

UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN INFORMACIJSKE TEHNOLOGIJE
ŠTUDIJSKI PROGRAM BIODIVERZITETA

Brina GABERŠEK

**UPORABA FILOGENETSKIH DREVES V
MOLEKULARNI EKOLOGIJI**

ZAKLJUČNA NALOGA
Univerzitetni študij

**USING PHYLOGENETIC TREES IN MOLECULAR
ECOLOGY**

GRADUATION THESIS
University studies

Koper, 2012

Mentorica: doc. dr. Elena Varljen Bužan
Somentorica: Martina Lužnik

SEZNAM OKRAJŠAV IN SIMBOLOV

A	adenin (angl. <i>Adenine</i>)
ABI	format datoteke primeren za obdelavo v programu CodonCode Aligner
ABI	format datoteke primeren za obdelavo v programu CodonCode Aligner
AIC	Akaikov vrednostni kriterij (angl. <i>Akaike information criterion</i>)
ALN	format datoteke ki ga bere program MEGA
BIC	Bayesov vrednostni kriterij v statistiki (angl. <i>Bayesian information criterion</i>)
BLAST	orodje za lokalno poravnavanje zaporedij (angl. <i>Basic Local Alignment Search Tool</i>)
bp	bazni par
C	citozin (angl. <i>Cytosine</i>)
DM	distančna metoda (angl. <i>Distance Method</i>)
DNK	deoksiribonukleinska kislina (anlg. <i>DeoxyriboNucleic Acid</i>)
FASTA	orodje za poravnavo zaporedij in format datoteke ki ga bere program MEGA
G	gvanin (angl. <i>Guanine</i>)
GCG	format datoteke ki ga bere program MEGA
GIS	geografski informacijski sistem (angl. <i>Geographic Information System</i>)
INSDB	mednarodna baza nukleotidnih zaporedij (angl. <i>International Nucleotide Sequence Database</i>)
IUPAC	mednarodno zveza za teoretično in uporabno kemijo (angl. <i>International Union of Pure and Applied Chemistry</i>)
MEGA	program za analizo molekularne evlucijske genetike (angl. <i>Molecular Evolutionary Genetics Analysis</i>)
MEG	format datoteke primeren za obdelavo v programu MEGA
ML	metoda največjega verjetja (angl. <i>Maximum Likelihood Approach</i>)

MP	metoda največje varčnosti (angl. <i>Maximum Parsimony Method</i>)
mtDNK ali mDNK	mitohondrijska DNK
NBRF	format datoteke ki ga bere program MEGA
NEXUS	format datoteke ki ga bere program MEGA
NJ	metoda združevanja sosedov (angl. <i>Neighbor-joining Method</i>)
PCR	verižna reakcija s polimerazo (angl. <i>Polymerase Chain Reaction</i>)
PHYLIP	format datoteke ki ga bere program MEGA
PIR	format datoteke ki ga bere program MEGA
RAPD	naključno pomnoževanje polimorfne DNK (angl. <i>Random Amplification of Polymorphic DNA</i>)
RNK	ribonukleinska kislina (angl. <i>RiboNucleic Acid</i>)
SCF	format datoteke primeren za obdelavo v programu CodonCode
T	timin (angl. <i>Thymine</i>)
UPGMA	metoda netehtane aritmetične sredine (angl. <i>Unweighted Pair Group Method with Arithmetic Mean</i>)

KAZALO VSEBINE

SEZNAM OKRAJŠAV IN SIMBOLOV	II
KAZALO VSEBINE.....	IV
KAZALO SLIK.....	VII
KAZALO PREGLEDNIC.....	VIII
1 UVOD.....	1
1.1 DEOKSIRIBONUKLEINSKA KISLINA – DNK (angl. <i>DeoxyriboNucleic Acid</i>).....	3
1.2 MITOHONDRIJSKA DNK (<i>mtDNK</i> ali <i>mDNK</i>)	3
1.3 CITOKROM <i>B</i>	4
2 PREDSTAVITEV PROBLEMA	7
2.1 FILOGENIJA	7
2.2 FILOGEOGRAFIJA.....	7
2.3 FILOGENETSKA DREVESA.....	8
2.4 METODA NAJVEČJE VARČNOSTI (MP)	10
2.5 DISTANČNA METODA (DM).....	11
2.6 METODA NAJVEČJEGA VERJETJA (ML)	12
2.7 MODELI MOLEKULARNE EVOLUCIJE.....	12
2.7.1 Markov model.....	13
2.7.2 Model zamenjave nukleotidov.....	13
2.7.3 Model zamenjave aminokislin.....	14
2.7.4 Stopnja heterogenosti.....	14
2.7.5 Združeni modeli.....	14
3 POSTOPEK RISANJA FILOGENETSKIH DREVES	15
3.1 PORAVNAVANJE ZAPOREDIJ	15
3.1.1 FASTA.....	15
3.1.2 BLAST (angl. <i>Basic Local Alignment Search Tool</i>).....	18

3.2	ANALIZA ZAPOREDIJ: CODON CODE ALIGNER	20
3.2.1	Koda IUPAC, ki ga uporablja CodonCode Aligner.....	21
3.2.2	Potek dela v programu.....	23
3.2.2.1	Glavno okno	23
3.2.2.2	Projektno okno – delo z nukleotidnimi zaporedji.....	24
3.2.2.3	Izris drevesa.....	27
3.3	PORAVNAVA ZAPOREDIJ IN IZRIS FILOGENETSKEGA DREVESA: MEGA (angl. <i>Molecular Evolutionary Genetics Analysis</i>).....	28
3.3.1	Koda IUPAC, ki jo podpira MEGA	30
3.3.2	Metoda vezanja ali test vezanja (angl. <i>Bootstrap</i>)	30
3.3.3	Potek dela v programu.....	31
3.3.3.1	Urejanje (nukleotidnih) zaporedij.....	31
3.3.3.2	Konstrukcija dreves	33
3.3.3.3	Ocenjevanje parnih razdalj	36
4	DISKUSIJA	39
5	ZAKLJUČEK.....	42
6	POVZETEK	44
6.1	POVZETEK.....	44
6.2	SUMMARY	45
7	VIRI IN LITERATURA.....	48

KAZALO SLIK

- Slika 1:** Slika prikazuje celico in DNK, ki se nahaja v njej. V jedru celice je jedrna DNK, v mitohondrijih pa mitohondrijska DNK. (Vir: <http://www.dnahandbook.com/pages/10176.htm>)...3
- Slika 2:** Shematski prikaz mitohondrijskega genoma. (Vir: <http://genesinfo.blogspot.com/>).....5
- Slika 3:** Trodimenzionalna struktura citokroma *b*. (Vir: <http://chemistry.umeche.maine.edu/MAT500/Proteins9.html>).....5
- Slika 4:** Grafični prikaz filogenetskega drevesa in njegovih delov. Obravnavani vrsti D in E z njunim razvejiščem v tem primeru tvorita monofiletsko skupino. B in C (brez A) tvorita parafiletsko skupino. E in A tvorita polifiletsko skupino (Trontelj, 2009).9
- Slika 5:** Nekaj načinov prikaza drevesa s petimi taksoni. **Zgoraj levo** je koreninjeno drevo, dolžina vej pa prikazuje, koliko so stopnje evolucije oddaljene med seboj. Korenina je skrajno levo. Vidimo, da sta bila taksona E in C deležna veliko sprememb, odkar sta divergirala iz zadnjega skupnega prednika. Taksona B in D pa sta bila deležna le malo sprememb. **Spodaj levo** je koreninjeno drevo, vendar veje nakazujejo samo sorodstvene odnose, njihova dolžina ne da nobene informacije. **Desno** je nekoreninjeno drevo, skupen prednik torej ni znan. (Vir: <http://www.talkorigins.org/faqs/comdesc/phylo.html>) 10
- Slika 6:** Primerjava nukleotidnih zaporedij *Panthera onca* in *Panthera pardus* v FASTA formatu. Pove, v koliko nukleotidih se zaporedji ujemata (v našem primeru 90 %). 17
- Slika 7:** Primer FASTA nukleotidnih zaporedij vrst *Uncia uncia*. V prvi vrsti so informacije o nukleotidnem zaporedju (vrstno ime in kateri del gena nukleotidno zaporedje predstavlja). 17
- Slika 8:** Orodje BLAST, prosto dostopno na spletu 19
- Slika 9:** Z BLAST-om primerjamo nukleotidno zaporedje *Panthera leo* z nukleotidnim zaporedjem *Panthera pardus*. 19
- Slika 10:** Pozdravno okno programa CodonCode Aligner 4.0.3 20
- Slika 11:** Program ob zagonu ponudi, da izberemo, kateri projekt bomo odprli. 24
- Slika 12:** Projektno okno, v katerega uvažamo datoteke nukleotidnih zaporedij. Tu jih lahko združimo ali jih zavržemo. Lahko jih preimenujemo itd. 25
- Slika 13:** Dvojni pritisk na *Contig* z levo tipko miške nam odpre okno, v katerem lahko obdelujemo in popravljamo naše končno nukleotidno zaporedje. V zgornjem delu okna vidimo dolžino celotnega nukleotidnega zaporedja in tudi, koliko so dolgi posamezni fragmenti. Če pa izberemo gumb *Show differences*, lahko vidimo vse baze, kjer se posamezna nukleotidna zaporedja

razlikujejo. V našem primeru je v končnem nukleotidnem zaporedju označena črka Y. To seveda ni baza, je simbol iz kode IUPAC (Preglednica 1). V primeru, da bi delali z aminokislinami, bi uporabili Preglednico 2.26

Slika 14: Tako izgleda okno s sledovi (angl. *trace*), ki ga vidimo, če imamo datoteke v ABI, AB1 ali SCF formatu. V tem primeru je označeno mesto, kjer je možno, da je prišlo do mutacije. (Vir: http://www.codoncode.com/aligner/quicktour/specify_regions.htm).....27

Slika 15: Pozdravno okno programa MEGA 5.05.....28

Slika 16: Opozorilno okno. Preveriti moramo uspešnost prevajanja formatov.....29

Slika 17: Okno za urejanje tekstovnih datotek in prevajanja formatov. V prvi vrstici mora biti ključna beseda #Mega. To pove, da je datoteka v MEGA formatu. V naslednji vrstici sledi !Title in nato ime datoteke z njenim formatom, ki smo jo izbrali (v našem primeru sekvence.txt). Stavček se zaključí z znakom »;«. Nato sledi nukleotidno zaporedje ali aminokislinsko zaporedje. Pišemo lahko tudi komentarje, ki lahko obsegajo poljubno število vrstic. Označeni morajo biti z znakoma »[komentar]«.29

Slika 18: Okno za urejanje poravnav.....31

Slika 19: Okno »*Alignment Explorer*«, v katerem lahko med drugim tudi poravnamo nukleotidna zaporedja.32

Slika 20: Prejšnje nukleotidno zaporedje je prevedeno v aminokislinsko zaporedje.....33

Slika 21: Izračunan model evolucije za ML.....34

Slika 22: Izrisano *Maximum Likelihood* drevo osebkov rodu *Panthera* in *Neofelis* na njihovih nukleotidnih zaporedjih gena citokroma *b*. Uporabili smo še *Bootstrap* metodo s 1500 ponovitvami in model *Kimura 2-parameter*. Skala prikazuje 2 % divergenco nukleotidnih zaporedij.....35

Slika 23: Če v oknu *TreeExplorer* pritisnemo File, Show Information (Ctrl+I), dobimo informacije o drevesu. V našem primeru je drevo nekoreninjeno, model je *Kimura 2-parameter*.....35

Slika 24: Za primerjavo s prejšnjim drevesom je tukaj izrisano *Neighbor-Joining* drevo osebkov rodu *Panthera* in *Neofelis* na njihovih nukleotidnih zaporedjih gena citokroma *b*. Uporabili smo še *Bootstrap* metodo z 1000 ponovitvami. Skala prikazuje 1 % divergenco nukleotidnih zaporedij.36

KAZALO PREGLEDNIC

Preglednica 1: Koda IUPAC za nukleotidna zaporedja (DNK/RNK).....	21
Preglednica 2: Koda IUPAC za aminokislinska zaporedja (protein)	22

1 UVOD

Ekologija je obširna in raznolika znanstvena veda. Pokriva zelo široko področje in je ena najpomembnejših znanosti, ki pomaga pri upravljanju okolja (Egerton, 2001). Proučuje značaj, razporeditev, številčnost, pomen in spremembo stanja organizma znotraj enega ekosistema in med različnimi ekosistemi. Ekologi se med drugim trudijo razložiti življenjske procese in prilagoditve organizmov, potovanje snovi in energije skozi ekosisteme ter razporeditev in bogastvo biodiverzitete (Chapman in Reiss, 1999).

Praktično uporabo ekologije lahko zasledimo tudi v varstveni biologiji, pri upravljanju območij (npr. z naravnimi viri), načrtovanju mest ter še na veliko drugih naravoslovnih in družboslovnih področjih. Na ta način se je razvilo več podpodročij ekologije (Chapman in Reiss, 1999): vedenjska ekologija, populacijska ekologija, ekologija ekosistemov, evolucijska ekologija in nenazadnje molekularna ekologija, ki bo predmet obravnave te naloge.

Taksonomija je veda, ki organizme razvrsti v sistem in tako omogoči lažji pregled nad organizmi. Taksoni so v sistemu hierarhično urejeni. Hierarhija se tvori na podlagi sorodstvenih odnosov taksonov (predniški takson in potomci). Z vidika določenega taksona, je ekologija skupek vseh interakcij tega taksona z živim in neživim okoljem.

Molekularna ekologija je pomembna zveza med ekologijo in evolucijo oz. genetsko dediščino. Pri odgovarjanju na ekološka vprašanja (vedenjske ekologije, odnosa med organizmom in okoljem ...), si molekularna ekologija pomaga s populacijsko genetiko, molekularno filogenijo in genomiko. S pomočjo analize zaporedij mitohondrijskih in kloroplastnih genov ter tehniko mikrosatelitne DNK, se določa genski pretok in hibridizacija med populacijami (Hadrys in sod., 1992).

Molekularna ekologija ugotavlja kateri vrsti oz. populaciji organizem pripada, proučuje dednost in genetske spremembe v evoluciji, spreminjanje in prilagajanje organizmov na genetskem nivoju, ugotavlja sorodstvene odnose med organizmi (taksoni). Sorodna je področjema ekološke genetike in varstvene genetike (Freeland in sod., 2011).

Varstvena biologija metode molekularne ekologije uporabi z namenom ohranjati vrste, njihove habitate in ekosisteme. Zadnjih nekaj let je velik poudarek na proučevanju DNK oz. nukleotidnih zaporedjih genov. Na tak način lahko izvemo pomembne informacije o določenem osebk ali/in populaciji. S pridobljenimi podatki lahko osebk/populacijo uvrstimo v filogenetsko drevo, sledimo lahko sorodstvenim odnosom osebka/populacije in ugotavljamo prednike. Te informacije potrebujemo pri načrtovanju upravljanja z biodiverzitetjo in ohranjanju vrst. Pomembno je skrbeti, da je v populaciji zadosten pretok genov in tako zadostna diverzitetja genov. Le tako ostaja določena populacija vitalna in stabilna ter ima možnost nadaljnje evolucije.

Vredno je omeniti, da je proučevanje DNK pomembno tudi v forenziki in medicini. Od začetka devetdesetih let prejšnjega stoletja se uporablja metoda verižne reakcije s polimerazo (angl. *Polymerase Chain Reaction* ali PCR). Forenziki uporabljajo molekularne genetske teste za identifikacijo bioloških sledi in preverjanje sorodstvenih povezav. Na ta način se lahko identificirajo storilci in žrtve kaznivih dejanj, pa tudi žrtve naravnih ali prometnih nesreč (Drobnič, 2012).

V medicini je analiza koristna predvsem za ugotavljanje bolezni ter preprečevanje in zdravljenje le-teh. V današnjem času, ko je večja možnost rakavih obolenj, je še posebej pomembno, da se bolezen dovolj zgodaj odkrije in se začne z zdravljenjem. Ugotavljajo pa se lahko tudi sorodstveni odnosi in predniki posameznikov.

Za dobre rezultate je potrebna ustrezna obdelava nukleotidnih zaporedij. Tu se varstvena biologija in z njo molekularna ekologija prepletata z bioinformatiko. Za obdelovanje in analizo nukleotidnih zaporedij obstaja veliko programov. Naloga opisuje nekaj izbranih programov in delo z njimi. Podrobneje sta opisani orodji FASTA in BLAST ter programa CodonCode Aligner 4.0.3 in MEGA 5.05. Okvirno je predstavljen potek obdelave nukleotidnih zaporedij od neobdelanih nukleotidnih zaporedij do izdelave filogenetskega drevesa. Nukleotidna zaporedja za primer so izbrana iz spletne baze podatkov GenBank. V nalogi so opisani postopki za analizo nukleotidnih zaporedij, ki oja večinoma veljajo tudi za aminokislinska zaporedja.

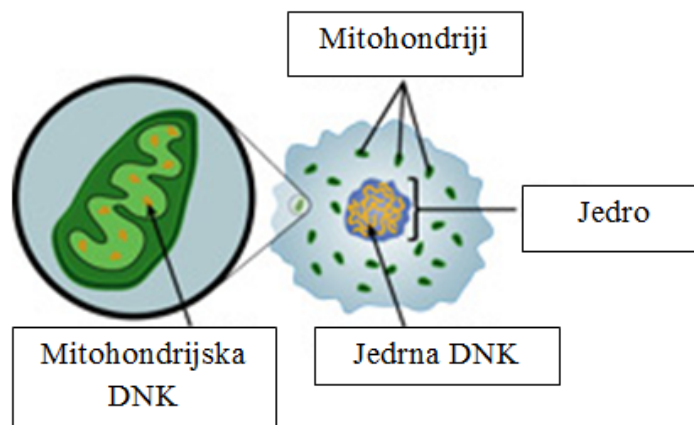
1.1 DEOKSIRIBONUKLEINSKA KISLINA – DNK (angl. *DeoxyriboNucleic Acid*)

Genske informacije v posameznih organizmih so zbrane v dolgi molekuli, imenovani DNK. Sestavljena je iz štirih nukleotidnih baz: adenina (A), timina (T), citozina (C) in gvanina (G). Adenin in gvanin sodita med purine, timin in citozin pa med pirimidine. Baze se med seboj vežejo v pare A in T z dvojno vezjo ter G in C s trojno vezjo. Vsaka nukleotidna baza pa je povezana s sladkorjem in fosfatom. Vse skupaj tvori dvojno vijačnico.

Pojavijo se lahko tudi mutacije, ki so gonilo evolucije. Na njih temeljijo analize DNK, si so opisane v nalogi. Zamenjava (oz. substitucija) je kadar se baze v zaporedju zamenjajo. Če adenin zamenja gvanin in obratno ali citozin zamenja timin in obratno, se zamenjava imenuje tranzicija. Če purin zamenja pirimidin in obratno, se zamenjava imenuje trasverzija.

Pri živalih poznamo jedrno in mitohondrijsko DNK. Za ugotavljanje sorodnosti organizmov kot so vretenčarji in nekateri nevretenčarji (Kocher in sod., 1989), je bolj pomembna mitohondrijska DNK, saj ji zaradi njenih lastnosti lažje sledimo. Lastnosti so opisane v nadaljevanju. Pri rastlinah se v raziskavah sorodnosti več uporablja kloroplastna DNK (Palmer, 1992).

1.2 MITOHONDRIJSKA DNK (*mtDNK* ali *mDNK*)



Slika 1: Slika prikazuje celico in DNK, ki se nahaja v njej. V jedru celice je jedrna DNK, v mitohondrijih pa mitohondrijska DNK. (Vir: <http://www.dnahandbook.com/pages/10176.htm>)

Nahaja se v mitohondrijih celice (Slika 1). Je krožna molekula in pri človeku vsebuje 16.569 baznih parov (bp) (Alexeyev in sod., 2004). Pri vretenčarjih in nevretenčarjih je dolga približno od 15.000 do 18.000 bp (Carr, 2009). Nekatere rastline lahko imajo tudi do 2.400.000 bp, vendar število genov ostaja isto kot pri rastlinah z manjšim številom bp (McClellan, 1998).

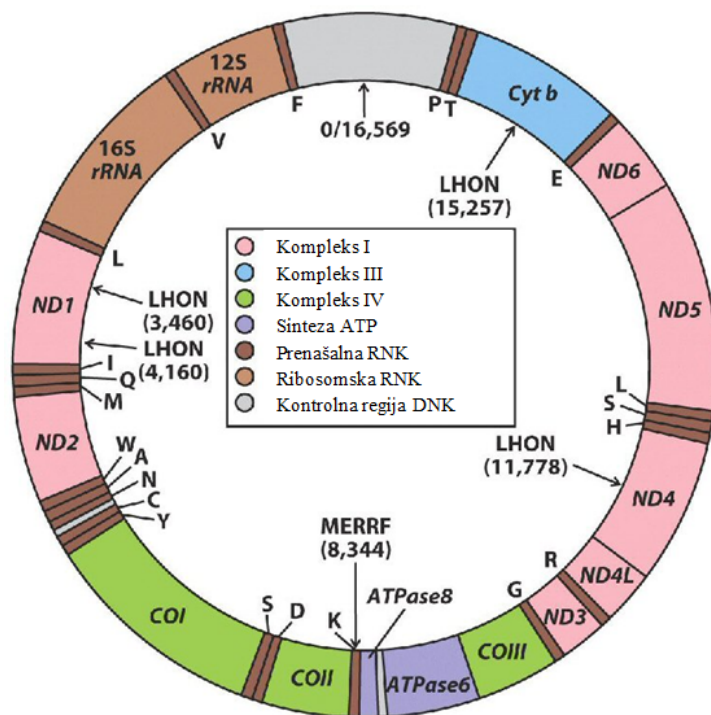
Mendelsko dedovanje za *mtDNK* ne velja. Deduje se samo po materi (maternalna dednost) in se prenese na vse potomce, ne glede na spol. Mitohondriji semenčic oz. spermijev so po oploditvi uničeni, tako ostanejo samo mitohondriji matere (Rokas in sod., 2003). V mitohondriju je prisoten večji polimorfizem, ker kopiči mutacije veliko hitreje kot jedrna DNK. Pri delitvi mitohondrijev se *mtDNK* prepíše, poleg nje pa tudi vse mutacije, ki so nastale do trenutka delitve. Novonastale celice poleg podedovanih mutacij pridobijo še nove mutacije. To se ponavlja iz generacije v generacijo mitohondrijev in mutacije se kopičijo.

Mutacije so pogoste, saj je *mtDNK* izpostavljena delovanju reaktivnih kisikovih spojin, kot sta O_2^- in H_2O_2 . Popravljalni mehanizem za *mtDNK* pa ni tako učinkovit, kot je za jedrni DNK. Zaradi tega tudi prihaja do staranja in mitohondrijskih bolezni (Wei in sod., 2001). *MtDNK* se ne rekombinira, zato je sledenje mutacijskim spremembam v primerjavi z jedrno DNK veliko lažje (Rokas in sod., 2003).

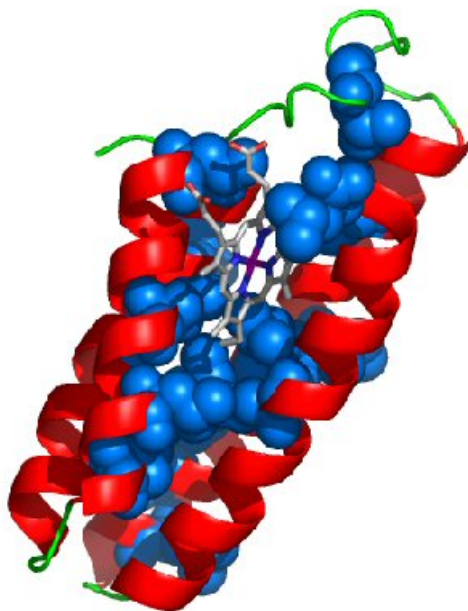
Osebek ima načeloma enako *mtDNK* v vseh telesnih celicah, čemur rečemo homoplazmija. Lahko pa pride do mutacije in ima isti osebek v telesu dve različni *mtDNK*, kar imenujemo heteroplazmija.

1.3 CITOKROM B

Beljakovino kodira *mtDNK* (Slika 2) in je centralna podenota citokroma *c* (*bc* kompleks). Spada v družino hemoproteinov. Sestavljen je iz 400 do 500 aminokislin in ene hem prostetične skupine. Trodimenzionalno strukturo citokroma *b* prikazuje slika 3. Obstajajo trije tipi citokromov: *a*, *b* in *c*. V ciklusu citronske verige oz. oksidativni dihalni verigi skrbi za prenos elektronov. Nahaja se na zunanji membrani mitohondrija. Citokrom *b* sodeluje v kompleksu III, ki je eden od štirih kompleksov dihalne verige (Degli Esposti in sod., 1993).



Slika 2: Shematski prikaz mitohondrijskega genoma. (Vir: <http://genesinfo.blogspot.com/>)



Slika 3: Trodimenzionalna struktura citokroma *b*. (Vir: <http://chemistry.umeche.maine.edu/MAT500/Proteins9.html>)

Najbolj priljubljena gena za delo v filogeniji vretenčarjev (Johns in Avise, 1998) in višjih rastlin (Asard in sod., 1989) sta gena za citokrom *b* in citokrom oksidazo 1. Citokrom *b* dovolj variira, da lahko s proučevanjem njegovega nukleotidnega zaporedja odgovorimo na filogenetska vprašanja na nivoju populacije. Hkrati pa se dovolj ohranja, da lahko z njim ugotavljamo daljne sorodstvene odnose. Zaradi naštetih lastnosti se zelo pogosto uporablja v genetskih raziskavah, predvsem pri proučevanju sesalcev. Različne študije so zato brez težav primerljive (Degli Esposti in sod., 1993).

2 PREDSTAVITEV PROBLEMA

2.1 FILOGENIJA

Paleontologi želijo razumeti, kako je potekala evolucija oz. kako se je razvijalo življenje skozi čas. Navadno se zanimajo za daljša časovna obdobja. Preden se lotijo natančnejšega proučevanja, morajo organizme uvrstiti v čas. To jim omogoča filogenija. Veja filogenije je tudi filogenetika, ki s pomočjo taksonomije pomaga pri sistematiki (Guralnick in sod., 1994).

»Filogenija je uvrščanje bioloških vrst v sistematske enote, ki so zasnovane na sorodstvenih odnosih med vrstami« (Svarog, Filogenija, 2012). Sorodstvene odnose ponazorimo s kladogrami.

Cilj filogenije je analizirati podatke, ki so na voljo, in z njihovo pomočjo ugotoviti potek evolucije tekom časa. Končni rezultati se prikažejo v obliki sistema. Kadar ni dovolj fosilov, je potrebno pridobiti več drugih podatkov s teoretičnim ali biološkim proučevanjem. Vsaka hipoteza, ki nastane tekom dela, mora nujno imeti dovolj trdne dokaze. Vseeno pa se moramo zavedati, da je filogenetsko drevo le domneva. Sčasoma se lahko spremeni, če se najdejo nove najdbe in novi dokazi (Svarog, Filogenija, 2012).

2.2 FILOGEOGRAFIJA

Filogeografija se ukvarja s proučevanjem ter analizo geografskih podatkov in genetske variabilnosti. Proučuje prostorsko razporeditev posameznih genotipov in upošteva dogajanja v preteklosti (Kidd in Ritchie, 2006).

Nepogrešljiva je pri pojasnjevanju geografskih vzorcev evolucije znotraj vrst ali med ozko sorodnimi vrstami (zato so osnovne enote filogeografije monofiletske skupine). Pomaga pri raziskovanju širokega spektra problemov, ki so povezani z biogeografijo (npr. genski pretok, učinek »ozkega grla«, širjenje populacij, poti kolonizacije in vikariantski dogodki) (Arbogast in Kenagy, 2001).

Pri analizah se uporablja geografski informacijski sistem (GIS). To je računalniški sistem, ki nam pomaga v celoto povezati različne geografske podatke in med njimi poišče medsebojne odnose oz. povezave (Kidd in Ritchie, 2006).

2.3 FILOGENETSKA DREVESA

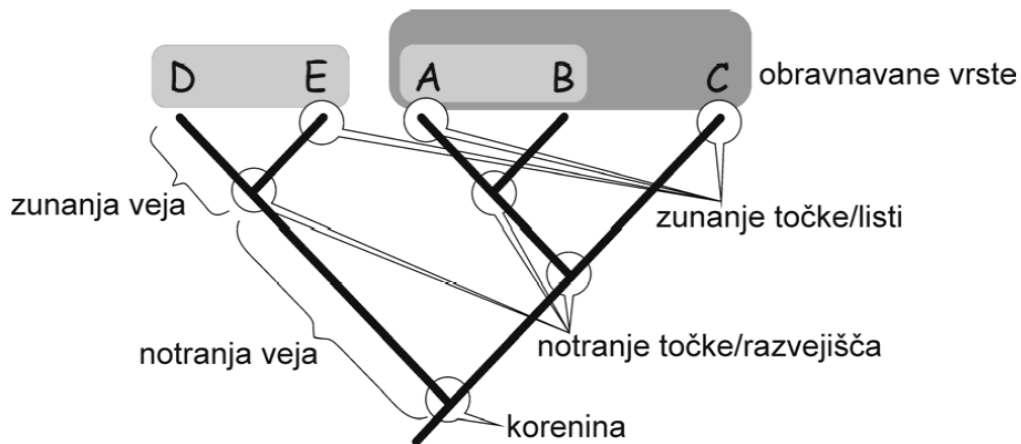
Filogenetsko drevo prikazuje sorodstvene odnose med organizmi v obliki dvodimenzionalnega grafa. Sorodstveni odnosi se danes ugotavljajo predvsem z molekularnimi metodami, včasih pa so sorodnost ugotavljali na podlagi morfoloških lastnosti, kar pa je manj zanesljivo. Sorodnost lahko ugotavljamo med organizmi različnih vrst (ali drugih višjih taksonov) ali pa med različnimi osebki znotraj ene vrste. Molekularne metode določijo podobnost na podlagi nukleotidnih zaporedij. Filogenetska drevesa tako danes prikazujejo večinoma odnose med nukleotidnimi zaporedji, kar pa ni nujno, da odraža dejanske sorodstvene odnose med organizmi.

Določeni organizem (oz. njegovo nukleotidno zaporedje) je na drevesu predstavljen kot sistematska enota ali takson. Klasična hierarhija taksonov se začne z najmanjšim taksonom vrsta (*species*), nadaljuje s taksonom rod (*genus*), družina (*familia*), red (*ordo*), razred (*classis*) in konča z največjim taksonom deblo (*phylum*). Vmesne kategorije se tvorijo tako, da se klasičnim taksonom doda predpona (*supra-*, *infra-*, *sub-* ...).

Sorodstveni odnosi med taksoni so v drevesu hierarhični. Veje se vedno cepijo, nikoli ponovno združujejo. Iz ene predniške vrste načeloma dobimo le dve hčerinski vrsti, redko več. Temu pravimo dihotomnost, drevo pa je binarno (Trontelj, 2009).

Deli filogenetskega drevesa (Trontelj, 2009) (Slika 4):

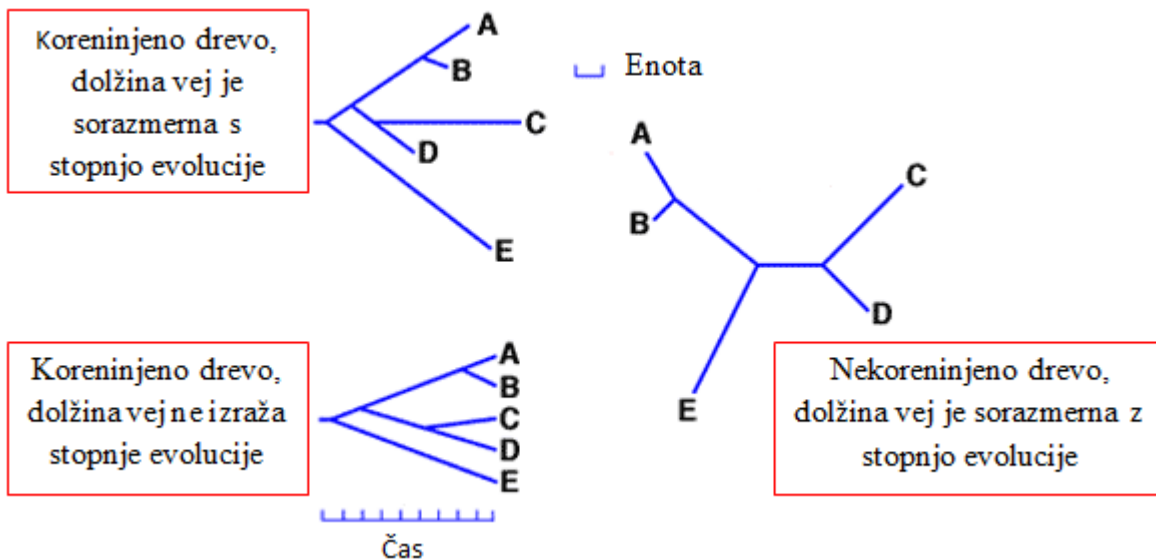
- korenina (*root*),
- veja (notranja, zunanja) (*outer branches = leaves, inner branches*),
- razvejišča (*nodes*),
- listi (na katerih so obravnavane vrste) (*leaves*).



Slika 4: Grafični prikaz filogenetskega drevesa in njegovih delov. Obravnavani vrsti D in E z njunim razvejiščem v tem primeru tvorita monofiletsko skupino. B in C (brez A) tvorita parafiletsko skupino. E in A tvorita polifiletsko skupino (Trontelj, 2009).

Korenina predstavlja skupnega prednika vseh taksonov, ki so na danem drevesu. Vsako razvejišče predstavlja skupnega prednika oz. neko predniško stanje, iz katerega sta se razvili novi, med seboj različni vrsti, ki sta reproduktivno izolirani (velja, če so taksoni vrste). Vse nadaljnje evolucijske spremembe so neodvisne od prejšnjih (Trontelj, 2009). Dolžina vej predstavlja število sprememb v nukleotidnih zaporedjih, ki so se pojavila v času od prednika do njegovega (neposrednega) potomca. Koliko časa je trajala takšna evolucija, navadno ni znano, lahko pa se oceni s filogenetsko analizo. Zunanja vozlišča ali listi pa predstavljajo proučevane taksone.

Za razliko od koreninjenega drevesa poznamo še nekoreninjeno drevo, vendar nam slednje ponuja manj podatkov (Slika 5). Iz nekoreninjenega drevesa se ne da razbrati skupnih prednikov skupin in ne vemo, kako si proučevane skupine evolucijsko sledijo. Takšno drevo nam poda informacije samo o odnosih med skupinami. Drevo lahko ukoreninimo z nukleotidnim zaporedjem fosilne vrste, če je poznan, ali pa z bližnjim sorodnikom, ki mora biti soroden vsem proučevanim skupinam, hkrati pa dovolj oddaljen, da se ne uvršča med njih.



Slika 5: Nekaj načinov prikaza drevesa s petimi taksoni. **Zgoraj levo** je koreninjeno drevo, dolžina vej pa prikazuje, koliko so stopnje evolucije oddaljene med seboj. Korenina je skrajno levo. Vidimo, da sta bila taksona E in C deležna veliko sprememb, odkar sta divergirala iz zadnjega skupnega prednika. Taksona B in D pa sta bila deležna le malo sprememb. **Spodaj levo** je koreninjeno drevo, vendar veje nakazujejo samo sorodstvene odnose, njihova dolžina ne da nobene informacije. **Desno** je nekoreninjeno drevo, skupen prednik torej ni znan. (Vir:

<http://www.talkorigins.org/faqs/comdesc/phylo.html>)

Metode za rekonstrukcijo dreves lahko v grobem razdelimo na tiste, ki temeljijo na izračunanih razvojnih razdaljah med danimi organizmi, in tiste, ki temeljijo na lastnostih oz. stanjih znakov danih organizmov. Največkrat se uporabljajo naslednje metode: metoda največje varčnosti (angl. *Maximum Parsimony Method*), distančna metoda (angl. *Distance Method*) in metoda največjega verjetja (angl. *Maximum Likelihood Approach*). Velja še omeniti, da so vse rekonstrukcije dreves v bistvu samo predvidevanja sorodstvenih odnosov med proučevanimi taksoni.

2.4 METODA NAJVEČJE VARČNOSTI (MP)

MP (angl. *Maximum parsimony* ali *Minimum evolution method*) preuči vsa možna drevesa, nato pa rekonstruira drevo, ki vsebuje najmanj mutacij oz. predvideva najmanj sprememb stanj znakov. Tako dobimo najbolj varčno drevo. Pri velikem številu taksonov lahko naletimo na težavo, saj je

takšna metoda zelo potratna (ravno zaradi tega, ker mora preveriti vse topologije dreves). Težavo lahko odpravimo s pomočjo hevrističnih oblik metode.

Pomembno se je zavedati, da lahko MP naredi veliko topoloških napak, kadar imamo opravka z velikim številom taksonov.

Metoda najprej določi informativne pozicije. Kot najbolj verjetno drevo določi tisto, do katerega pride z najmanjšim številom sprememb med temi pozicijami (najbolj varčno drevo) (Mount, 2004).

2.5 DISTANČNA METODA (DM)

Temelji na poznanih ali izračunanih razvojnih razdaljah med preučevanimi organizmi oz. pari nukleotidnih zaporedij. Organizma sta si sorodnejša, tem manjša je razdalja med njima.

DM uporablja algoritme, ki s pomočjo podatkov rekonstruirajo filogenetsko drevo. Algoritmi nukleotidna zaporedja oz. morfološke znake pretvorijo v matriko, ki predstavlja razdalje. Slednje pomenijo število razlik med preučevanimi taksoni. To je število različnih morfoloških znakov med določenima taksonoma oz. število različnih nukleotidov.

Cilj DM je sosednje organizme postaviti na prava mesta v drevesu in jih povezati z vejami prave dolžine (Mount, 2004).

DM uporabljajo metode (Mount, 2004):

- ❖ metoda Fitch in Morgoliash (angl. *Fitch and Margoliash Method*),
Nukleotidna zaporedja se najprej razdelijo v skupine po tri in nato iz združenih podatkov izračunamo dolžino vej drevesa. Nukleotidna zaporedja, ki sledijo, se nato k analizi dodajajo postopoma. Takšna analiza je najbolj primerna za drevesa s kratkimi vejami, saj pri daljših vejah rezultati niso zanesljivi. Cilj je predvideti drevo, ki najbolj ustreza tabeli razdalj.
- ❖ metoda združevanja sosedov (angl. *Neighbor-joining*),
Ta metoda je podobna prejšnji, le da uporablja drug algoritem. Uporabna je predvsem, kadar se stopnja evlucijskih sprememb posameznih vej pod pogoji spreminja. Kadar dolžine vej variirajo, so najbolj zanesljive metode *Neighbor-joining*, *Sattath* in *Taversky method*. Metoda izbere tista nukleotidna zaporedja, ki skupaj dajo najboljši približek oz.

oceno dolžin vej (in se s tem najbolj približa dejanski dolžini med nukleotidnimi zaporedji).

- ❖ metoda netehtane aritmetične sredine (angl. *Unweighted Pair Group Method with Arithmetic Mean*),

UPGMA privzame, da so stopnje sprememb (ki jih predstavljajo dolžine vej), konstantne. Poišče najbolj sorodna nukleotidna zaporedja in izračuna dolžino veje, ki ju povezuje. Nato izračuna povprečje med to razdaljo in razdaljo naslednjega para ali skupine. To se ponavlja, dokler niso v drevo vključena vsa nukleotidna zaporedja, na koncu pa predvidi še lego korenine.

- ❖ itd.

2.6 METODA NAJVEČJEGA VERJETJA (ML)

ML podobno kot MP primerja vsa možna drevesa, ki lahko nastanejo iz danih podatkov in nato izbere najbolj verjetno drevo oz. drevo z največjim verjetjem. Odloča se na podlagi največjega verjetja glede na izbrani model evolucije (npr. model stopenj evolucijskih sprememb v proteinskem ali nukleotidnem zaporedju). Algoritem v nukleotidnem zaporedju primerja stanje znakov na določeni poziciji.

ML najprej za dano število nukleotidnih zaporedij izriše drevo, nato naključno premesti veje in dobljeno drevo primerja s prejšnjim. Izmed dreves izbere tisto, ki bolj ustreza danim podatkom. Ta algoritem se ponavlja, dokler nastajajo »boljša« drevesa oz. dokler ne nastane najbolj verjetno drevo (Mount, 2004).

2.7 MODELI MOLEKULARNE EVOLUCIJE

Že leta 1965 sta Zuckerkandl in Pauling predlagala model molekularne ure (angl. *molecular clock*), ki čas divergence med geni in proteini meri na podlagi števila sprememb med zaporedji. Leta 1969 sta Jukes in Cantor predlagala stohastičen model za zamenjavo nukleotidov (substitucijo). Pri tem so vse zamenjave enako vredne, ne glede na to, za kateri nukleotid gre. Modela sta preprosta in upoštevata en sam faktor (Lio in Goldman, 1998).

Težko je ustvariti splošen model, ki bi predvidel vse vrste mutacij, vzorcev itd. Modeli zato pogosto temeljijo na preprostih zakonih ali izkustvenih opazovanjih, ki ne razumejo biologije v ozadju (Lio in Goldman, 1998). Nekaj modelov je naštetih v nadaljevanju.

2.7.1 Markov model

Model temelji na stohastičnem modelu evolucije nukleotidnih ali aminokislinskih zaporedij. Privzame, da so različni predeli zaporedij v evoluciji neodvisni med seboj. Vsak predel pogleda posebej in ga označi z verjetnostjo $P_{ij}(\mathbf{T})$ (baza i bo zamenjana z bazo j , pri času \mathbf{T}). Podpisana i in j lahko zavzameta vrednosti med 1 in 4, kadar imamo nukleotidno zaporedje. Vrednosti predstavljajo nukleotide A, T, C in G. Kadar imamo aminokislinsko zaporedje, lahko i in j zavzameta vrednosti od 1 do 20. Vrednosti predstavljajo aminokisliline (Lio in Goldman, 1998).

Verjetnost prehajanja baz iz ene v drugo lahko zapišemo tudi kot matriko $P(\mathbf{T})$.

$$P(\mathbf{T} + d\mathbf{T}) = P(\mathbf{T})(\mathbf{I} + \mathbf{Q}d\mathbf{T})$$

\mathbf{T} predstavlja čas, $d\mathbf{T}$ predstavlja spremembo v času, \mathbf{I} je identična matrika (identiteta) in \mathbf{Q} je matrika trenutnih vrednosti (nediagonalne vrednosti so vrednosti zamenjave i z j).

Model ima tri pomembne lastnosti: homogenost (vrednosti matrike so neodvisne od časa, kar pomeni, da vzorci zamenjav nukleotidov ali aminokislilin ostajajo enaki na različnih delih drevesa), nespremenljivost (proces je v ravnovesju, kar pomeni, da ostajajo frekvence nukleotidov približno ista tekom evolucije) in reverzibilnost (velja $\pi_i P_{ij}(\mathbf{T}) = \pi_j P_{ji}(\mathbf{T})$ za vse i, j in \mathbf{T} , π_i pa predstavlja frekvenco baze) (Lio in Goldman, 1998).

2.7.2 Model zamenjave nukleotidov

Model (Jukes in Cantor) je definiran z $Q_{ij} = \alpha$, za vse i in j (zavzameta lahko vrednosti od 1 do 4, pri čemer sta različna med seboj). To pomeni, da ostaja vrednost α enaka za vse baze, ki so zamenjane (z drugo bazo). Posledica tega je, da je za vse frekvence baz privzeta vrednost 0,25 (Lio in Goldman, 1998).

To uporablja tudi model *Kimura 2-parameter*, *Blaisdell*, *Felsenstein* idr. Modeli so parametrični (definirani so s parametri) in delujejo na nivoju posameznih nukleotidov. Model mutacij tripletov,

ki se dobro približa realnosti, sta predlagala Goldman in Yang. Upoštevala sta znanje o genskem kodu in posledice zamenjave nukleotidov v zaporedjih (Lio in Goldman, 1998).

2.7.3 Model zamenjave aminokislin

Modeli temeljijo večinoma na izkustvenem pristopu. Model evolucije proteinov so razvili Dayhoff in njegovi sodelavci. Vrednost zamenjave aminokislinske se dobi iz poravnave aminokislinskih zaporedij, ki so vsaj 85 % identična. Model pomaga pri iskanju zaporedij v bazah podatkov. Predvsem so znane matrike P(0,5), P(1) in P(2,5) (rečemo jim tudi matrike PAM50, PAM100 in PAM250). Te matrike pomagajo pri ugotavljanju ujemanja obravnavanega zaporedja z zaporedji iz baze podatkov. Matrika impliciranih vrednosti se uporablja v filogeniji (Lio in Goldman, 1998).

Podoben pristop sta uporabila tudi Jones in Gannet, pri svojih modelih. Adachi in Hasegawa sta uporabila obrnljiv Markov model. Preprost, neempirični model je predlagal Nei. Drugačen pristop sta si izbrala Henikoff in Henikoff pri svojem modelu. Matrike tega modela (BLOSUM50) delujejo boljše kot PAM (Lio in Goldman, 1998).

2.7.4 Stopnja heterogenosti

Vključitev heterogenosti v izračune, je omogočila boljše rezultate in s tem boljšo filogenetsko rekonstrukcijo. Uspešni so tisti modeli, ki temeljijo na sovisni razporeditvi vrednosti. Nei in Gojobori ter Yang in Tamura so predlagali modele na podlagi gama razporeditve (Lio in Goldman, 1998).

Uspeh gama razporeditve si lahko razlagamo v njeni prilagodljivosti. Vrednost zamenjave določenega predela je v modelu predstavljena s parametrom α . Če je parameter manjši od 1, pomeni, da je na tistem predelu velika stopnja variacije. Če je parameter večji od 1, pa je nizka stopnja variacije (Lio in Goldman, 1998).

2.7.5 Združeni modeli

Različna zaporedja, kot so aminokislinska z kodirajočimi in nekodirajočimi predeli ali različni predeli istega gena, lahko kažejo heterogenost v procesu evolucije. Parametri, ki kažejo na to heterogenost so frekvence nukleotidov, vrednosti tranzicij/transverzij idr. (Lio in Goldman, 1998).

3 POSTOPEK RISANJA FILOGENETSKIH DREVES

Potem ko iz sekvenatorja dobimo zaporedja, sledi obdelava. Zaporedja lahko dobimo tudi iz podatkovnih baz. Najstarejše podatkovne baze o nukleotidnih zaporedjih so zbrane na strani *International Nucleotide Sequence Database* (INSD), spletni naslov <http://www.insdc.org/>. S te strani lahko dostopamo do:

- *GenBank*, spletni naslov <http://www.ncbi.nlm.nih.gov/genbank/>,
- *European Nucleotide Archive* (ENA), spletni naslov <http://www.ebi.ac.uk/ena/>,
- *DNA Data Bank of Japan* (DDBJ), spletni naslov <http://www.ddbj.nig.ac.jp/>.

Podatkovne baze vsebujejo nukleotidna zaporedja organizmov. Zaporedja se dnevno dodajajo v baze. Najbolj priljubljena je baza podatkov GenBank.

Nekaj baz aminokislinskih zaporedij:

- UniProt (<http://www.uniprot.org/>),
- Pfam (<http://pfam.sanger.ac.uk/>),
- SUPERFAMILY (<http://supfam.cs.bris.ac.uk/SUPERFAMILY/>).

Na spletu obstaja še veliko podatkovnih baz, tudi za zaporedja genomov, struktur proteinov, nukleotidnih zaporedij RNK ipd. Nekaj od njih je zbranih v seznam na spletni strani: http://en.wikipedia.org/wiki/List_of_biological_databases#Genome_databases.

3.1 PORAVNAVANJE ZAPOREDIJ

Za poravnavo nukleotidnih zaporedij uporabljamo različna orodja, ki delujejo na podlagi številnih algoritmov. Nekaj jih je opisanih v nadaljevanju.

3.1.1 FASTA

Orodje je dostopno na spletni strani:

http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=select&pgm=fad.

FASTA je program za hitro poravnavanje parov proteinov in nukleotidnih zaporedij (Slika 6). Priporočena je uporaba zadnje verzije FASTA3, saj uporablja izboljšane metode poravnavanja in računanja statističnih vrednosti.

Program išče ujemajoče se vzorce ali besede, ki jim rečemo ang. *k-tuples*. Na podlagi ujemajočih vzorcev oz. besed, program naredi lokalno poravnavo. Zaradi hitrega algoritma kmalu dobimo rezultate. Kljub temu je BLAST še hitrejši (Mount, 2004).

FASTA dano nukleotidno zaporedje (aminokislinsko zaporedje) primerja z vsemi nukleotidnimi zaporedji (aminokislinski zaporedji) iz baze podatkov tarčnega nukleotidnega zaporedja (aminokislinskega zaporedja). Ugotovi in izpiše nukleotidno zaporedje, ki se najbolj ujema. Izpiše še lokalne poravnave teh nukleotidnih zaporedij z danim nukleotidnim zaporedjem. Tako z BLAST kot s FASTA programom lahko ugotovimo reprezentativne in evolucijske odnose med nukleotidnimi zaporedji. Lahko pa tudi ugotovimo identiteto članov posamezne družine genov (Mount, 2004).

Nukleotidno zaporedje moramo podati v FASTA formatu (Slika 7). Dobimo ga iz FASTA knjižnice nukleotidnih zaporedij oz. baze podatkov. Oblika obsega eno vrstico komentarjev, ki se začne s simbolom » >«, temu pa sledi določeno število vrstic, ki vsebujejo informacije o nukleotidnih zaporedjih. Razen zadnje vrstice so vse druge omejene na 80 znakov. Zaporedje je zapisano v standardni kodi za aminokislinske ali nukleotide (NCBI, 2007).

FASTA – primerja dano aminokislinsko zaporedje z aminokislinskimi zaporedji iz knjižnice aminokislinskih zaporedij. Išče podobno zaporedje. Prav tako nukleotidno zaporedje primerja z nukleotidnimi zaporedji iz knjižnice nukleotidnih zaporedij in med njimi išče podobno (Mount, 2004).

TFASTA – primerja dano aminokislinsko zaporedje z aminokislinskimi zaporedji iz knjižnice nukleotidnih zaporedij. Pred tem pa prevede knjižnico nukleotidnih zaporedij v aminokislinska zaporedja (Mount, 2004).

FASTF/TFASTF in FASTS/TFASTS: prvi primerja skupek kratkih peptidnih fragmentov z nukleotidnim zaporedjem iz baze podatkov aminokislinskih zaporedij, drugi pa skupek primerja s

prevedeno bazo podatkov nukleotidnih zaporedij (Mount, 2004).

Programi so zasnovani tako, da omogočajo poravnavo nukleotidnih zaporedij, prevedene v vse tribralne okvirje, ki se berejo naprej, in tiste tri, ki se berejo nazaj.

```
>>>gi|126256219|gb|EF437582.1|, 219 nt vs TMP.q2 library

>>gb|EF437590.1| Panthera pardus cytochrome b (Cytb) gene, part (219 nt)
  initn: 853 init1: 853 opt: 897 Z-score: 943.6 bits: 181.7 E(10000): 9.4e-51
banded Smith-Waterman score: 897; 90.0% identity (90.0% similar) in 219 nt overlap (1-219:1-219)
Entrez Lookup Re-search database General re-search

      10      20      30      40      50      60      70      80
gi|126 ATCAACCACTCATTGATCTTCCCGCTCCATCCAATATCTCAGCATGATGAACTTTGGTCCCTACTAGGAGTGTG
      :: :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
gb|EF4 ATTAATCACTCATTGATCTTCCCGCTCCATCCAACATCTCAACATGATGAACTTTGGCTCCCTATTAGGAGTATG
      10      20      30      40      50      60      70      80

      90     100     110     120     130     140     150     160
gi|126 TTTAATCCTACAAATTCTTACTGGTCTCTTCCCTAGCCATACACTACACGTCAGATAACAATAACCGCTTTCTCATCAGTCA
      :: :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
gb|EF4 TTTAATCCTACAAATTCTCACCGGCCTCTTTCTGGCCATACATTACACATCAGACACAACAACCGCTTTCTCATCAGTTG
      90     100     110     120     130     140     150     160

      170     180     190     200     210
gi|126 CCCATATCTGCCGCGACGTAAATTACGGCTGAATCATCCGATATTTACACGCCAATGGA
      :: :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
gb|EF4 CCCATATCTGCCGCGACGTAAATTACGGCTGAATTATCCGGTATTTACATGCCAACGGA
      170     180     190     200     210
```

Slika 6: Primerjava nukleotidnih zaporedij *Panthera onca* in *Panthera pardus* v FASTA formatu. Pove, v koliko nukleotidih se zaporedji ujemata (v našem primeru 90 %).

Uncia uncia cytochrome b (Cytb) gene, partial cds; mitochondrial

```
>gi|126256221|gb|EF437584.1| Uncia uncia cytochrome b (Cytb) gene, partial cds;
mitochondrial
ATCAATCACTCATTGATCTTCCCACTCCATCCAACATCTCCGCATGATGAACTTTGGCTCCCTGT
TAGGAGTATGTTAATCCTACAAATTCTCACCGGCCTCTTCTAGCCATACACTATACATCAGACACAAT
AACCGCTTTCTCGTCAGTCACCCACATCTGCCGCGACGTAAATTATGGCTGAATTATCCGATACCTACAC
GCCAACGGA
```

Slika 7: Primer FASTA nukleotidnih zaporedij vrst *Uncia uncia*. V prvi vrsti so informacije o nukleotidnem zaporedju (vrstno ime in kateri del gena nukleotidno zaporedje predstavlja).

3.1.2 BLAST (angl. *Basic Local Alignment Search Tool*)

BLAST je orodje, s katerim lahko primerjamo aminokislinska zaporedja (proteine) ali nukleotidna zaporedja (DNK). Je spletna aplikacija, na voljo vsem in zastonj. Obsega veliko genomsko knjižnico (človeka, miši, podgane in ostale organizme). Dostopen je na spletnem naslovu <http://blast.ncbi.nlm.nih.gov/Blast.cgi> (Slika 8).

Z BLAST-om najdemo regije, kjer so si nukleotidna zaporedja podobna. Program primerja nukleotidno ali aminokislinsko zaporedje z zaporedji iz genomskih knjižnic in statistično izračuna obseg ujemanja med zaporedji (Slika 9). Program nam je tako v pomoč pri ugotavljanju v katero družino genov spada naše nukleotidno zaporedje in tudi pri ugotavljanju sorodstvenih odnosov med vrstami (Mount, 2004).

Glede na to, kakšne rezultate želimo dobiti, lahko izbiramo med veliko različnimi algoritmi. V BLASTX podamo nukleotidno zaporedje, ki se nam nato pretvori v proteinsko zaporedje. To pa lahko naprej primerjamo z drugimi proteinskimi zaporedji iz knjižnic (BLASTP, PSI BLAST, PHI-BLAST). BLASTN, MEGABLAST vpisano nukleotidno zaporedje primerjata s knjižnico nukleotidnih zaporedij. Algoritmov je še veliko (Jakovac, 2009).

Za razliko od FASTA, ki v zaporedju išče »besede« iste dolžine, pa BLAST išče pomembne »besede«. Pomembnost besed izračuna s pomočjo algoritma. Priporoča se uporaba BLAST2, saj je hitrejši in rezultati so bolj pregledni (Mount, 2004).

Slika 8: Orodje BLAST, prosto dostopno na spletu

```

Panthera pardus cytochrome b (Cytb) gene, partial cds; mitochondrial
Length=219

Score = 294 bits (159), Expect = 8e-85
Identities = 199/219 (91%), Gaps = 0/219 (0%)
Strand=Plus/Plus

Query 1  ATTAACCACTCATT CATTGATCTT CCCACTCCACCCAATATCTCAGCATGATGAACTTT 60
      |||||  |||||||  |||||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Sbjct 1  ATTAATCACTCATT CATTGATCTT CCCGCTCCATCCAACATCTCAACATGATGGAACCTTT 60

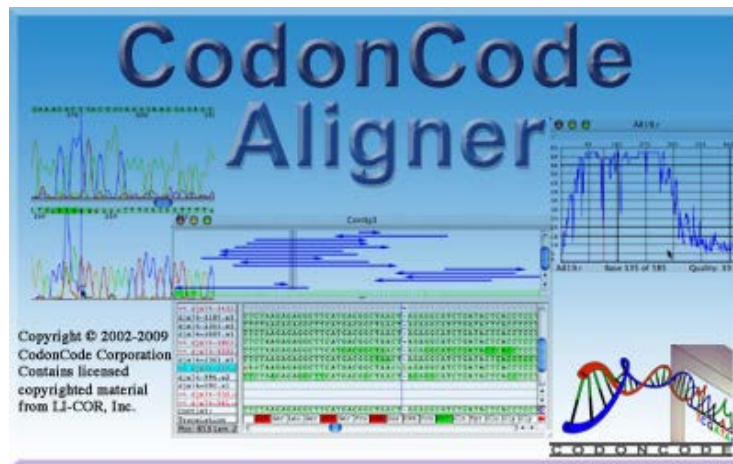
Query 61  GGCTCCTTATTAGGAGTATGTTTAATCCTACAAAATCTCACCGGCCTCTTTCTAGCCATA 120
      |||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Sbjct 61  GGCTCCCTATTAGGAGTATGTTTAATCCTACAAAATCTCACCGGCCTCTTTCTGGCCATA 120

Query 121  CATTACACACCAGACACAATAACC GCTTT CTCAT CAGTCACCCACATTTGCCGCGATGTA 180
      |||||  |||||||  |||||||  |||||||  |||||||  |||||  |||||  |||||  |||||  |||||
Sbjct 121  CATTACACATCAGACACAACAACC GCTTT CTCAT CAGTTGCCCATATCTGCCGCGACGTA 180

Query 181  AACTATGGCTGAAT TATCCGGTACCTACACGCCAACGGA 219
      |||  |||||||  |||||  |||||  |||||  |||||
Sbjct 181  AATTACGGCTGAAT TATCCGGTATTTACATGCCAACGGA 219
    
```

Slika 9: Z BLAST-om primerjamo nukleotidno zaporedje *Panthera leo* z nukleotidnim zaporedjem *Panthera pardus*.

3.2 ANALIZA ZAPOREDIJ: CODON CODE ALIGNER



Slika 10: Pozdravno okno programa CodonCode Aligner 4.0.3

To je programski paket za obravnavo neobdelanih nukleotidnih zaporedij. Program je plačljiv. S spleta ga lahko snamemo na naslovu <http://www.codoncode.com/aligner/>.

Deluje na operacijskih sistemih Mac OS X (10.3.4 ali novejši) in Windows (98, ME, 2000, XP ali novejši). Koliko prostora na pomnilniku zavzame, je odvisno od velikosti projekta. Potrebno je vedeti, da je Aligner javanski program, zato moramo imeti na računalniku naloženo primerno verzijo Java. Na Windows-ih bo namestitveni program samodejno naložil trenutno verzijo Java (CodonCode Corporation, 2012).

Ko naložimo Aligner in ga zaženemo, se ta zažene v »demo« verziji. Kar pomeni, da bomo program lahko uporabljali 30 dni, če si ne priskrbimo licence. V demo verziji ne moremo tiskati, shranjevati ali izvažati projektov. Da lahko začnemo uporabljati vse te funkcije, je potrebno pridobiti licenco.

Aligner lahko združuje, poravnava in ureja DNK in RNK nukleotidna zaporedja. Vse, kar delamo, nam shrani v projekte (ang. *projects*). V teh projektih so zbrana urejena in združena nukleotidna zaporedja naših vzorcev. Aligner najboljše dela z datotekami s končnicama SCF (.scf) in ABI (.abi). SCF dobimo s pomočjo programa PHRED, ABI pa z *KB base caller*. Zaporedja v datoteki ABI so neobdelana zaporedja, ki jih dobimo iz sekvenatorja. Uvozimo lahko tudi nukleotidna

zaporedja v tekstovni obliki.

Aligner štiti naša originalna nukleotidna zaporedja tako, da pri uvažanju naredi njihovo kopijo.

Tako pri urejanju in poravnavanju spreminjamo le kopije datotek.

3.2.1 Koda IUPAC, ki ga uporablja CodonCode Aligner

IUPAC (International Union of Pure and Applied Chemistry) je mednarodna organizacija, ki se ukvarja s kemijo, z biologijo in s fiziko. Dostopna je na spletni strani <http://www.iupac.org/>. Nastala je leta 1919, med drugim tudi zaradi potrebe po standardizaciji kod zaporedij (nukleotidnih ali aminokislinskih). Na ta način je zapis zaporedij krajši in zato bolj pregleden. Kode aminokislin so lahko sestavljene iz ene črke ali treh (npr. aminokislino cistein lahko označimo z C ali Cys) (IUPAC, 2012).

Genska koda nam pove, katere aminokislino tvorijo tripleti nukleotidov in je dostopna na <http://users.ox.ac.uk/~linc1775/blueprint.htm>.

Preglednica 1: Koda IUPAC za nukleotidna zaporedja (DNK/RNK)

DNK/RNK	
Koda nukleotida	Baza
A	adenin
C	citozin
G	gvanin
T (ali U)	timin (ali uracil)
R	A ali G
Y	C ali T/U
S	G ali C
W	A ali T/U

K	G ali T/U
M	A ali C
B	C ali G ali T/U
D	A ali G ali T/U
H	A ali C ali T/U
V	A ali C ali G
N	A ali C ali G ali T/U
. ali -	vrzel

(Vir: <http://www.bioinformatics.org/sms2/iupac.html>)

Preglednica 2: Koda IUPAC za aminokislinska zaporedja (protein)

PROTEIN		
Koda aminokislina	Trimestna črkovna koda	Aminokislina
A	Ala	alanin
B	Asx	asparaginska kislina ali asparagin
C	Cys	cistezin
D	Asp	asparaginska kislina
E	Glu	glutaminska kislina
F	Phe	fenilalanin
G	Gly	glicin
H	His	histidin
I	Ile	izolevcin
K	Lys	lizin
L	Leu	levcin
M	Met	metionin

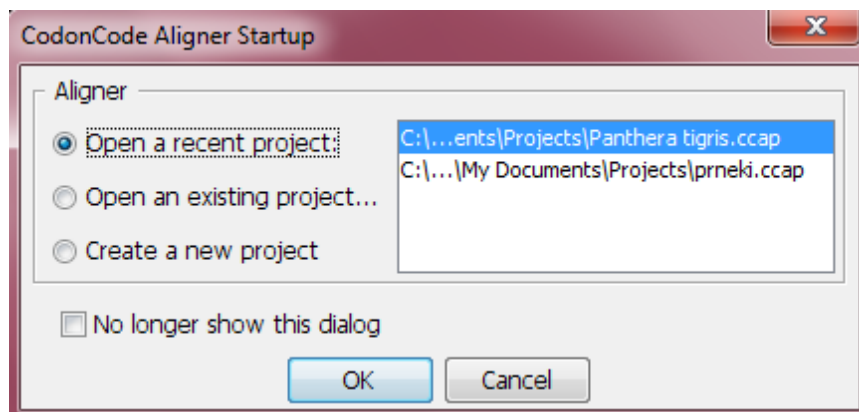
N	Asn	asparagin
P	Pro	prolin
Q	Gln	glutamin
R	Arg	arginin
S	Ser	serin
T	Thr	treonin
V	Val	valin
W	Trp	triptofan
X	Xaa	katerakoli aminokislina
Y	Tyr	tirozin
Z	Glx	glutamin ali glutaminska kislina

(Vir: <http://www.bioinformatics.org/sms2/iupac.html>)

3.2.2 Potek dela v programu

3.2.2.1 Glavno okno

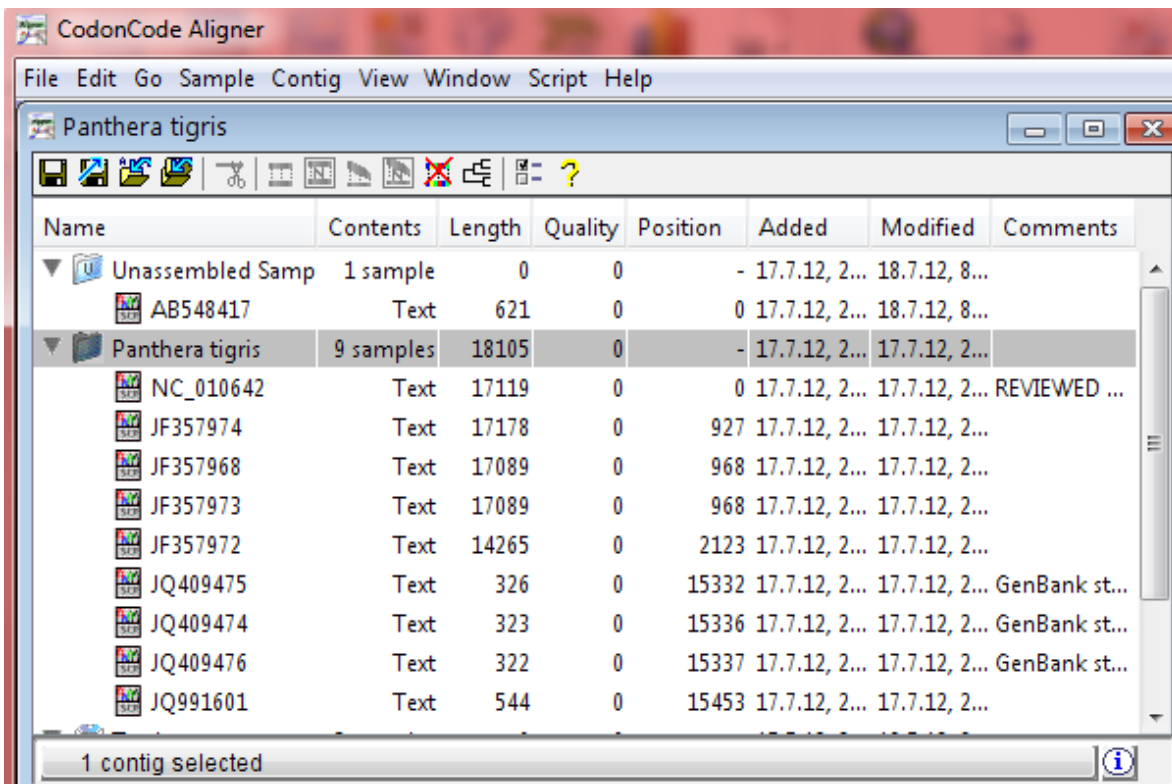
Najprej se nam odpre glavno okno (Slika 11) in nam ponudi izbiro med projektom, na katerem (če) smo pred kratkim delali (ang. *open a recent project*), med projektom, ki že obstaja (ang. *open an existing project*), ali pa začne z novim projektom (ang. *create a new project*). Izberemo želeno in potrdimo. Lahko tudi označimo »*No longer show this dialog*«, če ne želimo več prejemati tega okna ob zagonu programa.



Slika 11: Program ob zagonu ponudi, da izberemo, kateri projekt bomo odprli.

3.2.2.2 Projektno okno – delo z nukleotidnimi zaporedji

Karkoli prej izberemo, se nam bo po potrditvi odprlo projektno okno (Slika 12). Če smo izbrali že obstoječi projekt, se nam odpre vse, kar smo že naredili na njem, če pa smo izbrali nov projekt, bo okno prazno. Naslednji korak je, da uvozimo datoteke z želenimi nukleotidnimi zaporedji (File, Import, Add samples, nato poiščemo, kje imamo datoteke shranjene). Datoteke in mape lahko kar potegnemo v projektno okno tako, da datoteko primemo in med prenašanjem držimo tipko miške. Nukleotidna zaporedja lahko dobimo tudi s spleta (Genbank). Slednje storimo tako, da izberemo File, Import, From Genbank. Paziti moramo, da so izbrane datoteke v formatih ABI, AB1 ali SCF. Program pa prepozna tudi različne vrste besedilnih formatov. Na novo uvoženi vzorci bodo v mapi »*Unassembled Samples*«.



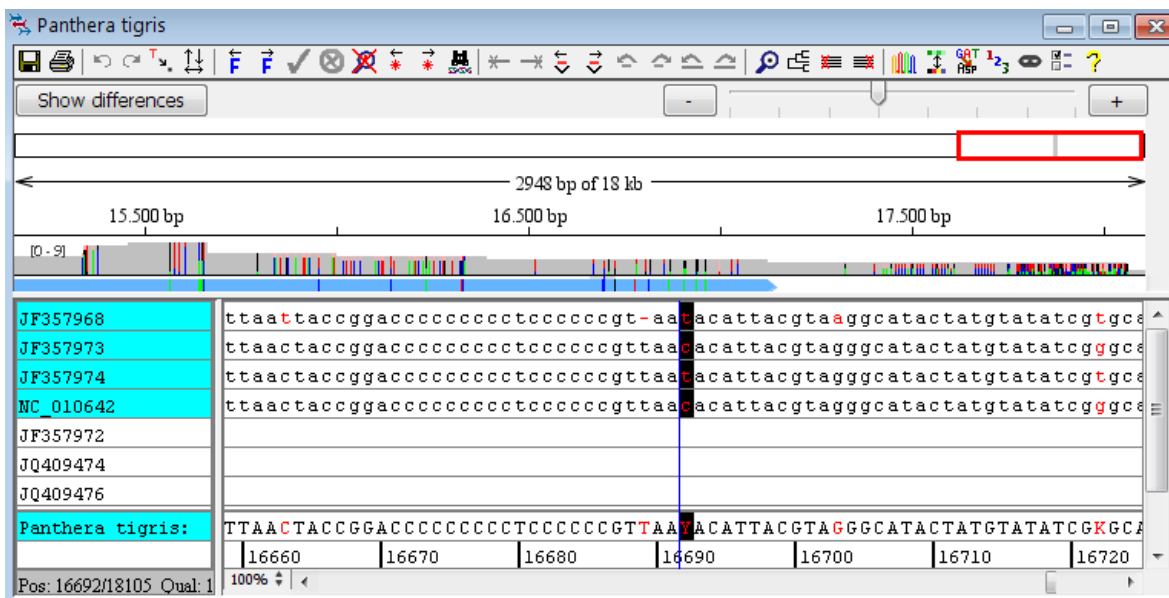
Slika 12: Projektno okno, v katerega uvažamo datoteke nukleotidnih zaporedij. Tu jih lahko združimo ali jih zavrzemo. Lahko jih preimenujemo itd.

Začetek in konec zaporedij sta navadno slabše kakovosti, zato ju je dobro odrezati, preden zaporedja združimo. Potem, ko to storimo, jih združimo (ang. *assemble*). To storimo tako, da označimo vsa nukleotidna zaporedja, ki jih želimo združiti (držimo tipko »ctrl« in z miško označimo zaporedja, ki jih želimo združiti), nato pritisnemo gumb »assemble« zgoraj ali pa kar »ctrl + Alt A«. Dobimo mapo »Contig«, ki jo lahko po želji preimenujemo. To storimo preprosto tako, da pritisnemo desno tipko miške, preimenujemo in potrdimo z vnašalko (angl. *enter*). Vsak skupek združenih nukleotidnih zaporedij se nam pojavi v svoji mapi.

Tista nukleotidna zaporedja, ki so res slabe kakovosti, lahko že takoj damo v koš (ang. *Trash*), saj se navadno ne morejo urejati. Kakovost zaporedij se vidi v četrtem stolpcu glavnega okna (ang. *Quality*). Ocenjena je s številko. Pod 500 je kakovost navadno slaba. Če si kasneje premislimo, jih lahko vzamemo iz koša.

Vedno je dobro imeti referenčno zaporedje, ki je v pomoč pri preverjanju. Še posebej je v pomoč, če so zaporedja slabe kakovosti. Najdemo jih na spletu v podatkovni bazi GenBank. Pogledamo, če že obstaja kakšno zaporedje, iste ali vsaj sorodne vrste preučevanega zaporedja, in ga dodamo k našim nukleotidnim zaporedjem za kontrolo.

Ko imamo vse to urejeno, lahko dvakrat pritisnemo »Contig« in začnemo urejati nukleotidna zaporedja. Odpre se nam okno, v katerem so izpisana vsa nukleotidna zaporedja, ki smo jih združili (Slika 13). Če pritisnemo na posamezno, se nam odpre okno s sledovi (angl. *trace*), kjer rdeče črte ponazarjajo timin, zelene adenin, modre citozin in črne gvanin. Barve lahko tudi spreminjamo (desni pritisk miške na nukleotidno zaporedje, Preferences, Base colors). V istem oknu lahko nastavimo še veliko drugih lastnosti (npr. kriterije algoritma za poravnavanje, kriterije algoritma za združevanje in druge tehnične stvari, kot so, kam naj program shranjuje naše delo, kje naj išče nukleotidna zaporedja, lahko si uredimo orodno vrstico po svojih potrebah in željah itd.).



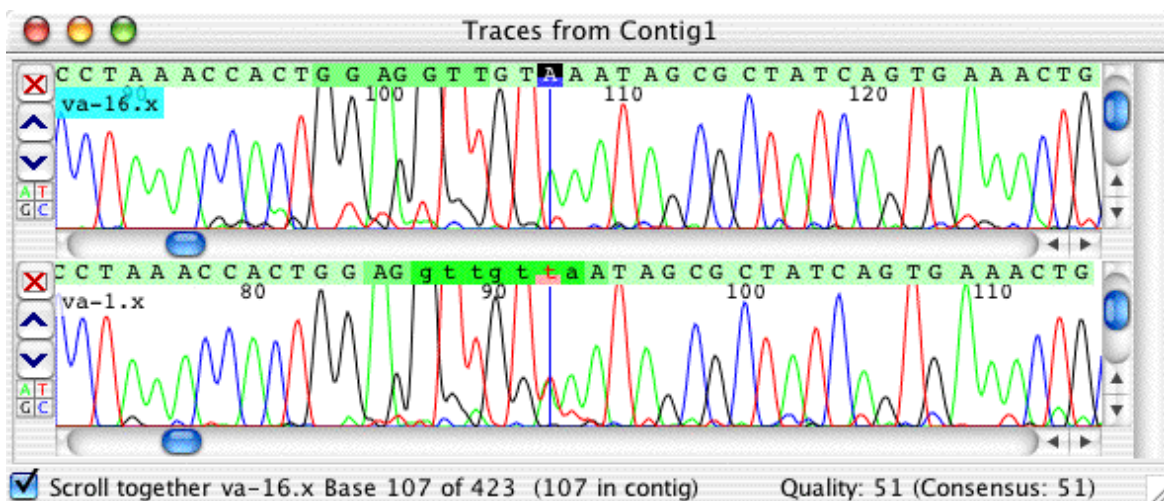
Slika 13: Dvojni pritisk na *Contig* z levo tipko miške nam odpre okno, v katerem lahko obdelujemo in popravljamo naše končno nukleotidno zaporedje. V zgornjem delu okna vidimo dolžino celotnega nukleotidnega zaporedja in tudi,

koliko so dolgi posamezni fragmenti. Če pa izberemo gumb *Show differences*, lahko vidimo vse baze, kjer se posamezna nukleotidna zaporedja razlikujejo. V našem primeru je v končnem nukleotidnem zaporedju označena črka Y. To seveda ni baza, je simbol iz kode IUPAC (Preglednica 1). V primeru, da bi delali z aminokislinami, bi uporabili

Preglednico 2.

Program CodonCode Aligner je najbolj uporaben, kadar imamo več zaporedij istega vzorca (od 5' do 3' in obratno). Program nam združi vse delčke zaporedij gena. Naša naloga je, da pregledamo, če se zaporedja ujemajo in popravimo tam, kjer je črka zapisana rdeče oz. se naša pridobljena nukleotidna zaporedja razlikujejo v tistem nukleotidu. Mi moramo ugotoviti, kateri nukleotid je pravi, saj so zaporedja iz istega vzorca, zato morajo biti identična. Končni rezultat (angl. *consensus*) je nukleotidno zaporedje gena.

Če so datoteke v ABI, AB1 ali SCF formatu, si lahko pomagamo z dvojnimi pritiskom na nukleotidno zaporedje z levo tipko miške. Odpre se nam okno s sledovi (Slika 14). Kjer pride do razlike med nukleotidi, pogledamo, kateri ima višji vrh. Ko pregledamo celo zaporedje, ga izvozimo v fasta formatu. Ta format uporabimo za nadaljnjo obdelavo.



Slika 14: Tako izgleda okno s sledovi (angl. *trace*), ki ga vidimo, če imamo datoteke v ABI, AB1 ali SCF formatu. V tem primeru je označeno mesto, kjer je možno, da je prišlo do mutacije. (Vir: http://www.codoncode.com/aligner/quicktour/specify_regions.htm)

3.2.2.3 Izris drevesa

Program omogoča tudi izris drevesa za določen skupek nukleotidnih zaporedij. Izberemo Contig,

Buid tree. Uporablja metodo NJ. Izbiramo lahko med modeloma *Number of Differences* in *p-distance*. Pri lastnosti »Gaps/Missing data« imamo na izbiro *Pairwise Deletion*, *Complete Deletion*, *Consider Interal Gaps* in *Consider All Gaps*. Nazadnje pa imamo še dve možnosti, kako naj bo drevo izrisano, *Topology only* ali *in Label branches*.

3.3 PORAVNAVA ZAPOREDIJ IN IZRIS FILOGENETSKEGA DREVESA: MEGA (angl. *Molecular Evolutionary Genetics Analysis*)



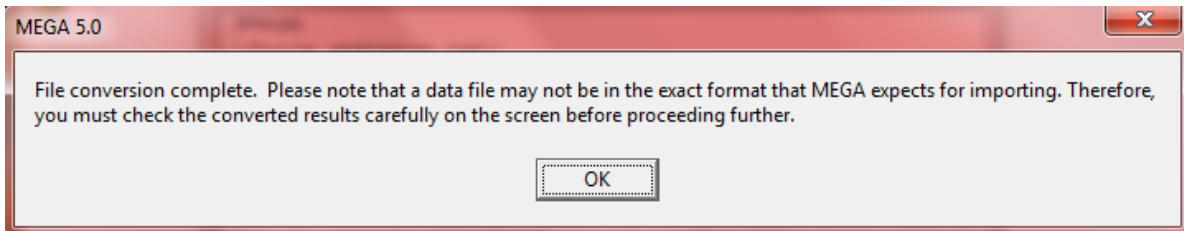
Slika 15: Pozdravno okno programa MEGA 5.05

Programski paket podpirajo Windows-i 95/98, NT, ME, 2000 in XP. Razvijalci priporočajo, da ima računalnik, na katerega želimo naložiti program, vsaj 64 MB RAM-a in 20 MB razpoložljivega prostora na trdem disku (Tamura in sod., 2008). Velika večina današnjih računalnikov to omogoča.

Na spletu je brezplačno dostopen, uporablja se za analize nukleotidnih zaporedij. Z njegovo pomočjo lahko izrišemo filogenetska drevesa, ocenjujemo evlucijske razdalje in računamo (osnovne) statistične podatke.

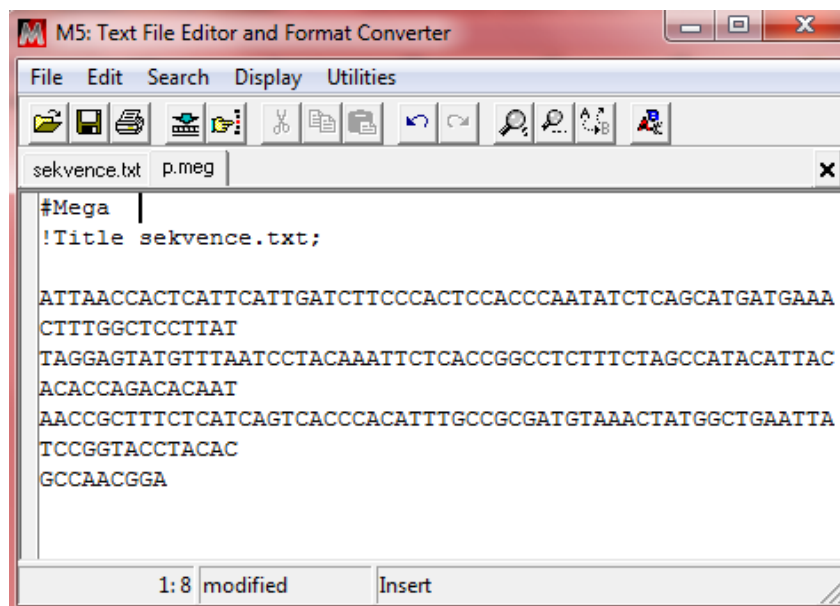
Kot vhodne datoteke za program MEGA uporabljamo ASCII tekstovne datoteke (.TXT format), ki vsebujejo nukleotidna zaporedja, proteinska zaporedja, statistične ali filogenetske podatke. Zaporedja so lahko zapisana v .ALN, .NEXUS, .PHYLIP, .GCG, .FASTA, .PIR ali .NBRF formatu. Za delo v programu MEGA moramo te datoteke nato spremeniti v .MEG format (File,

Convert File Format to MEGA, nato izberemo želeno datoteko ter v kakšnem formatu je zaporedje v datoteki zapisano). Po potrditvi nas program opozori, da moramo preveriti uspešnost prevajanja v format .MEG (Slika16).



Slika 16: Opozorilno okno. Preveriti moramo uspešnost prevajanja formatov.

Ko pritisnemo OK, se nam odpre okno za urejanje tekstovnih datotek in prevajanja formatov (angl. *Text File Editor and Format Converter*) (Slika17).



Slika 17: Okno za urejanje tekstovnih datotek in prevajanja formatov. V prvi vrstici mora biti ključna beseda #Mega. To pove, da je datoteka v MEGA formatu. V naslednji vrstici sledi !Title in nato ime datoteke z njenim formatom, ki smo jo izbrali (v našem primeru sekvence.txt). Stavki se zaključijo z znakom »;«. Nato sledi nukleotidno zaporedje ali aminokislinsko zaporedje. Pišemo lahko tudi komentarje, ki lahko obsegajo poljubno število vrstic. Označeni morajo biti z znakoma »[komentar]«.

Preden začnemo z analizo v programu MEGA, moramo poravnati nukleotidna zaporedja, če tega še nismo storili. Nato v meniju glavnega okna izberemo analizo, ki jo želimo izvesti. Program nam ponuja precejšnjo izbiro metod za izračun evolucijskih razdalj med nukleotidnimi in aminokislinskimi zaporedji. Za rekonstrukcijo filogenetskih dreves lahko izbiramo med metodami ML, UPGMA, NJ, MP itd. Rezultati se pojavijo nekaj trenutkov kasneje (odvisno od tega, koliko časa zahteva izbrana analiza). Glede na izbrano analizo so lahko v obliki drevesa, matrike, teksta itd.

Program nam ponuja vodnik, s katerim se lahko sprehodimo čez vse funkcije programa, vsebuje tudi že vnaprej pripravljene primere. Z vodnikom si lahko pogledamo poravnavo nukleotidnih zaporedij, ocenjevanje evolucijskih razdalj, rekonstrukcijo dreves itd.

3.3.1 Koda IUPAC, ki jo podpira MEGA

Koda je podobna prejšnji. Razlikuje se le v nekaj simbolih. Pri nukleotidnih zaporedjih je simbol za vrzel » - « (kot ga pozna tudi npr. CodonCode Aligner). Simbol » . « pa predstavlja identični nukleotid nukleotidu iz prvega zaporedja (različno od npr. CodonCode Aligner). Program pozna tudi dodaten simbol » ? «, ki nadomešča manjkajoče podatke.

Pri aminokislinskih zaporedjih, za razliko od programa CodonCode Aligner, program MEGA pozna le tiste simbole, ki pomenijo natanko eno aminokislino. Tako ni simbola B, ki označuje asparaginsko kislino ali asparagin. Ni simbola X, ki označuje katerokoli aminokislino in ni simbola Z, ki označuje glutaminsko kislino ali glutamin. Pojavi se dodaten simbol » * «, ki označuje konec oz. stop kodon.

3.3.2 Metoda vezanja ali test vezanja (angl. *Bootstrap*)

Ta statistična metoda odgovori na vprašanje, kako veliko je ujemanje med podatki in sorodstvenimi odnosi, upodobljenimi v drevesu. Metoda torej izmeri, kateri deli drevesa so slabo podprti s strani podatkov in pomaga predvideti, ali bi pri več podatkih bili rezultati enaki. Metoda

torej ne pove, ali so rezultati pravi. Slaba stran je velika obremenitev računalnika, saj mora vsako kopijo drevesa pregledati več stokrat (Holder in Lewis, 2003).

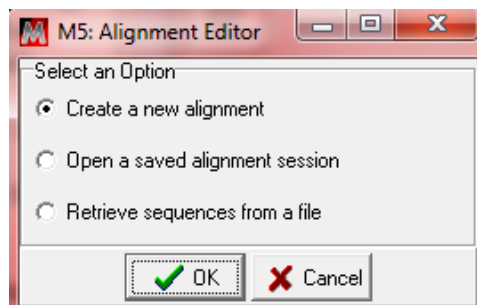
Algoritem: recimo, da je a nukleotidnih zaporedij (aminokislinskih zaporedij) in vsaka ima b nukleotidov (ali aminokislin). Uporabimo izbrano metodo za konstrukcijo drevesa. B nukleotidov vsakega nukleotidnega zaporedja je naključno izbranih z zamenjavo. Drevo je ponovno rekonstruirano z dobljenimi nukleotidnimi zaporedji. Metoda za izgradnjo drevesa ostaja ista.

Topologija drugega drevesa se primerja s topologijo prvega drevesa. Enake notranje veje označi z 1, različne pa z 0. To se več stokrat ponovi in vsakič, ko notranja veja dobi 1, se zabeleži. Končnemu številu rečemo vrednost vezanja. Če je ta vrednost posamezne veje 70 % ali več, se topologija te veje smatra za pravilno (Tamura in sod., 2008).

3.3.3 Potek dela v programu

3.3.3.1 Urejanje (nukleotidnih) zaporedij

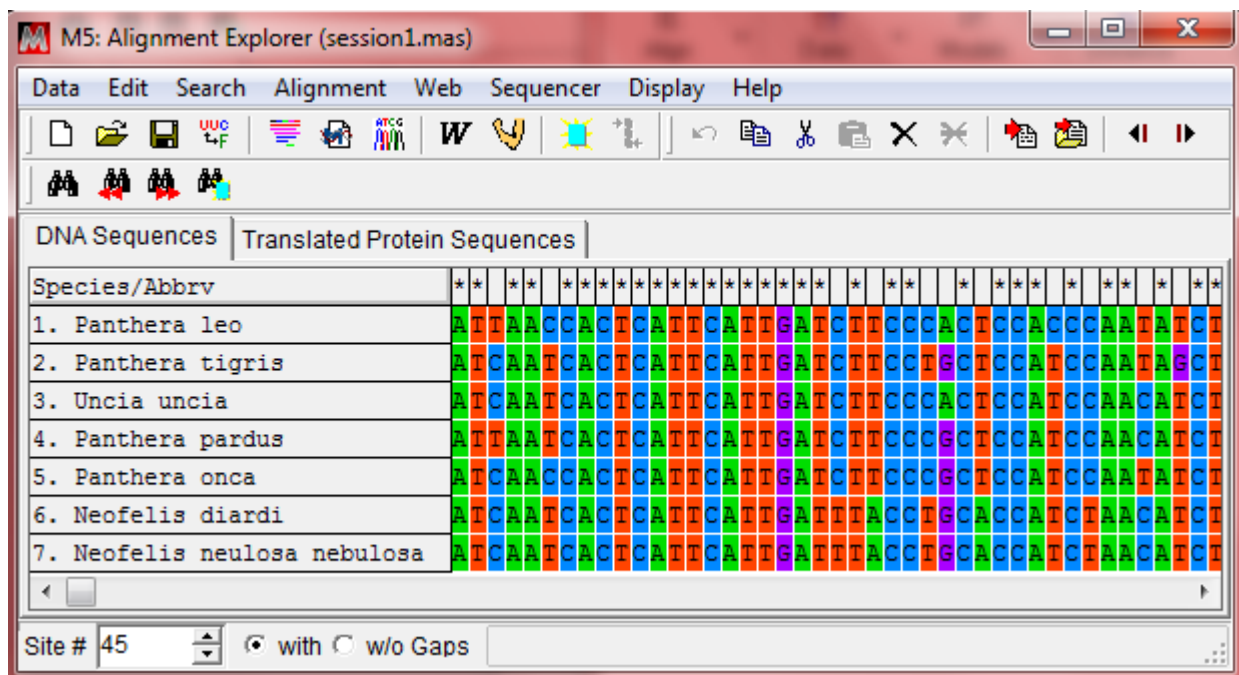
Z MEGA lahko poravnavamo nukleotidna zaporedja (pri tem program uporablja algoritma ClustalW ali MUSCLE). To naredimo tako, da izberemo Align, Edit/Build Alignment, nato izbiramo med Create a new alignment, Open a saved alignment session ali Retrieve sequences from a file (Slika 18).



Slika 18: Okno za urejanje poravnav

Recimo, da izberemo »Create a new alignment«. Odpre se nam okno, kjer izbiramo, ali bomo

poravnavali nukleotidna zaporedja ali aminokislinska zaporedja (Slika 19). Če želimo odpreti nukleotidno zaporedje, ki ga imamo na računalniku, izberemo Data, Open, Retrieve Sequences from a File in poiščemo datoteko, kjer imamo shranjeno datoteko z zaporedjem. Če že imamo shranjeno delo v *Alignment Session*, lahko izberemo tudi to. Imamo pa še možnost, da nukleotidna zaporedja poiščemo na svetovnem spletu, kar storimo tako, da izberemo Web, Show Browser, poiščemo želeno nukleotidno zaporedje in damo »Add to Alignment«.

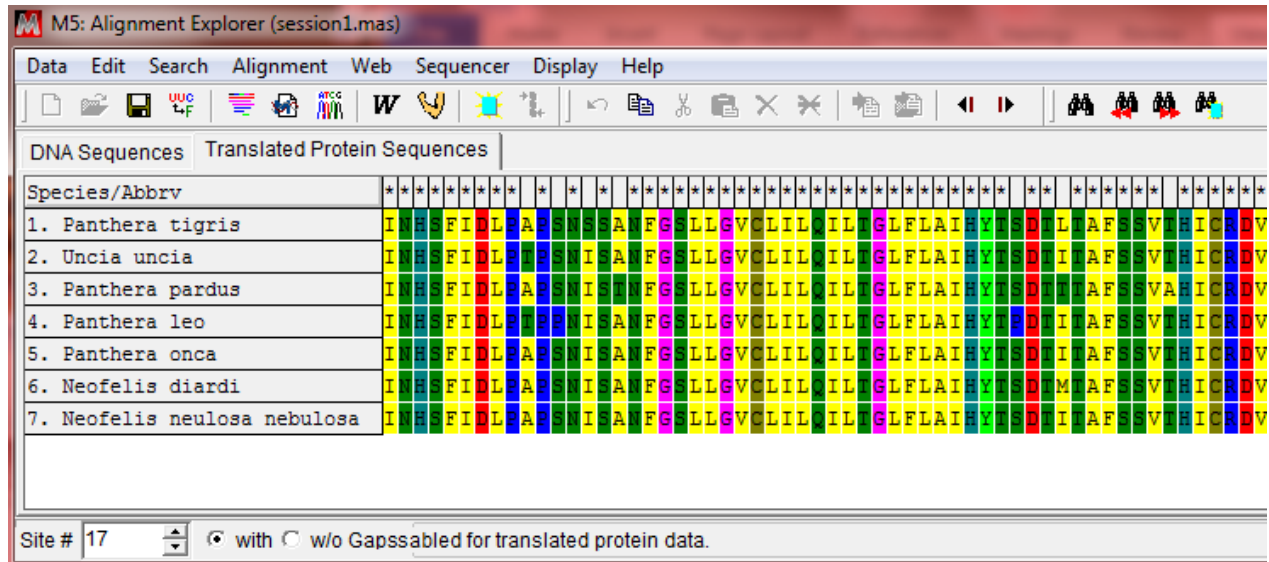


Slika 19: Okno »Alignment Explorer«, v katerem lahko med drugim tudi poravnamo nukleotidna zaporedja.

Ko izberemo vsa zelena nukleotidna zaporedja, jih poravnamo tako, da izberemo Edit, Select All, nato pa še Alignment, Align by ClustalW ali Align by Muscle. Kar smo dobili, je poravnava (angl. *Alignment session*), to nato shranimo v različnih formatih (format .mas je za poravnavo, format .meg je za nadaljnjo obdelavo v programu MEGA, format .fasta je za druge programe).

Nukleotidna zaporedja lahko prevedemo v aminokislinsko zaporedje (Slika 20), kar je koristno, kadar imamo kodirajočo regijo (npr. citokrom *b*). Pri tem je potrebno paziti, da se nukleotidno zaporedje začne s prvo pozicijo kodona, saj se v nasprotnem primeru znotraj aminokislinskega

zaporedja pojavijo STOP kodoni (ki so predstavljeni z znakom »*«). Če ni napak, so znaki »*« samo v zadnjem stolpcu.



Slika 20: Prejšnje nukleotidno zaporedje je prevedeno v aminokislinsko zaporedje.

3.3.3.2 Konstrukcija dreves

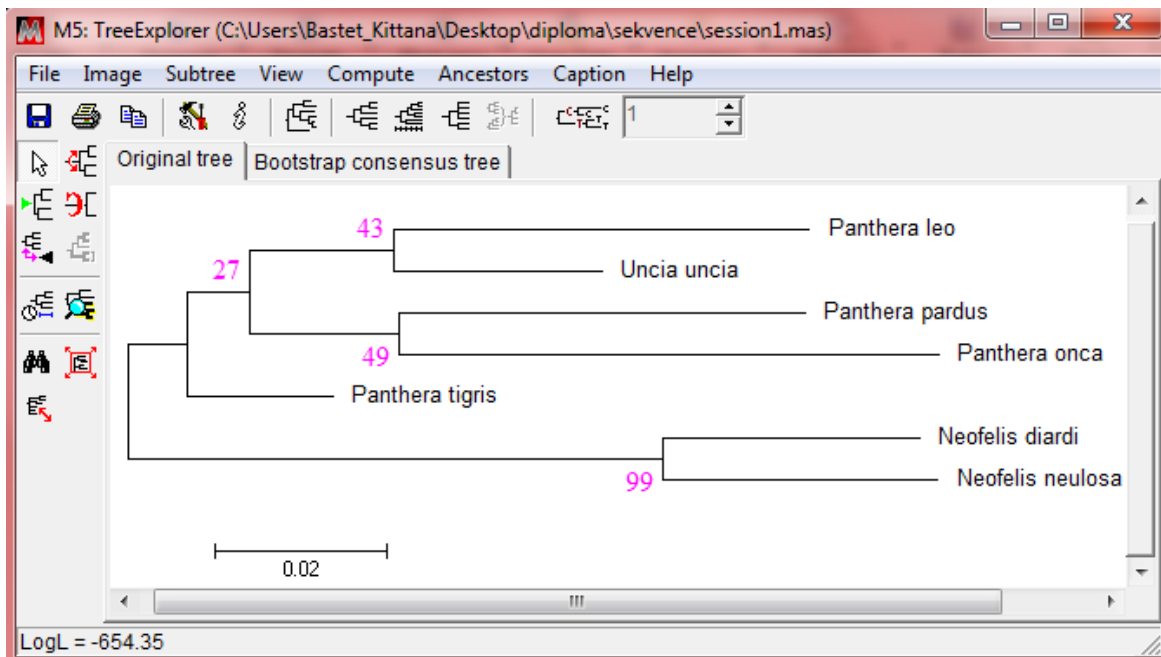
Najprej lahko izračunamo model evolucije (Slika 21). To storimo tako, da pritisnemo gumb *Models*, nato pa Find best DNA/Protein Models (ML), izberemo sejo zelenih zaporedij in potrdimo. Pritisnemo »Compute« in dobljeno shranimo (npr. podatkiML). Najboljši modeli, ki jih je smiselno uporabiti v nadaljnji analizi ML, so tisti z najmanjšo vrednostjo BIC oz. AIC.

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R
HKY+G	16	1376.497	1291.496	-629.568	n/a	0.26	12.06
HKY+I	16	1376.574	1291.573	-629.607	0.64	n/a	11.42
K2+I	13	1378.362	1309.247	-641.504	0.64	n/a	11.15
K2+G	13	1378.482	1309.367	-641.564	n/a	0.26	11.55
T92+I	14	1380.626	1306.213	-638.968	0.63	n/a	11.02
TN93+G	17	1383.360	1293.069	-629.333	n/a	0.26	12.26
T92+G	14	1383.718	1309.304	-640.514	n/a	0.54	10.48
HKY+G+I	17	1383.832	1293.541	-629.568	0.00	0.26	12.05
K2+G+I	14	1385.680	1311.267	-641.495	0.62	7.90	11.27
T92+G+I	15	1387.957	1308.248	-638.966	0.62	13.35	11.10
TN93+G+I	18	1390.695	1295.117	-629.333	0.00	0.26	12.26
HKY	15	1397.516	1317.808	-643.746	n/a	n/a	9.37

Slika 21: Izračunan model evolucije za ML

Ko imamo to, lahko konstruiramo drevo. Poleg *Maximum Likelihood* drevesa lahko konstruiramo še *Neighbor-Joining* drevo, *Minimum-Evolution* drevo, UPGMA drevo ali *Maximum Parsimony*.

Pritisnemo gumb *Phylogeny*, nato pa *Construct/Test Maximum Likelihood Tree*, izberemo sejo zelenih zaporedij in potrdimo. Odpre se nam okno *Analysis Preferences*. Kjer so vrstice obarvane rumeno, lahko izbiramo metode oz. kriterije. Pri *Initial Tree for ML* (v *Option* stolpcu) izberemo *Use tree from file* (v *Selection* stolpcu) in pri *Initial Tree File* (v *Option* stolpcu) poiščemo prej shranjeno datoteko *podatkiML.nwk* (v *Selection* stolpcu). Pritisnemo »Compute« in nato *Automatic Tree* ter počakamo nekaj trenutkov. Dobimo drevo, ki ga lahko poljubno preoblikujemo in obarvamo (Slika 22). Dostopamo lahko tudi do informacij o drevesu (Slika 23). Dobro je, da pogledamo drevesa različnih metod (Slika 24).

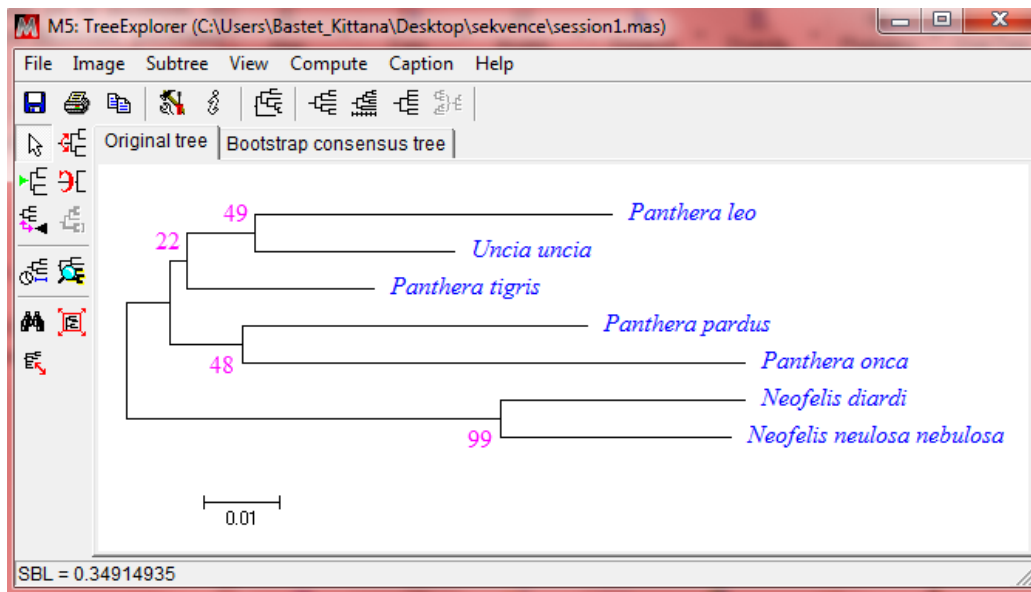


Slika 22: Izrisano *Maximum Likelihood* drevo osebkov rodu *Panthera* in *Neofelis* na njihovih nukleotidnih zaporedjih gena citokroma *b*. Uporabili smo še *Bootstrap* metodo s 1500 ponovitvami in model *Kimura 2-parameter*. Skala prikazuje 2 % divergenco nukleotidnih zaporedij.

The screenshot shows the 'M5: Information' dialog box with the 'Tree' tab selected. The 'Stay in Front?' checkbox is checked. The table below shows the tree information:

Type	Value
Type	Unrooted
SBL	0.37343461
Log Likelihood	-654.35
Model	K2
Ts/Tv	9.1606

Slika 23: Če v oknu *TreeExplorer* pritisnemo File, Show Information (Ctrl+I), dobimo informacije o drevesu. V našem primeru je drevo nekoreninjeno, model je *Kimura 2-parameter*.



Slika 24: Za primerjavo s prejšnjim drevesom je tukaj izrisano *Neighbor-Joining* drevo osebkov rodu *Panthera* in *Neofelis* na njihovih nukleotidnih zaporedjih gena citokroma *b*. Uporabili smo še *Bootstrap* metodo z 1000 ponovitvami. Skala prikazuje 1 % divergenco nukleotidnih zaporedij.

3.3.3.3 Ocenjevanje parnih razdalj

Parne razdalje nukleotidnih zaporedij (aminokislinskih zaporedij) se ponavadi merijo s številom zamenjav nukleotidov (aminokislin), ki se zgodijo tekom določenega časa. Te razdalje so osnova za preučevanje evolucije molekul, pomembne so tudi za filogenetske rekonstrukcije in pri ocenjevanju časa pojava divergence (Tamura in sod., 2008).

Program MEGA ponudi izbiro metod za ocenjevanje razdalj: nukleotidi (ang. *Nucleotide*), aminokisliline (ang. *Amino acid*) in sinonimno/nesinonimno (ang. *Syn-non-synonymous*).

Nukleotidi: nukleotidna zaporedja se primerjajo po posameznih nukleotidih, razdalja pa se računa tako na kodirajočih delih kot na nekodirajočih delih nukleotidnega zaporedja (Tamura in sod., 2008). Izbiramo lahko med naslednjimi modeli:

- *p-distance*,

Razdalja je rezultat deleža (p) mest, kjer se nukleotidni zaporedji, ki ju primerjamo,

razlikujeta (število nukleotidov, ki se razlikujejo delimo s številom vseh nukleotidov, ki jih primerjamo). Računa lahko za tranzicije, transverzije ali oboje.

- **Kimura 2-Parameter Model,**

Osnova modela Kimura 2 je model Jukes in Cantor (angl. *Jukes and Cantor basic model*). Razlikuje tranzicijo in trasverzijo. Model privzame, da je stopnja tranzicij različna od trasverzij.

Matrika stopenj. Stopnja transverzij je 1, κ predstavlja stopnjo razmerja med tranzicijami in transverzijami. Ta model predvideva, da se vse baze enako ponavljajo ($\pi_T = \pi_C = \pi_A = \pi_G = 0.25$). Matrika stopenj označena s Q .

$$Q = \begin{pmatrix} * & \kappa & 1 & 1 \\ \kappa & * & 1 & 1 \\ 1 & 1 & * & \kappa \\ 1 & 1 & \kappa & * \end{pmatrix}$$

Enačba za razdaljo (d) modela Kimura, kjer p pomeni delež območij, ki predstavljajo razliko tranzicij, q pa delež območij, ki predstavljajo razliko trasverzij.

$$\hat{d} = -\frac{1}{2} \ln(1 - 2p - q) - \frac{1}{4} \ln(1 - 2q)$$

- **Jukes-Cantor Model,**

Stopnja nukleotidnih zamenjav je ista za vse nukleotide. Oцени najbolj možno število (ML) zamenjav nukleotidov. Računa za tranzicije in transverzije.

- **Maximum Composite Likelihood Model,**

- **Tamura 3-Parameter Model,**

Upošteva razlike stopenj tranzicij in transverzij. Privzema, da so zamenjave enakomerne. Računa za tranzicije, transverzije ali oboje.

- **Tajima-Nei Model,**

Podobno kot prejšnji model, ocene so boljše. Računa za tranzicije in transverzije.

- **Tamura-Nei Model,**

- *Log-Det Method.*

Sinonimno/nesinonimno: aminokislinska zaporedja se primerjajo po posameznih kodonih.

Uporabljajo se torej samo za aminokislinska zaporedja (proteine) (Tamura in sod., 2008). Metode:

- *Nei-Gojobori Method,*
- *Li-Wu-Luo Method,*
- *Kumar Method,*
- *Modified Nei-Gojobori Method,*
- *Pamilo-Bianchi.Li Method.*

Aminokislina: aminokislinska zaporedja se primerjajo po posameznih aminokislinah. Razdalja se lahko računa za aminokislinska zaporedja ali nukleotidna zaporedja, ki kodirajo proteine (te se samodejno prevedejo) (Tamura in sod., 2008). Metode:

- *p-distance,*
- *Equal Input,*
- *Poisson Model,*
- *Dayhoff and JTT Models.*

4 DISKUSIJA

Pri sklepanju sorodstvenih odnosov med organizmi v molekularni filogeniji se privzame, da so genska drevesa neke skupine taksonov izomorfna filogenetskim drevesom. Če to privzamemo, lahko drevesi preoblikujemo iz genskega v filogenetsko in obratno. To storimo tako, da zamenjamo imena zaporedij z ustreznim organizmom (organizmom, ki mu pripada zaporedje) (Page in Holmes, 2009).

Potrebno se je zavedati, da v resnici (če ne privzamemo prej navedenega) ta preoblikovanja niso tako preprosta. Zaradi genskega (kromosomskega) podvajanja, izgube, razvrščanja linij in horizontalnega genskega prenosa ni nujno, da sta gensko in filogenetsko drevo izomorfni. Vzroki za neujemanje dreves so namreč različni od vzrokov za homoplazijo v zaporedjih (Page in Charleston, 1997).

Program GeneTree primerja genska in filogenetska drevesa s pomočjo usklajenih dreves. Program je brezplačno dostopen na spletnem naslovu <http://taxonomy.zoology.gla.ac.uk/rod/genetree/genetree.html>. Izračuna ceno pretvarjanja genskega drevesa v filogenetsko drevo. Vizualno predstavi lokacijo in število genskih podvajanj in izgub ter na podlagi tega poišče najboljše filogenetsko drevo (Page, 1998).

Razen CodonCode Alignerja, so ostala obravnavana orodja in programi brezplačno dostopni na spletu. FASTA in BLAST nam torej nudita poravnavanje parov nukleotidnih zaporedij ali aminokislinskih zaporedij. Na tak način lahko ugotavljamo, koliko in kje so si izbrani nukleotidni zaporedji podobni oz. različni. Iz tega lahko sklepamo sorodnost taksonov, podani so tudi odstotki ujemanja zaporedij.

Nukleotidna zaporedja lahko poravnavamo v programih CodonCode Aligner in MEGA. CodonCode Aligner je najbolj uporaben, kadar imamo opravka z neobdelanimi zaporedji v .abi formatu. Nudi nam veliko izbiro orodij za lažjo poravnavo (nukleotidnega) zaporedja in tudi rezultati so boljši. Težava, na katero lahko naletimo pri delu je, da včasih bazo težko določimo.

Razlogov za to je lahko več (slaba kakovost nukleotidnih zaporedij, premalo vzorcev itd.). Pri velikem številu vzorcev je program počasen. Za uporabnika je neugodno, da je program plačljiv.

Filogenetsko drevo najlepše izrisujemo v programu MEGA, čeprav lahko preprosto drevo (ki temelji na metodi NJ) konstruiramo tudi v CodonCode Alignerju. Ko imamo že obdelana nukleotidna zaporedja, je izris lahek. Program nam konstruira drevo, ki prikazuje sorodstvene odnose izbranih taksonov. Izbiramo lahko med velikim številom dreves oz. metod, s katerimi bomo konstruirali drevo. Rezultati različnih metod se lahko malce razlikujejo, saj se razlikujejo tudi cilji, ki jih posamezne metode želijo doseči. Dobro je uporabiti več metod.

V filogenetskih raziskavah, ki temeljijo na nukleotidnih zaporedjih, se poleg Bayesove metode (angl. *Bayesian*) večinoma uporablja metoda ML (Bayesove metode v nalogi niso obravnavane). Metoda je statistično boljša od drugih, saj pri izračunu uporablja predpostavljeni model evolucije. Slednje je bolj točno, kot če bi privzeli, da so mesta taksonov neodvisna. MP in UPGMA se bolj uporabljata za ocenjevanje filogenije na podlagi morfoloških znakov (Lewis, 2001).

Metoda ML je iz več vidikov najprimernejša. Omogoča rekonstrukcijo sorodstvenih odnosov zaporedij, ki so bila dalj časa ločena ali pa se razvijajo zelo hitro. Izriše drevo največjega verjetja (največja verjetnost, kako so zaporedja vpeta v drevo in najbolj verjetni sorodstveni odnosi med njimi). Pri tej metodi moramo najprej izračunati verjetnost danih zaporedij. Slaba stran metode je, da precej obremeni računalnik, saj mora analizirati mnogodimenzionalni prostor parametrov (Holder in Lewis, 2003).

Tudi metoda NJ je razširjena, algoritem je hiter. Dobro se obnese, če je divergenca zaporedij nizka. Iz zaporedij moramo najprej izračunati matriko razdalj. Pri tem se nekateri podatki izgubijo, zato se lahko zgodi, da opazovane razlike zaporedij ne prikažejo prave filogenije (to se na splošno dogaja pri distančnih metodah) (Holder in Lewis, 2003).

Metoda MP za razliko od distančnih metod doda drevesu tudi »zgodovino«^g genskih zaporedij (upoštevajo mutacije). Metoda nam izriše drevo z najmanjšim možnim številom mutacij, glede na dane podatke. Algoritem je hiter, če upoštevamo, da mora pregledati na stotine/tisoče dreves. Slaba stran je, da izrisano drevo temelji izključno na najmanjšem številu mutacij. Nastajanje mutacij je

nepredvidljivo in naključno, zato se lahko zgodi, da dobljeno drevo ne prikaže prave filogenije. Slaba stran metode pa je tudi to, da predvideva, da je število sprememb enako na vseh vejah (kar ni nujno, da drži) (Holder in Lewis, 2003).

Program MEGA nudi veliko funkcij, da lahko drevo izrišemo, kakor želimo. Nudi nam več tipov dreves, drevo lahko obarvamo. Pomembno je, da lahko izračuna za koliko se dve zaporedji razlikujeta med seboj. Število je napisano na posamezni veji drevesa, ki povezuje ustrezna taksona. Razdalja je v bistvu število, ki pove, za koliko se razlikujeta takson in njegov potomec.

5 ZAKLJUČEK

Ekologija pokriva veliko področje in je ena najpomembnejših znanosti, ki pomaga pri upravljanju živih bitij in okolja. Proučuje življenjske procese in prilagoditve organizmov, potovanje snovi in energije skozi ekosisteme ter razporeditev in bogastvo biodiverzitete.

Molekularna ekologija odgovarja na ekološka vprašanja. Z različnimi metodami (določanje nukleotidnega zaporedja, RAPD, mikrosateliti itd.) se ugotavlja podobnost med organizmi in populacijami. Ugotovi se lahko tudi kateri vrsti oz. populaciji organizem pripada, proučuje se lahko dednost in genetske spremembe v evoluciji itd. Sorodna je področjem ekološke genetike in varstvene genetike. Pri analizi genov si pomagamo z določenimi programi.

Analiza polimorfizma *mtDNK* je zelo priljubljena za ugotavljanje filogenetskih odnosov taksonov in njihove evolucijske zgodovine. Mitohondrije in z njimi *mtDNK* organizmi dedujejo po materi. Posledica tega je, da pri *mtDNK* ne prihaja do rekombinacij (to velja vsaj za višje razvite živali). Zanimivo je tudi, da se pri sesalcih razvije desetkrat hitreje kot jedrna DNK. Zaradi teh značilnosti je *mtDNK* zelo pomembna pri proučevanju molekularne genetike populacij in v sistematiki (Page in Holmes, 2009).

Dvodimenzionalni graf, ki prikazuje te sorodstvene odnose, je filogenetsko drevo. Odnosi so hierarhično urejeni. Veje se vedno cepijo, nikoli ponovno združujejo. Iz ene predniške vrste načeloma dobimo dve hčerinski vrsti. Poznamo koreninjeno drevo in nekoreninjeno drevo, slednje nam ponuja manj podatkov.

Metode za rekonstrukcijo dreves razdelimo na tiste, ki temeljijo na izračunanih razvojnih razdaljah med danimi organizmi, in tiste, ki temeljijo na lastnostih oz. stanjih znakov danih organizmov. Največkrat se uporabljajo: metoda največje varčnosti, distančna metoda in metoda največjega verjetja. Vse rekonstrukcije dreves so v bistvu samo predvidevanja, kako naj bi evolucija proučevanih organizmov potekala.

Za dobro analizo nukleotidnih zaporedij in dobre rezultate je potrebno dobro poznavanje programov. V nalogi sta predstavljeni orodji FASTA in BLAST ter programa CodonCode Aligner 4.0.3 in MEGA 5.05. Predstavljene so glavne funkcije programov, ostale si lahko pogledamo v priročnikih (ang. *Manual*) za posamezni program. Oba priročnika sta dostopna na spletu. Poleg tega oba programa nudita tudi možnost pomoči (ang. *Help*), kjer najdemo veliko koristnih informacij. Gumb Help navadno najdemo v orodni vrstici glavnega okna, skrajno desno. Za program MEGA je zelo uporabna tudi spletna stran http://www.megasoftware.net/mega4/manual_pdf.html.

Da ugotovimo filogenetske odnose, potrebujemo nukleotidna zaporedja taksonov, ki jih želimo obravnavati, in določene programe, s katerimi dobimo željeno. Nukleotidna zaporedja moramo najprej poravnati in nato konstruirati drevo. V nalogi smo uporabili nukleotidno zaporedje gena citikroma *b* vrst iz rodu *Panthera* in *Neofelis*. Opisana poravnava izbranih nukleotidnih zaporedij je narejena v CodonCode Alignerju. Filogenetsko drevo smo konstruirali v programu MEGA, pri čemer smo uporabili metodo ML in metodo vezanja s 1500 ponovitvami.

6 POVZETEK

6.1 POVZETEK

Ekologija je področje biologije, ki proučuje značaj, razporeditev, številčnost, pomen in spremembo stanja organizma znotraj enega ekosistema in med različnimi ekosistemi. Praktično uporabo ekologije lahko zasledimo med drugim tudi v varstveni biologiji. Molekularna ekologija je podpodročje ekologije in odgovarja na ekološka vprašanja. Ugotavlja tudi kateri vrsti oz. populaciji organizem pripada, proučuje dednost in genetske spremembe v evoluciji, spreminjanje in prilagajanje organizmov na genetskem nivoju itd. Sorodna je področjema ekološke genetike in varstvene genetike.

Varstvena biologija skrbi za ohranjanje biodiverzitete. V pomoč so ji informacije, pridobljene z analizo (nukleotidnih) zaporedij. Na podlagi teh informacij lahko nato ustrezno ukrepamo. Vrsti/populaciji lahko pomagamo preživeti, jo uvrstimo na Rdeči seznam ipd.

V molekuli DNK so zbrane genske informacije. Sestavljena je iz adenina, timina, citozina in gvanina. Baze se med seboj vežejo v pare A in T z dvojno vezjo ter G in C s trojno vezjo. Vsaka nukleotidna baza je povezana s sladkorjem in fosfatom. Poznamo jedrno in mitohondrijsko DNK. Za ugotavljanje sorodnosti organizmov je bolj pomembna slednja. Nahaja se v mitohondrijih celice. Tam je prisoten večji polimorfizem, ker se mutacije kopičijo veliko hitreje kot v jedru. Mutacijskim spremembam (v primerjavi z jedrno DNK) veliko lažje sledimo, ker se ne rekombinira.

Citokrom *b* je beljakovina na zunanji membrani mitohondrija. V ciklusu citronske verige skrbi za prenos elektronov. Gen citokroma *b* dovolj variira, da lahko odgovorimo na vprašanja na nivoju populacije, hkrati pa se dovolj ohranja, da lahko z njim ugotavljamo daljne sorodstvene odnose. Zaradi teh lastnosti se pogosto uporablja in zato lahko različne študije brez težav primerjamo.

Filogenetsko drevo prikazuje hierarhične sorodstvene odnose med organizmi v obliki dvodimenzionalnega grafa. Določeni organizem je na drevesu predstavljen kot sistematska enota ali takson. Hierarhija se začne z najmanjšim taksonom vrsta (*species*), večji takson je rod (*genus*),

sledi družina (*familia*), red (*ordo*), razred (*classis*), konča se z največjim taksonom debla (*phylum*). Veje drevesa se vedno cepijo, navadno na dve veji. Takšno drevo je binarno.

Nekatere metode za rekonstrukcijo dreves temeljijo na izračunanih razvojnih razdaljah med danimi organizmi, druge temeljijo na lastnostih oz. stanjih znakov danih organizmov. Največkrat uporabljene metode so: metoda največje varčnosti (angl. *Maximum Parsimony Method*), distančna metoda (angl. *Distance Method*) in metoda največjega verjetja (angl. *Maximum Likelihood Approach*). Rekonstrukcije dreves so le predvidevanja sorodstvenih odnosov taksonov.

Postopek risanja filogenetskih dreves

- FASTA in BLAST sta prosto dostopna na spletu. Nudita poravnavanje parov nukleotidnih zaporedij ali aminokislinskih zaporedij. Na tak način lahko ugotovljamo, koliko in kje sta si izbrani zaporedji podobni oz. različni. Iz tega lahko sklepamo sorodnost taksonov, podani so tudi odstotki ujemanja zaporedij.
- CODON CODE ALIGNER
V nalogi je obravnavana verzija 4.0.3. Program je plačljiv. Nudi nam več funkcij za poravnavo neobdelanih nukleotidnih zaporedij. Slaba stran je, da včasih baze ne moremo določiti. Razlogov za to je lahko več (slaba kakovost nukleotidnih zaporedij, prememalo vzorcev itd.).
- MEGA (Molecular Evolutionary Genetics Analysis)
V nalogi je obravnavana verzija 5.05. Programski paket je primeren predvsem za konstrukcijo filogenetskih dreves in računanje osnovnih statističnih podatkov. Program je prosto dostopen.

6.2 SUMMARY

Ecology is a branch of biology. It studies amount, distribution, composition, number and changing states of organisms within one ecosystem and among different ecosystems. Practical use of ecology can be found, among other fields, also in the conservation biology. Molecular ecology is branch of

ecology and it answers ecological questions. It also ascertains which species or population organism belong, it studies inheritance and genetic changes in evolution, modification and adaptation of organisms on genetic level etc. Molecular ecology is related to ecological genetics and conservation genetics.

Conservation biology takes care of Earth's biodiversity. Information from sequence analysis is needed to support that. With that information decisions can be made properly. We can help species/populations to survive, place them to Red list etc.

The DNA molecule contains genetic informations. Adenine, cytosine, guanine and thymine are four bases found in the DNA. These bases are attached to the sugar and phosphate. Nucleobases pair up with each other (base pairing). A is bonding exclusively to T and C is bonding only to G. There are two types of DNA: nuclear DNA and mitochondrial DNA. For phylogenetic researches the mitochondrial DNA is of great importance. It is found in the Mitochondria. Because it accumulates mutations faster than a cell, stronger polymorphism takes place. On the other hand it does not recombine, so we can easily track mutations.

Cytochrome *b* is an integral membrane protein. It is a component of respiratory chain and it is involved in the electron transport. We use its gene to study population due to its sequence variability, at the same time it is preserved well enough to determine phylogenetic relationships between organisms. That makes cytochrome *b* an extremely important study tool in molecular population genetics.

Phylogenetic tree is a two-dimensional graph of hierarchic relationships between organisms. They are called taxa. The lowest rank of hierarchy is occupied by *species*, followed by *genus*, *family*, *order*, *class*, and lastly on the highest rank, *phylum*. The branch of the tree usually always splits into two branches.

Phylogenetic trees are constructed using different methods. Some methods are based on calculating genetic distances from multiple sequence alignments, (and) others (methods) are based on general

similarity. Methods of great importance are Maximum Parsimony Method, Distance Method and Maximum Likelihood Approach. We have to know that all trees constructions are only a forecast.

Making phylogenetic tree procedure

- FASTA and BLAST are *freeware*. They are used for finding regions of local similarity between protein or DNA sequences. That way we can see how (much) and where the defined sequences are similar so we can predict taxa relationships.
- CODON CODE ALIGNER
The thesis describes software version 4.0.3. The program is *retail software*. It has many functions for the sequence alignment.
- MEGA (Molecular Evolutionary Genetics Analysis)
The thesis describes software version 5.05. This program package is *freeware*. We can construct filogenetic trees and calculate basic statistical data.

7 VIRI IN LITERATURA

- Alexeyev, M. F., LeDoux S. P., Wilson, G. L., 2004. Mitochondrial DNA and aging. *Clinical Science* 107, str: 355-364. URL: <http://www.clinsci.org/cs/107/0355/1070355.pdf>
- Arbogast, B. S., Kenagy, G. J., 2001. Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography* 28, str: 819-825. URL: <http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2699.2001.00594.x/pdf>
- Asard, H., Venken, M., Caubergs, R., Reijnders, W., Oltmann, F. L., De Greef, J. A., 1989. *b*-Type Cytochromes in Higher Plant Plasma Membranes. *Plant Physiology*, 90, str. 1077-1083.
- Carr, S. M., Vertebrate mitochondrial DNA (mtDNA). 2009. URL: http://www.mun.ca/biology/scarr/mtDNA_genome.html
- Chapman, J. L., Reiss, M. J., 1999. *Ecology: principles and applications*. Second edition. Cambridge, Cambridge University Press, 330 str.
- CodonCode Corporation, 2012. *CodonCode Aligner User Manual*. URL: <http://www.codoncode.com/aligner/AlignerHelp.pdf>
- Degli Esposti, M., De Vries, S., Crimi, M., Ghelli, A., Patarnello, T., Meyer, A., 1993. Mitochondrial cytochrome b: evolution and structure of protein. *Biochimica et Biophysica Acta*, 1143, str: 243-271.
- DNA Handbook. Mitochondrial DNA. URL: <http://www.dnahandbook.com/pages/10176.htm> (Citirano 5.6.2012)
- Dobrič, K., 2012. Ministrstvo za notranje zadeve, Policija, Center za forenzične preiskave. Ugotavljanje identitete posameznika z uporabo genetskih informacij, metode verižne reakcije s polimerazo in računalniško podprte tehnologije. URL: <http://www.zrss.si/bzid/geni/pdf/drobnic-clanek.pdf>
- Egerton, F. N., 2001. A History Of The Ecological sciences: early Greek origins. *Bulletin of the Ecological Society of America* 82(1), str. 93-97.
- Farias, I. P., Orti, G., Sampaio, I., Schneider, H., Meyer, A., 2001. The Cytochrome b Gene as a Phylogenetic Marker: The Limits of Resolution for Analyzing Relationships Among Cichlid Fishes. *Journal of Molecular evolution*, 53, str. 89-103. URL: http://www.evoamazon.net/Legal_papers/Farias%202001.pdf
- Felsenstein, J., 1981. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution* 17, str: 368-376. URL: <http://ctb.pku.edu.cn/main/2011course/Sequence%20analysis/Felsenstein%201981.pdf>
- Freeland, J. R., Peterson S. D., Kirk H., 2011. *Molecular ecology*. Second edition. Oxford, Wiley-Blackwell, 449 str.
- Genetics Information. Detecting DNA. 2010. URL: <http://genesinfo.blogspot.com/> (Citirano 5.6.2012)
- Guralnick, R., Collins, A., Waggoner, B., Speer, B., 1994. *Introduction to Phylogeny*. URL:

- <http://www.ucmp.berkeley.edu/exhibit/introphylo.html> (Citirano 9.8.2012)
- Hadrys, H., Balick, M., Schierwater, B., 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology* 1, str. 55-63.
- Hagstrom, G. I., Hang, D. H., Ofria, C., Torng, E., 2004. Using Avia to Test the Effects of Natural Selection on Phylogenetic Reconstruction Methods. *Artificial Life* 10, str. 157-166. URL: <http://authors.library.caltech.edu/13452/1/HAGal04.pdf>
- Holder, M., Lewis, P. O., 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nature reviews genetics* 4, str: 275-284. URL: <http://cgl.bioinfo.uqam.ca/bif7001/articles/BIF7001-Phylo-NatureReviewsGenetics4a.pdf>
- IUPAC, 2012. URL: <http://www.iupac.org/home/about.html> (Citirano 11.8.2012)
- Jakovac, M., Bioinformatika. BLAST. 2009. URL: <http://www.bioinformatika.si/2009/06/blast.html> (Citirano 23.6.2012)
- Johns, G. C., Avise, J. C., 1998. A Comparative Summary of Genetic Distances in the Vertebrates from the Mitochondrial Cytochrome *b* Gene. *Molecular Biology and Evolution*, 15(11), str. 1481-1490.
- Kim, J., Sanderson, M., 2008. Penalized Likelihood Phylogenetic Inference: Bridging the Parsimony-Likelihood Gap. *Systematic Biology*, 57(5), str. 667-674. URL: <http://sysbio.oxfordjournals.org/content/57/5/665.full.pdf>
- Kocher, T. D., 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceeding of the Natural Academy of Sciences of the United States of America (PNAS)*, 86, str. 6196-6200.
- Lewis, P. O., 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology*, 50(6), str: 913-925. URL: <http://www.biology.ualberta.ca/courses.hp/biol606/papers/Lewis.2001.pdf>
- Lio, P., Goldman N., 1998. Models of Molecular Evolution and Phylogeny. *Genome Research*, 8, str. 1233-1244.
- McClellan, P., 1998. Plant Genome Organization and Structure. Mitochondrial Genome Organization. URL: <http://www.ndsu.edu/pubweb/~mcclellan/plsc731/genome/genome8.htm> (Citirano: 8.8.2012)
- Mount, D., 2004. *Bioinformatics: Sequence and Genome Analysis* (second edition). Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, 692 str.
- NCBI/BLAST Home. URL: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- NCBI/BLAST Query Input and database selection. URL: <http://www.ncbi.nlm.nih.gov/BLAST/blastcgihelp.shtml> (Citirano 5.6.2012)
- Page, R. D. M., 1998. GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics applications note*, 14 (9), str: 819-820.
- Page, R. D. M., Holmes, E. C., 2009. *Molecular Evolution: A Phylogenetic Approach*. Blackwell

Publishing company, 346 str.

- Page, R. D. M., Charleston, M. A., 1997. From Gene to Organismal Phylogeny: Reconciled Trees and the Gene Tree/Species Tree Problem. *Molecular Phylogenetics and Evolution* 7(2), str: 231-240.
- Palmer, J. D., 1992. *Mitochondrial DNA in Plant Systematics: Applications and Limitations. Molecular Systematics of Plants* Chapman & Hall, New York, 173 str.
- Rokas, A., Ladoukakis E., Zouros E., 2003. Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution*, 18(8), str: 411-417.
- Svarog, 2012. *Sistematika živih organizmov, Filogenija* URL: http://www.svarog.si/biologija/MSS/index.php?page_id=11534 (Citirano 9.8.2012)
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2008. MEGA: Molecular Evolutionary Genetics Analysis (Version 4.0). Arizona State University, Center for Evolutionary Functional Genomics. URL: <http://www.megasoftware.net/mega4/mega4.pdf>
- Theobald, D., 29+ Evidences for Macroevolution. *Phylogenetics Primer*. URL: <http://www.talkorigins.org/faqs/comdesc/phylo.html> (Citirano 16.6.2012)
- Trontelj, P., 2009. Filogenetske osnove biološke sistematike. V: Krajšek S., Vičar M., Vilhar B. (ur.). *Biodiverziteta- raznolikost živih sistemov*, Mednarodni posvet Biološka znanost in družba. Ljubljana, Zavod RS za šolstvo, str. 29-35. URL: http://www.zrss.si/bzid/biodiverziteta/pdf/Zbornik_BZID_Biodiverziteta_2009_splet.pdf
- Wei, Y. H., Lu, C. Y., Wei, C. Y, Ma, Y. S., Lee, H. C., 2001. Oxidative Stress in Human Aging and Mitochondrial Disease-Consequences of Defective Mitochondrial Respiration and Impaired Antioxidant Enzyme System. *Chinese Journal of Physiology*, 44(1): str: 1-11.