CL7 prema usporedbi programom *Geneious* sličan je dvjema satDNA, CgiSat01b i CanSat01b, dok je analiza RE očitavanja iz tog klastera detektirala u tri vrste, *C. gigas*, *C. angulata* i *O. edulis*, iako je broj očitavanja *O. edulis* zanemariv jer iznosi samo tri dok je kod druge dvije vrste on veći od 13 100.





## **5. RASPRAVA**



U ovom je radu, u genomu vrste *C. gigas* iz očitavanja dobivenih sekvenciranjem NGS na platformi *Illumina*, određen ukupni set ponavljajućih sekvencija DNA, repeatom, *de novo*. Korištenjem analiza dostupnih unutar programskog paketa *RepeatExplorer*, utvrđene su ponavljajuće sekvencije DNA koje su u genomu prisutne u udjelu većem ili jednakom 0,01 %, a ukupno čine 20,69 % genoma. Analiza *RepeatExplorer* pogodna je za određivanje ponavljajućih sekvencija *de novo* koje su niskozastupljene u genomima (Goubert i sur., 2015.) te je ovakvom analizom bilo moguće detektirati ponavljajuće DNA koje bi drugim metodama ostale nedetektirane. Dvije najznačajnije skupine ponavljajućih DNA u genomu *C. gigas* jesu satDNA i PGE što je uobičajeno za većinu eukariotskih genoma (Biscotti i sur., 2015.). Najbrojnija satDNA u genomu kamenice *C. gigas* je Cg170/HindIII, a najbrojniji PGE retrotranspozon Ty3/Gypsy. Procijenjeni udio ponavljajućih DNA niži je u odnosu na onaj detektiran od strane Zhanga i sur. (2012.), koji je iznosio 36 %. Udio ponavljajućih sekvencija DNA u genomu školjkaša varira, primjerice za vrste *M. galloprovincialis*, *L. gigantea*, *P. fucata* i *A. californica* iznosi 22,47 – 43,70 % (Murgarella i sur., 2016.). Udio ponavljajućih sekvencija školjkaša znatno je niži u odnosu na neke biljne vrste u kojima udio ponavljajućih DNA može iznositi i do 80 % (López-Flores i Garrido-Ramos, 2012.; Brenchley i sur., 2012.; Qin i sur., 2014.). Velik udio detektiranih ponavljajućih sekvencija DNA u genomu *C. gigas* ostao je neklasificiran (gotovo 75 % ukupno detektiranih ponavljajućih DNA). U radu Zhanga i sur. (2012.) autori su drugim metodologijama određivali udio ponavljajućih DNA u genomu *C. gigas*, i tada je značajan udio sekvencija ponavljajuće DNA ostao neklasificiran, više od 62 %. Također, u prethodno spomenutim školjkaškim vrstama kojima je udio ponavljajućih DNA isto određivan *ab initio*, više od 70 % sekvencija ponavljajuće DNA nije se moglo klasificirati (Murgarella i sur., 2016.).

Analiza kromatina vezanog za protein Cg-CenH3, pokazala je heterogenost u sastavu ponavljajućih sekvencija DNA te se nijedna ne ističe kao dominantna, glavna skupina ponavljajućih DNA koje čine centromere svih kromosoma. Slična situacija pronalazi se primjerice kod trave *Sacchareum spontaneum* u kojoj su centromere građene od ponavljanja nalik satDNA SCEN i ponavljanja porijeklom od retrotranspozona Ty3/Gypsy (Zhang i sur., 2017.), a u ovoj vrsti sastav centromernih sekvencija varira ovisno o stupnju ploidijske koji može uzrokovati gubitak retrotranspozona u području centromere (Wang i sur., 2022.a). U genomu *D. melanogaster* centromere u svom središtu sadrže retroelemente koje okružuju regije satDNA (Chang i sur., 2019.). Kombinaciju satDNA i retroelementa u svojoj centromeri imaju i neke biljne vrste, primjerice riža, šećerna repa i pšenica (Yan i sur., 2006., Kowar i

sur., 2016., Su i sur., 2019.). Suprotno je uočeno kod uročnjaka (*Arabidopsis thaliana*) u čijem genomu glavnu komponentu centromera čini satDNA duga 180 pb koja se naziva AtCon, pAL1 ili, jednostavnije, 180 pb dugo ponavljanje (Kumekawa i sur., 2000., 2001.; Nagaki i sur., 2003.), dok su nekim vrstama iz roda *Candida* retrotranspozoni dominantna skupina ponavljajućih DNA (Lynch i sur., 2010.), kao i nekim životinjskim vrstama (na primjer: Carbone i sur., 2012., de Sotero-Caio i sur., 2017.).

Najbrojnija sekvencija u repeatu *C. gigas* jest satDNA Cg170 koja čini 1,76 % genoma i 68,75 % repeatoma ove vrste. Iako prema rezultatima analize *ChIP-Seq Mapper* ona nije pokazala obogaćenje prema zadanim kriterijima (br. *hit*-ova > 500, omjer *hit*-ova u uzorcima ulaznog i imunoprecipitiranog kromatina > 1,7), njezina prisutnost na centromerama potvrđena je citogenetičkom lokalizacijom na metafaznim kromosomima (Slika 19.). Anotacijom na kromosome *in silico* vidljiva je njena prisutnost duž krakova kromosoma u, većinom, kraćim nizovima koji nisu lako detektabilni metodom FISH na metafaznim kromosomima. Također, njezina prisutnost i izvan područja centromera vidljiva je u pokusima kolokalizacije sa Cg-CenH3 (Slika 20.). Ova satDNA pojavljuje se kao središnje ponavljanje nekih elemenata Helitron (Šatović i sur., 2016.), primjerice elementom Helitron-N55\_CGi. Raspršena prisutnost ove satDNA duž kromosoma, najvjerojatnije kao sastavni dio elemenata Helitron, potencijalni je razlog zašto ova satDNA nije pokazala značajno centromerno obogaćenje. Doduše, u našoj analizi, klaster C116 koji je pokazao obogaćenje pokazao je sličnost s očuvanim terminalnim regijama elementa Helitron-N55\_CGi. Uz njega, sličnost s tim elementom pokazao je i klaster C114, no, on nije pokazao centromerno obogaćenje. Nadalje, tip ponavljajućih sekvencija DNA za ta dva klastera ostao je neklasificiran u analizi *RepeatExplorer*. Sličan rezultat je prilikom analiziranja ponavljajućih DNA u genomima dviju vrsta roda *Drosophila* dobiven u radu Silva i sur. (2019.). U tom je radu satDNA 154TR, otprije poznata kao dio elementa Helitron (Abdurashitov i sur., 2013.), u analizi *RepeatExplorer* klasificirana kao satDNA *niske vjerojatnosti* (engl. low-confidence satellite). Tako su klasificirane još 3 najzastupljenije satDNA u genomima vrsta roda *Drosophila* (Silva i sur., 2019.), otprije poznate kao dijelovi različitih PGE (Zelentsova i sur., 1986.) ili dio intergenske razmaknice ribosomalnih gena (Abdurashitov i sur., 2013.).

Centromerno obogaćenje samo jednog klastera koji pokazuje sličnost s postranom regijom elementa Helitron-N55\_CGi, koji u središnjem dijelu sadrži Cg170 ponavljanja, ukazuje na mogućnost da se u području centromera ti elementi nalaze u skraćenom, krnjem obliku. Skraćeni Helitroni posljedica su

načina njihovog kretanja po genomu, procesa koji se naziva replikacija mehanizmom kotrljajućeg kruga (engl. rolling-circle replication). Njime mogu nastati skraćene kopije Helitrona pri čemu se elementi češće umnažaju bez 3' kraja (Pritham i Feschotte, 2007.). Do sada je otkrivena prisutnost Helitrona u području centromera različitih organizama. Primjerice, Xiong i sur. (2016.) otkrili su prisutnost PGE Helitron u centromerama 27 biljnih vrsta, a u centromerama kukuruza, elementi CentHel bili su skraćeni te ispresijecani s drugim tipovima ponavljajuće DNA. Prisutnost skraćenih elemenata Helitron u centromerama utvrđena je i u kromosomima pšenice (Wang i sur., 2022.b). Potencijalna uklopljenost i drugih centromerno obogaćenih sekvencija DNA u elemente poput elemenata Helitron i visok stupanj raspršenja po genomu mogući je razlog njihovog niskog obogaćenja.

Protein specifičan za centromerno područje vrste *C. gigas*, centromerni histon H3, Cg-CenH3, zbog tehničkih je razloga lokaliziran na preparatima gonadnih stanica mužjaka umjesto na metafaznim kromosomima. Za pripremu metafaznih kromosoma koristi se octena kiselina koja može uništiti proteinske epitope Cg-CenH3 za koje se antitijela vežu. Vjerodostojnost nenasumične i reproducibilne lokalizacije proteina Cg-CenH3 na jezgrama muških spolnih stanica može se potvrditi sličnim obrascima raspodjele koji su uočeni za protein RBMY (Colaco i Modi, 2018.). Taj protein paralog je proteina RBMX koji je uključen u regulaciju kohezije sestrinskih kromatida kao jedan dio centromernog kompleksa (Matsunaga i sur., 2012.; Cho i sur., 2018.). Centromerne ponavljajuće DNA na mišjim i ljudskim spolnim stanicama (Scherthan i sur., 1996.) pokazuju raspodjelu nalik na ovu uočenu za vrstu *C. gigas*. Također, slična lokalizacija centromera na spolnim stanicama uočena je i na ljudskim spolnim stanicama (Wiland i sur., 2016.) i na spolnim stanicama skakavaca (Gosálvez i sur., 2010.).

Zbog navedenih tehničkih razloga, pokusi kombinacije IF-a i FISH-a također su napravljeni na preparatima gonadnih stanica mužjaka. Većina signala u pokusima kombinacije IF-a i FISH-a koja dolazi od ukupnog Cg-CenH3-imunoprecipitiranog kromatina pokazuje preklapanje sa signalom porijeklom od proteina Cg-CenH3 (Slika 18.). Međutim, postoje područja u jezgrama samo sa signalom porijeklom od sekvencija DNA iz imunoprecipitiranog kromatina. Ova potonja predstavljaju područja duž kromosomskih krakova na kojima su te sekvencije smještene i samostalno.

Neke od centromerno obogaćenih sekvencija DNA (C1112, C1150, C1344, C1460 i C1485) lokalizirane su na metafaznim kromosomima. Tim pokusima utvrđena je prisutnost svake na centromerama samo 1-2 kromosomska para (Slike 22. i 23.), ali i prisutnost duž krakova kromosoma. Čak ni najbrojnija satDNA

u genomu vrste *C. gigas*, Cg170, za koju je otprije poznato da lokalizira u području centromera nekoliko kromosomskih parova (Clabby i sur., 1996.), nije pokazala isključivu prisutnost na centromerama već i raspršenu lokalizaciju duž krakova kromosoma. Smještaj i u ostatku genoma, ne samo u centromernom području, vjerojatni su razlog zašto pojedini klasteri ponavljajućih sekvencija nisu pokazali značajna obogaćenja u frakciji kromatina imunoprecipitiranog antitijelom na Cg-CenH3. Manja obogaćenja od očekivanih primijećena su i za sekvencije smještene u centromernom području brašnara *Tribolium castaneum*, gdje je glavna satDNA pokazala znatno manje centromerno obogaćenje od očekivanog (Gržan i sur., 2020.). Sukladno tomu, brojne obogaćene sekvencije pokazale su sličnost s mobilnim elementima te ponavljanjima nazvanima *root-related repeats* u bazi *Repbase* (Tablice 12. i 13.) koji imaju karakteristike satDNA. Oni su možebitno dijelovi do sada neopisanih elemenata Helitron, tj. njihovih središnjih dijelova.

Ovakav smještaj najobogaćenijih centromernih sekvencija po kromosomima i šarolikost skupina ponavljajućih DNA koje čine centromeru kamenice *C. gigas* pokazuju da centromera ove vrste nije uniformnog i specifičnog sastava već je odlikuje heterogenost sastava ponavljajućih sekvencija i kompleksnost njihove građe.

Neke od centromerno obogaćenih sekvencija u analizi *RepeatExplorer* (CL33, CL43, CL69) pokazale su sličnost sa sekvencijama 45S rDNA. Do sada je već primijećena prisutnost klastera sekvencija rDNA u blizini centromera u genomima školjkaša, primjerice u *Mytilus trossulus* (Perezv-vGarcia i sur., 2014.), *Clausinella fasciata* (García-Souto i sur., 2015.), *Chamelea gallina* i *Chamellea striatula* (García-Souto i sur., 2017.a), *Bosemprella incarnata*, *Macomangulus tenuis*, *Moerella donacina* i *Serratina serrata* (García-Souto i sur., 2017.b) te *Arctica islandica* (García-Souto i Pasantes, 2018.). Prisutnost sekvencija porijeklom od različitih sekvencija rDNA u blizini centromera opisana je i u genomima biljaka (Vittorazzi i sur., 2011., Layat i sur., 2012.), kvasaca (Tong i sur., 2019.) i kukaca (Chang i sur., 2019.). Prema navedenim literaturnim podacima, u centromerama i u blizini centromera, ribosomalni geni i sekvencije porijeklom od ribosomalnih gena postoje kao klasteri funkcionalnih ribosomalnih gena, nefunkcionalne kopije ribosomalnih gena te kao tandemska ponavljanja koja dijele regije sličnosti s rDNA. Robicheau i sur. (2017.) ističu da nepotpune, ali gotovo cjelovite sekvencije rDNA mogu istovremeno postojati na nekoliko različitih mjesta u genomu bez da stvaraju nizove uzastopnih ponavljanja te da se preferentno smještaju u blizinu centromera. Uz to, pokazuju da centromere čak posreduju u distribuciji sekvencija rDNA po genomu.

Neobično je da je ukupno 13,33 % centromerno obogaćenih sekvencija pokazalo sličnost s integriranim viralnim sekvencijama. Međutim, za neke viruse otkriveno je da mutacije u njihovom genomu mogu uzrokovati da im se preferentno mjesto integracije u domaćinskom genomu promijeni iz transkripcijski vrlo aktivnih područja u transkripcijski manje aktivno područje centromere (Winans i sur., 2022.).

Analizom sastava kromatina imunoprecipitiranog antitijelom na kromatinsku modifikaciju H3K9me3, utvrđeno je da gotovo 90 % sastava heterokromatina vrste *C. gigas* čine PGE (većim dijelom DNA transpozoni i manjim dijelom LTR-retrotranspozoni) dok su satDNA zastupljene sa samo 2,27 %. U drugim vrstama školjkaša, utvrđeno je da heterokromatin sadrži sekvencije rDNA, primjerice u vrstama *Spisula solida*, *Spisula subtruncata* i *Macrura stultorum* (García-Souto i sur., 2016.) te da se neke satDNA dominantno nalaze u GC-bogatim heterokromatinskim regijama koje mogu biti smještene interkalarno i subterminalno (Petrović i sur., 2008., García-Souto i sur., 2017.c). U odnosu na sekvencije prisutne u centromernom kromatinu, sekvencije detektirane u heterokromatinu pokazuju veću homogenost i raspoređenost u 49 genomskih klastera i tri grupe ponavljajućih DNA, sa spomenutom dominacijom PGE. U centromernom kromatinu, nema dominantne skupine ponavljajućih DNA, a sekvencije su raspoređene u 15 klastera i šest grupa ponavljajućih DNA.

U genomu vrste *C. gigas*, heterokromatin je prisutan u malim količinama, a na metafaznim kromosomima imunoprecipitirane sekvencije DNA asocirane s heterokromatinskom oznakom lokaliziraju (peri)centromerno na jednom paru kromosoma (slabi signali) i telomerno na drugom paru kromosoma (jaki signali), Slika 24. To je u skladu s dosadašnjim analizama heterokromatina u genomu *C. gigas*, metodom C-pruganja, kojom je utvrđen njegov smještaj telomerno na 10. paru kromosoma i (peri)centromerno na 5. paru kromosoma (Bouilly i sur., 2008.). Proteini Cg-CenH3 i H3K9me3 kolokaliziraju samo na malim dijelovima jezgri gonadnih stanica što je također u skladu s telomernim smještajem najvećeg dijela heterokromatina *C. gigas* (Slika 25.; Bouilly i sur., 2008.).

Istraživanja satelitoma dobila su na važnosti u zadnjih nekoliko godina, s pojavom novih metoda poput sekvenciranja NGS. Nove metode omogućile su bolje uvide u organizaciju, funkciju i evoluciju satelitoma različitih vrsta, a koji su se do sada temeljili na približnim vrijednostima i saznanjima (Garrido-Ramos, 2017.). Kombinacija sekvenciranja novih generacija i metoda poput fluorescencijske hibridizacije *in situ* omogućila je detekciju promjena u organizaciji satDNA u različitim organizmima te otkrivanje velikih razlika u rearanžmanima satDNA među pojedinim vrstama (Garrido-Ramos, 2017.).



Vrsta *C. gigas* predstavlja vrijedan nestandardni modelni organizam za proučavanje osobitosti i organizacije samog satelitoma, ali i ukupnih ponavljajućih DNA zbog nekoliko značajnih karakteristika svog genoma: nizak udio satDNA u odnosu na udio pokretnih genetičkih elemenata (Peñaloza i sur., 2021.), nizak udio heterokromatina koji je ograničen na centromerna područja jednog para kromosoma i telomerna područja drugog para kromosoma (ovaj rad, Bouilly i sur., 2008.) te povezanost nizova satDNA s PGE, Helitron/Helentron u kojima čine središnja ponavljanja (ovaj rad, Vojvoda Zeljko i sur., 2020., Peñaloza i sur., 2021.).

Do sada su bile poznate tri satDNA u genomu vrste *C. gigas*: Cg170, DTHS3 i BclI satDNA (pregled u Šatović i sur., 2018.). Cg170 naziva se još i satDNA HindIII zbog toga što je nakon inicijalnog otkrivanja u vrsti *C. gigas*, ta satDNA nakon restriksijske razgradnje enzimom HindIII otkrivena još u sedam vrsta roda *Ostrea* i *Crassostrea* (López-Flores i sur., 2004.). SatDNA DTHS3 otkrivena je nakon razgradnje enzimom HindIII u vrsti *Donax trunculus* (Linnaeus, 1758.), a zajedno s još tri satDNA čini porodicu satDNA DTHS u toj vrsti (Plohl i Cornudella, 1997.). SatDNA BclI je satDNA otkrivena u vrsti *O. edulis*, a kasnije u još pet vrsta roda *Ostrea* i *Crassostrea* (López-Flores i sur., 2010.).

Analizom ukupnih satDNA u genomu vrste *C. gigas*, utvrđen je satelitom te vrste, istraženi su načini povezanosti satDNA i PGE te su određeni najučestaliji organizacijski obrasci najistaknutijih satDNA u genomu ove kamenice.

Korištenjem sekvenciranja NGS niske pokrivenosti genoma i bionformatičkih analiza u programu *RepeatExplorer*, u ovom su radu, detektirane ukupno 52 satDNA koje čine satelitom vrste *C. gigas*. Tako značajan skok u detekciji broja satDNA primijećen je i za mnoge druge vrste primjenom sekvenciranja nove generacije, primjerice u genomima kukaca (Ruiz-Ruano i sur., 2016., Palacios-Gimenez, 2020.; Pita i sur., 2017.), rakova (Boštjančić i sur., 2021.) i riba (Utsunomiya i sur., 2019.).

Veličine monomera satDNA u genomu vrste *C. gigas* značajno variraju (od 21 pb do 2414 pb). Većina ih je kompleksna po veličini monomera (> 50 pb), njih 99,46 %, a značajan udio, njih 37,98 %, čine satDNA čija veličina monomera odgovara onoj koja se smatra evolucijski favoriziranom, 160-360 pb (Henikoff i sur., 2001., Heslop-Harrison i Schwarzacher, 2013.). Takav evolucijski favorizirani raspon duljina monomera primijećen je i u kratkim nizovima satDNA vrste *C. gigas* uklopljenih u pokretne elemente Helitron/Helentron (Vojvoda Zeljko i sur., 2020.). Iako njihov udio varira od 44 do 76,1 %, ipak je većina satelitoma, 98,95 %, obogaćena bazama A i T što je općenito karakteristika sekvencija

satDNA. Ukupni udio satDNA u genomu pacifičke kamenice iznosi 6,33 %, što je u skladu s udjelima satDNA i u drugim porodicama školjkaša, ali značajno manje od udjela satDNA u genomima mnogih drugih organizama (Garrido-Ramos, 2017.), uz iznimku satDNA PjHhaI čiji udio u genomu školjkaša *P. jacobaeus* iznosi 20 % (pregled u Šatović i sur., 2018.). Najbrojnija satDNA u genomu *C. gigas* je satDNA Cg170/HindIII, za koju dostupni literaturni podaci navode udio u genomu 1-4 % (Clabby i sur., 1996.). Analizom u ovom radu njezin je udio procijenjen na 1,76 % što je u skladu s dosadašnjim procjenama te je to gotovo trećina ukupnog satelitoma ove vrste. To je značajno niže od udjela glavne satDNA u genomima nekih vrsta u kojima može iznositi do 40 %, primjerice kod kukaca roda *Tribolium* (Plohl i sur., 1993., Juan i sur., 1993.; Ugarković i sur., 1996.a i b; Mravinac i Plohl 2007., 2010.; Mravinac i sur., 2004., 2005.).

Pokretni genetički elementi i satDNA dinamične su sekvencije i najbrojniji tip ponavljajuće DNA, a svojom evolucijom mijenjaju evoluciju genoma u cjelini te njegovu strukturu i funkciju (López-Flores i Garrido-Ramos, 2012.). Ta dva tipa ponavljajuće DNA razlikuju se u nekoliko bitnih karakteristika: strukturi, organizaciji, načinima premještanja po genomu te dinamici njihovih sekvencija (Šatović i Plohl, 2013.). Međutim, sve veći broj istraživanja pokazuje kako PGE i satDNA mogu biti povezani na različite načine (pregled u Meštrović i sur., 2015.), jedan tip sekvencija može nastajati iz drugog (Kapitonov i sur., 1998., Macas i sur., 2009., Sharma i sur., 2013., Dias i sur., 2015., Šatović i sur., 2013., 2016., Lower i sur., 2018., Vondrak i sur., 2020.), a PGE mogu sudjelovati u distribuciji satDNA po genomu (Alkan i sur., 2007., Dias i sur., 2015., Šatović i sur., 2013., 2016., Paço i sur., 2019.). Pregled njihove povezanosti kod školjkaša donešen je u Šatović i sur. (2018.).

U ovom je radu većina detektiranih satDNA pokazala sličnost s nekim tipom PGE te je potvrđeno da veze između uzastopno ponovljenih i raspršeno ponovljenih DNA mogu biti veoma bliske, složene i opsežne. Najveći udio satDNA pokazao je sličnost s elementima Helitron, čak 75,61 % satelitoma (Tablica 18.). Budući da su visok stupanj raspoređenja po genomu, odnosno prisutnost na većem broju kromosoma, pokazale samo satDNA slične Helitronima (Slike 21., 22., 27. i 31.), može se pretpostaviti da su upravo oni ključni za distribuciju satDNA u genomu *C. gigas*. Slično su utvrdili i Vojvoda Zeljko i sur. (2020.) u svom radu. Oni su otkrili da su sekvencije nalik satDNA u genomu pacifičke kamenice asociirane s elementima HINE (eng. *Helitron-associated INterspersed Elements*), neautonomnim Helitronima, strukturnim pandanima Helitrona. SatDNA koje su pokazale sličnost s drugim tipovima PGE, pokazale su sličnost samo manjim dijelom svojih monomera, te nisu bile sastavni dio PGE-a što

njihovu propagaciju uz pomoć PGE čini malo vjerojatnom. Takva djelomična sličnost može biti posljedica različitih rearanžmana sekvencija PGE, primjerice nepreciznog izrezivanja PGE s nekog mjesta u genomu pri čemu dio sekvencije PGE ostaje na starom mjestu, a veći dio premješta se na novu lokaciju ili uzastopnim ponavljanjem dijela PGE i susjedne genomske sekvencije. Budući da je u trenutno referentnom složenom genomu pacifičke kamenice još uvijek prisutan dio neposloženih okvira (eng. *scaffolds*), raspored satDNA koje su pokazale sličnost s PGE i onih koje nisu trebalo bi uzeti s rezervom jer on možda nije konačan.

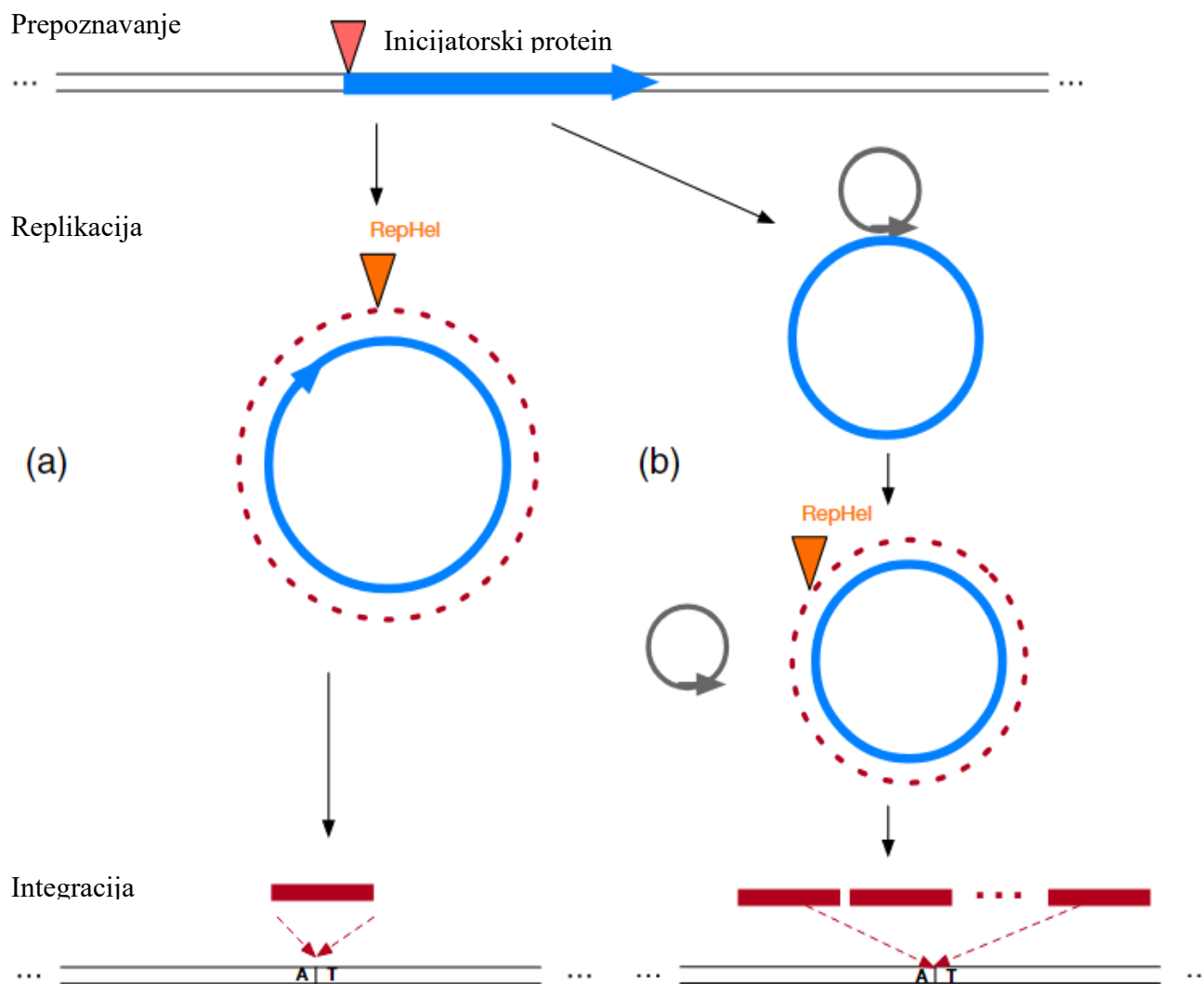
Nadalje, u ovom je radu istražen dominantni organizacijski oblik 11 satDNA koje su pokazale sličnost s PGE Helitron. Organizacija navedenih satDNA utvrđivana je kao jedan od sljedećih oblika u odnosu na prisutnost očuvanih postranih regija elemenata Helitron s kojima su pokazale sličnosti: samostalne satDNA, satDNA uklopljene u element i satDNA prisutne u oba oblika. Od 11 analiziranih satDNA, dominantni organizacijski obrazac kao samostani nizovi imaju CgiSat04 i CgiSat07. Monomeri samo jedne satDNA, CgiSat48, nalaze se isključivo uklopljeni u elemente u genomu *C. gigas*. Preostalih osam satDNA koje čine gotovo 50 % satelitoma u genomu dominantni organizacijski obrazac je uklopljenost u element. Rezultati pokazuju da organizacija satDNA u genomu nije ujednačena te da jedna satDNA može istovremeno postojati u više organizacijskih oblika. Prema tim rezultatima možemo pretpostaviti da su satDNA iz PGE u genomu *C. gigas* nastajale različitim mehanizmima. Nekoliko literaturno opisanih mehanizama može dovesti do takve distribucije nizova satDNA u genomu.

Izsvak i sur. (1999.) prvi su predložili mehanizam umnažanja neautonomnih elemenata MITE nazvanih *Angel* i detektiranih u genomu teleostnih riba. Mehanizam se temelji na specifičnosti građe elemenata MITE i procesu replikacije DNA tijekom kojega ti neautonomni elementi od stanice posuđuju sve potrebne enzime za svoje umnažanje, te reinserciju u genom. Elementi MITE građeni su od središnje domene bogate bazama A i T koju omeđuju terminalna palindromska inverzna ponavljanja. Takva palindromska terminalna ponavljanja mogu uzrokovati zastajkivanje replikacije, a područje bogato bazama A i T zbog slabijih veza među lancima DNA može uzrokovati odvajanje novonastajućeg lanca od replikacijskih rašlji. Također, to područje može formirati strukturu omče u odvojenom lancu. Na slobodnom 3'-kraju odvojenog lanca može ponovno započeti replikacija u kojoj kao kalup služi dio lanca sa sekvencijom inverznog ponavljanja. Tako se po završetku replikacije može formirati struktura ukosnice koja se kasnije nakon cirkularizacije može udvostručiti i reinsertirati na neko novo, homologno mjesto u genomu.

Još jedan mehanizam koji se temelji na procesu replikacije jest onaj koji su Hikosaka i Kawahara (2004.) predložili za nastajanje uzastopnih ponavljanja djelomičnim umnažanjem elemenata MITE. Autori mehanizam opisuju na primjeru sekvencije Xstir iz genoma triju vrsta žaba iz roda *Xenopus*. Sekvencija Xstir u genomima tih triju vrsta postoji kao uzastopni niz i kao dio elementa MITE. Sekvencije Xstir dobivene razgradnjom elementa MITE (nazvane Xmix) očuvane su u tim trima vrstama, uz manje vrsno specifične razlike. Međutim, niz koji Xstir svojim uzastopnim ponavljanjem tvori u genomima *X. laevis* i *X. borealis* i koji je zajednički tim dvjema vrstama, bitno se razlikuje od niza kojega sekvencija Xstir tvori u genomu vrste *X. tropicalis* (tropXstir). Autori pretpostavljaju da su nizovi sekvencija Xstir i tropXstir nastali djelomičnom uzastopnom duplikacijom središnje sekvencije elementa MITE nakon evolucijskog razdvajanja vrsta *X. tropicalis* i *X. laevis/X. borealis*. Od dva spomenuta, autori kao vjerojatniji mehanizam nastanka uzastopno ponovljenih sekvencija Xstir ističu onaj koji se temelji na jednolančanom lomu DNA i izrezivanju dvolančanog dijela strukture omče koja nastaje između dva susjedna elementa MITE na istom lancu DNA. Nastanak takve strukture tijekom replikacije moguć je zbog toga što dio zaostajućeg lanca u replikacijskim rašljama ostaje u jednolančanom obliku. Nakon izrezivanja peteljke, zaostajući se lanac spaja, a nakon završene replikacije jedna od nastalih molekula DNA ima istosmjerno uzastopno ponovljen interni segment elementa MITE. Tako ponovljen segment dalje se može tandemizirati nesimetričnim *crossing over*-om (Krüger i Vogel, 1975.; Smith, 1976.) ili replikacijskim proklizavanjem (Levinson i Gutman, 1987.) te premještati po genomu na nova, homologna mjesta. Ovakvi mehanizmi umnažanja i premještanja potencijalno su mogući i za elemente Helitron i Helentron budući da njihova terminalna inverzna ponavljanja mogu biti palindromska ili im se palindromske sekvencije nalaze u blizini 3' ili 5'-kraja (Kapitonov i Jurka, 2001., Feschotte i Wessler, 2001., Thomas i sur., 2014., Yang i Barbash, 2008.).

Uz mogućnost ovakvih mehanizama premještanja, za PGE Helitron pokazano je da se umnažaju i premještaju mehanizmom kotrljajućeg kruga, uz nastanak kružnog intermedijera (Grabundžija i sur., 2016.). Transpozicija Helitrona odvija se u tri koraka: prepoznavanje, replikacija i insercija (Slika 31., preuzeto iz Xiong i sur., 2016.). Enzim RepHel katalizira replikaciju koja može rezultirati cjelovito ili djelomično umnoženim Helitronom, a u iznimno rijetkim slučajevima mogu nastati dva cjelovita elementa Helitron jedan za drugim kao posljedica dva kruga replikacije. Replikacija kreće od 5'-kraja, a struktura ukosnice na 3'-kraju služi kao terminatorsko mjesto replikacije osim ako se nekim internim rekombinacijskim događajima to mjesto ne izreže iz kružnog intermedijera. U potonjem slučaju mogu

nastati samo djelomično umnoženi Helitroni pa se nizovi uzastopno ponovljenih krnjih Helitrona mogu ugraditi na neko novo mjesto u genomu. Ovakvim bi se ishodom replikacije Helitrona mogao objasniti velik broj pronađenih satDNA u ekstrahiranim regijama genoma vrste *C. gigas*, koji su imali konzerviranu postranu regiju prisutnu samo s jedne strane niza. Druga mogućnost nastanka takvih nizova jest mogućnost da se između nizova povezanih s elementom i samostalnih nizova dogode rekombinacije koje mogu stvoriti hibridne strukture. Daljnja produljivanja i skraćivanja takvih hibridnih, ali i samostalnih nizova, mogu biti posljedica kombinacije mehanizama poput nesimetričnog *crossing over*-a i reinsercije umnoženih elemenata u genom (pregled u Plohl i sur., 2008.).



Slika 31. Prikaz mehanizma kotrljajućeg kruga prema kojemu se umnažaju i premještaju elementi Helitron. Ovisno o konformaciji Helitrona, moguća su dva završetka mehanizma kotrljajućeg kruga, nastankom cjelovitog monomera elementa Helitron (a) ili skraćениh kopija elemenata Helitron (b). Slika prilagođena prema Xiong i sur. (2016.).

Scalvenzi i Pollet (2014.) na primjeru elementa MITE (miDNA4) koji u sebi sadrži satDNA predlažu dva moguća mehanizma nastanka samostalnih nizova uzastopno ponovljenih DNA i onih povezanih s elementima. Autori su pretpostavili postojanje predačke sekvencije elementa miDNA4 koji u sebi prvotno nije sadržavao sekvenciju satDNA, no ona je nastala nakupljanjem supstitucija tijekom

vremena, a sama predačka sekvencija omogućila je stvaranje biblioteke satDNA (Scalvenzi i Pollet, 2014.). Amplifikacija i premještanje takve satDNA odvija se ispreplitanjem klasičnog načina umnažanja satDNA (rekombinacija) i premještanjem samog elementa (Slika 4., preuzeto iz Scalvenzi i Pollet, 2014.). Element koji u sebi sadrži monomer sekvencije satDNA ili kraći niz satDNA, učestalije transponira pa je zbog toga veća vjerojatnost da će se naći u rekombinacijski aktivnom području, primjerice subtelomerne regije. Takva pozicija može rezultirati češćim umnažanjem sekvencije satDNA unutar elementa čime se njezin niz produljuje. Međutim, što je niz dulji, to je transpozicija elementa manje vjerojatna. Nadalje, nakupljanjem mutacija u rubnim regijama elementa MITE tijekom vremena mogu nastati samostalni nizovi satDNA bez postranih regija elementa. Sličnu situaciju nalazimo u vrsti *Locusta migratoria* u čijem genomu postoje satDNA koje su nastale iz elementa LINE te se nalaze u genomu i kao dugi uzastopno ponovljeni nizovi. Kao primjer satDNA koja je potencijalno nastala opisanim mehanizmom u genomu vrste *C.gigas* jest satDNA CgiSat01 (Cg170/HindIII). Budući da u genomu *C. gigas* postoji velik broj kopija elementa Helitron koji u svom središnjem dijelu sadrži monomer CgiSat01 ponovljen 1-10 puta, možemo pretpostaviti da je to posljedica učestalog premještanja elementa s kraćim nizovima. Što je dulji niz, to mu je broj kopija u genomu proporcionalno manji, a nizovi s najvećim brojem monomera isključivo su prisutni u samostalnom obliku (Slika 30.).

Uz CgiSat01, za još 10 satDNA u genomu *C. gigas* analizirana je prisutnost postranih konzerviranih regija uz nizove satDNA. Svaka od analiziranih satDNA imala je specifičan omjer ekstrahiranih sekvencija u kojima su bile prisutne konzervirane regije samo s jedne ili obje strane središnjeg niza ili potpuno odsutne. Postoji mogućnost da je takva situacija početni i završni dio u nastanku uzastopnih ponavljanja DNA prema modelu Scalvenzija i Polleta (2014.). Monomeri i nizovi koji imaju konzerviranu regiju s obje strane potencijalno su nastali produljivanjem nekog kraćeg niza ili skraćivanjem duljeg niza. Monomeri i nizovi monomera koji imaju prisutnu samo jednu konzerviranu regiju, potencijalno su izgubili drugu regiju nakupljanjem mutacija, rekombinacijskim događajima, nepotpunom replikacijom kotrljajućim krugom, ili izrezivanjima okolnih regija.

U radu iz 2019., Vondrak i sur., analizirali su satDNA u biljci *Lathyrus sativus* te također uočili neuobičajene duljine nizova satDNA. U toj vrsti, samo su dvije satDNA u genomu bile prisutne u duljinama uobičajenima za „klasične satDNA“, dok su preostale bile prisutne uglavnom u kraćim nizovima. Mjestimična prisutnost duljih uzastopnih nizova nije bila povezana s duljinom monomera. Takav su zaključak donijeli i Ruiz-Ruano i sur. (2016.) proučavajući evoluciju satDNA u vrsti *Locusta*

*migratoria* u čijem genomu satDNA duljina 5-400 pb podjednako tvore i kraće i dulje nizove. U genomu *C. gigas* samo jedna satDNA tvori nizove duljina poput klasičnih satDNA, CgiSat01 (Cgi170/HindIII), no, većina tih nizova (> 91 %) nije samostalna u genomu već vezana uz PGE Helitron. Jedine dvije satDNA koje su u genomu *C. gigas* prisutne isključivo u samostalnom obliku (CgiSat04 i CgiSat07), ne tvore nizove dulje od 19 i 13 monomera, a monomeri su im dugi po 2000 pb. U genomu vinske mušice (Kuhn i sur., 2012.; Larracuate, 2014.) uočeno je da iste satDNA mogu tvoriti nizove različitih duljina, međutim, dulji nizovi prisutni su u heterokromatinu, a kraći nizovi u eukromatinu što kod vrste *L. migratoria* nije slučaj (Camacho i sur., 2015.). To nije slučaj ni kod *C. gigas* što je pokazano detaljnim analizama centromernog kromatina i heterokromatina u ovom radu.

U ovom radu, određeni su satelitomi četiriju vrsta kamenica te međusobno uspoređeni. Za svaku vrstu kamenice najprije su određeni zasebni satelitomi u nekoliko krugova analiza TAREAN, kako bi se postigle veće dubine analiza i otkrio što veći broj satDNA u svakom od genoma (Novák i sur., 2020.). Za međusobnu usporedbu korištena je komparativna analiza dostupna u paketu analiza RE te su tako detektirani klasteri satDNA uspoređeni programom *Geneious* sa satDNA dobivenima u zasebnim analizama TAREAN. Velik broj satDNA koje su detektirane zasebnim određivanjima satelitoma u genomima četiriju vrsta kamenica, nisu klasificirane kao satDNA u komparativnoj analizi programom RE. Razlozi za to su dvojaki. Djelomično je to zbog tehničkih ograničenja komparativne analize RE i TAREAN, a djelomično zbog specifičnosti samih ponavljajućih DNA u genomima školjkaša. Za razliku od analize samo jedne vrste programom RE, komparativna analiza genoma više vrsta istovremeno iziskuje odabir jedne od dviju opcija odabira očitavanja za svaku vrstu. Moguće je odabrati jednak broj očitavanja za svaku vrstu, bez obzira na veličinu genoma pojedine vrste ili se može odabrati reprezentativan broj očitavanja koji predstavlja određenu pokrivenost genoma, jednaku za sve vrste koje se analiziraju (Novák i sur., 2020.). U ovom je radu odabrana druga opcija budući da je na taj način povećana sposobnost detekcije ponavljajućih DNA, ali je zbog dvostruko većeg genoma vrste *O. edulis* od ostalih vrsta moguće da je očitavanja iz te vrste bilo više u odnosu na ostale tri vrste. Nadalje, budući da se istovremeno analizira nekoliko vrsta, analiza nije toliko detaljna koliko je detaljna kad se analizira svaka vrsta zasebno. Zbog toga i sami tvorcii programskog paketa RE (Novák i sur, 2020.) predlažu da se najprije napravi analiza na svakoj vrsti zasebno, a zatim komparativna analiza. U ovom je radu napravljen i korak dalje kojim se u komparativnoj analizi kao jedan od ulaznih dokumenata koristila lokalno izrađena baza koja je sadržavala sve setove satDNA svake pojedine vrste dobivene zasebnim



analizama TAREAN. Na taj se način pokušalo doskočiti smanjenoj *dubini* komparativne analize te povezati klastere dobivene komparativnom analizom, a koji nisu bili klasificirani kao satDNA u komparativnoj analizi, i satDNA dobivene individualnim analizama TAREAN. Još jedan način kojim se pokušalo produbiti svaku od zasebnih analiza TAREAN jest nekoliko *krugova* analiziranja s odabirom različitih vrijednosti parametara svake analize za pojedinu vrstu te odabir one analize za daljnje usporedbe koja je detektirala najveći broj različitih satDNA. Na sličan su način i drugi autori pokušali doskočiti ovom problemu *dubine* analize (Novák i sur., 2020., de Lima i Ruiz-Ruano, 2022.).

Postoje slučajevi klastera, ukupno njih 29, koji su klasificirani kao satDNA komparativnom analizom te neki od njih nalazi dosta velik broj očitavanja, međutim, nisu pokazali sličnost ni sa jednom satDNA iz lokalne baze satDNA nakon usporedbe programom *Geneious*. Primjeri za to su CL92, CL95, CL110, CL117, CL146, CL181, CL191, CL197 i CL218, kojima za barem jednu vrstu broj očitavanja prelazi brojku 2000.

Vrste *C. gigas* i *C. angulata* najsirodnije su (Gagnaire i sur., 2018., Salvi i sur., 2014.) te su im udjeli satelitoma u genomu približno jednaki. Za vrstu *C. gigas* udio satDNA u genomu iznosi 6,33 %, a za vrstu *C. angulata* udio satDNA u genomu iznosi 7,70 % te je u njenom genomu individualnim analizama satelitoma detektirano 10 satDNA više nego u genomu vrste *C. gigas*. Očitavanja satDNA tih dviju vrsta istovremeno su detektirana u 52 klastera, a prema usporedbi programom *Geneious*, satDNA tih dviju vrsta zajedno pokazuju sličnost u 49 klastera. Iz toga proizlazi da te dvije vrste dijele najveći broj zajedničkih satDNA. S vrstom *C. virginica* kojoj je udio satDNA trostruko manji nego u vrstama *C. gigas* i *C. angulata*, te iznosi 2,33 % za 33 detektirane satDNA, vrsta *C. gigas* nema zajedničkih satDNA prema analizi RE, a prema usporedbi programom *Geneious*, satDNA te dvije vrste zajedno pokazuju sličnost u trima klasterima. Vrsta *C. virginica* s vrstom *C. angulata* istovremeno pokazuje sličnost u dva klastera, a prema usporedbi programom *Geneious* ne pokazuju zajedničku sličnost niti u jednom klasteru. S vrstom *O. edulis* čije 52 detektirane satDNA u genomu dvostruko većem od vrsta *C. gigas* i *C. angulata* imaju udio od 6,67 %, *C. virginica* istovremeno pokazuje sličnost u dva klastera prema analizi RE, a prema usporedbi programom *Geneious* u njih četiri. Vrsta *O. edulis* s vrstom *C. gigas* pokazuje zajedničku sličnost samo u jednom klasteru, i prema usporedbi s programom *Geneious* i prema analizi RE, a s vrstom *C. angulata* pokazuje zajedničku sličnost u četiri klastera prema usporedbi programom *Geneious*, a u dva prema analizi RE. Te tri vrste, *C. gigas*, *C. angulata* i *O. edulis* prema analizi RE pokazuju zajedničku sličnost u 24 klastera, a prema usporedbi programom *Geneious* u njih

19. Vrste *C. gigas*, *C. angulata* i *C. virginica* pokazuju zajedničku sličnost u 12 klastera prema analizi RE, a prema usporedbi programom *Geneious* u njih pet. Istovremenu sličnost u samo jednom klasteru pokazuju vrste *C. gigas*, *C. virginica* i *O. edulis* i vrste *C. angulata*, *C. virginica* i *O. edulis* prema usporedbi programom *Geneious* te vrste *C. gigas*, *C. virginica* i *O. edulis* prema analizi RE, dok prema analizi RE istovremenu sličnost u dva klastera pokazuju vrste *C. angulata*, *C. virginica* i *O. edulis*. Sve četiri vrste prema analizi RE istovremeno pokazuju sličnost u 57 klastera, a prema usporedbi programom *Geneious* u 21 klasteru. Potencijalno je vrsno specifičnih satDNA prema analizi RE u genomu vrste *C. virginica* njih 13, a u genomu vrste *O. edulis* njih 35. Prema usporedbi programom *Geneious*, vrsta *C. gigas* ima potencijalno osam vrsno specifičnih satDNA, vrsta *C. angulata* njih 15, *C. virginica* njih 19, a vrsta *O. edulis* njih 41.

Sve analizirane vrste dijele određen broj satDNA kao što predviđa i teorija satelitne biblioteke (Salser i sur., 1976., Fry i Salser, 1977., Meštrović i sur., 1998.). Vrste *C. gigas* i *C. angulata* najrodnije su i dijele najveći broj zajedničkih satDNA, a te dvije vrste dijele veći broj zajedničkih satDNA s vrstom *O. edulis* nego s vrstom *C. virginica* iako s potonjom pokazuju manju udaljenost u filogenetskim analizama baziranim na sekvencijama rDNA (Salvi i sur., 2014.). De Lima i Ruiz-Ruano (2022.) odredili su satelitome i potvrdili teoriju satelitne biblioteke za više od 30 vrsta roda *Drosophila* te su za neke primijetili pojavu novih satDNA u kratkim evolucijskim periodima, odnosno među srodnijim vrstama, zbog čega im je bilo teško potvrditi finalni set satDNA koji bi obuhvaćao bibliotku u svim analiziranim vrstama. Budući da su prema nekim autorima (López-Flores i sur., 2004., Reece i sur., 2008. Gagnaire i sur., 2018.) *C. gigas* i *C. angulata* zapravo podvrste iste vrste, ne čude rezultati analize RE koji pokazuju da dijele najveći broj zajedničkih satDNA prema analizi RE i prema usporedbi programom *Geneious*. U vrstama *C. gigas* i *C. angulata* satDNA Cg170/HindIII postoji u dvije varijante u genomu (duljina 164 i 166 pb), dok je u vrsti *O. edulis* prisutna u samo jednoj varijanti (duljine 166 pb). U vrsti *C. virginica* ta satDNA naizgled uopće nije detektirana, no unatoč tome što nije klasificirana kao satDNA, postojanje velikog broja očitanja iz vrste *C. virginica* u klasteru C11 kojeg dijeli s ostalima (Tablica 24.), ukazuje na to da sekvencije koje pripadaju toj satDNA postoje u genomu i ove vrste. Navedene razlike u satelitomima mogu biti posljedica nemogućnosti detekcije svih satDNA u pojedinačnim satelitomima zbog značajne prisutnosti *kombiniranih elemenata* (element Helitron/Helentron) u genomu kamenica (rezultati ovog rada, Šatović Vukšić i Plohl, 2021.a i b.), pojave novih sekvencija satDNA u genomima nekih vrsta, brže evolucije monomernih sekvencija

satDNA u pojedinim genomima (de Lima i Ruiz-Ruano, 2022.) ili je njihova zastupljenost u genomu niža od one koja se mogla detektirati analizama u ovom radu. Analizom TAREAN nije moguće razlikovati pojedinačne i malobrojne nukleotidne supstitucije u sekvencijama sličnih, no, ipak, različitih satDNA te se zbog toga može dogoditi da više takvih sličnih satDNA budu okarakterizirane kao jedna jedinstvena satDNA (de Lima i Ruiz-Ruano, 2022.). SatDNA koje u svojoj sekvenciji sadrže i dijelove postranih regija pokretnih elemenata (npr. Helitrona/Helentrona) čiji su sastavni dio mogu ostati neklasificirane te zbog toga ostanu izuzete iz ukupnog satelitoma u genomu neke vrste (Šatović Vukšić i Plohl, 2021.a). No, kao što je vidljivo u Tablici 24., prisutnost tih sekvencija u pojedinom genomu može biti utvrđena njihovom prisutnošću u pojedinom klasteru, ukoliko ga dijele sa sekvencijama već definiranim kao satDNA u (nekim ili svim) ostalim analiziranim vrstama.

Slično tome, u genomu vrste *Rhodnius prolixus* u kojoj je satelitom određivan analizom RE, četiri porodice satDNA otprije poznate za tu vrstu nisu detektirane ovim tipom analize (Montiel i sur., 2021.). To bi potencijalno mogla biti posljedica njihove povezanosti s nekim drugim sekvencijama, moguće onima koje imaju porijeklo od PGE, te su zbog toga klasificirane kao neki drugi tip ponavljajućih sekvencija (Šatović Vukšić i Plohl, 2021.a).

## **6. ZAKLJUČAK**



1. Satelitom vrste *C. gigas* čine 52 satDNA od kojih je najbrojnija Cg170 s udjelom u genomu od 1,76 %.
2. Sekvencije koje većim dijelom grade centromeru *C. gigas* u jednakim su udjelima satDNA i DNA transpozoni. Na centromerama se ne može pronaći samo jedna centromerno-specifična sekvencija koja bi obuhvaćala centromerna područja svih kromosoma ove vrste, nego se više različitih ponavljajućih sekvencija nalazi na limitiranom broju kromosoma. Sekvencije koje grade centromeru vrste *C. gigas* na metafaznim kromosomima lokaliziraju ne samo centromerno već i raspršeno po krakovima kromosoma.
3. Protein Cg-CenH3 pokazuje distribuciju specifičnu za stadij na različitim razvojnim fazama muških spolnih stanica vrste *C. gigas*.
4. Sekvencije koje grade heterokromatin obilježen trimetilacijom Lys9 u histonu H3 smještene su većim dijelom telomerno na jednom paru kromosoma i (peri)centromerno na drugom paru kromosoma. Heterokromatin vrste *C. gigas* obilježen trimetilacijom Lys9 na histonu H3 dominantno je građen od DNA transpozona.
5. Satelitomskim analizama utvrđeno je da satelitom vrste *C. angulata* čini 61 satDNA, vrste *C. virginica* 33, a satelitom vrste *O. edulis* 52 satDNA. Sve četiri vrste kamenica dijele neke zajedničke satDNA, uz postojanje potencijalno vrsno-specifičnih satDNA, a najsirođnije vrste, *C. gigas* i *C. angulata*, dijele najveći broj zajedničkih satDNA.
6. SatDNA koje su pokazale sličnost s elementima Helitron čine većinu satelitoma vrste *C. gigas*. Samo jedna satDNA u genomu se nalazi isključivo vezana uz element. Za dvije satDNA koje pokazuju sličnost s elementima Helitron utvrđeno je da su većim brojem svojih monomera/nizova u genomu prisutne u samostalnom obliku, a manjim vezane uz element. Preostalih devet koje su pokazale sličnost s elementima Helitron u genomu je prisutno dominantno vezano uz element.
7. U vrsti *C. gigas* nije otkriveno tipično nakupljanje satDNA na određenim kromosomskim pozicijama, čime je u ovoj vrsti ustanovljen novi (visokoraspršeni) obrazac organizacije satDNA na razini genoma.



## **7. REFERENCE**





- Adega, F., Guedes-Pinto, H. and Chaves, R. (2009) ‘Satellite DNA in the karyotype evolution of domestic animals-Clinical considerations’, *Cytogenetic and Genome Research*, 126(1–2), pp. 12–20. doi: 10.1159/000245903.
- Ahmad, S. F., Worapong, S., Jehangir, M., Suntronpong, A., Panthum, T., Malaivijitnond, S., Srikulnath, K. (2020) ‘Dark Matter of Primate Genomes: Satellite DNA Repeats and Their Evolutionary Dynamics’, *Cells*, 9(12), pp. 1–34. doi: 10.3390/cells9122714.
- Alkan, C., Ventura, M., Archidiacono, N., Rocchi, M., Cenk Sahinalp, S., Eichler, E. E. (2007) ‘Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data’, *PLoS Computational Biology*, 3(9), pp. 1807–1818. doi: 10.1371/journal.pcbi.0030181.
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., Gouil, Q. (2020) ‘Opportunities and challenges in long-read sequencing data analysis-Genome Biology-Full Text’, *Genome Biology*. *Genome Biology*, 21(30), pp. 1–16.
- Arkhipova, I. R., Pyatkova, K. I., Meselson, M., Evgenev, M. B. (2003) ‘Retroelements containing introns in diverse invertebrate taxa’, *Nature Genetics*, 33(2), pp. 123–124. doi: 10.1038/ng1074.
- Arkhipova, I. R., Yushenova, I. A. and Angert, E. (2019) ‘Giant Transposons in Eukaryotes: Is Bigger Better?’, *Genome Biology and Evolution*, 11(3), pp. 906–918. doi: 10.1093/gbe/evz041.
- Bachmann, L., Raab, M., Schibel, J., Sperlich, D. (1991) ‘Satellite DNA’, in Hewitt, G. M., Johnston, A. W. B., and Young, J. P. W. (eds) *Molecular Techniques in Taxonomy*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 133–143. doi: 10.1007/978-3-642-83962-7\_9.
- Brammer, S.P., Vasconcelos, S., Balvedi Poersch, L., Oliveira, A. R., Brasileiro-Vidal, A. C. (2013) ‘Genomic *in situ* hybridization in Triticeae: A methodological approach’, *Plant Breeding from Laboratories to Fields*, InTech. doi: 10.5772/52928.
- Bao, W. and Jurka, J. (2013) ‘Homologues of bacterial TnpB-IS605 are widespread in diverse eukaryotic transposable elements’, *Mobile DNA*. *Mobile DNA*, 4(1), p. 1. doi: 10.1186/1759-8753-4-12.
- Barbaglia, A. M., Klusman, K. M., Higgins, J., Shaw, J. R., Curtis Hannah, L., Lal, S. K. (2012) ‘Gene capture by Helitron transposons reshuffles the transcriptome of maize’, *Genetics*, 190(3), pp. 965–975. doi: 10.1534/genetics.111.136176.

- Barra, V. and Fachinetti, D. (2018) ‘The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA’, *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-06545-y.
- Batistoni, R., Pesole, G., Marracci, S., Nardi, I. (1995) ‘A tandemly repeated DNA family originated from SINE-related elements in the European plethodontid salamanders (Amphibia, Urodela)’, *Journal of Molecular Evolution*, 40(6), pp. 608–615. doi: 10.1007/BF00160508.
- Belyayev, A., Josefiová, J., Jandovaá, M., Kalendar, R., Krak, K., Mandák, B. (2019) ‘Natural history of a satellite DNA family: From the ancestral genome component to species-specific sequences, concerted and non-concerted evolution’, *International Journal of Molecular Sciences*, 20(5). doi: 10.3390/ijms20051201.
- Benedum, U. M., Neitzel, H., Sperling, K., Bogenberger, J., Fittler, F. (1986) ‘Organization and chromosomal distribution of a novel repetitive DNA component from *Muntiacus muntjak vaginalis* with a repeat length of more than 40 kb’, *Chromosoma*, 94(4), pp. 267–272. doi: 10.1007/BF00290855.
- Bieler, R., Mikkelsen, P. M., Collins, T. M., Glover, E. A., González, V. L., Graf, D. L., Harper, E. M., Healy, J., Kawauchi, G. Y., Sharma, P. P., Staubach, S., Strong, E. E., Taylor, J. D., Tëmkin, I., Zardus, J.D., Clark, S., Guzmán, A., McIntyre, E., Sharp, P., Giribet, G. (2014) ‘Investigating the Bivalve Tree of Life – an exemplar-based approach combining molecular and novel morphological characters’, *Invertebrate Systematics*, 28, pp. 32–115. doi: <http://dx.doi.org/10.1071/IS13010>.
- Birky, C. W. J. and Skavaril, R. V (1976) ‘Maintenance of genetic homogeneity in systems with multiple genomes.’, *Genetical research*. England, 27(2), pp. 249–265. doi: 10.1017/s001667230001644x.
- Biscotti, M. A., Olmo, E. and Heslop-Harrison, J. S. P. (2015) ‘Repetitive DNA in eukaryotic genomes.’, *Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*. Netherlands, pp. 415–420. doi: 10.1007/s10577-015-9499-z.
- Blower, M. D., Sullivan, B. A. and Karpen, G. H. (2002) ‘Conserved organization of centromeric chromatin in flies and humans.’, *Developmental cell*. United States, 2(3), pp. 319–330. doi:

10.1016/s1534-5807(02)00135-1.Boeke, J. D. *et al.* (1985) ‘Ty elements transpose through an RNA intermediate.’, *Cell*. United States, 40(3), pp. 491–500. doi: 10.1016/0092-8674(85)90197-7.

Boštjančić, L. L., Bonassin, L., Anušić, L., Lovrenčić, L., Besendorfer, V., Maguire, I., Grandjean, F., Austin, C. M., Greve, C., Hamadou, A. B., Mlinarec, J. (2021) ‘The *Pontastacus leptodactylus* (Astacidae) Repeatome Provides Insight Into Genome Evolution and Reveals Remarkable Diversity of Satellite DNA’, *Frontiers in Genetics*, 11(January), p. 611745. doi: 10.3389/fgene.2020.611745.

Bouilly, K., Chaves, R., Leitão, A., Benabdelmouna, A., Guedes-Pinto, H. (2008) ‘Chromosomal organization of simple sequence repeats in chromosome patterns’, *Journal of Genetics*, 87(2), pp. 119–125. Bourque, G. *et al.* (2018) ‘Ten things you should know about transposable elements 06 Biological Sciences 0604 Genetics’, *Genome Biology*. *Genome Biology*, 19(1), pp. 1–12. doi: 10.1186/s13059-018-1577-z.

Brajković, J. *et al.* (2012) ‘Satellite DNA-like elements associated with genes within euchromatin of the beetle *tribolium castaneum*’, *G3: Genes, Genomes, Genetics*, 2(8), pp. 931–941. doi: 10.1534/g3.112.003467.

Brajković, J., Feliciello, I., Bruvo-Madžarić, B., Ugarković, Đ. (2018) ‘Dispersion profiles and gene associations of repetitive DNAs in the euchromatin of the beetle *Tribolium castaneum*’, *G3: Genes, Genomes, Genetics*, 8(3), pp. 875–886. doi: 10.1534/g3.117.300267.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D’Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M. C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, J., Dvorak, J., McCombie, W. R., Hall, A., Mayer, K. F. X., Edwards, K. J., Bevan, M. W., Hall, N. (2012) ‘Analysis of the bread wheat genome using whole-genome shotgun sequencing’, *Nature*. Nature Publishing Group, 491(7426), pp. 705–710. doi: 10.1038/nature11650.

Britten, R. J. and Kohne, D. E. (1968) ‘Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms.’, *Science (New York, N.Y.)*. United States, 161(3841), pp. 529–540. doi: 10.1126/science.161.3841.529.

Brutlag, D., Fry, K., Nelson, T., Hung, P. (1977) ‘Synthesis of hybrid bacterial plasmids containing highly repeated satellite DNA’, *Cell*, 10(3), pp. 509–519. doi: 10.1016/0092-8674(77)90038-1.

- Camacho, J.P.M., Ruiz-Ruano, F.J., Martín-Blázquez, R., López-León, M. D., Cabrero, J., Lorite, P., Cabral-de-Mello, D. C., Bakkali, M. A step to the gigantic genome of the desert locust: chromosome sizes and repeated DNAs. (2015) *Chromosoma* **124**, 263–275. <https://doi.org/10.1007/s00412-014-0499-0>.
- Carbone, L., Harris, R. A., Mootnick, A. R., Milosavljević, A., Martin, D. I. K., Rocchi, M., Capozzi, O., Archidiacono, N., Konkel, M. K., Walker, J. A., Batzer, M. A., de Jong, P. J. (2012) ‘Centromere remodeling in Hoolock leuconedys (Hylobatidae) by a new transposable element unique to the gibbons’, *Genome Biology and Evolution*, *4*(7), pp. 648–658. doi: 10.1093/gbe/evs048.
- Chang, C. H., Chavan, A., Palladino, J., Wei, X., Martins, N. M. C., Santinello, B., Chen, C. C., Erceg, J., Beliveau, B. J., Wu, C. T., Larracunte, A. M., Mellone, B. (2019) ‘Islands of retroelements are major components of *Drosophila* centromeres’, *PLoS Biology*. doi: 10.1371/journal.pbio.3000241.
- Chaves, R., Ferreira, D., Mendes-da-Silva, A., Meles, S., Adegas, F. (2017) ‘FA-SAT is an old satellite DNA frozen in several bilateria genomes’, *Genome Biology and Evolution*, *9*(11), pp. 3073–3087. doi: 10.1093/gbe/evx212.
- Chen, C., Ara, T. and Gautheret, D. (2009) ‘Using Alu elements as polyadenylation sites: A case of retroposon exaptation’, *Molecular Biology and Evolution*, *26*(2), pp. 327–334. doi: 10.1093/molbev/msn249.
- Chen, L. L., DeCerbo, J. N. and Carmichael, G. G. (2008) ‘Alu element-mediated gene silencing’, *EMBO Journal*, *27*(12), pp. 1694–1705. doi: 10.1038/emboj.2008.94.
- Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C. R., Gu, M., Blattner, F. R., Jiang, J. (2002) ‘Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon’, *Plant Cell*, *14*(8), pp. 1691–1704. doi: 10.1105/tpc.003079.
- Cheung, S., Manhas, S. and Measday, V. (2018) ‘Retrotransposon targeting to RNA polymerase III-transcribed genes’, *Mobile DNA*. *Mobile DNA*, *9*(1), pp. 1–15. doi: 10.1186/s13100-018-0119-2.
- Cho, Y., Ideue, T., Nagayama, M., Araki, N., Tani., T. (2018) ‘RBMX is a component of the centromere noncoding RNP complex involved in cohesion regulation’, *Genes to Cells*, *23*(3), pp. 172–184. doi: 10.1111/gtc.12562.

- Clabby, C., Goswami, U., Flavin, F., Wilkins, N. P., Houghton, J. A., Powell, R. (1996) 'Cloning, characterization and chromosomal location of a satellite DNA from the pacific oyster, *Crassostrea gigas*', *Gene*, 168(2), pp. 205–209. doi: 10.1016/0378-1119(95)00749-0.
- Coates, B. S., Hellmich, R. L., Grant, D. M., Abel, C. A. (2012) 'Mobilizing the genome of lepidoptera through novel sequence gains and end creation by non-autonomous lep1 helitrons', *DNA Research*, 19(1), pp. 11–21. doi: 10.1093/dnares/dsr038.
- Cohen, S., Agmon, N., Sobol, O., Segal, D. (2010) 'Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells', *Mobile DNA*, 1(1), pp. 1–11. doi: 10.1186/1759-8753-1-11.
- Colaco, S. and Modi, D. (2018) 'Genetics of the human Y chromosome and its association with male infertility', *Reproductive Biology and Endocrinology*. *Reproductive Biology and Endocrinology*, 16(1), pp. 1–24. doi: 10.1186/s12958-018-0330-5.
- Cross, I., Díaz, E., Sánchez, I., Rebordinos, L. (2005) 'Molecular and cytogenetic characterization of *Crassostrea angulata* chromosomes', *Aquaculture*, 247(1–4), pp. 135–144. doi: 10.1016/j.aquaculture.2005.02.039.
- Dawe, R. K. and Henikoff, S. (2006) 'Centromeres put epigenetics in the driver's seat.', *Trends in biochemical sciences*. England, 31(12), pp. 662–669. doi: 10.1016/j.tibs.2006.10.004.
- De Lima, L. G., Svartman, M. and Kuhn, G. C. S. (2017) 'Dissecting the Satellite DNA Landscape in Three Cactophilic *Drosophila* Sequenced Genomes.', *G3 (Bethesda, Md.)*. England, 7(8), pp. 2831–2843. doi: 10.1534/g3.117.042093.
- De Lima, L. G. and Ruiz-Ruano, F. J. (2022) 'In-Depth Satellitome Analyses of 37 *Drosophila* Species Illuminate Repetitive DNA Evolution in the *Drosophila* Genus', *Genome Biology and Evolution*, 14(5), pp. 1–19. doi: 10.1093/gbe/evac064.
- De Silva, D. M. Z. A., Utsunomia, R., Ruiz-Ruano, F. J., Daniel, S. N., Porto-Foresti, F., Teruo Hashimoto, D., Oliveira, C., Camacho, J. P. M., Forest, F. (2017) 'High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*', *Scientific Reports*, 7(1), pp. 1–12. doi: 10.1038/s41598-017-12939-7.

- De Sotero-Caio, C. G., Cavalcanti Cabral-de-Mello, D., da Silva Calixto, M., Targino Valente, G., Martins, C., Loreto, V., de Souza, M. J., Santos, N. (2017) ‘Centromeric enrichment of LINE-1 retrotransposons and its significance for the chromosome evolution of Phyllostomid bats’, *Chromosome Research*. *Chromosome Research*, 25(3–4), pp. 313–325. doi: 10.1007/s10577-017-9565-9.
- Del Pilar Garcillán-Barcia, M., Bernales, I., Mendiola, M. V., de la Cruz, F. (2001) ‘Single-stranded DNA intermediates in IS91 rolling-circle transposition’, *Molecular Microbiology*, 39(2), pp. 494-501. doi: 10.1046/j.1365-2958.2001.02261.x.
- Dias, G. B., Heringer, P., Svartman, M., Kuhn, G. C. S. (2015) ‘Helitrons shaping the genomic architecture of *Drosophila*: enrichment of DINE-TR1 in  $\alpha$ - and  $\beta$ -heterochromatin, satellite DNA emergence, and piRNA expression’, *Chromosome Research*, 23(3), pp. 597–613. doi: 10.1007/s10577-015-9480-x.
- Dong, Y., Lu, X., Song, W., Shi, L., Zhang, M., Zhao, H., Jiao, Y., Lai, J. (2011) ‘Structural characterization of helitrons and their stepwise capturing of gene fragments in the maize genome’, *BMC Genomics*, 12(1), p. 609. doi: 10.1186/1471-2164-12-609.
- Dover, G. (1982) ‘Molecular drive: a cohesive mode of species evolution’, *Nature*, 299(5879), pp. 111–117. doi: 10.1038/299111a0.
- Du, C., Caronna, J., He, L., Dooner, H. K. (2008) ‘Computational prediction and molecular confirmation of Helitron transposons in the maize genome’, *BMC Genomics*, 9, pp. 1–10. doi: 10.1186/1471-2164-9-51.
- Eickbush, T. H. (1992) ‘Transposing without ends: the non-LTR retrotransposable elements’, *The New biologist*, 4(5), p. 430—440. Available at: <http://europepmc.org/abstract/MED/1325183>.
- Elder, J. F. J. and Turner, B. J. (1995) ‘Concerted evolution of repetitive DNA sequences in eukaryotes.’, *The Quarterly review of biology*. United States, 70(3), pp. 297–320. doi: 10.1086/419073.
- Elliott, T. A. and Gregory, T. R. (2015) ‘What’s in a genome? The C-value enigma and the evolution of eukaryotic genome content’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1678). doi: 10.1098/rstb.2014.0331.

- Eymery, A., Callanan, M. and Vourc'h, C. (2009) 'The secret message of heterochromatin: New insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription', *International Journal of Developmental Biology*, 53(2–3), pp. 259–268. doi: 10.1387/ijdb.082673ae.
- Ezgeta-Balić, D., Šegvić-Bubić, T., Stagličić, N., Lin, Y., Bojanić Varezić, D., Grubišić, L. i Briski, E. (2019). Distribution of non-native Pacific oyster *Magallana gigas*(Thunberg, 1793) along the eastern Adriatic coast. *Acta Adriatica*, 60 (2), 137-146. <https://doi.org/10.32582/aa.60.2.3>.
- Fang, Y., Coehlo, M. A., Shu, H., Schotanus, K., Thimmappa, B. C., Yadav, V., Chen, H., Malc, E. P., Wang, J., Mieczkowski, P. A., Kronmiller, B., Tyler, B. M., Sanyal, K., Dong, S., Nowrousian, M., Heitman, J. (2020) 'Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages', *PLoS Genetics*, 16(3), pp. 1–30. doi: 10.1371/journal.pgen.1008646.
- Feliciello, I., Picariello, O. and Chinali, G. (2006) 'Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: Molecular evidence of a new mechanism of DNA repair acting on satellite DNA', *Gene*, 383, pp. 81–92. doi: 10.1016/j.gene.2006.07.016.
- Feliciello, I., Akrap, I. and Ugarković, Đ. (2015) 'Satellite DNA Modulates Gene Expression in the Beetle *Tribolium castaneum* after Heat Stress', *PLoS Genetics*, 11(8), pp. 1–18. doi: 10.1371/journal.pgen.1005466.
- Feng, Q., Moran, J. V., Kazazian Jr., H. H., Boeke, J. D. (1996) 'Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition', *Cell*, 87(5), pp. 905–916. doi: 10.1016/S0092-8674(00)81997-2.
- Fernández Robledo, J. A., Yadavalli, R., Allam, B., Pales Espinosa, E., Gerdol, M., Greco, S., Stevick, R. J., Gómez-Chiarri, Zhang, Y., Heil, C. A., Tracy, A. N., Bishop-Bailey, D., Metzger, M. J. (2019) 'From the raw bar to the bench: Bivalves as models for human health', *Developmental and Comparative Immunology*, 92, pp. 260–282. doi: 10.1016/j.dci.2018.11.020.
- Ferree, P. M. and Barbash, D. A. (2009) 'Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*', *PLoS Biology*, 7(10). doi: 10.1371/journal.pbio.1000234.



- Feschotte, C. and Wessler, S.R. (2001) 'Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes', *Proceedings of the National Academy of Sciences*, 98(16):8923-4. doi: 10.1073/pnas.171326198.
- Feschotte, C., Jiang, N. and Wessler, S. R. (2002) 'Plant transposable elements: Where genetics meets genomics', *Nature Reviews Genetics*, 3(5), pp. 329–341. doi: 10.1038/nrg793.
- Finnegan, D.J. (1989) 'Eukaryotic transposable elements and genome evolution' *Trends in Genetics*, 5(4):103-7. doi: 10.1016/0168-9525(89)90039-5.
- Franco, A., Berthelin, C. H., Goux, D., Sourdain, P., Mathieu, M. (2008) 'Fine structure of the early stages of spermatogenesis in the Pacific oyster, *Crassostrea gigas* (Mollusca, Bivalvia)', *Tissue and Cell*, 40(4), pp. 251–260. doi: 10.1016/j.tice.2007.12.006.
- Fry, K. and Salser, W. (1977) 'Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents.', *Cell*. United States, 12(4), pp. 1069–1084. doi: 10.1016/0092-8674(77)90170-2.
- Gaffney, P. M., Pierce, J. C., Mackinley, A. G., Titchen, D. A., Glenn, W. K. (2003) 'Pearl, a novel family of putative transposable elements in bivalve mollusks.', *Journal of molecular evolution*, 56(3), pp. 308–316. doi: 10.1007/s00239-002-2402-5.
- García-Souto, D., Troncoso, T., Perez, M., Pasantes, J. J. (2015) 'Molecular cytogenetic analysis of the European hake *Merluccius merluccius* (Merlucciidae, Gadiformes): U1 and U2 snRNA gene clusters map to the same location', *PLoS ONE*, 10(12), pp. 1–11. doi: 10.1371/journal.pone.0146150.
- García-Souto, D., Qarkaxhija, V. and Pasantes, J. J. (2017a) 'Resolving the taxonomic status of *Chamelea gallina* and *C. striatula* (Veneridae, Bivalvia): A combined molecular cytogenetic and phylogenetic approach', *BioMed Research International*, p. 7638790. doi: 10.1155/2017/7638790.
- García-Souto, D., Sumner-Hempel, Fervenza, S., Pérez-García, C., Torreiro, A., González-Romero, R., Eirín-López, J. M., Morán, P., Pasantes, J. J. (2017b) 'Detection of invasive and cryptic species in marine mussels (Bivalvia, Mytilidae): A chromosomal perspective', *Journal for Nature Conservation*. Elsevier GmbH., 39, pp. 58–67. doi: 10.1016/j.jnc.2017.07.005.

- García-Souto, D., Mravinac, B., Šatović, E., Plohl, M., Morán, P., Pasantes, J. J. (2017c) ‘Methylation profile of a satellite DNA constituting the intercalary G+C-rich heterochromatin of the cut trough shell *Spisula subtruncata* (Bivalvia, Mactridae)’, *Scientific Reports*, 7(1), pp. 1–12. doi: 10.1038/s41598-017-07231-7.
- García-Souto, D. and Pasantes, J. J. (2018) ‘Cytogenetics in *Arctica islandica* (Bivalvia, arctidae): The longest lived non-colonial metazoan’, *Genes*, 9(6), p. 9060299. doi: 10.3390/genes9060299.
- Garrido-Ramos, M. A. (2017) ‘Satellite DNA: An evolving topic’, *Genes*, 8(9). doi: 10.3390/genes8090230.
- Garrigues, J. M., Tsu, B. V., Daugherty, M. D., Pasquinelli, A. E. (2019) ‘Diversification of the *Caenorhabditis* heat shock response by helitron transposable elements’, *eLife*, 8, pp. 1–25. doi: 10.7554/eLife.51139.
- Gemayel, R. *et al.* (2010) ‘Variable tandem repeats accelerate evolution of coding and regulatory sequences’, *Annual Review of Genetics*, 44, pp. 445–477. doi: 10.1146/annurev-genet-072610-155046.
- Gerdol, M. and Venier, P. (2015) ‘An updated molecular basis for mussel immunity.’, *Fish & shellfish immunology*. England, 46(1), pp. 17–38. doi: 10.1016/j.fsi.2015.02.013.
- Giles, N. H. and Riley, H. P. (1950) ‘Studies on the mechanism of the oxygen effect on the radiosensitivity of *Tradescantia* chromosomes.’, *Proceedings of the National Academy of Sciences of the United States of America*, 36(6), pp. 337–344. doi: 10.1073/pnas.36.6.337.
- Glöckner, G., Szafranski, K., Winckler, T., Dingermann, T., Quail, M. A., Cox, E., Eichinger, L., Noegel, A. A., Rosenthal, A. (2001) ‘The complex repeats of *Dictyostelium discoideum*’, *Genome Research*, 11(4), pp. 585–594. doi: 10.1101/gr.GR-1622RR.
- Glöckner, G. and Heidel, A. J. (2009) ‘Centromere sequence and dynamics in *Dictyostelium discoideum*.’, *Nucleic acids research*. England, 37(6), pp. 1809–1816. doi: 10.1093/nar/gkp017.
- Gomes-dos-Santos, A., Lopes-Lima, Castro, L. F. C., Froufe, E. (2020) ‘Molluscan genomics: the road so far and the way forward’, *Hydrobiologia*. Springer International Publishing, 847(7), pp. 1705–1726. doi: 10.1007/s10750-019-04111-1.

- Goodwin, T. J. D., Butler, M. I. and Poulter, R. T. M. (2003) ‘Cryptons: A group of tyrosine-recombinase- encoding DNA transposons from pathogenic fungi’, *Microbiology*, 149(11), pp. 3099–3109. doi: 10.1099/mic.0.26529-0.
- Gosálvez, J., Kjelland, M. E. and López-Fernández, C. (2010) ‘Sperm DNA in grasshoppers: Structural features and fertility implications’, *Journal of Orthoptera Research*, 19(2), pp. 243–252. doi: 10.1665/034.019.0210.
- Gotea, V. and Makałowski, W. (2006) ‘Do transposable elements really contribute to proteomes?’, *Trends in Genetics*, 22(5), pp. 260–267. doi: 10.1016/j.tig.2006.03.006.
- Goubert, C., Modolo, L., Vieira, C., Valiente Moro, C., Mavingui, P., Boulesteix, M. (2015) ‘De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*)’, *Genome Biology and Evolution*, 7(4), pp. 1192–1205. doi: 10.1093/gbe/evv050.
- Grabundzija, I., Hickman, A. B. and Dyda, F. (2018) ‘Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition’, *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-03688-w.
- Gržan, T., Despot-Slade, E., Meštrović, N., Plohl, M., Mravinac, B. (2020) ‘CenH3 distribution reveals extended centromeres in the model beetle *Tribolium castaneum*’, *PLoS Genetics*, 16(10), pp. 1–28. doi: 10.1371/journal.pgen.1009115.
- Gunter, G. (1950) ‘The Generic Status of Living Oysters and the Scientific Name of the Common American Species’, *American Midland Naturalist*, 43(2), p. 438. doi: 10.2307/2421912.
- Guo, X., Gao, J., Li, F., Wang, J. (2014) ‘Evidence of horizontal transfer of non-autonomous Lep1 Helitrons facilitated by host-parasite interactions’, *Scientific Reports*, 4, pp. 1–8. doi: 10.1038/srep05119.
- Han, M. J., Xu, H. E., Zhang, H. H., Feschotte, C., Zhang, Z. (2014) ‘Spy: A new group of eukaryotic DNA transposons without target site duplications’, *Genome Biology and Evolution*, 6(7), pp. 1748–1757. doi: 10.1093/gbe/evu140.

- Hancks, D. C. and Kazazian, H. H. (2016) ‘Roles for retrotransposon insertions in human disease’, *Mobile DNA*. *Mobile DNA*, 7(1). doi: 10.1186/s13100-016-0065-9.
- He, C. and Bonasio, R. (2017) ‘Chromatin Mapping: A cut above’, *eLife*, 6, e25000. doi: [10.7554/eLife.25000](https://doi.org/10.7554/eLife.25000)
- Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., Novák, P., Neumann, P., Taudien, S., Platzer, M., Houben, A. (2013) ‘The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization’, *Plant Journal*, 73(4), pp. 555–565. doi: 10.1111/tpj.12054.
- Heikkinen, E., Launonen, V., Müller, E., Bachmann, L. (1995) ‘The pvB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements’, *Journal of Molecular Evolution*, 41(5), pp. 604–614. doi: 10.1007/BF00175819.
- Heitz, E. (1928) ‘*Das heterochromatin der moose*’, *Bornträger*.
- Henikoff, S., Ahmad, K. and Malik, H. S. (2001) ‘The centromere paradox: Stable inheritance with rapidly evolving DNA’, *Science*, 293(5532), pp. 1098–1102. doi: 10.1126/science.1062939.
- Heslop-Harrison, J. S. P. and Schwarzacher, T. (2011) ‘Organisation of the plant genome in chromosomes.’, *The Plant journal : for cell and molecular biology*. England, 66(1), pp. 18–33. doi: 10.1111/j.1365-313X.2011.04544.x.
- Heslop-Harrison, J. S. P. and Schwarzacher, T. (2013) ‘Nucleosomes and centromeric DNA packaging.’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(50), pp. 19974–19975. doi: 10.1073/pnas.1319945110.
- Hewitt, J., Morrison, M., Spigel, R., Turner, S. (2006) ‘Oyster Farming in the Kaipara Harbour Proposed Aquaculture Management Area D’, *Ecological Assessment Prepared for Biomarine Ltd.* NIWA, Client Report: HAM2006-118.
- Hikosaka, A. and Kawahara, A. (2004) ‘Lineage-specific tandem repeats riding on a transposable element of MITE in *Xenopus* evolution: A new mechanism for creating simple sequence repeats’, *Journal of Molecular Evolution*, 59(6), pp. 738–746. doi: 10.1007/s00239-004-2664-1.

Hoskins, R. A., Smith, D. A., Carlson, J. W., Carvalho, A. B., Halpern, A., Kaminker, J. S., Kennedy, C., Mungall, C. J., Sullivan, B. A., Sutton, G. G., Yasuhara, J. C., Wakimoto, B. T., Myers, E. W., Celniker, S. E., Rubin, G. M., Karpen, G. H. (2002) ‘Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly.’, *Genome biology*, 3(12), pp. 1–16. doi: 10.1186/gb-2002-3-12-research0085.

Houben, A., Schroeder-Reiter, E., Nagaki, K., Nasuda, S., Wanner, G., Murata, M., Endo, T. R. (2007) ‘CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley’, *Chromosoma*, 116(3), pp. 275–283. doi: 10.1007/s00412-007-0102-z.

Hu, K., Xu, K., Yi, Shen, J., Ma, C., Fu, T., Ouyang, Y., Tu, J. (2019) ‘Helitron distribution in *Brassicaceae* and whole genome Helitron density as a character for distinguishing plant species’, *BMC Bioinformatics*. *BMC Bioinformatics*, 20(1), pp. 1–20. doi: 10.1186/s12859-019-2945-8.

Huang, K., Li, C. F., Wu, J., Wei, J. H., Zou, Y., Han, M. J., Zhou, Z. Y. (2016) ‘Enhancer activity of Helitron in sericin-1 gene promoter from *Bombyx mori*’, *Insect Science*, 23(3), pp. 396–405. doi: 10.1111/1744-7917.12347.

Huber, M. (2010) *Compendium of bivalves. A full-color guide to 3,300 of the world's marine bivalves. A status on bivalvia after 250 years of research. ConchBooks, p. 23.*

Izsvák, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., Hackett, P. B. (1999) ‘Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification’, *Journal of Molecular Evolution*, 48(1), pp. 13–21. doi: 10.1007/PL00006440.

Jenuwein, T. and Allis, C. D. (2001) ‘Translating the histone code.’, *Science (New York, N.Y.)*. United States, 293(5532), pp. 1074–1080. doi: 10.1126/science.1063127.

Jeppesen, P. and Turner, B. M. (1993) ‘The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression.’, *Cell*. United States, 74(2), pp. 281–289. doi: 10.1016/0092-8674(93)90419-q.

Juan, C., Vazquez, P., Rubio, J. M., Petitpierre, E., Hewitt, G. M. (1993) ‘Presence of highly repetitive DNA sequences in *Tribolium* flour-beetles’, *Heredity*, 70(1), pp. 1–8. doi: 10.1038/hdy.1993.1.

Jurka, J. (1997) 'Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons', *Proceedings of the National Academy of Sciences of the United States of America*, 94(5), pp. 1872–1877. doi: 10.1073/pnas.94.5.1872.

Jurka, J. (2012) 'Satellite DNA from the Pacific oyster genome SAT-1\_CGi', *Rebase Reports* 12:2722.

Kapitonov, V. V., Holmquist, G. P. and Jurka, J. (1998) 'L1 repeat is a basic unit of heterochromatin satellites in cetaceans.', *Molecular biology and evolution*, 15(5), pp. 611–612. doi: 10.1093/oxfordjournals.molbev.a025963.

Kapitonov, V. V. and Jurka, J. (1999) 'Molecular paleontology of transposable elements from *Arabidopsis thaliana*', *Genetica*, 107(1–3), pp. 27–37.

Kapitonov, V. V. and Jurka, J. (2001) 'Rolling-circle transposons in eukaryotes', *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp. 8714–8719. doi: 10.1073/pnas.151269298.

Kapitonov, V. V. and Jurka, J. (2006) 'Self-synthesizing DNA transposons in eukaryotes', *Proceedings of the National Academy of Sciences of the United States of America*, 103(12), pp. 4540–4545. doi: 10.1073/pnas.0600833103.

Kapitonov, V. V. and Jurka, J. (2007) 'Helitrons on a roll: eukaryotic rolling-circle transposons', *Trends in Genetics*, 23(10), pp. 521–529. doi: 10.1016/j.tig.2007.08.004.

Kapitonov, V. V. and Jurka, J. (2008) 'A universal classification of eukaryotic transposable elements implemented in Rebase', *Nature Reviews Genetics*, 9(5), pp. 411–412. doi: 10.1038/nrg2165-c1.

Kapitonov, V. V. and Jurka, J. (2008) 'A universal classification of eukaryotic transposable elements implemented in Rebase.', *Nature reviews. Genetics*. England, pp. 411–2; author reply 414. doi: 10.1038/nrg2165-c1.

Kapoor, S., Zhu, L., Froyd, C., Liu, T., Rusche, L. N. (2015) 'Regional centromeres in the yeast *Candida lusitanae* lack pericentromeric heterochromatin', *Proceedings of the National Academy of Sciences of the United States of America*, 112(39), pp. 12139–12144. doi: 10.1073/pnas.1508749112.

Kim, S., Cho, C. S., Han, K., Lee, J. (2016) 'Structural Variation of Alu Element and Human Disease.', *Genomics & informatics*. Korea (South), 14(3), pp. 70–77. doi: 10.5808/GI.2016.14.3.70.

- Kim, Y. B., Oh, J. H., McIver, L. J., Rashkovetsky, E., Michalak, K., Garber., H. R., Kang, L., Nevo, E., Korol, A. B., Michalak, P. (2014) ‘Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel’, *Proceedings of the National Academy of Sciences of the United States of America*, 111(29), pp. 10630–10635. doi: 10.1073/pnas.1410372111.
- Kipling, D. and Warburton, P.E. (1997) ‘Centromeres, CENP-B and Tigger too’, *Trends in Genetics*, 13(4), pp. 1–5. Kit, S. (1961) ‘Equilibrium sedimentation in density gradients of DNA preparations from animal tissues’, *Journal of Molecular Biology*. Academic Press Inc. (London) Ltd., 3(6), pp. 711–716. doi: 10.1016/S0022-2836(61)80075-2.
- Koga, A., Hirai, Y., Terada, S., Jahan, I., Baicharoen, S., Arsaithamkul, V., Hirai, H. (2014) ‘Evolutionary origin of higher-order repeat structure in alpha-satellite DNA of primate centromeres’, *DNA research: an international journal for rapid publication of reports on genes and genomes*. England, 21(4), pp. 407–415. doi: 10.1093/dnares/dsu005.
- Kourtidis, A., Drosopoulou, E., Pantzartzu, C. N., Chintiroglou, C. C., Scouras, Z. G. (2006) ‘Three new satellite sequences and a mobile element found inside HSP70 introns of the Mediterranean mussel (*Mytilus galloprovincialis*)’, *Genome*, 49(11), pp. 1451–1458. doi: 10.1139/g06-111.
- Kowar, T., Zakrzewski, F., Macas, J., Kobližková, Viehoveer, P., Weisshaar, B., Schmidt, T. (2016) ‘Repeat Composition of CenH3-chromatin and H3K9me2-marked heterochromatin in Sugar Beet (*Beta vulgaris*)’, *BMC Plant Biology*, 16(1), pp. 1–16. doi: 10.1186/s12870-016-0805-5.
- Kramerov, D. A. and Vassetzky, N. S. (2011) ‘Origin and evolution of SINEs in eukaryotic genomes’, *Heredity*. England, 107(6), pp. 487–495. doi: 10.1038/hdy.2011.43.
- Krüger, J., and Vogel, F. (1975) ‘Population genetics of unequal crossing over’, *Journal of Molecular Evolution*, 4, pp. 201–247. doi: 10.1007/BF01732983.
- Kuhn, G. C. S., Küttler, H., Moreira-Filho, O., Heslop-Harrison, J. S. (2012) ‘The 1.688 repetitive DNA of drosophila: Concerted evolution at different genomic scales and association with genes’, *Molecular Biology and Evolution*, 29(1), pp. 7–11. doi: 10.1093/molbev/msr173.
- Kuhn, G. C. S. (2015) ‘Satellite DNA transcripts have diverse biological roles in *Drosophila*’, *Heredity*. England, pp. 1–2. doi: 10.1038/hdy.2015.12.

- Kumekawa, N., Hosouchi, T., Tsuruoka, H., Kotani, H. (2000) ‘The size and sequence organization of the centromeric region of arabidopsis thaliana chromosome 5’, *DNA Research*, 31;7(6):315-21. doi: 10.1093/dnares/7.6.315.
- Kumekawa, N., Hosouchi, T., Tsuruoka, H., Kotani, H. (2001) ‘The size and sequence organization of the centromeric region of arabidopsis thaliana chromosome 4’, *DNA Research*, 8(6), pp. 285–290. doi: 10.1093/dnares/8.6.285.
- Lander, E. S. *et al.* (2001) ‘Erratum: Initial sequencing and analysis of the human genome: International Human Genome Sequencing Consortium (Nature (2001) 409 (860-921))’, *Nature*, 412(6846), pp. 565–566. doi: 10.1038/35087627.
- Larracuente, A. M. (2014) ‘The organization and evolution of the Responder satellite in species of the *Drosophila melanogaster* group: Dynamic evolution of a target of meiotic drive’, *BMC Evolutionary Biology*, 14(1), pp. 1–12. doi: 10.1186/s12862-014-0233-9.
- Lauria Sneiderman, M. P. and Meller, V. H. (2021) ‘Drosophila Satellite Repeats at the Intersection of Chromatin, Gene Regulation and Evolution.’, *Progress in molecular and subcellular biology*. United States, 60, pp. 1–26. doi: 10.1007/978-3-030-74889-0\_1.
- Layat, E., Sáez-Vásquez, J. and Tourmente, S. (2012) ‘Regulation of pol I-transcribed 45S rDNA and pol III-transcribed 5S rDNA in arabidopsis’, *Plant and Cell Physiology*, 53(2), pp. 267–276. doi: 10.1093/pcp/pcr177.
- Lee, H. R., Zhang, W., Langdon, T., Jin, W., Yan, H, Cheng, Z., Jiang, J. (2005) ‘Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), pp. 11793–11798. doi: 10.1073/pnas.0503863102.
- Leitão, A. and Chaves, R. (2008) ‘Banding for Chromosomal Identification in Bivalves: A 20-year History’, *Dynamic Biochemistry, Process Biotechnology and Molecular Biology*, 2, pp. 44–49.
- Levinson, G. and Gutman, G. A. (1987) ‘Slipped-strand mispairing: a major mechanism for DNA sequence evolution.’, *Molecular biology and evolution*. United States, 4(3), pp. 203–221. doi: 10.1093/oxfordjournals.molbev.a040442.



- Li, J. and Leung, F. C. (2006) 'A CR1 element is embedded in a novel tandem repeat ( HinfI repeat ) within the chicken genome', *Genome*, 49, pp. 97–103. doi: 10.1139/G05-090.
- Liu, Q., Li X., Zhou, X, Li, M., Zhang, F., Schwarzacher, T., Heslop-Harrison, J. S. (2019) 'The repetitive DNA landscape in *Avena* (Poaceae): Chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads', *BMC Plant Biology*. BMC Plant Biology, 19(1), pp. 1–17. doi: 10.1186/s12870-019-1769-z.
- Liu, Z., Yue, W., Li, D., Wang, R. R. C., Kong, X., Lu, K, Wang, G., Dong, Y., Jin, W., Zhang, X. (2008) 'Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres', *Chromosoma*, 117(5), pp. 445–456. doi: 10.1007/s00412-008-0161-9.
- Lohe, A. R. and Brutlag, D. L. (1986) 'Multiplicity of satellite DNA sequences in *Drosophila melanogaster*', *Proceedings of the National Academy of Sciences of the United States of America*, 83(3), pp. 696–700. doi: 10.1073/pnas.83.3.696.
- Lohe, A. R. and Brutlag, D. L. (1987) 'Identical satellite DNA sequences in sibling species of *Drosophila*', *Journal of Molecular Biology*, 194(2), pp. 161–170. doi: 10.1016/0022-2836(87)90365-2.
- Lohe, A.R. and Roberts, P. (1988) 'Evolution of satellite DNA sequences in *Drosophila*', In: Verma, R. (ed) Molecular and structural aspects of heterochromatin, *Cambridge University Press*, pp. 148–186.
- López-Flores, I., de la Herrán, R., Garrido-Ramos, M.A., Boudry, P., Ruiz-Rejón, C., Ruiz-Rejón, M. (2004) 'The molecular phylogeny of oysters based on a satellite DNA related to transposons.', *Gene*, 339, pp. 181–188. doi: 10.1016/j.gene.2004.06.049.
- López-Flores, I., Ruiz-Rejón, C., Cross, I., Rebordinos, L., Robles, F., Navajas-Pérez, R., de la Herrán, R. (2010) 'Molecular characterization and evolution of an interspersed repetitive DNA family of oysters', *Genetica*, 138(11), pp. 1211–1219. doi: 10.1007/s10709-010-9517-1.
- López-Flores, I. and Garrido-Ramos, M. A. (2012) 'The repetitive DNA content of eukaryotic genomes', *Genome Dynamics*, 7, pp. 1–28. doi: 10.1159/000337118.
- Lorite, P., Carrillo, J. A., Tinaut, A., Palomeque, T. (2002) 'Comparative study of satellite DNA in ants of the *Messor* genus', *Gene*, 297(1–2), pp. 113–122. doi: 10.1016/S0378-1119(02)00875-2.

- Louzada, S., Lopes, M., Ferreira, D., Adegas, F., Escudeiro, A., Gama-Carvalho, M., Chaves, R. (2020) 'Decoding the Role of Satellite DNA in Genome Architecture and Plasticity—An Evolutionary and Clinical Affair', *Genes*, 11, 72. <https://doi.org/10.3390/genes11010072>
- Lower, S. S., McGurk, M. P., Clark, A. G., Barbash, D. A. (2018) 'Satellite DNA evolution: old ideas, new approaches', *Current opinion in genetics & development*, England, 49, pp. 70–78. doi: 10.1016/j.gde.2018.03.003.
- Luan, D. D., Korman, M. H., Jakubczak, J. L., Eickbush, T. H. (1993) 'Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition', *Cell*, United States, 72(4), pp. 595–605. doi: 10.1016/0092-8674(93)90078-5.
- Luchetti, A., Šatović, E., Mantovani, B., Plohl, M. (2016) 'RUDI, a short interspersed element of the V-SINE superfamily widespread in molluscan genomes', *Molecular Genetics and Genomics*. Springer Berlin Heidelberg, 291(3), pp. 1419–1429. doi: 10.1007/s00438-016-1194-z.
- Lynch, M., Logue, M. E., Butler, G., Wolfe, K. H. (2007) 'The frailty of adaptive hypotheses for the origins of organismal complexity.', *Proceedings of the National Academy of Sciences of the United States of America*, United States, 104 Suppl 1(Suppl 1), pp. 8597–8604. doi: 10.1073/pnas.0702207104.
- Lynch, D. B. *et al.* (2010) 'Chromosomal G + C content evolution in yeasts: Systematic interspecies differences, and GC-poor troughs at centromeres', *Genome Biology and Evolution*, 2(1), pp. 572–583. doi: 10.1093/gbe/evq042.
- Macas, J., Neumann, P. and Navrátilová, A. (2007) 'Repetitive DNA in the pea (*Pisum sativum* L.) genome: Comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*', *BMC Genomics*, 8, pp. 1–16. doi: 10.1186/1471-2164-8-427.
- Macas, J., Novák, P., Pellicer, J., Čížková, J., Koblížková, A., Neumann, P., Fuková, I., Doležel, J., Kelly, L. J., Leitch, I. J. (2015) 'In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae', *PLoS ONE*, 10(11), pp. 1–23. doi: 10.1371/journal.pone.0143424.
- Maison, C. and Almouzni, G. (2004) 'HP1 and the dynamics of heterochromatin maintenance', *Nature Reviews Molecular Cell Biology*, 5(4), pp. 296–305. doi: 10.1038/nrm1355.

Malik, H. S. (2009) 'The centromere-drive hypothesis: a simple basis for centromere complexity.', *Progress in molecular and subcellular biology*, 48(February 2009), pp. 33–52. doi: 10.1007/978-3-642-00182-6\_2.

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., Kuhlmann, M., Brandt, R., Vanzela, A. L. L., Beseda, T., Šimková, H., Pedrosa-Harand, A., Houben, A. (2015) 'Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin', *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), pp. 13633–13638. doi: 10.1073/pnas.1512255112.

Martínez-Expósito, M. J., Pasantes, J. J. and Méndez, J. (1994) 'NOR activity in larval and juvenile mussels (*Mytilus galloprovincialis*, Lmk.)', *Journal of Experimental Marine Biology and Ecology*, 175(2), pp. 155–165. doi: 10.1016/0022-0981(94)90023-X.

Martínez-García, M. F., Ruesink, J. L., Grijalva-Chon, J. M., Lodeiros, C., Arreola-Lizárraga, de la Re-Vega, E., Varela-Romero, A., Chávez-Villalba, J. (2022) 'Socioecological factors related to aquaculture introductions and production of Pacific oysters (*Crassostrea gigas*) worldwide', *Reviews in Aquaculture*, 14(2), pp. 613–629. doi: <https://doi.org/10.1111/raq.12615>.

Martínez-Lage, A., González-Tizón, A. and Méndez, J. (1995) 'Chromosomal markers in three species of the genus *mytilus* (Mollusca: Bivalvia)', *Heredity*, 74(4), pp. 369–375. doi: 10.1038/hdy.1995.55.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., Okazaki, T. (1989) 'A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite', *The Journal of cell biology*, United States, 109(5), pp. 1963–1973. doi: 10.1083/jcb.109.5.1963.

Matsunaga, S., Takata, H., Morimoto, A., Hayashihara, K., Higashi, T., Akatsuchi, K., Mizusawa, E., Yamakawa, M., Ashida, M., Matsunaga, T. M., Azuma, T., Uchiyama, S., Fukui, K. (2012) 'RBMX: A Regulator for Maintenance and Centromeric Protection of Sister Chromatid Cohesion', *Cell Reports*, 1(4), pp. 299–308. doi: 10.1016/j.celrep.2012.02.005.

May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L., Martienssen, R. A. (2005) 'Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats', *PLoS genetics*. United States, 1(6), p. e79. doi: 10.1371/journal.pgen.0010079.

- McGurk, M. P. and Barbash, D. A. (2017) ‘Double insertion of transposable elements provides a substrate for the evolution of satellite DNA’, *bioRxiv*, (i), pp. 70–78. doi: 10.1101/158386.
- McKinley, K. L. and Cheeseman, I. M. (2016) ‘The molecular basis for centromere identity and function’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 17(1), pp. 16–29. doi: 10.1038/nrm.2015.5.
- Messing, J., Bharti, A. K., Karlowski, W. M., Gundlach, H., Kim, H. R., Yu, Y., Wei, F., Fuks, G., Soderlund, C. A., Mayer, K. F. X., Wing, R. A. (2004) ‘Sequence composition and genome organization of maize’, *Proceedings of the National Academy of Sciences of the United States of America*, 101(40), pp. 14349–14354. doi: 10.1073/pnas.0406163101.
- Meštrović, N., Plohl, M., Mravinac, B., Ugraković, Đ. (1998) ‘Evolution of satellite DNAs from the genus *Palorus*-Experimental evidence for the “library” hypothesis’, *Molecular Biology and Evolution*, 15(8), pp. 1062–1068. doi: 10.1093/oxfordjournals.molbev.a026005.
- Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko, T., Šatović, E., Plohl, M. (2015) ‘Structural and functional liaisons between transposable elements and satellite DNAs’, *Chromosome Research*, 23(3), pp. 583–596. doi: 10.1007/s10577-015-9483-7.
- Miga, K. H. (2019) ‘Centromeric satellite DNAs: Hidden sequence variation in the human population’, *Genes*, 10(5). doi: 10.3390/genes10050352.
- Montiel, E. E., Panzera, F., Palomeque, T., Lorite, P, Pita, S. (2021) ‘Satellitome Analysis of *Rhodnius prolixus*, One of the Main Chagas Disease Vector Species’, *International Journal of Molecular Sciences*, 22(11), p. 6052. doi: 10.3390/ijms22116052.
- Moran, J. V., DeBerardinis, R. J. and Kazazian, H. H. (1999) ‘Exon shuffling by L1 retrotransposition’, *Science*, 283(5407), pp. 1530–1534. doi: 10.1126/science.283.5407.1530.
- Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., Batzer, M. A., Moran, J. V. (2002) ‘DNA repair mediated by endonuclease-independent LINE-1 retrotransposition’, *Nature genetics*. United States, 31(2), pp. 159–165. doi: 10.1038/ng898.

- Mravinac, B. and Plohl, M. (2007) ‘Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*’, *Gene*, 394(1–2), pp. 45–52. doi: 10.1016/j.gene.2007.01.019.
- Mravinac, B., Plohl, M., Meštrović, N., Ugarković, Đ. (2002) ‘Sequence of PRAT satellite DNA “frozen” in some Coleopteran species’, *Journal of molecular evolution*. Germany, 54(6), pp. 774–783. doi: 10.1007/s00239-001-0079-9.
- Mravinac, B., Plohl, M. and Ugarković, Đ. (2005) ‘Preservation and high sequence conservation of satellite DNAs suggest functional constraints’, *Journal of Molecular Evolution*, 61(4), pp. 542–550. doi: 10.1007/s00239-004-0342-y.
- Mravinac, B. and Plohl, M. (2010) ‘Parallelism in evolution of highly repetitive DNAs in sibling species’, *Molecular Biology and Evolution*, 27(8), pp. 1857–1867. doi: 10.1093/molbev/msq068.
- Müller, W. E. G., Schröder, H. C. and Gutenberg, J. (2022) *Satellite DNAs in Physiology and Evolution*. Available at: <http://www.springer.com/series/388>.
- Murgarella, M., Puiu, D., Novoa, B., Figueras, A., Posada, D., Canchaya, C. (2016) ‘A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*’, *PloS one*, 11(7), pp. 1–22. doi: 10.1371/journal.pone.0151561.
- Nagaki, K., Talbert, P. B., Zhong, C. X., Dawe, R. K., Henikoff, S., Jiang, J. (2003) ‘Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres’, *Genetics*, 163(3), pp. 1221–1225. doi: 10.1093/genetics/163.3.1221.
- Negoua, A., Rouault, J. D., Chakir, M., Capy, P. (2013) ‘Internal deletions of transposable elements: the case of Lemi elements’, *Genetica*, 141(7), pp. 369–379. doi: 10.1007/s10709-013-9736-3.
- Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., Macas, J. (2012) ‘Stretching the rules: Monocentric chromosomes with multiple centromere domains’, *PLoS Genetics*, 8(6). doi: 10.1371/journal.pgen.1002777.
- Neumann, P., Novák, P., Hošťáková, N., Macas, J. (2019) ‘Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification’, *Mobile DNA*, 10(1), pp. 1–17. doi: 10.1186/s13100-018-0144-1.

- Nishihara, H., Plazzi, F., Passamonti, M., Okada, N. (2016) ‘MetaSINEs: Broad distribution of a novel SINE superfamily in animals’, *Genome Biology and Evolution*, 8(3), pp. 528–539. doi: 10.1093/gbe/evw029.
- Novák, P., Neumann, P. and Macas, J. (2010) ‘Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data’, *BMC Bioinformatics*, 11, pp. 1–12. doi: 10.1186/1471-2105-11-378.
- Novák, P., Neumann, P. and Macas, J. (2020) ‘Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2’, *Nature Protocols*, 15(11), pp. 3745–3776. doi: 10.1038/s41596-020-0400-y.
- Ohno, S. (1972) ‘So Much “Junk” DNA in Our Genome’, In: Smith, H.H., Ed., *Evolution of Genetic Systems*, Gordon and Breach, New York, pp. 366-370.
- Ohshima, K., Hamada, M., Terai, Y., Okada, N. (1996) ‘The 3’ Ends of tRNA-Derived Short Interspersed Repetitive Elements Are Derived from the 3’ Ends of Long Interspersed Repetitive Elements’, *Molecular and Cellular Biology*, 16(7), pp. 3756–3764. doi: 10.1128/mcb.16.7.3756.
- Ohzeki, J. I., Nakano, M., Okada, T., Masumoto, H. (2002) ‘CENP-B box is required for *de novo* centromere chromatin assembly on human alphoid DNA’, *Journal of Cell Biology*, 159(5), pp. 765–775. doi: 10.1083/jcb.200207112.
- Okada, N. and Ohshima, K. (1995) ‘Evolution of tRNA-derived SINEs’, in: Maraia, R. J. (ed.) *The Impact of Short Interspersed Elements (SINEs) on the Host Genome*, pp. 61–79. R. G. Landes, Austin, TX.
- Orgel, L. E. and Crick, F. H. C. (1980) ‘Selfish DNA: the ultimate parasite’, *Nature*, 284(5757), pp. 604–607. doi: 10.1038/284604a0.
- Paço, A., Freitas, R. and Vieira-Da-Silva, A. (2019) ‘Conversion of DNA sequences: From a transposable element to a tandem repeat or to a gene’, *Genes*, 10(12). doi: 10.3390/genes10121014.
- Palacios-Gimenez, O. M., Bellini Bardella, V., Lemos, B., Cavalcanti Cabral-de-Mello, D. (2018) ‘Satellite DNAs are conserved and differentially transcribed among *Gryllus* cricket species’, *DNA Research*, 25(2), pp. 137–147. doi: 10.1093/dnares/dsx044.

- Palacios-Gimenez, O. M., Milani, D., Song, H., Marti, D. A., López-Léon, M. D., Ruiz-Ruano, F. J., Camacho, J. P. M., Cabral-de-Mello, D. C. (2020) ‘Eight Million Years of Satellite DNA Evolution in Grasshoppers of the Genus *Schistocerca* Illuminate the Ins and Outs of the Library Hypothesis’, *Genome Biology and Evolution*, 12(3), pp. 88–102. doi: 10.1093/gbe/evaa018.
- Pavlek, M., Glefand, Y., Plohl, M., Meštrović, N. (2015) ‘Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms’, *DNA Research*, 22(6), pp. 387–401. doi: 10.1093/dnares/dsv021.
- Penagos-Puig, A. and Furlan-Magaril, M. (2020) ‘Heterochromatin as an Important Driver of Genome Organization.’, *Frontiers in cell and developmental biology*. Switzerland, 8, p. 579137. doi: 10.3389/fcell.2020.579137.
- Peñaloza, C., Gutierrez, A. P., Eöry, L., Wang, S., Guo, X., Archibald, A. L., Bean, T. P., Houston, R. D. (2021) ‘A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*’, *GigaScience*, 10(3), pp. 1–9. doi: 10.1093/gigascience/giab020.
- Pérez-García, C., Morán, P. and Pasantes, J. J. (2011) ‘Cytogenetic characterization of the invasive mussel species *Xenostrobus securis* Lmk. (Bivalvia: Mytilidae)’, *Genome*, 54(9), pp. 771–778. doi: 10.1139/g11-040.
- Peters, A. H. F. M., Kubicek, S., Mechtler, K., O’Sullivan, R. J., Derijck, A. A. H. A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, A. H. A., Jenuwein, T. (2003) ‘Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin’, *Molecular Cell*, 12(6), pp. 1577–1589. doi: 10.1016/S1097-2765(03)00477-5.
- Petkevičiūtė, R., Stunžėnas, V. and Stanevičiūtė, G. (2018) ‘Comments on species divergence in the genus *Sphaerium* (Bivalvia) and phylogenetic affinities of *Sphaerium nucleus* and *S. corneum* var. *mamillanum* based on karyotypes and sequences of 16S and ITS1 rDNA’, *PLoS ONE*, 13(1), pp. 1–17. doi: 10.1371/journal.pone.0191427.
- Petraccioli, A., Odierna, G., Capriglione, T., Barucca, M., Forconi, Olmo, E., Biscotti, M. A. (2015) ‘A novel satellite DNA isolated in *Pecten jacobaeus* shows high sequence similarity among molluscs’, *Molecular Genetics and Genomics*, 290(5), pp. 1717–1725. doi: 10.1007/s00438-015-1036-4.

- Petrović, V. and Plohl, M. (2005) ‘Sequence divergence and conservation in organizationally distinct subfamilies of *Donax trunculus* satellite DNA.’, *Gene*, 362, pp. 37–43. doi: 10.1016/j.gene.2005.06.044.
- Petrović, V., Pérez-García, C., Pasantes, J. J., Šatović, E., Prats, E., Plohl, M. (2009) ‘A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax trunculus*: A dominance of GC-rich heterochromatin’, *Cytogenetic and Genome Research*, 124(1), pp. 63–71. doi: 10.1159/000200089.
- Pezer, Ž. and Ugarković, Đ. (2012) ‘Satellite DNA-associated siRNAs as mediators of heat shock response in insects’, *RNA biology*. United States, 9(5), pp. 587–595. doi: 10.4161/rna.20019.
- Pita, S., Panzera, F., Mora, P., Vela, J., Cuadrado, Á., Sánchez, A., Palomeque, T., Lorite, P. (2017) ‘Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease’, *PloS one*, 12(7), p. e0181635. doi: 10.1371/journal.pone.0181635.
- Plohl, M., Lucijanić-Justić V., Ugarković, Đ., Petitpierre, E., Juan, C. (1993) ‘Satellite DNA and heterochromatin of the flour beetle *Tribolium confusum*’, *Genome*, 36(3), pp. 467–475. doi: 10.1139/g93-064.
- Plohl, M. and Cornudella, L. (1997) ‘Characterization of interrelated sequence motifs in four satellite DNAs and their distribution in the genome of the mollusc *Donax trunculus*.’, *Journal of molecular evolution*, 44(2), pp. 189–98.
- Plohl, M., Luchetti, A., Meštrović, N., Mantvani, B. (2008) ‘Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin.’, *Gene*. Netherlands, 409(1–2), pp. 72–82. doi: 10.1016/j.gene.2007.11.013.
- Plohl, M., Petrović, V., Luchetti, A., Ricci, A., Šatović, E., Passamonti, M., Mantovani, B. (2010) ‘Long-term conservation vs high sequence divergence: The case of an extraordinarily old satellite DNA in bivalve mollusks’, *Heredity*, 104(6), pp. 543–551. doi: 10.1038/hdy.2009.141.
- Plohl, M., Meštrović, N. and Mravinac, B. (2012) ‘Satellite DNA evolution’, *Genome Dynamics*, 7(May 2015), pp. 126–152. doi: 10.1159/000337122.
- Plohl, M., Meštrović, N. and Mravinac, B. (2014) ‘Centromere identity from the DNA point of view.’, *Chromosoma*. Austria, 123(4), pp. 313–325. doi: 10.1007/s00412-014-0462-0.



- Poulter, R. T. M., Goodwin, T. J. D. and Butler, M. I. (2003) 'Vertebrate helitrons and other novel Helitrons', *Gene*, 313(1–2), pp. 201–212. doi: 10.1016/S0378-1119(03)00679-6.
- Poulter, R. T. M. and Goodwin, T. J. D. (2005) 'DIRS-1 and the other tyrosine recombinase retrotransposons', *Cytogenetic and Genome Research*, 110(1–4), pp. 575–588. doi: 10.1159/000084991.
- Pritham, E. J. and Feschotte, C. (2007) 'Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*', *Proceedings of the National Academy of Sciences of the United States of America*, 104(6), pp. 1895–1900. doi: 10.1073/pnas.0609601104.
- Pritham, E. J., Putliwala, T. and Feschotte, C. (2007) 'Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses', *Gene*, 390(1–2), pp. 3–17. doi: 10.1016/j.gene.2006.08.008.
- Probst, A. V. and Almouzni, G. (2011) 'Heterochromatin establishment in the context of genome-wide epigenetic reprogramming', *Trends in Genetics*, 27(5), pp. 177–185. doi: 10.1016/j.tig.2011.02.002.
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., Yang, Y., Wu, Z., Mao, L., Wu, H., Ling-Hu, C., Zhou, H., Lin, H., González-Morales, S., Trejo-Saavedra, D.L., Tian, H., Tang, X., Zhao, M., Huang, Z., Zhou, A., Yao, X., Cui, J., Li, W., Chen, Z., Feng, Y., Niu, Y., Bi, S., Yang, X., Li, W., Cai, H., Luo, X., Montes-Hernández, S., Leyva-González, M.A., Xiong, Z., He, X., Bai, L., Tan, S., Tang, X., Liu, D., Liu, J., Zhang, S., Chen, M., Zhang, L., Zhang, L., Zhang, Y., Liao, W., Zhang, Y., Wang, M., Lv, X., Wen, B., Liu, H., Luan, H., Zhang, Y., Yang, S., Wang, X., Xu, J., Li, X., Li, S., Wang, J., Palloix, A., Bosland, P.W., Li, Y., Krogh, A., Rivera-Bustamante, R.F., Herrera-Estrella, L., Yin, Y., Yu, J., Hu, K., Zhang, Z. (2014) 'Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization', *Proceedings of the National Academy of Sciences of the United States of America*, 111(14), pp. 5135–5140. doi: 10.1073/pnas.1400975111.
- Reece, K. S., Cordes, J. F., Stubbs, J. B., Hudson, K. L., Francis, E. A. (2008) 'Molecular phylogenies help resolve taxonomic confusion with Asian *Crassostrea* oyster species', *Marine Biology*, 153(4), pp. 709–721. doi: 10.1007/s00227-007-0846-2.

- Richard, G.-F., Kerrest, A. and Dujon, B. (2008) ‘Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes’, *Microbiology and Molecular Biology Reviews*, 72(4), pp. 686–727. doi: 10.1128/mmbr.00011-08.
- Robicheau, B. M., Susko, E., Harrigan, A. M., Snyder, M. (2017) ‘Ribosomal RNA genes contribute to the formation of pseudogenes and junk DNA in the human genome’, *Genome Biology and Evolution*, 9(2), pp. 380–397. doi: 10.1093/gbe/evw307.
- Robillard, É., le Rouzic, A., Zhang, Z., Capy, P., Hua-Van, A. (2016) ‘Experimental evolution reveals hyperparasitic interactions among transposable elements’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(51), pp. 14763–14768. doi: 10.1073/pnas.1524143113.
- Romanova, L. Y., Deriagin, G. V., Mashkova, T. D., Tumeneva, I. G., Mushegian, A. R., Kisselev, L. L., Alexandrov, I. A. (1996) ‘Evidence for selection in evolution of alpha satellite DNA: The central role of CENP-B/pJa binding region’, *Journal of Molecular Biology*, 261(3), pp. 334–340. doi: 10.1006/jmbi.1996.0466.
- Roy-Engel, A. M., El-Sawy, M., Farooq, L., Odom, G. L., Perepelitsa-Belancio, V., Bruch, H., Oyeniran, O. O., Deininger, P. L. (2005) ‘Human retroelements may introduce intragenic polyadenylation signals’, *Cytogenetic and Genome Research*, 110(1–4), pp. 365–371. doi: 10.1159/000084968.
- Ruesink, J. L., Lenihan, H. S., Trimble, A. C., Heiman, K. W., Micheli, F., Byers, J. E., Kay, M. C. (2005) ‘Introduction of non-native oysters: Ecosystem effects and restoration implications’, *Annual Review of Ecology, Evolution, and Systematics*, 36, pp. 643–689. doi: 10.1146/annurev.ecolsys.36.102003.152638.
- Saavedra, C. and Bachère, E. (2006) ‘Bivalve genomics’, *Aquaculture*, 256(June), pp. 1–14.
- Salvi, D., Macali, A. and Mariottini, P. (2014) ‘Molecular phylogenetics and systematics of the bivalve family Ostreidae based on rRNA sequence-structure models and multilocus species tree’, *PLoS ONE*, 9(9), pp. 19–21. doi: 10.1371/journal.pone.0108696.

- Sanyal, K., Baum, M. and Carbon, J. (2004) ‘Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique’, *Proceedings of the National Academy of Sciences of the United States of America*, 101(31), pp. 11374–11379. doi: 10.1073/pnas.0404318101.
- Scalvenzi, T. and Pollet, N. (2014) ‘Insights on genome size evolution from a miniature inverted repeat transposon driving a satellite DNA’, *Molecular Phylogenetics and Evolution*. Elsevier Inc., 81(December), pp. 1–9. doi: 10.1016/j.ympev.2014.08.014.
- Schatz, M. C., Delcher, A. L. and Salzberg, S. L. (2010) ‘Assembly of large genomes using second-generation sequencing’, *Genome Research*, 20(9), pp. 1165–1173. doi: 10.1101/gr.101360.109.
- Scherthan, H., Weich, S., Schwegler, H., Heyting, C., Härle, M., Cremer, T. (1996) ‘Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing’, *Journal of Cell Biology*, 134(5), pp. 1109–1125. doi: 10.1083/jcb.134.5.1109.
- Schmidt, T. and Heslop-Harrison, J. S. (1998) ‘Genomes, genes and junk: The large-scale organization of plant chromosomes’, *Trends in Plant Science*, 3(5), pp. 195–199. doi: 10.1016/S1360-1385(98)01223-0.
- Schueler, M. G. and Sullivan, B. A. (2006) ‘Structural and functional dynamics of human centromeric chromatin’, *Annual Review of Genomics and Human Genetics*, 7, pp. 301–313. doi: 10.1146/annurev.genom.7.080505.115613.
- Sedlazeck, F.J., Lee, H., Darby, C.A., Schatz, M.C. (2018) Piercing the Dark Matter: Bioinformatics of Long-Range Sequencing and Mapping. *Nature Reviews Genetics*, 19, pp. 329–346. doi: [10.1038/s41576-018-0003-4](https://doi.org/10.1038/s41576-018-0003-4)
- Seilacher, A. (1985) ‘Bivalve Morphology and Function’, *Series in Geology, Notes for Short Course*. Cambridge University Press, 13, pp. 88–101. doi: 10.1017/S0271164800001111.
- Sharma, A., Wolfgruber, T. K. and Presting, G. G. (2013) ‘Tandem repeats derived from centromeric retrotransposons’, *BMC Genomics*, 14(1). doi: 10.1186/1471-2164-14-142.
- Shatskikh, A. S., Kotov, A. A., Adashev, V. E., Bazylev, S. S., Olenina, L. V. (2020) ‘Functional significance of satellite DNAs: Insights from *Drosophila*’, *Frontiers in Cell and Developmental Biology*, 8(May), pp. 1–19. doi: 10.3389/fcell.2020.00312.

- Shen, S., Lin, L., Cai, J. J., Jiang, P., Kenkel, E. J., Stroik, M. R., Sato, S., Davidson, B. L., Xing, Y. (2011) 'Widespread establishment and regulatory impact of Alu exons in human genes', *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), pp. 2837–2842. doi: 10.1073/pnas.1012834108.
- Silva, B. S. M. L., Heringer, P., Dias, G. B., Svartman, M., Kuhn, G. C. S. (2019) 'De novo identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines', *PLoS ONE*, 14(12), pp. 1–15. doi: 10.1371/journal.pone.0223466.
- Smith, G.P. (1976) 'Evolution of repeated DNA sequences by unequal crossover', *Science*, 191(4227), pp. 528-35. doi: 10.1126/science.1251186.
- Smith, K. M., Galazka, J. M., Phatale, P. A., Connolly, L. R., Freitag, M. (2012) 'Centromeres of filamentous fungi.', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*. Netherlands, 20(5), pp. 635–656. doi: 10.1007/s10577-012-9290-3.
- Spaller, T., Kling, E., Glöckner, G., Hillmann, F., Winckler, T. (2016) 'Convergent evolution of tRNA gene targeting preferences in compact genomes', *Mobile DNA*. Mobile DNA, 7(1), pp. 1–15. doi: 10.1186/s13100-016-0073-9.
- Stölting, M., Wiesner, C., van Vliet, V., Butt, E., Pavenstädt, H., Linder, S., Kremerskothen, J. (2012) 'Lasp-1 Regulates Podosome Function', *PloS one*, 7(4), pp. 1–10. doi: 10.1371/journal.pone.0035340.
- Strachan, T., Webb, D. and Dover, G. A. (1985) 'Transition stages of molecular drive in multiple-copy DNA families in *Drosophila* ', *The EMBO Journal*, 4(7), pp. 1701–1708. doi: 10.1002/j.1460-2075.1985.tb03839.x.
- Su, H., Liu, Y., Liu, C., Shi, Q., Huang, Y., Han, F. (2019) 'Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes', *Plant Cell*, 31(9), pp. 2035–2051. doi: 10.1105/TPC.19.00133.

- Sujiwattananarat, P., Thapana, W., Srikulnath, K., Hirai, Y., Hirai, H., Koga, A. (2015) 'Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids', *Scientific Reports*. Nature Publishing Group, 5(March). doi: 10.1038/srep10315.
- Sullivan, B. A. and Karpen, G. H. (2004) 'Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin', *Nature Structural & Molecular Biology*, 11(11), pp. 1076–1083. doi: 10.1038/nsmb845.
- Sultana, T., Zamborlini, A., Cristofari, G., Lesage, P. (2017) 'Integration site selection by retroviruses and transposable elements in eukaryotes', *Nature Reviews Genetics*, 18(5), pp. 292–308. doi: 10.1038/nrg.2017.7.
- Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., Boeke, J. D. (2002) 'Human L1 retrotransposition is associated with genetic instability *in vivo* similar sloppy structures without a clear preference for', *Cell*, 110, pp. 327–338. doi: [10.1016/s0092-8674\(02\)00839-5](https://doi.org/10.1016/s0092-8674(02)00839-5).
- Šatović, E. and Plohl, M. (2013) 'Tandem repeat-containing MITEs in the clam *Donax trunculus*', *Genome Biology and Evolution*, 5(12), pp. 2549–2559. doi: 10.1093/gbe/evt202.
- Šatović, E., Vojvoda Zeljko, T., Luchetti, A., Mantovani, B., Plohl, M. (2016) 'Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements', *BMC Genomics*, 17(1), pp. 1–12. doi: 10.1186/s12864-016-3347-1.
- Šatović, E. and Plohl, M. (2017) 'Two new miniature inverted-repeat transposable elements in the genome of the clam *Donax trunculus*', *Genetica*. Springer International Publishing, 145(4), pp. 379–385. doi: 10.1007/s10709-017-9973-y.
- Šatović, E. and Plohl, M. (2018) 'Distribution of DTHS3 satellite DNA across 12 bivalve species.', *Journal of genetics*, 97(2), pp. 575–580.
- Šatović, E., Vojvoda Zeljko, T. and Plohl, M. (2018) 'Characteristics and evolution of satellite DNA sequences in bivalve mollusks', *European Zoological Journal*. Taylor & Francis, 85(1), pp. 95–104. doi: 10.1080/24750263.2018.1443164.

- Šatović, E., Luchetti, A., Pasantes, J. J., García-Souto, D., Cedilak, A., Mantovani, B., Plohl, M. (2019) 'Terminal-Repeat Retrotransposons in Miniature ( TRIMs ) in bivalves', *Scientific Reports*, 9(19962). doi: 10.1038/s41598-019-56502-y.
- Šatović, E., Tunjić Cvitanić, M. and Plohl, M. (2020) 'Tools and databases for solving problems in detection and identification of repetitive DNA sequences', *Periodicum Biologorum*, 121–122(1–2), pp. 7–14. doi: 10.18054/pb.v121-122i1-2.10571.
- Šatović Vukšić, E. & Plohl, M. (2021a) 'Exploring satellite DNAs: Specificities of bivalve mollusks genomes', In: Ugarković, Đ. (ed.), *Satellite DNAs in Physiology and Evolution*. Cham, Springer, pp. 57–83. doi:10.1007/978-3-030-74889-0.
- Šatović-Vukšić, E. and Plohl, M. (2021b) 'Classification Problems of Repetitive DNA Sequences', *Dna*, 1(2), pp. 84–90. doi: 10.3390/dna1020009.
- Talbert, P. B. and Henikoff, S. (2018) 'Transcribing Centromeres: Noncoding RNAs and Kinetochores Assembly', *Trends in Genetics*. Elsevier Ltd, 34(8), pp. 587–599. doi: 10.1016/j.tig.2018.05.001.
- Talbert, P. B. and Henikoff, S. (2020) 'What makes a centromere?', *Experimental cell research*. United States, 389(2), p. 111895. doi: 10.1016/j.yexcr.2020.111895.
- Tautz, D. (1993) 'Notes on the definition and nomenclature of tandemly repetitive DNA sequences.', *Exs*, 67, pp. 21–28. doi: 10.1007/978-3-0348-8583-6\_2.
- Tek, A. L., Song, J., Macas, J., Jiang, J. (2005) 'Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences', *Genetics*, 170(3), pp. 1231–1238. doi: 10.1534/genetics.105.041087.
- Thomas, J., Schaack, S. and Pritham, E. J. (2010) 'Pervasive horizontal transfer of rolling-circle transposons among animals', *Genome Biology and Evolution*, 2(1), pp. 656–664. doi: 10.1093/gbe/evq050.
- Thomas, J. and Pritham, E. J. (2015) 'Helitrons, the eukaryotic rolling-circle transposable elements', *Mobile DNA III*, (1), pp. 891–924. doi: 10.1128/9781555819217.ch40.
- Tong, P., Piodux, A. L., Toda, N. R. T., Ard, R., Berger, H., Shukla, M., Torres-Garcia, J., Müller, C. A., Nieduszynski, C. A., Allshire, R. C. (2019) 'Interspecies conservation of organisation and function 213

between nonhomologous regional centromeres’, *Nature Communications*. Springer US, 10(1), pp. 1–13. doi: 10.1038/s41467-019-09824-4.

Trojer, P. and Reinberg, D. (2007) ‘Facultative heterochromatin: is there a distinctive molecular signature?’, *Molecular cell*, 28(1), pp. 1–13. doi: 10.1016/j.molcel.2007.09.011.

Tunjić Cvitanić, M.; Vojvoda Zeljko, T.; Pasantes, J.J.; García-Souto, D.; Gržan, T.; Despot-Slade, E.; Plohl, M.; Šatović, E. (2020) Sequence Composition Underlying Centromeric and Heterochromatic Genome Compartments of the Pacific Oyster *Crassostrea gigas*. *Genes*, 11, 695. <https://doi.org/10.3390/genes11060695>.

Tunjić-Cvitanić, M.; Pasantes, J.J.; García-Souto, D.; Cvitanić, T.; Plohl, M.; Šatović-Vukšić, E. (2021) Satellitome Analysis of the Pacific Oyster *Crassostrea gigas* Reveals New Pattern of Satellite DNA Organization, Highly Scattered across the Genome. *Int. J. Mol. Sci.*, 22, 6798. <https://doi.org/10.3390/ijms22136798>.

Ugarković, Đ., Podnar, M., Plohl, M. (1996a) ‘Satellite DNA of the red flour beetle *Tribolium castaneum*-comparative study of satellites from the genus *Tribolium*’, *Molecular Biology and Evolution*, 13(8), pp. 1059-66. doi: 10.1093/oxfordjournals.molbev.a025668.

Ugarković, Đ., Durajlija, S. and Plohl, M. (1996b) ‘Evolution of *Tribolium madens* (Insecta, Coleoptera) satellite DNA through DNA inversion and insertion’, *Journal of Molecular Evolution*, 42(3), pp. 350–358. doi: 10.1007/bf02337545.

Ugarković, Đ. and Plohl, M. (2002) ‘Variation in satellite DNA profiles-Causes and effects’, *EMBO Journal*, 21(22), pp. 5955–5959. doi: 10.1093/emboj/cdf612.

Ugarković, Đ. (2005) ‘Functional elements residing within satellite DNAs’, *EMBO Reports*, 6(11), pp. 1035–1039. doi: 10.1038/sj.embor.7400558.

Ullu, E. and Tschudi, C. (1984) ‘Alu sequences are processed 7SL RNA genes’, *Nature*, 312(5990), pp. 171–172. doi: 10.1038/312171a0. Usakin, L. et al. (2007) ‘Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in *Drosophila melanogaster* ovaries’, *Genetics*, 176(2), pp. 1343–1349. doi: 10.1534/genetics.107.071720.

- Utsunomiya, Y. T., Milanesi, M., Fortes, M. R. S., Porto-Neto, L. R., Utsunomiya, A. T. H., Silva, M. V. G. B., Garcia, J. F., Ajmone-Marsan, P. (2019) ‘Genomic clues of the evolutionary history of *Bos indicus* cattle’, *Animal Genetics*, 50(6), pp. 557–568. doi: 10.1111/age.12836.
- Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., Thermes, C. (2018) ‘The Third Revolution in Sequencing Technology’, *Trends in Genetics*, Elsevier Ltd, 34(9), pp. 666–681. doi: 10.1016/j.tig.2018.05.008.
- Vaughn, C. C. and Hoellein, T. J. (2018) ‘Bivalve Impacts in Freshwater and Marine Ecosystems’.
- Venner, S., Feschotte, C., Biémont, C. (2009) Dynamics of transposable elements: towards a community ecology of the genome. *Trends in Genetics*, 25(7) pp. 317–23. doi: 10.1016/j.tig.2009.05.003.
- Vittorazzi, S. E., Lourenco, L. B., Solé, M., Faria, R. G., Recco-Pimentel, S. M. (2016) ‘Chromosomal analysis of *Physalaemus kroyeri* and *Physalaemus cicada* (Anura, Leptodactylidae)’, *Comparative Cytogenetics*, 10(2), pp. 311–323. doi: 10.3897/compcytogen.v10i2.9319.
- Vojvoda Zeljko, T., Pavlek, M., Meštrović, N., Plohl, M. (2020) ‘Satellite DNA-like repeats are dispersed throughout the genome of the Pacific oyster *Crassostrea gigas* carried by Helentron non-autonomous mobile elements’, *Scientific Reports*, 10, pp. 1–12. doi: 10.1038/s41598-020-71886-y.
- Vondrak, T., Robledillo, L. A., Novák, P., Koblížková, A., Neumann, P., Macas, J. (2020) ‘Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats’, *Plant Journal*, 101(2), pp. 484–500. doi: 10.1111/tpj.14546.
- Vondrak, T., Oliveira, L., Novák, P., Koblížková, Neumann, P, Macas, J. (2021) ‘Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads’, *Computational and structural biotechnology journal*, 19, pp. 2179–2189. doi: 10.1016/j.csbj.2021.04.011.
- Wang, L., Wang, Y., Zhang, J., Feng, Y., Chen, Q., Liu, Z. S., Liu, C. L., He, W., Wang, H., Yang, S. F., Yang, S. F., Zhang, Y., Luo, Y., Tang, H. R., Wang, X. R. (2022) ‘Comparative analysis of transposable elements and the identification of candidate centromeric elements in the *Prunus* subgenus *Cerasus* and its relatives’, *Genes*, 13(4). doi: 10.3390/genes13040641.



- Wang, Y., Xu, Z. and Guo, X. (2001) ‘A centromeric satellite Sequence in the Pacific oyster (*Crassostrea gigas*, Thunberg) Identified by fluorescence *in situ* hybridization’, *Marine Biotechnology*, 3(5), pp. 486–492. doi: 10.1007/s10126-001-0063-3.
- Wang, S., Lorenzen, M. D., Beeman, R. W., Brown, S. J. (2008) ‘Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome’, *Genome Biology*, 9(3), pp. 1–14. doi: 10.1186/gb-2008-9-3-r61.
- Wang, Jun *et al.* (2012) ‘The oyster genome reveals stress adaptation and complexity of shell formation’, *Nature*, 490(7418), pp. 49–54. doi: 10.1038/nature11413.
- Wang, X., Xu, W., Wei, L., Zhu, C., He, C., Song, H., Cai, Z., Yu, W., Jiang, Q., Li, L., Wang, K., Feng, C. (2019) ‘Nanopore sequencing and *de novo* assembly of a black-shelled Pacific oyster (*Crassostrea gigas*) genome’, *Frontiers in Genetics*, p. 1211. doi: 10.3389/fgene.2019.01211.
- Wang, Z., Zhao, G., Yang, Q., Gao, L., Liu, C., Ru, Z., Wang, D., Jia, J., Cui, D. (2022) ‘Helitron and CACTA DNA transposons actively reshape the common wheat-AK58 genome’, *Genomics*, 114(2), p. 110288. doi: <https://doi.org/10.1016/j.ygeno.2022.110288>.
- Wasson, K., Zabin, C. J., Bedinger, L., Diaz, M. C., Pearse, J. S. (2001) ‘Biological invasions of estuaries without international shipping: The importance of intraregional transport’, *Biological Conservation*, 102(2), pp. 143–153. doi: 10.1016/S0006-3207(01)00098-2.
- Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., Boeke, J. D., Moran, J. V. (2001) ‘Human L1 retrotransposition: cis Preference versus trans complementation’, *Molecular and Cellular Biology*, 21(4), pp. 1429–1439. doi: 10.1128/mcb.21.4.1429-1439.2001.
- Weiner, A. M. (2002) ‘SINEs and LINEs: The art of biting the hand that feeds you’, *Contemporary Music Review*, 21(1), pp. 71–79. doi: 10.1080/07494460216644.
- Weiss-Schneeweiss H., Leitch A. R., McCann J., Jang T.-S., Macas J. (2015). ‘Employing next generation sequencing to explore the repeat landscape of the plant genome’, in *Next Generation Sequencing in Plant Systematics, Regnum Vegetabile 157*, eds Hörandl E., Appelhans M. (Königstein: Koeltz Scientific Books; ), 155–180.

Wiland, E., Fraczek, M., Olszewska, M., Kurpisz, M. (2016) ‘Topology of chromosome centromeres in human sperm nuclei with high levels of DNA damage’, *Scientific Reports*. Nature Publishing Group, 6(July), pp. 1–12. doi: 10.1038/srep31614.

Winans, S., Yu, H. J., de los Santos, K., Wang, G. Z., Ramani, V. N. K., Goff, S. P. (2022) ‘A point mutation in HIV-1 integrase redirects proviral integration into centromeric repeats’, *Nature Communications*. Springer US, 13(1), pp. 1–12. doi: 10.1038/s41467-022-29097-8.

Wong, L. H. and Choo, K. H. A. (2004) ‘Evolutionary dynamics of transposable elements at the centromere.’, *Trends in genetics : TIG*. England, 20(12), pp. 611–616. doi: 10.1016/j.tig.2004.09.011.

Xu, L., Mao, A., Liu, H., Gui, B., Choy, Huang, H., Yu, Q., Zhang, X., Chen, M., Lin, N., Chen, L., Han, J., Wang, Y., Zhang, M., Li, X., He., D., Lin, Y., Zhang, J., Cram, D. S., Cao, H. (2020) ‘Long-Molecule Sequencing’, *The Journal of Molecular Diagnostics*, 22(8), pp. 1087–1095. doi: 10.1016/j.jmoldx.2020.05.004.

Yan, H., Ito, H., Nobuta, K., Ouyang, S., Jin, W., Tian, S., Lu, C., Venu, R. C., Wang, G. L., Green, P. J., Wing, R. A., Buell, C. R., Meyers, B. C., Jiang, J. (2006) ‘Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere’, *Plant Cell*, 18(9), pp. 2123–2133. doi: 10.1105/tpc.106.043794.

Yang, H.P., Barbash, D.A. (2008) ‘Abundant and species-specific *DINE-1* transposable elements in 12 *Drosophila* genomes’, *Genome Biology* 9, R39, doi.org/10.1186/gb-2008-9-2-r39.

Yang, L. and Bennetzen, J. L. (2009a) ‘Structure-based discovery and description of plant and animal Helitrons’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), pp. 12832–12837. doi: 10.1073/pnas.0905563106.

Yang, L. and Bennetzen, J. L. (2009b) ‘Distribution, diversity, evolution, and survival of Helitrons in the maize genome’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(47), pp. 19922–19927. doi: 10.1073/pnas.0908008106.

Yuan, Y. W. and Wessler, S. R. (2011) ‘The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(19), pp. 7884–7889. doi: 10.1073/pnas.1104208108.

- Zelentsova, E.S., Vashakidze, R.P., Krayev, A.S., Evgen'ev, M. B. (1986) 'Dispersed repeats in *Drosophila virilis*: elements mobilized by interspecific hybridization', *Chromosoma* 93, 469–476 (1986). doi: 10.1007/BF00386786
- Zhang, W., Zuo, S., Li, Z., Meng, Z., Han, J., Song, J., Pan, Y. B., Wang, K. (2017) 'Isolation and characterization of centromeric repetitive DNA sequences in *Saccharum spontaneum*', *Scientific Reports*, 7, pp. 1–12. doi: 10.1038/srep41659.
- Zhong, C. X. *et al.* (2002) 'Centromeric retroelements and satellites interact with maize kinetochore protein CENH3', *The Plant cell*. England, 14(11), pp. 2825–2836. doi: 10.1105/tpc.006106.
- Zimmer, E. A., Marshall, J. B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J. A., Jiang, J., Dawe, R. K. (1980) 'Rapid duplication and loss of genes coding for the  $\alpha$  chains of hemoglobin', *Proceedings of the National Academy of Sciences of the United States of America*, 77(4 I), pp. 2158–2162. doi: 10.1073/pnas.77.4.2158.

## **8. PRILOZI**



Dodatak 1. Skripta *parser.py* napisana programskim jezikom *Python* uz pomoć koje su izrezivane postrane regije zajedno s monomerima ili nizovima satDNA sličnih elementima Helitron iz kromosoma trenutno referentnog genoma vrste *C. gigas*.

```
import os
```

```
def parser(filename,satname,d):
```

```
    f = open(filename, mode='r')
```

```
    f.readline()
```

```
    positions = []
```

```
    last_min = 0
```

```
    last_max = 0
```

```
    first = True
```

```
for line in f:
```

```
    parts = line.split(',')
```

```
    name = parts[0]
```

```
    if name == satname:
```

```
        #print("hit")
```

```
        current_max = int(parts[3])
```

```
        current_min = int(parts[2])
```

```

    positions.append( (current_min, current_max) )

positions.reverse()

sequences = []
sequence_start = 0
sequence_end = 0

for pos in positions:

    if first:

        sequence_start = pos[0]

        first = False

    else:

        if (pos[0]-last_max) < d:

            pass

        else:

            sequence_end = last_max

            sequences.append( (sequence_start, sequence_end) )

            sequence_start = pos[0]

last_min = pos[0]
last_max = pos[1]

sequence_end = last_max

```

```
sequences.append( (sequence_start, sequence_end) )
```

```
f.close()
```

```
return sequences
```

```
def finder(cfilename, sequences, L):
```

```
    cfile = open(cfilename, mode='r')
```

```
    foldername = "Output_files"
```

```
    if not os.path.exists(foldername):
```

```
        os.makedirs(foldername)
```

```
    nfile = open(foldername + '/' + cfilename + '_positions.txt','w+')
```

```
    for i,seq in enumerate(sequences):
```

```
        #ovdje promijeni zelis li zadržati sredinu ili ne:
```

```
        KEEP_MIDDLE = True #True ili False
```

```
        start = seq[0]-L + 10
```

```
        if start < 0:
```

```
            start = 0
```

```
        end = seq[1]+L + 10
```

```
        size = seq[1]-seq[0]
```



```

nfile.write(str(start) + ' ' + str(end) + ' ' + str(size) + '\n')

cfile.seek(start)

if KEEP_MIDDLE:

    txt = cfile.read(end-start)

else:

    txt = cfile.read(L)

    cfile.seek(seq[0])

    txt = txt + cfile.read(L)

sfile = open(foldername + '/' + cfilename + '_out' + str(i) + '.fasta','w+')

sfile.write('>' + str(i) + '\n')

sfile.write(txt)

sfile.close()

nfile.close()

cfile.close()

```

```

print("Test program")

```

```

ANOTNAMES = ["C1A.csv","C2A.csv", "C3A.csv", "C4A.csv", "C5A.csv", "C6A.csv", "C7A.csv",
"C8A.csv", "C9A.csv", "C10A.csv"]

```

```

CMSMNames = ["C1.csv","C2.csv", "C3.csv", "C4.csv", "C5.csv", "C6.csv", "C7.csv", "C8.csv",
"C9.csv", "C10.csv"]

```

```

if len(ANOTNAMES) == len(CMSMNames):

```

```

    for i in range(len(ANOTNAMES)):

```

```

print(ANOTNAMES[i],CMSMNames[i])

p = parser(filename=ANOTNAMES[i], satname='CgiSat01a', d=500)

print("Parser done")

print(p[0])

finder(CMSMNames[i], p, 1000)

print(CMSMNames[i]+" completed")

print("Program done")

```

Dodatak 2. Skripta *parser\_scaffolds.py* napisana programskim jezikom *Python* uz pomoć koje su izrezivane postrane regije zajedno s monomerima ili nizovima satDNA sličnih elementima Helitron iz nesvrstanih segmenata trenutno referentnog genoma vrste *C. gigas*.

```
import os
```

```
def parser_scaffold(filename,satname,scaffold, d):
```

```
    f = open(filename, mode='r')
```

```
    f.readline()
```

```
    positions = []
```

```
    last_min = 0
```

```
    last_max = 0
```

```
    first = True
```

```

for line in f:

    parts = line.split(',')

    line_scaffold = parts[0]

    name = parts[1]

    if scaffold == line_scaffold:

        if name == satname:

            #print("hit")

            try:

                current_max = int(parts[4])

                current_min = int(parts[3])

            except:

                continue

            positions.append( (current_min, current_max) )

positions.reverse()

sequences = []

sequence_start = 0

sequence_end = 0

for pos in positions:

    if first:

```

```

sequence_start = pos[0]

first = False

else:

    if (pos[0]-last_max) < d:

        pass

    else:

        sequence_end = last_max

        sequences.append( (sequence_start, sequence_end) )

        sequence_start = pos[0]

last_min = pos[0]

last_max = pos[1]

sequence_end = last_max

sequences.append( (sequence_start, sequence_end) )

f.close()

return sequences

def finder_scaffold(cfilename, sequences, scafflist, L):

    cfile = open(cfilename, mode='r')

    cfile.readline()

    foldername = "Output_files"

```

```

if not os.path.exists(foldername):

    os.makedirs(foldername)

nfile = open(foldername + '/' + cfilename + '_positions.txt','w+')

for scaf in sequences:

    cline = cfile.readline().split(',')

    scafname = cline[0]

    cline_strip = cline[1]

for i,seq in enumerate(scaf):

    #ovdje promijeni zelis li zadržati sredinu ili ne:

    KEEP_MIDDLE = True #True ili False

    start = seq[0]-L

    if start < 0:

        start = 0

    end = seq[1]+L

    size = seq[1]-seq[0]

    nfile.write(scafname + ' ' + str(start) + ' ' + str(end) + ' ' + str(size) + '\n')

    if KEEP_MIDDLE:

        txt = cline_strip[start:end]

```

**else:**

```
txt = cline_strip[start:(start+L)]
```

```
txt = txt + cline_strip[seq[0]:(seq[0]+L)]
```

```
sfile = open(foldername + '/' + scafname + '_out' + str(i) + '.fasta','w')
```

```
sfile.write('>' + scafname + ' ' + str(i) + '\n')
```

```
sfile.write(txt)
```

```
sfile.close()
```

```
nfile.close()
```

```
cfile.close()
```

```
def get_scaffolds(sfilename):
```

```
    f = open(sfilename, "r")
```

```
    f.readline() #skip header
```

```
    names = []
```

```
    for line in f:
```

```
        result = line.find(',')
```

```
        names.append(line[0:result])
```

```
    f.close()
```

```
    return names
```

```

if __name__ == "__main__":

    print("Test program")

    ANOTNAME = "anotacije scaffolds AT.csv"

    SCFLDNAME = "scaffolds.csv"

    scaffold_list = get_scaffolds(SCFLDNAME)

    p = []

    print("Starting parser...")

    for item in scaffold_list:

        #print("Starting parser for scaffold", item)

        p.append( parser_scaffold(filename=ANOTNAME, satname='CgiSat01b',scaffold=item, d=500) )

        #print("Parser done")

    print("parser done!")

    finder_scaffold(SCFLDNAME, p,scaffold_list, 1000)

    print("Program done!")

```

Dodatak 3. Skripta *boxer.py* napisana programskim jezikom *Python* uz pomoć koje je određivana prisutnost konzerviranih regija (*box-eva*) u izrezanim postranim regijama kromosoma trenutno referentnog genoma vrste *C. gigas*.

```
import os
```

```

def file_worker(filename, LIST, LIST_max):
    """ Parses file and fills LIST container"""
    f = open(filename, "r")
    f.readline() #get rid of first line (column names)
    for line in f:
        line = line.split(",")
        if line[0] == '\n':
            continue
        nr = int(line[0])
        index = line[2].find("Box")
        if index >= 0:
            box = int(line[2][index+4])-1

            LIST[nr][box] = 1

        if nr > LIST_max:
            LIST_max = nr
    f.close()
    return (LIST, LIST_max)

def first_writer(filename, LIST, LIST_max, header = True):

```



```
"""Writes data for LIST to first csv"""
```

```
f = open(filename, "w+")
```

```
if header:
```

```
    f.write("Position,Box 1,Box 2,Box 3,Box 4,Box 5,Box 6,Box 7\n")
```

```
for i in range(LIST_max+1):
```

```
    line = str(i)
```

```
    for j in range(7):
```

```
        line += ","+str(LIST[i][j])
```

```
    line += "\n"
```

```
    f.write(line)
```

```
f.close()
```

```
def second_writer(filename, LIST, LIST_max, header = True):
```

```
    """Writes data per boxes"""
```

```
    f = open(filename, "w+")
```

```
    boxes = ["Box 1", "Box 2", "Box 3", "Box 4", "Box 5", "Box 6", "Box 7"]
```

```
    f.write("Box,List of positions\n")
```

```
    for i, box in enumerate(boxes):
```

```

f.write(box)

for j in range(LIST_max):

    if LIST[j][i] == 1:

        f.write(","+str(j) )

f.write("\n")

f.close()

if __name__ == "__main__":

    FILENAME = "anotacije boxeva za cgisat01b na chr3.csv"

    FILENAME1 = FILENAME[:-4]+"_first.csv"

    FILENAME2 = FILENAME[:-4]+"_second.csv"

    LIST = [[0]*7 for i in range(10000)]

    LIST_max = 0

    LIST, LIST_max = file_worker(FILENAME,LIST, LIST_max)

    first_writer(FILENAME1, LIST, LIST_max)

    second_writer(FILENAME2,LIST,LIST_max)

```

Dodatak 4. Skripta *boxer\_scaffolds.py* napisana programskim jezikom *Python* uz pomoć koje je određivana prisutnost konzerviranih regija (*box*-eva) u izrežanim postranim regijama nesvrstanih segmenata trenutno referentnog genoma vrste *C. gigas*.

```
import os
```

```
def file_worker(filename, LIST, LIST_max):
```

```
    """ Parses file and fills LIST container"""
```

```
    f = open(filename, "r")
```

```
    f.readline() #get rid of first line (column names)
```

```
    for line in f:
```

```
        line = line.split(",")
```

```
        if line[0] == '\n':
```

```
            continue
```

```
        nr = int(line[0])-22994773
```

```
        index = line[2].find("Box")
```

```
        if index >= 0:
```

```
            box = int(line[2][index+4])-1
```

```
            LIST[nr][box] = 1
```

```
        if nr > LIST_max:
```

```
            LIST_max = nr
```

```
    f.close()
```

```
    return (LIST, LIST_max)
```

```

def first_writer(filename, LIST, LIST_max, header = True):
    """Writes data for LIST to first csv"""
    f = open(filename, "w+")
    if header:
        f.write("Position,Box 1,Box 2,Box 3,Box 4,Box 5,Box 6,Box 7\n")

    for i in range(LIST_max+1):
        line = str(i)
        for j in range(7):
            line += ","+str(LIST[i][j])
        line += "\n"
        f.write(line)
    f.close()

def second_writer(filename, LIST, LIST_max, header = True):
    """Writes data per boxes"""
    f = open(filename, "w+")

    boxes = ["Box 1", "Box 2", "Box 3", "Box 4", "Box 5", "Box 6", "Box 7"]

    f.write("Box,List of positions\n")

```

```

for i, box in enumerate(boxes):

    f.write(box)

    for j in range(LIST_max):

        if LIST[j][i] == 1:

            f.write( ","+str(j) )

        f.write("\n")

    f.close()

if __name__ == "__main__":

    FILENAME = "anotacije boxeva na scaffoldu cgisat03.csv"

    FILENAME1 = FILENAME[:-4]+"_first.csv"
    FILENAME2 = FILENAME[:-4]+"_second.csv"

    LIST = [[0]*7 for i in range(10000)]

    LIST_max = 0

    LIST, LIST_max = file_worker(FILENAME,LIST, LIST_max)

    first_writer(FILENAME1, LIST, LIST_max)

    second_writer(FILENAME2,LIST,LIST_max)

```

Dodatak 5. Tablica prikazuje klustere dobivene u svakoj od analiza TAREAN napravljenih za genomske sekvencije vrste *C. gigas*.

satDNA	1. krug analize		2. krug analize		3. krug analize		4. krug analize	
	1 779 522 očitavanja, 0,35 x pokrivenost genoma	% genoma	1 000 000 očitavanja, 0,2 x pokrivenost genoma	% genoma	2 000 000 očitavanja, 0,4 x pokrivenost genoma	% genoma	1 000 000 očitavanja, 0,2 x pokrivenost genoma	% genoma
CgiSat01a	CI1	0.76	CI1	0.79	CI1	0.79	CI2	0.52
CgiSat01b	CI2	0.57	CI2	0.57	CI2	0.58	CI1	0.57
CgiSat02	CI4	0.40	CI3	0.46	CI3	0.48	CI4	0.41
CgiSat03	CI5	0.38	CI4	0.39	CI7	0.39	CI9	0.36
CgiSat04	CI8	0.36	CI9	0.32	CI10	0.33	-	-
CgiSat05	CI9	0.35	CI5	0.38	CI6	0.39	CI7	0.40
CgiSat06	CI11	0.33	CI8	0.32	CI11	0.33	CI11	0.32
CgiSat07	CI28	0.20	CI22	0.21	CI26	0.20	CI23	0.21
CgiSat08	CI29	0.19	CI27	0.19	CI29	0.19	CI30	0.18
CgiSat09	CI40	0.16	CI48	0.13	CI40	0.16	CI41	0.14
CgiSat10	CI46	0.14	-	-	CI44	0.15	CI40	0.14
CgiSat11	CI48	0.14	CI43	0.14	CI48	0.14	CI39	0.14
CgiSat12	CI59	0.12	CI61	0.10	CI63	0.12	CI56	0.11
CgiSat13	CI60	0.11	CI59	0.11	-	-	-	-
CgiSat14	CI61	0.11	CI58	0.11	CI70	0.11	-	-
CgiSat15	CI86	0.09	CI80	0.08	CI85	0.09	CI77	0.08
CgiSat16	CI102	0.07	CI91	0.07	CI103	0.07	CI89	0.07
CgiSat17	CI119	0.06	CI122	0.06	CI128	0.06	-	-
CgiSat18	CI129	0.06	CI121	0.06	-	-	CI130	0.05
CgiSat19	CI139	0.05	CI180	0.04	CI199	0.04	CI206	0.03
CgiSat20	CI145	0.05	CI138	0.05	CI152	0.05	CI125	0.05
CgiSat21	CI152	0.05	CI137	0.05	CI148	0.05	CI142	0.05
CgiSat22	CI175	0.04	CI159	0.04	CI173	0.04	CI152	0.04
CgiSat23	CI182	0.04	CI175	0.04	-	-	-	-
CgiSat24	CI199	0.03	CI171	0.04	CI194	0.04	CI173	0.04
CgiSat25	CI205	0.03	-	-	CI211	0.03	CI204	0.03
CgiSat26	CI269	0.02	CI247	0.02	CI293	0.02	CI246	0.02
CgiSat27	CI296	0.02	CI266	0.02	CI316	0.02	CI262	0.02
CgiSat28	CI298	0.02	CI264	0.02	CI301	0.02	CI282	0.02
CgiSat29	CI308	0.02	-	-	-	-	CI300	0.01
CgiSat30	CI326	0.01	CI295	0.01	CI332	0.01	CI290	0.02

CgiSat31	CI343	0.01	-	-	CI362	0.01	CI344	0.01
CgiSat32	CI373	0.01	-	-	CI365	0.01	-	-
CgiSat33	-	-	CI139	0.05	CI147	0.05	-	-
CgiSat34	-	-	CI186	0.03	CI187	0.04	-	-
CgiSat35	-	-	CI234	0.02	CI274	0.02	-	-
CgiSat36	-	-	CI341	0.01	CI390	0.01	CI341	0.01
CgiSat37	CI10	0.34	-	-	-	-	-	-
CgiSat38	CI70	0.10	-	-	-	-	-	-
CgiSat39	CI142	0.05	-	-	-	-	-	-
CgiSat40	CI178	0.04	-	-	-	-	-	-
CgiSat41	CI312	0.02	-	-	-	-	-	-
CgiSat42	-	-	CI187	0.03	-	-	-	-
CgiSat43	-	-	CI342	0.01	-	-	-	-
CgiSat44	-	-	CI344	0.01	-	-	-	-
CgiSat45	-	-	-	-	CI93	0.08	-	-
CgiSat46	-	-	-	-	CI241	0.03	-	-
CgiSat47	-	-	-	-	CI282	0.02	-	-
CgiSat48	-	-	-	-	-	-	CI5	0.40
CgiSat49	-	-	-	-	-	-	CI172	0.04
CgiSat50	-	-	-	-	-	-	CI234	0.02
CgiSat51	-	-	-	-	-	-	CI319	0.01
CgiSat52	-	-	-	-	-	-	CI111	0.06

## **9. ŽIVOTOPIS**





Monika Tunjić Cvitanić rođena je 7. rujna 1991. u Zagrebu gdje završila osnovnoškolsko i srednjoškolsko obrazovanje. Preddiplomski studij molekularne biologije na Prirodoslovno-matematičkom fakultetu upisuje 2010. Diplomski rad izradila je u Francuskoj, u Orléansu u Institute National des Forêts (Državni šumarski institut) pod mentorsvom dr. sc. Caroline Teyssier i dr. sc. Brigitte Musch te je 2015. stekla titulu magistre molekularne biologije i magistre bioindustrijskih tehnika. Nakon devetomjesečnog radnog iskustva u tvrtki PrimeVigilance, 2016. upisuje doktorski studij biologije na Prirodoslovno-matematičkom fakultetu. Od listopada 2016. do veljače 2018. zaposlena je u Centru za istraživanje i prijenos znanja u biotehnologiji Sveučilišta u Zagrebu. Na Institutu Ruđer Bošković započinje s radom 2018. godine u Laboratoriju za strukturu i funkciju heterokromatina te započinje izradu doktorske disertacije pod mentorstvom prof. dr. sc. Miroslava Plohla i doc. dr. sc. Eve Šatović Vukšić. Znanstveno se usavršavala u Ljubljani, u laboratoriju prof. dr. sc. Igora Križaja na Institutu Jožef Štefan, te dvaput po mjesec dana na Sveučilištu u Vigu u Španjolskoj u laboratoriju prof. dr. sc. Juana Josea Pasantesa. Za oba usavršavanja u Španjolskoj dobila je stipendiju Instituta Ruđer Bošković za mobilnost mladih istraživača. Dobitnica je Stipendije Grada Zagreba za izvrsnost u studiranju za akad. god. 2014./2015. te Posebne rektorove nagrade za sudjelovanje na manifestaciji Noć biologije. Monika je do sada objavila sedam znanstvenih radova s međunarodnom recenzijom od čega su tri s prvim autorstvom te je sudjelovala na nekoliko domaćih i međunarodnih znanstvenih skupova s posterskim izlaganjima. Od rujna 2022. do listopada 2023. bila je na porodiljnom dopustu.