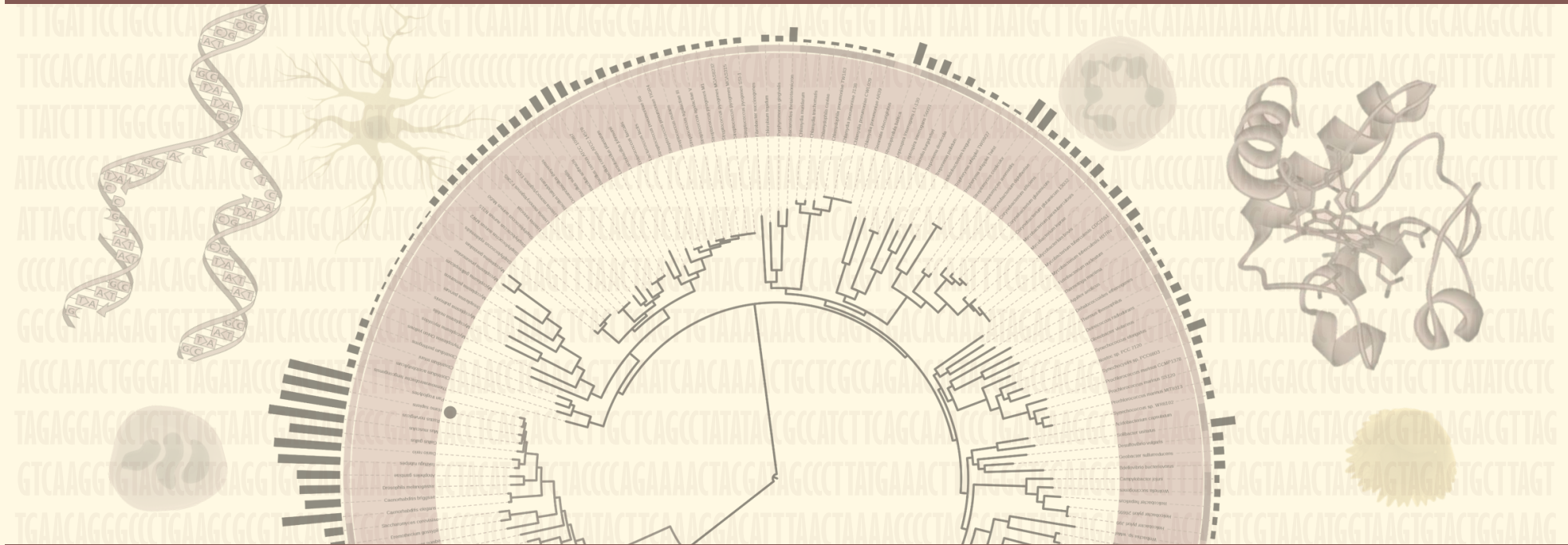


PETA VJEŽBA

MOLEKULARNA FILOGENETIKA



doc. dr. sc.
JOSIP SKEJO

SADRŽAJ

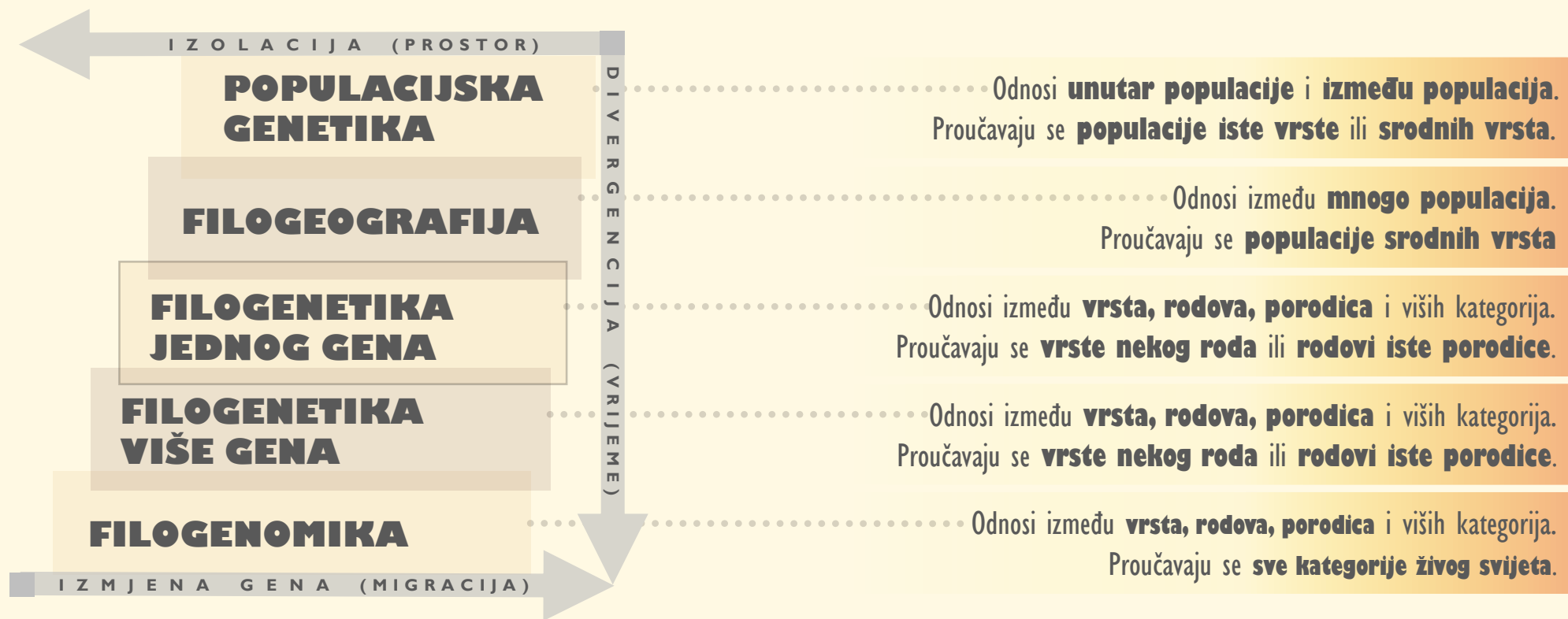
pretraga, identifikacija, organizacija, sravnjivanje, konstrukcija



- pretraživanje molekularnih sekvenc(ija)
- identifikacija molekularnih sekvenci
- organizacija molekularnih sekvenci
- sravnjivanje/poravnanje molekularnih sekvence
- matrica udaljenosti sekvenci (eng. *distance matrix*)
- konstrukcija stabala u programu MEGA
- najveća štedljivost, MP (*maximum parsimony*) stablo
- odabir najboljeg ML (*maximum likelihood*) modela
- najveća vjerodostojnost, ML (*maximum likelihood*) stablo
- interpretacija statističke podržanosti čvorova
- interpretacija duljine grana u filogramu

RAZINE

MOLEKULARNE FILOGENETIKE



POPULACIJSKA GENETIKA

Populacijska genetika je grana evol. biologije koja **proučava evoluciju na razini populacije**.

Važni pojmovi su frekvencija alela, mendelijansko nasljeđivanje, Hardy-Weibergov princip, genotip, fenotip, dominacija (*dominance*), reproduktivna sprema (*fitness*), relativni fitness, efektivna veličina populacije, mutacija, gensko odstupanje (*genetic drift*), srođivanje (*inbreeding*), migracija, neravnoteža vezanosti gena (*linkage disequilibrium*).

Hardy-Weibergov princip

$$(p + q)^2 = p^2 + 2pq + q^2$$

gdje su p i q frekvencije dvaju alela ili genotipova u populaciji, a p^2 i q^2 frekvencije homozigota, odnosno $2pq$ heterozigota.

Reproduktivna sprema ili fitness

$$w = 1 - s$$

gdje je w fitness, a s selektivni pritisak.

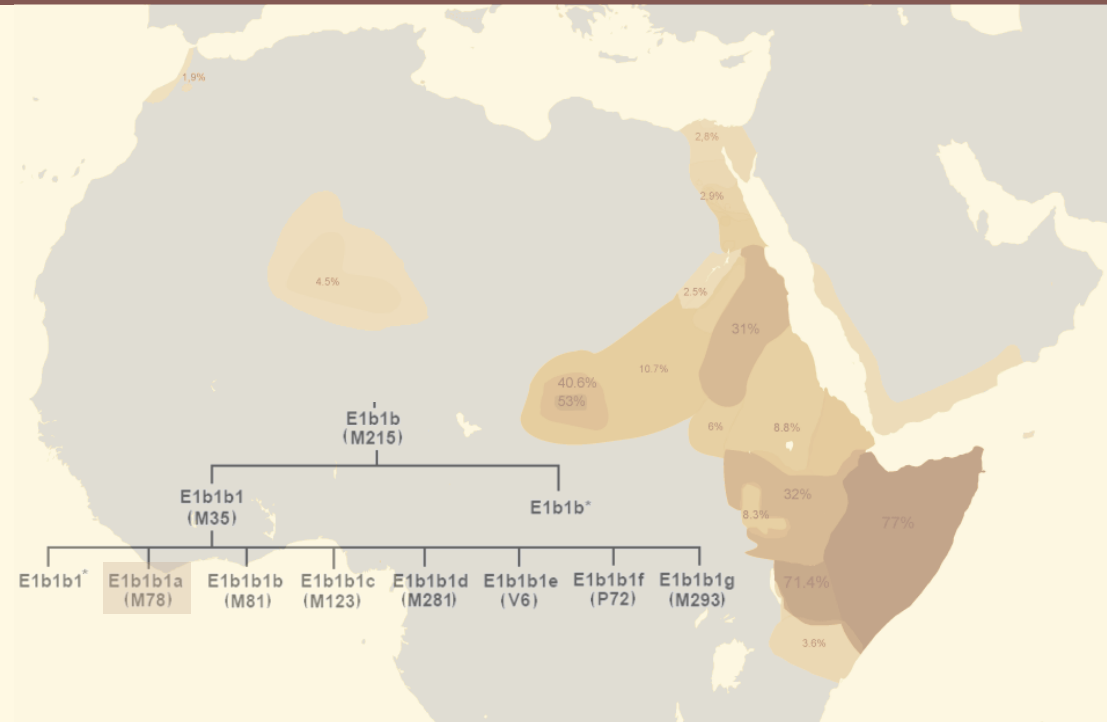
FILOGEOGRAFIJA

FILOGENOMIKA S GEOGRAFIJOM

Filogeografija je grana evolucijske biologije koja **proučava varijabilnost gena** u odnosu na geografsku rasprostranjenost.

Filogeografija povezuje događaje iz prošlosti s današnjom rasprostranjenošću alela i odgovara na pitanja zbog čega je negdje prisutna neka prilagodba (mutacija.)

Važni pojmovi filogeografije su: haplotip, haplogrupa, biogeografija, specijacija, radijacija, drift (*bottleneck*, *founder effect*).



Distribucija ljudske y-haplogrupe E1b1b1a1b (E-V13*), koja pripada M78 unutar M35.

FILOGENETIKA

JEDNOGA (UNI-) ILI VIŠE GENA (MULTILOKUSNA)

UNILOKUSNA FILOGENETIKA

Filogenija jednog gena (npr. 16S, ITS ili COI) predstavlja evoluciju tog gena, a ne vrsta kojima gen pripada.

MULTILOKUSNA FILOGENETIKA

Filogenija temeljena na kombinaciji evolucije više gena (npr. 16S+ITS+COI) daje finiju rezoluciju.

BILJEG (GEN)	SISTEMATSKA KATEGORIJA							
	CARSTVO	KOLJENO	RAZRED	RED	PORODICA	ROD	VRSTA	POPULACIJA
16S i 18S rRNA	—————	—————	—————	—————			
23S i 28S rRNA			—————	—————	—————		
5.8S rRNA	—————		—————				
IGS							—————	—————
ITS						—————	—————	—————
12S rRNA		—————	—————				
16S rRNA					—————		
ND1				—————		
ND2				—————		
COI				—————		
COII				—————		
Cyt b				—————		
D-Loop							—————	—————

Različiti geni imaju različit signal na različitim razinama. Primjerice, konzervirani geni poput 16S ili 5.8S dobri su za filogeniju carstava, koljena i razreda dok su varijabilni geni poput IGS, ITS ili D-Loop dobri za filogeniju rodova, vrsta i populacija.

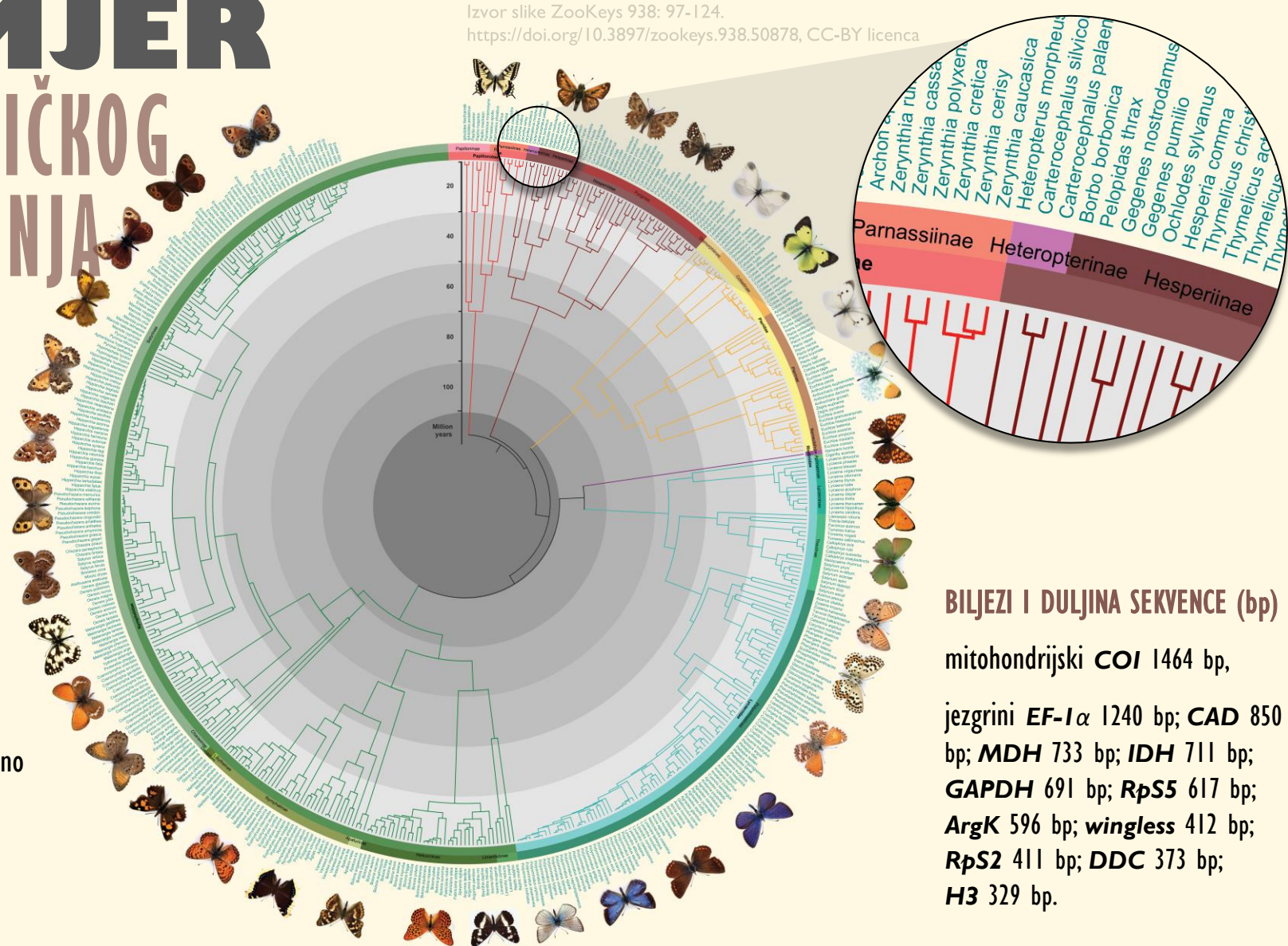
PRIMJER FILOGENETIČKOG ISTRAŽIVANJA

Wiemers et al. (2020)
A complete time-calibrated
multi-gene phylogeny of
the European butterflies.
Zookeys, 938: 97–124.

ISTRAŽIVANA SKUPINA
danji leptiri (Papilionoidea)

VANJSKE SKUPINE
za svaku porodicu pojedinačno

VRSTA DENDROGRAMA
kronogram



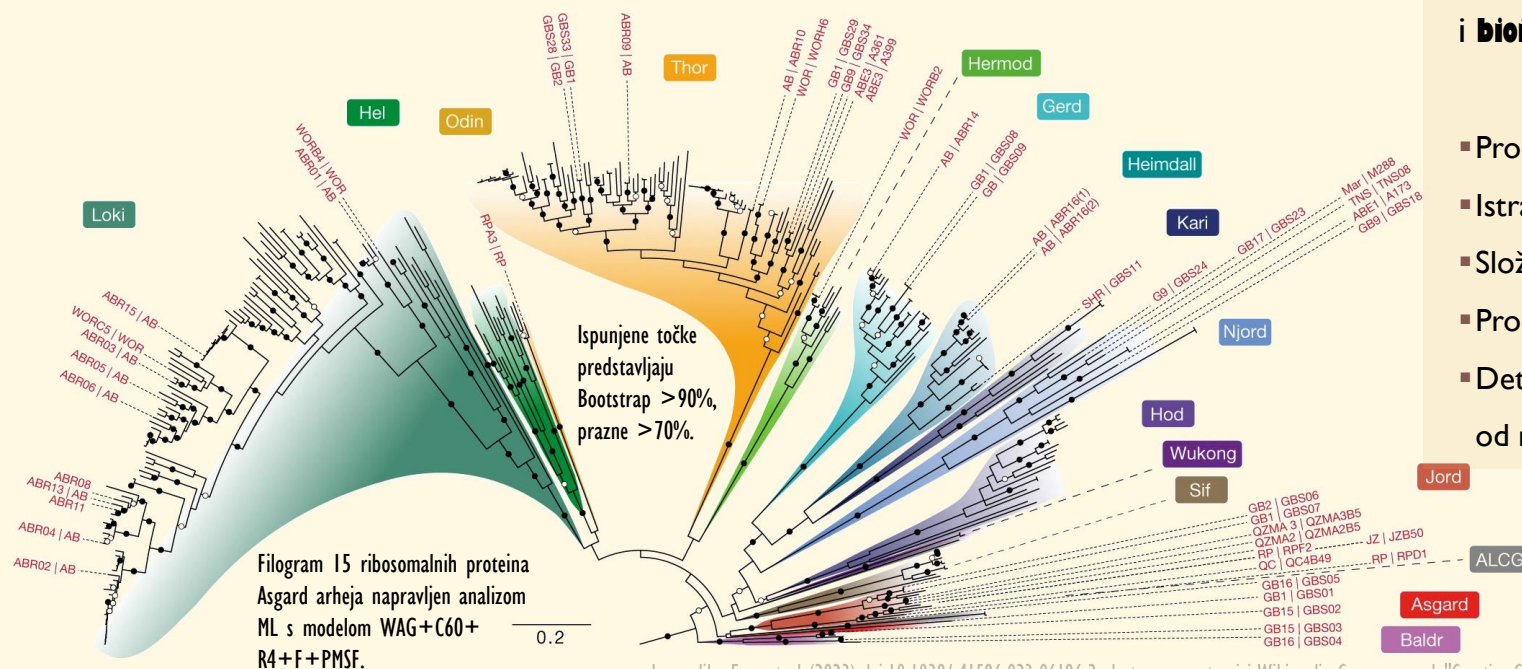
FILOGENOMIKA

FILOGENETIKA ČITAVIH GENOMA

Filogenomika je grana evlucijske biologije koja proučava **evoluciju** velikih dijelova genoma ili čak **čitavih genoma**.

Filogenomika kombinira saznanja iz **klasične evolucije**, **filogenetike** i **bioinformatike** kako bi

- Proučila evoluciju proteinskih porodica,
- Istražila lateralni prijenos gena (LGT),
- Složila do sada neistražene genome,
- Pronašla do sada nepoznate gene,
- Detaljno rekonstruirala evlucijske odnose od razine vrsta do razine domena.



Izvor slike Eme et al (2023) doi:10.1038/s41586-023-06186-2, dostupno na stranici Wikimedia Commons pod "Creative Commons Attribution 4.0 International" Licencom

Zadatak 1.

Nadi neki objavljeni rad koji sadrži filogenetičko istraživanje (npr. preko Google pretraživanja „phylogeny of”), primjerice na skupini biljaka ili životinja koja ti je najdraža. Provjeri kakvo je stablo objavljeno i na osnovu čega je napravljena pa potom popuni priloženu tablicu.

0 RADU	AUTORI RADA	
	GODINA OBJAVE	
	NASLOV RADA	
	ČASOPIS	
	VOLUMEN I STRANICE	
0 STABLU	ISTRAŽIVANA SKUPINA	
	VANJSKA/E SKUPINA/E	
	VRSTA DENDROGRAMA	
	NAZIV BILJEGA: GEN ILI REGIJA	
	DULJINA SEKVENCE (bp)	

PRETRAŽIVANJE

MOLEKULARNIH SEKVENCI



NCBI (*National Center for Biotechnology*)
Stranica sadrži sve objavljene sekvence.

<https://www.ncbi.nlm.nih.gov/>

Svaka sekvenca ima svoj jedinstveni kod i uz nju je navedena regija ili gen kojoj sekvenca pripada, duljina sekvence u parovima baza i mnogi drugi podatci.

Primjerice, kod NC_023100.1 krije mitohondrijski genom izumrlog čovjeka *Homo heidelbergensis* i dugačak je 16 568 bp (parova baza). Kad uđemo u podatke o toj sekvenci možemo naći i rad u kojem je objavljena: Meyer *et al.* (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*, 505(7483), 403–406. Otvorimo li rad, možemo čitati o značaju ove sekvence i saznati informacije o kvaliteti i metodologiji.

PROTOKOL PRETRAGE SEKVENCI

1. Otvori stranicu <https://blast.ncbi.nlm.nih.gov>
2. Upiši ime vrste ili više kategorije na latinskom ili engleskom i ime gena ili regije koji želiš pronaći, npr. „tomato chloroplast”.
3. Među rezultatima je [AC_000188.1/](https://blast.ncbi.nlm.nih.gov/AC_000188.1/), tj. čitav kloroplastni genom rajčice (*Solanum lycopersicum*), što je željeni rezultat.
4. Željenu sekvencu možemo zalijepiti u Word ako s njom želimo kasnije raditi ili ju skinuti u FASTA formatu.

IDENTIFIKACIJA MOLEKULARNIH SEKVENCI

BLAST – Basic Local Alignment Search Tool



PROTOKOL IDENTIFIKACIJE SEKVENCI

nukleotidne sekvence

1. Otvori stranicu <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
2. Odaberi „Nucleotide BLAST”
3. Zalijepi sekvencu „GTCTTAATCCCCATTTTCCAGATGAGGGAA” u „Nucleotide BLAST” te klikni na dnu opciju „BLAST” i sačekaj da analiza prikaže rezultat.
4. Odgovori kojoj skupini pripada sekvenca iz (3) i koji gen predstavlja.

aminokiselinske sekvence

5. Zalijepi sekvencu „AKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEM-KASEDLKKH” u „Protein BLAST” te klikni na dnu opciju „BLAST” i sačekaj rezultat.
6. Odgovori kojoj skupini pripada sekvenca iz (3) i koji protein predstavlja.

ORGANIZACIJA

MOLEKULARNIH SEKVENCI

Molekularne sekvence baza ili aminokiselina mogu se jednostavno organizirati u tablici (Microsoft Excel ili LibreOffice Calc) ili još jednostavnije, samo tekstualno kopirati u Notepad, Microsoft Word ili LibreOffice Writer. Treba paziti da se prilikom lijepljena sekvence u tekstualni dokument uvijek kopira i kod pod koji se sekvenca može naći u bazi. Otvorimo bazu BOLD, bazu sekvenci mitohondrijskog COI gena koji služi kao barkod sekvenca jer je utvrđeno da je vrsno specifičan.

PROTOKOL ORGANIZACIJE SEKVENCI

1. Otvori stranicu **Bold Systems**
2. Odaberi „Explore the data”
3. Pretraži skupinu/skupine koja te zanima.
4. Odaberi najmanje 8 unutarnjih sekvenci.
5. Skini sekvence tako da odabereš sekvencu i klikneš FASTA.
6. Svaku pojedinu sekvencu otvori u programu Notepad.
7. Zalijepi sve sekvence u jedan Notepad dokument.
8. Dodaj barem dvije vanjske skupine, tj. dvije vanjske sekvence.

BOLDSYSTEMS

<https://www.boldsystems.org/>

EXPLORE THE DATA

Tražim najveće životinje, kitove porodica Balaenopteridae.

Odabrat ću 14 unutarnjih sekvenci, *Balaenoptera*, *Megaptera*, *Eschrichtius*.

FASTA

Open with



Notepad

Spremam Notepad file na radnu površinu kao „Asistentovi_kitovi.txt”

Tražim srodnike ove porodice, tj. pripadnike Neobalaenidae i Balaenidae, ukupno 4 vanjske sekvence iz rodova *Caperea*, *Eubalaena* i *Balaena*.

PRIMJER ORGANIZIRANIH SEKVENCI kitova (Cetacea)

BOLD kod **ime vrste** **gen** **NCBI kod**
 >GBMNC74293-20|Balaenoptera edeni|COI-5P|MT895690

>GBMNC74293-20|Balaenoptera edeni|COI-5P|MT895690|GTAGGCACCTGGCCTAAGCTTATAATCCGTGCTGAGTTAGGTCAGCCCGGCACACTAATCGGAGATGACCAAGTC
 >GBMND52179-21|Balaenoptera edeni|COI-5P|MW571086|GCCCGGCACACTAATCGGAGATGACCAAGTCCTACAACGTATTAGTAACGCCACGCCTTTGTGATAATCTTCTT
 >GBMNF1500-22|Balaenoptera brydei|COI-5P|MW446661|ATTGGCACCTTATTTATTATTGGTGCATGAGCAGGAATAGTAGGCACTGGCCTAAGCTTATTAATTCGCGCTG
 >GBMNA13488-19|Balaenoptera brydei|COI-5P|AP006469|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATTGGCACCTATATTTACTATTTGGTGCATGAGC
 >GBMNA13489-19|Balaenoptera borealis|COI-5P|AP006470|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATTGGTACCCTATATTTACTATTTGGTGCATGAGC
 >GBMA25701-19|Balaenoptera borealis|COI-5P|MF409248|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATTGGTACCCTATATTTACTATTTGGTGCATGAGC
 >GBMNA13490-19|Balaenoptera omurai|COI-5P|AB201256|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATCGGCACCCTATACTTACTATTTGGTGCCTGAG
 >GBMNA13493-19|Balaenoptera musculus|COI-5P|X72204|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATCGGCACCCTATACTTACTATTTGGTGCCTGAGC
 >GBMA25696-19|Balaenoptera musculus|COI-5P|MF409242|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATCGGCACCCTATACTTACTATTTGGTGCCTGAGC
 >GBMA4282-13|Megaptera novaeangliae|COI-5P|GQ353285|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATCGGCACCCTATACTTACTATTTGGTGCCTGAGC
 >GBMA21221-19|Megaptera novaeangliae|COI-5P|KY001615|AAAGACATCGGCACCCTATATTTACTATTTGGTGCCTGAGCAGGAATAGTAGGCACTGGCCTAAGCTTATTAAT
 >GBMA7532-13|Balaenoptera physalus|COI-5P|KC572804|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATCGGCACCCTATACTTATTTATTCGGTGCCTGAGC
 >GBMA14706-17|Balaenoptera bonaerensis|COI-5P|LC106302|CTAATCGGAGACGACCAAGCTATAACGTATTAGTAACAGCCACGCCTTCGTGATAATCTTTTCATGGTTAT
 >GBMNA16637-19|Balaenoptera acutorostrata|COI-5P|AJ554054|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATCGGCACCCTGATTTACTATTTGGTGCCTGAGC
 >GBMA25698-19|Eschrichtius robustus|COI-5P|MF409244|ATGTTTCATAAACCGCTGACTATTCTCAACCAATCACAAGACATTGGCACCTATATTTACTATTTGGTGCCTGAGC
 >GBMA26816-19|Eschrichtius robustus|COI-5P|MH047022|CCTATATTTACTATTTGGTGCCTGAGCAGGAATAGTAGGCACTGGCCTAAGCTTATTAATTCGCGCTGAAGTAGGT
 >GBMNA13496-19|Caperea marginata|COI-5P|AJ554052|ATGTTTCATAAACCGCTGACTATTCTCAACCAATCACAAGACATTGGCACCTATACTTATTTACTATTTGGCGCCTGAGCA
 >GBMA0175-06|Caperea marginata|COI-5P|AP006475|AACCGCTGACTATTCTCAACCAATCACAAGACATTGGCACCTATACTTATTTACTATTTGGCGCCTGAGCAGGAATAGTAG
 >GBMA25766-19|Eubalaena glacialis|COI-5P|MF459656|ATGTTTCATAAACCGCTGACTATTCTCAACCAACCACAAAGACATTGGCACCTATATTTACTATTTGGCGCCTGAGCAC
 >GBMA0172-06|Balaena mysticetus|COI-5P|AP006472|AACCGCTGACTATTCTCAACCAACCACAAAGATATTGGCACCTATATTTACTATTTGGTGCCTGAGCAGGAATAGTAG

Balaenopteridae
 3 roda,
 10 vrsta
 16 sekvenci

vanjske skupine
 3 roda, 3 vrste
 4 sekvence

SRAVNJIVANJE MOLEKULARNIH SEKVENCI

Molekularne sekvence sraunjuju se programski (MAFFT, Clustal Omega MUSCLE), za razliku od kladističkih matrica koje se kodiraju ručno u tablicama

Sraunjivanje rezultira sekvencama kod kojih su slični nizovi baza jedni iznad drugih, tj. usporedivi su. Identični nizovi najčešće predstavljaju i homologne baze.

		0	0	0	0	0	0	0	0	1	1	1	1	
		1	2	3	4	5	6	7	8	9	0	1	2	3
>seq1	GACATTCATTA	-	G	A	C	-	A	T	T	C	A	T	T	A
>seq2	ACCGGTCCATTA	-	A	C	C	G	G	T	C	C	A	T	T	A
>seq3	AAATATTCA	A	A	A	T	-	A	T	T	C	-	-	-	A
>seq4	AACGTCATCA	-	A	A	C	-	G	T	T	C	A	T	C	A

LEGENDA:
KONZERVIRANO MJESTO (sivo)
DELECIJA (narančasto)
INSERCIJA (zeleno)

PRIMJER PROTOKOLA SRAVNJIVANJA SEKVENCI

1. Otvori stranicu **Clustal Omega**
2. Kopiraj sekvence iz Notepad-a i zalijepi ih na stranicu.
3. Odaberi „FASTA” pod „Output format” i klikni „Submit”
4. Sačekaj da program srauni sekvence.
5. Skini sraunjene sekvence u FASTA formatu pod „Result Files”.

Clustal Omega
www.ebi.ac.uk/jdispatcher/msa/clustalo

1. Input sequence (FASTA format) is pasted into the text area.

2. Sequence type is set to "Automatic".

3. The "Submit" button is clicked.

4. A confirmation message appears: "YOUR JOB IS QUEUED... please be patient!" and "The result of your job will appear in this browser window." A "QUEUED" status is shown.

MATRICA

UDALJENOSTI MEĐU SEKVENCAMA

PROTOKOL UNOSA SEKVENCI U PROGRAM MEGA

1. Otvori program MEGA (MEGA-X ili MEGA-II)
2. Ubaci sekvence u MEGA tako da povučesh FASTA datoteku.
3. Kad se sekvence ubacuju MEGA pita kakve su i čije su sekvence.
4. Kad su sekvence u programu klikni na ikonu sa sekvencama.
5. Opcija 'Highlight' odredi broj konzerviranih i varijabilnih mjesta.



PROTOKOL IZRADA MATRICE UDALJENOSTI

1. Klikni opciju „Distance” i odaberi „Compute Pairwise Distances”
2. MEGA pita „Would you like to use...”, klikni „Yes”
3. MEGA sada traži parametre, no ostavi sve kako jest i klikni „Ok” .

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 GBMNC74293-20 <i>Balaenoptera_edeni</i> COI-5P MT895690	-																		
2 GBMND52179-21 <i>Balaenoptera_edeni</i> COI-5P MW571086	0.00	-																	
3 GBMNF1500-22 <i>Balaenoptera_brydei</i> COI-5P MW446661	0.05	0.05	-																
4 GBMNA13488-19 <i>Balaenoptera_brydei</i> COI-5P AP006469	0.05	0.05	0.00	-															
5 GBMNA13489-19 <i>Balaenoptera_borealis</i> COI-5P AP006470	0.05	0.05	0.05	0.05	-														
6 GBMA25701-19 <i>Balaenoptera_borealis</i> COI-5P MF409248	0.04	0.04	0.05	0.04	0.01	-													
7 GBMNA13490-19 <i>Balaenoptera_omurai</i> COI-5P AB201256	0.08	0.08	0.11	0.10	0.08	0.08	-												
8 GBMNA13493-19 <i>Balaenoptera_musculus</i> COI-5P X72204	0.09	0.09	0.11	0.10	0.07	0.07	0.09	-											
9 GBMA25696-19 <i>Balaenoptera_musculus</i> COI-5P MF409242	0.09	0.08	0.10	0.10	0.07	0.07	0.08	0.00	-										
10 GBMA4282-13 <i>Megaptera_novaeangliae</i> COI-5P GQ353285	0.09	0.09	0.11	0.11	0.08	0.08	0.09	0.10	0.10	-									
11 GBMA21221-19 <i>Megaptera_novaeangliae</i> COI-5P KY001615	0.09	0.10	0.12	0.11	0.09	0.09	0.10	0.11	0.10	0.01	-								
12 GBMA7532-13 <i>Balaenoptera_physalus</i> COI-5P KC572804	0.08	0.09	0.11	0.11	0.08	0.08	0.09	0.10	0.09	0.07	0.07	-							
13 GBMA14706-17 <i>Balaenoptera_bonaerensis</i> COI-5P LC106302	0.10	0.10	0.11	0.11	0.09	0.10	0.10	0.10	0.10	0.09	0.09	0.08	-						
14 GBMNA16637-19 <i>Balaenoptera_acutorostrata</i> COI-5P AJ554054	0.10	0.11	0.10	0.11	0.11	0.10	0.11	0.12	0.12	0.11	0.11	0.11	0.04	-					
15 GBMA25698-19 <i>Eschrichtius_robustus</i> COI-5P MF409244	0.12	0.12	0.12	0.12	0.11	0.13	0.13	0.12	0.11	0.12	0.12	0.12	0.12	-					
16 GBMA26816-19 <i>Eschrichtius_robustus</i> COI-5P MH047022	0.11	0.12	0.12	0.12	0.12	0.11	0.12	0.13	0.13	0.11	0.12	0.12	0.11	0.12	0.00	-			
17 GBMNA13496-19 <i>Caperea_marginata</i> COI-5P AJ554052	0.15	0.15	0.13	0.13	0.14	0.14	0.17	0.16	0.16	0.13	0.13	0.17	0.16	0.16	0.16	0.16	-		
18 GBMA0175-06 <i>Caperea_marginata</i> COI-5P AP006475	0.15	0.15	0.13	0.13	0.13	0.13	0.18	0.16	0.16	0.13	0.13	0.17	0.16	0.16	0.16	0.15	0.00	-	
19 GBMA25766-19 <i>Eubalaena_glacialis</i> COI-5P MF459656	0.11	0.11	0.14	0.14	0.11	0.11	0.13	0.14	0.14	0.13	0.13	0.14	0.13	0.14	0.12	0.12	0.12	-	
20 GBMA0172-06 <i>Balaena_mysticetus</i> COI-5P AP006472	0.11	0.11	0.14	0.13	0.11	0.11	0.13	0.14	0.13	0.12	0.11	0.11	0.11	0.12	0.11	0.11	0.11	0.03	-

Zadatak 2.

Interpretirajući priloženu matricu udaljenosti sekvence nekog gena između različitih vrsta čovjekolikih majmuna odgovori na pitanja.

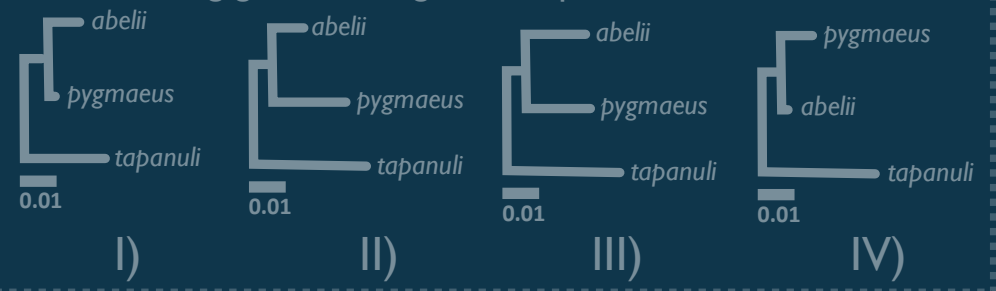
1	0.00									
2	0.01	0.00								
3	0.06	0.05	0.00							
4	0.05	0.07	0.02	0.00						
5	0.07	0.06	0.02	0.01	0.00					
6	0.10	0.09	0.11	0.12	0.10	0.00				
7	0.11	0.10	0.12	0.13	0.12	0.03	0.00			
8	0.19	0.19	0.21	0.22	0.21	0.15	0.14	0.00		
9	0.17	0.20	0.22	0.22	0.23	0.14	0.15	0.03	0.00	
10	0.17	0.18	0.20	0.21	0.22	0.13	0.14	0.01	0.04	0.00
	1	2	3	4	5	6	7	8	9	10

Br.	VRSTA	Br.	VRSTA
1	<i>Pan troglodytes</i>	6	<i>Gorilla gorilla</i>
2	<i>Pan paniscus</i>	7	<i>Gorilla beringei</i>
3	<i>Homo heidelbergensis</i>	8	<i>Pongo pygmaeus</i>
4	<i>Homo sapiens</i>	9	<i>Pongo tapanuliensis</i>
5	<i>Homo neanderthalensis</i>	10	<i>Pongo abelii</i>

PITANJA ZA INTERPRETACIJU MATRICE

- Kolike su razlike u sekvenci unutar vrsti istog roda?
- Kolika je razlika između ljudi i čimpanzi?
- Kolika je razlika između čimpanzi i gorila?
- Kolika je razlika između gorila i orangutana?

e) Koji od priloženih dendrograma točno prikazuje odnose sekvenci ovog gena u orangutanima prema matrici?



KONSTRUKCIJA

STABLA IZ MOLEKULARNIH SEKVENCI

Za konstrukciju stabla u molekularnoj filogenetici mogu se koristiti mnoge metode, a među najčešće korištenima su najveća štedljivost ili *Maximum Parsimony (MP)*, najveća vjerodostojnost ili *Maximum Likelihood (ML)* i bajezijske statistike, Bayesovog zaključivanja ili *Bayesian Inference (BI)*. Ne postoji netočno stablo, jer je svako stablo dobiveno svakom metodom točno budući da znamo iz kojeg je seta podataka i s kojim parametrima nastalo. Za razliku od toga, poznatim ili točnim evolucijskim scenarijem smatra se onaj kojem se interpretacija ne mijenja pri različitim metodama.

	Maximum Parsimony (MP)	Maximum Likelihood (ML)	Bayesian Inference (BI)
DEFINICIJA METODE	Traži stablo koje objašnjava evoluciju svih svojstava u najmanje koraka .	Traži stablo koje objašnjava evoluciju prema prethodno izračunatom modelu supstitucija.	Traži stabla na osnovi prethodne vjerojatnosti (prior), modela i nakladne vjerojatnosti (posterior).
PREDNOST(I)	Brza i (matematički) razumljiva analiza.	Detaljna analiza. Matematički razumljiva analiza.	Slična analiza kao ML, ali brža i koristi prethodno znanje.
NEDOSTATAK(-CI)	Nije uvijek najkraći evolucijski put stvaran.	Nekad jako spora analiza. Ovisna o modelu.	Kompleksna (MCMC). PP nekad puno veći od Bootstrapa.
STATISTIČKA PODRŽANOST ČVOROVA	Bootstrap vrijednost je postotak stabala u kojima čvor ima iste potomke.	Bootstrap vrijednost je postotak stabala u kojima čvor ima iste potomke.	Posterior probability (PP) je vjerojatnost da će u sljedećem stablu čvor imati iste potomke.

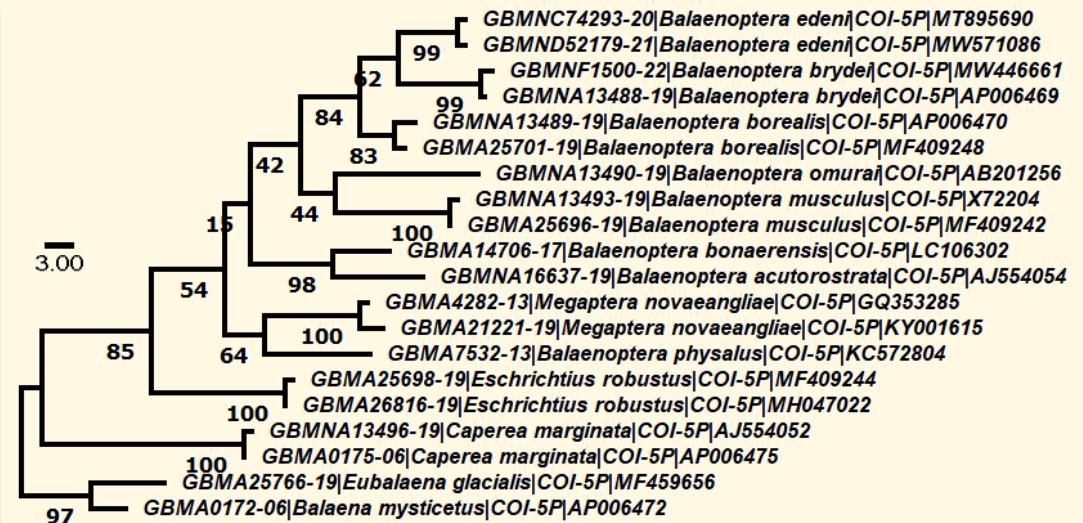
NAJVEĆA ŠTEDLJIVOST

MAXIMUM PARSIMONY, MP



PROTOKOL FILOGENIJE METODOM NAJVEĆE ŠTEDLJIVOSTI U PROGRAMU MEGA

1. Klikni Phylogeny
2. Klikni Construct/Test Maximum Parsimony Tree(s)
3. Kao test filogenije odaberi Bootstrap
4. Zadaj do 300 Bootstrap ponavljanja
5. Klikni OK i sačekaj da dobiješ stablo.



Opis stabla iz programa MEGA-11: **The evolutionary history was inferred using the Maximum Parsimony method.** Tree #2 out of 2 most parsimonious trees (length = 252) is shown. The consistency index is 0.527778, the retention index is 0.694872. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (400 replicates) are shown below the branches. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). **Branch lengths were calculated using the average pathway method and are in the units of the number of changes over the whole sequence.** This analysis involved 20 nucleotide sequences and there were a total of 469 positions in the final dataset.

MAXIMUM LIKELIHOOD

IZBOR NAJBOLJEG MODELA EVOLUCIJE

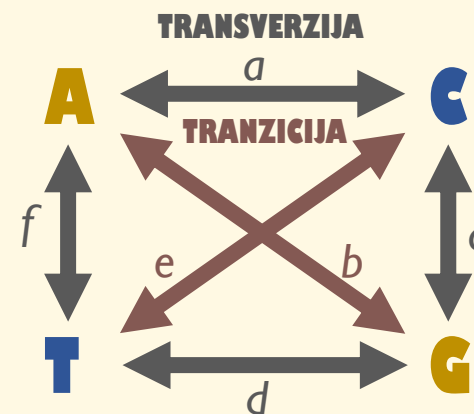
Izbor modela je jednostavna analiza koja odabire model evolucije, primjerice jedan od ovih dolje u tablici, a koji najbolje opisuje evoluciju naših sekvenci. Model za parametre ima učestalosti **TRANZICIJA** – mutacija purina u purin, pirimidina u pirimidin i **TRANSVERZIJA** – mutacija purina u pirimidin i obrnuto, kao i o količini određenih baza u setu sekvenci.

PROTOKOL IZBORA ML MODELA U PROGRAMU MEGA



1. Klikni „Models” pa „Find Best DNA/Protein Models (ML)”
2. Klikni „Compute”, tj. ne trebaš mijenjati parametre.
3. Model najnižim BIC brojem najbolje opisuje evoluciju.

MODEL	SUPSTITUCIJE	UČESTALOST BAZA	REFERENCA
JC69 (ili JC)	$a=b=c=d=e=f$	$\pi_A=\pi_C=\pi_G=\pi_T=0.25$	Jukes & Cantor (1969)
F81	$a=b=c=d=e=f$	sve π vrijednosti slobodne	Felsenstein (1981)
K2P (ili K80)	$a=c=d=f$ (TRANSVERZIJA), $b=e$ (TRANZICIJA)	$\pi_A=\pi_C=\pi_G=\pi_T=0.25$	Kimura (1980)
HKY85	$a=c=d=f$ (TRANSVERZIJA), $b=e$ (TRANZICIJA)	sve π vrijednosti slobodne	Hasegawa et al. (1985)
K3ST (ili K81)	$a=f$ (γ TRANSVERZIJA), $c=d$ (β TRANSV.), $b=e$ (TRANZICIJA)	$\pi_A=\pi_C=\pi_G=\pi_T=0.25$	Kimura (1981)
TN93	$a=c=d=f$ (TRANSV.), b (A ↔ G TRANZICIJA), e (C ↔ T TRANZ.)	sve π vrijednosti slobodne	Tamura & Nei (1993)
SYM	svi parametri slobodni, tj. svaki može biti drugačiji	$\pi_A=\pi_C=\pi_G=\pi_T=0.25$	Zharkikh (1994)
GTR (ili REV)	svi parametri slobodni, tj. svaki može biti drugačiji	sve π vrijednosti slobodne	Tavaré (1986)



PRIMJER IZBORA NAJBOLJEG MODELA

Kasnije u izradi ML dendrograma, biramo ML model koji je pri izboru modela imao najniži BIC broj.

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
GTR+G	46	3924.752	3596.484	-1752.010	n/a	0.16	17.28	0.271	0.292	0.286	0.151	0.006	0.014	0.132	0.006	0.287	0.005	0.013	0.293	0.000	0.235	0.010	0.000
GTR+G+I	47	3929.819	3594.425	-1749.971	0.59	0.92	18.37	0.271	0.292	0.286	0.151	0.005	0.015	0.124	0.005	0.299	0.004	0.014	0.306	0.000	0.221	0.008	0.000
GTR+I	46	3935.202	3606.934	-1757.235	0.71	n/a	15.86	0.271	0.292	0.286	0.151	0.008	0.013	0.147	0.007	0.264	0.005	0.013	0.269	0.000	0.263	0.010	0.001
K2+I	39	4034.980	3756.607	-1839.136	0.37	n/a	13.72	0.250	0.250	0.250	0.250	0.008	0.008	0.233	0.008	0.233	0.008	0.008	0.233	0.008	0.233	0.008	0.008
T92+I	40	4035.562	3750.060	-1834.854	0.37	n/a	13.81	0.281	0.281	0.219	0.219	0.009	0.007	0.204	0.009	0.204	0.007	0.009	0.263	0.007	0.263	0.009	0.007
TN93	42	4130.118	3830.359	-1872.986	n/a	n/a	13.07	0.271	0.292	0.286	0.151	0.011	0.011	0.100	0.010	0.320	0.006	0.010	0.327	0.006	0.179	0.011	0.011
HKY	41	4133.317	3840.686	-1879.159	n/a	n/a	13.01	0.271	0.292	0.286	0.151	0.011	0.010	0.140	0.010	0.265	0.006	0.010	0.271	0.006	0.251	0.011	0.010
GTR	45	4152.891	3831.750	-1870.653	n/a	n/a	13.13	0.271	0.292	0.286	0.151	0.011	0.018	0.100	0.010	0.321	0.004	0.017	0.328	0.001	0.180	0.007	0.003
K2	38	4157.461	3886.218	-1904.950	n/a	n/a	12.93	0.250	0.250	0.250	0.250	0.009	0.009	0.232	0.009	0.232	0.009	0.009	0.232	0.009	0.232	0.009	0.009
T92	39	4158.134	3879.761	-1900.713	n/a	n/a	12.95	0.281	0.281	0.219	0.219	0.010	0.008	0.203	0.010	0.203	0.008	0.010	0.262	0.008	0.262	0.010	0.008
JC+G	38	4332.467	4061.224	-1992.453	n/a	0.17	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G+I	39	4341.641	4063.268	-1992.467	0.00	0.17	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	38	4433.966	4162.722	-2043.202	0.37	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC	37	4548.385	4284.272	-2104.985	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

GTR—General Time Reversible, **K2**—Kimura 2-parametarski, **JC**—Jukes-Cantor, **HKY**—Hasegawa-Kishino-Yano, **T92**—Tamura 3-parametarski, **TN93**—Tamura-Nei. **+G**—diskretna Gamma razdioba (Discrete Gamma Distribution), **+I**—postoje evolucijski nevarijabilna mjesta (Invariable Sites).

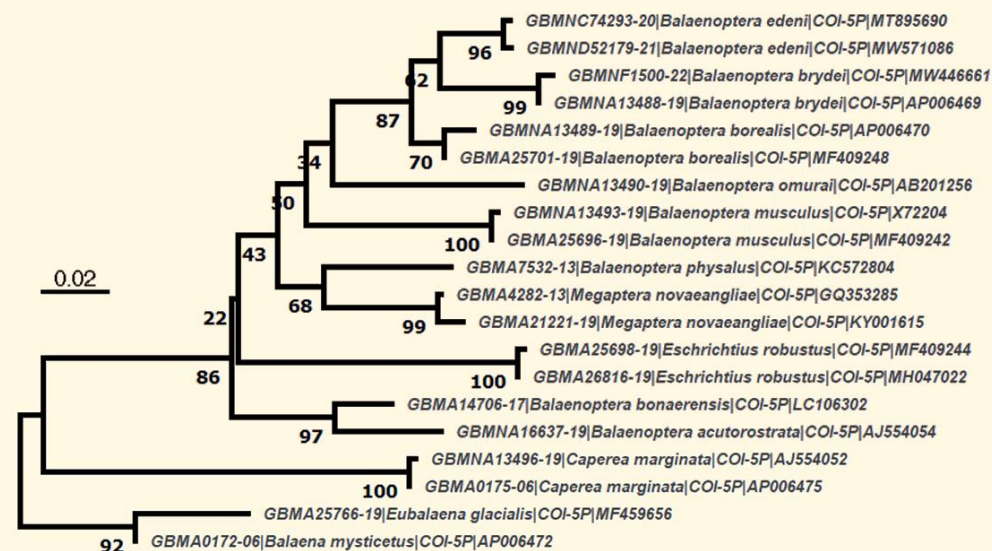
MAXIMUM LIKELIHOOD

NAJVEĆA VJERODOSTOJNOST, ML



PROTOKOL FILOGENIJE METODOM NAJVEĆE VJERODOSTOJNOSTI U PROGRAMU MEGA

1. Klikni Phylogeny
2. Klikni Construct/Test Maximum Likelihood Tree(s)
3. Kao test filogenije odaberi Bootstrap
4. Zadaj do 300 Bootstrap ponavljanja
5. Odaberi parametre modela koje si dobila/dobio i pazi da uključiš i +G i +I uz model ako je takav bio rezultat izbora najboljeg modela.
6. Klikni OK i sačekaj da dobiješ stablo.

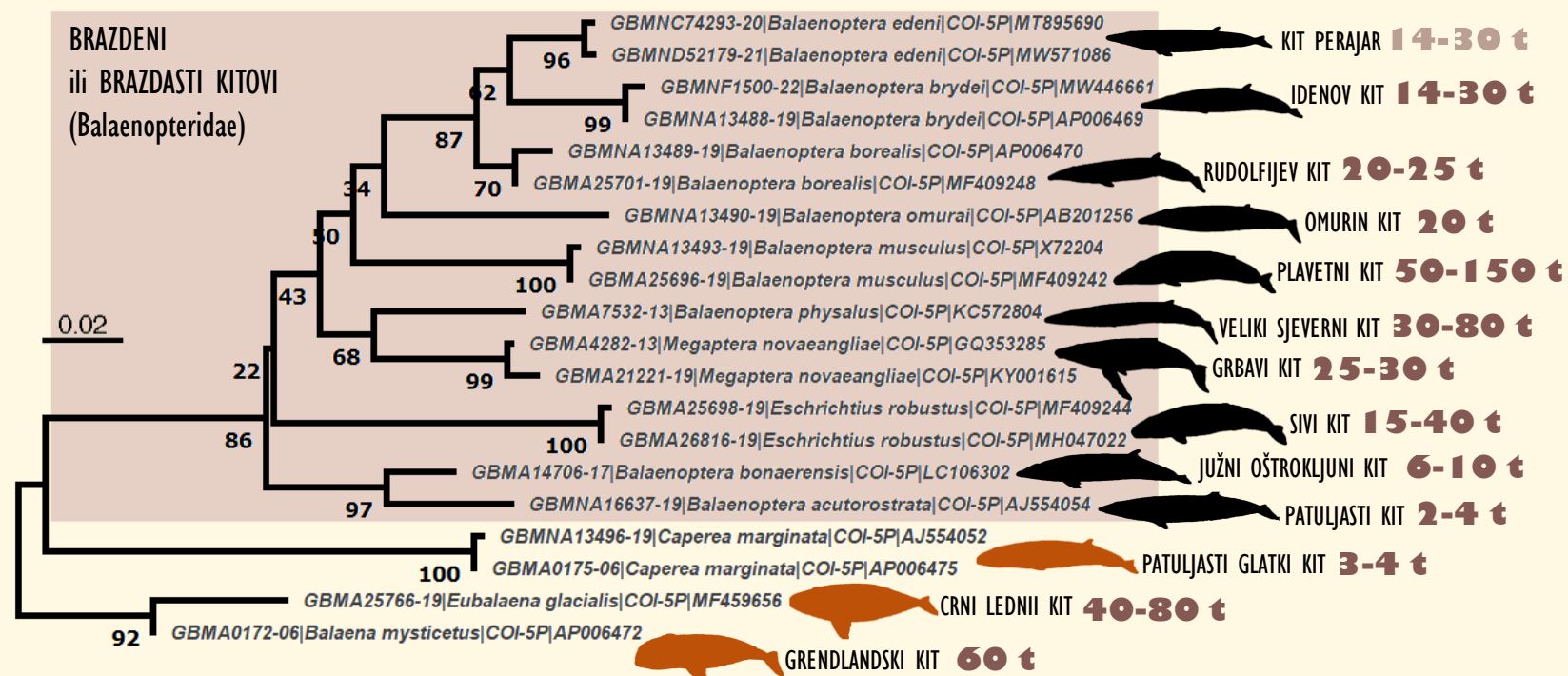


ML filogram konstruiran u programu MEGA-II uz GTR+G model
Bootstrap je izračunat na temelju 400 ponavljanja.

UREDIVANJE STABLA

U PREZENTACIJI ILI NEKOM PROGRAMU

ML filogram konstruiran u programu MEGA-II uz GTR+G model Bootstrap je izračunat na temelju 400 ponavljanja.



Gen COI rekonstruirao je rod *Balaenoptera* kao parafiletski jer su unutar njega ugniježđeni rodovi *Megaptera* i *Eschrichtius*,

Izvor silueta je rad Ren et al. (2022) DNA Barcoding Technology Used to Successfully Sub-Classify a Museum Whale Specimen as *Balaenoptera edeni*. *Frontiers in Ecology and Evolution*. www.frontiersin.org/articles/10.3389/fevo.2022.921106/full. CC-BY licenca.

Zadatak 3.

Skiciraj stabla dobivena metodama najveće štedljivosti (*maximum parsimony*) i najveće vjerodostojnosti (*maximum likelihood*) i ne zaboravi na čvorovima naznačiti bootstrap vrijednost, kao ni prikazati duljinu grana i skalu uz filogram. Potom odgovori na postavljenih 10 pitanja.

MP	ML

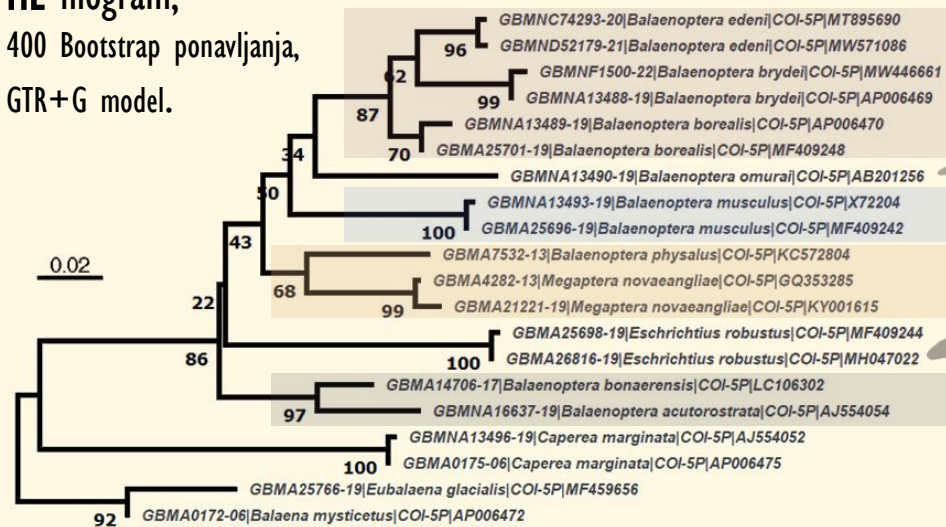
PITANJA:

- Na kojem je genu temeljena analiza i kolika je duljina sekvence?
- Koji je model izabran u ML filogeniji?
- Kolika je duljina dobivenog MP stabla u koracima?
- Je li vanjska skupina bila sestrinska ostalima?
- Jesu li sve vrste monofiletske i holofiletske?
- Koje su skupine iste na MP stablu i ML stablu?
- Koje su skupine različite na MP stablu i ML stablu?
- Jesu li podržanosti na MP stablu i ML stablu slične?
- Jesu li grane slične duljine u oba stabla?
- Odgovaraju li dobiveni odnosi objavljenim filogenijama?

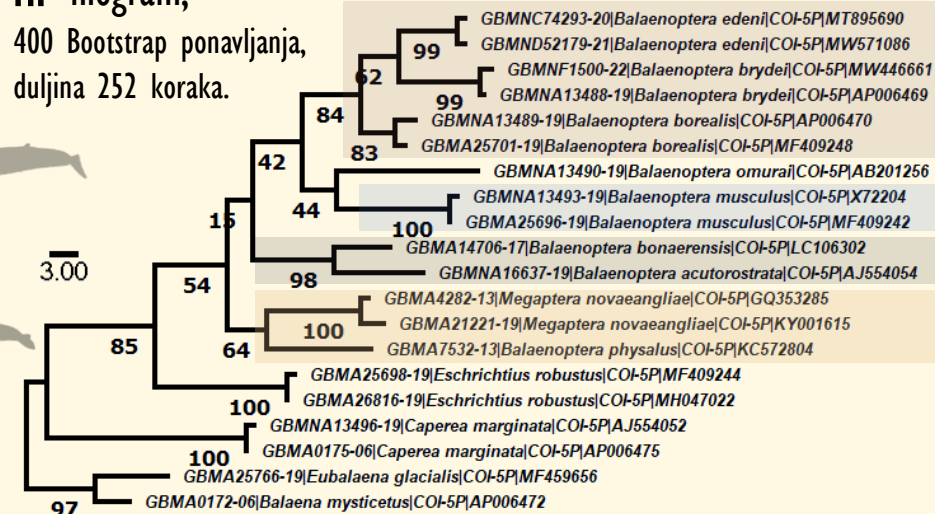
PRIMJER USPOREDBE

DVAJU DENDROGRAMA

ML filogram,
400 Bootstrap ponavljanja,
GTR+G model.



MP filogram,
400 Bootstrap ponavljanja,
duljina 252 koraka.



Najprije definiramo koje su **holofiletske** skupine na oba stabla iste. Potom gledamo skupine koje imaju različit položaj u dvama stablima i pokušavamo zaključiti zašto tako da interpretiramo **podržanosti** čvorova, **duljinu** grana i samu matricu **udaljenosti**.

LEGENDA

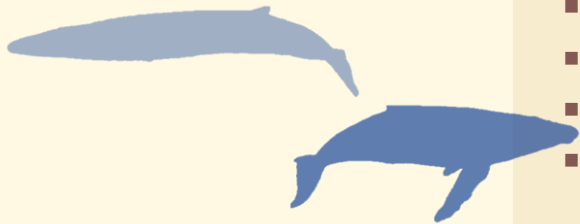
((<i>Balaenoptera_edeni</i> , <i>Balaenoptera_brydei</i>), <i>Balaenoptera_borealis</i>)
<i>Balaenoptera_musculus</i>
(<i>Balaenoptera_physalus</i> , <i>Megaptera_novaeangliae</i>)
(<i>Balaenoptera_bonaerensis</i> , <i>Balaenoptera_acutorostrata</i>)

PODSJETIMO SE I PONOVIAMO

ŠTO ZNAMO O MOLEKULARNOJ FILOGENETICI

NAUČILI SMO

- pretraživati molekularne sekvenc(ij)e
- identificirati molekularne sekvence
- organizirati molekularne sekvence
- sravnjivati/poravnavati sekvence
- izraditi matricu udaljenosti sekvenci u programu MEGA
- konstruirati kladograme i filograme u programu MEGA
- konstruirati MP (maximum parsimony) stablo
- odabrati najbolji ML (maximum likelihood) model
- konstruirati ML (maximum likelihood) stablo
- interpretirati statističku podržanost čvorova
- interpretirati duljine grana u filogramu



Izvor silueta je rad Ren et al. (2022) DNA Barcoding Technology Used to Successfully Sub-Classify a Museum Whale Specimen as Balaenoptera edeni edeni. *Frontiers in Ecology and Evolution*. www.frontiersin.org/articles/10.3389/fevo.2022.921106/full. CC-BY licenca.



Kraj pete vježbe

Sliku uslikao Mike Doherty, dostupna na linku https://unsplash.com/photos/whale-tail-on-blue-sea-during-daytime-JRsl_wfC-9A