

Identifikacija i karakterizacija niskoizastupljenih satelitnih DNA u genomu vrste *Tribolium castaneum*

Dombi, Mira

Master's thesis / Diplomski rad

2018

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

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:938878>

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

Download date / Datum preuzimanja: **2024-12-23**



Repository / Repozitorij:

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



Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Biološki odsjek

Mira Dombi

**Identifikacija i karakterizacija nisko zastupljenih satelitnih DNA u
genomu vrste *Tribolium castaneum***

Diplomski rad

Zagreb, 2018.

Ovaj rad, izrađen u Laboratoriju za strukturu i funkciju heterokromatina Instituta Ruđer Bošković, pod vodstvom dr. sc. Brankice Mravinac, predan je na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistra molekularne biologije.

Neizmjernu zahvalnost, prvenstveno, dugujem mentorici dr. sc. Brankici Mravinac na strpljenju, podršci, vodstvu i ogromnom znanju koje mi je prenijela pri izradi ovog diplomskog rada. Također zahvaljujem svim djelatnicima Laboratorija za strukturu i funkciju heterokromatina Instituta Ruđer Bošković na pomoći i suradnji.

Zahvaljujem se svim profesorima, kolegama i prijateljima koji su studentske dane učinili zanimljivijima i uspješnima.

Naposlijetku najveću zaslugu za sva moja postignuća dugujem svojim roditeljima koji su u svakom trenutku bili uz mene pružajući podršku i ohrabrenje i bez kojih danas ne bih bila tu gdje jesam.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Diplomski rad

Identifikacija i karakterizacija nisko zastupljenih satelitnih DNA u genomu vrste *Tribolium castaneum*

Mira Dombi

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Satelitne DNA su uzastopno ponavljajuće sekvence raširene u heterokromatinskim regijama eukariotskih genoma, a karakterizira ih još uvijek nedovoljno rasvijetljena funkcija. Kukac kestenjasti brašnar *Tribolium castaneum* je izvrstan modelni organizam koji se desetljećima koristi u molekularnim istraživanjima. Genom vrste *T. castaneum* sekvenciran je prije deset godina pristupom „shotgun“, no u sastavljenom dijelu genoma nedostaje 20% ukupne DNA primarno sastavljene od satelitnih sekvenca. U ovom radu genom vrste *T. castaneum* sekvenciran je tehnologijom sekvenciranja nove generacije (NGS) kako bi se pronašle dosad neotkrivene satelitne DNA. U analizi satelitoma korišten je bioinformatički alat TAREAN koji omogućava detekciju potencijalnih satelitnih sekvenca na temelju kratkih, neposloženih NGS sljedova. Bioinformatičkom analizom utvrđena su tri moguća satelita: TCst154, TCst266 i TCst309, koji nisu bili detektirani u prethodnim istraživanjima. Premda ova tri satelita čine mali udio u genomu (<0,5%), njihova uzastopno ponovljena organizacija uspješno je dokazana metodom Southern blot. Metodom fluorescencijske hibridizacije *in situ* također je određena njihova kromosomska lokalizacija. Kombinirajući bioinformatičke alate i eksperimentalne metode, ovim je istraživanjem potvrđena primjenjivost kratkih, neposloženih NGS sljedova u otkrivanju nisko zastupljenih satelita, a također je proširena i spoznaja o sadržaju satelitoma vrste *T. castaneum*.

(59 stranica, 15 slika, 7 tablica, 73 literaturna navoda, jezik izvornika: hrvatski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Ključne riječi: *Tribolium castaneum*, satelitna DNA, sekvenciranje nove generacije

Voditelj: Dr. sc. Brankica Mravinac, viši znanstveni suradnik

Ocjenitelji: Dr. sc. Višnja Besendorfer, red. prof.

Dr. sc. Nada Oršolić, red. prof.

Dr. sc. Željka Vidaković-Cifrek, izv. prof.

Rad prihvaćen: 06.09.2018.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Graduation Thesis

Identification and characterization of low-copy satellite DNA in the genome of the species *Tribolium castaneum*

Mira Dombi

Rooseveltovo trg 6, 10000 Zagreb, Croatia

Satellite DNAs are tandemly repeated sequences spread in the heterochromatic regions of eukaryotic genomes, and are characterized by still poorly understood functional roles. The red flour beetle *Tribolium castaneum* is an excellent model organism that has been used in molecular studies for many decades. Its genome was sequenced ten years ago by shotgun approach; however, 20% of the genome is omitted from the current assembly due to the repetitive set-up based primarily on satellite DNAs. In this work, the *T. castaneum* genome was sequenced by Next Generation Sequencing (NGS) technology in order to detect yet unidentified satellite sequences. The satellitome was analyzed by the computational pipeline TAREAN, which enables detection of putative satellite DNAs from unassembled short reads. Bioinformatic analysis revealed three probable satellites: TCst154, TCst266 i TCst309 which were not detected in previous studies. Although these three satellite DNAs comprise small amount of the total genomic content (<0.5%), Southern blot method successfully confirmed their tandemly repeated organization. Fluorescence *in situ* hybridization also determined their chromosomal localization. By combining bioinformatic tools and experimental methods, this work has confirmed the applicability of short, unassembled NGS reads in detecting low-copy satellite DNAs, as well as broadened the knowledge on the content of the *T. castaneum* satellitome.

(59 pages, 15 figures, 7 tables, 73 references, original in: Croatian)

Thesis deposited in the Central Biological Library

Key words: *Tribolium castaneum*, satellite DNA, Next Generation Sequencing

Supervisor: Brankica Mravinac, PhD, Senior Research Associate

Reviewers: Višnja Besendorfer, PhD, Prof.

Nada Oršolić, PhD, Prof.

Željka Vidaković-Cifrek, PhD, Assoc. Prof.

Thesis accepted: 06.09.2018.

SADRŽAJ

1. UVOD	1
1.1. Satelitne DNA.....	1
1.2. Kestenjasti brašnar <i>Tribolium castaneum</i>	5
1.3. Satelitne DNA vrste <i>T. castaneum</i>	6
1.4. Cilj istraživanja.....	7
2. MATERIJALI I METODE	8
2.1. Materijali	8
2.1.1. Životinjski materijal.....	8
2.1.2. Osnovne kemikalije.....	8
2.1.3. Pufferi i otopine.....	8
2.1.4. Kompleti reagencija	10
2.1.5. Enzimi	11
2.1.6. Elektroforetski biljezi.....	11
2.1.7. Plazmidni vektor	12
2.1.8. Bakterijski soj	12
2.1.9. Tehnička oprema i uređaji	12
2.1.10. Ostali materijali.....	13
2.2. METODE.....	13
2.2.1. Izolacija genomske DNA	13
2.2.2. Razgradnja DNA restrikcijskim endonukleazama.....	14
2.2.3. Umnažanje odsječaka DNA metodom lančane reakcije polimerazom.....	15
2.2.3.1. Početnice	15
2.2.3.2. Reakcijske smjese PCR-a i uvjeti reakcija	16
2.2.4. Elektroforetsko razdvajanje odsječaka DNA u agaroznom gelu	17
2.2.5. Izolacija odsječaka DNA iz agaroznog gela	17

2.2.6. Kloniranje odsječaka DNA	18
2.2.6.1. Ligacija odsječaka DNA i plazmidnog vektora	18
2.2.6.2. Transformacija kompetentnih bakterijskih stanica	18
2.2.6.3. PCR na bakterijskim kolonijama.....	18
2.2.6.4. Izolacija plazmidne DNA.....	19
2.2.6.5. Određivanje nukleotidnih sljedova kloniranih odsječaka DNA.....	19
2.2.7. Obilježavanje hibridizacijskih sonda	19
2.2.8. Hibridizacijska analiza prema Southernu	20
2.2.8.1. Prijenos odsječaka DNA na najlonsku membranu.....	21
2.2.8.2. Prehibridizacija i hibridizacija membrana.....	21
2.2.8.3. Otkrivanje hibridizacijskog signala.....	21
2.2.9. Fluorescencijska hibridizacija <i>in situ</i> (FISH).....	22
2.2.9.1. Priprema kromosomskih preparata.....	22
2.2.9.2. Prehibridizacija i hibridizacija <i>in situ</i>	22
2.2.9.3. Otkrivanje fluorescencijskog signala	23
2.2.9.4. Mikroskopska analiza kromosomskih preparata	24
2.2.10. Kapljična analiza.....	24
2.2.11. Bioinformatička analiza	24
2.2.11.1. Analiza genoma vrste <i>T. castaneum</i>	24
2.2.11.2. Analiza kloniranih odsječaka DNA.....	25
3. REZULTATI	27
3.1. Identifikacija satelitnih DNA u genomu vrste <i>T. castaneum</i> sekvenciranom metodom NGS	27
3.2. Kloniranje i karakterizacija odsječaka satelitnih DNA TCst154, TCst266 i TCst309..	30
3.3. Genomska organizacija satelitnih DNA TCst154, TCst266 i TCst309.....	39
3.4. Kromosomska lokalizacija satelitnih DNA TCst154, TCst266 i TCst309.....	42
3.5. Udio satelitnih DNA TCst154, TCst266 i TCst309	45

4. RASPRAVA.....	47
5. ZAKLJUČAK.....	52
6. LITERATURA	53
7. PRILOZI	I

POPIS KRATICA

KRATICA	OBJAŠNJENJE
BLAST	engl. <i>Basic Local Alignment Search Tool</i>
CNI	engl. <i>Close-Neighbor-Interchange</i>
DAPI	engl. <i>4',6-diamidino-2-phenylindole dihydrochloride</i>
EDTA	engl. <i>Ethylenediaminetetraacetic Acid</i>
FISH	fluorescencijska hibridizacija <i>in situ</i> ; engl. <i>fluorescence in situ hybridization</i>
HOR	engl. <i>Higher Order Repeat</i>
IPTG	izopropil- β -D-tiogalaktozid
LB	engl. <i>Luria Broth</i>
ML	engl. <i>Maximum Likelihood</i>
MP	engl. <i>Maximum Parsimony</i>
NGS	engl. <i>Next Generation Sequencing</i>
NNI	engl. <i>Nearest-Neighbor-Interchange</i>
NJ	engl. <i>Neighbor Joining</i>
pb	parova baza
PBS	engl. <i>Phosphate Buffered Saline</i>
PCR	lančana reakcija polimeraze; engl. <i>polymerase chain reaction</i>
satDNA	satelitna DNA
SDS	engl. <i>Sodium Dodecyl Sulfate</i>
SOC	engl. <i>Super Optimal Broth with Catabolite Repression</i>
SSC	engl. <i>Saline Sodium Citrate</i>
TAE	engl. <i>Tris-Acetate-EDTA</i>
TAREAN	engl. <i>Tandem Repeat Analyzer</i>
TCst154	engl. <i>Tribolium castaneum Satellite 154</i>
TCst266	engl. <i>Tribolium castaneum Satellite 266</i>
TCst309	engl. <i>Tribolium castaneum Satellite 309</i>
TE	engl. <i>Tris-EDTA</i>
X-gal	5-brom-4-klor-3-indolil- β -D-galaktozid

1. UVOD

1.1. Satelitne DNA

Satelitne DNA (satDNA) su nekodirajuće sekvence koje uzastopnim nizanjem ponavljajućih jedinica, takozvanih monomera, grade vrlo dugačke nizove DNA prvenstveno u heterokromatinskim područjima eukariotskih genoma (Garrido-Ramos 2017). SatDNA su uglavnom strukturirane na način da su monomeri ponovljeni jedan za drugim u organizaciji glava-rep (engl. *head-to-tail*), no također neke satDNA pokazuju složenu organizaciju gdje monomeri (ili više monomera u nizu) formiraju novu jedinicu ponavljanja, koja se naziva jedinicom organizacije višeg reda (engl. *higher-order repeat*, HOR) (Palomeque i Lorite 2008).

Na kromosomima, satDNA se najčešće nalaze u područjima (peri)centromernog ili subtelomernog heterokromatina, a iznimno i u eukromatinskim područjima (Martinez-Lage i sur. 1994). SatDNA smještene u centromernim i pericentromernim heterokromatinskim područjima povezuju se s jednom od osnovnih bioloških funkcija, pravilnim razdvajanjem kromosoma u mejozi i mitozu (Schueler i sur. 2001, Sun i sur. 2003). Predloženo je da promjene u centromernoj DNA preko promjena u centromernim proteinima mogu dovesti do specijacije, odnosno odvajanja vrsta (Henikoff i sur. 2001). SatDNA su tradicionalno smatrane potpuno transkripcijski neaktivnima, no u novije vrijeme transkripti satDNA pronađeni su u mnogim organizmima (Biscotti i sur. 2015). Pokazano je kako kratki transkripti nekih satDNA upravljaju procesima izgradnje i održavanja heterokromatina (Volpe i sur. 2002, Verdel i sur. 2004) te postoje indikacije kako satDNA mogu imati regulatornu ulogu u eukariotskim genomima (Ugarković 2005).

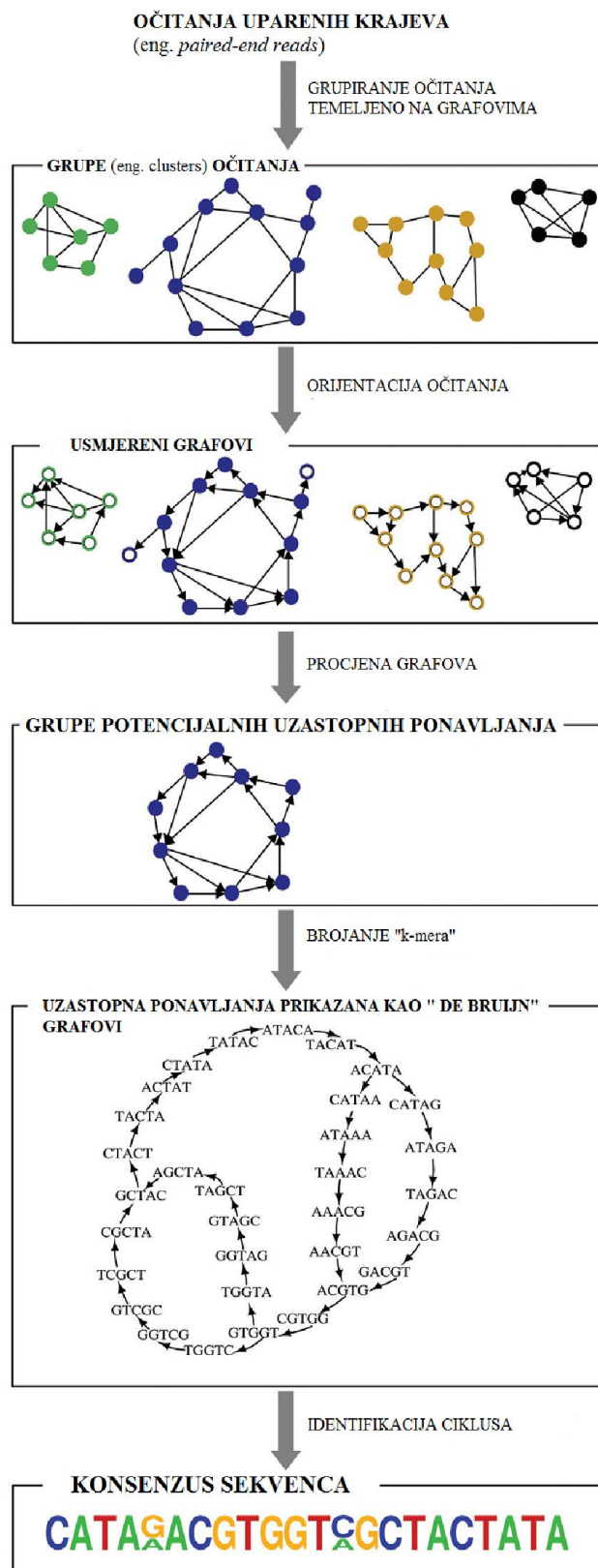
U eukariotskim genomima istovremeno postoji veći broj različitih satDNA koje se međusobno razlikuju s obzirom na slijed nukleotida osnovnih ponavljajućih jedinica, broj kopija u genomu, kromosomsku lokaciju, te broj osnovnih ponavljajućih jedinica (Garrido-Ramos 2017), a zajedno čine vrsno specifičan satelitni profil (Ugarković i Plohl 2002). S obzirom da recentna literatura bilježi sve više studija visokog protoka (engl. *high-throughput analysis*) kojima se otkriva i do nekoliko desetaka različitih satDNA unutar pojedinih genoma (Ruiz-Ruano i sur. 2016, 2017, Palacios-Gimenez i sur. 2017), nedavno je u području za obuhvaćanje skupa različitih satDNA prisutnih u genomu pojedine vrste predložen naziv „satelitom“ (Ruiz-Ruano i sur. 2016).

U evolucijskom smislu, postavlja se pitanje koliko je satelitni profil ili satelitom jedne vrste uistinu vrsno specifičan. Hipoteza takozvane „satelitne knjižnice“ (engl. „*library hypothesis*“), prvi puta predložena od strane Fry i suradnika (1977), pretpostavlja kako srodne vrste dijele čitav skup satelitnih sekvenci različitih konzerviranih satDNA porodica, prisutnih u zajedničkom pretku, koje su zbog mehanizma usklađene evolucije prisutne u različitom broju kopija kod različitih vrsta, formirajući tako profil satDNA specifičan za određenu vrstu (Meštrović i sur. 2006). Prvi eksperimentalni podaci koji u potpunosti podržavaju tu pretpostavku dobiveni su proučavanjem satDNA u rodu *Palorus*. Usporedbom varijanti osnovnih ponavljajućih jedinica satDNA zabilježena je visoka konzerviranost nukleotidne sekvence među srodnim vrstama, te odsutnost vrsno-specifičnih mutacija (Meštrović i sur. 1998). Iako postoje znatne varijacije stupnja sličnosti satDNA između srodnih vrsta, evolucija ovih sekvenci slijedi načela usklađene evolucije (Dover 1986). Usklađena evolucija ponovljenih sekvenci rezultira većom srodnošću satelitnih monomera unutar vrste, nego što je sličnost između monomera porijeklom iz različitih vrsta. Takvo stanje posljedica je homogenizacije novonastalih varijanti monomera unutar genoma jedinke, kao i istovremenog fiksiranja novonastale varijante među jedinkama reproduktivno povezane zajednice (Dover, 1986). Među mehanizmima koji dovode do umnožavanja i širenja satDNA u genomu svakako su najvažniji nesimetrično prekriženje (engl. *crossing-over*) i genska konverzija, no također je značajno i umnažanje mehanizmom "kotrljajućeg kruga" (engl. *rolling-circle mechanism*) kao i mehanizmi zasnovani na transpoziciji (Plohl i sur. 2012). Genska konverzija je proces kojim se mutacije nastale u jednom monomeru šire na drugi, ne mijenjajući time broj ponovljenih jedinica. Nesimetrično prekriženje je međutim recipročni rekombinacijski proces koji širi promjene mijenjajući pritom i broj kopija te se smatra odgovornim za česte promjene u duljini ponavljajućih nizova (Smith 1976). Budući da su otkrivene i izvankromosomske kružne molekule DNA sastavljene od satelitnih jedinica (Cohen i sur. 2003), pretpostavlja se da bi se dugi satelitni nizovi mogli umnažati izvankromosomski, putem replikacije "kotrljajućeg kruga", nakon koje bi uslijedilo njihovo ponovno umetanje u genom (Ohki i sur. 1995).

Zbog ograničenja najčešće korištenih tehnologija za sekvenciranje i softverskih alata za analizu, satDNA i dalje spadaju u najneistraženije dijelove eukariotskog genoma (Treangen i sur. 2011). Istraživanje satDNA predstavlja izazov iz nekoliko razloga. Većina modernih tehnologija za sekvenciranje kao rezultat daje kratka očitavanja sekvence koja onemogućavaju uključivanje dugačkih ponavljajućih nizova prilikom sastavljanja genoma. Dodatno, prilikom

pripremanja biblioteka koristi se metoda lančane reakcije polimerazom (engl. *polymerase chain reaction*) koja je pristrana s obzirom na umnažanje sekvenci bogatih GC parovima baza što rezultira nepravilnom zastupljenošću nekih satDNA prije samog sekvenciranja. Nadalje, mnoge računalne metode dizajnirane su za analizu repetitivnih sekvenci u već sastavljenim genomima kao npr. Tandem Repeat Finder (Benson 1999) ili koriste strategije za detekciju koje se oslanjaju na sličnosti s već poznatim repetitivnim sekvencama kao npr. RepeatMasker (Smit i sur. 2015), što onemogućava detekciju novih satDNA. Zbog navedenih problema razvili su se različiti novi pristupi za analiziranje satDNA.

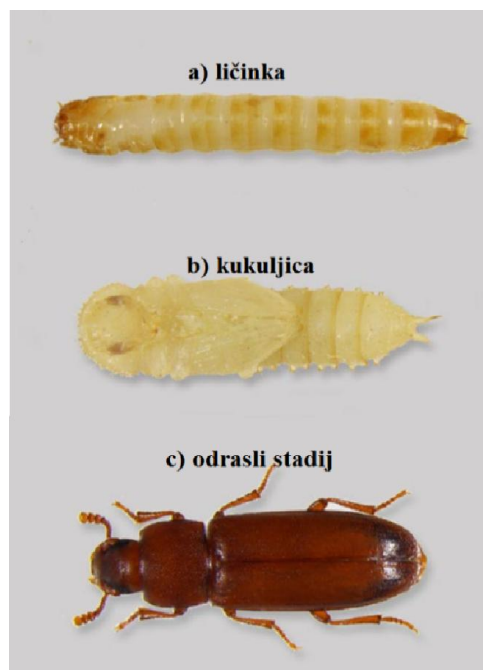
Danas se analiza satDNA uglavnom oslanja na sekvenciranje s velikim protokom podataka (engl. *High-throughput Sequencing*) koristeći sekvenciranje nove generacije (engl. *Next Generation Sequencing*, NGS) (Garrido-Ramos 2015). Kod vrsta kojima nedostaje referentni genom, *de novo* identifikaciju repetitivnih sekvenca DNA omogućava bioinformatički alat RepeatExplorer (Novák i sur. 2010, 2013). Nadalje, iz računalnog programa RepeatExplorer razvijena je njegova inačica TAREAN (akronim od engl. *Tandem Repeat Analyzer*) (Slika 1), koja omogućava identifikaciju potencijalnih satDNA iz kratkih, neposloženih odječaka DNA dobivenih metodom NGS, a također je izrazito učinkovita u otkrivanju repetitivnih sekvenci zastupljenih u malom broju kopija (Novák i sur. 2017).



Slika 1. Shematski prikaz tijeka rada računalnog programa TAREAN (preuzeto i modificirano iz Novák i sur. 2017).

1.2. Kestenjasti brašnar *Tribolium castaneum*

Kukac *Tribolium castaneum* (Herbst 1797), poznat pod nazivom kestenjasti brašnar, pripadnik je porodice crnokrilaca (Tenebrionidae) iz reda kornjaša (Coleoptera). Njegova veličina varira od 2,3 do 4,4 mm. Tijekom razvoja prolazi tri stadija: ličinku, kukuljicu i odrasli stadij (Slika 2). Generacijsko vrijeme ovisno je o temperaturi i vlažnosti te traje najkraće 20 dana na temperaturi od 37,5 °C do više od 140 dana na temperaturi manjoj od 20 °C (Howe 1956). Ovaj kozmopolitski rasprostranjen štetnik skladištenih žitarica i hrane predmet je laboratorijskih i genetičkih istraživanja više od pet desetljeća (Sokoloff 1972). Slično kao vinska mušica *Drosophila melanogaster* koja pripada redu Diptera, brašnar *T. castaneum* ima sve potrebne karakteristike modelnog organizma: lak uzgoj, kratko generacijsko vrijeme, prolazak holometabolije, brojno potomstvo te je lako podložan genetičkim manipulacijama. Potencijal kukca *T. castaneum* za genetičke analize dokazan je eksperimentima interferencije RNA (Tomoyasu i sur. 2008), molekularnim kartiranjem na razini cijelog genoma te klasičnim mutacijskim istraživanjima (Lorenzen i sur. 2005). Genom vrste *T. castaneum* sekvenciran je prije deset godina pristupom „shotgun“ sekvenciranja (Richards i sur. 2008) što je uvelike olakšalo molekularna i genetička istraživanja ove vrste. Međutim, zbog eksperimentalne strategije primijenjene prilikom sekvenciranja, otprilike 20% genoma, prvenstveno zahvaćajući (peri)centromerni heterokromatin, još uvijek nije sastavljeno.



Slika 2. Različiti životni stadiji kukca *Tribolium castaneum* (preuzeto i modificirano iz Khan i sur. 2016).

1.3. Satelitne DNA vrste *T. castaneum*

Porodica crnokrilaca Tenebrionidae, čiji je pripadnik i kukac *T. castaneum*, poznata je po visokom udjelu satDNA u genomu svojih vrsta (Ugarković i sur. 1995). Zajednička karakteristika vrsta porodice Tenebrionidae je istaknuta regija pericentromernog heterokromatina koja je zabilježena na svim kromosomima te varira od 25% do 58% cijele duljine kromosoma (Smith 1952, Weith 1985, Juan i Petitpierre 1989). Unutar roda *Tribolium* veliki dio genoma pojedinih vrsta sačinjen je od uzastopno ponavljajućih sekvenci. Primjerice, udio glavne satDNA čini 31% genoma vrste *Tribolium freemani* (Juan i sur. 1993), 40% genoma vrste *Tribolium confusum* (Plohl i sur. 1993), 13% genoma vrste *Tribolium anaphe* (Mravinac i sur. 2004) te 21% genoma vrste *Tribolium brevicornis* (Mravinac i sur. 2005). Dvije dominantne satDNA pronađene su u vrsti *Tribolium madens*, ukupno čineći 34% genoma (Ugarković i sur. 1996, Mravinac i Plohl 2007), dok njima srodne satDNA u vrsti *Tribolium audax* čine gotovo 60% ukupne genomske DNA (Mravinac i Plohl 2010). Zajedničko je svim spomenutim satDNA da se nalaze u regijama pericentromernog heterokromatina. Ugarković i sur. (1996) klasičnim su eksperimentalnim metodama detektirali i okarakterizirali glavni satelit vrste *T. castaneum*, procijenivši njegov udio na 17% ukupne genomske DNA. Otkrivena satDNA odlikovala se karakteristikama uobičajenim za satDNA opisanim u drugim vrstama iz roda *Tribolium*: zastupljenost u velikom broju kopija, smještaj u (peri)centromernom heterokromatinu svih kromosoma te visoki udio AT parova baza (Ugarković i sur. 1996). Sekvenciranje genoma vrste *T. castaneum* metodom shotgun sekvenciranja (Richards i sur. 2008) otvorilo je mogućnost za daljnje bioinformatičke analize. Wang i sur. (2008) analizirali su novosekvencirani genom vrste *T. castaneum* koristeći tri komplementarna pristupa uz upotrebu računalnih platformi: Tandem Repeat Finder (Benson 1999), TEpipe (Biedler i Tu 2003) i RepeatScout (Price i sur. 2005). Pokazano je kako otprilike 30% asembliranog genoma vrste *T. castaneum* čini repetitivna DNA, s time da 17% čine uzastopna ponavljanja, a 83% ponavljanja raspršena po genomu uključujući i mobilne elemente te da dolazi do nakupljanja repetitivne DNA u područjima pericentromernog heterokromatina kao što je i zabilježeno u Juan i Petitpierre (1989). Iako je centromerna satDNA nedovoljno zastupljena u sastavljenom genomu vrste *T. castaneum*, Wang i sur. (2008) otkrili su značajan udio uzastopnih ponavljanja s osnovnim ponavljajućim jedinicama dužim od 100 pb. Ovi rezultati ukazuju na postojanje određenog broja satDNA kandidata koje su raspoređene izvan centromernih regija, konkretno u eukromatinskim dijelovima kromosoma. Uzimajući u obzir generalni nedostatak informacija o uzastopnim ponavljanjima u eukromatinskim regijama, Pavlek i sur. (2015) analizirali su devet najzastupljenijih porodica

uzastopnih ponavljanja (Cast1-Cast9) koje su zabilježene u sastavljenom genomu te su kombinacijom bioinformatičkih i eksperimentalnih analiza otkrili kako one čine 4,3% ukupnog genoma, a pretežno su izgrađene od monomera duljine 170 pb u ≥ 5 uzastopnih ponavljanja. Međutim, zbog eksperimentalne strategije primijenjene prilikom sekvenciranja genoma vrste *T. castaneum* na temelju kojeg je sastavljen trenutno dostupan genom ove vrste, postoji velika vjerojatnost da neke od satelitnih sekvenca iz heterokromatinskih (peri)centromernih regija nisu detektirane. Do sada genom vrste *T. castaneum* nije sekvenciran pomoću metode NGS. Metoda NGS ima mogućnost sekvenciranja DNA odsječaka neovisno o njihovom smještaju unutar genoma te time omogućava detekciju sekvenca koje se nalaze unutar heterokromatinskih regija. Razvoj suvremenih bioinformatičkih alata specijaliziranih za analizu NGS podataka omogućava pak identifikaciju i karakterizaciju satDNA iz neposloženih kratkih odsječaka te se time otvara mogućnost *in silico* identifikacije uzastopno ponovljenih sekvenca, zastupljenih u relativno niskom broju kopija.

1.4. Cilj istraživanja

Cilj ovog istraživanja bio je bioinformatičkom analizom kratkih odsječaka DNA, dobivenih sekvenciranjem genoma vrste *T. castaneum* pomoću metode NGS, identificirati potencijalne, dosad neutvrđene satelitne DNA sekvence. Nastavno bioinformatičkoj analizi, upotrebom niza klasičnih molekularnih metoda željelo se istražiti genomsku organizaciju i kromosomski smještaj najprominentnijih satelitnih kandidata dobivenih analizom *in silico*. Ovakvim pristupom, kombinirajući bioinformatičke alate i eksperimentalne metode, uz karakterizaciju satelitoma vrste *T. castaneum*, cilj je bio potvrditi primjenjivost kratkih, neposloženih NGS sljedova u otkrivanju niskozastupljenih satelita, koji pak mogu biti značajni za razumijevanje strukture i funkcije genoma.

2. MATERIJALI I METODE

2.1. Materijali

2.1.1. Životinjski materijal

U eksperimentalnom dijelu rada korišteni su kukci vrste *Tribolium castaneum* Herbst, soj GA2. Spomenuti soj potječe iz SAD-a, sakupljen je 1982. godine u državi Georgia, a korišten je kao referentni soj prilikom sekvenciranja genoma vrste *T. castaneum* (Richards i sur. 2008). Inicijalna kultura u Laboratoriju za strukturu i funkciju heterokromatina dobivena je ljubaznošću Merrilee Susan Haas iz inektarija Agricultural Research Service (Manhattan, Kansas, SAD) pri United States Department of Agriculture te se potom nastavila propagirati kao laboratorijska kultura.

2.1.2. Osnovne kemikalije

Osnovne kemikalije korištene u ovom radu:

- etidijev bromid (Serva)
- etanol, NaOH, očetena kiselina, KCl, NaCl, Na-citrat, Na-acetat, Na₂HPO₄, KH₂PO₄ (Kemika)
- agaroz, orange G, etilendiamintetraoctena kiselina (EDTA), Tris, formaldehid, formamid (Sigma)
- 5-brom-4-klor-3-indolil-β-D-galaktozid (X-gal), izopropil-β-D-tiogalaktozid (IPTG) (GIBCO BRL)

2.1.3. Puferi i otopine

Puferi i otopine korištene u ovom radu:

- puffer 50x TAE (akronim od engl. *Tris-Acetate-EDTA*) (pH 8,0): 2 M Tris, 1 M octena kiselina, 50 mM EDTA
- tekuća hranjiva podloga LB (akronim od engl. *Luria Broth*) (pH 7,0): bakto tripton (10 g/L), ekstrakt kvasca (5 g/L), NaCl (10 g/L)

- tekuća hranjiva podloga SOC (akronim od engl. *Super Optimal Broth with Catabolite Repression*): 2% bakto tripton, 0,5% ekstrakt kvasca, 10 mM NaCl, 2,5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glukoza
- kruta hranjiva podloga LB s antibiotikom (pH 7,0): LB agar (40 g/L) i ampicilin (0,1 µg/L)
- pufer G (pufer za izolaciju genomske DNA): 0,1 M NaCl, 0,01 M Tris-HCl (pH 8,0), 25 mM EDTA (pH 8,0), 0,5% SDS (akronim od engl. *Sodium Dodecyl Sulfate*)
- fenol:kloroform:izoamilni alkohol (25:24.1)

Otopine za fluorescencijsku hibridizaciju *in situ* (FISH):

- denaturacijska otopina: 70% formamid u puferu 2x SSC (akronim od engl. *Saline Sodium Citrate*)
- pufer DeSO₄ (pH 7,0): 4x SSC, 20% dekstran-sulfat, 50 mM Na-fosfat
- hibridizacijska otopina: 60% deionizirani formamid, 40% DeSO₄ pufer
- pufer za ispiranje: 50% formamid u puferu 2x SSC
- pufer 4M: 4x SSC, 5% blokirajući reagens
- pufer 4T: 4x SSC, 0,05% Tween 20
- fiksativ (mješavina octene kiseline i etanola u omjeru 1:3)
- otopina DAPI-a (akronim od engl. *4',6-diamidino-2-phenylindole dihydrochloride*) u 2x SSC-u: 50 ng/ml, 2x SSC
- pufer 20x SSC (pH 7,0): 3M NaCl, 0,3 M Na-citrat
- pufer PBS (akronim od engl. *Phosphate Buffered Saline*) (pH 7,4): 137 mM NaCl, 2,7 mM KCl, 10,1 mM Na₂HPO₄, 1,8 mM KH₂PO₄
- 70% hladni etanol
- 90% hladni etanol
- 100% hladni etanol
- 2x SSC u 50% formamidu

- pufer 1x TE (akronim od engl. *Tris-EDTA*): 10 mM Tris-HCl (pH 8,0), 1 mM EDTA (pH 8,0)

Otopine za hibridizaciju metodom prema Southernu:

- hibridizacijski pufer: 0,25 M fosfatni pufer (pH 7,2), 1 mM EDTA (pH 8,0), 20% SDS, 0,5% blocking reagent

- pufer za ispiranje: 20 mM Na₂HPO₄, 1 mM EDTA, 1% SDS

- pufer 1 (pH 8,0): 0,1 M maleinska kiselina, 3 M NaCl, 0,3% Tween 20

- pufer 2: u puferu 1 otopljen blocking reagent finalne količine 1%

- pufer 3: 0,1 M Tris-HCl (pH 9,5), 0,1 M NaCl

- pufer za skidanje sonde: 0,2 M NaOH, 0,1% SDS

- otopina za kemiluminiscenciju CDP-star (akronim od engl. *Disodium 2-chloro-5-(4-methoxySpiro[1,2-dioxetane-3,2'-(5-chlorotricyclo[3.3.1.1^{3,7}]decan])-4-yl]-1-phenyl phosphate*) razrijeđen u puferu 3 u omjeru 1:100

Otopine za kapljičnu analizu:

- denaturacijska otopina: 0,4 M NaOH

- neutralizacijska otopina: 0,5 M Tris-HCl (pH 7,4), 1,5 M NaCl

2.1.4. Kompleti reagensija

Kompleti reagensija korišteni u ovom radu:

- komplet „QIAquick Gel Extraction Kit“ za pročišćavanje odsječaka DNA iz agaroznog gela (QIAGEN)

- komplet „GoTaq Flexi DNA Polymerase“ za lančanu reakciju polimerazom, koji sadržava: DNA polimerazu GoTaq (5U/μl), 5x koncentrirani pufer GoTaq Flexi i 25 mM MgCl₂ (Promega)

- komplet „GoTaq Green Master Mix“ za lančanu reakciju polimerazom, koji sadržava u 2x koncentriranom reakcijskom puferu Green GoTaq: DNA polimerazu GoTaq , 400 µM smjesu dNTP i 3mM MgCl₂ (Promega)
- komplet „QIAquick PCR Purification Kit“ za pročišćavanje produkata reakcije PCR-a (QIAGEN)
- komplet „pGEM-T Easy Vector System I“ za kloniranje produkata PCR-a koji sadržava: plazmidni vektor pGEM-T (50 ng/µl), enzim T4 DNA ligazu (3 Weiss U/µl) i 2x koncentrirani pufer za T4 DNA ligazu (Promega)
- komplet „High Pure Plasmid Isolation Kit“ za izolaciju plazmida (Roche Applied Science)

2.1.5. Enzimi

Enzimi korišteni u ovom radu:

- restrikcijska endonukleaza Fast Digest *EcoRI* s pripadajućim puferom za restrikcijsku analizu (Fermentas)
- restrikcijske endonukleaze s pripadajućim puferima za restrikcijsku analizu *MvaI* (*BstNI*) (Thermo Fisher Scientific), *AluI* (New England Biolabs), *RsaI* (New England Biolabs), *PvuII* (Fermentas), *HincII* (Roche Applied Science), *HaeIII* (*BsuRI*) (Fermentas), *DraI* (Roche Applied Science), *KpnI* (Fermentas)
- proteinaza K, RNaza A, pepsin (Roche Applied Science)

2.1.6. Elektroforetski biljezi

Elektroforetski biljezi korišteni u ovom radu:

- O'GeneRuler DNA Ladder Mix (Fermentas), koji sadrži smjesu odsječaka DNA duljine 100 pb, 200 pb, 300 pb, 400 pb, 500 pb, 600 pb, 700 pb, 800 pb, 900 pb, 1000 pb, 1200 pb, 1500 pb, 2000 pb, 2500 pb, 3000 pb, 3500 pb, 4000 pb, 5000 pb, 6000 pb, 8000 pb i 10000 pb
- 2-Log DNA Ladder (New England Biolabs), koji sadrži smjesu odsječaka DNA duljine 100 pb, 200 pb, 300 pb, 400 pb, 500 pb, 600 pb, 700 pb, 800 pb, 900 pb, 1000 pb, 1200 pb, 1500 pb, 2000 pb, 3000 pb, 4000 pb, 5000 pb, 6000 pb, 8000 pb i 10000 pb

2.1.7. Plazmidni vektor

Za kloniranje produkata PCR-a korišten je plazmidni vektor pGEM-T iz kompleta „pGEM-T Easy Vector System I“ koji omogućuje AT-kloniranje. Taq polimeraza ugrađuje na 3' krajevima produkata PCR-a dodatan dATP, koji se u reakciji ligacije sparuje s komplementarnim timidinom na krajevima linearnog T-vektora.

Vektor pGEM-T sadrži gen za otpornost na antibiotik ampicilin (apm^r) što omogućuje selekciju transformiranih bakterija. Mjesto za kloniranje fragmenata se nalazi unutar gena lacZ koji kodira enzim β -galaktozidazu pa je putem α -komplementacije omogućena plavo-bijela selekcija transformanata.

2.1.8. Bakterijski soj

Za umnažanje plazmida korištene su ultrakompetentne bakterijske stanice vrste *Escherichia coli* soja XL10-Gold ($\text{Tet}^r \Delta(\text{mcrA})183 \Delta(\text{mcrCB-hsdSMR-mrr})173\text{endA1supE44 thi-1recA1gyrA96relA1lac Hte [F' proAB lacI}^q\Delta\text{MI5 Tn10 (Tet}^r \text{) Amy Cam}^r\text{] (Stratgene)}$).

2.1.9. Tehnička oprema i uređaji

Tehnička oprema i uređaji korišteni u ovom radu:

- uređaj PCR 2720 Thermal Cycler (Applied Biosystems)
- uređaj za horizontalnu elektroforezu DNA u agaroznom gelu (Bio-Rad)
- UV transiluminator sa sustavom za snimanje agaroznih gelova G:BOX (Syngene)
- uređaj za kapljičnu analizu Minifold I Dot-Blotter (Schleicher i Schuell)
- hibridizacijska pećnica OV1 (Biometra)
- stolna mikrocentrifuga Minispin (Eppendorf)
- stolna tresilica Vibramax (Heidolph)
- termoblok (Bioblock Scientific)
- uređaj za brzo sušenje uzoraka Speed vac DNA120 (Thermo)

- sušionik (Shel Lab)
- termostatski inkubator (Sutjeska)
- vodena kupelj (Inko)
- uređaj za zataljivanje plastičnih vrećica (Gorenje)
- fotografsko crveno svjetlo (Kaiser)
- uređaj za fluorimetrijsko mjerenje koncentracije DNA Qubit (Invitrogen)
- konfokalni laserski pretražni mikroskop Leica TCS SP8 X (Leica Microsystems)

2.1.10. Ostali materijali

Ostali materijali korišteni u ovom radu:

- pozitivno nabijene najlonske membrane (Roche Applied Science)
- papir Whatman 3MM
- papirnati ubrusi
- RF4 medicinski Röntgen filmovi NIF 18x 43,2 cm (Fotokemika)
- EFKE FR-210 razvijач za ručnu obradu Röntgen filmova (Fotokemika)
- EFKE FF-210 koncentrirani tekući fiksir za ručnu obradu svih vrsta Rontgen filmova (Fotokemika)

2.2. METODE

2.2.1. Izolacija genomske DNA

Ukupna genomska DNA izolirana je iz odraslih jedinki vrste *Tribolium castaneum* prema protokolu za izolaciju genomske DNA iz tkiva kukaca pomoću organskih otapala. Tkivo je smrznuto u tekućem dušiku, smrvljeno u tarioniku te inkubirano ~18 h na temperaturi od 50 °C u vodenoj kupelji u puferu za izolaciju genomske DNA (1 ml pufera/100 mg tkiva) uz dodatak proteinaze K (0,1 mg/ml). Nakon dodatka jednakog volumena fenola (pH 8,0) otopina je promiješana invertnim okretanjem dok se nije stvorila

emulzija, a potom centrifugirana 9 min na 12 000 rpm. Nakon centrifugiranja vodena je faza prebačena u novu mikroeprevetu te je isti postupak ponovljen još dva puta. Sljedeće ponavljanje provedeno je s jednakim volumenom otopine fenol:kloroform:izoamilni alkohol u omjeru 25:24:1. Genomska DNA iz gornje, vodene faze uzorka istaložena je pomoću 0,1 početnog volumena 3 M Na-acetata (pH 5,2) i 2 početna volumena 100%-tnog ledenog etanola. Nakon centrifugiranja u trajanju od 15 min na 12 000 rpm, uklonjen je supernatant, dodan je 75%-tni etanol te je uzorak ponovno centrifugiran 7 min na 12 000 rpm. Supernatant je uklonjen vakuum sisaljkom, a talog je osušen na sobnoj temperaturi (~30 min) te otopljen u 1x TE puferu. Nakon inkubacije uzorka u otopini RNaze A (10 µg/ml) i SDS-a (0,1%) tijekom jednog sata u vodenoj kupelji na 37 °C, provedena je ekstrakcija pomoću otopine fenol:kloroform:izoamilni alkohol. Uzorak je ponovno istaložen, kao što je prethodno opisano 0,1 volumenom 3 M Na-acetata i 2 volumena etanola, 10 min na sobnoj temperaturi te centrifugiran 15 min na 12 000 rpm. Nakon centrifugiranja supernatant je uklonjen vakuum sisaljkom, dodan je 75%-tni etanol, nakon čega je uzorak ponovno centrifugiran 7 min na 12 000 rpm. Nakon centrifugiranja supernatant je uklonjen te je talog osušen na sobnoj temperaturi i otopljen u puferu 1x TE. Količina i kakvoća izolirane ukupne genomske DNA procijenjene su elektroforezom u 1% agaroznom gelu usporedbom s DNA poznate koncentracije. Dodatno je koncentracija DNA određena fluorometrijskim mjerenjem na uređaju Qubit.

2.2.2. Razgradnja DNA restrikcijskim endonukleazama

Razgradnja genomske DNA restrikcijskim endonukleazama provedena je prema uputama proizvođača, uz korištenje 10 U enzima po µg DNA. Uzorci genomske DNA, razgrađeni restrikcijskim endonukleazama razdvojeni su u 1% agaroznom gelu u puferu 1x TAE uz dodatak etidijevog bromida (0,5 µg/ml). Uzorci DNA, nanoseni u jažice zajedno s otopinom Orange G u omjeru 6:1, podvrgnuti su elektroforezi pri naponu od 60 V. Kao elektroforetski biljezi korišteni su O'GeneRuler DNA Ladder Mix (Fermentas) i 2-Log DNA Ladder (New England Biolabs). Nakon elektroforetskog razdvajanja uzoraka, agarozni gelovi fotografirani su pomoću uređaja G:BOX, koji također omogućava i računalnu pohranu fotografije gela.

2.2.3. Umnažanje odsječaka DNA metodom lančane reakcije polimerazom

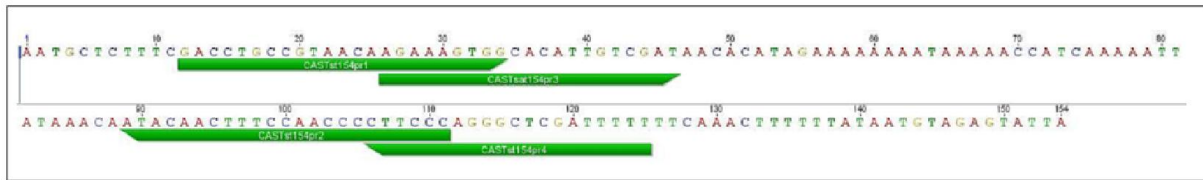
Odsječci potencijalnih satelitnih DNA umnoženi su metodom lančane reakcije polimerazom (PCR).

2.2.3.1. Početnice

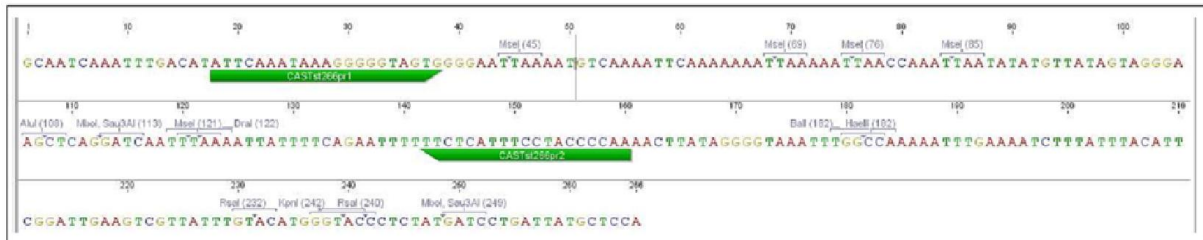
Korištene su početnice za umnažanje odsječaka tri različite potencijalne satelitne DNA: TCst154, TCst206 i TCst309. Nukleotidni slijedovi početnica navedeni su u tablici 1. Mjesta vezanja početnica duž satelitnih monomera TCst154, TCst206 i TCst309 prikazane su na slici 3, a u tablici 1 navedene su očekivane duljine monomernih, dimernih i trimernih odsječaka DNA umnoženih metodom PCR. Navedene oligonukleotidne početnice sintetizirane su u korisničkom servisu Macrogen Europe (Amsterdam, Nizozemska). Pomoću gradijentnog PCR-a određene su optimalne temperature sparivanja pojedinog para početnica s kalupom DNA.

Tablica 1. Parovi početnica korišteni pri umnažanju satelitnih DNA TCst154, TCst266 i TCst309 vrste *Tribolium castaneum* metodom PCR te pretpostavljene duljine umnoženih odsječaka.

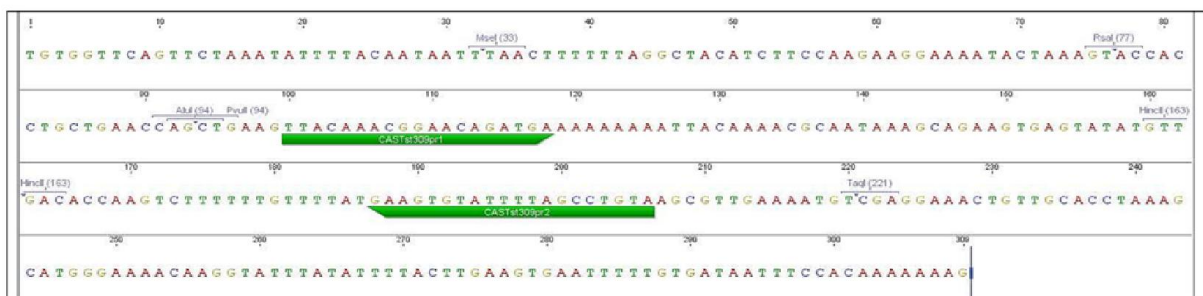
satelitna DNA	nukleotidni slijed početnica	duljina umnoženih odsječaka
TCst154	CASTst154pr3 5'-AGAAAGTGGCACATTGTCGAT-3'	99 pb (TCst154 PCR „monomer“)
	CASTst154pr4 5'-AAAAATCGAGCCCTGGGAAG-3'	253 pb (TCst154 PCR „dimer“)
		407 pb (TCst154 PCR „trimer“)
TCst266	CASTst266pr1 5'-ATTCAAATAAAGGGGGTAGTG-3'	143 pb (TCst266 PCR „monomer“)
	CASTst266pr2 5'-TTGGGGTAGGAAATGAGAA-3'	409 pb (TCst266 PCR „dimer“)
		675 pb (TCst266 PCR „trimer“)
TCst309	CASTst309pr1 5'-TTACAAACGGAACAGATGA-3'	107 pb (TCst309 PCR „monomer“)
	CASTst309pr2 5'-TACAGGCTAAAATACACTTC-3'	416 pb (TCst309 PCR „dimer“)
		725 pb (TCst309 PCR „trimer“)



TCst154



TCst266



TCst309

Slika 3. Mjesta vezanja parova početnica na monomere satelitnih DNA TCst154, TCst206 i TCst309.

2.2.3.2. Reakcijske smjese PCR-a i uvjeti reakcija

Umnažanja satelitnih DNA metodom PCR-a provedena su u ukupnom volumenu reakcijskih smjesa od 10 μ l. Smjesa od 10 μ l je sadržavala: 5x GoTaq Flexi pufer, 2,5 mM MgCl₂, 0,2 mM smjese dNTP, 0,4 μ M specifičnih početnica, 0,25 U GoTaq polimeraze i 10 ng genomske DNA.

Programi ciklusa umnažanja pojedinih satelitnih DNA navedeni su u tablicama 2, 3 i 4.

Tablica 2. Program ciklusa umnažanja satelitne DNA TCst154 vrste *Tribolium castaneum* metodom PCR.

PCR korak	temperatura	vrijeme	broj
početna denaturacija	95 °C	3 min	1
denaturacija	95 °C	15 s	35
sparivanje	43 °C	15 s	35
produljenje	72 °C	30 s	35
završno produljenje	72 °C	10 min	1

Tablica 3. Program ciklusa umnažanja satelitne DNA TCst266 vrste *Tribolium castaneum* metodom PCR.

PCR korak	temperatura	vrijeme	broj
početna denaturacija	94 °C	3 min	1
denaturacija	94 °C	15 s	35
sparivanje	46 °C	15 s	35
produljenje	72 °C	40 s	35
završno produljenje	72 °C	10 min	1

Tablica 4. Program ciklusa umnažanja satelitne DNA TCst309 vrste *Tribolium castaneum* metodom PCR.

PCR korak	temperatura	vrijeme	broj
početna denaturacija	94 °C	3 min	1
denaturacija	94 °C	15 s	35
sparivanje	53 °C	15 s	35
produljenje	72 °C	40 s	35
završno produljenje	72 °C	10 min	1

2.2.4. Elektroforetsko razdvajanje odsječaka DNA u agaroznom gelu

Odsječci DNA nakon reakcije PCR razdvojeni su elektroforezom u 1%-tnom agaroznom gelu. Gel je pripremljen otapanjem agaroze u puferu 1x TAE uz dodatak etidijevog bromida (0,5 µg/ml) za vizualizaciju DNA. Elektroforeza se odvijala u istom puferu pri sobnoj temperaturi. Duljina odsječaka procijenjena je prema elektroforetskom bilježu „O'GeneRuler DNA Ladder Mix“, dok se za procjenu količine DNA koristila λ-DNA poznate koncentracije.

Nakon elektroforetskog razdvajanja gelovi su fotografirani pomoću uređaja G:BOX.

2.2.5. Izolacija odsječaka DNA iz agaroznog gela

Dijelovi gela s odabranim odsječcima DNA su izrezani na transiluminatoru i izvagani, a zatim je DNA izolirana korištenjem kompleta „QIAquick Gel Extraction Kit“ prema uputama proizvođača.

2.2.6. Kloniranje odsječaka DNA

Za kloniranje produkata PCR-a korišten je plazmidni vektor pGEM-T iz kompleta „pGEM-T Easy Vector System I“ koji omogućuje AT-kloniranje.

2.2.6.1. Ligacija odsječaka DNA i plazmidnog vektora

Reakcije ligacije izoliranih odsječaka DNA i vektora pGEM-T Easy provedene su pomoću enzima T4 DNA ligaza u 2x Rapid ligacijskom puferu na 4 °C preko noći. Omjer fragmenata DNA i vektora u reakciji iznosio je 3:1.

2.2.6.2. Transformacija kompetentnih bakterijskih stanica

Kompetentne bakterijske stanice vrste *E. coli* transformirane su vektorom pGEM-T Easy metodom temperaturnog šoka. U 25 µl otopine ultrakompetentnih stanica na ledu dodano je 2 µl ligacijske smjese i inkubirano 30 minuta. Stanice su podvrgnute temperaturnom šoku u trajanju od 30 sekundi inkubacijom otopine stanica i ligacijske smjese u vodenoj kupelji na temperaturi od 42 °C te ponovnim vraćanjem na led 2 minute. Stanice su regenerirane sat i pol vremena u 250 µl prethodno ugrijanog SOC medija uz neprekidnu trešnju od 250 rpm-a na temperaturi od 37 °C, nakon čega su nasadene na selektivne krute podloge LB, s dodatkom ampicilina (100 µl/ml), 40 µl induktora IPTG (100 mM) te 40 µl supstrata X-gal (20 mg/ml).

2.2.6.3. PCR na bakterijskim kolonijama

Korištenje plazmidnog vektora pGEM-T easy omogućava plavo-bijelu selekciju kolonija. U svrhu određivanja veličina ugrađenih odsječaka provodio se PCR na bakterijskim kolonijama (engl. *colony PCR*). Sterilnim nastavkom dotaknuta je određena bakterijska kolonija te je uronjena u mikroeprevetu s 10 µl mQ-H₂O i dobro propipetirana. Uzorak je denaturiran u uređaju za PCR pri 95 °C 10 minuta, zatim je ohlađen na ledu i kratko centrifugiran. Reakcija PCR-a provedena je u volumenu od 5 µl reakcijske smjese koja je sadržavala 2,5 µl 2x GoTaq Green Master Mix, 0,5 µl bakterijske denaturirane kolonije, 10 µM M13F početnice (5'-GTAAAACGACGGCCAGT-3') i M13R-pUC(-40) početnice (5'-CAGGAAACAGCTATGAC-3') te 1,6 µl mQ-H₂O. Program umnažanja započinjao je početnom denaturacijom na 94 °C 2 minute, zatim je uslijedilo 30 ciklusa koji su se sastojali

od denaturacije 15 sekundi na 94 °C, sparivanja početnica 15 sekundi na 54 °C i produljivanja lanaca 30 sekundi na 72 °C. Produkti reakcije PCR-a na bakterijskim kolonijama analizirani su u 1% agaroznom gelu. Kolonije koje su posjedovale odsječke DNA očekivane duljine uzgojene su u 4 ml tekuće selektivne podloge LB s dodatkom ampicilina (100 µg/ml) preko noći na temperaturi od 37 °C uz neprekidnu vrtnju od 225 rpm.

2.2.6.4. Izolacija plazmidne DNA

Bakterije s ugrađenim željenim odsječkom uzgojene su preko noći na tekućoj selektivnoj podlozi LB te oborene centrifugiranjem. Plazmidna DNA zatim je izolirana upotrebom kompleta „High Pure Plasmid Isolation Kit“ prema uputama proizvođača, a finalno isprana s kolone sa 100 µl elucijskog pufera. Koncentracija i izgled izolirane plazmidne DNA provjereni su elektroforezom u agaroznom gelu usporedbom s λ-DNA poznate koncentracije i elektroforetskim biljgom „O'GeneRuler DNA Ladder Mix“.

2.2.6.5. Određivanje nukleotidnih sljedova kloniranih odsječaka DNA

Nukleotidni sljedovi kloniranih odsječaka DNA određeni su u servisu za sekvenciranje Macrogen Europe (Amsterdam, Nizozemska) uz upotrebu spomenutih plazmidnih početnica M13F i M13R-pUC(-40).

2.2.7. Obilježavanje hibridizacijskih sonda

Sonde za detekciju potencijalnih satelitnih DNA metodom hibridizacije prema Southernu obilježene su digoksininom uz pomoć metode lančane reakcije polimerazom gdje se kao kalup DNA koristila smjesa određenih klonova specifičnih za pojedinu satelitnu DNA. Konkretno, za TCst154 u smjesi klonova nalazili su se klonirani odsječci cl1, 2, 16, 25 i 32, za TCst266 klonirani odsječci cl27, 32, 40, 47 i 50 te za TCst309 klonirani odsječci cl2, 10, 31, 32, 34. Princip ove metode temelji se na korištenju smjese dNTP-ova u kojoj je molekula digoksinina vezana na dUTP, koji se enzimatski ugrađuje u nosivost sintetizirane lance DNA sonde. Umnažanje i obilježavanje potencijalnih satelitnih DNA izvedeno je u reakcijskoj smjesi volumena 20 µl koja je sadržavala 5x GoTaq Colorless pufer, 2,5 mM MgCl₂, 0,5 mM dATP, 0,5 mM dGTP, 0,5 mM dCTP, 0,375 mM dTTP, 0,125 mM DIG-

dUTP, 5 U GoTaq DNA polimeraze, 1 ng kalupa DNA i 10 μ M specifičnih početnica. Programi umnažanja za pojedine satelitne DNA navedeni su u tablicama 2, 3 i 4.

Sonde za detekciju potencijalnih satelitnih DNA metodom fluorescencijske *in situ* hibridizacije obilježavane su biotinom uz pomoć metode lančane reakcije polimerazom, pri čemu se kao kalup DNA koristila smjesa određenih klonova specifičnih za pojedinu satelitnu DNA (korištene su iste smjese klonova kao i za obilježavanja sonda za hibridizaciju prema Southernu, kako je navedeno u prethodnom odlomku). Princip ove metode temelji se na korištenju smjese dNTP-ova u kojoj je molekula biotina konjugirana na dUTP te se takvi obilježeni nukleotidi enzimatski ugrađuju u novosintetizirane lance DNA sonde. Umnažanje i obilježavanje potencijalnih satelitnih DNA izvedeno je u reakcijskoj smjesi volumena 50 μ l koja je sadržavala 5x GoTaq Colorless pufer, 2,5 mM MgCl₂, 5 U GoTaq DNA polimeraze, 1 ng kalupa DNA, 10 μ M specifičnih početnica, 0,5 mM dATP, 0,5 mM dGTP, 0,5 mM dCTP, 0,32 mM dTTP, 0,18 mM biotin-dUTP. Program umnažanja započet je dvominutnom početnom denaturacijom na 94 °C, a nastavio se kroz 40 ciklusa unutar kojih su se izmjenjivala 3 koraka: denaturacija na 94 °C 15 sekundi, sparivanje početnica kroz 15 sekundi te reakcija sinteze DNA na 72 °C 30 sekundi. Ovisno o satDNA temperatura sparivanja početnica varirala je od 43 °C za TCst154, 46 °C za TCst266 do 53 °C za TCst309. Program je okončan produženom reakcijom sinteze DNA na 72 °C kroz 10 minuta, kako bi se dovršila sinteza svih započetih lanaca. Nakon završetka reakcije 25 μ l PCR smjese pročišćeno je kompletno za pročišćavanje produkata PCR reakcije QIAquick PCR Purification Kit (Qiagen), čime su uklonjene neugrađene početnice i nukleotidi koji bi mogli smetati u kasnijoj hibridizacijskoj analizi.

2.2.8. Hibridizacijska analiza prema Southernu

Hibridizacijska analiza prema Southernu metoda je kojom se utvrđuje postojanje, smještaj i duljina sekvenci DNA od interesa u npr. ukupnoj genomske DNA. Dvolančani fragmenti DNA (dobiveni restrikcijskom razgradnjom ukupne genomske DNA, plazmidne DNA ili umnožavanjem pomoću metode lančane reakcije polimerazom) razdvajaju se elektroforezom u agaroznom gelu s obzirom na razlike u duljini na način opisan u poglavlju 2.2.4. Nakon elektroforetskog razdvajanja na gel se položi najlonska membrana jednakih dimenzija te se lužnatim, kapilarnim prijenosom odsječi DNA prenose iz gela na membranu. Tijekom prijenosa, lužnati pufer, prolazeći kroz gel, denaturira dvolančane odsječke DNA te

do pozitivno nabijene najlonske membrane donosi i čvrsto veže jednolančane odsječke negativnog naboja. Skidanjem membrane s gela fragmenti DNA zadržavaju međusobno jednak položaj kao što su ga imali na gelu. Najlonska membrana s prenesenim fragmentima DNA izlaže se tijekom prekonoćne inkubacije specifičnoj sondi DNA, koja ciljano sjeda na isključivo sebi komplementarne odsječke DNA. U procesu detekcije utvrđuje se položaj specifično vezane sonde DNA tj. odsječka od interesa.

2.2.8.1. Prijenos odsječaka DNA na najlonsku membranu

Nakon elektroforetskog razdvajanja, odsječci DNA preneseni su na pozitivno nabijenu najlonsku membranu (Roche Applied Science) lužnatim transferom u otopini 0,4 M NaOH. Nakon prekonoćnog prijenosa, membrana je neutralizirana ispiranjem u 2x SSC puferu te ostavljena da se potpuno osuši na sobnoj temperaturi. Nakon što se membrana potpuno osušila, stavljena je 30 min u suhi sterilizator na 120 °C kako bi se odsječci dodatno fiksirali na najlonsku membranu.

2.2.8.2. Prehibridizacija i hibridizacija membrana

Membrane su stavljene u hibridizacijske cilindre u kojima se provodila prehibridizacija u ~15 ml prethodno zagrijanog prehibridizacijskog pufera tijekom 2 sata na 65 °C. Nakon prehibridizacijskog koraka, sama hibridizacija provedena je u svježem, prethodno zagrijanom hibridizacijskom puferu u koji je dodana prethodno denaturirana sonda DNA specifična za pojedini satelit u koncentraciji od 20 ng/ml hibridizacijskog pufera. Hibridizacija se odvijala preko noći na temperaturi od 65 °C.

2.2.8.3. Otkrivanje hibridizacijskog signala

Nakon hibridizacije membrane su isprane u prethodno zagrijanom puferu za ispiranje 3x 20 min na 63 °C u hibridizacijskom cilindru. Nakon toga je uslijedila detekcija signala. Svi koraci detekcije provedeni su u plastičnim kadicama na tresilici pri sobnoj temperaturi. Detekcija hibridizacijskog signala započeta je inkubacijom membrana 2x 5 min u puferu 1, nakon koje je uslijedila inkubacija 30 min u puferu 2. Sljedeća inkubacija provedena je 30 min u svježem puferu 2 u kojem je prethodno razrijeđen anti-DIG-AP konjugat (Roche Applied Science) u omjeru 1:20 000. Višak antitijela ispran je u puferu 1 tijekom višekratnog

ispiranja (5x 10 min). Nakon dvije inkubacije od 5 min u puferu 3, sljedeća inkubacija napravljena je u tamnoj komori. Membrane su ponovno inkubirane 10 min u svježem puferu 3 u kojem je prethodno razrijeđen supstrat za alkalnu-fosfatazu CDP-Star (Roche Applied Science) u omjeru 1:50 te zataljene u prozirnoj plastičnoj vrećici. Membrane su nakon toga izložene Röntgen filmu kroz potrebno vrijeme, ovisno o intenzitetu signala. Ekspozirani Röntgen film razvijen je provlačenjem kroz fotografski razvijatelj, vodu te fiksir.

2.2.9. Fluorescencijska hibridizacija *in situ* (FISH)

Fluorescencijska hibridizacija *in situ* je citogenetička metoda koja omogućava izravnu identifikaciju i lokalizaciju određenih sekvenci nukleinskih kiselina u stanici, odnosno na kromosomima. Princip metode počiva u specifičnom sparivanju dviju komplementarnih DNA sekvenca, od kojih jedna (sonda DNA) traži drugu (ciljno, komplementarno mjesto na kromosomu) na kromosomskom preparatu.

2.2.9.1. Priprema kromosomskih preparata

Kromosomski preparati pripremljeni su iz gonada kukuljica isključivo mužjaka vrste *T. castaneum*. Izolirano tkivo inkubirano je 1 sat u kolcemidu (10 µg/ml). Stanje hipotoničnog šoka koje dovodi do rasprsnuća stanica postignuto je inkubacijom 15 min u 0,075 M KCl, nakon čega je tkivo fiksirano 15 min u fiksativu (octena kiselina:etanol u omjeru 1:3). Trajni preparati pripremljeni su tzv. „squash“ tehnikom u 45%-tnoj otopini octene kiseline. Nakon toga preparati su u potpunosti uronjeni u tekući dušik te je s predmetnog stakalca odstranjena pokrovnica. Preparati su ostavljeni da se osuše na sobnoj temperaturi, a potom do ulaska u eksperiment FISH-a čuvani na temperaturi od -20 °C.

2.2.9.2. Prehibridizacija i hibridizacija *in situ*

Preparati su prvo 5 minuta ispirani u otopini 2x SSC na 37 °C u vodenoj kupelji, a potom tretirani 1 sat na 37 °C u otopini 2x SSC u kojoj je prethodno otopljen enzim RNaza A (100 µg/ml) kako bi se s preparata uklonile (degradirale) molekule RNA. Nakon ispiranja u otopini 2x SSC 3x 5 min na 37 °C uslijedio je tretman pepsinom (100 µg/ml) u otopini 10 mM HCl kroz 10 min na 37 °C. Nakon ispiranja, prvo u otopini PBS 2x 5 min, a potom u otopini PBS koja sadrži otopljeni 50 mM MgCl₂ na sobnoj temperaturi, slijedila je inkubacija

preparata 10 min u otopini formaldehida (2,7 ml formaldehida u 100 ml pufera PBS s dodatkom 50 mM MgCl₂). Nakon toga, preparati su isprani 5 min u otopini PBS, a potom dehidrirani prolaskom kroz seriju hladnih etanola (80% → 90% → 100%, 3 min svaki) te ostavljeni da se osuše na sobnoj temperaturi. Prije same hibridizacije bilo je potrebno denaturacijom razdvojiti dvolančane odsječke obilježenih sonda kao i odsječke DNA na samom kromosomskom preparatu. Kromosomski preparati zagrijani su nekoliko minuta na 50-60 °C na poklopcu vodene kupelji te denaturirani u denaturacijskoj otopini (70% formamid u 2x SSC) točno 2 min na 70 °C, a potom odmah dehidrirani prolaskom kroz seriju hladnih etanola (80% → 90% → 100%, 3 min svaki) te ostavljeni da se u potpunosti osuše na sobnoj temperaturi. Sto ng liofiliziranih sonda otopljeno je u 15 µl hibridizacijske otopine (60% formamid, 40% DeSO₄ pufer) te denaturirano 5 min na 75 °C. Nakon denaturacije sonde su ohlađene na ledu kroz nekoliko minuta i nanešene na kromosomske preparate. Preparati su preneseni u vlažnu komoricu, a hibridizacija se odvijala preko noći na 37 °C. Posthibridizacijsko ispiranje provedeno je u predgrijanom puferu 2x SSC koji je sadržavao 50% formamid tijekom 4x 5 min na 37 °C, nakon čega je uslijedilo ispiranje u puferu 2x SSC 3x 5 min na 37 °C.

2.2.9.3. Otkrivanje fluorescencijskog signala

Otkrivanje fluorescencijskog signala za sonde obilježene biotinom započeto je inkubacijom u puferu 4M u vlažnoj komorici 30 min na 37 °C, a zatim inkubacijom u istim uvjetima u 4M puferu koji sadrži avidin-FITC u razrjeđenju 1:500. Nakon ispiranja 3x 5 min u puferu 4T na sobnoj temperaturi provedena je amplifikacija signala inkubacijom od 20 min u vlažnoj komorici u puferu 4M koji sadrži razrijeđeno protutijelo α -avidin^{biotin} u omjeru 1:100. Suvišak protutijela uklonjen je ispiranjem u puferu 4T 3x 5 min na sobnoj temperaturi, a potom je preparat ponovno inkubiran kroz 20 min u vlažnoj komorici na sobnoj temperaturi u puferu 4M u kojem je razrijeđen avidin-FITC u omjeru 1:2000. Nakon završnih ispiranja 3x 5 min u puferu 4T i 5 min u otopini PBS, preparati su dehidrirani prolaskom kroz seriju hladnih etanola (80% → 90% → 100%, 5 min svaki), a zatim ostavljeni da se potpuno osuše na zraku. Bojanje preparata provedeno je 20-minutnom inkubacijom u otopini 2x SSC koja je sadržavala DAPI (50 ng DAPI/ml 2x SSC), nakon čega je višak boje uklonjen prvo ispiranjem u komorici s protočnom vodom 1 min, a zatim nekoliko minuta u destiliranoj vodi. Nakon toga preparati su ostavljeni na sobnoj temperaturi da se u potpunosti osuše. Prije

mikroskopiranja preparati su uklopljeni u tzv. *antifade* reagens (Mowiol 4-88 antifade medium).

2.2.9.4. Mikroskopska analiza kromosomskih preparata

Preparati uklopljeni u *antifade* reagens analizirani su korištenjem konfokalnog laserskog pretražnog mikroskopa Leica TCS SP8 X. Dobivene slike procesirane su i analizirane programima ImageJ (Abramoff i sur. 2004) i Adobe Photoshop CS5.

2.2.10. Kapljična analiza

Udio satDNA TCst154, TCst266 i TCst309 u ukupnoj genomskoj DNA vrste *T. castaneum* određen je kapljičnom analizom (engl. *dot blot*). Uzorak genomske DNA vrste *T. castaneum* te umnoženi odsječci satDNA TCst154, TCst266 i TCst309 razrijeđeni su u određenim koncentracijama u puferu 1x TE u ukupnom volumenu 100 µl. Nakon denaturacije od 5 min na 100 °C u termobloku, uzorci su ohlađeni na ledu 5 min. Nakon dodavanja 80 µl 20x SSC pufera svaki uzorak nanesen je na najlonsku membranu uz pomoć uređaja za kapljičnu analizu (engl. *dot blotter*). Membrana je zatim u potpunosti osušena na sobnoj temperaturi te inkubirana u suhom sterilizatoru tijekom 20 min na 120 °C čime su fragmenti DNA dodatno fiksirani na najlonsku membranu. Uslijedio je proces denaturacije membrane inkubacijom u denaturirajućem puferu 10 min, a potom je membrana neutralizirana ispiranjem 3x 5 min u neutralizacijskom puferu. Daljnji postupci prehibridizacije, hibridizacije, ispiranja i detekcije jednaki su kao i u postupku hibridizacijske analize prema Southernu (2.2.8). Hibridizacijski signali denzitometrirani su programom ImageJ 1.37v, dostupnom na internetskoj stranici National Institutes of Health (<https://imagej.nih.gov/ij/>). Broj kopija satDNA TCst154, TCst266 i TCst309 u genomu vrste *T. castaneum* određen je prema baždarnoj krivulji dobivenoj na temelju serije razrjeđenja različitih količina satDNA TCst154, TCst266 i TCst309 uz pomoć programa Microsoft Office Excel 2013.

2.2.11. Bioinformatička analiza

2.2.11.1. Analiza genoma vrste *T. castaneum*

Genom vrste *T. castaneum* inicijalno je sekvenciran u servisu za sekvenciranje Admera Health (South Plainfield, SAD) metodom sekvenciranja nove generacije uz korištenje platforme Illumina 1.9 NextSeq 500/550. Genomska DNA je fragmentirana na odsječke

duljine 500 pb, a genomski biblioteka pripremljena u servisu Admera Health upotrebom kompleta Kapa Hyper Prep Kit (Kapa Biosystems). Metodom sekvenciranja sparenih krajeva (engl. *pair-end*) očitano je 100 pb s jednog i drugog kraja odsječka. Sekvenciranje je rezultiralo s 2x 4 061 259 očitavanja (engl. *reads*) duljine od 35-101 pb, čime je dobivena prosječna pokrivenost genoma (engl. *genome coverage*) od 4x. Očitavanja, dobivena u elektronskom formatu FASTAQ, inicijalno su analizirana i procesirana alatima računalne platforme RepeatExplorer (<https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>). Očitavanja su potkraćena na uniformnu duljinu od 89 pb, a filtrirana s obzirom na kvalitetu pod sljedećim parametrima: ocjena kvalitete ≥ 10 , preko 95% baza, nisu dopuštene nepoznate baze (engl. „*quality score ≥ 10 over 95% of bases, no Ns allowed*“).

Za detekciju i analizu uzastopno ponovljenih sekvenca programom TAREAN (Novák i sur. 2017) iz procesiranih očitavanja upotrebom alata odabiranje sekvenci (engl. „*Sequence sampling*“) nasumičnim odabirom uz parametar: generator slučajnih brojeva =10 (engl. *random number generator seed =10*) izdvojeno je 400 000 očitavanja. TAREAN analiza provedena je pod sljedećim uvjetima: prag veličine klastera 0,001, spajanje klastera, bez filtriranja obilnih satelitnih ponavljanja, duga i detaljna analiza (engl. *cluster size threshold 0,001, perform cluster merging, no automatic filtering of abundant satellite repeats, long and slow queue*).

2.2.11.2. Analiza kloniranih odsječaka DNA

Program BioEdit v.7.2.5 (Hall 1999) korišten je za uređivanje i osnovne manipulacije nukleotidnih sljedova, dobivenih kapilarnim sekvenciranjem u korisničkom centru Macrogen. Sljedovima su određeni krajevi te odstranjena vezna mjesta za početnice. Budući da su klonirani odsječci sadržavali satelitne monomere ili dimere uz dodatak omeđujućih krnjih satelitnih fragmenata submonomerne duljine, uz pomoć računalnog programa Geneious 11.1.4. (Biomatters Ltd) iz nukleotidnih sljedova kloniranih odsječaka su *in silico* ekstrahirani sljedovi cjelovitih monomera korišteni u daljnjim analizama.

Računalni program Geneious 11.1.4. također je korišten za anotaciju monomera TCst154, TCst266 i TCst309 unutar postojećeg sastavljenog genoma vrste *T. castaneum*, genomski sklop (engl. *genome assembly*) Tcas5.2 ([https://metazoa.ensembl.org/Tribolium castaneum/Info/Index](https://metazoa.ensembl.org/Tribolium%20castaneum/Info/Index)). Anotirani monomeri morali su zadovoljiti kriterij minimalne sličnosti

od 80% s konsenzus sekvencom pojedine satDNA. Anotirani monomeri su *in silico* ekstrahirani iz genoma te korišteni u daljnjim usporedbama s kloniranim odsječcima.

Za filogenetske analize korišteno je više različitih programa. Višestruko sravnjivanje nukleotidnih sljedova napravljeno je programom ClustalX v. 2.21. (Larkin i sur. 2007). Program MEGA6 v. 06 (Tamura i sur. 2013) korišten je u analizi najveće vjerojatnosti (engl. *Maximum Likelihood*, ML), analizi udruživanja susjeda (engl. *Neighbor Joining*, NJ) te analizi najveće parsimonije (engl. *Maximum Parsimony*, MP). U analizi ML korištena je heuristička metoda odabira stabla zamjene najbližih susjeda (engl. *Nearest-Neighbor-Interchange*, NNI). U analizi NJ korištene su udvojene udaljenosti (engl. *pairwise distance*, *p-distance*), dok je u analizi MP za pretraživanje stabala korištena metoda zamjene bliskih susjeda (engl. *Close-Neighbor-Interchange*, CNI). Statistička podrška za topologiju filogenetskih stabala određena je na temelju 500 replikacija (*bootstrap* ponavljanja). Filogenetska stabla grafički su uređena u programu FigTree v. 1.4.3. (<http://tree.bio.ed.ac.uk/software/figtree/>).

Nukleotidni sljedovi kloniranih odsječaka satelitnih DNA TCst154, TCst266 i TCst309 uspoređeni su s nukleotidnim sljedovima pohranjenima u genskoj bazi NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). U tu svrhu korišten je *on-line* alat BLAST (engl. *Basic Local Alignment Search Tool*) (Altschul i sur. 1990) za pretraživanje po sličnosti prema algoritmu *blastn*, koji omogućava pronalaženje nukleotidnih sljedova i s manjim stupnjem sličnosti.

3. REZULTATI

3.1. Identifikacija satelitnih DNA u genomu vrste *T. castaneum* sekvenciranjem metodom NGS

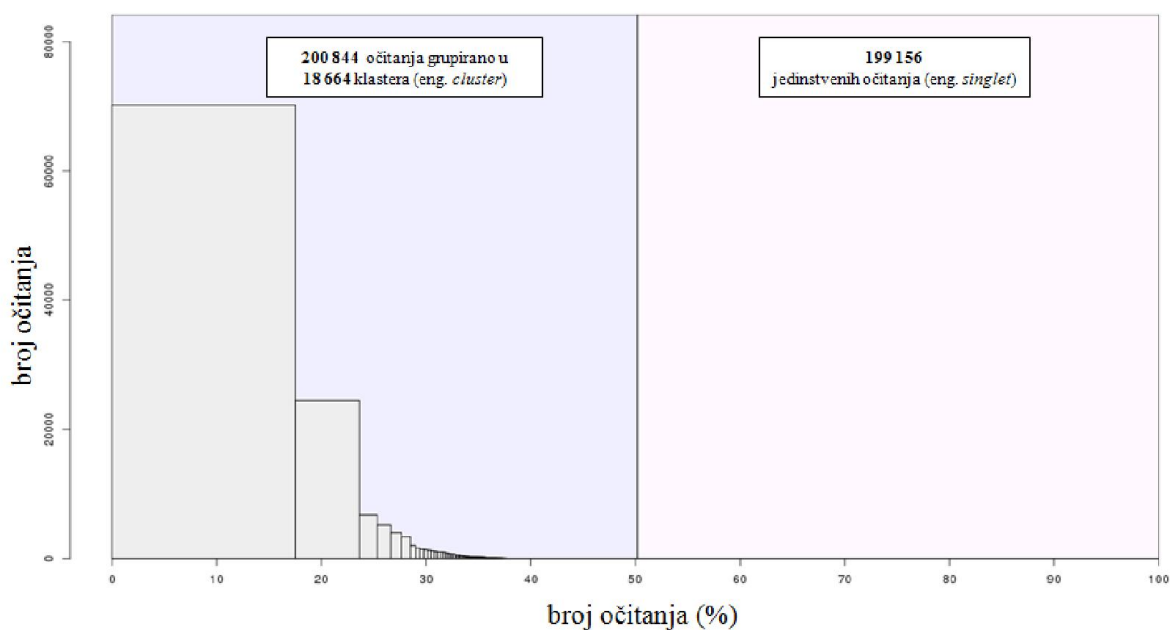
U svrhu detekcije dosad neotkrivenih satDNA u genomu vrste *T. castaneum*, 400 000 nasumično odabranih NGS očitavanja pojedinačne duljine 89 pb analizirano je bioinformatičkim programom TAREAN. S obzirom na veličinu genoma vrste *T. castaneum* od 204 Mb (Brown i sur. 1990), spomenuti broj očitavanja ukupnom duljinom prosječno pokriva genom približno 0,17x, što je unutar preporučenog raspona genomske pokrivenosti ulaznih podataka za TAREAN analizu (optimalan raspon kreće se od 0,01-0,50x).

Prema TAREAN analizi, od ukupno 400 000 nasumično odabranih očitavanja, njih 200 844 (50,02%) grupiralo se u 18 664 klastera (engl. *cluster*), a 199 156 (49,98%) ostalo je nesvrstano te se smatra da predstavljaju jedinstvena očitavanja (engl. *singlet*) (slika 4).

Nadalje, unutar 18 664 klastera (CL1-CL18664) TAREAN analiza definirala je 20 potencijalnih satDNA, od kojih je 5 svrstala u satDNA visoke vjerojatnosti (engl. *high confidence*), dok je preostalih 15 definirala kao satDNA sekvence niske vjerojatnosti (engl. *low confidence*) (prilog 1, slika P1).

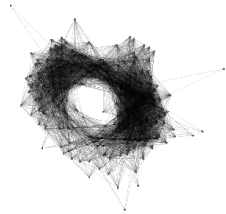
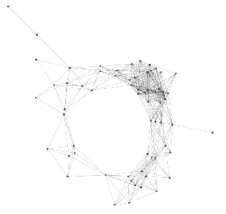
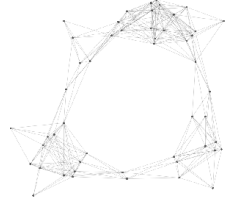
Konsenzus sekvence potencijalnih satDNA uspoređene su s podacima pohranjenim u genskoj bazi NCBI GenBank kako bi se utvrdilo koje od 20 potencijalnih satDNA pripadaju već poznatim, prethodno opisanim satDNA vrste *T. castaneum*. BLAST analiza pokazala je da konsenzus sekvenca klastera CL1, koji podržava 18% ukupnih NGS očitavanja, svojim nukleotidnim slijedom duljine 363 pb odgovara sekvenci glavnog satelita u vrsti *T. castaneum* opisanim od strane Ugarković i sur. (1996). Pet klastera (CL2, CL3, CL6, CL18 i CL20) pronalazi sličnost s prethodno opisanim satelitnim sekvencama vrste *T. castaneum*, koji se prvenstveno smještaju u eukromatinskim kromosomskim regijama (Pavlek i sur. 2015). Šest klastera (CL10, CL16, CL52, CL127, CL158, CL160) pokazuje sličnost s različitim nukleotidnim sljedovima porijeklom iz genoma vrste *T. castaneum*, kao što su pretpostavljeni proteini, nekodirajuće molekule RNA te MEDEA (engl. *Maternal Effect Dominant Embryonic Arrest*) elementi. Konsenzus sekvence klastera CL33, CL36, CL61, CL104, CL145, CL195, CL199 i CL248 nisu pokazale sličnost sa sekvencama vrste *T. castaneum* pohranjenim u genskoj bazi NCBI GenBank.

Isključivši iz daljnjeg razmatranja satDNA koje su opisane u prethodnim istraživanjima (Ugarković i sur. 1996, Pavlek i sur. 2015), za eksperimentalnu provjeru u ovom radu odabrani su klasteri čiji je TAREAN grafički prikaz odgovarao kružnoj strukturi karakterističnoj za uzastopno ponovljene sekvence te koji su bili podržani s minimalno 50 očitavanja. Pritom su odabrani klasteri za koje je vjerojatnost da se radi o satelitim sekvencama iznosila $>0,05$. Navedene kriterije zadovoljili su klasteri CL52, CL104 i CL127, u daljnjem tekstu imenovani kao satDNA TCst154, TCst266 i TCst309 (tablica 5, prilog 1). Naziv satelita izveden je akronim iz opisa „TRI^BOLIUM CASTANEUM satelit“, a pridružena numeracija odgovara duljini konsenzus nukleotidnog slijeda dotične satDNA.



Slika 4. Histogram prikazuje rezultate TAREAN analize 400 000 očitavanja dobivenih metodom sekvenciranja nove generacije genoma vrste *Tribolium castaneum*.

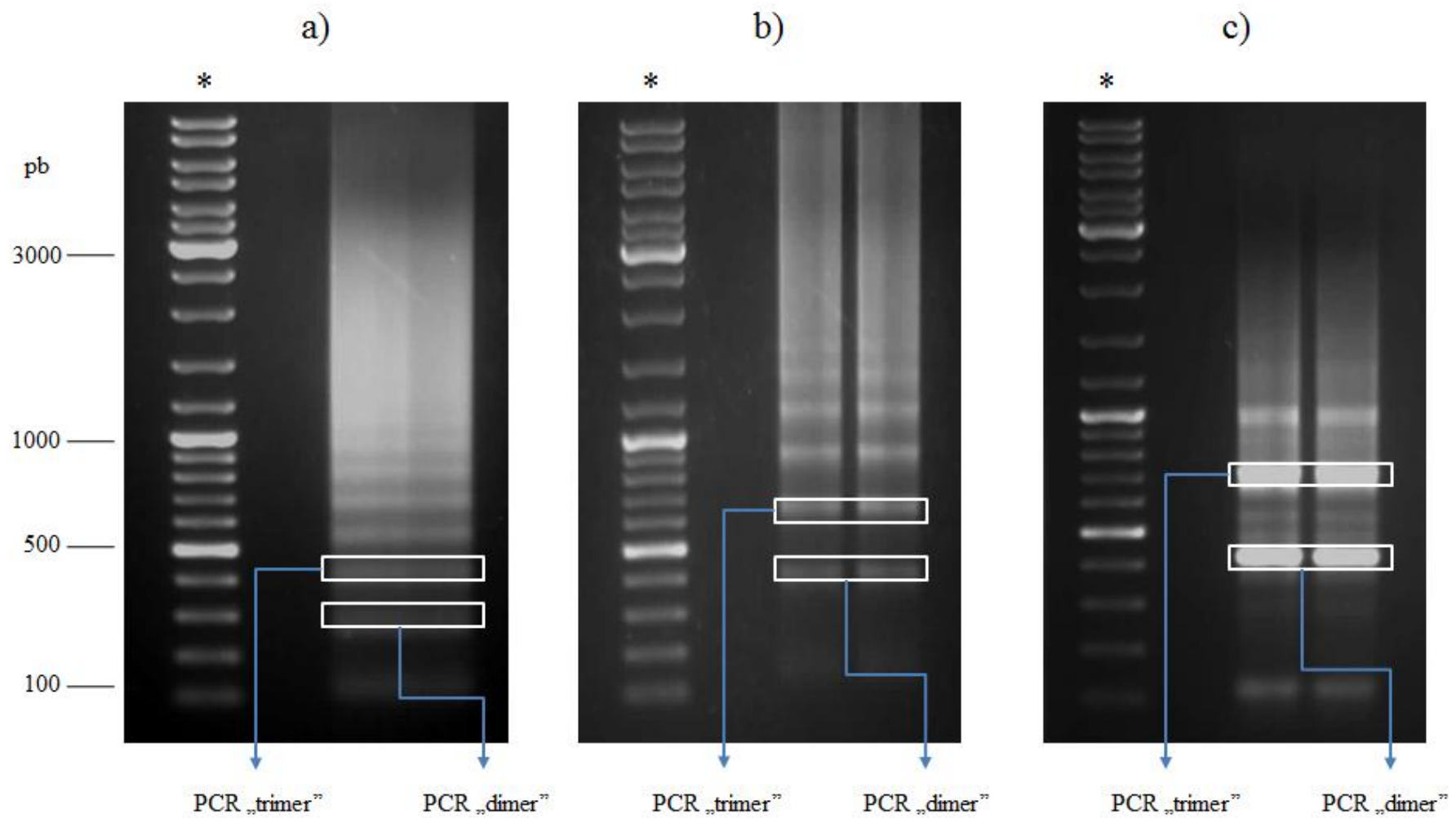
Tablica 5. Sumarni prikaz klastera dobivenih TAREAN analizom NGS podataka genoma vrste *Tribolium castaneum* odabranih za daljnju eksperimentalnu provjeru.

ime satelita	nukleotidni slijed satelitnog konsenzusa	duljina monomera	grafički prikaz klastera	TAREAN klaster
TCst154	AATGCTCTTTTCGACCTGCCGTAACAAGAAAAGTGGCACATTGT CGATAACACATAGAAAAAATAAAAAACCATCAAAAATTATA AACAAACAACCTTTCCAACCCCTCCAGGGCTCGATTTTTT CAAACCTTTTTTATAATGTAGAGTATTA	154 pb		CL52
TCst266	GCAATCAAATTTGACATATTCAAATAAAGGGGGTAGTGGGG AATTA AAAATGTCAAAATTCAAAAAATTA AAAATTAACCAA ATTAATATATGTTATAGTAGGGAAGCTCAGGATCAATTTAAA ATTATTTTCAGAAATTTTTCTCATTTCCTACCCCAAACTTAT AGGGGTAAATTTGGCCAAAAATTTGAAAACTTTTATTACAT TCGGATTGAAGTCGTTATTTGTACATGGGTACCCTCTATGAT CCTGATTATGCTCCA	266 pb		CL104
TCst309	TGTGGTTCAGTTCTAAATATTTTACAATAATTTAACTTTTTTA GGCTACATCTTCCAAGAAGGAAAAACTAAAGTACCACCTG CTGAACCAAGCTGAAGTTACAAACGGAACAGATGAAAAA ATTACAAAACGCAATAAAGCAGAAGTGAGTATATGTTGACA CCAAGTCTTTTTGTTTTATGAAGTGATTTTAGCCTGTAAGC GTTGAAAAATGTCGAGGAACTGTTGCACCTAAAGCATGGGA AAACAAGGTATTTATATTTTACTTGAAGTGAATTTTGTGAT AATTTCCACAAAAAAG	309 pb		CL127

3.2. Kloniranje i karakterizacija odsječaka satelitnih DNA TCst154, TCst266 i TCst309

Konsenzus sekvence TCst154, TCst266 i TCst309 dobivene TAREAN analizom predstavljaju uprosječene nukleotidne sljedove. Da bi se dobio uvid u nukleotidne sljedove pojedinih, izvornih monomera u genomu, potrebno je klonirati i sekvencirati genomske odsječke koji sadrže satDNA sljedove od interesa. S obzirom na to da je bioinformatička analiza procijenila udio sekvenca TCst154, TCst266 i TCst309 u genomu vrlo niskim ($\leq 0,05\%$), odustalo se od tradicionalno upotrebljavane strategije kloniranja satelitnih odsječaka dobivenih restriksijskom razgradnjom genomske DNA. Umjesto toga, odabran je pristup umnažanja satelitnih odsječaka metodom PCR. Na temelju NGS konsenzus sekvence, kreirane su i sintetizirane početnice specifične za potencijalne satDNA TCst154, TCst266 i TCst309 (slika 3) nakon čega su metodom PCR iz izolirane genomske DNA vrste *T. castaneum* umnoženi odsječci spomenutih sekvenca.

Nakon elektroforetskog razdvajanja umnoženih odsječaka DNA u agaroznom gelu uočeno je da su sve tri reakcije rezultirale ljestvičastim elektroforetskim obrascem, a elektroforetski profili pokazali su očekivane veličine elektroforetskih vrpca (engl. *band*) za svaku od satelitnih DNA TCst154 (slika 5a), TCst266 (slika 5b) i TCst309 (slika 5c), kao što je navedeno u tablici 1 (poglavlje 2.2.3.1.). Na temelju elektroforetskog razdvajanja umnoženih odsječaka zaključeno je kako su sekvence TCst154, TCst266 i TCst309 uistinu prisutne u genomu vrste *T. castaneum* u obliku uzastopnih ponavljanja što je vidljivo iz ljestvičastog obrasca pojavljivanja vrpca koje odgovaraju monomeru određene satDNA i njegovim višekratnicima.



Slika 5. Prikaz elektroforetskog razdvajanja odsječaka satDNA TCst154 (a), TCst266 (b) i TCst309 (c) umnoženih metodom PCR iz genomske DNA vrste *Tribolium castaneum*. Znakom * označen je položaj elektroforetskog biljega O'GeneRuler DNA Ladder Mix (Fermentas). Bijelim pravokutnicima označeni su izolirani i klonirani fragmenti.

Nakon umnažanja PCR-om i elektroforetskog razdvajanja, odsječci DNA koji su duljinom odgovarali takozvanim PCR „dimerima“ i PCR „trimerima“ (slika 5) izolirani su iz agaroznog gela te klonirani u plazmidni vektor pGEM-T Easy. Putem metode PCR-a na bakterijskim kolonijama (engl. *colony PCR*) napravljen je odabir kolonija koje su sadržavale plazmide s kloniranim odsječcima očekivane duljine. Iz pozitivnih kolonija izolirana je plazmidna DNA te su sekvenciranjem određeni nukleotidni sljedovi kloniranih odsječaka. Pokazalo se, prema očekivanju, da PCR „dimeri“ sadrže satelitne monomere u integralnoj duljini uz dodatak omeđujućih krnjih satelitnih fragmenata submonomerne duljine, dok PCR „trimeri“ u svom sastavu posjeduju satelitne dimere također omeđene krnjim satelitnim fragmentima.

Za satDNA TCst154 ukupno su sekvencirana 22 klona iz kojih je za potrebe daljnjih bioinformatičkih analiza ekstrahirano *in silico* 31 monomer. Za satDNA TCst266 sekvencirano je 29 klonova iz kojih su ekstrahirana 34 monomera, a za TCst309 sekvencirano je 29 klonova iz kojih je ekstrahirano 46 monomera (tablica 6). Duljine kloniranih odsječaka kao i duljine ekstrahiranih monomera navedene su u prilogima 2, 3 i 4.

Tablica 6. Sumarni prikaz broja sekvenciranih monomera ekstrahiranih iz satDNA TCst154, TCst266 i TCst309 PCR „dimeri“ i „trimeri“ vrste *Tribolium castaneum*.

	TCst154	TCst266	TCst309
PCR „dimeri“	13	24	12
PCR „trimeri“	9	5	17
ukupan broj ekstrahiranih monomera	31	34	46

Bioinformatičkom analizom ukupno 31 *in silico* ekstrahiranog monomera za satDNA TCst154 zabilježeno je kako se duljine monomera kreću u rasponu od 147 do 156 pb, a AT sastav pojedinih monomera varira od 65,6 % do 70,3%. Varijabilnost sekvenca zabilježena je

u rasponu od 0-12%, dok je prosječna *p*-udaljenost za 31 monomer iznosila 6%. Na temelju sravnjenja 31 monomera za satDNA TCst154 izrađeno je filogenetsko stablo metodom najveće vjerojatnosti te je utvrđeno kako se monomeri ekstrahirani iz PCR „trimer“ ne grupiraju skupa te sukladno tomu ne pokazuju složenu organizaciju tj. ne dolazi do formiranja jedinica organizacije višeg reda. Pojedinačni monomeri u odnosu na konsenzus sekvencu satDNA TCst154 pokazuju varijabilnost u rasponu 0,6%-7%.

Duljina 34 ekstrahirana monomera za satDNA TCst266 kreće se u rasponu od 265 do 288 pb, a njihov AT sastav varira od 68,4% do 72,2%. Varijabilnost sekvencu zabilježena je u rasponu od 0-13,8% , dok je prosječna *p*-udaljenost sva 34 monomera iznosila 5%. Nakon sravnjenja 34 *in silico* ekstrahirana monomera za satDNA TCst266 i izrade filogenetskog stabla nije zabilježeno grupiranje monomera ekstrahiranih iz PCR „trimer“ čime se može zaključiti kako monomeri TCst266 satDNA također ne formiraju jedinice organizacije višeg reda. Konsenzus sekvencu satDNA TCst266 u odnosu na nukleotidne sljedove pojedinačnih monomera pokazuje razlike u rasponu od 1%-10%.

Ukupno 46 *in silico* ekstrahiranih monomera za satDNA TCst309 duljinom varira od 308 do 312 pb. AT sastav kreće su rasponu od 67,4% do 68,7%. Varijabilnost sekvencu zabilježena je u rasponu od 0-2,2% , dok je prosječna *p*-udaljenost između 46 monomera iznosila 0,6%. Jednako kao i za TCst154 i TCst266, sravnjenje monomera za TCst309 i posljedična izrada filogenetskog stabla nije zabilježila grupiranje monomera ekstrahiranih iz PCR „trimer“ te samim time niti formiranje jedinica organizacije višeg reda. Varijabilnost pojedinačnih monomera spram konsenzus sekvencu satDNA TCst309 kreće se od 0%-1,6%.

Nakon bioinformatičke analize varijabilnosti između konsenzus sekvencu za TCst154, TCst266 i TCst309 dobivenih TAREAN analizom i nukleotidnih sljedova pojedinih, izvornih monomera spomenutih satDNA u genomu može se zaključiti kako niti jedan monomer za satDNA TCst154 i TCst266 nije u potpunosti identičan konsenzus slijedu pojedine satDNA, dok za razliku od njih tri monomera za satDNA TCst309 (TCst309c112a, 29a i 34) imaju identičan nukleotidni slijed konsenzus sekvencu za satDNA TCst309.

Od ukupno 204 Mb, koliko se procjenjuje da iznosi genom vrste *T. castaneum*, u trenutno dostupnom genomskom sklopu Tcas5.2 nalazi se posloženo i anotirano cca 160 Mb. Preostalih neposloženih ~20% genoma (cca 40 Mb) djelomično je dostupno kroz 305 neposloženih skafolda (engl. *unplaced scaffolds*) te 1848 neposloženih jedinstvenih sekvencu (engl. *unplaced singletons*). Želeći vidjeti u kojem dijelu genoma se preferencijalno nalaze

satDNA TCst154, TCst266 i TCst309, njihovim konsenzus sekvencama pretražene su baze sastavljenog genoma, neposloženih skafolda i neposloženih jedinstvenih sekvenca. Broj pronađenih monomera, koji su pokazivali sličnost >80% sa spomenutim satelitima, naveden je u tablici 7.

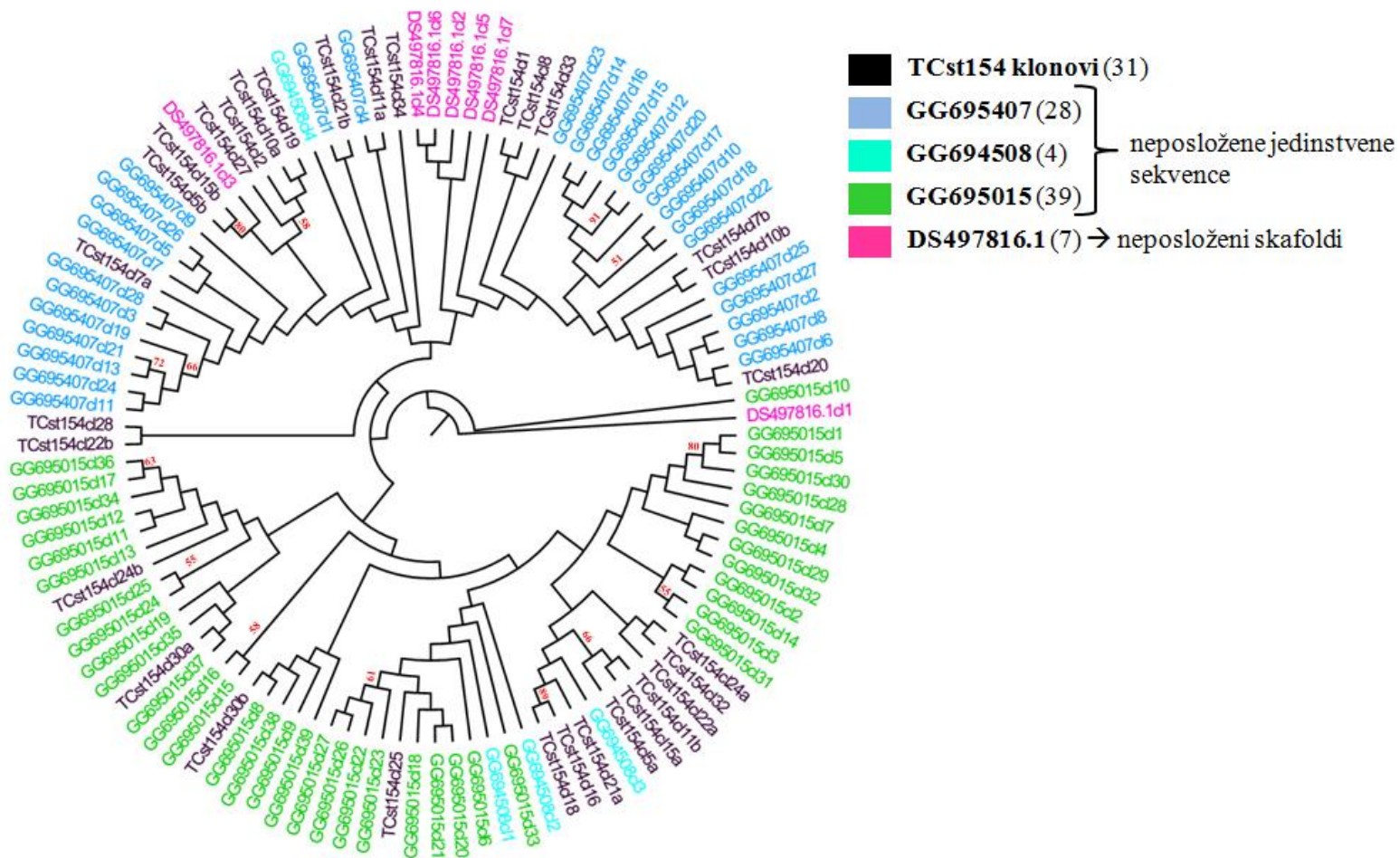
Tablica 7. Broj satelitnih monomera TCst154, TCst266 i TCst309 ekstrahirani iz postojećeg sastavljenog genoma vrste *Tribolium castaneum* (genomski sklop Tcas5.2), neposloženih skafolda i neposloženih jedinstvenih sekvenca.

satDNA	sastavljeni genom	neposloženi skafoldi	neposložene jedinstvene sekvence
TCst154	-	7	71
TCst266	-	19	31
TCst309	36	1	-

Značajno je da satDNA TCst154 i TCst266 nisu pronađene u sastavljenom dijelu genoma, već samo u neposloženim skafoldima i neposloženim jedinstvenim sekvencama. Za razliku od spomenuta dva satelita, satDNA TCst309 zabilježena je u sastavljenom genomu, a u neposloženom skafoldu zabilježena je tek jedna monomerna kopija. Dodatnu zanimljivost predstavlja činjenica da je satDNA TCst309, uz izuzetak jedne kopije na kromosomu 3, gotovo izolirano pronađena na kromosomu 7, distribuirajući se u dva poteza duljine 3,4 kb (11 monomera) i 7,4 kb (24 monomera), međusobno udaljena 65 kb.

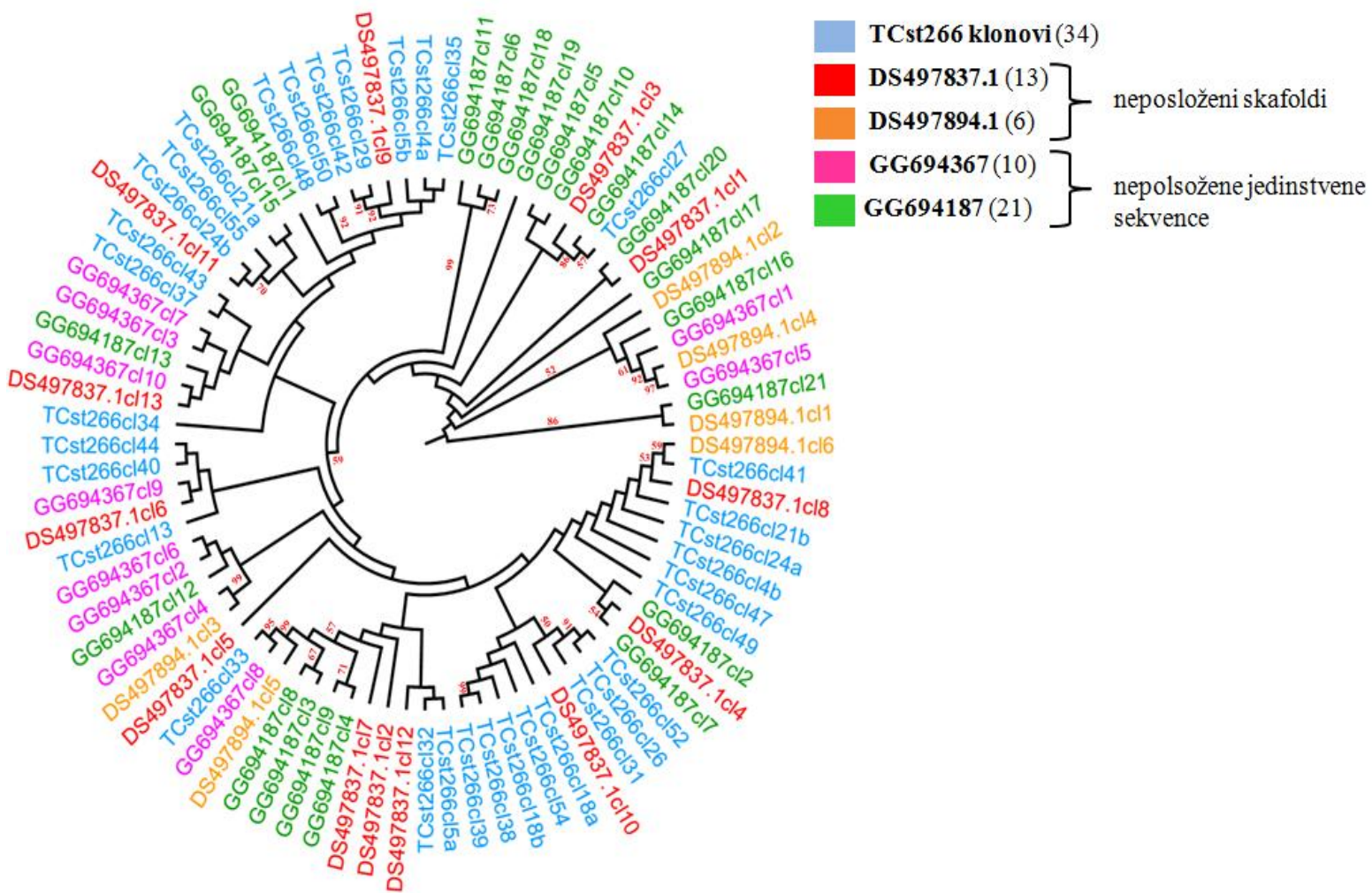
Kako bi se dobio uvid u odnos satelitnih monomera smještenih u različitim dijelovima genoma te kako bi ih se usporedilo s monomerima porijeklom iz kloniranih PCR odsječaka, pomoću programa Geneious ekstrahirani su monomerni nukleotidni sljedovi detektirani u sastavljenom genomu, neposloženim skafoldima i neposloženim jedinstvenim sekvencama. Za svaki od satelita napravljeno je sravnjenje svih ekstrahiranih monomernih kopija: 109 monomera za satDNA TCst154 (prilog 5, slika P5), 84 monomera za satDNA TCst266 (prilog 6, slika P6) te naposljetku 83 monomera za satDNA TCst309 (prilog 7, slika P7). Na temelju sravnjenja izrađena su filogenetska stabla metodama najveće vjerojatnosti, udruživanja susjeda i najveće parsimonije. Sva tri pristupa rezultirala su prikazom stabala koja su podržavala istu osnovnu topologiju pa su zbog jednostavnosti prikaza u rezultatima predložena samo stabla dobivena metodom ML (slike 6, 7, 8).

Bioinformatička analiza pokazala je kako sekvence svih 109 ekstrahiranih monomernih kopija za TCst154 međusobno variraju u rasponu od 0-18%. Filogenetsko stablo temeljeno na sravnjenju svih 109 ekstrahiranih monomernih kopija za satDNA TCst154 nije pokazalo značajnije grupiranje sekvenca s obzirom na njihovo porijeklo (slika 6).



Slika 6. Filogenetski odnosi monomera satDNA TCst154 vrste *Tribolium castaneum* dobiveni metodom najveće vjerojatnosti prema Tamura-3-parametarskom modelu uz dodatak gamma distribucije (T92+G). Brojevi na granama predstavljaju postotak stabala koja su podržala prikazanu topologiju u nasumičnoj statističkoj analizi u 500 navrata (prikazane su samo *bootstrap* podrške koje prelaze 50%). Popratnom legendom objašnjeno je porijeklo monomera. Monomeri označeni s TCst154 predstavljaju *in silico* ekstrahirane monomere iz umnoženih, kloniranih i sekvenciranih klonova. Oznaka GG i njoj pridružena numeracija predstavlja monomere pronađene u neposloženim jedinstvenim sekvencama, a oznaka DS predstavlja monomere koji se nalaze u neposloženim skafoldima. Broj analiziranih monomera naznačen je u zagradi.

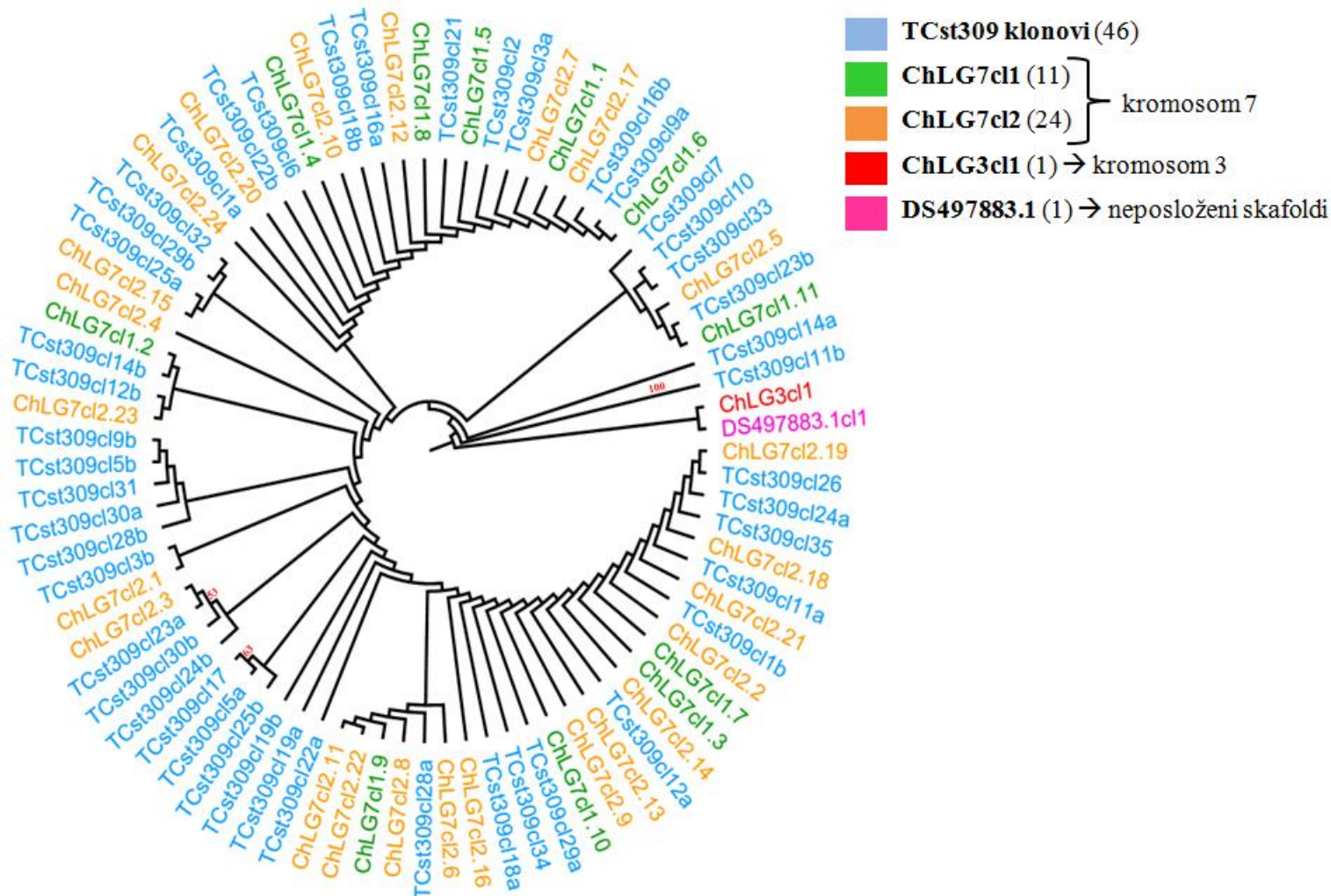
Osamdeset četiri monomera satDNA TCst266 međusobno se u nukleotidnom slijedu razlikuju od 0-38,2%. Nešto širi raspon varijabilnosti uzrokovan je dvjema sekvencama, DS497894.1cl1 i GG694187cl21, koje značajnije otklanjaju od ostatka analiziranih monomera. Izdvajanje dviju spomenutih sekvencu vidljivo je i na filogenetskom stablu (slika 7), na kojem se one izdvajaju u zasebnu grupu podržanu s vjerojatnošću od 86%. Povećana divergencija dvaju monomera može se objasniti njihovim rubnim položajima u tandemsom nizu. Naime, veća varijabilnost rubnih monomernih kopija česta je karakteristika satelitnih nizova. Kao što je vidljivo iz topologije filogenetskog stabla, ostatak analiziranih monomera TCst266 ne pokazuje značajnije grupiranje.



Slika 7. Filogenetski odnosi monomera satDNA TCst266 vrste *Tribolium castaneum* dobiveni metodom najveće vjerojatnosti prema Tamura-3-parametarskom modelu uz dodatak gamma distribucije (T92+G). Brojevi na granama predstavljaju postotak stabala koja su podržala prikazanu topologiju u nasumičnoj statističkoj analizi u 500 navrata (prikazane su samo *bootstrap* podrške koje prelaze 50%). Popratnom legendom objašnjeno je porijeklo monomera. Monomeri označeni s TCst266 predstavljaju *in silico* ekstrahirane monomere iz

umnoženih, kloniranih i sekvenciranih klonova. Oznaka GG i njoj pridružena numeracija predstavlja monomere pronađene u neposloženim jedinstvenim sekvencama, a oznaka DS predstavlja monomere koji se nalaze u neposloženim skafoldima. Broj analiziranih monomera naznačen je u zagradi.

Osamdeset tri monomera satDNA TCst309 međusobno pokazuju varijabilnost od 0-6%, a najveće otklanje pokazuje monomer pronađen izolirano u sastavljenom genomu na kromosomu 3 te zaseban monomer detektiran unutar neposloženog skafolda. Nešto veća varijabilnost ovih dviju zasebnih kopija, koje međusobno dijele čak 98,7% sličnosti, nije neobična s obzirom da se radi o monomerima koji su u genomu nađeni kao „samostojeće“ kopije van tandemskog ustroja. No, budući da odvajanje ovih dvaju monomera u zasebnu granu nije podržano s *bootstrap* vrijednošću većom od 50% (slika 8), moguće je kako spomenuto odvajanje nije značajno. Također, većina grana unutar stabla nije podržana s *bootstrap* vrijednostima većim od 50%, što se može objasniti malom divergencijom monomernih kopija TCst309.



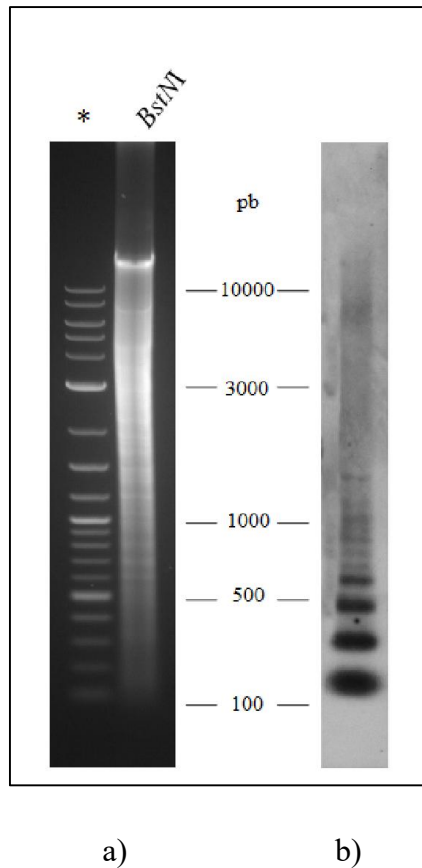
Slika 8. Filogenetski odnosi monomera satDNA TCst309 vrste *Tribolium castaneum* dobiveni metodom najveće vjerojatnosti prema Tamura-3-parametarskom modelu (T92). Brojevi na granama predstavljaju postotak stabala koja su podržala prikazanu topologiju u nasumičnoj statističkoj analizi u 500 navrata (prikazane su samo *bootstrap* podrške koje prelaze 50%). Popratnom legendom objašnjeno je porijeklo monomera. Monomeri označeni s TCst309 predstavljaju *in silico* ekstrahirane monomere iz umnoženih, kloniranih i sekvenciranih klonova. Oznake ChLG3 i ChLG7 i njima pridružene numeracije predstavljaju monomere pronađene u sastavljenom dijelu genoma, a oznaka DS predstavlja monomere koji se nalaze u neposložnim skafoldima. Broj analiziranih monomera naznačen je u zagradi.

Usporedbom varijabilnosti monomernih sekvenca koje su *in silico* ekstrahirane iz pozitivnih klonova i svih monomernih sekvenca pojedinih satDNA ekstrahiranih iz trenutno dostupnog genomskog sklopa Tcas5.2 može se zaključiti kako prilikom PCR amplifikacije satDNA TCst154, TCst266 i TCst309 nije bilo pristranosti s obzirom na njihov smještaj u genomu te da je uzorak PCR umnoženih odsječaka uistinu bio reprezentativan.

3.3. Genomska organizacija satelitnih DNA TCst154, TCst266 i TCst309

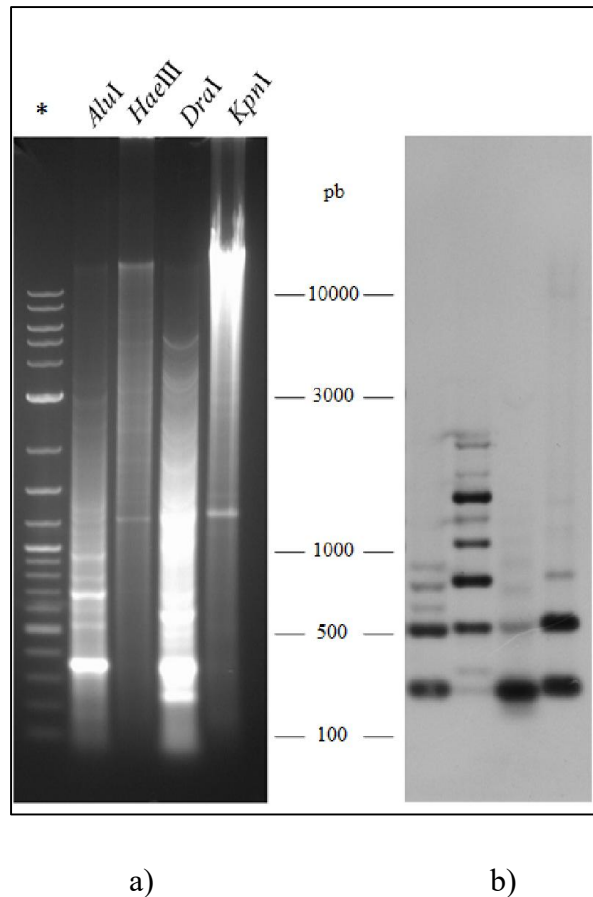
Kako bi se eksperimentalno potvrdila uzastopno ponavljajuća organizacija novootkrivenih satDNA TCst154, TCst266 i TCst309 u genomu vrste *T. castaneum*, genomska DNA razgrađena je restrikcijskim endonukleazama te potom analizirana metodom prema Southernu. Za restrikcijsku razgradnju su odabrani takozvani „konsenzus enzimi“, odnosno restrikcijske endonukleaze koje, prema bioinformatičkoj analizi, u konsenzus nukleotidnim sljedovima pronalaze samo jedno ciljno mjesto za svoje djelovanje.

U slučaju satDNA TCst154 od komercijalno dostupnih restrikcijskih endonukleaza taj kriterij zadovoljila je samo endonukleaza *Bst*NI. Genomska DNA razgrađena enzimom *Bst*NI (slika 9a) te potom hibridizirana sondom specifičnom za satDNA TCst154 (slika 9b) uistinu je pokazala ljestvičast profil hibridizacijskih signala, zasnovan na monomeru duljine 154 pb te njegovim višekratnicima. Naime, uslijed pojave točkastih mutacija uzastopni monomeri mogu se neznatno razlikovati u nukleotidnoj sekvenci, čime može biti promijenjeno mjesto prepoznavanja restrikcijskog enzima. Izostanak rada enzima na pojedinom mjestu dovodi do stvaranja dimernih odsječaka, odnosno daljnjih multimera osnovne jedinice ponavljanja. Eksperimentom hibridizacije sa specifičnom sondom dokazana je tandemska organizacija sekvence TCst154 u genomu vrste *T. castaneum*.



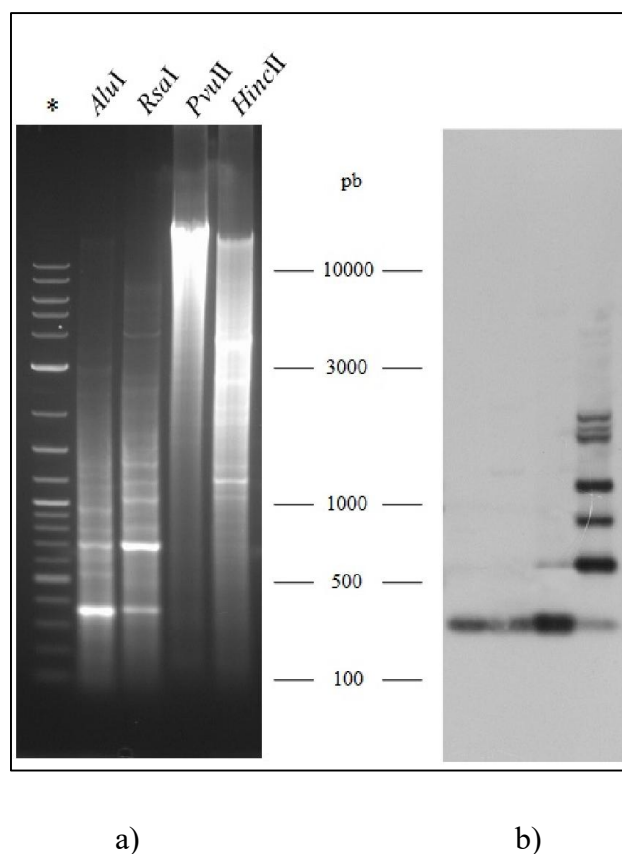
Slika 9. Elektroforetsko razdvajanje genomske DNA vrste *Tribolium castaneum* razgrađene restriktivnim enzimom *Bst*NI (a). Hibridizacija razgrađene genomske DNA sa sondom specifičnom za satelitnu DNA TCst154 (b). Znakom * označen je položaj elektroforetskog biljega 2-Log DNA Ladder.

Za satDNA TCst266 izbor endonukleaza koje unutar konsenzus sekvence pronalaze jedinstveno mjesto bio je veći pa su za analizu odabrana četiri enzima: *Alu*I, *Hae*III, *Dra*I i *Kpn*I (slika 10a). Kao što je vidljivo na slici 10b, razgradnje sa sva četiri enzima u hibridizaciji sa sondom specifičnom za satDNA TCst266 rezultirale su profilom signala karakterističnim za satelitne sekvence. Iako se sva četiri profila baziraju na monomeru duljine 266 pb i njegovim višekratnicima, vidljive su razlike u intenzitetu pojedinih signala, što je objašnjivo različitim stupnjem očuvanosti pojedinih restriktivskih mjesta između monomera. Pojava međusignala u linijama restriktivskih razgradnja *Alu*I i *Hae*III objašnjive su pak pojavom dodatnih restriktivskih mjesta kod određenog broja monomera (slika 10b).



Slika 10. Elektroforetsko razdvajanje genomske DNA vrste *Tribolium castaneum* razgrađene restriktivnim enzimima *AluI*, *HaeIII*, *DraI* i *KpnI* (a) te potom hibridizirane sa sondom specifičnom za satDNA TCst266 (b). Znakom * označen je položaj elektroforetskog biljega 2-Log DNA Ladder.

Kako bi se dokazala tandemska organizacija sekvence TCst309, genomska DNA razgrađena je enzimima *AluI*, *RsaI*, *PvuII* i *HincII* (slika 11a). Hibridizacija sa specifičnom sondom TCst309 otkrila je ljestvičast profil signala baziran na višekratnicima monomera duljine 309 pb samo u razgradnji s enzimom *HincII* (slika 11b). U reakcijama razgradnja enzimima *AluI* i *RsaI* vidljiv je signal isključivo u području monomera, dok je u razgradnji enzimom *PvuII* uz dominantan monomer prisutan i dimerni signal znatno slabijeg inteziteta. Ljestvica koja se sastoji isključivo od monomernih i dimernih signala ukazuje na nizove sastavljene od visokohomolognih monomera, što odgovara niskoj varijabilnosti satelita TCst309 već prethodno uočenoj analizom monomernih sekvenca. Pojava multimerne ljestvice u razgradnji enzimom *HincII* može se objasniti većom permisivnošću njegovog restriktivnog mjesta (GTY↓RAC).



Slika 11. Elektroforetsko razdvajanje genomske DNA vrste *Tribolium castaneum* razgrađene restriktivnim enzimima *AluI*, *RsaI*, *PvuII* i *HincII* (a) te potom hibridizirane sa sondom specifičnom za satDNA TCst309 (b). Znakom * označen je položaj elektroforetskog biljega 2-Log DNA Ladder.

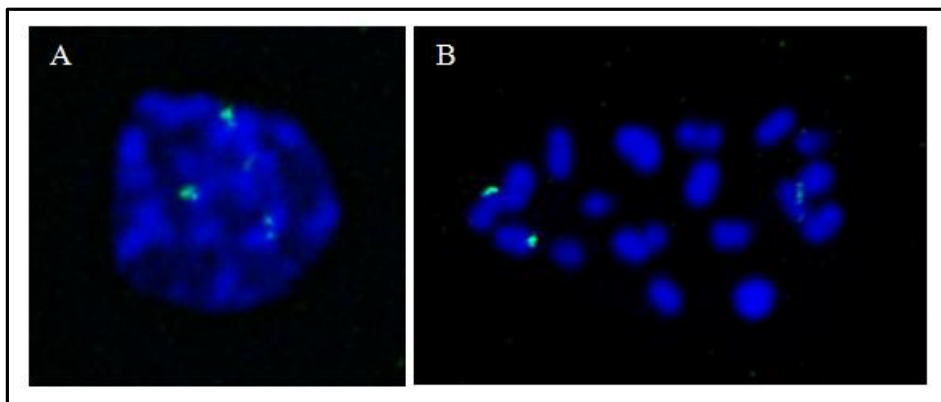
3.4. Kromosomska lokalizacija satelitnih DNA TCst154, TCst266 i TCst309

Položaj sljedova satDNA TCst154, TCst266 i TCst309 na kromosomima vrste *T. castaneum* utvrđen je metodom FISH uz korištenje biotinom obilježenih sonde specifičnih za pojedinu satDNA.

Preparati su pripremljeni iz muških gonada razvojnog stadija kukuljice kako bi se dobio što veći broj stanica u mejozi. Vrsta *T. castaneum* ima diploidan broj kromosoma $2n=20$ te mejotsku formulu mužjaka $9+Xy_p$ (Juan i Petitpierre 1989, 1991). Morfološko razlučivanje kromosoma mikroskopom znatno je otežano činjenicom da ova vrsta ima male kromosome duge 1-2 μm , koji se međutim intenzivno boje fluorescencijskom bojom DAPI, potvrđujući velik udio AT bogatih sekvenca.

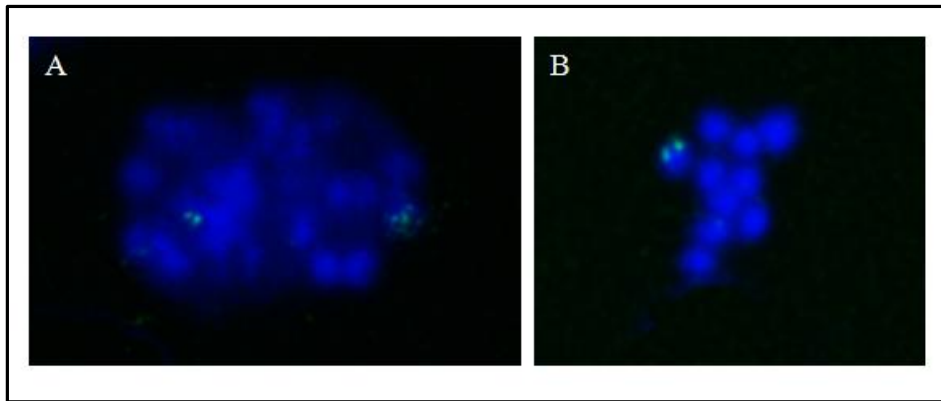
Fluorescencijska hibridizacija *in situ* sa sondom specifičnom za satDNA TCst154 (slika 12) rezultirala je u najvećem broju pregledanih mjesta na kromosomskim preparatima

pojavom dvaju jakih hibridizacijskih signala te dvaju signala manjeg intenziteta. Sukladno tomu, u haploidnom setu kromosoma satDNA TCst154 zabilježena je na dva kromosoma, prilikom čega je na jednom kromosomu prisutna u većem broju kopija, dok je na drugom kromosomu zastupljena s manjim brojem kopija, što je rezultiralo različitim intenzitetom hibridizacijskog signala. S obzirom na nisku zastupljenost satDNA TCst154 u genomu vrste *T. castaneum* nije isključena mogućnost postojanja dodatnih TCst154 regija izgrađenih od manjeg broja TCst154 monomera koje su u metodološkom kontekstu ispod razine osjetljivosti korištene metode. Budući da satDNA TCst154 nakon inicijalne bioinformatičke analize nije pronađena u sastavljenom dijelu genoma, već samo u neposloženim sklafoldima i neposloženim jedinstvenim sekvencama, vrlo je vjerojatno da se satDNA TCst154 pretežito nalazi unutar heterokromatinskih regija kromosoma.



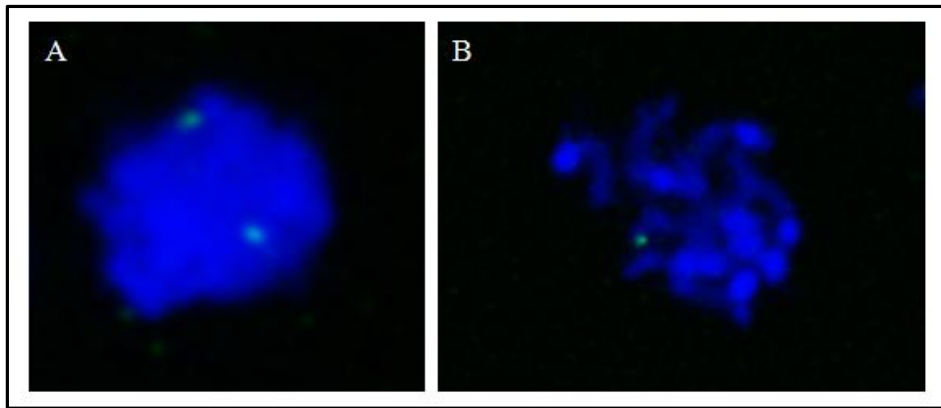
Slika 12. Kromosomi vrste *Tribolium castaneum* u profazi (A) i metafazi (B) nakon bojanja fluorescencijskom bojom DAPI i fluorescencijske hibridizacije *in situ* sa sondom specifičnom za satDNA TCst154.

Rezultati FISH analize sa sondom specifičnom za satDNA TCst266 najčešće su pokazali prisutnost dva signala u interfaznoj jezgri (slika 13a), odnosno postojanje regije satDNA TCst266 samo na jednom kromosomu u haploidnom setu (slika 13b). Međutim, postoji mogućnost postojanja dodatnih TCst266 regija izgrađenih od manjeg broja TCst266 monomera koje su ispod razine osjetljivosti metode FISH. Kao i u slučaju satDNA TCst154, bioinformatička analiza nije zabilježila satDNA TCst266 u sastavljenom dijelu genoma već samo u neposloženim sklafoldima i neposloženim jedinstvenim sekvencama pa se može pretpostaviti da su potezi satDNA TCst266 uglavnom smješteni u heterokromatinskim područjima.



Slika 13. Kromosomi vrste *Tribolium castaneum* u interfazi (A) i metafazi (B) nakon bojanja fluorescencijskom bojom DAPI i fluorescencijske hibridizacije *in situ* sa sondom specifičnom za satDNA TCst266.

Analiza FISH sa sondom specifičnom za satDNA TCst309 rezultirala je pojavom dvaju signala u interfaznoj jezgri, odnosno signalom na jednom kromosomu haploidnog seta (slika 14). Bioinformatičkom analizom detektirani su monomeri satDNA TCst309 u sastavljenom dijelu genoma i to gotovo izolirano na kromosomu 7 pa je za pretpostaviti da hibridizacijski signal dobiven nakon FISH analize predstavlja upravo regiju satDNA TCst309 smještenu na kromosomu 7. Iako je prema bioinformatičkoj analizi satDNA TCst309 na kromosomu 7 prisutna u dva zasebna poteza duljine 3,4 kb (11 monomera) i 7,4 kb (24 monomera), na slikama dobivenim FISH analizom nisu vidljiva dva odvojena hibridizacijska signala. Budući da su dva spomenuta poteza na kromosomu smještena relativno blizu, na međusobnoj udaljenosti od 65 kb, moguće je da zbog kondenziranosti kromatina na obrađenim preparatima hibridizacijski signali nisu vidljivi zasebno, već se stapaju u jednu točku.



Slika 14. Kromosomi vrste *Tribolium castaneum* u interfazi (A) i prometafazi (B) nakon bojanja fluorescencijskom bojom DAPI i fluorescencijske hibridizacije *in situ* sa sondom specifičnom za satDNA TCst309.

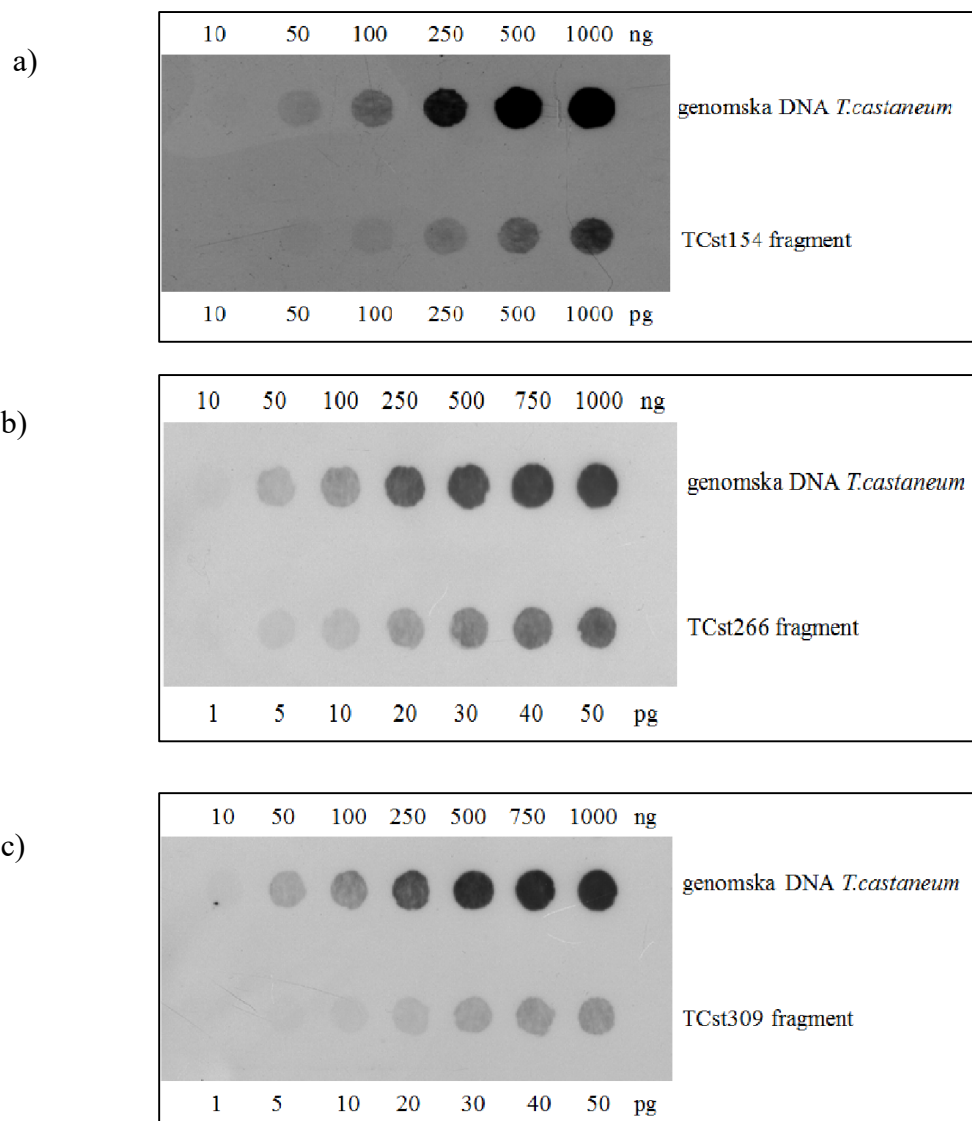
3.5. Udio satelitnih DNA TCst154, TCst266 i TCst309

Udio satelitnih DNA TCst154, TCst266 i TCst309 u genomu vrste *T. castaneum* određen je kapljičnom analizom te usporednim denzitometriranjem hibridizacijskih signala razrjeđenja genomske DNA i satelitnih monomera.

Prema rezultatima kapljične analize zastupljenost slijeda satDNA TCst154 u ukupnom genomu vrste *T. castaneum* procijenjena je na 0,33%, što je gotovo za red veličine više od bioinformatičke TAREAN procjene koja je iznosila 0,054%. S obzirom na veličinu haploidnog genoma vrste *T. castaneum*, koja iznosi 0,21 pg (Juan i Petitpierre 1991), odnosno 204 Mb, a uzimajući u obzir duljinu slijeda od 154 nukleotida, izračunat broj ponavljanja satDNA TCst154 iznosi približno 4370 kopija.

Procijenjena zastupljenost slijeda satDNA TCst266 u ukupnom genomu vrste *T. castaneum* na temelju kapljične analize iznosi 0,02%. Ovaj postotak u skladu je s bioinformatičkom procjenom od 0,016%. S obzirom na duljinu monomera, broj ponavljanja satDNA TCst266 u genomu vrste *T. castaneum* iznosio bi ~150 kopije.

Prema kapljičnoj analizi slijed satDNA TCst309 čini 0,05% ukupnog genoma vrste *T. castaneum* što iznosi nešto više nego što je pokazala prethodna bioinformatička analiza, kojom je udio satelitne DNA TCst309 procijenjen na 0,012%. Uzimajući u obzir duljinu slijeda od 309 pb, broj ponavljanja satDNA TCst309 u genomu vrste *T. castaneum* iznosi ~330 kopija.



Slika 15. Određivanje udjela satDNA TCst154 (a), TCst266 (b) i TCst309 (c) u genomu vrste *Tribolium castaneum* pomoću kapljične hibridizacijske analize. Količine genomske DNA te satelitnih fragmenata naznačene su uz pojedine filmove. Jačina hibridizacijskih signala odgovara udjelu satelitnih fragmenata.

4. RASPRAVA

Satelitne DNA sekvence, iako čine značajan dio eukariotskog genetičkog materijala još uvijek predstavljaju nedovoljno poznatu genomsku komponentu nejasne evolucije i funkcionalnog značaja (Plohl i sur. 2012). Jedan od razloga tome je i činjenica da genomski projekti tek fragmentarno obuhvaćaju ponovljene sekvence DNA, prvenstveno zbog poteškoća u sklapanju dugih nizova međusobno vrlo sličnih jedinica ponavljanja (Garrido-Ramos 2017, Eichler i sur. 2004). Još donedavno, satelitne DNA eksperimentalno su se otkrivale restrikcijom razgradnjom genomske DNA, no takav pristup imao je svoja ograničenja. Naime, takvim se načinom omogućilo otkrivanje samo dominantnih, brojem kopija najzastupljenijih satDNA porodica, dok su one prisutne u malom broju kopija ostale nezabilježene. Međutim, razvojem tehnologije sekvenciranja nove generacije te popratnih bioinformatičkih alata došlo je do revolucije u analizama satelitoma. Tako su primjerice, *high-throughput* analizom satelitoma skakavca *Locusta migratoria* otkrivene čak 62 satDNA porodice, od kojih je većina prisutna u genomu ispod 0,01% (Ruiz-Ruano i sur. 2016). Nadalje, Palacios-Gimenez i suradnici (2017) zabilježili su 45 satDNA porodica analizirajući satelitom skakavca *Eneoptera surinamensis* koje variraju od 0,05-1,43% ukupnog genoma. Dodatno je primjenjivost kratkih NGS očitavanja u identifikaciji niskoizastupljenih satelitnih DNA dokazana i prilikom analize genoma vrste *Pyrgomorpha conica* unutar koje je otkriveno 76 satDNA porodica čiji udio u genomu većinom nije prelazio 0,1% (Ruiz-Ruano i sur. 2018). Kao što je vidljivo iz navedenih primjera, studije podataka NGS računalnim programima specijaliziranim za analize ponavljajućih DNA omogućile su detekciju niskoizastupljenih satelitnih sekvenca koje bi tradicionalnim metodama ostale neotkrivene.

U ovom je radu genom vrste *T. castaneum* po prvi puta sekvenciran metodom NGS u svrhu otkrivanja niskoizastupljenih satDNA sekvenca. Kukac *T. castaneum* izvrstan je modelni organizam koji se već dugi niz godina koristi u molekularnim istraživanjima (Brown i sur. 2009). Jednako tako odličan je model za izučavanje ponavljajućih sekvenca budući da mu je više od petine genoma izgrađeno od satDNA (Ugarković i sur. 1996, Feliciello i sur. 2015, Pavlek i sur. 2015). Uz glavnu satDNA TCAST čiji udio iznosi oko 17% (Ugarković i sur. 1996), Pavlek i sur. (2015) su u sastavljenom genomu, dobivenom pristupom „shotgun“ (Richards i sur. 2008), detektirali još 9 satDNA porodica, koje u ukupnom zbiru čine 4,3% genoma, a prisutne su prvenstveno u eukromatinskim regijama. No, 20% genoma koji još

uvijek nije posložen, a zahvaća velik dio heterokromatinskih regija, potencijalno sadrži još dosta neotkrivenih satelitnih sekvenca.

U okviru ovog diplomskog rada na temelju podataka NGS sekvenciranog genoma vrste *T. castaneum* uz upotrebu bioinformatičkog programa TAREAN, detektirane su, uz već prethodno opisane, i nove, niskozastupljene satDNA. Među novootkrivenim satelitima za daljnju eksperimentalnu karakterizaciju odabrane su satDNA TCst154, TCst266 i TCst309. Za sva tri kandidata hibridizacijska analiza prema Southernu pokazala je uzastopno ponavljajuću organizaciju monomera, potvrđujući da se uistinu radi o tipičnim satelitnim sekvencama. Sve tri novoopisane satDNA pokazuju strukturne karakteristike sukladne satelitnim DNA sekvencama prethodno opisanim unutar vrsta roda *Tribolium*: duljinu monomera koja najčešće varira između 150-360 pb te visok udio AT parova baza (Ugarković i sur. 1996, Mravinac i sur. 2004, Pavlek i sur. 2015). Preferencijalna duljina osnovnih ponavljajućih jedinica u korelaciji je s duljinom monomera prethodno opisanih porodica satelitnih sekvenci te se može povezati s učinkovitim pakiranjem molekule DNA. Naime, dobro je poznato kako kromatin u eukariotskim genomima ima nukleosomalnu strukturu prilikom koje se DNA 1,67 puta omata oko histonskog oktamera (147 pb). Udaljenost između dva nukleosoma može varirati od 10-70 pb te se ta DNA naziva vezna DNA (Richmond i Davey 2003). Također, visok udio AT parova baza, zabilježen i u satDNA ostalih crnokrilaca, određuje zakrivljenost molekule DNA uzrokovanu nukleotidnom sekvencom te potencijalno olakšava tijesno pakiranje heterokromatina (Meštović i sur. 2000, Barcelo i sur. 1998).

Bioinformatičkom analizom monomerne kopije TCst154 i TCst266 nisu zabilježene u sastavljenom genomu vrste *T. castaneum*, već su prisutne isključivo u neposloženim skafoldima i neposloženim jedinstvenim sekvencama te se može pretpostaviti da spomenute satDNA pripadaju heterokromatinskoj frakciji genoma. Zanimljivo je da TCst154 i TCst266, koje bi se, dakle, mogle smatrati tipičnim heterokromatinskim satelitima, svojim karakteristikama (udio u genomu <0,5%, AT sastav >65%, nepostojanje HOR tandema, nepostojanje subvarijanata) u potpunosti prate obilježja takozvanih „eukromatinskih“ satelita opisanih u prethodnim istraživanjima vrste *T. castaneum* (Pavlek i sur. 2015). Na temelju spomenute konstatacije moguće je spekulirati da kod vrste *T. castaneum* mehanizmi koji utječu na genuzu i evoluciju satDNA ne diskriminiraju satelitne sekvence ovisno o njihovom kromosomskom položaju.

Za razliku od satDNA TCst154 i TCst266, satelit TCst309 prema svom kromosomskom smještaju priklanja se grupi „eukromatinskih“ satelita Cast1-Cast9 (Pavlek i sur. 2015), a također i ostalim karakteristikama prati njihova obilježja. Položaj ovog satelita na kromosomu 7 podržan je bioinformatičkom analizom kao i hibridizacijom *in situ*. S obzirom na ekskluzivan smještaj na samo jednom kromosomu, ovaj satelit bi mogao poslužiti kao jedinstveni kromosomski bar kod s obzirom na teško morfološko razlikovanje kromosoma vrste *T. castaneum*. Na sličan način otkrivena je u genomu vrste *Mus musculus* satDNA TRPC-21-MM, identificirana *in silico* te potvrđena *in situ*, koja može poslužiti kao citogenetička sonda za identifikaciju mišjih kromosoma 3 i 17 (Komissarov i sur. 2011).

Uz kromosom-specifičnost, satDNA TCst309 odlikuje se također niskom varijabilnošću svojih monomera, bilo da se uspoređuju monomeri kloniranih PCR odsječaka ili pak monomeri ekstrahirani iz trenutno dostupnog genomskog sklopa Tcas5.2. Na temelju velike homogenosti TCst309 monomera, može se zaključiti kako je satDNA TCst309 evolucijski gledano „mlada“ satelitna sekvenca potencijalno u početnom stadiju svoje propagacije unutar genoma vrste *T. castaneum*. U prilog takvom razmišljanju govori i njezin položaj na samo jednom kromosomu. Subirana i Messeguer (2017) su nakon usporedbe satelitoma dvaju srodnih vrsta *Caenorhabditis nigoni* i *Caenorhabditis briggsae* uz satDNA porodice prisutne u obje vrste pronašli i satDNA porodice specifične za pojedinu vrstu koje također pokazuju izrazitu homogenost osnovnih ponavljajućih jedinica te samim time ukazuju na njihov nedavni nastanak. S druge pak strane, homogenost nekog satelita ne mora nužno odražavati njegovu evolucijsku starost, već može biti i odraz eventualnog evolucijskog pritiska na očuvanje nukleotidnog slijeda. Poznati su primjeri satelitnih sekvenci koje su dobro očuvane unutar različitih vrsta. Primjerice, dodeca satelit inicijalno pronađen u vrsti *Drosophila melanogaster* detektiran je i u vrstama *Arabidopsis thaliana* i *Homo sapiens* (Abad i sur. 1992). Jedan od najdrastičnijih primjera sačuvanosti satDNA prema literaturnim podacima je porodica satDNA BIV160 za koju se procjenjuje da je unutar razreda Bivalvia prisutna oko 500 milijardi godina (Plohl i sur. 2010). Prema hipotezi knjižnice (Fry i Salser 1977), postoji mogućnost da se satDNA TCst309 također nalazi u genomima drugih vrsta koje pripadaju rodu *Tribolium*.

Program TAREAN pokazao se učinkovitim u detekciji niskozastupljenih satelitnih DNA sekvenci u genomu vrste *T. castaneum*, kako satDNA prisutnih u eukromatinu, tako i onih smještenih u heterokromatinskim regijama. Međutim, nije isključeno da su neke satDNA, zastupljene u izrazito malom broju kopija, i dalje ostale nezabilježene. Naime, s

obzirom na visok udio koji glavni satelit zauzima u genomu vrste *T. castaneum* (~17%), moguće je da je glavni satelit prisutan u tako izrazito velikom broju kopija opteretio algoritam za pronalazak uzastopno ponovljenih sekvenci te su sekvence prisutne u izrazito malom broju kopija promaknule detekciji. Upravo iz spomenutog razloga, u analizama satelitoma često se koristi pristup iterativnog „pročišćavanja“ genoma, odnosno, eliminacije dominantnih i/ili prethodno opisanih satelitnih sekvencija iz inicijalnog skupa NGS očitavanja te se time omogućava sveobuhvatnija i preciznija detekcija niskozastupljenih satelita. Tako su primjerice Ruiz-Ruano i suradnici (2016, 2017), analizirajući satelitome dvaju vrsta skakavaca *Locusta migratoria* i *Eumigus monticola*, koristeći spomenuti pristup pronašli veći broj satDNA nego što su pokazale inicijalne bioinformatičke analize. Ovakva strategija zasigurno bi pridonijela otkrivanju još većeg broja niskozastupljenih satDNA i u genomu vrste *T. castaneum*.

Računalni program TAREAN uz identifikaciju tandemskih sekvencija također procjenjuje i njihov potencijalni udio u genomu. Na uzorku od opisanih tri satelita, računalno procijenjen udio podržan je eksperimentalnom analizom za satDNA TCst266 te približno podržan za satDNA TCst309. Međutim, u slučaju satDNA TCst154 računalni je program za red veličine podcijenio udio ovog satelita u odnosu na eksperimentalnu procjenu metodom dot blot. Diskrepancija u računalnoj procjeni također je uočena za satelite Cast1-Cast9, čije su genomske udjele Pavlek i sur. (2015) eksperimentalno procijenili višima. Moguće objašnjenje je da eksperimentalna metoda dot blot varirajući temperaturu hibridizacije ima mogućnost veće fleksibilnosti u pronalaženju varijanata određene sekvence, dok je računalni program možebitno rigidniji te u obzir uzima samo vrlo slične sljedove i na taj način podcijenjuje udio heterogenijih satelitnih porodica.

Razvoj NGS tehnologije te popratni razvoj bioinformatičkih alata specijaliziranih za analizu repetitivnih sekvencija bez sumnje predstavljaju prekretnicu u istraživanju satelitoma. Međutim, unatoč otvorenim mogućnostima detekcije satelitnih sekvencija prisutnih čak i u izrazito malom broju kopija, još uvijek ostaje neriješen problem preciznog asembliranja repetitivnih regija. Čak i najbolje asembliranim genomima nedostaju regije sačinjene od uzastopno ponovljenih sekvencija, odnosno satDNA, što ih čini nepotpunima (Miga 2015, 2017). U tom smislu svakako hrabri recentna studija Jain i suradnika (2018), koji su, koristeći metodu sekvenciranja nanoporama (The Oxford Nanopore Technologies MinION, ONT), uspjeli asemblirati centromeru ljudskog Y kromosoma, uglavnom sačinjenu od satDNA DYZ3. Iako je asembliranje spomenute centromere bilo olakšano činjenicom da se radi o

relativnoj maloj centromernoj regiji, dobro okarakteriziranoj strukturi HOR-a satDNA DYZ3, postojanju određenog broja eksperimentalnih biljega u vidu varijanti sekvence te već postojećim podacima koji su nastali kao rezultat fizičkog kartiranja, ovakav rezultat predstavlja izrazito važan napredak u anotaciji uzastopno ponovljenih sekvenca te rasvjetljavanju njihove funkcije i evolucije.

Uzimajući sve u obzir, može se zaključiti kako će kompletiranje eukariotskih genoma biti moguće tek kada budemo imali usavršenu tehnologiju za precizno asembliranje i anotaciju repetitivnih regija odnosno njihovo visoko-kvalitetno sekvenciranje. Postojanje cjelovite informacije o strukturi genoma zasigurno će osigurati platformu za bolje razumijevanje i točniju interpretaciju uloge satDNA u pravilnom djelovanju centromera, kromosomskoj stabilnosti te povezanosti s bolestima.

5. ZAKLJUČAK

Istraživanje nisko zastupljenih satelitnih DNA u genomu vrste *Tribolium castaneum*, analiziranih u sklopu ovog rada, rezultira sljedećim zaključcima:

1. Bioinformatičkom analizom kratkih odsječaka DNA, dobivenih sekvenciranjem genoma vrste *T. castaneum* pomoću metode NGS, identificirane su dosad neopisane satelitne DNA čiji pojedinačni udjeli u genomu iznose <0,5%.
2. Tri *in silico* identificirane satelitne DNA TCst154, TCst266 i TCst309 karakterizirane su dodatno eksperimentalnim metodama kako bi se potvrdila njihova satelitna autentičnost. Metodom Southern blot hibridizacije dokazana je uzastopno ponavljajuća organizacija ovih sekvenca, a fluorescencijskom *in situ* hibridizacijom pokazana njihova kromosomska lokalizacija.
3. Kombinirajući bioinformatičke alate i eksperimentalne metode, ovim je istraživanjem potvrđena primjenjivost kratkih, neposloženih NGS sljedova u otkrivanju nisko zastupljenih satelita, a također je proširena i spoznaja o sadržaju satelitoma vrste *T. castaneum*.

6. LITERATURA

- Abad J. P., Carmena M., Bars S., Saunders R. D., Glover D. M. (1992): Dodeca satellite: a conserved G+C-rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. **89**: 4663-4667.
- Abramoff M. D., Magalhães P. J., Ram S. J. (2004): Image processing with ImageJ. Biophotonics Intern. **11**: 36-42.
- Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990): Basic local alignment search tool. J. Mol. Biol. **215**: 403-10.
- Barcelo F., Gutierrez F., Barjau I., Portugal J. (1998): A theoretical persual of the satellite DNA curvature in tenebrionid beetles. J. Biomol. Struct. Dyn. **16**: 41-50.
- Benson G. (1999): Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**: 573-580.
- Biedler J., Tu Z. (2003): Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. Mol. Biol. Evol. **20**: 1811-1825.
- Biscotti M. A., Canapa A., Forconi M., Olmo E., Barucca M. (2015): Transcription of tandemly repetitive DNA: functional roles. Chromosome Res. **23**: 463-477.
- Brown S. J., Henry J. K., Black W. C., Renell R. E. (1990): Molecular genetic manipulation of the red flour beetle: genome organization and cloning of ribosomal protein gene. Insect Biochem. **20**: 185-193.
- Brown S. J., Shippy T. D., Miller S., Bolognesi R., Beeman R. W., Lorenzen M. D., Bucher G., Wimmer E. A., Klingler M. (2009): The red flour beetle, *Tribolium castaneum* (Coleoptera): a model for studies of development and pest biology. Cold Spring Harb. Protoc. **8**: e126.
- Cohen S., Yacobi K., Segal D. (2003): Extrachromosomal circular DNA of tandemly repeated genomic sequences in *Drosophila*. Genome Res. **13**: 1133-1145.

- Dover G. A. (1986): Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet.* **2**: 159-165.
- Eichler E. E., Clark R. A., She X. (2004): An assessment of the sequence gaps: unfinished business in finished human genome. *Nature Rev. Genet.* **5**: 345-354.
- Feliciello I., Akrap I., Ugarković Đ. (2015): Satellite DNA modulates gene expression in beetle *Tribolium castaneum* after heat stress. *PLoS Genet.* **11**: 1005466.
- Fry K., Salser W. (1977): Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* **12**: 1069-1084.
- Garrido-Ramos M. A. (2017): Satellite DNA: an evolving topic. *Genes* **18**: E230
- Garrido-Ramos M.A. (2015): Satellite DNA in plants: more than just rubbish. *Cytogenet. Genome Res.* **146**:153-170.
- Hall T. A. (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95-98.
- Heinkoff S., Ahmad K., Malik H. S. (2001): The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**: 1098-1102.
- Howe R. W. (1956): The effect of temperature and humidity on the rate of development and mortality of *Tribolium castaneum* (Herbst) (Coleoptera, Tenebrionidae). *Ann. Appl. Biol.* **44**: 356-368.
- Jain M., Olsen H. E., Turner D. J., Stoddart D., Bulazel K. V., Paten B., Haussler D., Willard H. F., Akeson M., Miga K. H. (2018): Linear assembly of a human centromere on the Y chromosome. *Nat. Biotechnol.* **36**: 321-323.
- Juan C., Petitpierre E. (1989): C-banding and DNA content in seven species of Tenebrionidae (Coleoptera). *Genome* **32**: 834-839.
- Juan C., Petitpierre E. (1991): Karyological differences among Tenebrionidae (Coleoptera). *Genetica* **80**: 101-108.
- 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**: 1-8.

- Khan I., Prakash A., Agashe D. (2016): Divergent immune priming responses across flour beetle life stages and populations. *Ecol. Evol.* **6**: 7847-7855.
- Komissarov A. S., Gavrilova E. V., Demin S. J., Ishov A. M., Podgornaya O. I. (2011): Tandemly repeated DNA families in the mouse genome. *BMC Genomics* **12**: 531.
- Larkin M. A., Blackshields G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H., Valentin F., Wallace I. M., Wilm A., Lopez R., Thompson J. D., Gibson T. J., Higgins D. G. (2007): Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.
- Lorenzen M. D., Doyungan Z., Savard J., Snow K., Crumly L. R., Shippy T. D., Stuart J. J., Brown S. J., Beeman R. W. (2005): Genetic linkage maps of the red flour beetle *Tribolium castaneum* based on bacterial artificial chromosomes and expressed sequence tags. *Genetics* **170**: 741-747.
- Martinez-Lage A., Gonzalez-Tizon A., Mendez J. (1994): Characterization of different chromatin types in *Mytilus galloprovincialis* L. after C-banding, fluorochrome and restriction endonuclease treatments. *Heredity* **72**: 242-249.
- Meštrović N., Castagnone-Sereno P., Plohl M. (2006): Interplay of selective pressure and stochastic events direct evolution of the MEL172 satellite DNA library in root-knot nematodes. *Mol. Biol. Evol.* **23**: 2316-2325.
- Meštrović N., Mravinac B., Juan C., Ugarković Đ., Plohl M. (2000): Comparative study of satellite sequences and phylogeny of five species from the genus *Palorus* (Insecta, Coleoptera). *Genome* **43**: 776-785.
- Meštrović N., Plohl M., Mravinac B., Ugarković Đ. (1998): Evolution of satellite DNAs from the genus *Palorus*-experimental evidence for the library hypothesis. *Mol. Biol. Evol.* **15**: 1062-1068.
- Miga K. H. (2015): Completing the human genome: the progress and challenge of satellite DNA assembly. *Chromosome Res.* **23**: 421-426.
- Miga K. H. (2017): The promises and challenges of genomic studies of human centromeres. *Prog. Mol. Subcell. Biol.* **56**: 285-304.
- Mravinac B., Plohl M. (2007): Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. *Gene* **394**: 45-52.

- Mravinac B., Plohl M. (2010). Parallelism in evolution of highly repetitive DNAs in sibling species. *Mol. Biol. Evol.* **27**:1857-1867.
- Mravinac B., Plohl M., Ugarković Đ. (2004): Conserved patterns in the evolution of *Tribolium* satellite DNAs. *Gene* **332**: 169-177.
- Mravinac B., Ugarković Đ., Franjević D., Plohl M. (2005): Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA. *J. Mol. Evol.* **60**: 513-525.
- Novák P., Neumann P., Macas J. (2010): Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**: 378.
- Novák P., Neumann P., Pech J., Steinhaisl J., Macas J. (2013): RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. *Bioinformatics* **29**: 792-793.
- Novák P., Robledillo A. L., Koblížková A., Vrbová I., Neumann P., Macas J. (2017): TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **45**: e111.
- Ohki R., Oishi M., Kiyama R. (1995): Preference of the recombination sites involved. *Nucleic Acids Res.* **3**: 4971-4977.
- Palacios-Gimenez O.M., Dias G. B., de Lima L. G., Kuhn G. C. S., Ramos E., Martins C., Cabral de Mello D. C. (2017): High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*. *Sci. Rep.* **7**: 6422
- Palomeque T., Lorite P. (2008): Satellite DNA in insects: a review. *Heredity* **100**: 564-573.
- Pavlek M., Gelfand 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 Res.* **2**: 387-401.
- Plohl M., Lucijanić-Justić V., Ugarković Đ., Petitpierre E., Juan C. (1993): Satellite DNA and heterochromatin of the flour beetle *Tribolium confusum*. *Genome* **36**: 489-495.

- Plohl M., Meštrović N., Mravinac B. (2012): Satellite DNA evolution. U: Garrido-Ramos M. (ur.) Repetitive DNA. Basel, Genome Dynamics, str. 126-52.
- Plohl M., Petrović V., Luchetti A., Ricci A., Šatović E. (2010): Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity* **104**: 543-551.
- Price A. L., Jones N. C., Pevzner P. A. (2005): *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**: 351-358.
- Richards S. et al. (2008): The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949-955.
- Richmond T. J., Davey C. A. (2003): The structure of DNA in the nucleosome core. *Nature* **423**: 145-150.
- Ruiz-Ruano F. J., Cabrero J., Lopez-Leon M. D., Camacho J. P. M. (2016): Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome. *Chromosoma* **126**: 487-500.
- Ruiz-Ruano F. J., Castillo-Martínez J., Cabrero J., Gómez R., Camacho J. P. M., López-León M. (2018): High-throughput analysis of satellite DNA in the grasshopper *Pyrgomorpha conica* reveals abundance of homologous and heterologous higher-order repeats. *Chromosoma* **127**: 323-340.
- Ruiz-Ruano F. J., Cabrero J., López-León M.D., Sanchez A., Camacho J. P. M. (2017): Quantitative sequence characterization for repetitive DNA content in the supernumerary chromosome of the migratory locust. *Chromosoma* **127**: 45-57.
- Ruiz-Ruano F. J., Lopez-Leon M. D., Cabrero J., Camacho J. P. (2016): High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* **6**: 28333.
- Schueler M. G., Higgins A. W., Rudd M. K., Gustashaw K., Willard H. F. (2001): Genomic and genetic definition of a functional human centromere. *Science* **294**: 109-115.
- Smit A., Hubley R., Green P. (2015): RepeatMasker Open-4.0. 2013–2015. Institute for Systems Biology.

- Smith G. (1976): Evolution of repeated DNA sequences by unequal crossover. *Science* **191**: 528-535.
- Smith S. G. (1952): The evolution of heterochromatin in the genus *Tribolium* (Tenebrionidae: Coleoptera). *Chromosoma* **4**: 585-610.
- Sokoloff A. (1972) The biology of *Tribolium*, with special emphasis on genetic aspects. Vol I, Oxford University Press, Oxford.
- Subirana J. A., Messeguer X. (2017): Evolution of tandem repeat satellite sequences in two closely related *Caenorhabditis* species. Diminution of satellites in hermaphrodites. *Genes* **8**: 351.
- Sun X., Le H. D., Wahlstrom J. M., Karpen G. H. (2003): Sequence analysis of a functional *Drosophila* centromere. *Genome Res.* **13**: 182-194.
- Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. (2013): MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725-2729.
- Tomoyasu Y., Miller S. C., Tomita S., Schoppmeier M., Grossman D., Buchner G. (2008): Exploring systemic RNA interference in insects; a genome-wide survey for RNAi in *Tribolium*. *Genome Biol.* **9**: R10.
- Treangen T. J., Salzberg S.L. (2011): Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**:36-46.
- Ugarković Đ., Podnar M., Plohl M. (1996): Satellite DNA of the red flour beetle *Tribolium castaneum*-comparative study of satellites from the genus *Tribolium*. *Mol. Biol. Evol.* **13**: 1059-1066.
- Ugarković Đ., Petitpierre E., Juan C., Plohl M. (1995): Satellite DNAs in tenebrionid species- structure, organization and evolution. *Croat. Chem. Acta* **68**: 627-638.
- Ugarković Đ. (2005): Functional elements residing within satellite DNAs. *EMBO Rep.* **6**: 1035-1039.
- Ugarković Đ., Plohl M. (2002): Variation in satellite DNA profiles-causes and effects. *EMBO J.* **21**: 5955-5959.

Verdel A., Jia S., Gerber S., Sugiyama T., Gygi S., Grewal S. I., Moazed D. (2004): RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672-676.

Volpe T. A., Kidner C., Hall I. M., Teng G., Grewal S. I., Martienssen R. A. (2002): Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833-1837.

Wang S., Lorenzen M. D., Beeman R. W., Brown S. J. (2008): Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome Biol.* **9**: R61

Weith A. (1985): The fine structure of euchromatin and centromeric heterochromatin in *Tenebrio molitor* chromosomes. *Chromosoma* **91**: 287-296.

7. PRILOZI

Prilog 1.

Slika P1. Sumarni rezultati TAREAN analize 400 000 očitavanja dobivenih metodom sekvenciranja nove generacije genoma vrste *Tribolium castaneum*. Prikazani su izlazni podaci za potencijalne satelitne sekvence svrstane u kategorije visoke (*high confidence*) ili niske (*high confidence*) vjerojatnosti.

Prilog 2.

Tablica P2. Prikaz duljina kloniranih odsječaka i duljina ekstrahiranih monomera satDNA TCst154 vrste *Tribolium castaneum*.

Prilog 3.

Tablica P3. Prikaz duljina kloniranih odsječaka i duljina ekstrahiranih monomera satDNA TCst266 vrste *Tribolium castaneum*.

Prilog 4.

Tablica P4. Prikaz duljina kloniranih odsječaka i duljina ekstrahiranih monomera satDNA TCst309 vrste *Tribolium castaneum*.

Prilog 5.

Slika P5. Poravnanje sekvenci monomera satelitne DNA TCst154 ekstrahiranih iz izlaznih podataka sekvenciranog genoma vrste *Tribolium castaneum* (Tcas5.2) i monomera ekstrahiranih iz klonova analiziranih u ovom radu.

Prilog 6.

Slika P6. Poravnanje sekvenci monomera satelitne DNA TCst266 ekstrahiranih iz izlaznih podataka sekvenciranog genoma vrste *Tribolium castaneum* (Tcas5.2) i monomera ekstrahiranih iz klonova analiziranih u ovom radu.

Prilog 7.

Slika P7. Poravnanje sekvenci monomera satelitne DNA TCst309 ekstrahiranih iz izlaznih podataka sekvenciranog genoma vrste *Tribolium castaneum* (Tcas5.2) i monomera ekstrahiranih iz klonova analiziranih u ovom radu.

Prilog 1.

Tandem Repeat Analyzer

Run statistics:

Number of input sequences: 400000
 Number of analyzed sequences: 400000
 Cluster merging: Yes

Consensus files - fasta format:

[Putative satellite \(high confidence\) - total 5 found](#)
[Putative satellite \(low confidence\) - total 15 found](#)

Documentation

For the explanation of TAREAN output see [the help section](#)

Putative satellite (high confidence)

Cluster	Genome Proportion [%]	Genome Proportion Adjusted [%]	Size real	Satellite probability	Consensus length	Consensus	Graph layout	Kmer analysis	Connected component index C	Pair completeness index P	Kmer coverage	V	E	Pbs score	Similarity based annotation
1	CL1	18.000	18.000	0.976	363	TACAGTTTTCTTATGGTTTTTGTCTATATCTCGAAAAAGTTACTCCATCAATTTTTATCTTTTACTAAAAATAAA GCTGATAAAATTTCCACAAAATAGTTATTTGCATTTTTCTGTAGGACCAACCTAAGCGAGTATAGAGTGGATAA AATAATATTTAAACAAAAATTTGCTTTAAAAATAAATGTTGAAAAACGAAATTAACATAAATTAATGTCTAAACCT TTAGTAGAGGAAAAAGGAGAGACATAAAAGTCACTAAAGCTTCAGAGTCGTTTTAAATGCTGCATTTTCCTCTCCT CTCAAACTTGAAAACTGAATTAACATTTTGTTCAGTAACAGT		report	0.987	0.941	0.416	20610	19900000	0	0.00% organelle/plastid
6	CL6	0.860	0.860	0.908	311	GATATTTTGGGTGACACATTTTAAATACAGTAACATAAAATAAAATAAACTCAAATAATTTTATGCCAGGAAATTT GCCTTTAAATTTCAATGAATGAAAGTGGCTGCTCAAAATTAATGCAATAAATCAATTTTTTGGTTTCATTTTATAT TATTTTTTAAATGCTACAAACGGTAGCCCAAGCTGGAAGCATATGAAGTAGGTGATTTTATTGCTGTGTCTGATTTA TTTATTTACAAACCAATAATTTCAATAAACTAATTGCAATTAATACTTAAATGCGGGTGTGAAAAAT		report	0.956	0.919	0.462	3442	490000	0	0.06% organelle/plastid
20	CL20	0.170	0.170	0.846	180	ATTAAATATCCCGCATTTAATTAAGAAAAACAAAAAGCCGACGACTCCAATTTTTAATATGTTGTAGAGCAAGTCT TGTAATGTTCTCCACAAAATGAAAGCAATGCTGCATGCAATCCACACGCGCATGGTCAATCAAAACCGTGGTTTT ATGCGAATTTTCAAAAAAT		report	0.946	0.925	0.547	691	13000	0	
36	CL36	0.076	0.076	0.739	721	ACGTTAAAACTCACCCCTAAGGGTGAATCTAAAAAATATATAAGAGCCATTATCTAGGCATAAAAACACAACAT ATTTAAATTTGGTTAATTTTGTTTTTTTTAAATTTTACATTTTTACCCTTAACTACCCCAATACAAAGGGTGGT TTTTGCTAAAAATTTTTTATTAATATTTTAGGGGC TACCTATATGCAAAAAACCGTATCGATTAGGCTTTATTTGTA TTTTGGTTTTTTTGTCAATTTTAGGGTGTTTTTCGAATGAAAGGGTGGTCAATCCACATTTAATAAAATTTTTTT TAATCTTACGGCGCATGTCTATCGATTAAAGTTAAAAAATTTAAAAATTTAATCATGACTTTTGTGACTTTTATCCAGCTC AAAGTCCGCATTAGTGGTCCGGGTGGTGGGGGAAGTGACTTTAGATCTAATGAGATAAATACTGTTTTTTTTCAT ATGTCGGGGGAGCATGACTAAGCCTTTAGAAAGTTGTTGATGATGACGCCCTTTAGCTTACGGTATCATGGGACT TTGGTGGGAAAAATGACAGTCAAGCGCATGCGCATGCGCTCTGGACCAATGAAAAACGAAAGCTAAGAGAAATTCG GAGTTCTTTTTGAGGCTAAAAATTTTAAAAACATCTTGACATTGTTGAATTTTTGAATTTAGTGGTCTTTAA T		report	0.974	0.854	0.439	304	3450	0	
52	CL52	0.054	0.054	0.815	154	GAAAGAGCATTAACTCTACATATAAAAAAGTTAAAAAATCGAGCCCTGGGAGGGGTTGAAAAGTTGATTTGT TTAATAATTTTGATGGTTTTTATTTTTTCTATGTTGTTATCGAATGTCACACTTCTGTTACGGCAGGTC		report	0.945	0.912	0.657	218	7240	0	

Putative satellite (low confidence)

Cluster	Genome Proportion	Size real	Satellite probability	Consensus length	Consensus	Graph layout	Kmer analysis	Pair completeness	Kmer coverage	[V]	[E]	Pbs score	Similarity based annotation						
	Adjusted [%]						Connected component index C	Index P											
2	CL2	6.1000	6.1000	24466	0.0494	1460	AGTTTTAAAAAGTATTAATTTTANGCAATTTCTGSCCTAAAAACGCAATCACAGATGATCTTGGCTAGAAATGA GAAAAGCGTCAGATAGCACAATAAAGCAATCTTTTCTGTAAGAAAACGANGCTTAAATTTAGAAAAGAAATTCATTAAT ATAGATATATATCATATATATATATATATATATATATAAAAATTAATTAATTAACAGACAGAAAACCTGATTTCT GACAAAAAAACCAATACGTTATAGCTTTGACAGAAACGGCTATCTGAGACTCTTATAGGACTCTTATAGGGTCAAA AGACTACACTTTGGATTAAGAAATTAATTTTGTGACTCTGACTCGAATTTCTGACTCTCAACCAAAAACGGCTT ATTTTANGAGATCTTTGGCTTTGACCAAAAACCTTACATCGAACAGCAAAAAGCTGATTAATCTGGCATAAAAGGCA ATTTATGATCAGCTTGGATAAGGGGAGCTAATCTCAATCTGCTTCACTGCTTAATAAATAGCAAAAATTTGTAATAAATCTA ACAAAAACGCTGATTTCTAGCACAATAAATCAATTAATTAATTAATTTGAGCAGAAAGCTTATTTGGAAATGATCTAGTG CTGCAAAACGAGTAATTTTATTAACAGATAGCAGAACTCATGTTTGGCTATTAAGGACATTTTATATTTGCGATTTG TATTAAGCACAATAATTTGCTGATCTTTTGTAGAGAGCAGCAAACTTACATTAAGAAATCTGAAAATTTGCTATTTG GACAAAAACCGCAATTTGATCAAGTTGACAAAAACGGTGAATTTGACAAATATTTTGTGTTTGAACAAAACCT TATATCAAACTGACAAAAAGCTCTAATTTAGCACAAAAACAAAGATTTGGCTTATAGTCTAAAGCTGATCTAGGCA CAGTTGGATAAAACCAATTAATTTTGGTAAATCACTGTTACAAAACAAAACCTTACATCAAAAACGTAAGAAACCC TTATTCAGGACAGAAATTCGAAATTAACCTTCACTGGACAAAAACCTGCTATTTAGACGATCTTTTGTGTTGACAC GAAAACCTATATTAAGAAATTAAGAAACCTTATCTGGCTGAAACCTGCAATTTATATTTACTTAAAGTAAGAAACAA TTTATCTTAAATTAACCTTACTTGGATTAAGAAACCTTATTAAGAAACCTTAAATTTGCTTCAAAAAG AATTAATTTAGTTATAGTTTGGACAGCACTGTTTATTAAGCAAAATTAATGATTCGAACAAATTAAGCTTTTATGGGG CAGATGAAAACCAATTAATCATATAAGCGTTGTTTAAAGCAAAATTAATCTTTTCAAAAACAAATTTGACCT AAAGAACTGTTTTATAC		EBSCO1	0.780	0.850	0.450	20435	6980000	0.375	0.13%	organelle/plastid	0.00%	organelle/mitochondria
3	CL3	1.7000	1.7000	6781	0.3190	1110	ACTTCAAAAGAGTGAAMAAAAAATTTTTTAAACCACTCCAGTAAATTTTTTGGACTTCACTATGCTTAAACACCC ACAAAATTTGTTAAAGTCCACAAAAAGCCATATGTTCTGATTTTTGATGTTTGAATTTTTTCAATTTTCGTCGAGTTT TTTGGATTTTTGGGTACCAGGAAATTTTGGTGGCAATAGCTCCGGAACATCAGAGATACCCCAATGAGGTTTATTC GTTGAAACCTCTTAAATATCTATTTTTTCAAAAAAAATTAATTTGCTCCGACATATGTTTTCGAGCAAAATTCGAG ATAAAATGCAAAAATTTGTAATAATTTTTAATAAATTCATACAAAAAACCTATTTGGAAATTTGGCAATTTTTCGATGCCAA TGGATTTGGCGGATCATTTGCTATAGGATGGATCAAAATTAATTTGCTTCTGCAAAATTTGAGCAAAAATGAGAAA TAAAABAAATTAAGAAAGTTAAACCCCCCTTAAATCGGCTAAATTTTTAAGTGTCTTAAATCAAAATTAAGCCG ATTTACTCTATAGATTTTGAATGCTTCTTGGATCAAAAAGCGTTTCTCTAGCTCTTTAGTTTGGGCTGATTC GGGTTTGAATAATGAAAAATTTTTTTCGAGATATGCTCCGAGTCTATGCTAATTTGATAGATTTTTTATTTCCGTTG TCTTCCATTTTTCCGTTTCTGATAACCTTAATAGTTTCCGCTAATTTGGCGGTTTGAATAATGAAAABAAATGAAATGG AAACCTTTTTTTTGGTTTTTCTGGAACTTAAGACTTAAAGCAAAACCACTTGAATAGCTTGAATAGCTTAAATTTTA TTTATTTTTTTCAGCAAAACGGGAGCTTCCGGCAGCAAGGATGATTAAGTTTTTCTAAATTAAGAA ATTTCAAAAABAAACCAAAAATTTGATAGCAATGCGGTTTGTCTCAATTTGCAATTTGCAAGCTTCTTCTTATATAC CAACTTTTCAAGAGATTAATAATTTTTCAATTTTTTTCGCAAAATTTTCTGATTAATTAATTAACACTTA		EBSCO1	0.922	0.819	0.381	6781	700000	0.000				
10	CL10	0.3500	0.3500	1416	0.0148	772	ATTAATAATTAACAGAGCTTCTGGTGAACCTTTGAACTATGTTGAAACAGTAAAGAAATTTTCTGATGCTCAGATT TAAGTACATTTAAATGATTTGCAACAAATTTTCTGATATAGTTTAAATTAATTAACATTAAGAAATTAATTTAAAG TACTGTTGTAATAACTAGCGTTCATTTGGTTCGATGTTTCTGTTGTTTGGTTTAAAGTTTATTTGGAATTTGAAAG TGGATTAAGATGAAAATAATTAATAATTTGATTAACAAATGTTATTAACAACTTTGCTAATTAAGATTAATTAAGT GATCAAGCAACCAAGTTTGGCTCAAAAACCAATAATTAATTAAGTAAATTTAGTTTAAAGTTAAAGTTTTCAC TGAAAAATCTATGAATGCTGGGTAATTTTCAAAAACCTAATTAAGTAAATTAAGTAAATTTAAAGCTAAAT TATGCAAAAAGTGGCTTTTCAAACTCTTTGTAATTAATTTCCAAATTAAGAAATTAAGCAAACTTATGCTTTATC GTTTTTCTCTCTCATCTCTCAATTTTCAACATAAGTTTCTGATCAAAAGTCAAAATTAATTTCCGATAGCAATTT AATAATTTGAAAGCCCAAAATTTGCAACAGCTCAAAATTAATTAAGCAATTTTCCGCAATTTATGTTACTGTTT TGAACCTAGCTAGTGGGCCCCCAAGGATTAATTTACTTCCGAATTTGAAA		EBSCO1	0.861	0.556	0.834	1416	67100	0.000				
16	CL16	0.2500	0.2500	986	0.0451	528	GTTTTGAGAAAAGAAAGATGAACTTAAGGTAAGATTAACCTCAACTTTAGTTGACTTATAGCCCGGATCTTTTGT ATGAAATCTGCGAATAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTTTCTGAGCTAATTCGCGACTTT TGAACCTGAATGATTTTGGCTTTTAAAGTAAATTTATGCAATTAATTAATTTGTAATTTGATTTTAAAGAAATTTTGT TAAATTAATCTGAAAAGAGCGTCAATGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT GGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT AAATTAAGGATTAAGGATGCTTAAAGGCTCAAAATTTGTAACGACTTTTTT		EBSCO1	0.949	0.742	0.400	986	23000	0.227				
18	CL18	0.2000	0.2000	810	0.4410	166	ACTCAATTTAGGACATAGGCTTAAATTAAGAAAGAAATAATTAAGGACTTATGTTGTTGATTAATTTGACAGGCTAA ATTTTTGCGCATATTTTCCAAAGCTTAACTGCTAGGCACTGATCTTGGCAAAATTTGATTTCCGCGGCTACTTCTAGCT AGAAC		EBSCO1	0.923	0.862	0.599	810	68800	0.000				
33	CL33	0.0880	0.0880	351	0.0116	73	AATTTAAAAATGGCGGAGAGCTTGTAAACAGTTGGCAACATAAAAATTCGACGAAATCGGTAATTA		EBSCO1	0.704	0.729	0.389	351	5310	0.000				
61	CL61	0.0400	0.0400	161	0.0485	142	GTTTTCAATTTAAATTCGAGGCGGAATCACTTTAAGTTTAAATTTAAACAAATCAAAATGAAAAGAACTTGAATGGCC TTTCAGGGCGGTTTGAATTTAATTAATAATTTAATTTCCGATTAAGTAATTAATTTT		EBSCO1	0.826	0.750	0.494	161	2190	0.000				
104	CL104	0.0160	0.0160	64	0.0806	266	AGTTGGTTTTTTACATGGATAGCTTTATGGCTCTGATATGCTCCAGCACTAAATTTGACATATTCAAAATAAGGG GGTATGGGGAATTAAGATTCAAAATTCAAAATAATTAATAATCAACCAATTAATAATGATTAATAGTGGGAACTCA GATCAATTTAAATTAATTTTTCAGAAATTTATCTCACTTCTACCCAAAACCTATAGGGGTAATTTGGCCAAAATTT GAAATTTTTATTAACATTCGGATCG		EBSCO1	0.875	0.882	0.700	64	453	0.000				

127	CL127	0.0120	0.0120	50	0.1670	309	GTATTATATTTACTTGAAGTGAATTTTGTGATAATTCACAAAAAAGTGGTTCAGTCTAAATTTTACAAT AATTTAACTTTTTAGGCTACATCTTCCAAGAAGGAAAAATACAAAGTACCACCTGCTGAACCAGCTGAAGTTACAAACG GAACAGATGAAAAAATAACAAAACGCAATAAGCAGAAAGTGAATATGTTGACACCAAGCTTTTTTGTTTTATGA AGTGTATTTAGCCTGTAAGCGTTGAAAATGTCGAGGAAACTGTTGCACCTAAAGCATGGAAAAACAG	report	0.900	0.852	0.894	50	292	0.000
145	CL145	0.0100	0.0100	41	0.0119	167	GTCGGAAACAATGTTAAAAATGTTATGTTGTCACAGTCGCAATGCACTATTGCAAGTTTCGCCCATTTACCCCGAA ATCGGTTTTGTCACACAATGCCCTTTTGACCACAATTCACCAATAAAAAAGACGTTTTTAAAAATAAAATATGGCT TTTAGTT	report	0.976	0.577	0.853	41	345	0.000
158	CL158	0.0092	0.0092	37	0.0152	164	GTAATTTTC TTTTGTGACGCAATTAACCCATTCCACATGAAAGCCATTAAATACTACCTTTCTGAAAATTTCA ACCCCTAAAAGTTTGGGAAGACATCCAACCTTAGTTGAAATATGTCACACTTTTATTAATAATTAATATTTTTT TCTA	report	0.703	0.609	0.618	37	162	0.000
160	CL160	0.0090	0.0090	36	0.0417	125	TTTCTGTAATGCCAACCTAAAAATCGCACTAAATAAATTAATAAGGGCGGACACGTTTGAAGTAGACCACCC TAAAAATCGCACTAAATCCGTAACATAAATTTAAAAATGCCACGTG	report	0.861	0.714	0.694	36	362	0.000
195	CL195	0.0075	0.0075	30	0.0885	102	TTAAATTTATTTTCAAATTAACAAATTTATGCGACTGAAACGAATTAACATTGAACGTGAATGCGGTCGAGTACAT TTTTATTAATTTGAAGCGCAA	report	0.800	1.000	0.741	30	215	0.000
199	CL199	0.0072	0.0072	29	0.4870	178	GGTTTTCTTAAACAATGAAATTTGTACACTTATAACCTTGAATGGTGTCTACCATCTGCCAAAATTCGTTGAAAAA TTAATGGAACTTTGCGAAAAAATTCGATTTATGTTGGGAGGAGGAAAAACGTAAGCTGCCTCACTTAAATAAAAAA TTCATATCTCCGAAACC	report	0.897	0.933	0.816	29	118	0.000
248	CL248	0.0058	0.0058	23	0.0514	120	CTCTGAGATTAACCTAAACCTGACGAAATCACGGAAAAAGAAAAATCACCTGAGAAGTAGAACTAAAAGAGGACAAGG CTCCTTCTCTAAAGACAAAACCTCACACCCGAAAAAC	report	0.826	0.769	0.703	23	97	0.000

Slika P1. Sumarni rezultati TAREAN analize 400 000 očitavanja dobivenih metodom sekvenciranja nove generacije genoma vrste *Tribolium castaneum*. Prikazani su izlazni podaci za potencijalne satelitne sekvence svrstane u kategorije visoke (*high confidence*) ili niske (*low confidence*) vjerovatnosti.

Prilog 2.

Tablica P2. Prikaz duljina kloniranih odsječaka i duljina ekstrahiranih monomera satDNA TCst154 vrste *Tribolium castaneum*.

naziv kloniranog odsječka	duljina kloniranog odsječka (pb)	naziv monomera ekstrahiranih iz kloniranih odsječaka	duljina ekstrahiranih monomera (pb)
TCst154cl 1	255	TCst154cl 1	155
TCst154cl 2	253	TCst154cl 2	154
TCst154cl 5	377	TCst154cl 5a	155
		TCst154cl 5b	154
TCst154cl 7	408	TCst154cl 7a	154
		TCst154cl 7b	155
TCst154cl 8	209	TCst154cl 8	154
TCst154cl 10	408	TCst154cl 10a	154
		TCst154cl 10b	154
TCst154cl 11	408	TCst154cl 11a	154
		TCst154cl 11b	154
TCst154cl 15	377	TCst154cl 15a	155
		TCst154cl 15b	154
TCst154cl 16	256	TCst154cl 16	154
TCst154cl 18	256	TCst154cl 18	154
TCst154cl 19	254	TCst154cl 19	154
TCst154cl 20	245	TCst154cl 20	155
TCst154cl 21	407	TCst154cl 21a	153
		TCst154cl 21b	154
TCst154cl 22	401	TCst154cl 22a	154
		TCst154cl 22b	147
TCst154cl 24	403	TCst154cl 24a	154
		TCst154cl 24b	148
TCst154cl 25	255	TCst154cl 25	156
TCst154cl 27	254	TCst154cl 27	154
TCst154cl 28	225	TCst154cl 28	154
TCst154cl 30	386	TCst154cl 30a	154
		TCst154cl 30b	154
TCst154cl 32	254	TCst154cl 32	154
TCst154cl 33	254	TCst154cl 33	156
TCst154cl 34	252	TCst154cl 34	153

Prilog 3.

Tablica P3. Prikaz duljina kloniranih odsječaka i duljina ekstrahiranih monomera satDNA TCst266 vrste *Tribolium castaneum*.

naziv kloniranog odsječaka	duljina kloniranog odsječaka (pb)	naziv monomera ekstrahiranih iz kloniranih odsječaka	duljina ekstrahiranih monomera (pb)
TCst266cl 4	674	TCst266cl 4a	266
		TCST266cl 4b	265
TCst266cl 5	675	TCst266cl 5a	266
		TCst266cl 5b	266
TCst266cl 13	409	TCst266cl 13	266
TCst266cl 18	675	TCst266cl 18a	266
		TCst266cl 18b	266
TCst266cl 21	674	TCst266cl 21a	266
		TCst266cl 21b	265
TCst266cl 24	650	TCst266cl 24a	266
		TCst266cl 24b	266
TCst266cl 26	408	TCst266cl 26	265
TCst266cl 27	408	TCst266cl 27	265
TCst266cl 29	409	TCst266cl 29	266
TCst266cl 31	408	TCst266cl 31	265
TCst266cl 32	409	TCst266cl 32	266
TCst266cl 33	431	TCst266cl 33	288
TCst266cl 34	409	TCst266cl 34	266
TCst266cl 37	408	TCst266cl 37	265
TCst266cl 38	375	TCst266cl 38	266
TCst266cl 39	375	TCst266cl 39	266
TCst266cl 40	407	TCst266cl 40	266
TCst266cl 41	397	TCst266cl 41	266
TCst266cl 42	409	TCst266cl 42	266
TCst266cl 43	408	TCst266cl 43	265
TCst266cl 44	407	TCst266cl 44	266
TCst266cl 47	409	TCst266cl 47	266
TCst266cl 48	409	TCst266cl 48	266
TCst266cl 49	408	TCst266cl 49	266
TCst266cl 50	409	TCst266cl 50	266
TCst266cl 52	408	TCst266cl 52	265
TCst266cl 54	409	TCst266cl 54	265
TCst266cl 55	390	TCst266cl 55	266

Prilog 4.

Tablica P4. Prikaz duljina kloniranih odsječaka i duljina ekstrahiranih monomera satDNA TCst309 vrste *Tribolium castaneum*.

naziv kloniranog odsječka	duljina kloniranog odsječka (pb)	naziv monomera ekstrahiranih iz kloniranih odsječaka	duljina ekstrahiranih monomera (pb)
TCst309cl 1	726	TCst309cl 1a	309
		TCst309cl 1b	310
TCst309cl 2	417	TCst309cl 2	310
TCst309cl 3	725	TCst309cl 3a	310
		TCst309cl 3b	308
TCst309cl 5	728	TCst309cl 5a	312
		TCst309cl 5b	309
TCst309cl 6	416	TCst309cl 6	309
TCst309cl 7	416	TCst309cl 7	309
TCst309cl 9	729	TCst309cl 9a	311
		TCst309cl 9b	309
TCst309cl 10	416	TCst309cl 10	309
TCst309cl 11	726	TCst309cl 11a	310
		TCst309cl 11b	309
TCst309cl 12	724	TCst309cl 12a	309
		TCst309cl 12b	310
TCst309cl 14	728	TCst309cl 14a	311
		TCst309cl 14b	310
TCst309cl 16	727	TCst309cl 16a	309
		TCst309cl 16b	312
TCst309cl 17	416	TCst309cl 17	308
TCst309cl 18	727	TCst309cl 18a	311
		TCst309cl 18b	309
TCst309cl 19	726	TCst309cl 19a	310
		TCst309cl 19b	309
TCst309cl 21	417	TCst309cl 21	310
TCst309cl 22	726	TCst309cl 22a	310
		TCst309cl 22b	309
TCst309cl 23	727	TCst309cl 23a	309
		TCst309cl 23b	311
TCst309cl 24	727	TCst309cl 24a	310
		TCst309cl 24b	310
TCst309cl 25	726	TCst309cl 25a	311
		TCst309cl 25b	309
TCst309cl 26	417	TCst309cl 26	310
TCst309cl 28	727	TCst309cl 28a	311
		TCst309cl 28b	309
TCst309cl 29	725	TCst309cl 29a	309
		TCst309cl 29b	310
TCst309cl 30	727	TCst309cl 30a	311
		TCst309cl 30b	309
TCst309cl 31	417	TCst309cl 31	310
TCst309cl 32	418	TCst309cl 32	311
TCst309cl 33	416	TCst309cl 33	309
TCst309cl 34	416	TCst309cl 34	309
TCst309cl 35	417	TCst309cl 35	310

Prilog 5.

```
10 20 30 40 50 60 70 80 90 100
GG695015c115 TCAAAATTTATAAACCAATACAACCTTCCAAACCCTTCCAGGGCTCGATTTTTTTCA-AACTTTTTATATGTAGAGTATTAATGCTCTTCGACCTG 99
GG695015c116 .....C.....A.....T.....G..... 99
TCst154c121a .....C..... 98
TCst154c15a .....C.....A 99
TCst154c115a .....C.....A 99
GG69508c13 .....A.....C.....A..... 93
TCst154c116 .....C.....C..... 99
TCst154c118 .....C..... 99
GG695015c18 .....C.....A.....C.....T..... 99
TCst154c130b .....T.....A.....C.....T..... 99
GG695015c138 .....T.....C.....T..... 99
GG695015c124 .....T.....C.....A.....T..... 99
GG695015c125 .....C.....A.....G..... 99
GG695015c111 .....T.....C.....A.....T.....C.....A..... 93
GG695015c112 .....T.....C.....A.....T.....C.....A..... 93
GG695015c117 .....C.....T.....C.....A.....T.....C.....A..... 93
GG695015c136 .....C.....T.....C.....A.....T.....C..... 93
GG695015c134 .....T.....C.....A.....T.....C.....T..... 93
GG695015c113 .....T.....C.....A.....C..... 93
TCst154c124b .....C.....A.....C..... 93
GG695015c137 .....C.....A..... 93
GG695015c119 .....C.....A..... 99
GG695015c135 .....C.....A..... 99
TCst154c130a .....C.....A..... 99
GG695015c19 .....A.....C.....ATC..... 99
GG695015c139 .....G.....C..... 99
TCst154c124a .....G.....C.....T.....T..... 99
TCst154c132 .....G.....C..... 99
TCst154c122a .....G.....A.....C..... 99
GG695015c120 .....C.....C.....C..... 93
GG695015c121 .....C.....C.....C..... 93
GG695015c118 .....C.....C.....C..... 93
GG695015c123 .....C.....C.....C..... 93
GG695015c127 .....C.....C.....C..... 93
GG695015c126 .....C.....C.....C..... 93
GG695015c122 .....A.....C.....C.....C..... 93
TCst154c125 .....C.....T.....C.....C.....T..... 99
GG695015c16 .....C.....C.....C.....C..... 93
GG695015c133 .....C.....A.....C.....C.....T..... 99
GG69508c11 .....T.....C.....C.....C..... 99
GG69508c12 .....C.....C.....C..... 99
GG695015c15 .....A.....T.....C..... 88
GG695015c130 .....A.....T.....C..... 88
GG695015c11 .....A.....T.....C..... 88
GG695015c128 .....T.....C.....A.....A..... 99
GG695015c17 .....C.....C.....A..... 99
GG695015c14 .....C.....C.....C..... 99
GG695015c129 .....C.....C.....C..... 99
GG695015c132 .....G.....C.....C..... 99
GG695015c13 .....C.....C.....C..... 93
GG695015c131 .....C.....C.....C..... 93
GG695015c114 .....C.....C.....C..... 99
GG695015c12 .....C.....C.....C..... 99
GG695047c12 .....T.....C.....C.....T.....A.....A.....A.....T..... 99
GG695047c18 .....G.....C.....C.....T.....A.....A.....A.....T..... 99
GG695047c122 .....G.....C.....C.....T.....A.....A.....A.....T..... 99
GG695047c125 .....G.....C.....C.....TT.....C.....A.....A.....T..... 99
GG695047c127 .....G.....C.....C.....TT.....A.....A.....T..... 99
GG695047c16 .....G.....A.....C.....A.....T.....TT.....AC.....A.....T..... 99
GG695047c110 .....G.....C.....C.....TT.....A.....CA.....A.....T..... 99
GG695047c118 .....G.....C.....C.....TT.....A.....CA.....A.....T..... 99
GG695047c112 .....G.....A.....C.....TT.....AG.....CA.....A.....T..... 99
GG695047c120 .....G.....A.....C.....TT.....AG.....CA.....A.....T..... 99
GG695047c115 .....G.....C.....C.....TT.....A.....CA.....A.....T..... 99
GG695047c116 .....G.....C.....C.....TT.....A.....CA.....A.....T..... 99
GG695047c114 .....G.....C.....C.....TT.....A.....CA.....A.....CT..... 99
TCst154c120 .....G.....C.....C.....TTT.....CA.....A.....T..... 100
GG695047c117 .....G.....C.....C.....TT.....A.....CA.....A.....T..... 99
TCst154c133 .....G.....C.....C.....T.....A.....CT.....A.....A.....T..... 100
TCst154c17b .....G.....C.....C.....TC.....A.....A.....A.....T..... 100
TCst154c110b .....G.....C.....C.....TC.....A.....A.....A.....T..... 99
TCst154c11 .....G.....A.....C.....C.....T.....A.....CT.....A.....A.....T.....A..... 100
TCst154c18 .....G.....A.....G.....C.....A.....AT.....T.....CA.....A.....A.....T.....A..... 99
GG695047c123 .....T.....G.....C.....C.....A.....AT.....T.....CA.....A.....A.....T..... 100
DS497816.1c16 .....G.....C.....C.....G.....T.....T.....T..... 97
DS497816.1c12 .....C.....C.....A.....C.....T..... 99
DS497816.1c17 .....C.....C.....A.....T.....T..... 98
TCst154c111b .....C.....A.....C.....A..... 99
TCst154c122b .....C.....C.....T.....C..... 93
TCst154c128 .....T.....C.....TC.....T.....G.....A..... 99
GG695047c13 .....T.....C.....TC.....T.....A.....G.....A..... 99
GG695047c128 .....TC.....TTA.....A.....G.....A..... 100
GG695047c15 .....C.....A.....T.....A.....G.....A..... 99
GG695047c126 .....C.....T.....A.....G.....A..... 99
GG695047c17 .....C.....A.....C.....T.....A.....G.....A.....T..... 99
TCst154c17a .....C.....A.....C.....T.....A.....G.....A.....T..... 99
GG695047c113 .....G.....C.....C.....T.....A.....G.....A..... 99
GG695047c111 .....G.....C.....C.....T.....A.....G.....A..... 99
GG695047c119 .....G.....C.....C.....T.....A.....G.....A..... 99
GG695047c124 .....G.....T.....C.....C.....T.....A.....G.....A.....T..... 99
TCst154c110a .....G.....C.....C.....T.....A.....T.....G.....T.....T.....A..... 99
TCst154c119 .....C.....C.....C.....T.....A.....T.....G.....T.....T.....A..... 99
TCst154c12 .....G.....C.....C.....T.....A.....T.....G.....T.....A..... 99
TCst154c127 .....G.....T.....C.....C.....T.....T.....G.....T.....A.....G..... 99
TCst154c15b .....C.....C.....C.....T.....A.....G.....A..... 99
```

TCst154c115bC.....C.....T-A.....G.....A..... 99
GG695407c19C.....C.....T-A.....G.....A..... 99
GG695408c14C.....A.....T-A.....G.....G.....A..... 99
DS497816.1c13 T.....C.....C.....T-A.....G.....T..... 99
DS497816.1c14C.....T-A.....G.....T..... 99
DS497816.1c15A.....A.....C.....C.....A-A.....C.....G.....T..... 99
TCst154c134A.....C.....C.....C.....A.....G.....A..... 98
GG695407c11G.....G.....C.....T-A.....G.....A..... 99
TCst154c121bG.....C.....C.....T-A.....G.....G..... 99
GG695407c14 .A.....G.....A.....C.....T-A.....G.....A..... 99
TCst154c111aG.....A.....C.....T-A.....G.....G..... 99
GG695015c110T.C.....A.....T.....CGC..... 98
DS497816.1c11C..T.....C..C..TT..T.....G..A.....C..... 99

110 120 130 140 150
GG695015c115 CCGTAACAAGAAAGTGGCACATTGTCGATAACACATAGAAAAAATAAACCAC 155
GG695015c116T..... 154
TCst154c121a .TT.....C..... 153
TCst154c15a 155
TCst154c115a 155
GG695408c13 149
TCst154c116A..... 154
TCst154c118A..... 154
GG695015c18 154
TCst154c130b 154
GG695015c138 154
GG695015c124C..... 154
GG695015c125C..... 154
GG695015c111T..... 149
GG695015c112 149
GG695015c117T..... 149
GG695015c136A..... 149
GG695015c134T..... 149
GG695015c113T..... 148
TCst154c124b 148
GG695015c137 149
GG695015c119 155
GG695015c135 154
TCst154c130a 154
GG695015c19 155
GG695015c139 154
TCst154c124a 154
TCst154c132 154
TCst154c122aG..... 154
GG695015c120T..... 149
GG695015c121T..... 149
GG695015c118T..... 149
GG695015c123T.....C..... 148
GG695015c127T.....C..... 148
GG695015c126T.....C..... 148
GG695015c122T.....C..... 148
TCst154c125T.....A.....G..... 156
GG695015c16 148
GG695015c133 155
GG695408c11 154
GG695408c12 154
GG695015c15T.....T.....A..... 145
GG695015c130T.....T.....A..... 145
GG695015c11T.....T.....A..... 145
GG695015c128T.....T.....A..... 156
GG695015c17T.....T.....AA..... 157
GG695015c14T.....T.....A..... 156
GG695015c129T.....T.....AA..... 157
GG695015c132T.....T.....A..... 156
GG695015c13T.....T.....T.....A..... 150
GG695015c131T.....T.....T.....A..... 150
GG695015c114T.....T.....AA..... 157
GG695015c12T.....T.....A..... 156
GG695407c12 A.T.....C..... 154
GG695407c18 A.T.....C..... 154
GG695407c122T..... 154
GG695407c125T.....C.....A..... 154
GG695407c127T.....A.....C..... 154
GG695407c16 T.T.....C..... 154
GG695407c110T.....C..... 154
GG695407c118T.....C..... 154
GG695407c112T..... 154
GG695407c120T..... 154
GG695407c115T..... 154
GG695407c116T..... 154
GG695407c114T..... 154
TCst154c120T.....C..... 155
GG695407c117T..... 154
TCst154c133TT.....C.....G..... 156
TCst154c17bT..... 155
TCst154c110b A.T.....C..... 154
TCst154c11 155
TCst154c18 154
GG695407c123 155
DS497816.1c16 T.....AG..... 152
DS497816.1c12T.....T.T.....G..... 154
DS497816.1c17T..... 153
TCst154c111bG.A.....T..... 154
TCst154c122bG.A..... 147
TCst154c128T.A.....G.A..... 154
GG695407c13T.....C.....T..... 154
GG695407c128T.....C.....T.....A..... 155
GG695407c15T.....C.....T..... 154
GG695407c126GT.....C.....T..... 154
GG695407c17T.....C.....T..... 154
TCst154c17aT.....C.....T..... 154
GG695407c113T.....G.C.....T..... 154

```

GG695407c121 ..T.....G.C...T.....- 154
GG695407c111 ..T.....C...T.....- 154
GG695407c119 ..T.....C...T.....- 154
GG695407c124 ..T.....C...T.....- 154
TCst154c110a A..T.....C.....A.....- 154
TCst154c119 ..TT.....C.....- 154
TCst154c12 ..TT.....C.....- 154
TCst154c127 ..T.....C...C.....- 154
TCst154c15b ..TT.....C.....- 154
TCst154c115b ..TT.....C.....- 154
GG695407c19 ..T.....C...T.....- 154
GG694508c14 ..T.....C.....- 153
DS497816.1c13 ..T.....C.....- 154
DS497816.1c14 ..T.....G.....ACC----- 149
DS497816.1c15 ..T.....G.....G...A.- 154
TCst154c134 ..T.....G.....- 153
GG695407c11 ..T...A.G.....- 154
TCst154c121b ..T.....C.....- 154
GG695407c14 ..T.....C.....- 154
TCst154c111a ..T.....C.....- 154
GG695015c110 ..T...AAA.....- 153
DS497816.1c11 ..T...A...A...G.T.....C...GT 154

```

Slika P5. Nukleotidni sljedovi monomera satDNA TCst154 ekstrahiranih iz izlaznih podataka sekvenciranog genoma vrste *Tribolium castaneum* (Tcas5.2) i monomera ekstrahiranih iz PCR klonova analiziranih u ovom radu. Slovimu su označene promjene unutar pojedinih uzoraka u odnosu na prvi nasumično odabrani monomer, točke predstavljaju podudarnost s prvom sekvencom, a crtice označavaju delecije unutar pojedinih odsječaka.

DS497837.1c18	..T.....C.G.A.....T.....T.....	175
TCst266c121b	..T.....C.G.....T.....T.....	174
TCst266c124a	..T.....C.G.....T.....T.....	175
TCst266c14b	..T.....C.G.A.....T.....TA.T.....	175
TCst266c147G.....T.....T.....	175
DS497837.1c15T.....T.....C.AA.....	175
TCst266c149C.....A.....T.....TA.T.....	175
TCst266c138T.....T.....TA.T.....	175
TCst266c139T.....T.....TA.T.....	175
TCst266c118bT.....T.....T.....	175
TCst266c154T.....T.....T.....	175
TCst266c118aT.....T.....T.....	175
GG694367c18G.....T.....T.....C.....C.AATCCCTGATATGCTGGAGCATA.....	197
TCst266c133T.G.....T.....T.....C.....C.AATCCCTGATATGCTGGAGCATA.....	197
DS497894.1c15G.....T.....T.....T.....C.....C.AATCCCTGATATGCTGGAGCATA.....	197
GG694187c13T.....T.....T.....T.....	175
GG694187c18T.....T.....T.....T.....C.....	175
GG694187c14	.A.....A.....T.....T.....T.....	175
GG694187c19A.....T.....T.....T.....	175
DS497837.1c17T.....T.....T.....T.....	175
DS497837.1c12T.....T.....T.....T.....	175
TCst266c15aT.....T.....T.....T.....	175
TCst266c132T.....T.....T.....T.....	175
DS497837.1c112T.....T.....T.....T.A.A.....	175
DS497837.1c16T.....T.....T.....T.....	175
GG694367c19A.....C.....T.....T.....T.....	175
TCst266c140A.....C.....T.....T.....T.....	175
TCst266c144A.....C.....T.....T.....T.....	175
TCst266c113A.....C.....T.....T.....T.....	175
DS497894.1c13A.....C.....T.....T.....T.....	175
GG694367c14A.....C.....T.....T.....T.....	175
GG694367c12A.....T.....T.....T.....A.....G.....	175
GG694367c16A.....T.....T.....T.....A.....G.....	175
GG694187c112A.....T.....T.....T.....T.....	174
DS497837.1c19A.....C.....T.AG.....TA.....T.....C.....AA.....	175
TCst266c142A.....C.....T.AG.....TA.....T.....C.....AA.....	175
TCst266c129A.....C.....T.AG.....TA.....T.....C.....AA.....	175
TCst266c15bA.....C.....T.AG.....TA.....T.....C.....AA.....	175
TCst266c14aA.....C.....T.AG.....TA.....T.....C.....AA.....	175
TCst266c135A.....C.....T.AG.....TA.....T.....A.....C.....AA.....	175
TCst266c148A.....T.....T.AG.....TA.....T.....C.....AA.....	175
TCst266c150A.....T.....T.AG.....TA.....T.....C.....AA.....	175
GG694187c11A.....T.....T.AG.....TA.....T.....C.....AA.....	172
DS497837.1c111A.....T.AG.....TA.....T.....C.....A.....T.....	175
TCst266c155A.....T.AG.....TA.....T.....C.....A.....T.....	175
TCst266c124bA.....T.AG.....TA.....T.....C.....A.....T.....	175
TCst266c121aA.....T.AG.....TA.....T.....C.....A.....T.....	175
GG694187c115A.....T.AG.....GTA.....T.....C.....A.....G.....	175
TCst266c134A.....T.AG.....TA.....T.....C.....A.....T.....	175
TCst266c137A.....C.....T.AG.....TA.....T.....C.....AA.....C.....	174
TCst266c143A.....C.....T.AG.....TA.....T.....C.....AA.....C.....	174
GG694367c13A.....T.AG.....T.....T.....C.....A.....T.....	175
GG694367c17A.....T.AG.....T.....T.....C.....A.....T.....	175
GG694187c113A.....T.AG.....TA.....G.....T.....C.....A.....	175
GG694187c16C.....TA.TT.....A.TG.....T.....C.TGC.....C.....A.....	175
GG694187c118C.....TA.TC.....TG.....G.....C.TGC.....C.....A.....	175
GG694187c111C.....TA.TC.....TG.....T.G.....C.TGC.....C.....	175
GG694187c119A.....TA.....TG.....T.....T.....C.....	175
DS497837.1c113A.....T.AG.....TA.....C.T.....C.....A.....T.....	175
GG694367c110A.....T.AG.....TA.....C.T.....C.....A.....T.....	175
GG694187c120T.....TA.....TG.....T.....C.TGC.....C.....	175
GG694187c114	.C.....G.....TA.....T.AG.....T.....C.T.C.....C.T.....T.....	175
TCst266c127	.C.....G.....TA.....T.AG.....T.....C.T.C.....C.T.....T.....	174
DS497837.1c13	C.....G.....TA.....T.G.A.T.....C.T.C.....T.C.T.....C.....T.....	175
GG694187c15	.C.....G.....TA.....T.G.....T.....C.T.C.....C.T.....T.....	175
GG694187c110	.C.....G.....TA.....T.G.....T.....C.T.C.T.....C.T.....T.....	175
DS497894.1c14TG.T.....G.....T.....C.T.....T.....T.....	175
GG694367c15TG.T.....G.....T.....C.T.....T.....T.....	175
GG694367c11TG.....T.....TG.....T.....C.T.C.....T.....T.....	175
GG694187c116TGG.....G.....T.....C.T.C.....T.....T.....	175
DS497894.1c12T.....TCA.....TG.A.T.....C.TGC.....T.....C.....A.....	175
GG694187c117A.....TA.....C.....T.G.....T.....C.T.....T.....T.....	175
DS497837.1c11	.C.....G.....TA.....T.G.....T.....C.T.C.A.....C.T.....T.....	175
GG694187c121G.A.....TA.....T.....A.....T.....C.T.C.A.....C.T.....A.....T.A.....	175
DS497894.1c11A.G.....TA.....AG.....T.....C.A.....G.....C.....A.....G.....C.....	175

	210	220	230	240	250	260	270	280	290		
TCst266c131	CAATCAAA	TTTGACATAT	TCAAATAA	AGGGGGT	AGTGGGAAT	TAAAAATG	TCAAATTT	CAAAAAATA	AAAAAT	CAACCAAATTAATATAT	265
TCst266c152	265
TCst266c126	265
DS497837.1c110	266
DS497837.1c14	265
GG694187c17	266
GG694187c12	266
DS497894.1c16	266
TCst266c141	266
DS497837.1c18	266
TCst266c121b	265
TCst266c124a	266
TCst266c14b	265
TCst266c147	266
DS497837.1c15	266
TCst266c149	266
TCst266c138	266
TCst266c139	266
TCst266c118b	266
TCst266c154	266
TCst266c118a	266
GG694367c18	288
TCst266c133	288
DS497894.1c15	288
GG694187c13	266

```

GG694187c18 .....C.....-...T...T..... 266
GG694187c14 .....G.....C.....TT...G..... 266
GG694187c19 .....C.....A.....T...TT...G..... 267
DS497837.1c17 .....C.....A.....T...G..... 266
DS497837.1c12 .....T.....A.....T...G..... 266
TCst266c15a .....T.....T.....G..... 266
TCst266c132 .....T.....T..... 266
DS497837.1c112 .....T.....T..... 266
DS497837.1c16 .....T.....T..... 266
GG694367c19 .....A.....T..... 266
TCst266c140 .....T.....T...T... 266
TCst266c144 .....T.....T...T... 266
TCst266c113 .....T.....T..... 266
DS497894.1c13 .....C.....T.....G..... 266
GG694367c14 .....C.....T.....G..... 266
GG694367c12 .....C.....T...T... 266
GG694367c16 .....C.....T...T... 266
GG694187c112 .....T.....AAA... 266
DS497837.1c19 .....T.....T... 266
TCst266c142 .....T.....T... 266
TCst266c129 .....T.....T... 266
TCst266c15b .....C.....T... 266
TCst266c14a .....T.....T... 266
TCst266c135 .....T.....T... 266
TCst266c148 .....T.....T... 266
TCst266c150 .....T.....T... 266
GG694187c11 .....T.....T... 263
DS497837.1c111 .....T.....T... 266
TCst266c155 .....T.....T... 266
TCst266c124b .....T.....T... 266
TCst266c121a .....T.....T... 266
GG694187c115 .....G.....T.A.....T.T.....C... 266
TCst266c134 .....T.....T... 266
TCst266c137 .....T.....T... 265
TCst266c143 .....T.....T... 265
GG694367c13 .....T.....T... 266
GG694367c17 .....G.....T.....T... 266
GG694187c113 .....G.....G.....T.....T... 266
GG694187c16 .....T.....G.....A.....T.T...G... 266
GG694187c118 .....T.....G.....A.....T.....A...T.T...G... 266
GG694187c111 .....T.....G.....T.....T... 266
GG694187c119 .....GG.....A.....T.T.T...G...C.A... 266
DS497837.1c113 .....T.....T... 266
GG694367c110 .....T.....T... 266
GG694187c120 .....G.G.....C...T.T...T.TC...A...A...T.T.T... 265
GG694187c114 .....G.G.....T.....T... 266
TCst266c127 .....G.G.....T.....T... 265
DS497837.1c13 .....G.G.....T.T... 266
GG694187c15 .....G.....T...T... 266
GG694187c110 .....G.....T...T... 266
DS497894.1c14 T...G.G.....T...T.G.....A.....T...T... 266
GG694367c15 T...G.G.....T...T.G.....A.....T...T... 266
GG694367c11 T...G.G.....T...T.G.....A.....T...T.T...T... 266
GG694187c116 TG...CG.G.....T.G.....A.....A.....T.T.T...T...T... 266
DS497894.1c12 .....T.G.....T.G.....A.....T.T...T...C... 266
GG694187c117 T...G.G.....TTG.....A.....T...T... 266
DS497837.1c11 T...G.G.....T.G.....C...A.....T.T...T...T... 266
GG694187c121 T...G.G.....T...T.G.....A.....TTTT.T...GTTGTT.T...T-TTGT.TG...TTC... 271
DS497894.1c11 T...G.G.....G...T.G...A.....G...CG.TCG.ATCG.TG...-...GTTTTTGCAC.G.GGCA.TC.TTT...CG...- 266

```

Slika P6. Nukleotidni sljedovi monomera satelitne DNA TCst266 ekstrahiranih iz izlaznih podataka sekvenciranog genoma vrste *Tribolium castaneum* (Tcas5.2) i PCR monomera ekstrahiranih iz klonova analiziranih u ovom radu. Slovimama su označene promjene unutar pojedinih uzoraka u odnosu na prvi nasumično odabrani monomer, točke predstavljaju podudarnost s prvom sekvencom, a crtice označavaju delecije unutar pojedinih odsječaka.

Prilog 7.

	10	20	30	40	50	60	70	80	90	100
TCst309c17	TTACAAAACGCAATAAAGCAGAAGTGAGTATATGTTGCCACCAAGTCTTTTTGTTTTATGAAGTGATTTTAGCCTGTAAAGCGTTGAAAATGTCGAGGA									100
TCst309c110										100
TCst309c133										100
ChLG7c12.24										100
TCst309c11a										100
ChLG7c12.20										100
ChLG7c12.5										100
TCst309c122b										100
TCst309c16										100
ChLG7c11.4										100
ChLG7c12.10										100
TCst309c118b										100
TCst309c116a										100
ChLG7c12.12										100
ChLG7c11.8										100
TCst309c121										100
ChLG7c11.5										100
TCst309c12										100
TCst309c13a										100
ChLG7c12.7										100
ChLG7c11.1										100
ChLG7c12.17										100
TCst309c114a										100
TCst309c116b										100
TCst309c123b										100
ChLG7c11.11										100
ChLG7c12.23				A						100
TCst309c114b				A					G	100
TCst309c112b				A						100
ChLG7c11.2				A						100
TCst309c15b										100
TCst309c19b										100
TCst309c131										100
ChLG7c12.4										100
ChLG7c12.16				A						100
TCst309c118a				A						100
ChLG7c12.1				A						99
TCst309c134				A						100
TCst309c129a				A						100
ChLG7c11.10				A						100
ChLG7c12.9				A						100
ChLG7c12.13				A						100
TCst309c112a				A						100
ChLG7c12.14				A						100
ChLG7c11.3				A						100
ChLG7c11.7				A						100
TCst309c122a			G	A						100
ChLG7c12.2				A						100
TCst309c11b				A						100
ChLG7c12.21				A						100
TCst309c111a				A						100
ChLG7c12.18				A						100
TCst309c135				A						100
TCst309c124a				A						100
ChLG7c12.19				A						100
TCst309c126				A						100
ChLG7c12.6				A						99
TCst309c128a				A						100
ChLG7c12.8				A						100
TCst309c119a				A						100
ChLG7c11.9				A						100
ChLG7c12.22				A						100
ChLG7c12.11				A						100
TCst309c15a				A						100
TCst309c117				A						100
TCst309c128b				A						100
TCst309c130a			G	A						100
TCst309c119b				A						100
TCst309c13b				A				G		100
TCst309c125b				A						100
TCst309c123a				A					G	100
TCst309c130b				A					G	100
ChLG7c12.3				A					G	100
TCst309c124b								G	G	100
ChLG7c12.15										100
TCst309c125a										100
TCst309c129b				A						99
TCst309c132				A						100
TCst309c19a										100
ChLG7c11.6										100
TCst309c111b									C	100
ChLG3c11										100
DS497883.1c11A										100

	110	120	130	140	150	160	170	180	190	200
TCst309c17	AACCTGTTCACCTAAAGCATGGGAAACCAAGGTATTTATATTTACTTGAAGTGAATTTTGTGATAATTTCCACAAAAAAA--GTGTCGTTTCAGTTCT									197
TCst309c110										197
TCst309c133										197
ChLG7c12.24										197
TCst309c11a										197
ChLG7c12.20										197
ChLG7c12.5										197
TCst309c122b										197
TCst309c16										197
ChLG7c11.4										197

ChLG7c12.10G..... 197
 TCst309c118bG..... 197
 TCst309c116aG..... 197
 ChLG7c12.12G..... 197
 ChLG7c11.8G..... 197
 TCst309c121A.....G..... 198
 ChLG7c11.5A.....G..... 198
 TCst309c12A.....G..... 198
 TCst309c13aA.....G..... 198
 ChLG7c12.7AA.....G..... 199
 ChLG7c11.1AA.....G..... 199
 ChLG7c12.17AA.....G..... 199
 TCst309c114aAA.....G..... 199
 TCst309c116bAAA.....G..... 200
 TCst309c123bAA.....CG..... 199
 ChLG7c11.11AA.....G..... 199
 ChLG7c12.23A.....G..... 198
 TCst309c114bC.....A.....G..... 198
 TCst309c112bA.....G..... 198
 ChLG7c11.2A.....AA.....G..... 199
 TCst309c15bG..... 197
 TCst309c19bG..... 197
 TCst309c131A.....A.....G..... 198
 ChLG7c12.4A.....G..... 198
 ChLG7c12.16AC.....G..... 199
 TCst309c118aAC.....G..... 199
 ChLG7c12.1C.....G..... 196
 TCst309c134G..... 197
 TCst309c129aG..... 197
 ChLG7c11.10G..... 197
 ChLG7c12.9G..... 197
 ChLG7c12.13G..... 197
 TCst309c112aG..... 197
 ChLG7c12.14G..... 197
 ChLG7c11.3A.....G..... 198
 ChLG7c11.7A.....G..... 198
 TCst309c122aA.....G..... 198
 ChLG7c12.2A.....G..... 198
 TCst309c11bA.....G..... 198
 ChLG7c12.21A.....G..... 198
 TCst309c111aA.....G..... 198
 ChLG7c12.18A.....G..... 198
 TCst309c135A.....G..... 198
 TCst309c124aA.....G..... 198
 ChLG7c12.19A.....G..... 198
 TCst309c126A.....G..... 198
 ChLG7c12.6AA.....G..... 198
 TCst309c128aAA.....G..... 199
 ChLG7c12.8AA.....G..... 199
 TCst309c119aAA.....G..... 199
 ChLG7c11.9AAA.....G..... 200
 ChLG7c12.22AAA.....G..... 200
 ChLG7c12.11AAA.....G..... 200
 TCst309c15aC.....AAA.....G..... 200
 TCst309c117G..... 196
 TCst309c128bG..... 197
 TCst309c130aAA.....G..... 199
 TCst309c119bG.....G.....G..... 197
 TCst309c13bG..... 197
 TCst309c125bG..... 197
 TCst309c123aG..... 197
 TCst309c130bG..... 197
 ChLG7c12.3G..... 197
 TCst309c124bA.....G..... 198
 ChLG7c12.15AA.....GC..... 199
 TCst309c125aAA.....GC..... 199
 TCst309c129bAA.....GC..... 198
 TCst309c132AA.....GC..... 199
 TCst309c19aAA.....G..... 199
 ChLG7c11.6G..... 196
 TCst309c111bA.....G..... 198
 ChLG3c11A.....A.....A.....G..... 198
 DS497883.1c11A.....AA.....G..... 199

210 220 230 240 250 260 270 280 290 300

TCst309c17AAATATTTTACAAATAATTTAACTTTTTTAGGCTACATCTTCCAAGAGGAAATACTAAAGTACCACCTGCTGAACCCAGTGAAGTTACAAACGGAAACAG 297
 TCst309c110 297
 TCst309c133 297
 ChLG7c12.24 297
 TCst309c11a 297
 ChLG7c12.20 297
 ChLG7c12.5C..... 297
 TCst309c122b 297
 TCst309c16 297
 ChLG7c11.4 297
 ChLG7c12.10 297
 TCst309c118b 297
 TCst309c116a 297
 ChLG7c12.12 297
 ChLG7c11.8 297
 TCst309c121 298
 ChLG7c11.5 298
 TCst309c12 298
 TCst309c13a 298
 ChLG7c12.7 299
 ChLG7c11.1 299
 ChLG7c12.17 299
 TCst309c114a 299
 TCst309c116b 300
 TCst309c123b 299
 ChLG7c11.11G..... 299
 ChLG7c12.23 298

```

TCst309c114b ..... 298
TCst309c112b ..... 298
ChLG7c11.2 ..... 299
TCst309c15b .....A..... 297
TCst309c19b .....C.....A..... 297
TCst309c131 .....A..... 298
ChLG7c12.4 ..... 298
ChLG7c12.16 ..... 299
TCst309c118a ..... 299
ChLG7c12.1 ..... 296
TCst309c134 ..... 297
TCst309c129a ..... 297
ChLG7c11.10 ..... 297
ChLG7c12.9 ..... 297
ChLG7c12.13 ..... 297
TCst309c112a ..... 297
ChLG7c12.14 ..... 297
ChLG7c11.3 ..... 298
ChLG7c11.7 ..... 298
TCst309c122a ..... 298
ChLG7c12.2 ..... 298
TCst309c11b ..... 298
ChLG7c12.21 ..... 298
TCst309c111a ..... 298
ChLG7c12.18 ..... 298
TCst309c135 ..... 298
TCst309c124a ..... 298
ChLG7c12.19 ..... 298
TCst309c126 ..... 298
ChLG7c12.6 ..... 298
TCst309c128a ..... 299
ChLG7c12.8 ..... 299
TCst309c119a .....T..... 299
ChLG7c11.9 ..... 300
ChLG7c12.22 ..... 300
ChLG7c12.11 ..... 300
TCst309c15a ..... 300
TCst309c117 .....T..... 296
TCst309c128b .....C..... 297
TCst309c130a .....A..... 299
TCst309c119b .....G..... 297
TCst309c13b ..... 297
TCst309c125b ..... 297
TCst309c123a .....C..... 297
TCst309c130b ..... 297
ChLG7c12.3 ..... 297
TCst309c124b .....T..... 298
ChLG7c12.15 ..... 299
TCst309c125a ..... 299
TCst309c129b .....G..... 298
TCst309c132 ..... 299
TCst309c19a ..... 299
ChLG7c11.6 ..... 295
TCst309c111b ..... 297
ChLG3c11 .....A.....GT.A.....G.GA.....AG...A 298
DS497883.1c11 .....GT.A.....G.GA.....AG...A 299

```

```

310
TCst309c17 .....|.....|.. 309
TCst309c110 ATGAAAAAAAAA 309
TCst309c133 ..... 309
ChLG7c12.24 ..... 309
TCst309c11a ..... 309
ChLG7c12.20 ..... 309
ChLG7c12.5 .....- 308
TCst309c122b ..... 309
TCst309c16 ..... 309
ChLG7c11.4 ..... 309
ChLG7c12.10 ..... 309
TCst309c118b ..... 309
TCst309c116a ..... 309
ChLG7c12.12 ..... 309
ChLG7c11.8 ..... 309
TCst309c121 ..... 310
ChLG7c11.5 ..... 310
TCst309c12 ..... 310
TCst309c13a ..... 310
ChLG7c12.7 ..... 311
ChLG7c11.1 ..... 311
ChLG7c12.17 ..... 311
TCst309c114a ..... 311
TCst309c116b ..... 312
TCst309c123b ..... 311
ChLG7c11.11 ..... 311
ChLG7c12.23 ..... 310
TCst309c114b ..... 310
TCst309c112b ..... 310
ChLG7c11.2 ..... 311
TCst309c15b ..... 309
TCst309c19b ..... 309
TCst309c131 ..... 310
ChLG7c12.4 ..... 310
ChLG7c12.16 ..... 311
TCst309c118a ..... 311
ChLG7c12.1 ..... 308
TCst309c134 ..... 309
TCst309c129a ..... 309
ChLG7c11.10 ..... 309
ChLG7c12.9 ..... 309
ChLG7c12.13 ..... 309
TCst309c112a ..... 309
ChLG7c12.14 ..... 309

```

ChLG7c11.3	310
ChLG7c11.7	310
TCst309c122a	310
ChLG7c12.2	310
TCst309c11b	310
ChLG7c12.21	310
TCst309c111a	310
ChLG7c12.18	310
TCst309c135	310
TCst309c124a	310
ChLG7c12.19	310
TCst309c126	310
ChLG7c12.6	310
TCst309c128a	311
ChLG7c12.8	311
TCst309c119a	310
ChLG7c11.9	312
ChLG7c12.22	312
ChLG7c12.11	312
TCst309c15a	312
TCst309c117	308
TCst309c128b	309
TCst309c130a	311
TCst309c119b	309
TCst309c13b	308
TCst309c125bT	309
TCst309c123a	309
TCst309c130b	309
ChLG7c12.3	309
TCst309c124b	310
ChLG7c12.15	311
TCst309c125a	311
TCst309c129b	310
TCst309c132	311
TCst309c19a	311
ChLG7c11.6	307
TCst309c111b	309
ChLG3c11	CC.....G	310
DS497883.1c11	CC.....G	311

Slika P7. Nukleotidni sljedovi monomera satelitne DNA TCst309 ekstrahiranih iz izlaznih podataka sekvenciranog genoma vrste *Tribolium castaneum* (Tcas5.2) i PCR monomera ekstrahiranih iz klonova analiziranih u ovom radu. Slovimama su označene promjene unutar pojedinih uzoraka u odnosu na prvi nasumično odabrani monomer, točke predstavljaju podudarnost s prvom sekvencom, a crtice označavaju delecije unutar pojedinih odsječaka.

ŽIVOTOPIS

OSOBNI PODACI:

Ime: **Mira**
Prezime: **Dombi**
Datum rođenja: 17.10.1994.
Mjesto rođenja: Sombor, Srbija

KONTAKTNI PODACI:

E-mail: mira.dombi94@gmail.com

OBRAZOVANJE:

2013-2016 *Preddiplomski studij:* Odjel za biologiju, Sveučilište Josipa Jurja
Strossmayera u Osijeku, Preddiplomski sveučilišni
studij biologije

2009-20013 *Srednja škola:* II. Gimnazija u Osijeku, Osijek

VJEŠTINE I KOMPETENCIJE:

Strani jezici: Engleski (izvrsno)
Španjolski (dobro)
Računalne vještine: Odlično poznavanje MS Office aplikacija (Word, Excel, PowerPoint)
Osnovno poznavanje programa BioEdit, MEGA