Segmenti mitohondrijske DNA u genomu stanične jezgre goveda

Andjel, Lucija

Master's thesis / Diplomski rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Agriculture / Sveučilište u Zagrebu, Agronomski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:204:777132

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

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



Repository / Repozitorij:

Repository Faculty of Agriculture University of Zagreb







Sveučilište u Zagrebu
Agronomski fakultet
University of Zagreb
Faculty of Agriculture

SEGMENTI MITOHONDRIJSKE DNA U GENOMU STANIČNE JEZGRE GOVEDA

DIPLOMSKI RAD

Lucija Andjel

Zagreb, srpanj, 2019.

Sveučilište u Zagrebu
Agronomski fakultet
University of Zagreb
Faculty of Agriculture

Diplomski studij:
Genetika i oplemenjivanje životinja

SEGMENTI MITOHONDRIJSKE DNA U GENOMU STANIČNE JEZGRE GOVEDA

DIPLOMSKI RAD

Lucija Andjel

Mentor:

prof. dr. sc. Ino Čurik

Zagreb, srpanj, 2019.

Sveučilište u Zagrebu
Agronomski fakultet

University of Zagreb
Faculty of Agriculture

IZJAVA STUDENTA
O AKADEMSKOJ ČESTITOSTI

Ja, **Lucija Andjel** JMBAG 0178097149, rođen/a dana 27.11.1994. u Zagrebu, izjavljujem da sam samostalno izradila/izradio diplomski rad pod naslovom:

SEGMENTI MITOHONDRIJSKE DNA U GENOMU STANIČNE JEZGRE GOVEDA

Svojim potpisom jamčim:

- da sam jedina autorica/jedini autor ovoga diplomskog rada;
- da su svi korišteni izvori literature, kako objavljeni tako i neobjavljeni, adekvatno citirani ili parafrazirani, te popisani u literaturi na kraju rada;
- da ovaj diplomski rad ne sadrži dijelove radova predanih na Agronomskom fakultetu ili drugim ustanovama visokog obrazovanja radi završetka sveučilišnog ili stručnog studija;
- da je elektronička verzija ovoga diplomskog rada identična tiskanoj koju je odobrio mentor;
- da sam upoznata/upoznat s odredbama Etičkog kodeksa Sveučilišta u Zagrebu (Čl. 19).

U Zagrebu, dana _____

_____
Potpis studenta / studentice

Sveučilište u Zagrebu
Agronomski fakultet

University of Zagreb
Faculty of Agriculture

IZVJEŠĆE

O OCJENI I OBRANI DIPLOMSKOG RADA

Diplomski rad studenta/ice **Lucije Andjel**, JMBAG 0178097149, naslova

SEGMENTI MITOHONDRIJSKE DNA U GENOMU STANIČNE JEZGRE GOVEDA

obranjen je i ocijenjen ocjenom _____ , dana _____ .

Povjerenstvo:			potpisi:
1.	prof. dr. sc. Ino Čurik	mentor	_____
2.	izv. prof. dr. sc. Vlatka Čubrić Čurik	član	_____
3.	doc.dr.sc. Maja Ferenčaković	član	_____

Zahvala

Ovime zahvaljujem:

- Prof. dr. sc. Ini Čuriku što mi je kroz ove tri godine bio i profesor i mentor, kako u završnome, tako i sada u diplomskome radu, što je vidio potencijal u meni, bodrio ga i pomogao omogućiti moj napredak u struci i znanosti te na pomoći, strpljenju i razumijevanju prilikom izrade ovog rada,
- Izv. prof. dr. sc. Vlatki Čubrić-Čurik kao profesorici i članici povjerenstva, koja je također uvidjela moju sposobnost i želju za radom te na savjetima i moralnoj podršci tijekom studiranja,
- Univ. prof. dipl.-ing. dr. rer. nat. Johannu Sölkneru i Assoc. prof. dr. Gaboru Meszarosu (BOKU, Beč) što su me prihvatili na studiranje jedan semestar tijekom Erasmus+ razmjene te na pomoći i savjetima pri započinjanju istraživanja za diplomski rad,
- dr. sc. Vladimiru Brajkoviću za pomoć i savjete tijekom izrade ovog rada i studija općenito,
- dr. sc. Dinku Novoselu, dr. vet. med. za pomoć i savjete tijekom izrade ovog rada,
- doc. dr. sc. Maji Ferenčaković kao članici povjerenstva te pomoći i savjetima tijekom studija,
- svojim prijateljima koji su mi obogatili život, kako van studija, tako i tijekom,
- svojoj baki Bari koja mi je također i najbolja prijateljica,
- svom psu Donni koja je kroz cijeli svoj život bila neiscrpan izvor ljubavi i sreće
- i svojim roditeljima na daru života; na ljubavi, brizi, potpori te omogućavanju cijelog dosadašnjeg školovanja do završetka studija.

Sadržaj

1. Uvod i pregled literature 1
   1.1. Hipoteza 2
   1.2. Cilj istraživanja 2
2. Materijali i metode 3
   2.1. Sekvence 3
   2.2. BLAST softver 4
      2.2.1. BLAST+ 5
         2.2.2.1. Output 6
   2.3. MEGA softver 7
3. Rezultati i rasprava 8
   3.1. Referentne sekvence 8
      3.1.1. Podudaranje referentnih kromosoma s referentom mtDNA 10
         3.1.1.1. Kromosomi Bos taurusa u usporedbi s mtDNA Bos taurusa i Bos indicusa .. 10
         3.1.1.2. Kromosomi Bos indicusa u usporedbi s mtDNA Bos taurusa i Bos indicusa 10
   3.2. MitoTAUROmics sekvence 11
      3.2.1. *No hits* rezultati 11
      3.2.2. *Numt*-ovi 11
4. Zaključak 25
5. Popis literature 26

Životopis 31

Sažetak

Diplomskog rada studenta/ice **Lucije Andjel**, naslova

SEGMENTI MITOHONDRIJSKE DNA U GENOMU STANIČNE JEZGRE GOVEDA

U mnogim vrstama životinja (i ljudi) otkrivene su stotine sekvenci koje potječu od insercija mtDNA u nuklearni genom. Danas se te sekvence zovu *numt*-ovi (nuklearne mitohondrijske DNA) ili mitohondrijski pseudogeni. Međutim, distribucija *numt*-ova kod goveda se tek odnedavno počela istraživati. Cilj rada bio je identificirati i istražiti učestalost pojave segmenata mitohondrijske DNA u genomu jezgre goveda. Za istraživanje su korištene 33 sekvence iz projekta MitoTAUROmics te referentni kromosomi i mtDNA *Bos taurusa* i *Bos indicusa* koji su dostupni u javnim bazama podataka. Usporedbe su se radile uz pomoć BLAST+ softvera, a poravnanje sekvenci se radilo pomoću MEGA 7.0.26. softvera. Pronađeno je ukupno 156 potencijalnih *numt*-ova, od kojih je 86 potvrđeno. Iako su više kromosoma *Bos indicusa* imali podudaranja sa traženim sekvencama, sama duljina segmenata je bila veća kod *Bos taurusa*. Najduži *numt* ima 6393 pb, dok najkraći 27 pb. Podudarnost između numt-ova i mtDNA varira od 70% do 100%. Ukupna duljina fragmenata kod *Bos taurusa* je bila 57189 pb, što je 3,5 puta dulje od ukupne duljine mitogenoma. Duljina fragmenata *Bos indicusa* je 30609 pb, što je 1,9 puta dulje od ukupne duljine mitogenoma. U mtDNA *Bos taurusa* identificirano je 5 potpunih gena koji kodiraju proteine, a to su 12s rRNA, COX1, COX2, COX3, ATP6 koji su pronađeni u 9 *numt*-ova. Kod *Bos indicusa* identificirano je 6 potpunih gena: ND3, ND6, ATP6, ATP8, COX3, ND4L u 5 *numt*-ova. Većina ostalih *numt*-ova kodira mt-tRNA gene. Iako su otkriveni geni koji kodiraju proteine, potrebna su daljnja istraživanja radi provjere funkcionalnosti *numt*-ova koji kodiraju mt-tRNA molekule.

Ključne riječi: mtDNA, *numt*, sekvenca, *Bos taurus*, *Bos indicus*

Summary

Of the master's thesis – student **Lucija Andjel,** entitled

MITOCHONDRIAL DNA SEGMENTS IN CATTLE NUCLEAR GENOME

In many animal species (and humans), hundreds of sequences, which originated from the insertion of mtDNA into the nuclear genome, have been discovered. Today, these sequences are called numts (nuclear mitochondrial DNAs) or mitochondrial pseudogenes. However, the distribution of cattle numts has only recently begun to be investigated. The aim of the study was to identify and investigate the incidence of mitochondrial DNA segments in the cattle nuclear genome. For this study 33 mitochondrial sequences, which were available from the MitoTAUROmics project and reference chromosomes and mtDNA of *Bos taurus* and *Bos indicus* that are available in public databases, were used. Comparisons were made with BLAST+ software, and alignment of sequences was performed using MEGA 7.0.26. software. A total of 156 potential numts were found, of which 86 were confirmed. Although more chromosomes of *Bos indicus* had matched sequences, the segment length itself was larger in *Bos taurus*. The longest numt was 6393 bp long, while the shortest was 27 bp. The correspondence between numt and mtDNA varies from 70% to 100%. The total length of fragments in *Bos taurus* was 57189 bp, which is 3.5 times longer than the total length of the mitogenome. The length of the *Bos indicus*'s segments fragment is 30609 pb, which is 1.9 times longer than the total length of the mitogenome. In the mtDNA of *Bos taurus*, 5 full genes coding for proteins were identified, namely 12s rRNA, COX1, COX2, COX3, ATP6, and they were found in 9 numts. In total, 6 full genes: ND3, ND6, ATP6, ATP8, COX3, ND4L in 5 numts were identified with mtDNA of *Bos indicus*. Most other numts encode the mt-tRNA genes. Although protein-encoding genes have been discovered, further investigations are required to verify the functionality of numts encoding mt-tRNA molecules.

Keywords: mtDNA, *numts*, sequence, *Bos taurus, Bos indicus*

1. Uvod i pregled literature

Endosimbioza je bitno povezana s evolucijom eukariota, pa su zbog toga transferi gena od organela do jezgre važan mehanizam genetske varijacije koji je pomogao u stvaranju tranzicije prokariota do eukariota (Gould i sur., 2008). Iako DNA može biti eksperimentalno preseljena iz organela u jezgru u laboratoriju (Ricchetti i sur., 1999), dalekosežniji eksperiment je onaj koji se događa u prirodi tijekom evolucijskog vremena. Sve sekvence genoma od eukariota koji imaju DNA u njihovim mitohondrijima sadrže dokaze za tekući proces prijenosa organela do nuklearne DNA u obliku nuklearnih kopija mitohondrijske i, u slučaju biljaka, kloroplastne DNA (Hazkani-Covo i sur., 2010). U većini životinjskih vrsta, uključujući ljude, mitohondrijski genom (mtDNA) naslijeđen je strogo od majke (Ladoukakis and Zouros, 2017).

Stotine sekvenci DNA koje su homogene za fragmente mtDNA, često nazivane nuklearnim mitohondrijskim DNA (*numt*-ovi) ili mitohondrijskim pseudogenima (Lopez i sur., 1994), potvrđene su u nuklearnim genomima. *Numt*-ovi su identificirani u brojnim vrstama od njihovog prvog otkrića (Du Buy i Riley, 1967), a sve veći broj istraživača sada se usredotočuje na njihovu evoluciju, moguću funkciju te utjecaj u istraživanjima mitohondrijskih gena (Ricchetti i sur., 2004; Thalmann i sur., 2004; Schmitz i sur., 2005).

Umetanje mtDNA fragmenata u nuklearni genom je neprekidan i više-manje kontinuirani proces kod kvasca (Ricchetti i sur., 1999) i primata (Ricchetti i sur., 2004), s brojem umetanja *numt*-ova kod primata od približno 5,6 *numt*-ova u milijun godina (Gherman i sur., 2007; Hazkani-Covo 2009). Budući da se za većinu *numt*-ova smatra da proizlaze iz pojedinačnih događaja insercija, isključujući mogućnost sličnosti zbog konvergentne evolucije, mogu se stoga smatrati lokusima bez molekularne homoplazmije (Zischler 2000, Hazkani-Covo 2009), što ih čini vrlo dobrim kandidatima za pouzdane filogenetske markere. Ovisno o mjestu umetanja, *numt* može ometati funkciju gena (Gaziev i Shaikhaev, 2010). Osim toga, *de novo* integracija *numt* pseudogena u nuklearni genom ima negativan učinak u nekim slučajevima, promovirajući različite poremećaje i starenje (Dayama i sur., 2014).

U području genetike životinja, sve je više studija koje se odnose na polimorfizam mtDNA i genetsku varijaciju. Zbog njihove homologije s komplementarnim fragmentom mtDNA, *numt*-ovi bi mogli značajno utjecati na rezultate analiza koji uključuju lančanu reakciju polimerazom (PCR) i mogli bi čak dovesti do potpuno pogrešnog zaključka (Wallace i sur., 1997). Dobar primjer slučaja je argument o odnosu Alzheimerove bolesti i polimorfizama mitohondrijske DNA (Davis i sur., 1997; Hirano i sur., 1997; Wallace i sur., 1997).

Kod ljudi je otkriveno da se cjelokupni mitohondrijski genom može naći oko 600 različitih fragmenata, očito distribuiran nasumce preko ljudskog nuklearnog genoma i najvjerojatnije umetnut od strane nehomolognog spajanja DNA krajeva (NHEJ) tijekom popravka dvostrukih prekida DNA (DSBs) (Hazkani-Covo i Covo, 2008, Lang i sur., 2011). *Numt*-ovi se mogu izgubiti ili se događaji insercija mogu pojaviti od *de novo* dolaska DNA iz mitohondrija ili ovisiti o dupliciranju koje uključuju već fiksirane fragmente, stvarajući tako složenije preuređivanje u cijelom nuklearnom genomu (Woischnik i Moraes, 2002; Triant i DeWoody, 2007). Neki izvještaji sugeriraju da se insercije *numt*-ova nastoje kolokalizirati s ponavljajućim

elementima (Tsuji i sur., 2012; Mishmar i sur., 2004), dok su drugi ustanovili da su nedovoljno zastupljeni (Gherman i sur., 2007). Dodatno, postoje dokazi da se *numt*-ovi radije umeću u otvorene kromatinske regije, tipično blizu A+T oligomernih sekvenci (Tsuji i sur., 2012). Budući da su ove studije prvenstveno utemeljene na starijim, fiksnim umetanjima u ljudsko podrijetlo, moguće je da su oni bili zbunjeni evolucijskim mutacijskim procesima koji su se dogodili od vremena tih umetanja (Dayama i sur., 2014).

Drugi važan aspekt *numt*-ova je njihov potencijalni učinak na studije mitohondrijske heteroplazmije, koje su razlike u razini stanica ili tkiva u pojedinim mitohondrijskim genomima zbog dijela visoke stope mutacije unutar tih sekvenci (Meyer i sur., 2014). Niske razine heteroplazmija su tipične za zdrave pojedince, a nedavna izvješća pokazala su da svaka osoba nosi između 1 i 14 heteroplazmija. Međutim, veće razine heteroplazmije su bile uključene u starenje i različite bolesti kao što su Leberova nasljedna optička neuropatija, dijabetes, gluhoća pa čak i rak. Prisutnost *numt*-ova može zbuniti proučavanje i dijagnosticiranje tih bolesti kroz pogrešnu identifikaciju nuklearno-specifične mutacije *numt*-ova kao heteroplazmije. Razvijeni su računski i molekularni pristupi kako bi se smanjio njihov učinak na te studije, ali koriste samo poznate *numt*-ove koji su već prisutni u referentnom slijedu i ne uzimaju u obzir nedavne insercije koje mogu još uvijek prevladavati u značajnom dijelu populacije (Dayama i sur., 2014).

1.1. Hipoteza

Null hipoteza (H_0) – u nuklearnom genomu postoje segmenti mitohondrijske DNA. Alternativna hipoteza (H_1) – u nuklearnom genomu ne postoje segmenti mitohondrijske DNA. S obzirom da točni mehanizmi, pomoću kojih su mtDNA sekvence integrirane u nuklearni genom, još uvijek nisu u potpunosti shvaćeni, ove su hipoteze postavljene sa svrhom boljeg poznavanja evolucije genoma.

1.2. Cilj istraživanja

Cilj rada je identificirati i istražiti učestalost pojave segmenata mitohondrijske DNA u genomu jezgre goveda. Analiza u ovom diplomskom radu temelji se na usporedbi segmenata mtDNA dostupnih iz projekta MitoTAUROmics (voditelj: prof. dr. sc. Ino Čurik) s genomom jezgre goveda (dostupne sekvence cijelog genoma goveda u javnim bazama podataka).

2. Materijali i metode

2.1. Sekvence

Za istraživanje su korištene sekvence DNA goveda. Kao referentne sekvence, koristila sam 29 autosoma *Bos taurusa* (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1) i *Bos indicusa* (https://www.ncbi.nlm.nih.gov/assembly/GCF_000247795.1), kao i njihove spolne kromosome i mtDNA koji su dostupni u javnim bazama podataka. Mitohondrijske sekvence koje su se koristile za usporedbu su uzete iz projekta MitoTAUROmics koji je bio pod vodstvom prof. dr. sc. Ine Čurika (http://mitotauromics.agr.hr). Iz tog projekta uzete su 33 sekvence koje su služile kao primjer skoro svih haplotipova goveda (Tablica 1). Od te 33 sekvence, 29 sekvenci su *Bos taurus*, dvije su *Bos indicus* i dvije *Bos primigenius*.

Tablica 1. Prikazuje informacije o sekvencama iz projekta MitoTAUROmics koje su se koristile za usporedbu. V00654 je referentna mtDNA *Bos taurusa* prema kojoj su ostale poravnane. Accession_number – pristupni broj korištenih sekvenci, Breed_code – kod pasmine (https://www.gov.uk/guidance/official-cattle-breeds-and-codes), ISO alpha-3 counry code – ISO alpha-3 kod države iz koje je sekvencirana životinja (https://www.iban.com/country-codes) i Haplotype – haplotip životinje/sekvence. Sivom bojom su označene sekvence *Bos indicusa*, a svijetlo sivom *Bos primigeniusa*. Ostale sekvence su *Bos taurus.*

Accession_number	Breed_code	ISO alpha-3 counry code	Haplotype
V00654	BRS		T3r
KF163063	NGU	ZMB	T1b
KF163065	NGU	ZMB	T1b
KF163066	NGU	ZMB	T1b
KT184462	DAI	EGY	T1a
JN817298	ABI	ETH	T1d
EU177850	CAB	ITA	T2
EU177851	CAB	ITA	T2d
KT184458	MEN	EGY	T2
KT184459	MEN	EGY	T2
EU177852	GRE+	GRC	T2d
KJ709682	GRH	DEU	T3r
AY676861	CHA	FRA	T3d

Tablica 1. nastavak

Accession_number	Breed_code	ISO alpha-3 counry code	Haplotype
GQ129207	HGY	HUN	T3r
EU177823	CIN	ITA	T3i
AY676857	ABA	GBR	T3n
HQ184045	MCG	ITA	R2
FJ971087	ROM	ITA	R2
FJ971084	AGR	ITA	R1a
HQ184043	ROM	ITA	R1
HQ184044	ROM	ITA	R1
HQ184036	GRA	ITA	Q1
HQ184037	GRA	ITA	Q1
FJ971081	CHI	ITA	Q2
KP637147	PIR	ESP	Q1
KT184471	DAI	EGY	Q1
EU177870	IRN+	IRN	I2
EU177868	IRQ+	IRQ	I1
GU256940	DIQ	CHN	I1
NC_005971	NEL	BRA	I1
FJ971088	TAU	MNG	I1
NC_013996	PRI	GBR	P1
DQ124389	TAU	KOR	P2
JQ437479	PRI	POL	P1

2.2. BLAST softver

Za pronalazak i usporedbu mtDNA s nuklearnom DNA, korišten je BLAST softver (https://www.ncbi.nlm.nih.gov/books/NBK279690/). BLAST je osnovni lokalni alat za pronalaženje područja podudaranja koji nalazi područja lokalne sličnosti između sekvenci. Program uspoređuje sekvence nukleotida ili proteina s bazama podataka sekvenci i izračunava statističku značajnost podudaranja. BLAST se može koristiti za zaključivanje funkcionalnih i evolucijskih odnosa između sekvenci, kao i za pomoć pri identifikaciji članova obitelji gena (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.2.1. BLAST+

NCBI pruža skup alata naredbenog retka za pokretanje BLAST-a pod nazivom BLAST+. To korisnicima omogućuje izvršavanje BLAST pretraživanja na vlastitom poslužitelju bez ograničenja veličine, volumena i baze podataka. BLAST+ se može koristiti s komandnom linijom tako da se može integrirati izravno u tijek rada (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download).

Primjer usporedbe sekvenci miša (*Mus musculus*) i ribe zebrice (*Danio rerio*) se može koristiti za provjeru funkcionalnosti kodova iz zadanog priručnika (https://angus.readthedocs.io/en/2018/running-command-line-blast.html). Treba obratiti pažnju kod usporedbi sekvenci jesu li one proteinske ili nukleotidne jer BLAST+ ima različite komande za različite vrste sekvenci (https://www.ncbi.nlm.nih.gov/books/NBK279668/, https://www.ncbi.nlm.nih.gov/books/NBK279684/#appendices.Options_for_the_commandline_a, vizualni primjer: https://www.ncbi.nlm.nih.gov/books/NBK279682/).

Prethodno spomenute i *downloadane* sekvence moraju biti u istom *folderu* kao i BLAST+ softver kako bi se mogle koristiti. Softver razlikuje dvije vrste sekvenci: *query*, odnosno onu sekvencu koju uspoređujemo (u ovom slučaju 33 sekvence iz projekta MitoTAUROmics, Tablica 1) i *subject* sekvencu, odnosno onu s kojom uspoređujemo (u ovom slučaju referentni kromosomi *Bos taurusa* i *Bos indicusa*). Postoje određene razlike između *query* i *subject* sekvenci. Prva je da datoteka *subject* sekvence može biti jedino *.fas* formata, dok *query* može biti u više njih (primjerice *.txt, .fas, .rtf* itd.). Druga je da *subject* sekvenca mora imati prethodno napravljenu databazu kako bi BLAST+ nju mogao prepoznati kao *subject* sekvencu (pobliže objašnjeno u poglavlju 2.2.2. Komandne linije).

Da bi BLAST+ prepoznao sekvencu ili njen segment kao valjane, minimalni broj nukleotida koje ona mora imati je 30, ispod toga ne dobivamo nikakve rezultate podudaranja. Također, *defaultno* podudaranje je zadano na 70%, ispod toga isto nema rezultata, no to se može modelirati, ovisno o željenom pragu.

2.2.2. Komandne linije

U vlastitom poslužitelju, u ovom slučaju *Bash* terminalu na OS operativnom sustavu, prvo treba biti u folderu u kojem želimo raditi usporedbe. To se radi pomoću komande *cd* koja znači „*change directory*“, odnosno promjena radnog direktorija do željenog. Nakon toga se radi databaza *subject* sekvence. U slučaju da se ne koristi cijela sekvenca, nego samo njen segment, s naredbom *head -n,* gdje *-n* označava broj redova nukleotida koje želimo vidjeti, vidimo samo taj željeni segment. Ista stvar se može postići i s naredbama *cat* i *less.*

U idućem koraku se radi databaza *subject* sekvence. Primjer koda:

```
./makeblastdb -in chr1.fas -dbtype nucl
```

U ovoj komandnoj liniji „./“ označava da želimo izvršiti određenu radnju u trenutnom direktoriju; služi za Unix, Linux i OS sustave, kod Windowsa se ne koristi. Nadalje, `makeblastdb -in` označava da želimo napraviti databazu za određenu sekvencu koja je imenovana nakon i `-dbtype nucl` označava da je ta sekvenca nukleotid. U slučaju da je sekvenca proteinska, piše se `-dbtype prot`. Ako je kod uspješan, u radnom direktoriju se dobiju tri nova *filea* koji imaju isti naziv kao sekvenca, ali ekstenzije su *.nhr, .nin* i *.nsq*. Oni su blastn *database fileovi*. Nakon što je databaza napravljena, može se raditi usporedba sekvenci.

2.2.2.1. Output

Ovisno o kvantiteti traženih informacija koje želimo dobiti u rezultatima, *output* može biti tekstualnog ili tabularnog formata. Shodno, komandne linije su različite.

Primjer linije koda za tekstualni ispis rezultata:

```
192:~ lucijaandjel$ ./blastn -query chr19.fasta.txt -db Dominette_wgs_mtDNA.fas
-out tauchr1.txt
```

`blastn` označava BLAST+-u da radimo s nukleotidnim sekvencama, s naredbom `-query` imenujemo *query* sekvencu, a `-db` *subject* sekvencu, odnosno databazu. `-out` označava da želimo output ove usporedbe i u ovom slučaju je u tekstualnom obliku, `tauchr1.txt` (slika 1a. i 1b.). U slučaju da imamo proteinsku sekvencu, koristi se `blastp`.

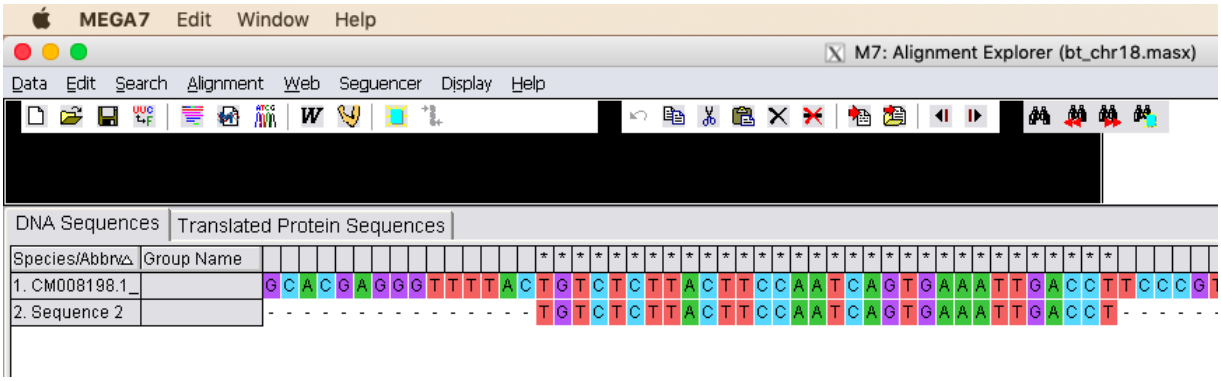
Linija koda za tabularni ispis rezultata:

```
192:~ lucijaandjel$ ./blastn -query Set_1_preim_al_bizon_pr.fas -db chr1.fasta.txt
-outfmt "6 qseqid sseqid qlen slen pident nident length qstart qend sstart send
mismatch gaps evalue bitscore score sstrand" -out tauchr1.xls
```

U ovoj komandnoj liniji naredba `-outfmt 6` označava da želimo da nam *output* bude tablični (https://www.ncbi.nlm.nih.gov/books/NBK279684/). Ostatak naredbi u navodnim znacima označava koje sve informacije točno želimo imati u samoj tablici i kojim redoslijedom želimo da one budu, a ovdje su to redom: (*qseqid*) naziv *query* sekvence, (*sseqid*) naziv *subject* sekvence, (*qlen*) dužina *query* sekvence, (*slen*) dužina *subject* sekvence, (*pident*) postotak identičnih podudaranja baza, (*nident*) broj identičnih baza, (*length*) dužina poravnanja, (*qstart*) početno mjesto podudaranja u *query* sekvenci, (*qend*) krajnje mjesto podudaranja u *query* sekvenci, (*sstart*) početno mjesto podudaranja u *subject* sekvenci, (*send*) krajnje mjesto podudaranja u *subject* sekvenci, (*mismatch*) broj nepodudarnih mjesta unutar poravnanja, (*gaps*) broj praznina uzrokovanih adicijom ili delecijom, (*evalue) E-value*, (*bitscore*) *bitscore*, (*score*) „sirovi“ rezultat i (*sstrand*) smjer lanca DNA subject sekvence (http://www.metagenomics.wiki/tools/blast/blastn-output-format-6) (Tablica 2.). *Score* je numerički rezultat koji proizlazi iz procjene usklađenosti sekvenci; izračunat je iz zbroja rezultata supstitucija i *gapova* (https://www.tititudorancea.com/z/blast_raw_score.htm). *E-value* je broj očekivanih pogodaka slične kvalitete koji se mogu naći slučajno. *Bitscore* označava potrebnu veličinu baze podataka sekvenci u kojoj se trenutna podudarnost može naći slučajno. (http://www.metagenomics.wiki/tools/blast/evalue).

2.3. MEGA softver

Svaki je *numt* zasebno poravnat na referentnu mtDNA u MEGA 7.0.26. softveru. Zasebno sam ih poravnala kako bih osigurala što veću točnost pri određivanju gena. Usklađivanje sam radila pomoću *Clustal W* opcije u kojoj sam odredila poravnanje DNA te sam u parametrima „*Gap Opening Penalty*“ smanjila na 2 kako bih pronašla što točnije podudaranje. Primjer segmenta kromosoma 18 *Bos taurusa* koji se poklapa s referentnom mtDNA je prikazan na Slici 3. Oni segmenti koji sadrže „ne“ u nazivu su pronađeni samo kod sekvenci haplogrupe I. Ta haplogrupa je tipična za *Bos indicus* pasmine, stoga nije ni čudno da je veći broj i postotak poklapanja s referentnom mtDNA *Bos indicusa.* Xia X. i sur. (2018) u svome radu potvrđuju da sve sekvence zebu goveda spadaju pod haplogrupu I.



Slika 3. Prikazuje rezultat poravnanja segmenta kromosoma 18 *Bos taurusa* koji se poklapa s referentnom mtDNA *Bos taurusa.* Poravnanje je rađeno u MEGA 7.0.26. softveru pomoću *Clustal W* opcije.

3. Rezultati i rasprava

3.1. Referentne sekvence

Prvotno sam usporedila referentne autosome i spolne kromosome *Bos taurusa* i *indicusa* s njihovom mtDNA na način da su mtDNA bile *query* sekvence, a kromosomi su bili *subject* sekvence.

Uspoređujući mtDNA *Bos taurusa* s njegovim kromosomima, nisam dobila nikakva podudaranja na 5, 7, 13, 19, 20, 21, 23, 25, 27. i 29. kromosomu, dok uspoređujući istu mtDNA s kromosomima *Bos indicusa*, nisam dobila nikakva podudaranja na 5, 6, 7, 19, 23, 27. i 29. kromosomu. Jedina razlika je što je kromosom 13 *Bos indicusa* dao rezultat u usporedbi s mtDNA *Bos taurusa* (Slike 1a. i 1b.). Uspoređujući mtDNA *Bos indicusa* s njegovim kromosomima, nisam dobila nikakva podudaranja na 5, 7, 16, 19, 23, 27. i 29. kromosomu. Identičan rezultat sam dobila uspoređujući istu mtDNA s kromosomima *Bos taurusa.*

```
BLASTN 2.7.1+


Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb
Miller (2000), "A greedy algorithm for aligning DNA sequences", J
Comput Biol 2000; 7(1-2):203-14.


Database: chr13.fasta.txt
           1 sequences; 83,472,345 total letters


Query= CM008198.1 Bos taurus isolate L1 Dominette 01449 registration number
42190680 breed Hereford mitochondrion, complete sequence, whole
genome shotgun sequence

Length=16340


***** No hits found *****


Lambda      K        H
    1.33    0.621     1.12

Gapped
Lambda      K        H
    1.28    0.460    0.850

Effective search space used: 1361516946276


  Database: chr13.fasta.txt
    Posted date:  Jan 12, 2019  6:49 PM
  Number of letters in database: 83,472,345
  Number of sequences in database:  1


Matrix: blastn matrix 1 -2
Gap Penalties: Existence: 0, Extension: 2.5
```

Slika 1a.

```
BLASTN 2.7.1+


Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb
Miller (2000), "A greedy algorithm for aligning DNA sequences", J
Comput Biol 2000; 7(1-2):203-14.



Database: chr13.fasta.txt
           1 sequences; 83,472,345 total letters



Query= AY126697.1 Bos indicus breed Nellore mitochondrion, complete genome

Length=16341
                                                                      Score     E
Sequences producing significant alignments:                          (Bits)  Value

  CM008180.2 Bos taurus isolate L1 Dominette 01449 registration n...   143     1e-31


> CM008180.2 Bos taurus isolate L1 Dominette 01449 registration
number 42190680 breed Hereford chromosome 13, whole genome
shotgun sequence
Length=83472345

 Score = 143 bits (77),  Expect = 1e-31
 Identities = 275/365 (75%), Gaps = 35/365 (10%)
 Strand=Plus/Plus

Query  1778      AGTAGAGGTGACATGCCTAACGAGCCTGGTGATAGCTGGTTGTCCA-GAAGATGAATCTA  1836
                  || |||||||||| |||| || |||||||||||| |||  |||||| |||  | |||  ||
Sbjct  11595647  AGAAGAGGTGACAAGCCTCACAAGCCTGGTGATATCTGACTGTCCAAGAA-AAGAAACTT  11595705

Query  1837      AGTTCAGCTTTAAAGATACCAAAAATACAAATAAATCCCACTGTAGCTTTAAAAGTTAGT  1896
                  |||||| |||     | |  |||| |||||| |||    | | ||  |||  || ||  |
Sbjct  11595706  AGTTCAACTT-----A-A--AAAATTACAAA-AAA----ATTATAAATTT-CAACTT--T  11595749

Query  1897      CTAAAAAGGTACAGCCTTTTAGAAACGGATACAACCTTGACTAGAGAGTAAAATTTAACA  1956
                   || |||||||||| ||||||||| || ||| |||||  ||||||||||||  |||||
Sbjct  11595750  -TACAAAGGTACAGCTTTTTAGAAATGGGTACGACCTTATCTAGAGAGTAAAA-ATAACA  11595807

Query  1957      CTACCATAGTAGGCCTAAAAGCAGCCATCAATTAAGAAAGCGTTAAAGCTCAACA---AC  2013
                   |  |||||  ||| |||||||| || ||||||||||  | || ||||||||||   ||
Sbjct  11595808  -T--CATAGCTGGCTTAAAAGCAGTCACCAATTAAGAAGTCATTCAAGCTCAACAATGAC  11595864

Query  2014      AAAAACTAAATAGATTCCAACAACAA--ATGATTAACTCCTAGCCCCAATACTGGACTAA  2071
                    | || ||||   | |||||| ||  || || |||||||||     | ||||||| |
Sbjct  11595865  TTTATCTTAATA---T-CAACAATAAGCAT-ATCAACTCCTAG-TTTTACACTGGACTGA  11595918

Query  2072      TCTATTATAGAATAGAAGCAATAATGTTAACATGAGTAACAAGAAAAATTTTCTCCTTGC  2131
                  |||||||   || ||||||||||||||||||||||||||||||  | ||||||||||
Sbjct  11595919  TCTATTAGCCAAAAGAAGCAATAATGTTAACATGAGTAACAAGAATTA-TTTCTCCTTGT  11595977

Query  2132      ATAAG  2136
                  |||||
Sbjct  11595978  ATAAG  11595982


Lambda      K        H
    1.33    0.621     1.12

Gapped
Lambda      K        H
    1.28    0.460    0.850

Effective search space used: 1361600418592


  Database: chr13.fasta.txt
    Posted date:  Jan 12, 2019  6:49 PM
  Number of letters in database: 83,472,345
  Number of sequences in database:  1



Matrix: blastn matrix 1 -2
Gap Penalties: Existence: 0, Extension: 2.5
```

Slika 1b.

Slika 1. Prikazuje primjer tekstualnog *outputa*. a) Prikazuje usporedbu mtDNA *Bos taurusa* s 13.-im kromosomom *Bos taurusa* (nema podudaranja; *no hits found*), b) prikazuje usporedbu mtDNA *Bos taurusa* s 13.-im kromosomom *Bos indicusa* (ima podudaranja, 75% *identity*). | označava da su baze između *query* i *subject* sekvence jednake, praznina da je došlo do supstitucije, a minus (-) da je došlo do adicije ili delecije. Brojevi označavaju točnu poziciju početka i kraja segmenta.

3.1.1. Podudaranje referentnih kromosoma s referentom mtDNA

3.1.1.1. Kromosomi *Bos taurusa* u usporedbi s mtDNA *Bos taurusa* i *Bos indicusa*

Prvi kromosom je imao dva bloka koja su se podudarala s mtDNA, jedan od 80% identiteta, tj. podudaranja, a drugi 77%. Drugi kromosom je imao najviše blokova, točnije njih 10, koji imaju identitete između 79% i 88%. Treći kromosom ima samo jedan blok koji ima 89% sličnosti s mtDNA. Četvrti ima četiri bloka koji imaju 78-87% podudaranja. Ovdje je iznimka kod *Bos indicusa* jer postoji i peti blok koji ima 100%-tno podudaranje s mtDNA od 28 baza. Osmi kromosom ima dva bloka koji imaju identitete 86% i 93%, iako je prvi blok veći od drugoga. Deveti kromosom ima šest blokova, prvih pet imaju oko 83% podudaranja, dok zadnji ima 95%. Deseti i jedanaesti imaju također šest blokova. Dvanaesti ima jedan mali blok od 69 baza identičnih od ukupno pronađenih 71, što čini 97% podudaranja s mtDNA. Trinaesti, već spomenuto, razlikuje se jer kod *Bos taurusa* nema nikakvih podudaranja, a kod *Bos indicusa* postoji jedan blok od 275 baza koje čine 75% identiteta od ukupnih 365 pronađenih baza koje se podudaraju. Razliku u identitetu čine adicije, delecije i eventualne supstitucije baza. Slka 1b prikazuje razlike između baza koje se podudaraju i onih koje se ne podudaraju. Četrnaesti kromosom također ima samo jedan blok od oko 260 baza. Petnaesti i sedamnaesti imaju zajedno po 3 manja bloka. Šesnaesti kromosom ima dva bloka, ali prvi blok ima čak preko 10430 identičnih baza, što ga čini najvećim. Kod referentnog šesnaestog kromosoma *Bos indicusa* nije bilo nikakvog podudaranja ni s jednom mtDNA. Osamnaesti ima jedan mali blok od 32 baze koje su 100% identične s bazama mtDNA. Kromosomi 22 i 24 imaju po dva bloka, s time da su blokovi puno veći kod 24. kromosoma, a 26. i 28. imaju oba po jedan blok, jedan od 77%, a drugi od 87% sličnosti. Kromosom X ima tri bloka, no prvi blok ima čak preko 5300 identičnih baza od 6450 koje čine 82% sličnosti.

3.1.1.2. Kromosomi *Bos indicusa* u usporedbi s mtDNA *Bos taurusa* i *Bos indicusa*

Postoje određene razlike kod rezultata kromosoma *Bos indicusa* u odnosu na rezultate *Bos taurusa*. Šesti kromosom ima jedan blok od oko 2040 identičnih baza koje čine 73% podudaranja s mtDNA od ukupnih 2800 pronađenih. Trinaesti kromosom ima jedan blok od 251 identične baze s mtDNA *Bos taurusa*, a dva bloka s mtDNA *Bos indicusa*, od kojih je jedan

blok ta ista 251 baza, a drugi od 274 druge baze. Dvadeseti i 21. kromosom imaju jedan blok od 64 i 91 identičnu bazu koje čine 90% identiteta s mtDNA od pronađenih 71 i 101 baze. Kromosom 25 ima također jedan blok od 116 baza koje su 100% identične s mtDNA *Bos taurusa*, a 112 identičnih od 116 s mtDNA *Bos indicusa*, što čini 97% identiteta.

3.2. MitoTAUROmics *sekvence*

3.2.1. *No hits* rezultati

Kada sam usporedila mitohondrijske sekvence s referentnim kromosomima *Bos taurusa*, nikakve rezultate nisam dobila na istim kromosomima kao i spomenuti u sekciji 3.1. Iznimka su pet sekvenci haplogrupe I (EU177870, EU177868, GU256940, NC_005971 i FJ971088) koje su imale 1 segment poduaranja na 13. kromosomu od 365 baza, od kojih je 275 bilo identičnih sa mito- sekvencama, odnosno segment je imao 75% podudarnosti sa sekvencama (Tablica 2).
Kada sam iste sekvence usporedila s referentnim kromosomima *Bos indicusa*, također isti brojevi kromosoma nisu dali nikakve rezultate kao u prethodno spomenutoj sekciji. U ovom slučaju nije bilo nikakvih iznimaka.

3.2.2. *Numt*-ovi

Na ostalim kromosomima, odnosno onima na kojima sam dobila podudaranja s mito-sekvencama, usporedila sam sličnosti i razlike između onih od *Bos taurusa* i onih od *Bos indicusa*. Usporedila sam broj blokova, odnosno segmenata koje oni imaju, ovisno o broju kromosoma, zatim broj međusobno podudarnih blokova i njihov postotak međusobnog podudaranja (Tablica 3).

Sveukupno je bilo 156 potencijalnih *numt*-ova, od kojih je 63 bilo od *Bos taurusa*, a preostalih 96 od *Bos indicusa*. Iako su kromosomi *Bos indicusa* imali više podudarnih segmenata s mito- sekvencama, oni kod *Bos taurusa* su bili dulji. Razlike u duljini mogu se objasniti međuvrsnim varijacijama u čimbenicima kao što su stabilnost zametne linije i broj mitohondrija (Richly, E., i Leister, D., 2004). U ovom slučaju različitost u duljini *numt*-ova između *Bos taurusa* i *Bos indicusa* nije statistički značajna, s obzirom da su u istraživanju korištene samo dvije *Bos indicus* sekvence. Drugi je kromosom imao najviše podudarnih segmenata, no mali broj njih (točnije samo tri segmenta) su se poklapali među obje vrste. Kromosom 16 kod *Bos taurusa* je imao najveće preklapanje sa sekvencama, čak i više od referentnog kromosoma 16, preko 11700 identičnih baza. S obzirom da kromosom *Bos indicusa* nije dao nikakav rezultat na tom kromosomu, taj segment nije uzet za daljnu analizu. Samo oni segmenti koji su homogeni s mtDNA kod obje vrste su uzeti za daljne razmatranje. Tablice 4 i 5 prikazuju duljinu te mjesto početka i kraja potencijalnih *numt*-ova koji su bili pronađeni unutar sve 33 sekvence kod obje vrste goveda. Razlika između tablica je u tome što su u Tablici

4 prikazani samo *numt*-ovi koji imaju više od 900 baznih parova, dok u Tablici 5 oni koji imaju manje od 900 bp. Ta granica je postavljena prema radu od Liu i Zhao, 2007. Segmenti koji imaju manje od 900 bp su ostavljeni radi provjere uzrokuju li pogreške u određivanju mitogenoma.

Tablica 2. Pokazuje primjer tabularnog *outputa*, u ovom slučaju informacije o pet sekvenci haplogrupe I koje su dale rezultate podudaranja s kromosomom 13 *Bos taurusa*. Značenje kratica iz prve kolone je objašnjeno u poglavlju 2.2.2.1.

qseq ID	sseq ID	qlength	slength	pident	nident	length	qstart	qend	sstart	send	gaps	evalue	score	strand
27_IRN_IRN_I2	CM008180.2	16339	83472345	75,34	275	365	1776	2134	11595647	11595982	35	$9{,}83x10^{-32}$	77	+
28_IRQ_IRQ_I1	CM008180.2	16339	83472345	75,34	275	365	1776	2134	11595647	11595982	35	$9{,}83x10^{-32}$	77	+
29_DIQ_CHN_I1	CM008180.2	16338	83472345	75,34	275	365	1775	2133	11595647	11595982	35	$9{,}83x10^{-32}$	77	+
30_NEL_BRA_I1	CM008180.2	16341	83472345	75,34	275	365	1778	2136	11595647	11595982	35	$9{,}83x10^{-32}$	77	+
31_TAU_MNG_I1	CM008180.2	16339	83472345	75,34	275	365	1776	2134	11595647	11595982	35	$9{,}83x10^{-32}$	77	+

Tablica 3. Prikazuje preklapanje između blokova potencijalnih *numt*-ova *Bos taurusa* s *Bos indicusom. Chr* označava broj kromosoma; *BT* broj blokova kod *Bos taurusa*; *BI* broj blokova kod *Bos indicusa*; *blok overlap* preklapanje između blokova *Bos taurusa* i *Bos indicusa*; *% blok* postotak preklapanja između blokova *Bos taurusa* i *Bos indicusa*; *% BT* postotak preklapanja unutar blokova *Bos taurusa* s obzirom na *Bos indicusa* i *% BI* postotak preklapanja unutar blokova *Bos indicusa* s obzirom na *Bos taurusa*. Blago siva boja označava postotak preklapanja između blokova 70%-79%; siva postotak 80%-89% i tamno siva boja označava postotak >90% između blokova *Bos taurusa* i *Bos indicusa*.

			Overlap			
Chr	**BT**	**BI**	**Blok *overlap***	**% blok**	**% BT**	**% BI**
1.	3	4	3/4	75,00%	55,80%	37,20%
2.	10	14	3/14	21,43%	31,06%	30,71%
3.	1	4	1/4	25,00%	24,00%	24,00%
4.	5	5	4/5	80,00%	100,00%	100,00%
6.	0	1	0/1	0,00%	0,00%	0,00%
8.	2	3	2/3	66,67%	71,00%	71,00%
9.	6	8	3/8	37,50%	81,00%	81,00%
10.	6	10	2/10	20,00%	78,50%	78,50%
11.	6	8	6/8	75,00%	84,35%	83,92%
12.	1	1	1	100,00%	100,00%	100,00%
13.	1	2	1/2	50,00%	100,00%	100,00%
14.	1	1	1/1	100,00%	100,00%	100,00%
15.	3	3	2/3	66,67%	86,03%	83,20%
16.	2	0	0/2	0,00%	0,00%	0,00%
17.	3	3	2/3	66,67%	99,50%	99,50%
18.	1	2	1/2	50,00%	100,00%	100,00%
20.	0	1	0/1	0,00%	0,00%	0,00%
21.	0	1	0/1	0,00%	0,00%	0,00%
22.	2	3	2/3	66,67%	98,00%	98,00%

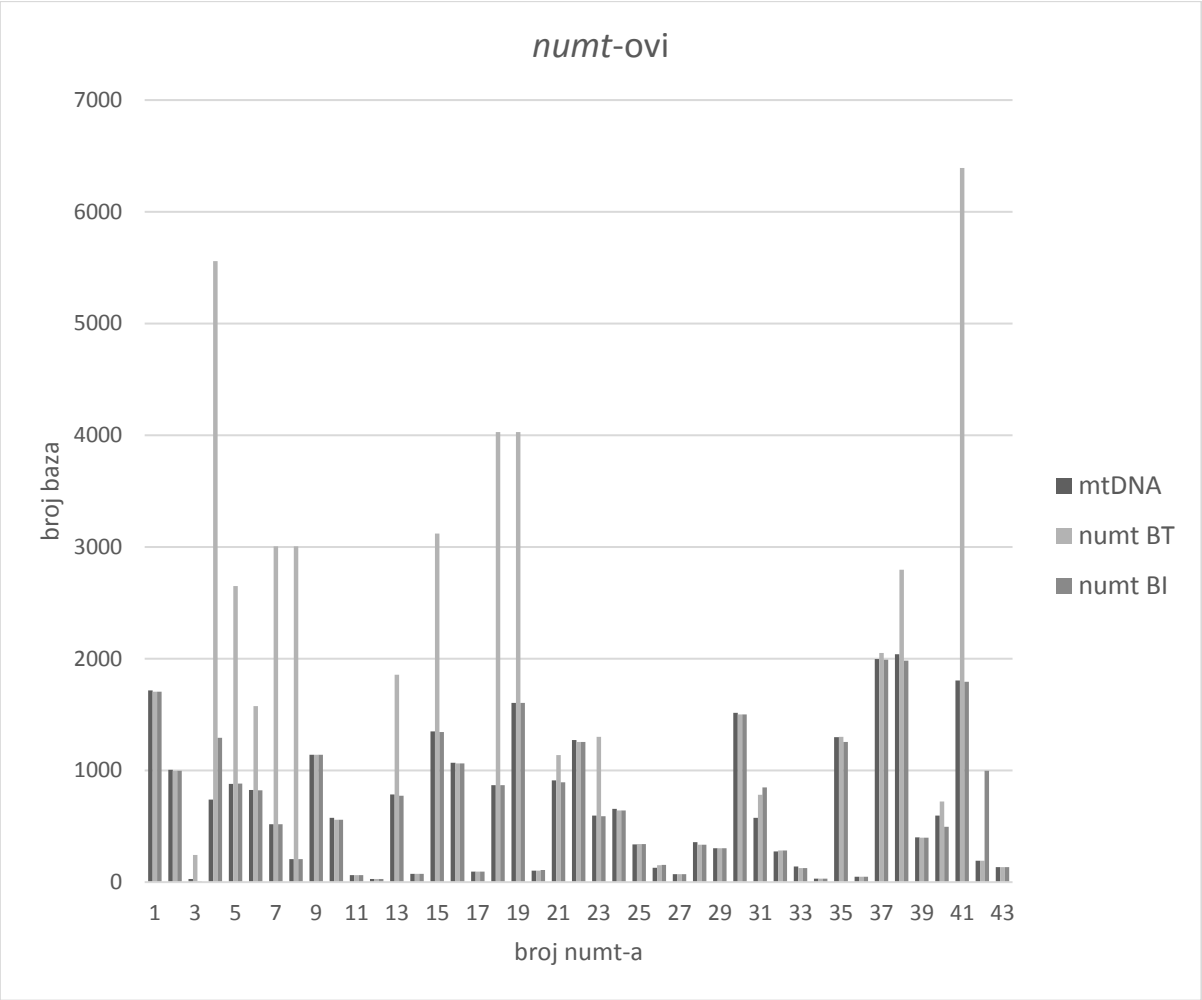
Tablica 3. nastavak

				Overlap		
Chr	**BT**	**BI**	**Blok *overlap***	**Chr**	**BT**	**BI**
24.	2	3	2/3	66,67%	84,00%	84,00%
25.	0	1	0/1	0,00%	0,00%	0,00%
26.	1	3	1/3	33,33%	100,00%	100,00%
28.	3	5	1/5	20,00%	90,27%	98,14%
X	3	4	3/4	75,00%	49,00%	49,00%
Y	-	2	0/2	0,00%	0,00%	0,00%
mtDNA	1	1	1/1	66,67%	84,00%	84,00%

Pronađeno je 43 *numt*-a koji su zajednički kod obje vrste, odnosno 86 s obzirom da se nalaze na različitim lokacijama na samim autosomima. Duljina varira od 27 do 6393 pb. Nadalje, 44 *numt*-a imaju duljinu veću od 900 bp, dok su ostalih 42 kraći od 900 p. Skoro pola, točnije 40 njih, imaju identičnu duljinu kod *Bos taurusa* i *Bos indicusa* te istu ili sličnu duljinu s mito-sekvencama. Iako imaju istu ili sličnu duljinu, *numt*-ovi su na različitim lokacijama u kromosomima *Bos taurusa* i *Bos indicusa*. Grafikon 1 prikazuje da su *numt*-ovi kod *Bos indicusa* u većini slučajeva vrlo slični s duljinom podudarnog segmenta u mtDNA, dok isti segmenti kod *Bos taurusa* mogu biti izrazito dulji.

Tablica 4. Prikazuje dijelove nuklearne DNA *Bos taurusa* i *Bos indicusa* koji su homogeni sa sekvencama iz projekta *MitoTAUROmics* i koji su veći od 900 pb. *Chr No.* označava broj kromosoma; *corrb BT;BI* odgovarajuće blokove *Bos taurusa* naspram *Bos indicusa*; *mtstart* pozicija na mito- sekvencama koje je početna zajednička s *numt*-ovima; *mtend* pozicija na mito-sekvencama koja je zadnja zajednička s numt-ovima; *mtlen* dužinu mito-sekvenci; *snumt BT* početna pozicija *numt*-a kod *Bos taurusa*; *enumt BT* krajnja pozicija *numt*-a kod *Bos taurusa*; *numtlen BT* dužina *numt*-a kod *Bos taurusa*; *snumt BI* početna pozicija *numt*-a kod *Bos indicusa*; *enumt BI* krajnja pozicija *numt*-a kod *Bos indicusa*; *numtlen BI* dužina *numt*-a kod *Bos indicusa*; *E value* označva E vrijednost, *score* „sirovi“ rezultat i *sstrand* smjer lanca subject sekvence. Siva boja označava da su oba *numt*-a iste dužine i iste ili slične dužine s dužinom podudarnog mito- segmenta. Svijetlo siva označava da je jedan od *numt*-ova slične dužine kao i mito- segment. *diff* označava da su lanci *Bos taurusa* i *Bos indicusa* različitog smjera.

Chr No.	corrb BT-BI	mtstart	mtend	mtlen	snumt BT	enumt BT	numtlen BT	snumt BI	enumt BI	numtlen BI	E value	Score	sstrand
1	1-1	383	2098	1715	116542420	116540716	1704	118062463	118060759	1704	0	691,83	-
1	2-2	13497	14503	1006	116548332	116547334	998	118068375	118067377	998	$1,564x10^{-33}$	258,30	-
2	1-1	9585	10323	738	2085710	2091268	5558	3367593	3368884	1291	0	1733,86	+
2	2-3	7652	8532	880	3386219	3388868	2649	3366428	3367309	881	0	933,23	+
2	3-2	1935	2760	825	2098836	2100413	1577	2092395	2093216	821	0	727,98	+
3	1-2	7004	7521	517	90809110	90812116	3006	97313043	97313560	517	0	1186,23	+
3	1-3	9802	10009	207	90809110	90812116	3006	97316144	97316350	206	$3,1814x10^{-72}$	1069,95	+
4	1-1	3517	4657	1140	37326902	37325762	1140	38609804	38608664	1140	0	698,36	-
8	1-1	6372	7157	785	54721065	54722921	1856	57136325	57137100	775	0	729,11	+
9	2-2	3805	5155	1350	4630335	4627214	3121	3978123	3976779	1344	0	1224,24	-
9	4-3	10764	11833	1069	4621110	4620046	1064	3970557	3969493	1064	0	507,61	-
10	1-2	3743	4612	869	5035087	5039115	4028	4582426	4583294	868	0	1850,36	+
10	1-1	5698	7302	1604	5035087	5039115	4028	4583747	4585350	1603	0	1624,71	+
11	1-2	11478	12389	911	45805993	45807129	1136	47578512	47579406	894	8,653x10-^{143}	327,77	+
11	2-1	2601	3874	1273	45788792	45790046	1254	47559159	47560413	1254	4,606x10-^{175}	342,59	+
15	1-1	6514	8030	1516	33197210	33195709	1501	31988682	31987181	1501	$3,274x10^{-65}$	323,00	-
22	1-1	849	2147	1298	47146955	47145654	1301	48234793	48233538	1256	$1,865x10^{-127}$	258,02	-
24	1-1	2845	4841	1996	61001514	61003566	2052	63649175	63651167	1992	0	1044,32	+
24	2-2	8238	10279	2041	43657378	43654582	2796	45245416	45243434	1982	0	533,05	-
28	1-1	5587	6183	596	27869954	27869232	722	27505776	27505282	494	$4,8333x10^{-46}$	408,27	-
X	1-1	9980	11785	1805	75132746	75139139	6393	45161229	45163023	1794	0	1950,50	+
X	2-2	13355	13546	191	108948576	108948384	192	45164623	45165619	996	$8,1315x10^{-50}$	278,64	diff



Grafikon 1. Prikazuje duljinu *numt*-ova u odnosu na odgovarajuću duljinu mito- sekvenci. X os prikazuje broj *numt*-a, dok Y os prikazuje broj podudarnih baznih parova. Tamnosive boje je prikazana duljina mito- sekvenci, svijetlosive duljine *numt*-ova kod *Bos taurusa*, a sive kod *Bos indicusa*.

Tablica 5. Prikazuje dijelove nuklearne DNA *Bos taurusa* i *Bos indicusa* koji su homogeni sa sekvencama iz projekta *MitoTAUROmics* i koji su manji od 900 pBos *Chr No*. označava broj kromosoma; *corrb BT;BI* odgovarajuće blokove *Bos taurusa* naspram Bos indicusa; *mtstart* pozicija na mito- sekvencama koje je početna zajednička s *numt*-ovima; *mtend* pozicija na mito-sekvencama koja je zadnja zajednička s numt-ovima; *mtlen* dužinu mito-sekvenci; *snumt BT* početna pozicija *numt*-a kod *Bos taurusa*; *enumt BT* krajnja pozicija *numt*-a kod *Bos taurusa*; *numtlen BT* dužina *numt*-a kod *Bos taurusa*; *snumt BI* početna pozicija *numt*-a kod *Bos indicusa*; *enumt BI* krajnja pozicija *numt*-a kod *Bos indicusa*; *numtlen BI* dužina *numt*-a kod *Bos indicusa*; *E value* označva E vrijednost, *score* „sirovi“ rezultat i *sstrand* smjer lanca subject sekvence. Siva boja označava da su oba *numt*-a iste dužine i iste ili slične dužine s dužinom podudarnog mito- segmenta. *diff* označava da su lanci *Bos taurusa* i *Bos indicusa* različitog smjera.

Chr No.	corrb BT-BI	mtstart	mtend	mtlen	snumt BT	enumt BT	numtlen BT	snumt BI	enumt BI	numtlen BI	E value	Score	sstrand
1	3-3	8296	8325	29	116556378	116556134	244	127879338	127879332	6	$2x10^{-10}$	63,65	-
4	3-2	2189	2766	577	98713549	98712990	559	102396714	102396155	559	$1,888x10^{-99}$	204,68	-
4	4-3	10486	10550	64	42448143	42448079	64	43909520	43909456	64	$7,361\ x10^{-17}$	51,50	-
4	5-5	12681	12708	27	110666394	110666421	27	114865256	114865283	27	0,0016225	27,00	+
8	2-3	423	498	75	65836070	65836145	75	68818688	68818763	75	6,1872x10-^{23}	62,47	+
9	6-6	390	485	95	20772005	20771910	95	21160851	21160756	95	$3,7328x10^{-34}$	83,00	-
10	6-10	1	105	104	63736058	63736162	104	65554253	65554145	108	$4,7119x10^{-29}$	72,50	diff
11	3-3	4914	5509	595	45791666	45792966	1300	47562033	47562623	590	$6,3477x10^{-89}$	233,17	+
11	4-4	3901	4556	655	45790278	45790919	641	47560645	47561286	641	$2,6777\ x10^{-84}$	174,05	+
11	5-5	12847	13186	339	46424093	46424433	340	48199539	48199879	339	$1,1289x10^{-41}$	104,23	+
11	6-6	12051	12179	128	66151410	66151561	151	66781692	66781847	156	$5,435x10^{-39}$	91,62	+
12	1-1	5541	5611	70	14863352	14863282	70	13847915	13847845	70	$2,4191x10^{-25}$	67,00	-
13	1-2	1776	2134	358	11595647	11595982	335	10050257	10050592	335	$2,3642x10^{-30}$	75,50	+
14	1-1	1331	1634	303	6298255	6297951	304	5663185	5662881	304	$2,6896x10^{-79}$	167,17	-
15	2-2	14368	14943	575	56633947	56634728	781	56430150	56430997	847	$3,5455x10^{-29}$	220,48	+
17	2-1	422	698	276	56664390	56664672	282	59484643	59484926	283	$2,7714x10^{-53}$	118,59	+
17	3-3	5318	5459	141	73028945	73028819	126	76507101	76506975	126	$2,625x10^{-20}$	56,50	-
18	1-2	2451	2482	31	5067400	5067431	31	3947588	3947619	31	$8,04x10^{-07}$	32,00	+
22	2-3	5329	5376	47	6914589	6914542	47	6929632	6929585	47	$2,8444x10^{-09}$	36,50	-
26	1-1	3886	4286	400	31757766	31758163	397	32151301	32151698	397	$5,1695x10^{-56}$	119,91	+
X	3-4	1	135	134	75139140	75139274	134	45167821	45167955	134	$1,374x10^{-44}$	104,52	+

Evolucija eukariotskih genoma pogođena je stjecanjem novih sljedova različitih izvora u nuklearnu DNA. Iako je donedavno bilo znano da se mtDNA naslijeđuje od majke, postoje istraživanja koja dokazuju da se mtDNA može naslijediti i od oca. Balciuniene J. i Balciunas D. (2019.) navode da su opažanja Luo i sur. (2018.) konzistentna s prisutnošću konkatamera mtDNA s više kopija, koji je integriran u nuklearni genom ("mega-*numt*") i razdvaja se autosomno dominantnim načinom bez obzira na njegovo roditeljsko podrijetlo, neovisno o prijenosu i selektivnom pojačanju očinskih mitohondrija. *Numt*-ovi mogu često nehotice biti otkriveni u potrazi za *bona fide* mtDNA. Oni predstavljaju fosile sekvenci prisutne u nuklearnom genomu mnogih eukariota, što čini oko 0,001 - 0,1% njihove DNA. Ako nisu prepoznati u stvarnim mtDNA studijama, oni mogu kompromitirati interpretaciju rezultata u otkrivanju heteroplazmijskih ili mtDNA patogenih varijanti ili dovode do pogrešnih filogenetskih rekonstrukcija (Schiavo, G i sur., 2017). Do danas, točni mehanizmi pomoću kojih su mtDNA sekvence integrirane u nuklearni genom nisu u potpunosti shvaćeni. Najviše prihvaćena hipoteza sugerira da u prisustvu mutagenih agensa ili stresnih uvjeta, fragmenti mtDNA mogu pobjeći iz mitohondrija i njihova integracija u nuklearni genom može nastati tijekom popravka dvostrukih prekida DNA (Ricchetti, M. i sur., 1999.; Hazkani-Covo E. i sur, 2010).

Kod ljudi, model životinja i domaćih životinja, izneseno je da noviji mutageni insercijski događaji, polimorfni unutar vrsta, uzrokuju brojne bolesti i nove fenotipe. Većina do sada analiziranih eukariotskih genoma pokazala je prisutnost mtDNA sekvenci u nuklearnom genomu (Schiavo, G i sur., 2017). Također, u nuklearnom genomu goveda, pronađeni su dokazi prijenosa iz mtDNA. Tablice 6 i 7 prikazuju točne pozicije mtDNA koji se podudaraju sa segmentima nuklearnih kromosoma te koje gene oni kodiraju.

Tablica 6. Prikazuje *numt*-ove koji su pronađeni kod *Bos taurusa*, gene mtDNA s kojima se oni poklapaju te točne pozicije baza mtDNA s kojima se poklapaju. Tamno siva boja označava da su geni identični kod *Bos taurusa* i *Bos indicusa*, siva da su geni slični kod *Bos taurusa* i *Bos indicusa*, a svijetlo siva da su geni slični, ali se nalaze na drugom numt-u istog kromosoma.

Numtovi BT	**mtDNA geni**	**mtDNA baze**
BT_CHR1_1[a]	tRNA-Phe, 12S rRNA	385-2111
BT_CHR1_2	tRNA-Glu	13408-14510
BT_CHR1_3ne[b]	tRNA-Lys	8078-8266
BT_CHR2_1	COX3, tRNA-Gly, -Arg, -His, -Ser, -Leu, -Glu, CYTB	9586-15174
BT_CHR2_2	COX2, tRNA-Lys, ATP6, COX3, -Gly, -Arg	7652-10361
BT_CHR2_3	12s rRNA, tRNA-Val, 16s rRNA	1183-2769
BT_CHR3	COX1, tRNA-Ser, -Asp, COX2, -Lys,ATP6, COX3,-Gly	7004-10029
BT_CHR4_1	tRNA-Ile	3517-4658

Tablica 6. nastavak

Numtovi BT	mtDNA geni	mtDNA baze
BT_CHR4_2	16s rRNA	2189-2767
BT_CHR4_3	"unnamed protein product;U.R.F. 4L"	10486-10550
BT_CHR4_4ne	"unnamed protein product; U.R.F. 5"	12681-12708
BT_CHR8_1	COX1, tRNA-Ser, -Asp, COX2, -Lys	6372-8246
BT_CHR8_2	tRNA-Phe, 12s rRNA	423-498
BT_CHR9_1	16s rRNA, tRNA-Leu, -Ile, -Gln, -Met	2029-5161
BT_CHR9_2	"unnamed protein product; U.R.F. 4"	10764-11833
BT_CHR9_3	tRNA-Phe, 12s rRNA	390-485
BT_CHR10_1	tRNA-Ile, -Gln, -Met, -Trp, -Ala, -Asn, -Cys, Tyr, COX1, -Ser, -Asp, COX2	3743-7779
BT_CHR10_2ne	D-loop početak	1-105
BT_CHR11_1	tRNA-His, -Ser, -Leu	11481-12639
BT_CHR11_2	16s rRNA, tRNA-Leu	2601-3886
BT_CHR11_3	tRNA-Trp, -Ala, -Asn, -Cys, Tyr, COX1	4910-6255
BT_CHR11_4	tRNA-Ile, -Gln, -Met	3901-4561
BT_CHR11_5ne	"unnamed protein product; U.R.F. 5"	12871-13220
BT_CHR11_6	tRNA-Ser, -Leu	12027-12179
BT_CHR12	tRNA-Cys	5541-5611
BT_CHR13	16s rRNA	1777-2136
BT_CHR14	12s rRNA, tRNA-Val, 16s rRNA	1331-1636
BT_CHR15_1	COX1, tRNA-Ser, -Asp, COX2	6514-8040
BT_CHR15_2ne	tRNA-Glu, CYTB	14368-15171
BT_CHR17_1	tRNA-Phe, 12s rRNA	422-704
BT_CHR17_2	tRNA-Trp, -Ala	5318-5459
BT_CHR18	16s rRNA	2451-2482
BT_CHR22_1	12s rRNA, tRNA-Val, 16s rRNA	849-2224
BT_CHR22_2	tRNA Trp	5329-5376
BT_CHR24_1	16s rRNA, tRNA-Leu, -Ile, -Gln, -Met	2845-4906
BT_CHR24_2	COX2, tRNA-Lys, ATP6	7407-8896
BT_CHR26	tRNA-Ile, -Gln, -Met	3885-4289
BT_CHR28ne	COX1	5703-6457

Tablica 6. nastavak

Numtovi BT	mtDNA geni	mtDNA baze
BT_CHRX_1	tRNA-Arg, -His, -Ser, -Leu, Glu, CYTB, -Thr, -Pro, D-loop kraj	9974-16412
BT_CHRX_2	"unnamed protein product; U.R.F. 5"	13355-13547
BT_CHRX_3	D-loop početak	1-135

[a] broj označava broj numt-a na određenom kromosomu, a [b] „ne“ označava da se ne podudara sa svim sekvencama, nego samo onima haplogrupe I.

Tablica 7. Prikazuje *numt*-ove koji su pronađeni kod *Bos indicusa*, gene mtDNA s kojima se oni poklapaju te točne pozicije baza mtDNA s kojima se poklapaju. Tamno siva boja označava da su geni identični kod *Bos taurusa* i *Bos indicusa*, siva da su geni slični kod *Bos taurusa* i *Bos indicusa*, a svijetlo siva da su geni slični, ali se nalaze na drugom numt-u istog kromosoma.

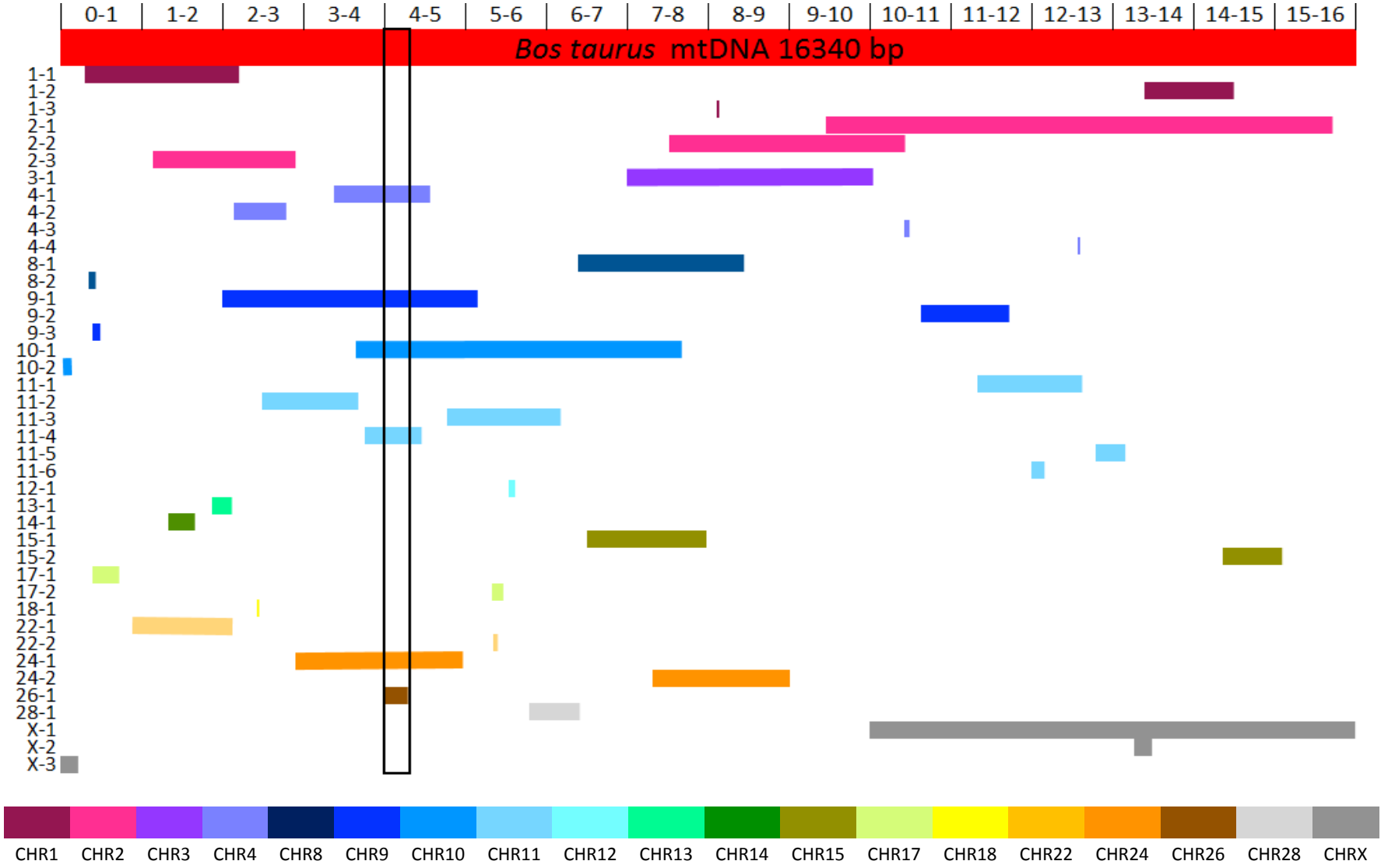
Numtovi BI	mtDNA geni	mtDNA baze
BI_CHR1_1[a]	tRNA-Phe, 12S rRNA	385-2111
BI_CHR1_2	ND5, ND6, tRNA-Glu	13481-14511
BI_CHR1_3ne[b]	ATP6	8341-8383
BI_CHR2_1	COX3, tRNA-Gly, ND3, -Arg, ND4L	9010-10324
BI_CHR2_2	16S rRNA	1936-2761
BI_CHR2_3	COX2, tRNA-Lys, ATP8, ATP6	7653-8534
BI_CHR3_1	COX1	7005-7522
BI_CHR3_2	tRNA-Gly, ND3	9804-10010
BI_CHR4_1	ND1, tRNA-Ile, -Gln, -Met, ND2	3518-4659
BI_CHR4_2	16S rRNA	2190-2769
BI_CHR4_3	ND4L, ND4	10487-10551
BI_CHR4_4	ND5	12955-13013
BI_CHR8_1	COX1	6373-7160
BI_CHR8_2	tRNA-Phe, 12S rRNA	425-600
BI_CHR9_1	ND1, tRNA-Ile, -Gln, -Met, ND2	3806-5157
BI_CHR9_2	"unnamed protein product; U.R.F. 4"	10765-11834
BI_CHR9_3	tRNA-Phe, 12S rRNA	392-487
BI_CHR10_1	tRNA-Ile, -Gln, -Met	3744-4613
BI_CHR10_2	COX1, tRNA-Ser	5699-7305
BI_CHR10_3	tRNA-Glu, CYTB	12817-14875

Tablica 7. nastavak

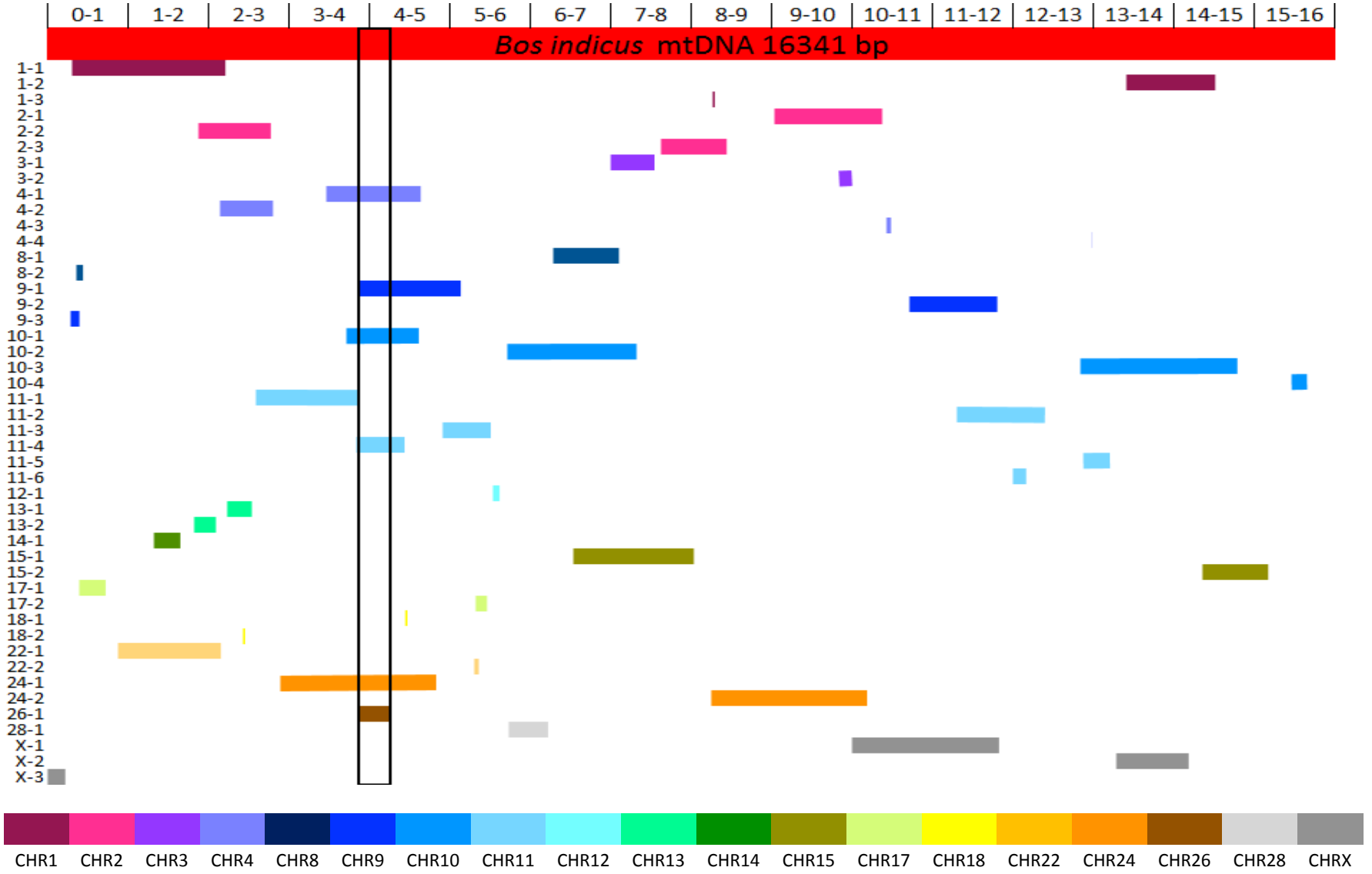
Numtovi BI	mtDNA geni	mtDNA baze
BI_CHR10_4ne	tRNA-Thr, - Pro, D-loop kraj	15630-15796
BI_CHR11_1	16S rRNA, tRNA-Leu	2602-3887
BI_CHR11_2	ND4, tRNA-His, -Ser, -Leu, ND5	11482-12394
BI_CHR11_3	ND2, tRNA-Trp, -Ala, -Asn	4911-5500
BI_CHR11_4	ND1, tRNA-Ile, -Gln, -Met, ND2	3902-4562
BI_CHR11_5	ND5	12872-13221
BI_CHR11_6	tRNA-Ser, -Leu, ND5	12028-12180
BI_CHR12	tRNA-Cys	5542-5612
BI_CHR13_1ne	16s rRNA	2212-2506
BI_CHR13_2ne	16s rRNA	1778-2137
BI_CHR14	12S rRNA, tRNA-Val, 16s rRNA	1333-1637
BI_CHR15_1	COX1, tRNA-Ser, COX2	6515-8041
BI_CHR15_2	ND6, tRNA-Glu, CYTB	14369-15232
BI_CHR17_1	tRNA-Phe, 12S rRNA	424-679
BI_CHR17_2	tRNA-Trp, -Ala, -Asn	5319-5460
BI_CHR18_1ne	ND2	4512-4591
BI_CHR18_2	16S rRNA	2452-2483
BI_CHR22_1	12S rRNA, tRNA-Val, 16s rRNA	851-2225
BI_CHR22_2	tRNA-Trp	5330-5377
BI_CHR24_1	16S rRNA, tRNA-Leu, ND1, -Ile, -Gln, -Met, ND2	2846-4847
BI_CHR24_2	ATP8, ATP6, COX3, tRNA-Gly, ND3, -Arg, ND4L	8240-10291
BI_CHR26	ND1, tRNA-Ile, -Gln, -Met	3886-4290
BI_CHR28ne	COX1	5762-6294
BI_CHRX_1	ND3, tRNA-Arg, ND4L, ND4	9983-11790
BI_CHRX_2	ND5, ND6	13265-14259
BI_CHRX_3	D-loop početak	1-135

[a] broj označava broj numt-a na određenom kromosomu, a [b] „ne“ označava da se ne podudara sa svim sekvencama, nego samo onima haplogrupe I.

Ukupna duljina fragmenata kod *Bos taurusa* je bila 57189 pb, što je 3,5 puta dulje od ukupne duljine mitogenoma. Duljina fragmenata *Bos indicusa* je 30609 pb, što je 1,9 puta dulje od ukupne duljine mitogenoma. Iako je bilo više segmenata podudarnih s *Bos indicusom,* oni su bili dosta kraći duljinom (Slike 4a i 4b). Na slikama 4a i 4b se mogu vidjeti lokacije *numt*-ova u odnosu na referentne mtDNA. Iako je većina *numt*-ova je odgovarala fragmentima mitohondrijskih gena te su njihove lokacije bile u većini slučajeva vrlo slične, postoje slučajevi u kojima kodiraju za različite gene. Također se može vidjeti i da je onaj dio koji je označen da ima najviše segmenata koji se preklapaju zajednički i kod numt-ova *Bos taurusa* i *Bos indicusa,* iako je kod *Bos indicusa* taj isti dio malo veći.


0-1
1-2
2-3
3-4
4-5
5-6
6-7
7-8
8-9
9-10
10-11
11-12
12-13
13-14
14-15
15-16
Bos taurus mtDNA 16340 bp
1-1
1-2
1-3
2-1
2-2
2-3
3-1
4-1
4-2
4-3
4-4
8-1
8-2
9-1
9-2
9-3
10-1
10-2
11-1
11-2
11-3
11-4
11-5
11-6
12-1
13-1
14-1
15-1
15-2
17-1
17-2
18-1
22-1
22-2
24-1
24-2
26-1
28-1
X-1
X-2
X-3
CHR1
CHR2
CHR3
CHR4
CHR8
CHR9
CHR10
CHR11
CHR12
CHR13
CHR14
CHR15
CHR17
CHR18
CHR22
CHR24
CHR26
CHR28
CHRX

Slika 4a.


0-1
1-2
2-3
3-4
4-5
5-6
6-7
7-8
8-9
9-10
10-11
11-12
12-13
13-14
14-15
15-16
Bos indicus mtDNA 16341 bp
1-1
1-2
1-3
2-1
2-2
2-3
3-1
3-2
4-1
4-2
4-3
4-4
8-1
8-2
9-1
9-2
9-3
10-1
10-2
10-3
10-4
11-1
11-2
11-3
11-4
11-5
11-6
12-1
13-1
13-2
14-1
15-1
15-2
17-1
17-2
18-1
18-2
22-1
22-2
24-1
24-2
26-1
28-1
X-1
X-2
X-3
CHR1
CHR2
CHR3
CHR4
CHR8
CHR9
CHR10
CHR11
CHR12
CHR13
CHR14
CHR15
CHR17
CHR18
CHR22
CHR24
CHR26
CHR28
CHRX

Slika 4b.

Slika 4. Prikazuje raspoređenost *numt*-ova u odnosu na referentnu mtDNA. a) Prikazuje raspoređenost *numt*-ova u odnosu na referentnu mtDNA *Bos taurusa*, a b) Prikazuje raspoređenost *numt*-ova u odnosu na referentnu mtDNA *Bos indicusa*. Svaka boja označava posebni kromosom, kao što je prikazano na legendi ispod. Brojevi iznad referentne mtDNA (x os) označavaju broj kilobaza (kbp). Brojevi na lijevoj strani (y os) označavaju broj kromosoma (prvi broj) te broj segmenta u tom kromosomu (drugi broj). Crnim pravokutnikom je označen dio koji ima najviše zajedničkih podudarnih segmenata s mtDNA.

Identificirano je 5 potpunih gena koji kodiraju proteine u mitohondriju kod *Bos taurusa*, a to su 12s rRNA, COX1, COX2, COX3, ATP6, koji su pronađeni u 9 *numt*-ova. Kod *Bos indicusa* identificirano je 6 potpunih gena: ND3, ND6, ATP6, ATP8, COX3, ND4L u 5 *numt*-ova. Kod ostalih *numt*-ova, geni se nalaze ili na početku ili na kraju segmenta, stoga nisu potpuni. Velika većina *numt*-ova, točnije 60 njih, kodira mt-tRNA gene. Pronađeni su svi geni kod *Bos taurusa*, a kod *Bos indicusa* svi, osim $tRNA^{Asn}$ i $tRNA^{Tyr}$. Još uvijek se ne može sa sigurnošću reći da se ti geni mogu presaviti u strukture koje vode do aktivne funkcije (Liu i Zhao, 2007). Postojanje mitohondrijskih tRNA-dvojnika u nuklearnom genomu je samo po sebi važna tema s obzirom da su mutacije u mitohondrijskim tRNA povezane s bolestima (Telonis, A. G. i sur., 2014). Mehanizmi za unos tRNA u mitohondrij su složeni i njihovo potpuno razumijevanje u reprezentativnim organizmima ostaje izazov (Salinas-Giege, T. i sur., 2015). Da bi se odgovorilo jesu li funkcionalne molekule tRNA proizvedene, potrebna su dodatna i dugotrajna eksperimentalna istraživanja.

4. Zaključak

U ovom radu istraživala se učestalost pojave segmenata mitohondrijske DNA u genomu jezgre goveda. Tijekom istraživanja *null* hipoteza nije odbačena, odnosno dokazana je postojanost segmenata mtDNA u nuklearnom genomu. Za istraživanje su korištene 33 sekvence iz dostupne iz projekta MitoTAUROmics te referentni kromosomi i mtDNA *Bos taurusa* i *Bos indicusa* koji su dostupni u javnim bazama podataka. Pronađeno je ukupno 156 potencijalnih *numt*-ova, od kojih je 86 potvrđeno. Iako su više kromosoma *Bos indicusa* imali podudaranja sa traženim sekvencama, sama duljina segmenata je bila veća kod *Bos taurusa*. Najduži *numt* ima 6393 pb, dok najkraći 27 pb. Podudarnost između *numt*-ova i mtDNA varira od 70% do 100%. U mtDNA *Bos taurusa* identificirano je 5 potpunih gena koji kodiraju proteine, a to su 12s rRNA, COX1, COX2, COX3, ATP6 koji su pronađeni u 9 *numt*-ova. Kod *Bos indicusa* identificirano je 6 potpunih gena: ND3, ND6, ATP6, ATP8, COX3, ND4L u 5 *numt*-ova. Većina (60) ostalih *numt*-ova kodira mt-tRNA gene. S obzirom na količinu mt-tRNA gena, postoji sumnja da kodiraju nuklearne tRNA, no još uvijek se ne može sa sigurnošću reći da su te molekule funkcionalne. Stoga, veća je vjerojatnost točnosti gena koji kodiraju proteine. S obzirom da je u ovom istraživanju korišten mali broj jedinki te da su, od tih jedinki, samo dvije *Bos indicus*, varijabilnost u ovom slučaju nije statistički značajna.

Kako bi se provjerila usporedna varijabilnost *Bos taurusa* i *Bos indicusa*, trebao bi se koristiti veći uzorak sekvenci obiju vrsta. Također, u budućim istraživanjima potrebno je provjeriti funkcionalnost *numt*-ova koji kodiraju transportne RNA molekule.

5. Popis literature

1. 1,M., Fu,Q., Aximu-Petri,A., Glocke,I., Nickel,B., Arsuaga,J.L., Martinez,I., Gracia,A., de Castro,J.M., Carbonell,E. et al. (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature, 505, 403–406.

2. Balciunas, D., & Balciuniene, J. (2019). A nuclear mtDNA concatemer (Mega-NUMT) could mimic paternal inheritance of mitochondrial genome. *Frontiers in Genetics*, *10*, 518.

3. Blanchard, J. L. & Schmidt, G. W. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. Mol. Biol. Evol. 13(3), 537–48 (1996).

4. Chen, N., Cai, Y., & Chen, Q. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun. 2018; 9: 2337.

5. Davis R.E., Miller S., Herrnstadt C., Ghosh S.S., Fahy E., Shinobu L.A., Galasko D., Thal L.J., Beal M.F., Howell N., Parker W.D. Jr. (1997) Mutations in mitochondrial cytochrome c oxidase genes segregate with late-onset Alzheimer disease. Proc. Natl. Acad. Sci. U.S.A., 94, 4526–4531.

6. Dayama, G., Emery, S. B., Kidd, J. M., & Mills, R. E. (2014). The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic acids research*, *42*(20), 12640-12649.

7. Du Buy H.G., Riley F.L. (1967) Hybridization between the nuclear and kinetoplast DNA's of *Leishmania enriettii* and between nuclear and mitochondrial DNA's of mouse liver. Proc. Natl. Acad. Sci. U.S.A., 57, 790–797.

8. Gaziev, A. I., & Shaikhaev, G. O. (2010). Nuclear mitochondrial pseudogenes. *Molecular Biology*, *44*(3), 358-368.

9. Gherman A, Chen PE, Teslovich TM, Stankiewicz P, Withers M, Kashuk CS, Chakravarti A, Lupski JR, Cutler DJ, Katsanis N (2007) Population bottlenecks as a potential major shaping force of human genome architecture. PLoS Genet 3(7):e119

10. Gherman,A., Chen,P.E., Teslovich,T.M., Stankiewicz,P., Withers,M., Kashuk,C.S., Chakravarti,A., Lupski,J.R., Cutler,D.J. and Katsanis,N. (2007) Population bottlenecks as a potential major shaping force of human genome architecture. PLoS Genet., 3, e119.

11. Gould SB, Waller RF, McFadden GI (2008) Plastid evolution. Annu Rev PlantBiol 59: 491–517

12. Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. PLoS Genet 6(2): e1000834. doi:10.1371/journal.pgen.1000834

13. Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet 6(2):e1000834

14. Hazkani-Covo, E. & Covo, S. Numt-mediated double-strand break repair mitigates deletions during primate genome evolution. PLoS Genet. 4(10), e1000237 (2008).

15. Hirano M., Shtilbans A., Mayeux R., Davidson M.M., DiMauro S., Knowles J.A., Schon E.A. (1997) Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. Proc. Natl. Acad. Sci. U.S.A., 94, 14894–14899.

16. Jensen-Seaman, M. I., Wildschutte, J. H., Soto-Calderón, I. D. & Anthony, N. M. A comparative approach shows differences in patterns of numt insertion during hominoid evolution. J. Mol. Evol. 68(6), 688–99 (2009).

17. Ladoukakis, E. D., & Zouros, E. (2017). Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. *Journal of Biological Research-Thessaloniki*, *24*(1), 2.

18. Lang, M., Sazzini, M., Calabrese, F. M., Simone, D., Boattini, A., Romeo, G., ... & Gasparre, G. (2012). Polymorphic NumtS trace human population relationships. *Human genetics*, *131*(5), 757-771.

19. Lopez J.V., Yuhki N., Masuda R., Modi W., O'Brien S.J. (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol., 39, 174–190.

20. Luo, S., Valencia, C. A., Zhang, J., Lee, N. C., Slone, J., Gui, B., ... & Chen, S. M. (2018). Biparental inheritance of mitochondrial DNA in humans. *Proceedings of the National Academy of Sciences*, *115*(51), 13039-13044.

21. Mishmar,D., Ruiz-Pesini,E., Brandon,M. and Wallace,D.C. (2004) Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. Hum. Mutat., 23, 125–133.

22. Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs doublestrandbreaks in yeast chromosomes. Nature 402: 96–100

23. Ricchetti M., Tekaia F., Dujon B. (2004) Continued colonization of the human genome by mitochondrial DNA. PLoS Biol., 2, E273.

24. Ricchetti, M., Fairhead, C. and Dujon, B. 1999, Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature, 402, 96–100.

25. Richly, E., & Leister, D. (2004). NUMTs in sequenced eukaryotic genomes. *Molecular biology and evolution*, *21*(6), 1081-1084.

26. Salinas-Giegé, T., Giegé, R., & Giegé, P. (2015). tRNA biology in mitochondria. *International journal of molecular sciences*, *16*(3), 4518-4559.

27. Schiavo, G., Hoffmann, O. I., Ribani, A., Utzeri, V. J., Ghionda, M. C., Bertolini, F., ... & Fontanesi, L. (2017). A genomic landscape of mitochondrial DNA insertions in the pig nuclear genome provides evolutionary signatures of interspecies admixture. *DNA Research*, *24*(5), 487-498.

28. Schmitz J., Piskurek O., Zischler H. (2005) Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in primates. J. Mol. Evol., 61, 1–11.

29. Telonis, A. G., Loher, P., Kirino, Y., & Rigoutsos, I. (2014). Nuclear and mitochondrial tRNA-lookalikes in the human genome. *Frontiers in genetics*, *5*, 344.

30. Thalmann O., Hebler J., Poinar H.N., Paabo S., Vigilant L. (2004) Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol. Ecol., 13, 321–335.

31. Triant, D. A. & DeWoody, J. A. Extensive mitochondrial DNA transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications. Gene 401(1–2), 61–70 (2007).

32. Tsuji,J., Frith,M.C., Tomii,K. and Horton,P. (2012) Mammalian NUMT insertion is non-random. Nucleic Acids Res., 40, 9073–9088.

33. Wallace D.C., Stugard C., Murdock D., Schurr T., Brown M.D. (1997) Ancient mtDNA sequences in the human nuclear genome: A potential source of errors in identifying pathogenic mutations. Proc. Natl. Acad. Sci. U.S.A., 94, 14900–14905.

34. Woischnik, M. & Moraes, C. T. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res. 12(6), 885–93 (2002).

35. Xia, X., Qu, K., Zhang, G., Jia, Y., Ma, Z., Zhao, X., ... & Lei, C. (2019). Comprehensive analysis of the mitochondrial DNA diversity in Chinese cattle. *Animal genetics*, *50*(1), 70-73.

36. Zischler H (2000) Nuclear integrations of mitochondrial DNA in primates: inference of associated mutational events. Electrophoresis 21(3):531–536

Internet:

1. NCBI (2019). The National Center for Biotechnology Information – Bos taurus complete genome. [https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1]

2. NCBI (2019). The National Center for Biotechnology Information – Bos indicus complete genome. [https://www.ncbi.nlm.nih.gov/assembly/GCF_000247795.1]

3. MitoTAUROmics project - Utilisation of the whole mitogenome in cattle breeding and conservation genetics – 2015. Author: prof. Ino Čurik. Date last modified: June 6th, 2015. [http://mitotauromics.agr.hr]

4. Guidance to official cattle breed and codes. [https://www.gov.uk/guidance/official-cattle-breeds-and-codes]

5. IBAN country codes alpha2 & alpha3 database for cattle. BIC data used with permission of S.W.I.F.T. SCRL. Database Rights Reserved, 2019. [https://www.iban.com/country-codes]

6. BLAST® Command Line Applications User Manual (2008). This manual documents the BLAST (Basic Local Alignment Search Tool) command line applications developed at the National Center for Biotechnology Information (NCBI). Created: June 23, 2008; Last Update: April 1, 2019. [https://www.ncbi.nlm.nih.gov/books/NBK279690/]

7. How to run command-line in BLAST. Copyright 2010 onwards, C. Titus Brown et al.. Created using Sphinx. [https://angus.readthedocs.io/en/2018/running-command-line-blast.html]

8. BLAST® Command Line Applications User Manual (2008). BLAST+ features. [https://www.ncbi.nlm.nih.gov/books/NBK279668/]

9. BLAST® Command Line Applications User Manual (2008). BLAST Appendices. Created: June 23, 2008; Last Update: November 14, 2018. [https://www.ncbi.nlm.nih.gov/books/NBK279684/#appendices.Options_for_the_commandline_a]

10. BLAST® Command Line Applications User Manual (2008). Display BLAST search results with custom output format. [https://www.ncbi.nlm.nih.gov/books/NBK279682/]

11. BLAST (Basic Local Alignment Search Tool, 2008). BLAST is a registered trademark of the National Library of Medicine. [https://blast.ncbi.nlm.nih.gov/Blast.cgi]

12. Metagenomic. BLASTN output format 6. Author: Matthias Scholz. [http://www.metagenomics.wiki/tools/blast/blastn-output-format-6]

13. Metagenomic. E-value & Bit-score. Author: Matthias Scholz. [http://www.metagenomics.wiki/tools/blast/evalue]

14. The Titi Tudorancea Bulletin, Encyclopedia. BLAST Raw Score. [https://www.tititudorancea.com/z/blast_raw_score.htm]

15. Komatsu, M., Yasuda, Y., Matias, J. M., Niibayashi, T., Abe Nishimura, A., Kojima, T., ... & Yamamoto, N. (2004). Mitochondrial DNA polymorphisms of D-loop and three coding regions (ND2, ND4, ND5) in three Philippine native cattle: indicus and taurus maternal lineages. *Animal Science Journal (Japan)*. [http://agris.fao.org/agris-search/search.do?recordID=JP2005001841]

16. Zhang, V. W., & Wong, L. J. C. (2017). Comprehensive Analyses of the Mitochondrial Genome. In *Next Generation Sequencing Based Clinical Molecular Diagnosis of Human Genetic Disorders*(pp. 287-304). Springer, Cham, str. 287-302. [https://books.google.hr/books?id=OCjiDgAAQBAJ&pg=PA300&lpg=PA300&dq=problems+with+numts&source=bl&ots=icrmgNgepf&sig=xLHy3sE2ywoyRSFginI1l_fPYgY&hl=hr&sa=X&ved=0ahUKEwiL1J7J3ebYAhUDPFAKHR7vBbcQ6AEIcjAI#v=onepage&q=problems%20with%20numts&f=false]

Životopis

Lucija Andjel rođena je 27.11.1994. godine u Zagrebu, Hrvatskoj. Od 2009. do 2013. godine pohađa Prirodoslovnu školu Vladimira Preloga, smjer Prirodoslovna gimnazija. Položivši maturu, nakon završetka srednje škole upisuje Agronomski fakultet, preddiplomski studij Agroekologija. Zadnju godinu preddiplomskog studija upisuje izborni modul „Njemački jezik za akademske potrebe I i II“ na Filozofskom fakultetu u Zagrebu gdje stječe A2 razinu njemačkog jezika. 2016.-e godine završava taj isti studij te stječe naziv univ. bacc. ing. agr. Na jesen iste godine upisuje diplomski studij Genetika i oplemenjivanje životinja. U 2017.-oj odlazi na Erasmus+ studiranje u *Czech University of Life Sciences (CULS) Prague* na jedan semestar gdje je na engleskom jeziku pohađala diplomski studij *Animal sciences*. U 2018. godini odrađuje stručnu praksu u Institutu „Ruđer Bošković“ u Zavodu za evolucijsku genetiku u trajanju od tri mjeseca. Iste godine odslušala je tečaj „*Detecting introgression using genomic data: ABBA BABA and related analyses*“. Iste godine odlazi po drugi puta na Erasmus+ studij u *University of Natural Resources and Life Sciences, Vienna (BOKU*), također na jedan semestar gdje je na engleskom jeziku pohađala diplomski studij *Animal genetics and breeding*. Tijekom razmjene u Beču, položila je modul „*German conversation*“ gdje je poboljšala svoje govorne vještine njemačkog jezika na B1 razinu. 2019. godine polaže IELTS ispit u kojem je dobila *band* 7.5 (slušanje 8.0, čitanje 8.5, pisanje 6.0, govorenje 7.0), odnosno prema CEFR-u C1 razinu znanja engleskog jezika. Tijekom diplomskog studija naučila je raditi u sljedećim programima/softverima: SAS, R, R Studio, MEGA, DNAsp, Network, Wombat, VanRaden, CFC, Endog, Pedig, Magellan, SnEP, PLINK, STRUCTURE, Python. Tijekom razmjene u Beču stekla je znanje u radu i programiranju u Linux operativnom sustavu kroz slušanje predmeta "*Essentials for bioinformatics data analysi*s". U slobodno vrijeme pohađala je početni i napredni tečaj astronomije u Zvjezdarnici Zagreb i bavi se yogom.