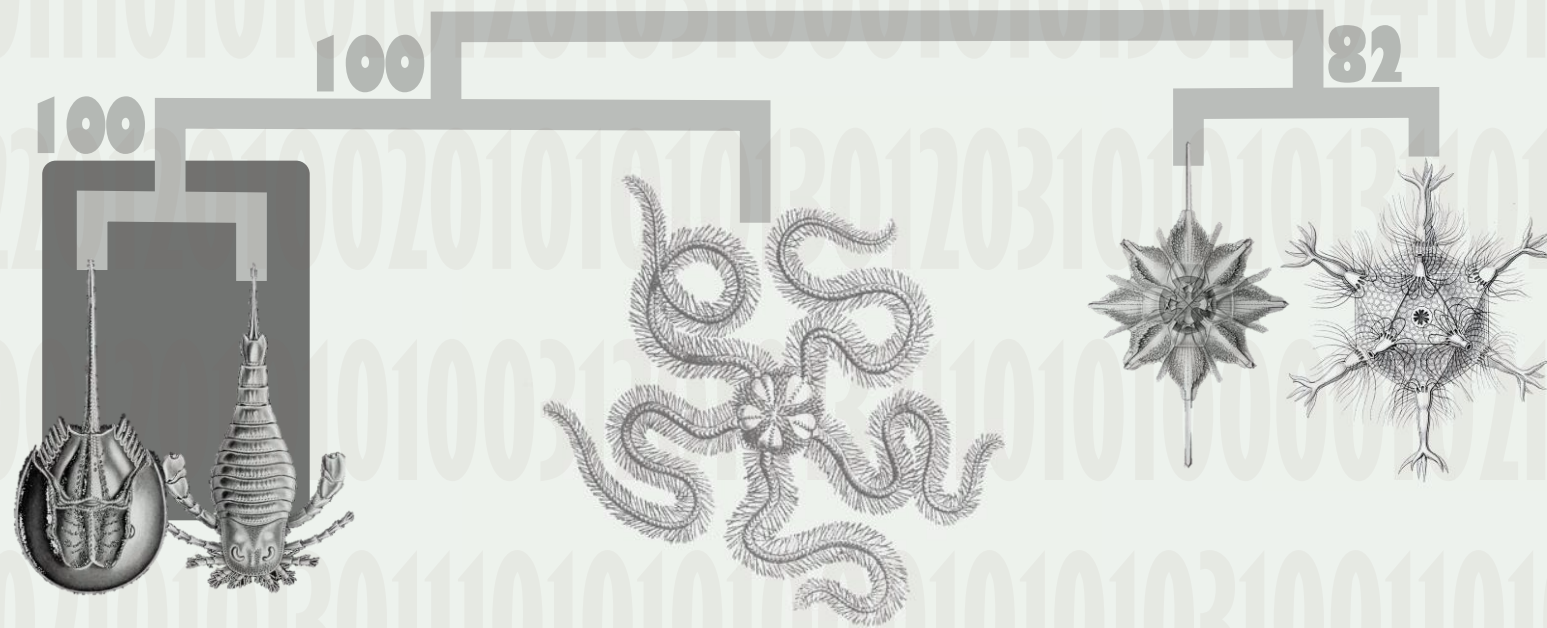


ČETVRTA VJEŽBA

KLADISTIČKA ANALIZA

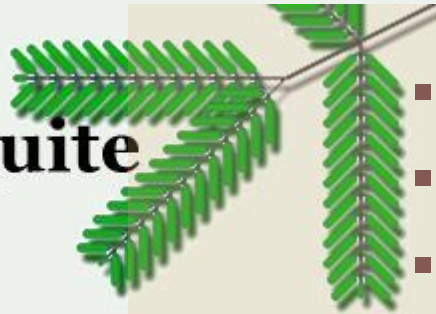


doc. dr. sc.
JOSIP SKEJO

SADRŽAJ

KONSTRUKCIJA, REKONSTRUKCIJA, RAČUNANJE

Mesquite



- Dovršavanje matrice svojstava i stanja za analizu
- Unos svojti, svojstava i stanja u program Mesquite
- **Konstrukcija** kladograma u programu Mesquite
- **Rekonstrukcija** predačkih stanja u programu Mesquite
- **Računanje** količine homoplazija i sinapomorfija na stablu

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

KLADISTIKA

grčki κλάδος, grana

Kladistika ili **filogenetička sistematika** pristup je sistematizaciji živog svijeta u kojem se empirijski matematičkim metodama kvantificira sličnost između svojti (npr. vrsta) i iz toga se utvrđuje srodnost.

ORGANIZIRANA TABLICA PODATAKA

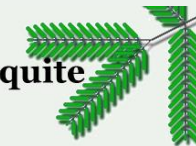
GRUPINA (latinski)	LOBAVA	HRKAJEŽICA	ČELJOST	SKELET	ŠKRLJONI OTVORI
MYXINI	PRISUTNA	MOGUĆE PRISUTNA	ODSUTNA	HRSKAVICA	OD SEDAM DO MNOGO
PETROMYZONTIDA	PRISUTNA	PRISUTNA	ODSUTNA	HRSKAVICA	SEDAM
CHONDRICHTHYES	PRISUTNA	PRISUTNA	PRISUTNA	HRSKAVICA	PET I ŠEST
ACTINOPTERYGII	PRISUTNA	PRISUTNA	PRISUTNA	KOST	ČETIRI
CROSSOPTERYGII	PRISUTNA	PRISUTNA	PRISUTNA	KOST	ČETIRI
DIPNOI	PRISUTNA	PRISUTNA	PRISUTNA	KOST	ČETIRI
AMPHIBIA	PRISUTNA	PRISUTNA	PRISUTNA	KOST	NIJEDAN
MAMMALIA	PRISUTNA	PRISUTNA	PRISUTNA	KOST	NIJEDAN
SQUAMATA	PRISUTNA	PRISUTNA	PRISUTNA	KOST	NIJEDAN
CROCODILIA	PRISUTNA	PRISUTNA	PRISUTNA	KOST	NIJEDAN
AVES	PRISUTNA	PRISUTNA	PRISUTNA	KOST	NIJEDAN
TESTUDINES	PRISUTNA	PRISUTNA	PRISUTNA	KOST	NIJEDAN
TUNICATA	ODSUTNA	ODSUTNA	ODSUTNA	ODSUTAN	-
CEPHALOCHORDATA	ODSUTNA	ODSUTNA	ODSUTNA	ODSUTAN	-

KODIRANJE

MATRICA SVOJSTAVA I STANJA

GRUPINA (latinski)	1	2	3	4	5
MYXINI	1	0/1	0	1	0&1
PETROMYZONTIDA	1	1	0	1	1
CHONDRICHTHYES	1	1	1	1	2&3
ACTINOPTERYGII	1	1	1	2	4
CROSSOPTERYGII	1	1	1	2	4
DIPNOI	1	1	1	2	4
AMPHIBIA	1	1	1	2	5
MAMMALIA	1	1	1	2	5
SQUAMATA	1	1	1	2	5
CROCODILIA	1	1	1	2	5
AVES	1	1	1	2	5
TESTUDINES	1	1	1	2	5
TUNICATA	0	0	0	0	-
CEPHALOCHORDATA	0	0	0	0	-

Mesquite



**KONSTRUKCIJA
STABLA IZ MATRICE**

POTREBNA SAMO MATRICA

**REKONSTRUKCIJA
PREDAČKIH STANJA**

NEWICKOVO STABLO + MATRICA

Podsjetimo se! Za današnje analize potrebna je **matrica svojstava** i stanja te **šest newickovih stabala**.

MATRICA

SVOJSTAVA I STANJA

SKUPINA (latinski)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MYXINI	1	0/1	0	1	0&1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
PETROMYZONTIDA	1	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
CHONDRICHTHYES	1	1	1	1	2&3	0&1	0	0	0	0	0	0	0	1	1	0	0	0	1	0
ACTINOPTERYGII	1	1	1	2	4	1	1	0&1	0	0	0	0	0	1	1	0	0	0	1	0
CROSSOPTERYGII	1	1	1	2	4	1	1	0	0	0	0	0	0	1	1	0	0	0	1	0
DIPNOI	1	1	1	2	4	1	1	0	0	0	0	0	0	1	1	0	1	0	1	0/1
AMPHIBIA	1	1	1	2	5	0	1	0	1	0	0	0	0	1	1	0	1	0	1	0&1
MAMMALIA	1	1	1	2	5	0	1	1	2	0&1	0	1	0	1	1	0	1	0	1	1
SQUAMATA	1	1	1	2	5	0	1	0	1	0&1	0	0	0	1	1	0	1	1	1	0&1
CROCODILIA	1	1	1	2	5	0	1	0	2	1	0	0	0	1	1	1	1	1	1	1
AVES	1	1	1	2	5	0	1	1	2	1	0	0	1	1	1	1	1	1	1	1
TESTUDINES	1	1	1	2	5	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1
TUNICATA	0	0	0	0	-	0	0	0	-	0	1	0	0	0	0/1	0	0	0	0	0
CEPHALOCHORDATA	0	0	0	0	-	0	0	0	-	0	1	0	0	0	0	0	0	0	0	0

Legenda svojstava i kodova stanja

NAZIV SVOJSTVA	STANJA
1 LUBANJA	0 odsutna, 1 prisutna
2 KRALJEŽNICA	0 odsutna, 1 prisutna
3 ČELJUST	0 odsutna, 1 prisutna
4 SKELET	0 odsutan, 1 hrskavica, 2 kost
5 ŠKRŽNIH OTVORA	0 >7, 1 7, 2 6, 3 5, 4 4, 5 0
6 ŠKRŽNI POKLOPAC	0 odsutan, 1 prisutan
7 PARNE PERAJE	0 odsutne, 1 prisutne
8 TEMPERATURA	0 poikilotermija, 1 endotermija
9 SRCE	0 dvo-, 1 tro-, 2 četverodijelno
10 AMNIOTSKO JAJE	0 odsutno, 1 prisutno
11 ŠKRŽNO ŽDRIJELO	0 odsutno, 1 prisutno
12 DLAKE	0 odsutne, 1 prisutne
13 PERJE	0 odsutno, 1 prisutno
14 ZUBI	0 odsutni, 1 prisutni
15 REP	0 odsutan, 1 prisutan
16 2 RUPE IZA OKA	0 odsutne, 1 prisutne
17 PLUĆA	0 odsutna, 1 prisutna
18 KERATINSKE LJUSKE	0 odsutne, 1 prisutne
19 UNUTARNJE UHO	0 odsutno, 1 prisutno
20 ČETIRI NOGE	0 odsutne, 1 prisutne

MESQUITE

PROGRAM ZA KLADISTIKU I FILOGENETIKU

mesquiteproject.org

- Organizacija podataka
 - Morfološki
 - Morfometrijski
 - Molekularni
 - Konstrukcija stabala
 - Vizualizacija stabala
 - Rekonstrukcija predačkih stanja
- Analiza principnih komponenti (PCA)
 - Mnogi dodatni paketi

Mesquite



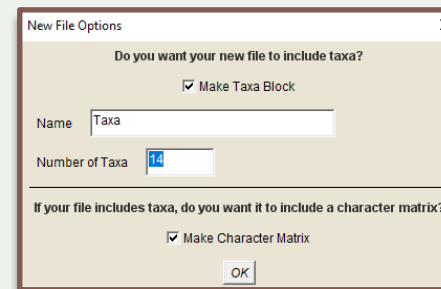
UNOS MATRICE

U PROGRAM MESQUITE

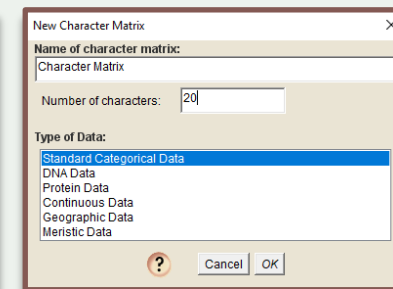
Protokol za unos matrice u Mesquite

1. Otvori Mesquite.exe, klikni „File” pa „New”
2. Daj ime projektu, npr. kladistika001 pa klikni „Save” i otvorit će se prozor.
3. Pod „Number of taxa” upiši broj svojti, 14 u našoj analizi i **klikni** uz „Make Character Matrix” pa klikni „OK”
4. Otvorit će se prozor u koji treba upisati broj svojstava, u našoj analizi 20
5. Otvorit će se prazna matrica. Iz Excel tablice **kopiraj nazive svojti** i **zalijepi ih tako da držiš tipku** **klikneš na „taxon 1”** pa „taxon 14”, **držeći** **onda zalijepi naredbom „ctrl+v”**. Na isti način možeš dodati i nazive svojstava.
6. Stanja svojstava kopiraš i zalijepiš na običan način.

PRIPAZI! Ako imaš problem sa dodavanjem naziva u matricu, provjeri da je u Mesquitu aktivna naredba  a ne 



PROZOR IZ KORAKA 2. I 3.



PROZOR IZ KORAKA 4.

Taxon \ Character	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
	LUBANJA	KRALJEVNICA	DELJIVST	SRŽANJ OTVORA	SRŽANJ POKLOPAK	PARNE PERAJE	TEMPERATURA	SRCE	ANIOGTSKO JAJE	SRŽNO ŽRUPJELO	DLAKE	PERAJE	ZUBI	REP	2. RUPJE DLA OVA	PUČA	KERATINSKE LUSKE	UNUTRANJE UNO	LETIRI UNO		
1 MYXINI	1	0/1	0	1	0&1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	
2 PETROMYZONTIDA	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
3 CHONDRICHTHYES	1	1	1	1	2&3	0&1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0
4 ACTINOPTERYGII	1	1	1	2	4	1	1	0&1	0	0	0	0	0	1	1	0	0	0	1	0	0
5 CROSSOPTERYGII	1	1	1	2	4	1	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0
6 DIPNOI	1	1	1	2	4	1	1	0	0	0	0	0	0	1	1	0	1	0	1	0/1	0
7 AMPHIBIA	1	1	1	2	5	0	1	0	1	0	0	0	0	1	1	0	1	0	1	0&1	0
8 MAMMALIA	1	1	1	2	5	0	1	1	2	0&1	0	1	0	1	1	0	1	0	1	1	1
9 SOAMATA	1	1	1	2	5	0	1	0	1	0&1	0	0	0	1	1	0	1	1	1	1	0&1
10 CROCODILIA	1	1	1	2	5	0	1	0	2	1	0	0	0	1	1	1	1	1	1	1	1
11 AVES	1	1	1	2	5	0	1	1	2	1	0	0	1	1	1	1	1	1	1	1	1
12 TESTUDINES	1	1	1	2	5	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1
13 TUNICATA	0	0	0	0	0	0	0	0	0	1	0	0	0	0/1	0	0	0	0	0	0	0
14 CEPHALOCHORDATA	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

MATRICA IZ KORAKA 5. I 6.

KONSTRUKCIJA

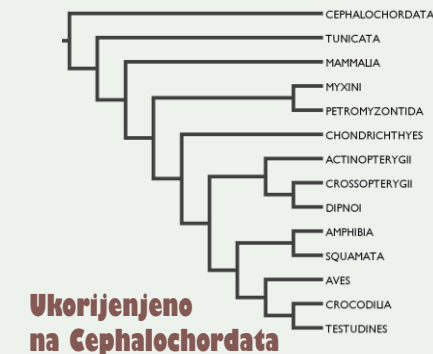
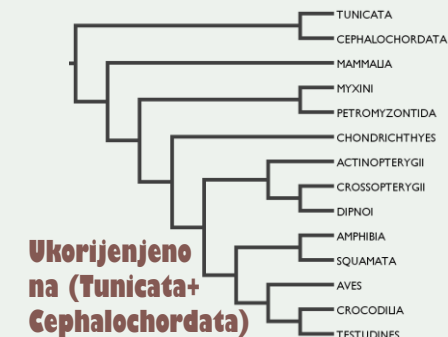
KLADOGRAMA U PROGRAMU MESQUITE

PROTOKOL ZA KONSTRUKCIJU STABLA METODOM UPGMA

1. **KLIKNI Analysis** (skroz gore nakon File, Edit...)
2. **KLIKNI Tree inference**
3. **KLIKNI Cluster Analysis**
4. **ODABERI Distances from Character Matrix i KLIKNI Ok**
5. **ODABERI Uncorrected distance (General) i KLIKNI Ok**
6. **ODABERI UPGMA analizu i klikni OK**
7. **MAXTREES broj odaberi 100 i klikni OK, pa OK**

Dobiveno stablo je neukorijenjeno i trebaš ga ukorijeniti naredbom

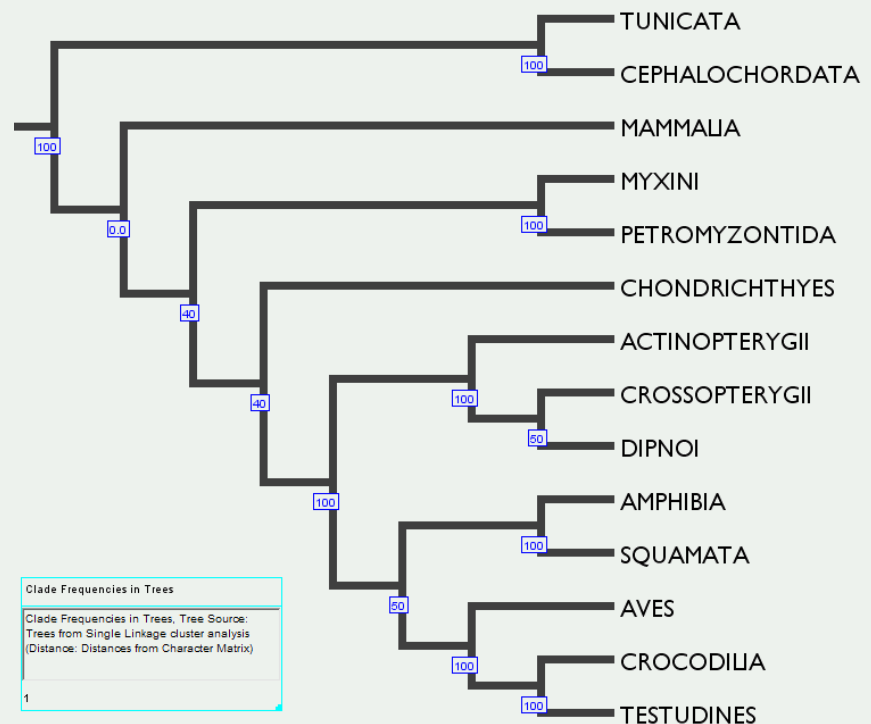
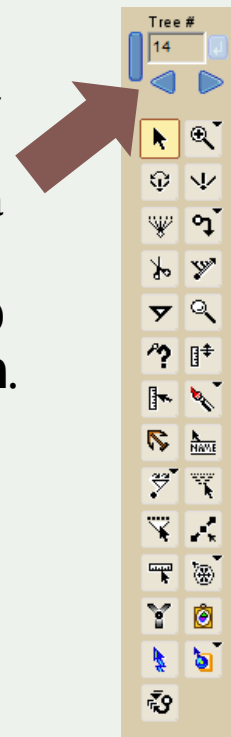
- ili na Cephalochordata
- ili na pretka Cephalochordata i Tunicata.



VIŠE STABALA KOJA SU NAJBOLJA

Budući da smo programu zadali da spremi do 100 najboljih stabala, odnosno da analizu izvrsti 100 puta i da pohrani sva najbolja stabla, sada preko plavih strelica **možemo vidjeti sva jednako dobra stabla.**

Sljedeće je pitanje kako na stablu označiti u koliko je od svih najboljih stabala neki čvor rekonstruiran na isti način.



PODRŽANOST

ČVOROVA U STABLU

Nakon što imamo skup najboljih stabala, želimo usporediti koliko često određeni čvor ima iste potomke između različitih stabala.

PROTOKOL ZA PRIKAZ PODRŽANOSTI ČVOROVA

1. **KLIKNI Analysis: Tree**

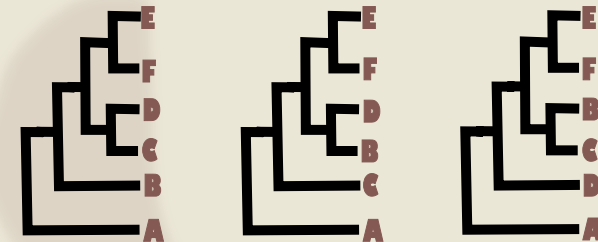
2. **KLIKNI Values for Nodes**

2. **KLIKNI Clade Frequencies in Trees** (stored trees)

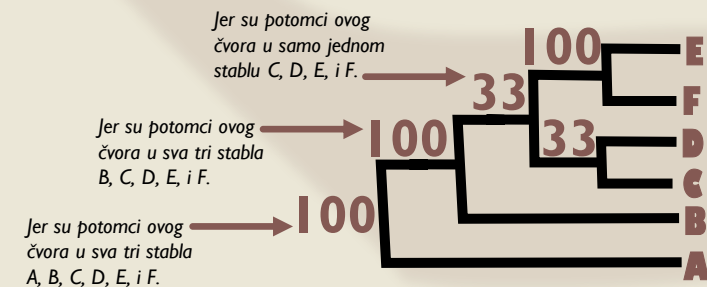
U analizi rijetko ili nikad dobijemo samo jedno stablo. Nakon što program rekonstruira mnogo stabala na temelju broja ponavljanja koje smo mu zadali, **bootstrap vrijednost** iskazuje **postotak stabala u kojem čvor ima iste potomke**.

>95 izvrsna podržanost, >90 dobra, 70-90 solidna, <70 upitna, problematična

Primjerice, ako je analiza dala ova tri stabla kao najbolja



Prvo stablo imat će ove vrijednosti podržanosti čvorova



Zadatak 1.

Nacrtaj jedan od kladograma dobivenih analizom UPGMA sa 100 ponavljanja izrađen na temelju svoje matrice svojstava i stanja u programu Mesquite i označi statističku podržanost čvorova.

REKONSTRUKCIJA

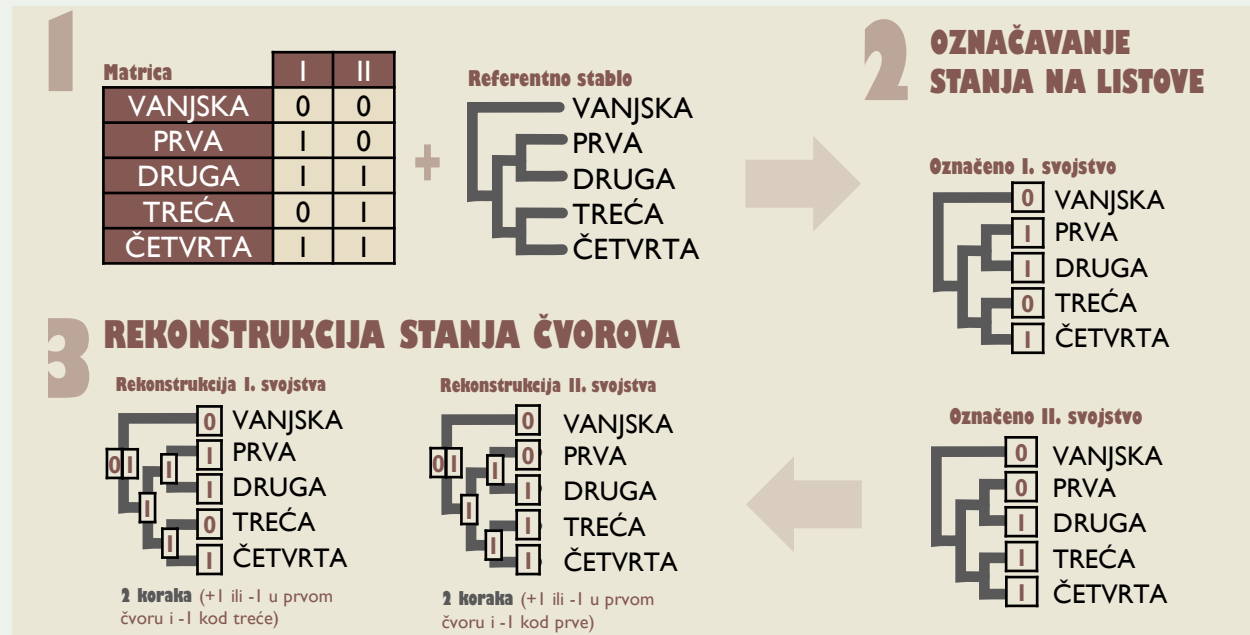
PREDAČKIH STANJA

engleski **ANCESTRAL (CHARACTER) STATE RECONSTRUCTION**

1) Za analizu rekonstrukcije predačkih stanja potrebna je numerički kodirana **matrica** svojstava i stanja i **skup referentnih** (objavljenih, tj. poznatih ili točnih) evolucijskih **dendrograma** u newickovom formatu.

2) Stanja se iz matrice označavaju **na listove** stabla i tako za svako svojstvo.

3) Ovisno o analizi koju koristimo (MP, ML) program **rekonstruira** stanja za sve **čvorove** u stablu (odnosno pretke).



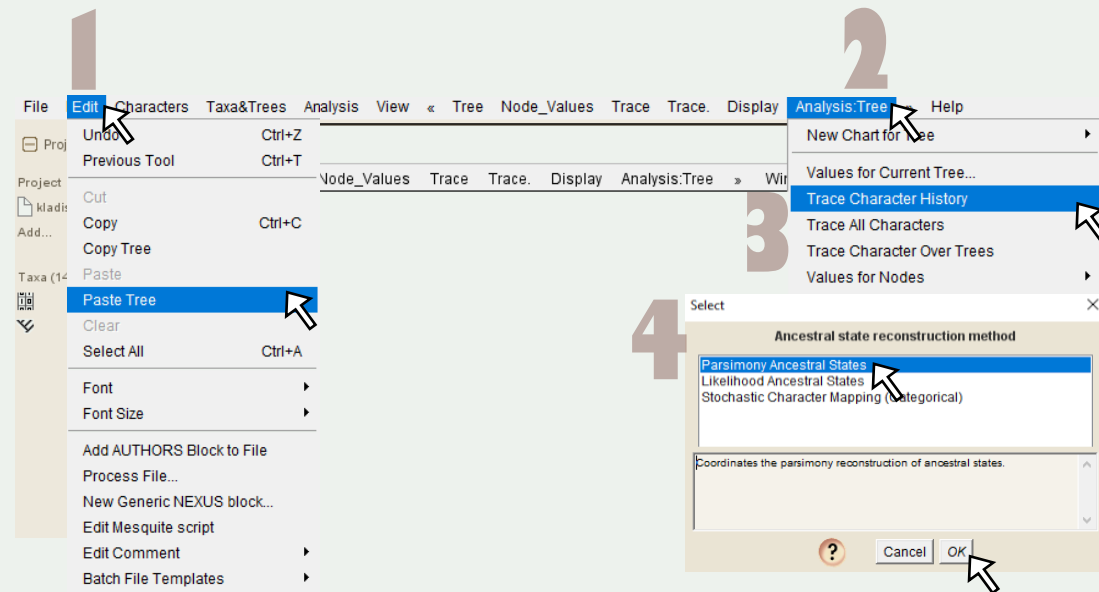
PROTOKOL

REKONSTRUKCIJE PREDAČKIH STANJA

Referentno stablo iz newickovog formata (**jedno od 6**) **treba kopirati** i pritom paziti da su **nazivi u newicku i matrici** u Mesqitu **identični**. Ako su nazivi u newicku nrodni, neka takvi budu i u programu, a ako su stručni, neka isti budu i u Mesqitu.

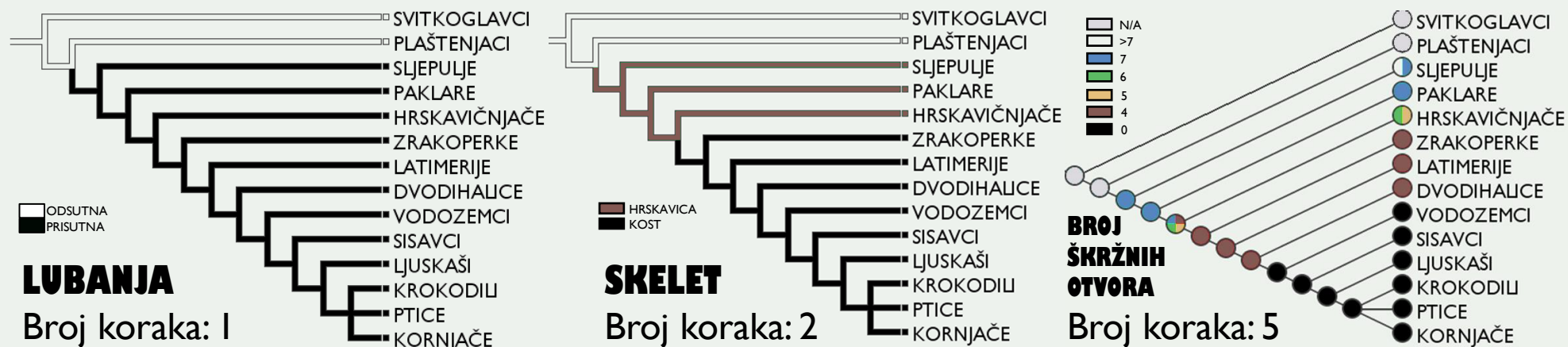
PROTOKOL ZA REKONSTRUKCIJU PREDAČKIH STANJA U PROGRAMU MESQUITE

1. Kopiraj newickovo stablo i zalijepi ga u Mesquite tako da klikneš „Edit” pa „Paste tree”. Ako su nazivi u matrici i newicku isti nema problema.
2. Klikni „Analysis: Tree” (skroz gore iza Display)
3. Klikni „Trace character history”
4. Odaberi „Parsimony Ancestral States” pa klikni „OK”.

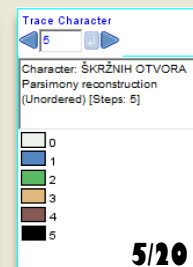
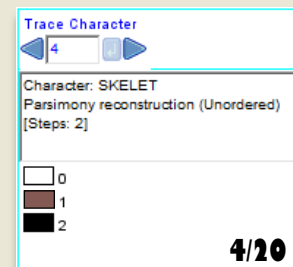
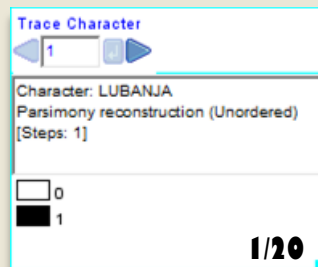


PREGLED ANALIZE

REKONSTRUKCIJA SVAKOG SVOJSTVA



Za svako svojstvo možemo pogledati rekonstrukciju predačkih stanja. Tomu nam služi plava strelica



Uza svako svojstvo u rekonstrukciji predačkih stanja navedeno je koliko je najmanje koraka bilo u evoluciji tog svojstva.

Zadatak 2.

Označi evoluciju svojstava i stanja koristeći oznake:

- + 1 dvodijelno srce
- 1 trodijelno srce
- + 1' četverodijelno srce
- + 2 parne peraje
- 2 četveronožnost
- + 3 škrge
- 3 pluća
- + 4 poikilotermija
- 4 endotermija



KLADISTIČKI INDEKSI

CONSISTENCY (CI) I RETENTION INDEX (RI)

Indeks konzistencije (CI) i indeks retencije (RI) korisne su mjere u kladističkoj analizi najveće štedljivosti (eng. *Maximum parsimony, MP*) jer nam brzo govore koliko u nekom stablu koje smo analizom dobili ima homoplazija, a koliko sinapomorfija. Analiza s mnogo homoplazija daje varljivije rezultate.

$$CI = \frac{m}{s}$$

[0,1]

0 = PUNO HOMOPLAZIJE
1 = NEMA HOMOPLAZIJE

CI mjeri količinu **homoplazija** u stablu; a
RI mjeri količinu **sinapomorfija** u stablu.

m—najmanji mogući broj koraka (promjena stanja) očekivan prema matrici
s—stvarni broj koraka na jednom promatranom stablu
g—najveći mogući broj koraka (promjena stanja) izveden prema matrici

$$RI = \frac{g - s}{g - m}$$

[0,1]

0 = NEMA SINAPOMORFIJA
1 = PUNO SINAPOMORFIJA

RAČUNANJE

INDEKSA KONZISTENCIJE I RETENCIJE

m —najmanji mogući broj koraka (promjena stanja) očekivan prema matrici
 s —stvarni broj koraka na jednom promatranom stablu
 g —najveći mogući broj koraka (promjena stanja) izveden prema matrici

	SVOJSTVA				
	I	II	III	IV	V
VANJSKA_S	0	0	0	0	0
SVOJTA_A	I	I	0	I	I
SVOJTA_B	I	I	0	0	I
SVOJTA_C	I	I	I	I	0
SVOJTA_D	I	I	I	I	I

Iz **matrice svojstava i stanja** najprije izračunamo najmanji mogući broj koraka (m) i najveći mogući broj koraka (g).

Vrijednost m izračunamo tako da popišemo sve očekivane promjene.

	PROMJENA (KORAK)
SVOJSTVO I	0 → I
SVOJSTVO II	0 → I
SVOJSTVO III	0 → I
SVOJSTVO IV	0 → I
SVOJSTVO V	0 → I

Ukupno pet koraka, dakle **$m = 5$**

Vrijednost g izračunamo tako da za svako svojstvo napišemo najveći broj koraka (promjena) stanja.

Jednostavno, u matrici za svako svojstvo pobrojimo koliko ima stanja 0, a koliko I. Broj onih kojih je manje je najveći broj koraka.

Ako svojstvo kod tri svojte ima stanje 0, a kod dvije I, upisujemo 2 za broj maksimalnih koraka, tj. računamo da je I iz 0 nastao dvaput nezavisno.

	NAJVEĆI MOGUĆI BROJ KORAKA
SVOJSTVO I	1
SVOJSTVO II	1
SVOJSTVO III	2
SVOJSTVO IV	2
SVOJSTVO V	2

Najviše 8 koraka, dakle **$g = 8$** .

PRIMJER

RAČUNANJA INDEKSA RETENCIJE I KONZISTENCIJE

Iz **matrice svojstava i stanja** imamo vrijednosti $m = 5$, odnosno $g = 8$.

	I	II	III	IV	V
VANJSKA_S	0	0	0	0	0
SVOJTA_A	I	I	0	I	I
SVOJTA_B	I	I	0	0	I
SVOJTA_C	I	I	I	I	0
SVOJTA_D	I	I	I	I	I

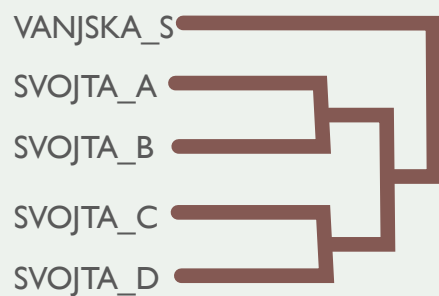
INDEKS KONZISTENCIJE

$$CI = \frac{m}{s}$$

INDEKS RETENCIJE

$$RI = \frac{g - s}{g - m}$$

Uzmimo dva stabla dobivena iz matrice svojstava i stanja s prethodnog prikaza i izračunajmo vrijednost s da izračunamo i CI i RI za oba.



	OPSERVIRANI BROJ KORAKA
SVOJSTVO I	1 (0 → I)
SVOJSTVO II	1 (0 → I)
SVOJSTVO III	1 (0 → I)
SVOJSTVO IV	2 (I → 0 dvaput)
SVOJSTVO V	2 (I → 0 dvaput)

Ukupno 7 koraka, dakle **$s = 7$** .

$$CI = \frac{5}{7} = 0.7143$$

$$RI = \frac{8 - 7}{8 - 5} = \frac{1}{3} = 0.33$$



	OPSERVIRANI BROJ KORAKA
SVOJSTVO I	1 (0 → I)
SVOJSTVO II	1 (0 → I)
SVOJSTVO III	1 (0 → I)
SVOJSTVO IV	1 (I → 0)
SVOJSTVO V	2 (I → 0 dvaput)

Ukupno 6 koraka, dakle **$s = 6$** .

$$CI = \frac{6}{7} = 0.8571$$

$$RI = \frac{8 - 6}{8 - 5} = \frac{2}{3} = 0.66$$

Zadatak 3.

Koristeći priloženu matricu svojstava i stanja izračunaj koje od priložena dva dendrograma ima više sinapomorfija, odnosno koje ima manje homoplazija.

	I	II	III	IV
VANJSKA	1	1	1	1
ALFA	0	0	0	1
BETA	0	0	1	1
GAMA	0	1	0	0
DELTA	0	0	1	0



PODSJETIMO SE I PONOVIAMO

ŠTO SMO NAUČILI O KLADISTICI

NAUČILI SMO

- odabrati svojstva potrebna za analizu i numerički im kodirati stanja
- problematiku, odnosno subjektivnost numberičkog kodiranja
- organizaciju matrice svojstava i stanja
- baratati stablom zapisanim u newickovom formatu
- (re)konstruirati stablo iz matrice svojstava i stanja u programu Mesquite
- odrediti i interpretirati statističku podržanost čvorova (Bootstrap)
- rekonstruirati predačka stanja ako znamo današnja i ako imamo stablo
- Izračunati i interpretirati indekse konzistencije i retencije za neko stablo

Kraj četrte vježbe



Izvor slike je stranica Unsplash, autor Pascal van de Vendel. Slika je dostupna na linku <https://unsplash.com/photos/AoUvQIL4FH4>