

Prof. dr. sc. Miroslav Plohl
Emeritus Instituta Ruđer Bošković, Zagreb
plohl@irb.hr

00-00

RAZVOJ TEHNIKA SEKVENCIRANJA I ISTRAŽIVANJE SATELITNIH DNK

Sažetak: Sekvence molekule deoksiribonukleinske kiseline (DNK), ponovljene uzastopno u velikom broju kopija u genomima biljaka, životinja i gljiva nazivamo satelitne DNK. Istraživanje genomskog sadržaja i točnog slijeda satelitnih DNK najviše je ograničeno tehnikama sekvenciranja, tj. 'čitanja' molekule DNK. Uzrok tomu je velika sličnost jedinica ponavljanja satelitnih DNK, koja predstavlja poteškoću pri preklapanju sekvenciranih segmenata, njihovog pravilnog postavljanja u odnosu na ostale sekvence DNK u genomu te u konačnici, pri sastavljanju kromosomske i genomske DNK kao neprekinute cjeline. Razvojem tehnika sekvenciranja i specijaliziranih računalnih pristupa dramatično su se mijenjali i koncepti razumijevanja broja, rasporeda, evolucije i uloge satelitnih DNK. Prva sekvenciranja DNK u Hrvatskoj, prije gotovo 40 godina, bila su povezana s istraživanjem satelitnih DNK kukca žutog brašnjara *Tenebrio molitor* i njegovih srodnika. U ovom izlaganju biti će prikazana povezanost napretka u tehnikama sekvenciranja i istraživanjima satelitnih DNK, zahvaljujući kojima se mijenjao naš pogled na ove sekvence, od genomskog balasta do važnih funkcionalnih komponentata.

Ključne riječi: DNK, genom, tehnike sekvenciranja, satelitne DNK, žuti brašnar

Key words: DNA, genome, sequencing techniques, satellite DNA, yellow mealworm

Uvod

Istraživanje strukture gena¹ i genoma² razvijalo se je usporedno s tehnologijom čitanja nukleotidne sekvence molekule deoksiribonuklinske kiseline (DNK)³ i bioinformatičkim alatima koji su potrebni za pohranu i obradu prikupljenih podataka. U istraživanjima genoma poseban problem predstavlja sekvenciranje i obrada sekvenci DNK koje se ponavljaju u velikom broju kopija. Kako bi se moglo razjasniti kakve su to sekvence i zašto njihova analiza predstavlja poteškoću te kako je postupan razvoj specifičnih metoda doveo do sve boljeg razumijevanja njihove organizacije, evolucije i uloga, potrebno je prvo osvrnuti se na sam fenomen biološkog nasljeđivanja.

Najkraće rečeno, molekula DNK sadrži osnovni nacrt kako će neki organizam izgledati i kako će funkcionirati⁴. Cjelokupni genom predstavlja dobro uređen sustav, bez obzira što ulogu pojedinih genomskih komponenti (još uvijek) ne razumijemo u potpunosti. Ili, kao što je još 1968. godine naglasio biolog i filozof znanosti Ludwig von Bertalanffy⁵ u svojoj knjizi *Opća teorija sustava*⁶ [1]: '*...genom kao cjelina proizvodi organizam kao cjelinu. Određeni geni, međutim, u prvom redu određuju smjer razvoja određenih značajki – tj. djeluju kao glavni dijelovi*' (str. 63 u hrvatskom prijevodu). Također naglašava: '*Mendelovi zakoni bili su korijeni genetike, ali – nasljednost, križna razmjena, učinak položaja koda i još mnogo toga – bio je to samo mali dio genetičkog iskustva koje se može opisati klasičnim zakonima*' (str. 166). Ovim promišljanjima, naglasio je kako proučavanje samih gena nije dovoljno da bi se objasnilo funkcioniranje složenih eukariotskih⁷ organizama. Ova, u tom trenutku tek intuitivna pretpostavka, pokazala se točnom. Cjelokupno nasljeđe Bertalanffyjevih ideja i Opća teorija sustava (OTS) revitalizirane su osobito u zadnjim godinama te se primijenjuju i proučavaju u različitim područjima znanosti [2, 3].

U referatu na simpoziju PIFT 2023 [4] opisao sam problem odnosa veličine eukariotskih genoma, broja gena i negenske (nekodirajuće) DNK, često zvane i genomska tamna materija. Veličina genoma značajno se razlikuje među vrstama, a zanimljivo je da ne slijedi evolucijsko stablo, odnosno složenost organizma. Geni zauzimaju iznenađujuće mali udio u genomu, u pravilu < 2%. S obzirom na činjenicu da geni kodiraju funkcionalne molekule koje sudjeluju u različitim strukturama i procesima, dugo vremena bilo je otvoreno pitanje: čemu bi uopće mogao služiti ostatak (odnosno velika većina) genomske DNK? Dio znanstvenika pretpostavljao je da je genomska tamna materija nefunkcionalna te

¹ Gen je segment molekule DNK odgovoran za neku funkciju ili osobinu. Geni u najvećem broju (oko 80%) kodiraju proteine ili bjelančevine, a ostatak kodira molekule ribonukleinske kiseline (RNK) koje su već kao takve funkcionalni elementi staničnih struktura i procesa.

² Genom je sveukupna DNK sadržana u nekom organizmu.

³ Molekulu DNK izgrađuju osnovni elementi ili nukleotidi, skraćeno nazvani A, T, G, i C. Raznolikost u njihovom uzastopnom slijedu osnova je informativnosti ove molekule. Određuje se metodama sekvenciranja DNK.

⁴ Genomska DNK u asocijaciji s proteinima 'pakirana' je u mikroskopskim strukturama koje zovemo kromosomi. Svaki kromosom sadrži jednu cjelovitu molekulu DNK. Genom svake vrste podijeljen je na točno definiran broj kromosoma prepoznatljive morfologije.

⁵ Karl Ludwig von Bertalanffy, Beč, Austrija, 1901. – Buffalo, SAD, 1972.; biolog i filozof širokog obrazovanja: u matematici, teoriji informacija, termodinamici, biokemiji, biofizici, medicini, psihologiji, lingvistici i drugim. Nominaciju za Nobelovu nagradu prekinula je iznenadna smrt 1972. godine. Bertalanffyjeve jednadžbe (1934) koje opisuju rast organizma, populacija, biocenoza, stanica (npr. malignih stanica) koriste se i danas.

⁶ U knjizi *General System Theory: Foundations, Development, Applications* (1968) Ludwig von Bertalanffy donosi kompilaciju svojih prethodnih radova i ideja kojima ukazuje na opće zakonitosti koji povezuju prirodna, tehnička i društvena područja istraživanja. Knjigu su prvi put na hrvatski jezik preveli prof.dr.sc. Zvonko Benčić i prof.dr.sc. Franjo Jović, u izdanju Kiklos – krug knjige, Zagreb, 2022.

⁷ Eukarioti su svi organizmi u čijim je stanicama genom organiziran u strukturi stanične jezgre i kromosoma; ovaj oblik organizacije genoma nemaju bakterije koje se označavaju kao prokarioti.

da predstavlja genomski balast koji je sam sebi svrha, a gomila se kao „sebična“ ili „parazitska DNK“. U najvećem dijelu predstavljaju ju sekvence DNK koje se ponavljaju, neke čak i više stotina tisuća puta. Možemo ih podijeliti u dvije velike skupine. U prvoj su sekvence koje se ponavljaju u uzastopnom slijedu te tako čine duge nizove generičkog naziva satelitne DNK. Genom svake vrste sadrži sebi svojstven komplet satelitnih DNK, koji se već među bliskim srodnicima razlikuje po sadržaju i/ili po broju kopija te time predstavlja evolucijski najvarijabilniju genomsku komponentu. Drugu skupinu čine sekvence DNK koje imaju sposobnost premještanja i ugradnje na nove genomske lokacije, čime nastaju raspršeno ponovljene sekvence DNK. Nazivaju se pokretni genetički elementi, ili transponirajući elementi⁸ [4, 5].

Hipoteza svrsishodnosti negenskih sekvenci DNK u skladu je s pretpostavkama Opće teorije sustava [1] prema kojima uložena materija i energija u biološkim organizmima mora pružati bar neku prednost u neprestanoj evolucijskoj kompeticiji. Upravo razdoblje od približno zadnja dva desetljeća obilježava skokovit napredak u istraživanju genomske tamne materije, zahvaljujući napretku u tehnologijama sekvenciranja DNK i bioinformatičkih analiza [6]. Dobiveni rezultati sve više potvrđuju da sve genomske komponente (kao i organizmi u cjelosti, populacije i cijeli ekosustavi) predstavljaju, sukladno Općoj teoriji sustava, *'veliku organizaciju, strukturiranu hijerarhijski u međudjelujuće sustave koji se ponašaju podjednako bez obzira o čemu se radilo'* što odgovara integrativnom, ili kako je von Bertalanffy nazvao, *'organizmičkom pristupu'* kojim treba promatrati živa bića i sve fenomene vezane uz njih, povezujući prirodna, ali i tehnička i društvena područja znanosti.

U poglavljima koja slijede opisujem kako je razvoj metoda sekvenciranja molekule DNK i specijaliziranih računalnih programa omogućio bolje razumijevanje satelitnih DNK. Određivanje nukleotidne sekvence satelitne DNK za nas ima poseban značaj, jer su upravo jedinice ponavljanja satelitne DNK kukca žutog brašnara *Tenebrio molitor* bile među prvim određenim u Hrvatskoj, 1986. / 1987. godine, na Institutu Ruđer Bošković u Zagrebu. To je prvobitno provedeno na *'home made'* izrađenom laboratorijskom uređaju. Kako je sekvenciranje molekule DNK u svijetu u to vrijeme već bio nezaobilazan alat u svim vrstama genomskih istraživanja, ubrzo su se pojavili specijalizirani servisi u koje je postalo brže i isplativije slati uzorke na sekvenciranje, uz visoku točnost dobivenih rezultata. Uz automatizaciju procesa, značajno povećanje broja istovremeno obrađivanih uzoraka te pad cijena, razvijali su se i novi pristupi, kao što su sekvenciranje sljedeće generacije⁹ i sekvenciranje dugih segmenata DNK¹⁰. Svaki od ovih pomaka u metodologiji otvorio je nove mogućnosti u sveobuhvatnijoj analizi i boljem razumijevanju satelitnih DNK. Posebno je tome doprinio razvoj baza podataka i računalnih programa specijaliziranih za analizu satelitnih DNK [6]. U odnosu na 'ručne' analize pojedinih jedinica ponavljanja najzastupljenijih satelitnih DNK neke vrste, što je bila standardna metoda u svijetu pred 30 i više godina, danas možemo relativno brzo opisati ukupni sadržaj ponovljenih sekvenci DNK u nekoj vrsti, a također sve bolje razumijeti i njihov linearan slijed i funkcionalni doprinos u strukturiranju genoma.

1. Molekula DNK

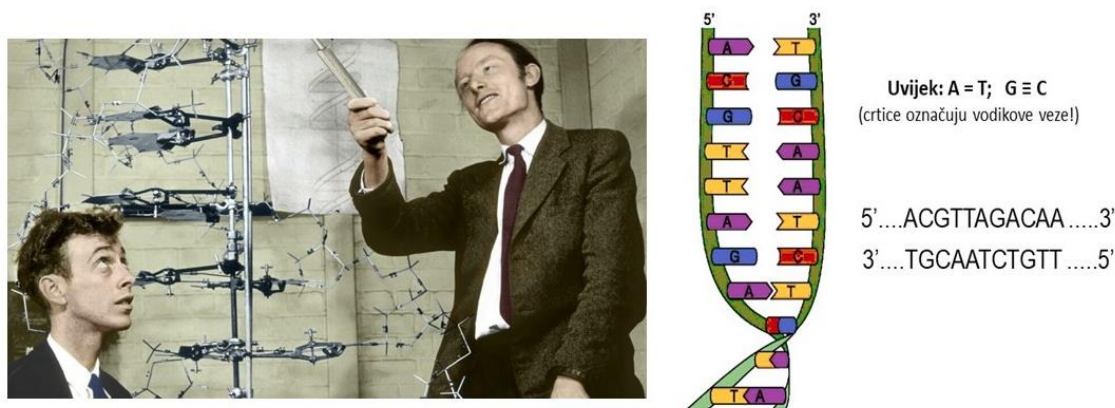
1.1. Osnovna svojstva molekule DNK

⁸ Pokretne genetičke elemente je 1950. godine otkrila Barbara McClintock (1902-1992). Istražujući regulaciju gena kukuruza dokazala je dinamičnu strukturu genoma. U to vrijeme smatralo se da je genom stabilan te su njezini rezultati dočekani s nevjericom. Nobelovu nagradu za fiziologiju i medicinu dobila je 1983. godine.

⁹ Engl. *next generation sequencing* (NGS) omogućuje procesiranje velikog broja uzoraka.

¹⁰ Engl. *long-read sequencing*. Do sada su najpoznatije metode koje koriste Pacific Biosciences (PacBio) i Oxford Nanopore Technologies (ONT).

Strukturu molekule DNK prvi su 1953. godine objasnili Francis Crick i James Watson (Sl. 1), u znanstvenom radu objavljenom u časopisu *Nature* [7], napisanom na osnovu kristalografske analize molekule DNK X-zrakama na kojoj su radili Rosalind Franklin i Maurice Wilkins¹¹



Sl. 1.: Francis Crick i James Watson na poznatoj fotografiji¹², na kojoj raspravljaju o strukturi molekule DNK¹³. Pojednostavljena shema molekule DNK, na kojoj su baze nanizane na uzdužnoj osi sastavljenoj od šećera deoksiriboze i fosfata. Desno su prikazani komplementarni parovi baza te ispis sekvence dvolančane DNK sa središnje sheme. U bazama podataka¹⁴ i pri bioinformatičkim analizama dovoljno je prikazati slijed baza samo jednog lanca, jer je drugi nužno komplementaran prvom prema pravilima sparivanja. Oznake 5' i 3' predstavljaju orijentaciju uzdužne osi s obzirom na povezivanje fosfatne skupine s deoksiribozom. Ukoliko nije posebno naznačeno, uvijek se prikazuje niz u smjeru 5' → 3'.

Molekula DNK je sastavljena od dva lanca koji se obavijaju jedan oko drugoga i tvore dvostruku uzvojniciu (Sl. 1). Svaki lanac sastoji se od naizmjenice poredanih šećera deoksiriboze i fosfatne skupine koji grade uzdužnu os molekule. Na svakoj deoksiribozi vezana je jedna od četiri baze koje nazivamo adenin, timin, guanin i citozin (A, T, G, C). Njihov uzastopni slijed je onaj koji svojom raznolikošću sadrži informaciju. Baze su u dvostrukoj uzvojnici okrenute jedna prema drugoj i specifično se sparuju putem slabih vodikovih veza kao parovi A=T (dvostruka veza) i G≡C (trostruka). Zagrijavanje otopine molekule DNK, kemijski agensi (npr. urea) ili enzimi u stanici mogu razdvojiti lance dvostruke uzvojnice. Razdvajanje lanaca je reverzibilno te se, npr. hlađenjem zagrijane otopine, struktura uzvojnice rekonstruira u potpunosti, sa sačuvanim izvornim slijedom baza. U stanici, enzimatsko razdvajanje lanaca omogućuje replikaciju, odnosno umnažanje molekule DNK, koje je potrebno kako bi se genetički materijal mogao raspodijeliti pri staničnoj diobi. Prema svakom od razdvojenih lanaca se, kao prema kalupu, sintetizira drugi, komplementaran, lanac¹⁵, čime nastaju dvije molekule s identičnim slijedom baza. One se potom raspodjeljuju u stanice koje nastaju procesom stanične diobe. Na ovom svojstvu zasniva se i enzimatsko sekvenciranje molekule DNK, kao što će biti objašnjeno u poglavlju 1.3.

¹¹ Za ovo otkriće su Maurice Wilkins, Francis Crick i James Watson 1962. godine dobili Nobelovu nagradu za fiziologiju i medicinu. Rosalind Franklin je umrla ranije, 1958. godine.

¹² <https://dnascience.plos.org/2023/04/20/seventy-years-since-watson-and-cricks-paper-introduced-dna-a-brief-history-of-the-molecule-of-life/>

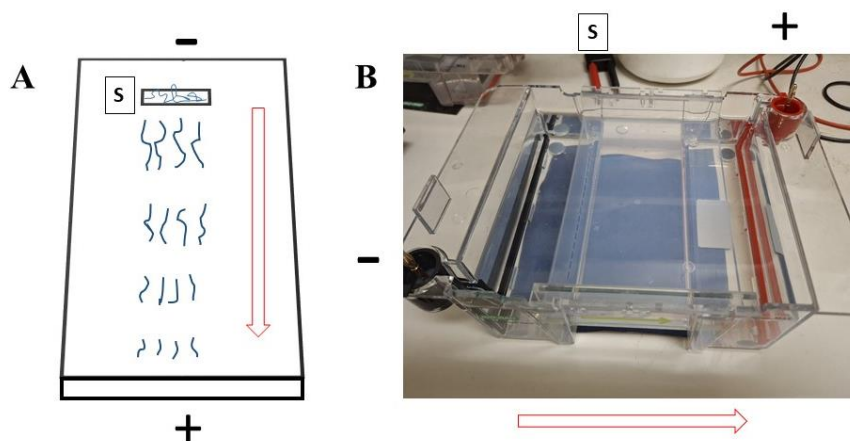
¹³ Detaljno, npr. na <https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/>

¹⁴ Važna javna baza podataka za sekvence DNK je GenBank, dio *The National Center for Biotechnology Information* (NCBI), <https://www.ncbi.nlm.nih.gov/nucleotide/>

¹⁵ Zbog ovog načina, kaže se da je replikacija molekule DNK semikonzervativna.

1.2. Elektroforetska analiza molekule DNK

Baza s pripadajućim šećerom i fosfatom¹⁶ naziva se nukleotid i čini osnovnu jedinicu svakog lanca molekule DNK. Veličina molekule DNK može se tako izraziti brojem nukleotida, kao i brojem baza, odnosno parova baza (pb)¹⁷. Negativni naboj fosfatne skupine osnova je razdvajanja fragmenata molekule DNK u električnom polju. Kako zbog negativno nabijenih fosfata svi fragmenti imaju jednaki naboj po jedinici mase, brzina migriranja u električnom polju isključiva je funkcija duljine fragmenta. Elektroforeza fragmenata molekule DNK odvija se u poroznom matriksu, odnosno gelu. Uz poznate standarde, time je moguće odrediti veličinu fragmenta, odnosno izrezati dio gela s fragmentom točno željene veličine radi daljnjeg kloniranja ili neke druge svrhe. Matriks može biti načinjen kao pločica gela izlivena od agaroze¹⁸. DNK se nanese u jažicu, 'bunarić' koji je ishodište migracije na gelu uronjenom u elektrolitsku otopinu. Molekule DNK se u električnom polju probijaju kroz pore u gelu: što je fragment kraći, lakše će i brže migrirati (Sl. 2). Preciznost razlučivanja ovisi o strukturi matriksa, odnosno što su pore manje, razlučivanje malih fragmenata je veće. Pri određivanju nukleotidne sekvence molekule DNK pore su takve da se može vidjeti razlika u duljini od jednog nukleotida. U tom slučaju gel se priprema od poliakrilamida, polimerne kemijske supstancije.



Sl. 2. A: Princip elektroforetskog razdvajanja fragmenata DNK. Negativno nabijeni fragmenti DNK migriraju kroz pore u gelu prema pozitivnoj elektrodi. Manji fragmenti se 'provlače' kroz pore brže nego veliki, dok oni najveći ostaju u jažici u koju se nanosi uzorak, na slici označena kao S (start). B: Suvremena kadica za elektroforezu molekule DNK tvrtke Bio-Rad. Znakovima + i - označena su mjesta na kojima se elektrode u kadici spajaju na ispravljač (nije prikazan). Agarozni gel (Slika 2B) mora biti uronjen u elektroforetsku otopinu. Ovakva elektroforeza omogućava razlučivanje molekula DNK u rasponu veličina od <100 pb do nekoliko tisuća pb, pri naponu od oko 100 V.

1.3. Sekvenciranje DNK prema Sangeru

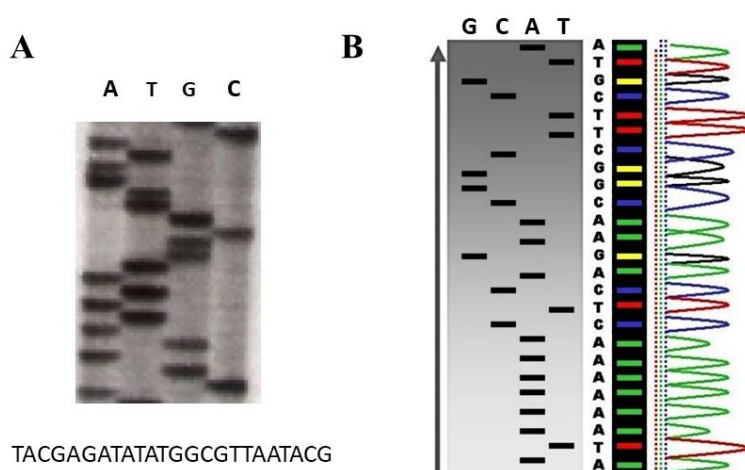
¹⁶ Svaki fosfat nosi negativni naboj na šećer-fosfatnoj osi molekule DNK.

¹⁷ Na primjer, veličina ljudskog genoma je 6,27 giga baza, raspodijeljenih na 23 pari kromosoma. Pri standardnim metodama izolacije DNK iz organizma ili iz tkiva ovako duge molekule popucaju na duljinu od oko 20 kb.

¹⁸ Agaroz je složeni oblik šećera, a dobiva se ekstrahiranjem iz crvenih algi. Za elektroforezu molekule DNK agaroz se priprema u rasponu koncentracija od 0,5-2%, ovisno o željenom razlučivanju fragmenata DNK. Zagrijavanjem se otopi i izlije u kalup u kojem se oblikuje elektroforetska pločica.

S obzirom da slijed baza u molekuli DNK nosi biološku informaciju, u suvremenim genetičkim i genomskim istraživanjima, kao i u primijenjenim biomedicinskim, forenzičkim ili drugim analizama od presudnog je značaja sekvencirati molekulu DNK, drugim riječima odrediti slijed baza. Vremenski najdulje, a uključivo u uporabi i danas, je metoda dideoksi sekvenciranja molekule DNK. Ova metoda je poznata kao sekvenciranje prema Sangeru¹⁹, koji ju je zajedno sa suradnicima razvio 1977. godine [8].

Metoda se sastoji od enzimatske replikacije fragmenata DNK *in vitro*, u prisutnosti niske koncentracije modificiranih nukleotida koji nasumičnom ugradnjom na tom mjestu zaustavljaju reakciju i rast novog lanca. Reakcija se provodi u četiri epruvete, a uz nemodificirane, u svakoj je jedan od nukleotida koji zaustavljaju reakciju, poimence, ddA (dideoksi A), ddG, ddT i ddC. Produkti reakcije se razdvajaju elektroforezom na poliakrilamidnom gelu u četiri linije (Sl. 3). U izvornom protokolu, produkt je obilježen radioaktivnim izotopom (³²P ili ³⁵S), što omogućuje vizualizaciju fragmenata eksponiranjem rendgenskog filma prislanjanjem na gel. Razlučivost na gelu je u razmaku od jedne baze, a sekvencirani fragment čita se odozdo prema gore. Na ovaj način bilo je moguće pročitati i >500 baza fragmenta DNK u kontinuitetu. Daljnjim razvojem ove metode radioaktivni biljezi zamjenjuju se fluorescencijskim pri čemu se svaki dideoksi nukleotid obilježava drugom bojom. Ovo je pak omogućilo provođenje reakcije u jednoj epruveti i automatsko očitavanje fluorescencijskog signala elektroforezom u kapilarnoj cjevčici. Razvijeni su automatski uređaji povezani s računalom u koje se izravno učitava sekvenca DNK. Tako se značajno povećala učinkovitost metode i brzo prikupljanje velikog broja podataka. Nadalje, automatizacija metode omogućila je istovremeno sekvenciranje velikog broja fragmenata DNK duljine do oko 1400 baza, što je pak omogućilo određivanje sekvence DNK cijelih genoma, uključivo i ljudskog, u svrhu provođenja genomskih projekata²⁰. Metoda prema Sangeru naziva se još i sekvenciranje DNK prve generacije i još uvijek je jedna od metoda od izbora u radu uključivo i na genomskim projektima.



Sl. 3.: Čitanje sekvence DNK sekvencirane metodom prema Sangeru. A: Radioaktivno obilježeni fragmenti vizualiziraju se eksponiranjem rendgenskog filma i čitaju odozdo prema gore. B: Shematski prikaz napredne metode koja uključuje fluorescencijsko obojene nukleotide: reakcija se odvija u jednoj epruveti, a očitavanje je

¹⁹ Frederick Sanger (1918 – 2013), britanski znanstvenik koji je u Cambridgeu razvio učinkovitu metodu sekvenciranja molekule DNK, za što je 1980. godine dobio Nobelovu nagradu za kemiju. To mu je bila druga Nobelova nagrada, prvu je dobio 1958. godine za određivanje aminokiselinskog slijeda u inzulinu.

²⁰ Krajnji cilj genomskih projekata je rekonstruirati genom u cijelosti, tj. pročitati sekvencu molekule DNK u kontinuitetu, od kraja do kraja svakog kromosoma. Ukratko, genska DNK se nasumično fragmentira na segmente takve duljine koji se mogu pročitati u cijelosti te potom preklapanjima dobivenih očitavanja rekonstruirati genom u izvornom obliku.

automatsko u odnosu na prvobitne četiri reakcije (lijeva strana u B). Strelica označava smjer čitanja. Detaljno objašnjenje je u tekstu.

2. Rana istraživanja satelitnih DNK

2.1. Što su satelitne DNK?

Satelitne DNK su sekvence DNK koje se ponavljaju jedna iza druge u gotovo identičnom obliku, izgrađujući duge nizove ponavljanja u istoj orijentaciji ('glava – rep'). Jedinice ponavljanja su satelitni monomeri. Nemaju potencijal za kodiranje pa ih se naziva i nekodirajuća ili negenska DNK. Naziv satelitna DNK nastao je u 60-tim godinama prošlog stoljeća, u pokusima centrifugiranja mišje genomske DNK u gradijentu gustoće cezijeveog klorida. Pri velikim brzinama dio sekvenci DNK bi se odvojio od glavne genomske DNK zbog razlika u nukleotidnom sadržaju²¹ [9]. Izdvajanje satelitnih DNK centrifugiranjem pri velikim brzinama u gradijentu gustoće vrlo je zahtjevno, a rezultat nesiguran, jer se samo neke razlikuju od ostatka genoma po gustoći.

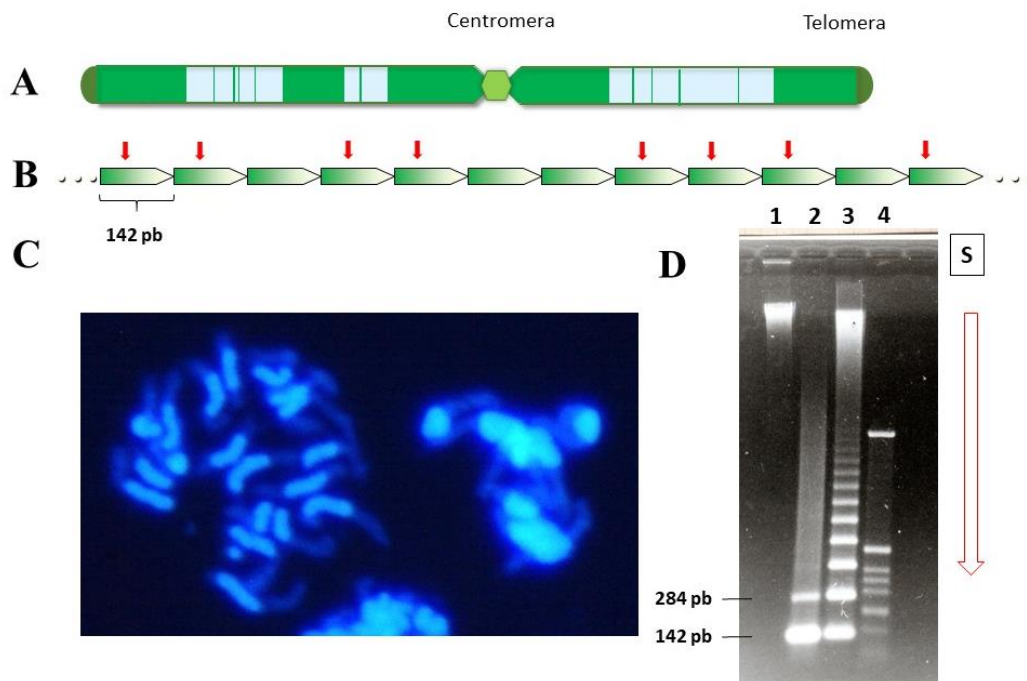
Puno učinkovitija metoda detektiranja satelitnih DNK pojavila se upotrebom restriksijskih endonukleaza kojima se cijepa ili razgrađuje sveukupna genomska DNK. Restriksijske endonukleaze su enzimi koji prepoznaju specifičan slijed u molekuli DNK, najčešće duljine 4-6 baza, i na tom mjestu 'prerežu' oba lanca molekule DNK, djelujući kao specifične molekularne škare. Ima ih više od 100 vrsta. Ukoliko je neka ponovljena sekvenca DNK organizirana u uzastopnom nizu, cijepanje molekule DNK s odgovarajućom restriksijskom endonukleazom oslobodit će veliki broj jedinica ponavljanja koje nazivamo satelitni monomeri. Ipak, zbog malih varijacija u nukleotidnoj sekvenci, enzim neće moći prepoznati sva mjesta te će se pojaviti i fragmenti dvostruke, trostruke, odnosno višestruke duljine monomera satelitne DNK. Pri razdjeljivanju produkata restrikcije u električnom polju, zbog toga se dobiveni fragmenti vide kao karakteristična 'ljestvica' genomske DNK. Ovo je najjednostavniji, ali ujedno i nepobitan dokaz da je neka sekvenca DNK organizirana kao niz uzastopnih ponavljanja, odnosno kao satelitna DNK (Sl. 4B, D).

Analize lokalizacije satelitnih DNK na kromosomima²² pokazale su da često zauzimaju velika područja koja su značajna za opstojnost kromosoma, kao i genoma u cjelini. Tako redovito predstavljaju dominantnu komponentu molekule DNK u području centromere²³ te na kromosomskim krajevima gdje sudjeluju u održavanju integriteta kromosoma. Također, u duljim ili kraćim nizovima prisutne su i u dijelovima kromosoma između kromosomskih krajeva i centromernog područja (Sl. 4A). Nije opažena nikakva izravna (uzročna) veza kojom bi se povezale zastupljenost i raspored satelitnih DNK u genomu i svojstva organizma.

²¹ A=T parovi baza lakši su od G≡C parova. Ovisno o njihovoj zastupljenosti u pojedinim segmentima molekule DNK u odnosu na glavnu genomske segmenta, razlikovat će se po gustoći plutanja i pojaviti kao odvojena, 'satelitna' frakcija pri centrifugiranju. Ovo je upravo bio slučaj kod uzastopno ponovljenih sekvenci DNK u mišjem genomu.

²² Satelitna DNK (ili bilo koja druga sekvenca DNK) se na kromosomima lokalizira metodom fluorescencijske hibridizacije *in situ* (FISH). Željena sekvenca DNK obilježi se fluorescentnim biljezima te pod određenim uvjetima hibridizira, tj. veže se na komplementarni slijed nukleotida na kromosomima. Signal se vizualizira pomoću fluorescencijskog mikroskopa, pod svjetlošću odgovarajuće pobudne valne duljine.

²³ Centromera je dio kromosoma odgovaran za pravilno razdvajanje genetičkog materijala u staničnoj diobi.



Sl. 4. A: Shematski prikaz kromosoma: dijelovi u kojima se najčešće mogu naći satelitne DNK označeni su zeleno. B: Niz jedinica ponavljanja (monomera) satelitne DNK. Crvene strelice iznad pojedinih monomera predstavljaju mjesto u sekvenci DNK koje prepoznaje neka restrikcijska endonukleaza. Zbog varijabilnosti u sekvenci DNK, takva mjesta ne postoje u svim monomerima. Jedinica ponavljanja glavne satelitne DNK kukca žutog brašnara *Tenebrio molitor* duljine je 142 pari baza (pb). C: Kromosomi žutog brašnara obojeni fluorescentnom bojom DAPI (4',6-diamidino-2-fenilindol) i snimljeni fluorescencijskim mikroskopom. Intenzivno obojeni dijelovi odgovaraju području oko centromere, u kojem se u ovom organizmu nalazi satelitna DNK duljine monomera 142 pb, koja izgrađuje do 40% genomske DNK. D: Elektroforetsko razdvajanje fragmenata satelitne DNK žutog brašnara. U liniji 1 je nanosena DNK prije razgradnje restrikcijskom endonukleazom. U liniji 2 je DNK razgrađena s endonukleazom koja ju cijepa u gotovo svakom monomeru, dok je razgradnja u 3 djelomična pa je izražena ljestvica osnovnog fragmenta. U 4 je kontrola razgradnje. Napon elektroforeze bio je 80 V uz struju 130 mA, u trajanju oko 4 sata. Elektroforeza je rađena na 1,2% agaroznom gelu.

2.2. Prva sekvenciranja u Hrvatskoj

Monomeri satelitne DNK žutog brašnara *Tenebrio molitor* su prema spoznajama autora ovog teksta prvi fragmenti molekule DNK koji su bili sekvencirani u Hrvatskoj²⁴. Ukratko, monomer satelitne DNK bio je prvo izrezan iz gela poput onog prikazanog na slici 2, DNK se ekstrahirala iz izrezanog djelića gela i klonirala kako bi se fragmenti umnožili. Sekvenciranje je rađeno metodom prema Sangeru, opisanom u poglavlju 1.3. Fragmenti DNK bili su obilježeni radioaktivnim izotopom i detektirani eksponiranjem gela na filmu. Elektroforeza je rađena na poliakrilamidnom gelu debljine 0,3 mm, izlivenom između dvije staklene ploče²⁵. Za ovu elektroforezu bio je potreban visokonaponski

²⁴ Prva sekvenciranja DNK radila su se 1986. / 1987. godine na Institutu Ruđer Bošković u Zagrebu. Autor ovog teksta sudjelovao je u uspostavljanju sustava sekvenciranja, uz Đurđicu Ugarković i Veru Gamulin (1948 - 2006), tada seniornu znanstvenicu koja je po povratku s poslijedoktorskog usavršavanja na Sveučilištu Yale (New Haven, SAD) prva kod nas uvela metodologiju rekombinantne DNK u svakodnevni laboratorijski rad.

²⁵ Radi se o gelu koji prilikom elektroforeze stoji vertikalno, što je razlika od elektroforeze na agaroznom gelu koja je horizontalna, kako je prikazano na Sl. 2B. Nakon elektroforeze na poliakrilamidnom gelu, jedno staklo se ukloni, gel se prekrije plastičnom folijom na koju se stavi rendgenski film i umetne u kazetu u kojoj se film

su nastavljena i istom metodom određene su jedinice ponavljanja najzastupljenijih satelitnih DNK u nizu drugih vrsta kukaca kornjaša (Coleoptera), srodnika vrste *Tenebrio molitor*. Neke od istraživanih vrsta su i danas značajne kao modelni organizmi za genomska istraživanja nekodirajuće DNK (na primjer, *Tribolium castaneum* [11] i *Palorus subdepressus* [12]). Svi istraživani kukci kornjaši bili su od interesa i zato jer predstavljaju štetnike u skladištima hrane. Iako također izvorno štetnik u skladištima hrane, žuti brašnar *Tenebrio molitor* je svojevrсна iznimka jer se uzgaja i kao hrana za životinje, a u zadnje vrijeme odobrena mu je upotreba u ljudskoj prehrani kao dopunskom izvoru hranjivih tvari²⁷. Koristi se također i u biomedicinskim istraživanjima. Prednost žutog brašnjara je lak uzgoj te veličina, odraslih do 2 cm, dok su njegove ličinke, koje su komercijalno iskoristive i do oko 3 cm duljine (Sl. 6). Ostali spomenuti istraživani kukci su veličine svega nekoliko milimetara. Osim kod kukaca kornjaša, moj vlastiti interes za istraživanjem satelitnih DNK proširio se na još dvije skupine organizama, školjkaše (Bivalvia) i oblice korijenovih kvržica (Nematoda), koji se po organizaciji genoma te po nekim strukturnim i organizacijskim svojstvima satelitnih DNK značajno razlikuju od prethodno navedenih kukaca te usporedno istraživani, zajedno s rezultatima drugih istraživačkih grupa, daju općenitiju sliku [13, 14].

Različite istraživačke grupe u svijetu su u tom razdoblju na sličan način analizirale satelitne DNK različitih organizama, uključivo i čovjeka. Iako je tako određen veliki broj satelitnih DNK, ovom metodom najčešće se detektiralo jednu ili samo nekoliko najzastupljenijih u genomu. Zaključak o reprezentativnoj 'konsenzus' sekvenci pojedine satelitne DNK zasnivao se na uspoređivanju malog broja sekvenciranih monomera, najviše do po nekoliko desetaka individualno izoliranih i kloniranih. Ipak, već ovakva istraživanja dala su značajne informacije i pregled o zastupljenosti satelitnih DNK, njihovoj organizaciji u genomima, strukturi, mogućim ulogama i evolucijskim karakteristikama u velikom broju vrsta biljaka i životinja [13, 15].



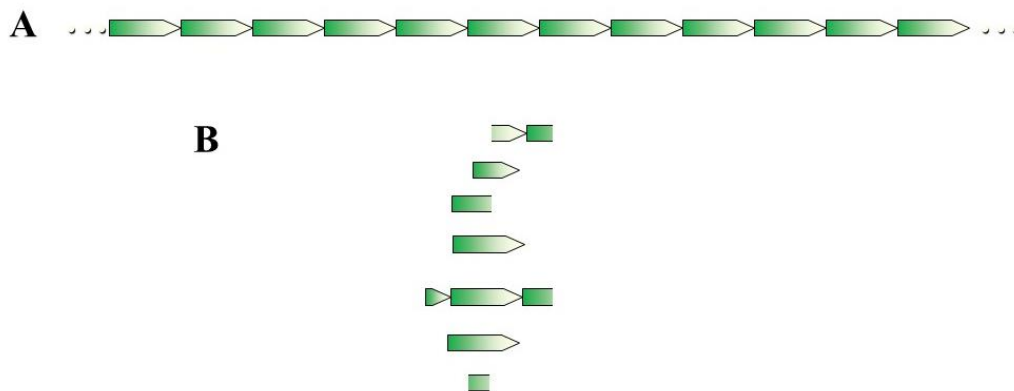
Sl. 6.: Odrasla jedinka i ličinka kukca kornjaša žutog brašnjara *Tenebrio molitor*. U komercijalne svrhe koriste se ličinke.

²⁷ Prva dozvola upotrebe sušenog žutog brašnjara u ljudskoj prehrani u EU bila je donesena na preporuku agencije EFSA (European Food Safety Authority) 22. lipnja 2021., a 20. siječnja 2025. EU Komisija odobrila je stavljanje na tržište UV-sterilizirano 'brašno' od mljevenih cijelih ličinki žutog brašnjara kao novog izvora ljudske prehrane. Može se koristiti i kao hranjivi dodatak pojedinim proizvodima, uključivo kruhu, kolačima, tjestenini, proizvodima od krumpira, sira itd.

3. Istraživanja satelitnih DNK visokoprotlačnim metodama sekvenciranja

3.1. Problemi genomske analize satelitnih DNK

Iako je prethodno opisana metoda pružila veliki broj podataka o strukturi i evoluciji satelitnih DNK i rezultirala velikim brojem objavljenih radova, bila je ograničena samo na one najzastupljenije u genomu. Ovaj pristup tako nije mogao odgovoriti na osnovno pitanje: koliko u genomu neke vrste postoji različitih tipova uzastopno ponovljenih sekvenci DNK? Sve one koje su prisutne u relativno malim udjelima²⁸ te koje se nisu mogle vizualizirati razgradnjom genomske DNK restrikcijским endonukleazama i elektroforezom na agaroznom gelu (Sl. 4) ostale bi neotkrivene. Drugo ograničenje u analizama satelitnih DNK bila je nemogućnost rekonstruiranja nizova monomera i njihovog pravilnog postavljanja u odnosu na ostale sekvence DNK u genomskim projektima. Osnovni uzrok tomu je mala razlika među sekvencama monomera pojedine satelitne DNK koje je zbog toga nemoguće jednoznačno nadovezati u neprekinuti niz (Sl. 7). Ukoliko se gotovo isti monomeri nalaze na više lokacija u genomu, nemoguće je odrediti odakle točno potječu. Posljedica je da su u rezultatima genomskih projekata koji se zasnivaju na sekvenciranju prema Sangeru satelitne DNK izrazito podzastupljene, a često i izostavljene. Drugim riječima, tako rekonstruirani genomi su isprekidani i nepotpuni s obzirom na ponovljene sekvence DNK [16]. Primjere podzastupljenih satelitnih DNK u sastavljenim genomskim projektima u odnosu na eksperimentalnu detekciju i analizu nalazimo u svim do sada sekvenciranim vrstama pa tako npr. i kod školjkaša [17]. Kod žutog brašnara *Tenebrio molitor*, opisanim eksperimentalnim metodama bilo je procijenjeno da glavna satelitna DNK izgrađuje > 40% genomske DNK [18], dok je u 'cijelom' sekvenciranom i sastavljenom genomu udio svih satelitnih DNK brašnara bio samo 0,15% [19]. Detaljna analiza metodom sekvenciranja nove generacije dala je procjenu zastupljenosti glavne satelitne DNK znatno bližu prvobitnoj eksperimentalnoj, 26,5% [20].



Sl. 7. A: Niz satelitne DNK sastoji se od uzastopno ponovljenih gotovo identičnih jedinica (monomera). B: Fragmenti sekvence satelitne DNK dobiveni sekvenciranjem. S obzirom na malu razliku u nukleotidnoj sekvenci između jedinica ponavljanja, preklapanjem fragmenata nemoguće je rekonstruirati originalni niz, što u sastavljanju sekvenciranih genoma dovodi do podzastupljenosti, potpunog izostanka ili pogrešno složenih nizova ovih važnih genomskih elemenata.

²⁸ Npr. koje izgrađuju < 2-3% genomske DNK. Osobito je to slučaj kod vrsta kod kojih u genomu postoji jedna (ili nekoliko) visokozastupljenih satelitnih DNK, koje će elektroforezom na agaroznom gelu (kao npr. na Sl. 4D) zbog svoje količine onemogućiti detekciju elektroforetskih pruga nisko zastupljenih satelitnih DNK.

3.2. Određivanje satelitnih DNK sekvenciranjem sljedeće generacije

Sekvenciranje sljedeće²⁹ (druge, nove) generacije omogućuje istovremeno sekvenciranje milijuna malih fragmenata DNK (50 do 400 pari baza) uz veliku točnost dobivenih podataka i po niskoj cijeni. Cijeli genom nekog organizma moguće je tako na jednom uređaju sekvencirati u jednom danu. Sekvenciranje ovom metodom zasniva se na kemijskoj reakciji dodavanja fluorescencijski obilježenog nukleotida na lanac koji se sintetizira prema genomskom fragmentu imobiliziranom u protočnoj komori³⁰. Po očitavanju, fluorescencijski biljeg ukloni se s nukleotida, a na rastući lanac doda se slijedeći obilježeni nukleotid. Svaki od četiri nukleotida obilježen je svojom fluorescencijskom bojom. Fragmentima DNK koje se sekvencira prethodno su dodani adapteri³¹, koji služe za vezivanje na stijenku protočne komore te za identifikaciju, kao svojevrsan bar-kod. Na taj način moguće je sekvencirati veliki broj fragmenata DNK istovremeno, uključivo i iz raznih vrsta, a potom ih razvrstati u računalu. Poseban izazov predstavlja zadnji korak, računalno preklapanje ovako kratkih fragmenata u cjelovite sekvence DNK, pri čemu ponovljene sekvence DNK i dalje predstavljaju nepremostivu prepreku.

Iako nepodesna za sastavljanje nizova monomera satelitnih DNK u kontinuitetu, metodom sekvenciranja sljedeće generacije moguće je odrediti ukupan sadržaj svih satelitnih DNK u genomu, odnosno satelitom [21]. S obzirom da se radi o sekvencama DNK koje se u genomu pojavljuju u višestrukim kopijama, cijepanjem ukupne genomske DNK na kratke fragmente ponovljene sekvence će biti najzastupljenije u ukupnoj masi segmenata DNK. Zbog toga je za određivanje satelitoma dovoljno pročitati segmente DNK u malom uzorku, manjem nego što je to veličina samog genoma kojeg istražujemo, odnosno dovoljna je tzv. pokrivenost genoma³² od samo 0,2 – 0,5 puta. Ovim pristupom izdvajaju se ponovljene sekvence DNK od ostalih, jedinstvenih, čime se istovremeno uklanja i 'šum' u analizi. Razvoj cijelog niza specijaliziranih bioinformatičkih alata omogućio je detektiranje svih (ili gotovo svih) ponovljenih sekvenci u nekom genomu [6]. Najčešće se koristi *RepeatExplorer*³³ [22]. Sekvencirane kratke genomske fragmente DNK *RepeatExplorer* grupira prema sličnosti. Metoda je zasnovana na pretvaranju očitanih sličnosti u grafove³⁴, u kojima gustoća i način rasporeda čvorova omogućuje identifikaciju pojedinih ponovljenih sekvenci. Izlazna lista programa uključuje procijenjen udio ponovljene sekvence u genomu, duljinu i konsenzus sekvencu jedinice ponavljanja te procjenu radi li se o satelitnoj DNK ili je sekvenca raspršeno ponovljena. Ovim pristupom moguće je identificirati i satelitne DNK koje su u genomu prisutne u malom broju kopija i izgrađuju < 1% genomske DNK. Velika prednost je da za analizu nije potrebno imati sastavljen genom, dok je nedostatak da metoda daje

²⁹ Engl. *Next generation sequencing*, NGS, istovremeno procesira veliki broj uzoraka, koristeći, visokoprotočne metode (engl. *high-throughput*). Jedna od najpoznatijih tehnologija koja se koristi i u sekvenciranju satelitnih DNK je Illumina.

³⁰ Engl. *flow cell*

³¹ Radi se o umjetno sintetiziranim kratkim segmentima DNK koji se dodaju na krajeve analiziranog genomskeg fragmenta.

³² Zamislamo da iz neke knjige želimo izdvojiti riječi koje se u njoj često ponavljaju: za tako nešto nije potrebno pročitati cijelu knjigu. Već nasumičnim listanjem uočiti ćemo riječi koje se često koriste. Pokrivenost genoma (engl. *genome coverage*) omjer je broja sekvenciranih nukleotida i veličine genoma. Tako je za točno i detaljno sastavljanje genoma s obzirom na jedinstvene sekvence DNK (npr. gene), genom potrebno pročitati 50 pa i više puta. Za analizu svih sekvenci DNK koje se ponavljaju, dovoljno je da pokrivenost genoma bude do 0,5 puta.

³³ Ovdje se referira na unaprijeđenu verziju *RepeatExplorer 2*, prva verzija bila je objavljena 2013. godine. Alat je dostupan na platformi Galaxy <https://repeatexplorerelixir.cerit-sc.cz/galaxy/>.

³⁴ U algoritmu se koriste De Bruijn-ovi grafovi. Program iscrtava njihovu strukturu za fragmente svake ponovljene sekvence. Kružni prikaz znak je da su monomeri ponovljeni u nizu jedan iza drugog, odnosno da predstavljaju satelitnu DNK. Zrakasti prikaz grafa ukazuje da su sekvence DNK raspršeno ponovljene u genomu.

samo katalog ponovljenih sekvenci, ali ne govori ništa o njihovom rasporedu u genomu, odnosno na kromosomima. Za potonje informacije koriste se eksperimentalne metode lokalizacije i mapiranja.

Istraživanje ukupnog broja satelitnih DNK u genomima velikog broja biljnih i životinjskih vrsta donijelo je značajan napredak u razumijevanju njihove zastupljenosti, organizacije, evolucije kao i mogućih uloga [15, 23]. Kao što se naslućivalo već i ranije, ne postoji nikakav jedinstveni obrazac u smislu strukture nukleotidne sekvence, broja i zastupljenosti u genomima, što je sada postalo i jasno pokazano (Tablica 1).

U okviru rada na genomskom projektu žutog brašnara *Tenebrio molitor*, satelitom ovog organizma određen je metodom sekvenciranja sljedeće generacije i opisanim računalnim pristupom [20]. Satelitom izgrađuje 11 satelitnih DNK koje zajedno zauzimaju oko 28% genoma. Duljine jedinica ponavljanja (monomera) su u rasponu od 93 *pb* do 735 *pb*. Najzastupljenija od svih satelitnih DNK u genomu brašnara je ona duljine monomera 142 *pb* i čiji je uzorak monomera već bio sekvenciran metodom prema Sangeru [10, 18]. Ova satelitna DNK sada je nazvana TmSat1, a izgrađuje 26,5% genomske DNK. Preostalih 10 satelitnih DNK prisutno je u vrlo malom broju kopija i izgrađuju od 0,45% do 0,01% genoma [20]. Jedina razlika u odnosu na prijašnji rezultat je upravo u genomskoj zastupljenosti TmSat1: dok je prvobitna eksperimentalna procjena bila > 40% [18], u satelitskoj analizi iznosila je 26,5% genoma. Iako određivanje točne vrijednosti ostaje otvoreno, vrijedi spomenuti kako i ova niža predstavlja jedan od najvećih udjela samo jedne satelitne DNK u nekom genomu.

Vrsta	Broj satelitnih DNK	% genoma	Najzastupljenija satelitna DNK (%)
<i>Crassostrea gigas</i>	52	6.33	1.29
<i>Cydalima perspectalis</i>	1	0.14	0.14
<i>Diatraea postlineella</i>	2	0.06	0.04
<i>Eneoptera surinamensis</i>	45	14	1.41
<i>Gryllus assimilis</i>	13	4	1.35
<i>Hippodamia variegata</i>	30	14.93	9.37
<i>Locusta migratoria</i>	60	2.14	0.97
<i>Locusta migratoria</i>	62	2.39	0.98
<i>Megaleporinus macrocephalus</i>	164	13.47	2.78
<i>Meloidogyne arenaria</i>	81	5.07	0.31
<i>Meloidogyne floricola</i>	38	1.55	0.14
<i>Meloidogyne incognita</i>	56	3.52	0.24
<i>Meloidogyne javanica</i>	73	4.27	0.21
<i>Oedaleus decorus</i>	58	2.53	0.62
<i>Ostrinia nubilalis</i>	4	0.27	0.15
<i>Pontastacus leptodactylus</i>	258	27.57	10.91

Tablica 1.: Usporedba broja satelitnih DNK, njihovog ukupnog udjela u genomu i udjela najzastupljenije satelitne DNK kod nekolicine vrsta beskralješnjaka (odabir iz [23]). U ovom uzorku, broj satelitnih DNK kreće se od samo jedne do 258, a ukupni udio od < 1% do oko 27%. U ovom uzorku, samo jedna satelitna DNK nađena je kod šimširovog moljca, a najveći broj satelitnih DNK i njihov udio u genomu je kod jedne vrste riječnog raka.

U organizacijskom smislu, analizom satelitoma promijenjena je paradigma o grupiranju satelitnih DNK isključivo u velikim blokovima, kakvi su često (ali ne uvijek) u području centromere [23]. Satelitne DNK mogu formirati i nizove koji odstupaju od ovog pravila i nalaze se raspršeni po genomu. Primjerice, takav obrazac raspršene organizacije relativno kratkih nizova uzastopnih ponavljanja sekvence DNK odredili smo kod školjkaša koje smo proučavali [24]. Zaključno,

određivanje satelitnih DNK sekvenciranjem sljedeće generacije, osim što je značajno povećalo broj podataka, kvalitativno je unaprijedilo naše znanje, posebno zahvaljujući jednostavnom istraživanju satelitoma u velikom broju vrsta koje nisu modelni organizmi³⁵. Također, ova istraživanja unaprijedila su i spoznaje o povezanosti satelitnih DNK i drugih ponovljenih sekvenci DNK, pokretnih genetičkih elemenata, izuzetno značajnih za izgradnju i evoluciju genoma i prilagođavanju promjenama u okolišu [5]. Pojedini pokretni genetički elementi mogu u svojoj strukturi sadržavati kratke nizove uzastopnih ponavljanja koje sličje satelitnim DNK, koji se na taj način mogu širiti po genomu i potom formirati 'klasične' satelitne nizove. Upravo je ova povezanost satelitnih DNK i pokretnih genetičkih elemenata posebno izražena kod školjkaša [24]. Pokazalo se međutim da korištenjem opisane *high-throughput* metodologije prisutnost segmenata satelitnih DNK u pokretnim genetičkim elementima može 'zbuniti' računalne algoritme i dovesti do nepotpunih ili čak krivih klasifikacija pojedinih sekvenci. Problem se može riješiti opetovanim analizama podgrupa sekvenciranih fragmenata, uključivo i 'ručnim' podešavanjem sravnjivanja uspoređivanih sekvenci DNK³⁶ [25].

3.3. Treća generacija sekvenciranja i satelitne DNK

Tehnologije koje nude rješenje za sastavljanje genoma u cjelosti, od kraja do kraja svakog kromosoma, su metode sekvenciranja dugih segmenata DNK, općenito poznate kao treća generacija sekvenciranja³⁷ [26]. Najpoznatije su one koje koriste Oxford Nanopore Technologies (ONT) i Pacific Biosciences (PacBio). Obje su ušle u uporabu približno 2015. godine. Od tada se ovakvi pristupi sve više i više koriste u genomskim istraživanjima. Čitanjem dugih segmenata genomske DNK (više stotina *kb* u kontinuitetu) smanjuje se broj preklapanja pri rekonstruiranju genoma u cjelini, a time i problemi koje nose dijelovi genoma obogaćeni ponovljenim sekvencama DNK. Princip sekvenciranja Oxford Nanopore tehnologijom (ONT)³⁸ zasniva se na mjerenju promjena u električnoj struji u ovisnosti o bazi u molekuli DNK koja u tom trenutku prolazi kroz nanoporu proteina umetnutog u membranu. Napravljena ovom metodom, 2022. godine bila je objavljena nova, gotovo cjelokupna verzija genoma čovjeka³⁹ [27]. Dovršena je 21 godinu nakon objavljivanja prve verzije, sastavljene preklapanjem genomskih fragmenata sekvenciranih prema Sangeru. Osim oko 200 *Mb* novododanih sekvenci DNK koje su u prvoj verziji promakle, sada su ispravljene i brojne greške koje su tada bile načinjene pri sastavljanju genoma. Pročitana su i sastavljena još i neka funkcionalno značajna područja, kao što su to centromerna, izgrađena od složenih oblika satelitnih DNK koja su sada mogla biti proučavana u stvarnom linearnom nizu. Danas se sve više vrsta uključuje u određivanje genoma tehnologijom sekvenciranja DNK treće generacije.

Trećom generacijom sekvenciranja i daljnjim razvojem bioinformatičkih alata otvorene su do sada neslućene mogućnosti u istraživanjima satelitnih DNK i genomske tamne materije općenito, što je dovelo do novih pomaka u razumijevanju njihove strukture i uloge [28]. Ovaj pristup omogućuje detaljno istraživanje linearnog rasporeda pojedinih satelitnih DNK u dijelovima genoma koji su do sada bili nedostupni [29]. Posebno je korisno kombiniranje različitih metoda [23]. Na primjer, satelitnim DNK identificiranim sekvenciranjem druge generacije mogu biti određene točne duljine nizova monomera, kao i položaj svakog takvog niza u odnosu na druge sekvence DNK u sastavljenom genomu.

³⁵ Dobro poznati modelni organizam je npr. vinska mušica, *Drosophila melanogaster*.

³⁶ Ovom napomenom želi se istaknuti kako niti ova napredna tehnologija još ne može detektirati sve situacije koje se pojavljuju u genomima. Potrebne su daljnje metodološke nadogradnje i usavršavanja. Detaljan opis ovakvih situacija i njihovih razrješavanja nadilazi ciljeve ovog članka.

³⁷ Engl. *third-generation sequencing*, odnosno *long-read sequencing*.

³⁸ Npr. <https://nxtgnt.ugent.be/nanopore-sequencing/>.

³⁹ Neki dijelovi su i u ovoj unaprijeđenoj verziji ostali nepotpuni.

Satelitne DNK djelomično ipak i dalje izmiču točnoj karakterizaciji pa genomi sastavljeni metodologijom treće generacije mogu ostati nepotpuni, posebice oni iznimno bogati satelitnim DNK. Sastavljanje genoma žutog brašnara *Tenebrio molitor* jedan je takav primjer. Genom brašnara sekvenciran tehnologijom treće generacije PacBio i sastavljen od tako dobivenih genomskih segmenata predstavlja tek oko 53% eksperimentalno određene veličine genoma [20]. Vjerojatno je uzrok tomu ogromna količina dugih nizova vrlo homogenih monomera glavne satelitne DNK. Primijećeno je da se s obzirom na duljinu nizova i sličnost monomera oni i dalje ne uspijevaju pravilno nadovezati između sebe, ili na druge sekvence genomske DNK. Potrebno je stoga daljnje unaprijeđivanje pristupa i metodologije koje će omogućiti rješavanje i ovakvih problema.

Zaključak

Satelitne DNK nazivamo segmente DNK ponovljene jedan iza drugog u uzastopnom slijedu. Ovakvi nizovi još uvijek predstavljaju najzagonetnije genomske komponente. Istraživanje satelitne DNK žutog brašnara *Tenebrio molitor* bilo je pokretač sekvenciranja DNK i istraživanja genomike ponovljenih sekvenci u Hrvatskoj.

Napredak u tehnikama sekvenciranja i obrade genomskih podataka posljednjih godina razjasnili su mnoga pitanja vezana uz broj, genomsku lokaciju i evoluciju satelitnih DNK [15, 28]. Komparativna satelitomika omogućuje istraživanje razlika u sadržaju, broju kopija i lokacijama između jedinki i među vrstama. Danas znamo da satelitne DNK imaju značajnu ulogu u (re)organizaciji genoma, njegovom funkcioniranju i evoluciji. Bez obzira što ne postoji gotovo nikakva sličnost između pojedinih satelitnih DNK (osim organizacije u nizove uzastopnih ponavljanja), one su najčešći oblik sekvenci DNK u ključnim dijelovima kromosoma, odgovornim za staničnu diobu (centromerno područje) [30], kao i na kromosomskim krajevima (telomerno područje). Ove sekvence sudjeluju u formiranju strukture stanične jezgre [31], kao i u odgovoru na stres [32]. Promjene u duljinama nizova satelitnih DNK u blizini nekih gena predstavljaju biljege koji se povezuju s pojavom patoloških stanja, kao što su shizofrenija [33], ili tumori i druge bolesti [34], a mogu utjecati i na ponašanje životinja (voluharice) [35]. Razlike u broju nekih kratkih jedinica ponavljanja individualno su specifične i koriste se u forenzici kao genomski 'otisak prsta' te se mogu koristiti i kod utvrđivanja srodnosti (primjerice očinstva). Moglo bi se nabrajati još puno primjera njihovih funkcionalnih uloga.

Satelitne DNK su prošle dugi put od genomskih komponenata koje su se u početku smatrale beskorisnim viškom, do značajnih graditelja i organizatora genoma. Treba naglasiti da je poznavanje strukture, rasporeda i uloga satelitnih DNK ključ razumijevanja genoma složenih višestaničnih organizama. Iako su se paradigme organizacije i uloga satelitnih DNK u zadnjim godinama znatno mijenjale, mnoge tajne vezane uz ove sekvence DNK očito još nisu razjašnjene! Zbog toga je osvježavanje o tomu o kakvim se sekvencama DNK radi i kako se one istražuju od velikog značaja za sve koje zanima biološka informacija kao takva, kao i razvoj metodologija kojima se ona može proučavati.

Zahvala

Najljepše zahvaljujem prof. Franji Joviću na pozivu za sudjelovanjem na 12. simpoziju PIFT 2025 i poticaju za pripremu ovog članka te dr.sc. Evi Šatović-Vukšić s Instituta Ruđer Bošković na kritičkom čitanju teksta te na pomoći u izradi dijela slika.

Literatura

- [1] L. von Bertalanffy, *Opća teorija sustava: osnove, razvoj, primjene*, Kiklos – krug knjige, Zagreb, lipanj 2022. ISBN: 978-953-7992-16-3.
- [2] D. Hammond, The legacy of Ludwig von Bertalanffy and its relevance for our time, *Systems Res. and Behavioral Sci.*, 36 (3), svibanj/lipanj 2019, stranice 301-307, posebno izdanje: Ludwig von Bertalanffy: legacies, traces, reverberations and amplifications. Commemorating the man and the work 50 years after 'General System Theory', <https://doi.org/10.1002/sres.2598>.
- [3] N. Brunner, M. Kühleitner, K. Renner-Martin. Bertalanffy-Pütter models for avian growth, *PLoS One*, 16(4), travanj 2021, e0250515, <https://doi.org/10.1371/journal.pone.0250515>.
- [4] M. Plohl, Kratki pregled istraživanja genoma u svjetlu opće teorije sustava, Povijest i filozofija tehnike, 11. simpozij PIFT 2023, ur. Franjo Jović, HRO CIGRE, Zagreb 2023, str. 19-42.
- [5] E. Šatović-Vukšić, M. Plohl, The secret world of „jumping“ DNA, *Frontiers for Young Minds*, 12, travanj 2024, 1-9, članak 1279209, <https://doi.org/10.3389/frym.2024.1279209>.
- [6] E. Šatović, M. Tunjić-Cvitanić, M. Plohl, Tools and databases for solving problems in detection and identification of repetitive DNA sequences, *Period. Biol.*, 121–122, prosinac 2020, 7–14, <https://doi.org/10.18054/pb.v121-122i1-2.10571>.
- [7] J.D. Watson, F.H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature*, 171(4356), travanj 1953, 737-738, <https://doi.org/10.1038/171737a0>.
- [8] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA*, 74(12), prosinac 1977, 5463-5476, <https://doi.org/10.1073/pnas.74.12.5463>.
- [9] S. Kit, Equilibrium sedimentation in density gradients of DNA preparations from animal tissues, *J. Mol. Biol.*, 3, prosinac 1961, 711–716, [https://doi.org/10.1016/S0022-2836\(61\)80075-2](https://doi.org/10.1016/S0022-2836(61)80075-2).
- [10] Đ. Ugarković, M. Plohl, V. Gamulin, Sequence variability of satellite DNA from the mealworm *Tenebrio molitor*, *Gene*, 83, travanj 1989, 181-183, [https://doi.org/10.1016/0378-1119\(89\)90417-4](https://doi.org/10.1016/0378-1119(89)90417-4).
- [11] Đ. Ugarković, M. Podnar, M. Plohl, Satellite DNA of the red flour beetle *Tribolium castaneum*--comparative study of satellites from the genus *Tribolium*, *Mol. Biol. Evol.*, 13(8), listopad 1996, 1059-1066, <https://doi.org/10.1093/oxfordjournals.molbev.a025668>.
- [12] M. Plohl, N. Meštrović, B. Bruvo, Đ. Ugarković, Similarity of structural features and evolution of satellite DNAs from *Palorus subdepressus* (Coleoptera) and related species, *J. Mol. Evol.* 46(2), veljača 1998, 234-239, <https://doi.org/10.1007/pl00006298>.
- [13] M. Plohl, Those mysterious sequences of satellite DNAs, *Periodicum Biologorum*, 112(4), prosinac 2010, 403-410, <https://hrcak.srce.hr/63710>.
- [14] M. Plohl, V. Petrović, A. Luchetti, A. Ricci, E. Šatović, M. Passamonti, B. Mantovani, Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks, *Heredity*, 104, lipanj 2010, 543-551, <https://doi.org/10.1038/hdy.2009.141>.
- [15] M.A. Garrido-Ramos, Satellite DNA: an evolving topic, *Genes* (Basel), 8(9):230, rujan 2017, <https://doi.org/10.3390/genes8090230>.
- [16] O.K. Tørresen, B. Star, P. Mier, M.A. Andrade-Navarro, A. Bateman, P. Jarnot, A. Gruca, M. Grynberg, A.V. Kajava, V.J. Promponas, M. Anisimova, K.S. Jakobsen, D. Linke, Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases, *Nucleic Acids Res.* 47(21), prosinac 2019, 10994-11006, <https://doi.org/10.1093/nar/gkz841>.
- [17] E. Šatović-Vukšić, M. Plohl, Exploring Satellite DNAs: Specificities of Bivalve Mollusks Genomes, *Prog. Mol. Subcell. Biol.* 60, 2021, 57-83, https://doi.org/10.1007/978-3-030-74889-0_3.
- [18] C.A. Davis, G.R. Wyatt, Distribution and sequence homogeneity of an abundant satellite DNA in the beetle, *Tenebrio molitor*, *Nucleic Acids Res.*, 17(14), srpanj 1989, 5579–5586, <https://doi.org/10.1093/nar/17.14.5579>.
- [19] S. Kaur, S.A. Stinson, G.C. diCenzo, Whole genome assemblies of *Zophobas morio* and *Tenebrio molitor*, *G3 Genes|Genomes|Genet.*, 13(6), lipanj 2023, jkad079, <https://doi.org/10.1093/g3journal/jkad079>.
- [20] B. Oppert, A.T. Dossey, F.C. Chu, E. Šatović-Vukšić, M. Plohl, T.P.L. Smith, S. Koren, M.L. Olmstead, D. Leierer, G. Ragan, J.S. Johnston, The genome of the yellow mealworm, *Tenebrio molitor*: it's bigger than you think. *Genes* (Basel), 14(12), prosinac 2023, 2209, <https://doi.org/10.3390/genes14122209>.

- [21] F.J. Ruiz-Ruano, M.D. López-León, J. Cabrero, J.P.M. Camacho, High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific Reports*, 6, srpanj 2016, 28333, <https://doi.org/10.1038/srep28333>.
- [22] P. Novák, P. Neumann, J. Macas, Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2, *Nature Protocols*, 15(11), studeni 2020, 3745–3776. <https://doi.org/10.1038/s41596-020-0400-Y>.
- [23] E. Šatović-Vukšić, M. Plohl, Satellite DNAs-From Localized to Highly Dispersed Genome Components, *Genes (Basel)*, 14(3), ožujak 2023, 742. <https://doi.org/10.3390/genes14030742>.
- [24] M. Tunjić-Cvitanić, J.J. Pasantes, D. García-Souto, T. Cvitanić, M. Plohl, E. Šatović-Vukšić, Satellitome analysis of the Pacific Oyster *Crassostrea gigas* reveals new pattern of satellite DNA organization, highly scattered across the genome, *Int. J. Mol. Sci.*, 22 (13), lipanj 2021, 6798, <https://doi.org/10.3390/ijms22136798>.
- [25] E. Šatović-Vukšić, M. Plohl, Classification Problems of Repetitive DNA Sequences. *DNA*, 1, 2021, 84–90, <https://doi.org/10.3390/dna1020009>.
- [26] E.L. van Dijk, Y. Jaszczyszyn, D. Naquin, C. Thermes, The third revolution in sequencing technology, *Trends Genet.*, 34(9), rujun 2018, 666–681, <https://doi.org/10.1016/j.tig.2018.05.008>.
- [27] S. Aganezov, S.M. Yan, D.C. Soto, M. Kirsche, S. Zarate i sur., A complete reference genome improves analysis of human genetic variation, *Science*, 376(6588), travanj 2022, eabl3533, <https://doi.org/10.1126/science>.
- [28] F.J. Sedlazeck, H. Lee, C.A. Darby, M.C. Schatz, Piercing the dark matter: bioinformatics of long-range sequencing and mapping, *Nature Rev. Genet.*, 19, lipanj 2018, 329–346, <https://doi.org/10.1038/s41576-018-0003-4>.
- [29] T. Vondrak, L. Oliveira, P. Novák, A. Koblíž, P. Neumann, J. Macas, Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads. *Comput. Struct. Biotechnol. J.*, 19(2), travanj 2021, 2179–2189, <https://doi.org/10.1016/j.csbj.2021.04.011>.
- [30] S. Henikoff, K. Ahmad, H.S. Malik, The centromere paradox: stable inheritance with rapidly evolving DNA, *Science* 293(5532), kolovoz 2001, 1098–1102, <https://doi.org/10.1126/science.1062939>.
- [31] M. Lopes, S. Louzada, M. Gama-Carvalho, R. Chaves, Pericentromeric satellite RNAs as flexible protein partners in the regulation of nuclear structure, *Wiley Interdiscip Rev RNA*, 15(4), srpanj-kolovoz 2024, e1868, <https://doi.org/10.1002/wrna.1868>.
- [32] D.C. Cabral-de-Mello, O.M. Palacios-Gimenez, Repetitive DNAs: the 'invisible' regulators of insect adaptation and speciation, *Curr. Opin. Insect. Sci.*, 67:101295, veljača 2025, <https://doi.org/10.1016/j.cois.2024.101295>.
- [33] E.S. Ershova, O.N. Agafonova, N.V. Zakharova, L.V. Bravve, E.M. Jestkova, V.E. Golimbet, T.V. Lezheiko, A.Y. Morozova, A.V. Martynov, R.V. Veiko, P.E. Umriukhin, G.P. Kostyuk, S.I. Kutsev, N.N. Veiko, S.V. Kostyuk, Copy number variation of satellite III (1q12) in patients with schizophrenia, *Front Genet.*, 10, studeni 2019, 1132, <https://doi.org/10.3389/fgene.2019.01132>.
- [34] S. Louzada, M. Lopes, D. Ferreira, F. Adegá, A. Escudeiro, M. Gama-Carvalho, R. Chaves, Decoding the role of satellite DNA in genome architecture and plasticity—an evolutionary and clinical affair, *Genes*, 11(1), siječanj 2020, 72, <https://doi.org/10.3390/genes11010072>.
- [35] J.M. Sadino, Z.R. Donaldson, Prairie voles as a model for understanding the genetic and epigenetic regulation of attachment behaviors, *ACS Chem Neurosci.*, 9(8), kolovoz 2018, 1939–1950, <https://doi.org/10.1021/acchemneuro.7b00475>.

DEVELOPMENT OF DNA SEQUENCING METHODOLOGY AND EXPLORATION OF SATELLITE DNAS

Abstract: Sequences of deoxyribonucleic acid (DNA) molecule, repeated consecutively in a large number of copies in genomes of plants, animals and fungi are called satellite DNAs. Research of the genomic content and exact composition of satellite DNAs is mostly limited by sequencing techniques, i.e. by 'reading' the DNA molecule. The reason for this is great similarity among the repeat units of a satellite DNA, which represents difficulty in overlapping

the sequenced segments, their correct placement in relation to other DNA sequences in the genome and, ultimately, in uninterrupted assembly of chromosomal and genomic DNA of an organism. With the advancements in sequencing techniques and specialized computational approaches, the concepts of understanding the number, arrangement, evolution and roles of satellite DNAs have changed dramatically. The first DNA sequencing in Croatia, almost 40 years ago, was associated with the work on satellite DNAs of the yellow mealworm *Tenebrio molitor* and related species. This presentation shows the liaisons between advancements in sequencing techniques and satellite DNA research, thanks to which we changed our view on these sequences, from genomic ballast to important functional components.

Miroslav Plohl