



ФИЛОГЕНЕТИЧЕН АНАЛИЗ

ПРОФ. ПЛАМЕНКА БОРОВСКА

ЕВОЛЮЦИОННА БИОЛОГИЯ

EVOLUTIONARY BIOLOGY

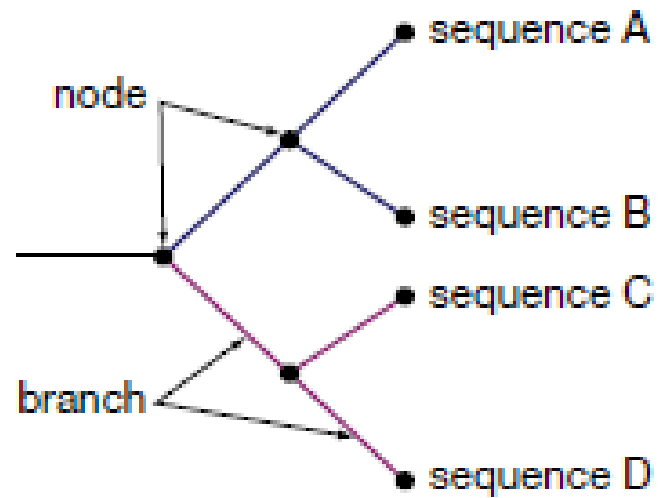
- Еволюционната биология е област от биологията, изучаваща произхода и развитието на видовете, както и тяхната промяна, умножаването и разнообразието във времето.
- От гледна точка на биологията, еволюцията представлява промяната в наследените черти на популацията от поколение на поколение.
- Тези черти са израз на гени, които са копирани и предавани на поколението по време на репродукцията.
- Мутациите в тези гени могат да произвеждат нови или променени черти, в резултат на наследствени разлики (генетична вариация) между организмите.
- Нови черти могат да бъдат получени и от прехвърляне на гени между популациите, както при миграцията, така и между видовете, при хоризонталния пренос на гени.
- Налице е еволюция, когато тези наследствени разлики станат по-чести или редки в дадена популация, или на неслучаен принцип чрез естествен подбор или на случаен принцип чрез генетично отклонение.

КОНЦЕПЦИЯТА НА ЕВОЛЮЦИОННИТЕ ДЪРВЕТА

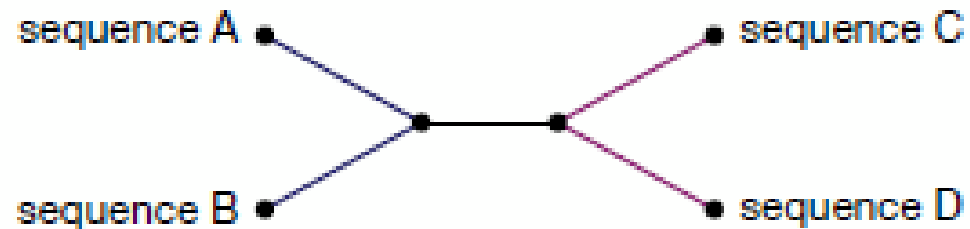
THE CONCEPT OF EVOLUTIONARY TREES

- Еволюционното дърво представлява двумерен граф, отразяващ еволюционните връзки между организмите или в случай на секвенции, в някои гени на отделни организми
- Отделните секвенции се наричат **таксони** (ед.ч. таксон), и се дефинират като отделни филогенетични единици на дървото.
- Дървото е съставено от външни клони (или листа), представящи таксоните и възлите, и също така клони, представящи взаимоотношенията между таксоните

A. Rooted tree



B. Unrooted tree



СКОРОСТИ НА ЕВОЛЮЦИЯТА

RATES OF EVOLUTION

- Някои видове филогенетични анализи предполагат, че темповете на еволюция в клоните на дърветата са едни и същи, а други предполагат, че те се различават.
- Допускането за постоянна скорост на мутацията в клоните на дървото е известно като *хипотезата за молекулярния часовник (the molecular clock hypothesis)* и обикновено се прилага за тясно свързани видове
- Дори да има еднаква скорост на еволюционните промени, статистическите вариации от един клон към друг клон, може да окажат влияние върху анализа.
- Броят на заместванията на нуклеотиди във всеки клон обикновено се приема, че варира в зависимост от разпределението на Поасон

ЕВОЛЮЦИОННО ВРЕМЕ

EVOLUTIONARY TIME

- В общия случай, еволюционното време, за което се е извършило разделянето на секвенция А от секвенция В, не е известно.
- Това, което може да бъде оценено на основата на филогенетичния анализ, е броят на измененията в секвенцията между възела А/В и възела А, и също така между възела А/В и възела В.
- Следователно, като се съди по дължината на клоните от този възел до А и В, се вижда, че броят на възникналите промени е един и същ.
- Освен това, също така е доста вероятно, че поради някакви биологични причини или под влияние на околната среда, уникални за всеки вид, единият таксон може да е претърпял повече мутации при разклоняването от прародителя отколкото другия таксон. В този случай, в еволюционното дърво ще бъдат маркирани различни дължини на клоните.

ЕВОЛЮЦИОННИЯТ ЧАСОВНИК

THE EVOLUTIONARY CLOCK

- Молекулярният часовник (базиран на **molecular clock hypothesis (MCH)**) представлява техника при молекулярната еволюция, която използва ограниченията на фосилите (изкопаемите) и темповете на молекулярните промени за оценка на времето в геоложката история, когато е възникнало разклоняването на два вида или два таксона.
- Молекулярният часовник се използва за оценка на времето на възникване на събития, наречени „образуване на видове“ (speciation) или „радиация“ (radiation).
- Молекулярните данни, използвани за тези изчисления, обикновено представляват секвенции от нуклеотиди за ДНК или секвенции от аминокиселини за протеини.
- Молекулярният часовник понякога се нарича генен часовник (**gene clock**) или еволюционен часовник (**evolutionary clock**).

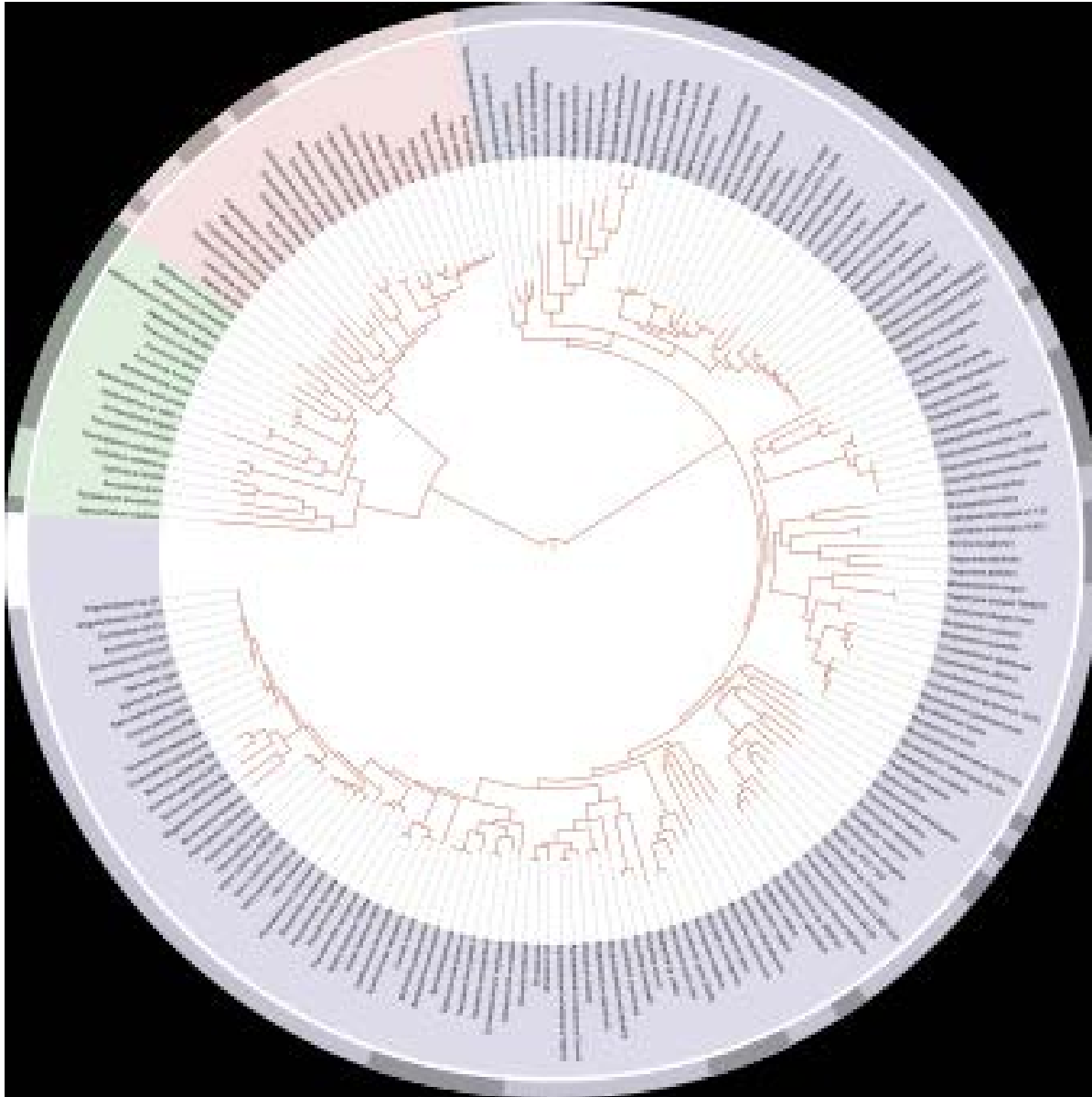
ДЪРВОТО НА ЖИВОТА

TREE OF LIFE



- Идеята за "дърво на живота" произтича от древните представи за стълбовидното преминаване от по-нисши към по-висши форми на живот
- **Charles Darwin (1859)** създава една от първите илюстрации и има изключителен принос за популяризирането на идеята за еволюционното "дърво" с неговата книга „Произход на видовете“ (*The Origin of Species*).
- Повече от един век, еволюционните биолози все още използват диаграми на дървото, за да изобразяват еволюцията, защото такива диаграми ефективно представят идеята, че **образуването на видовете възниква посредством адаптивно и случайно разцепване (раздвояване) на потомствените линии (родословие)**.
- Потомствена линия (lineage) - последователност от видове, всеки от които се счита, че е еволюирал от своя предшественик (напр., "the chimpanzee and gorilla lineages").
- С течение на времето, класификацията на видовете е станала по-малко статична и по-динамична.

АВТОМАТИЧНО ГЕНЕРИРАНО ДЪРВО НА ЖИВОТА, НА ОСНОВАТА НА НАПЪЛНО СЕКВЕНИРАНИ ГЕНОМИ

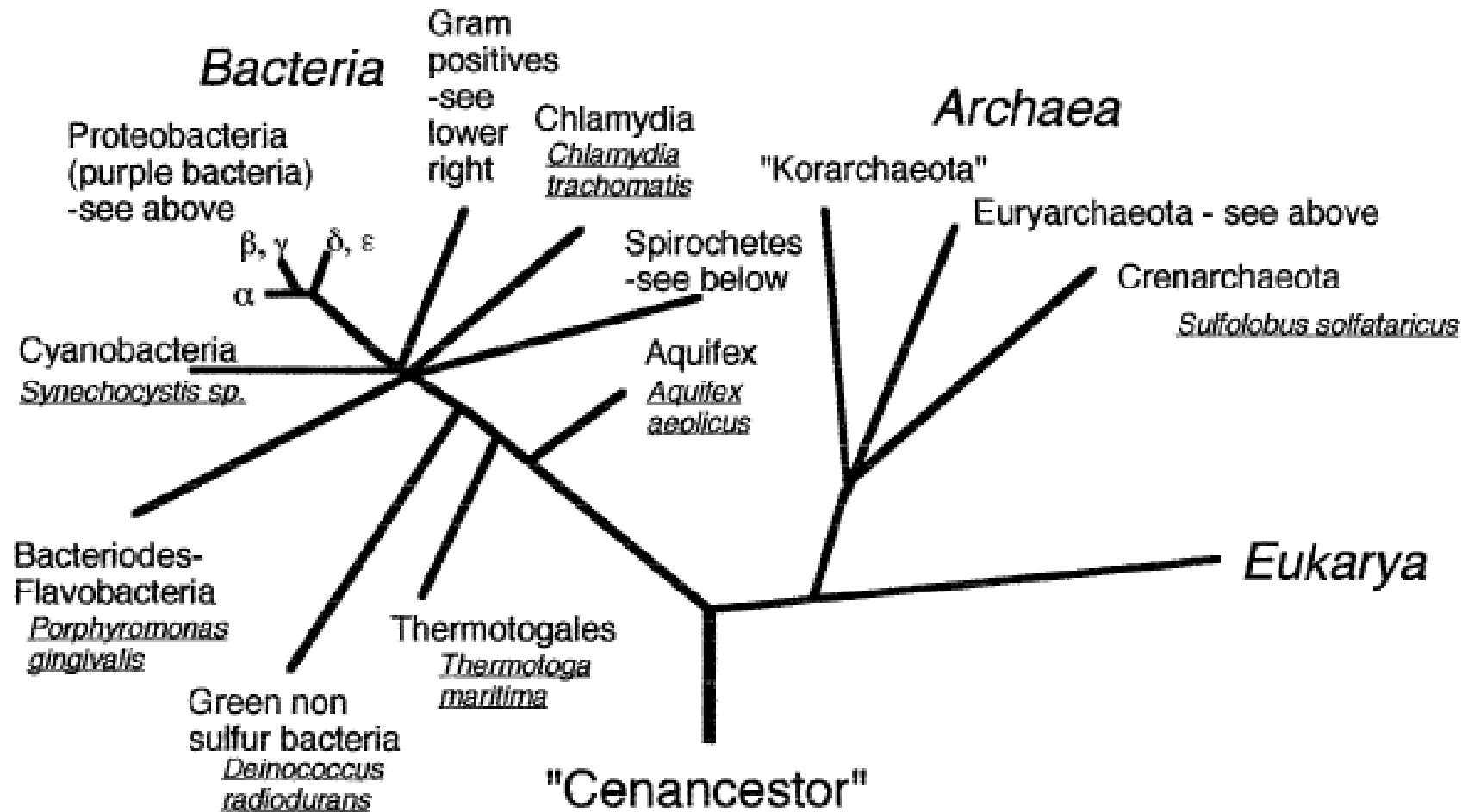


ФИЛОГЕНЕТИЧНО ДЪРВО (A PHYLOGENETIC TREE)

- **Филогенетичното дърво** или **еволюционното дърво** представлява разклонена диаграма или “дърво” показващо еволюционните отношения между различни биологични видове или други субекти въз основа на сходствата и различията в техните физически и / или генетични характеристики
- Таксоните, свързани помежду си в рамките на дървото, се предполага, че са произлезли от общ прародител
- При филогенетичното дърво с корен, всеки възел с потомци представя последния общ предшественик на потомците, и дължините на клоните в някои от дърветата, може да се интерпретират като оценки на еволюционното време.
- **Таксономията е клон на биологията, занимаващ се с идентифициране и обозначаване на организми**
- Всеки възел се нарича таксономична единица.
- В общия случай вътрешните възли се наричат хипотетични таксономични единици (hypothetical taxonomic units (HTUs)) тъй като не могат да бъдат наблюдавани директно.
- Дърветата са полезни в области на биологията като таксономия и сравнителна филогенетика.

PHYLOGENETIC TREE OF LIFE

ФИЛОГЕНЕТИЧНОТО ДЪРВО НА ЖИВОТА



ТИПОВЕ ФИЛОГЕНЕТИЧНИ ДЪРВЕТА

- *Дървета без корен – некоренени дървета (Unrooted trees)* илюстрират свързаността на листата без изобщо да се правят предположения относно произхода.
- Докато дърветата без корен винаги могат да бъдат генерирани от дърветата с корен като просто се пропусне корена, обратното не е възможно без да се използват средства за проследяване на потеклото (родословието);
- Тези средства обикновено предполагат включване на извън-груповите взаимоотношения във входните данни или въвеждане на допълнителни допускания за относителната скорост на еволюцията за всеки клон, като приложение на хипотезата за молекулярния часовник.

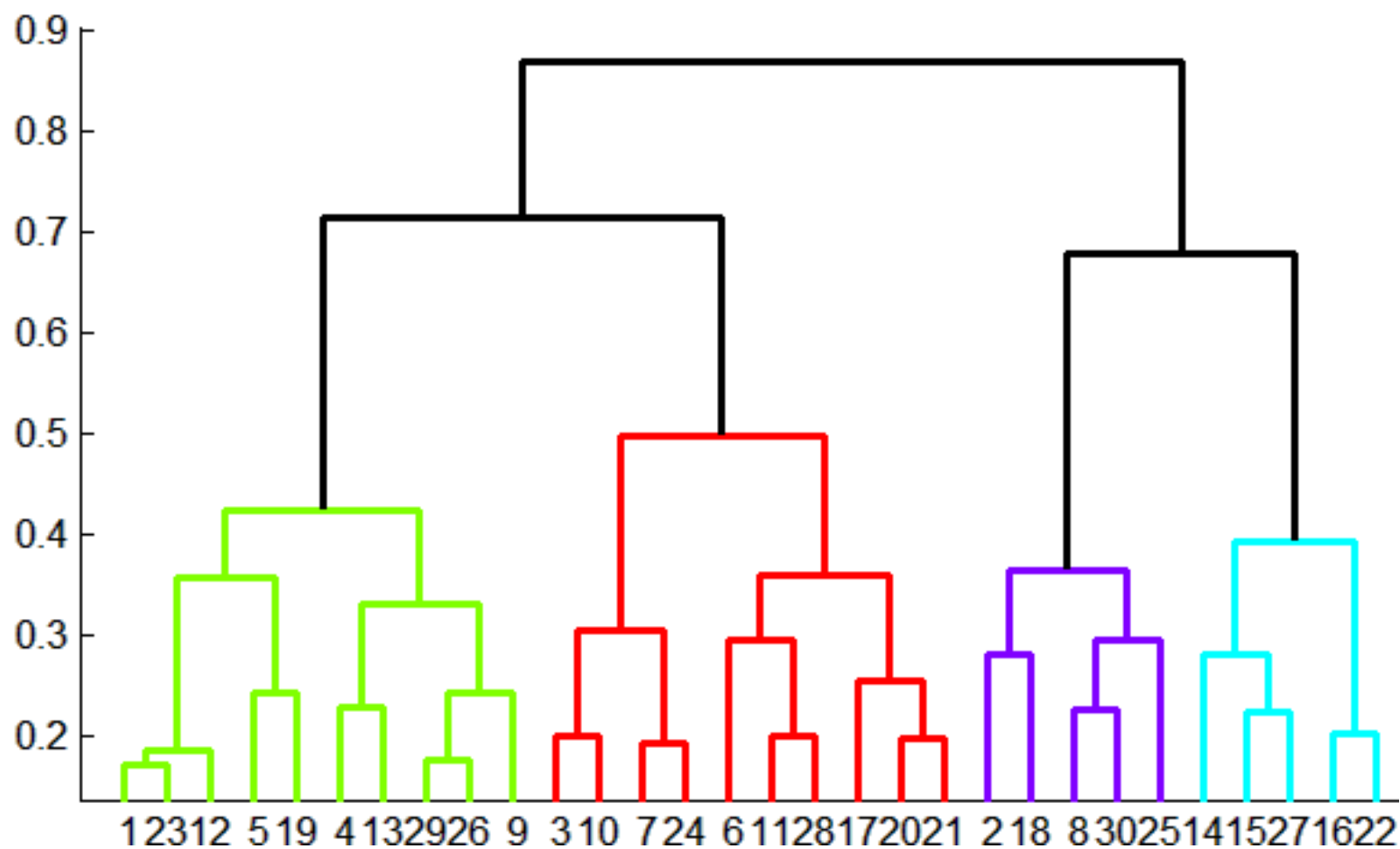
ФИЛОГЕНЕТИЧНИ ДЪРВЕТА

- Както вкоренените, така и некоренените филогенетични дървета могат да бъдат разклонени на две (чатал) – двоични дървета, или с много разклонения, също така могат да бъдат етикетирани или неетикетирани (без етикети).
- Вкорененото двоично дърво (чатал) има точно два потомъка, произтичащи от всеки вътрешен възел.
- Некорененото разклонено на две дърво приема формата на некоренено двоично дърво, свободно дърво с точно три съседа във всеки вътрешен възел.
- За разлика от това, вкорененото разклонено дърво може да има повече от два потомъка при някои от възлите, както и некорененото разклонено дърво може да има повече от три съседа при някои от възлите.
- При етиктираните дървета листата са маркирани със специфични стойности.
- Неетиктираните дървета, понякога наричани дървовидни форми, дефинират само топология.

ОСНОВНИ ОПРЕДЕЛЕНИЯ

- **Дендограмата** е широко понятие за представянето на филогенетичното дърво под формата на диаграма.
- Терминът „дендограма“ има гръцки произход: *dendron* - „дърво“, *gramma* - „чертеж“
- Дендограмата е дървовидна диаграма, често използвана за илюстриране на подравняването на клъстерите, получени посредством йерархично клъстериране.
- Дендограмите често се използват в изчислителната биология да илюстрират групирането на гени или образци.

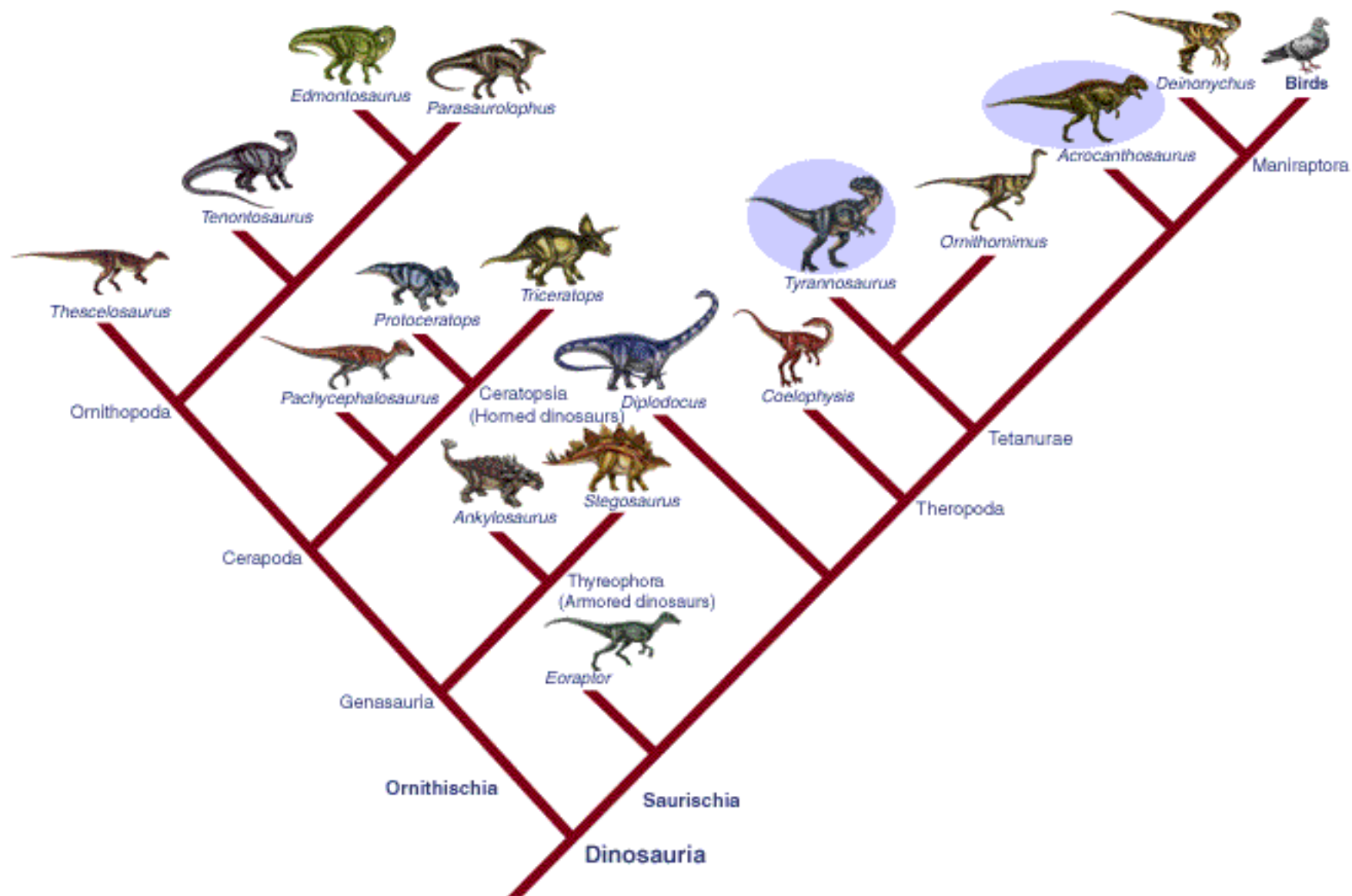
ДЕНДОГРАМА



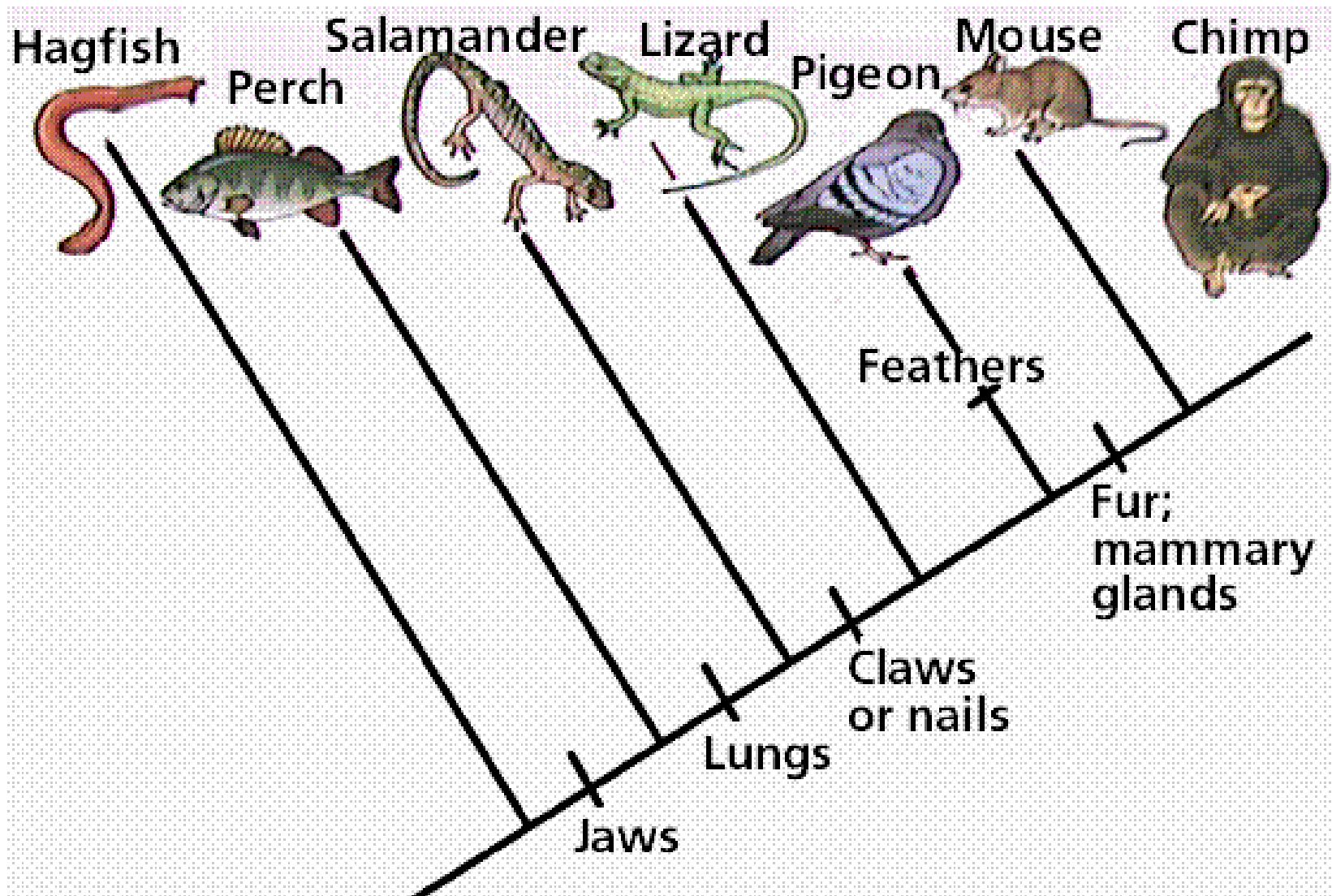
- *Филогенезата* (филогения) е еволюционното развитие на видовете.
- В биологията, *Филогения* е изучаването на еволюционните връзки между различни групи организми, по-голямата част от чиито потомствени организми са отдавна изчезнали
- *Кладистиката* е метод за класификация на организмите въз основа на техните филогенетични връзки.
- Клонът (**clade**) обхваща даден организъм и всичките му потомци
- *Клонът* (още клад на старогръцки: κλάδος клон) е група от организми, състояща се от един вид (изчезнал или съществуващ) и всички произлезли от него видове. Идеята е, че това е „естествена еволюционна група“ от организми, които трябва да бъдат групирани заедно.
- *Кладограмата* представлява дърво, формирано с използване на кладистични методи. Кладограмите представят само модели на разклоняването, при което дължината на клоните не представят еволюционно време.
- Например, птиците, динозаврите, крокодилите, и всичките потомци (живи или изчезнали) на техния последен прародител, формират клон.
- Software for phylogeny and cladistics –

<http://www.crescentbloom.com/ii/l/17.htm>

КЛАДОГРАМА



КЛАДОГРАМА



PHYLOGRAMS & CHRONOGRAMS

ФИЛОГРАМИ И ХРОНОГРАМИ

- Филограмата (**phylogram**) е филогенетично дърво, което в явен вид представя броя на промените на символите посредством дължината на клоните.
- Хронограмата (**chronogram**) е филогенетично дърво, което в явен вид представя еволюционното време посредством дължината на клоните

ОГРАНИЧЕНИЯ

- Въпреки, че филогенетичните дървета, конструирани на основата на секвенирани гени или геномни данни на различни биологични видове, могат да подпомогнат вникване в еволюционното развитие, те имат важни ограничения.
- Те не винаги отразяват акуратно еволюционната история на видовете.
- Данните, на които те се базират, са “зашумени” (“noisy”); евентуалните причини могат да бъдат хоризонталният трансфер на гени, хибридизацията на видове, които не са били най-близки съседи в дървото, конвергентна еволюция, и консервативни секвенции.

HORIZONTAL GENE TRANSFER (HGT)

ХОРИЗОНТАЛЕН ТРАНСФЕР НА ГЕНИ

- *Хоризонтален трансфер на гени (HGT)* е произволен процес, при който даден организъм включва генетичен материал от друг организъм, без да е потомък на този организъм.
- От друга страна, *вертикален трансфер на гени (vertical gene transfer)* е процес, когато един организъм получава генетичен материал от свой предшественик (прародител), например, неговата майка или видове, от които се е развил.
- Изследванията в генетиката са фокусирани предимно върху вертикалния трансфер на гени, но все повече се засилва убеждението, че хоризонталният пренос на гени е значимо явление и сред едноклетъчните организми е може би най-основната форма на генетичен трансфер.
- Изкуственият хоризонтален пренос на гени е форма на генетично инженерство.

ФИЛОГЕНЕТИЧНО ПРЕДСКАЗВАНЕ

- Филогенетичният анализ на семейство от свързани нуклеинови киселини или протеинови секвенции представлява разкриване на начина, по който се е развивало това семейство в хода на еволюцията
- Еволюционните взаимоотношения между секвенциите се представят посредством поставяне на секвенциите като външни клони на дървото.
- Разклоненията във вътрешната част на филогенетичното дърво отразяват степента на свързаност на секвенциите.
- Две много сходни секвенции се разполагат на съседни външни клони и се свързват с общ клон помежду им.
- Целта на филогенетичния анализ е да се открият всички взаимоотношения на разклоняване в дървото, както и дължината на клоните.

MEGA – MOLECULAR EVOLUTIONARY GENETIC ANALYSIS SOFTWARE

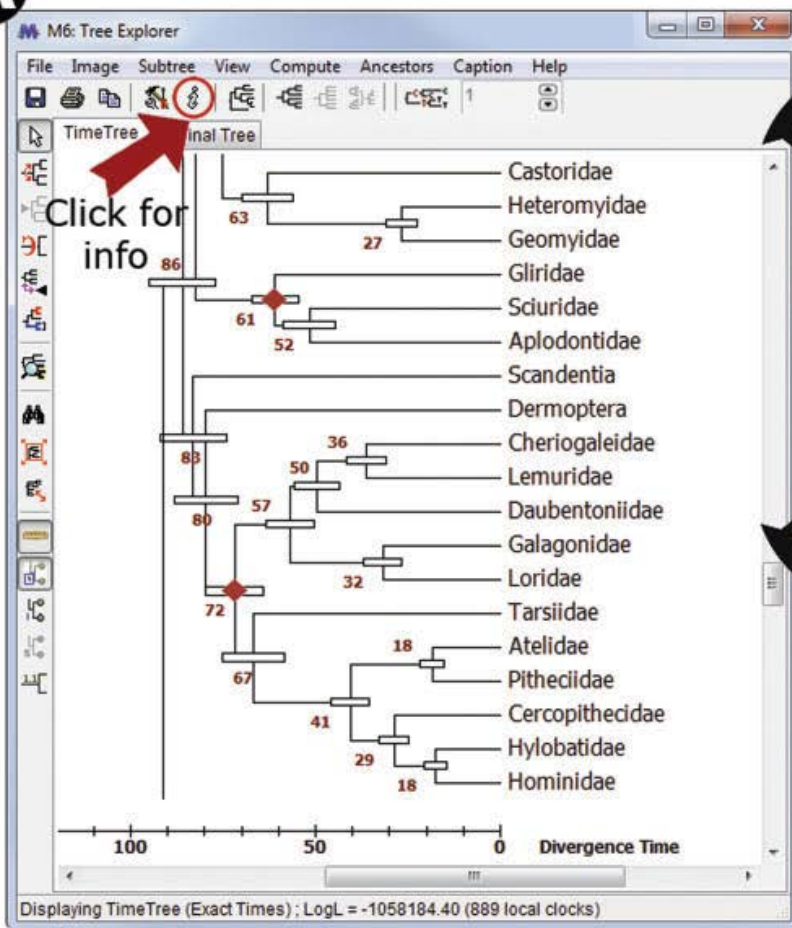
- MEGA – софтуер за анализ на молекулярна еволюционна генетика
- Свободно достъпен софтуер за генериране на дендограми, или филогенетични дървета, използвайки нуклеотидни или протеинови секвенции.
- Софтуерът MEGA е създаден в Pennsylvania State University

<http://www.megasoftware.net/>

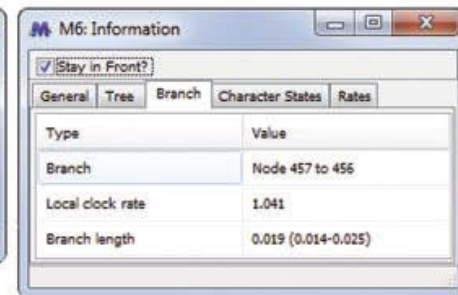
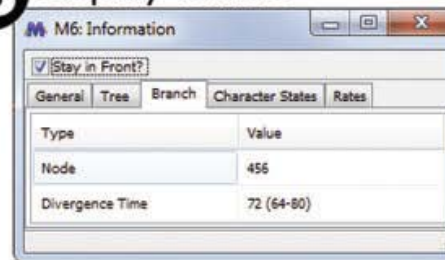
- MEGA е интегриран софтуерен инструмент за автоматизирано и ръчно подравняване на секвенции, генериране на филогенетични дървета, търсене в уеб-базирани биологични бази данни, оценка на темповете на молекулярната еволюция, генериране на секвенциите на предците, както и за тестване на еволюционната хипотеза.
- MEGA е многонишково приложение за Windows.

MEGA

- Софтуерният пакет MEGA (на език за програмиране C++) е разработен с цел оценка на еволюционната дистанция, реконструкция на филогенетични дървета и генериране на статистики на база молекулярни данни.
- Софтуерът имплементира различни методи за оценка на еволюционните дистанции на нуклеотидни и протеинови секвенции, 3 различни метода за генериране на филогенетични дървета (UPGMA, метод „свързване със съсед“ и метода на максималната пестеливост), както и два статистически теста за топологични разлики.
- За метода на максималната пестеливост са имплементирани нови алгоритми за комбинаторно търсене по метода на клоните и границите, както и за евристично търсене.
- В допълнение, MEGA изчислява статистически стойности като честоти на присъствие на нуклеотиди и аминокислини в секвенциите, отклонения при преходи и пресичания, статистическа честота на кодоните (codon usage tables), както и броя на вариабилните сайтове в специфицирани сегменти от нуклеотидни и протеинови секвенции.

A**B**

Display Panels

**C**

Download Results

	A	B	C	D	E	F	G	H	I	J
1	NodeLabel	NodeId	Des1	Des2	RelTime	StdErr	Rate	DivTime	CI_Lower	CI_Upper
440	Scylliorhinidae1	439	-	-	-	-	0.350	-	-	-
441	Proscylliidae	440	-	-	-	-	1.432	-	-	-
442	Pseudotriakidae	441	-	-	-	-	1.252	-	-	-
443	Leptochariidae	442	-	-	-	-	1.388	-	-	-
444	Triakidae2	443	-	-	-	-	0.261	-	-	-
445	Triakidae1	444	-	-	-	-	0.856	-	-	-
446	Hemigaleidae	445	-	-	-	-	1.345	-	-	-
447	Carcharhinidae	446	-	-	-	-	1.124	-	-	-
448	myNodeLabel1	447	195	196	0.058	0.008	0.835	36.9	26.3	47.5
449	-	448	447	194	0.074	0.006	0.564	47.3	39.9	54.7
450	myNodeLabel2	449	192	193	0.055	0.008	1.023	34.8	24.5	45.1
451	-	450	448	449	0.078	0.005	0.794	49.9	43.5	56.3

MEGA

A**Upload Alignment + Topology**

Data is automatically validated

Specify Time Constraints

(Optional)

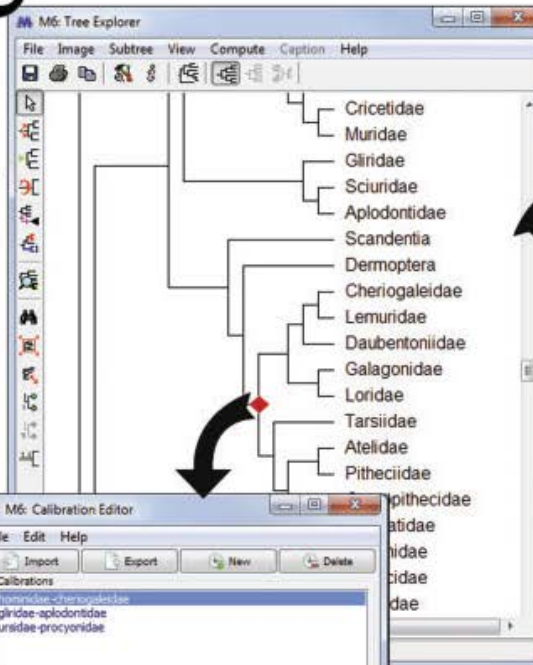
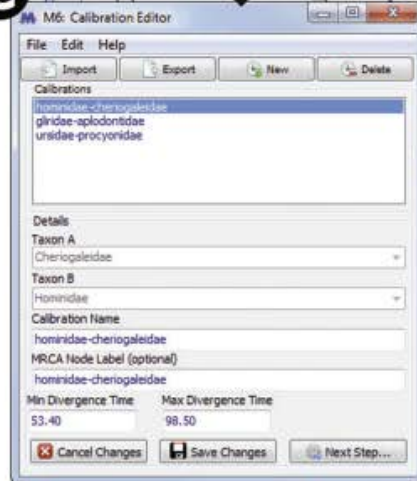
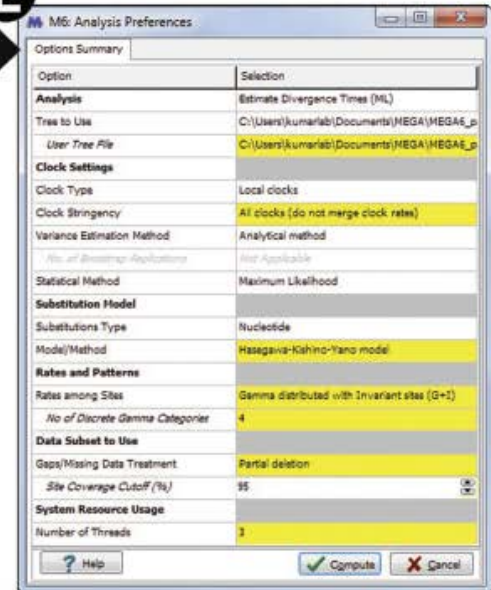
Select Analysis Preferences

Clock settings
Substitution model
Rates and Patterns
Gaps/Missing Data

Launch Analysis

Relative times are computed

If applicable, constrained nodes
are anchored and times are
recomputed

B**C****E****D**

```
File Edit Format View Help
!MRCA='hom-cher' TaxonA='Cherogaleidae' TaxonB='Hominidae'
!MRCA='glir-aplod' TaxonA='Gliridae' TaxonB='Apodontidae'
!MRCA='ursid-proc' TaxonA='Procyonidae' TaxonB='Ursidae'
!NodeName='user-defined name' M1Time=1.75 MaxTime=2.25;
```

TXT

**View the Timetree in
Tree Explorer**

PHYLIP (PHYLOGENY INFERENCE PACKAGE)

- Свободно достъпен софтуерен пакет за филогенетика за конструиране на еволюционни дървета.
- Обхваща 35 преносими програми (сорс кода е на C), налични са версии за системи под Windows и Linux
- Авторът на софтуерния пакет е професор Joseph Felsenstein, Department of Genome Sciences and the Department of Biology, University of Washington, Seattle



МЕТОДИ ЗА ОПРЕДЕЛЯНЕ НА ЕВОЛЮЦИОННИ ДИСТАНЦИИ

- Тези методи предполагат изчисляване на броя на промените между всяка двойка бази в група от секвенции с цел генериране на филогенетично дърво на групата
- Двойки бази в секвенциите, които имат минимален брой промени помежду си, се наричат „съседни“ (neighbors).
- В рамките на дървото, тези секвенции споделят общ възел или обща позиция за прародител, като се свързват към този общ възел посредством клон.
- Целта на тези методи е да се генерира дърво, в което „съседите“ са позиционирани коректно и също така, да се формират дължините на клоните по такъв начин, че да отразяват по възможност най-точно наличните входни данни.
- Откриването на най-близките съседи в рамките на група от секвенции с прилагането на методите за дистанцията е обикновено първата стъпка при множественото подравняване на секвенции

МЕТОДИ ЗА ОПРЕДЕЛЯНЕ НА ЕВОЛЮЦИОННИ ДИСТАНЦИИ

- Методът на дистанцията за пръв път е имплементиран от Feng и Doolittle, и колекция програми от тези автори правят подравняване и генерират дърво за множество от протеинови секвенции
- Софтуерът за множествено подравняване на секвенции CLUSTALW използва метода за дистанции от вида свързване на съседите за направляване на подравняването.
- RAUP version 4 има опции за филогенетичен анализ на основата на методите за дистанции.
- Програмите на пакета PHYLIP осъществяват анализ на дистанциите, като автоматично въвеждат секвенциите във PHYLIP infile format и автоматично генерират файл outfile, съдържащ таблица на дистанциите.

МЕТОДИ ЗА ОПРЕДЕЛЯНЕ НА ЕВОЛЮЦИОННИ ДИСТАНЦИИ

- **DNADIST** изчислява дистанциите между входни секвенции от аминокиселини. При него може да се направи избор от различни модели на еволюцията.
- **PROTDIST** изчислява мярка за дистанция за протеиновы секвенции, на основата на модела Dayhoff PAM или други модели на еволюционните промени на протеините.

МЕТОДИ ЗА ОПРЕДЕЛЯНЕ НА ЕВОЛЮЦИОННИ ДИСТАНЦИИ

- При подравняването на секвенции, обикновено се изчислява оценката за сходство
- При филогенетичния анализ се използва оценката за дистанция между две секвенции
- Тази оценка обхваща броя на различните символи в еднакви позиции при подравняването или броя на позициите в едната секвенция, които трябва да бъдат променени за да получим втората секвенция.
- Празните позиции се игнорират или може да се третираат като замествания.

ГЕНЕРИРАНЕ НА ФИЛОГЕНЕТИЧНО ДЪРВО - ПРИМЕР

A. Sequences

sequence A A C G C G T T G G G C G A T G G C A A C
sequence B A C G C G T T G G G C G A C G G T A A T
sequence C A C G C A T T G A A T G A T G A T A A T
sequence D A C A C A T T G A G T G A T A A T A A T

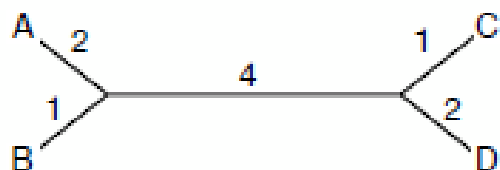
B. Distances between sequences, the number of steps required to change one sequence into the other.

n_{AB} 3
 n_{AC} 7
 n_{AD} 8
 n_{BC} 6
 n_{BD} 7
 n_{CD} 3

C. Distance table

	A	B	C	D
A	–	3	7	8
B	–	–	6	7
C	–	–	–	3
D	–	–	–	–

D. The assumed phylogenetic tree for the sequences A-D showing branch lengths. The sum of the branch lengths between any two sequences on the trees has the same value as the distance between the sequences.



MULTIPLE SEQUENCE ALIGNMENT

ACGCGTTGGGCGATGGCAAC	SEQUENCE A
ACGCGTTGGGCGACGGTAAT	SEQUENCE B
ACGCATTGAATGATGATAAT	SEQUENCE C
ACACATTGAGTGATAATAAT	SEQUENCE D

AC - C - TTG - - - GA - - - AA -

ACGCGTTGGGCGATGGCAAC
 ACGCGTTGGGCGACGGTAAT
 ACGCATTGAATGATGATAAT
 ACACATTGAGTGATAATAAT

SEQUENCE A

SEQUENCE B

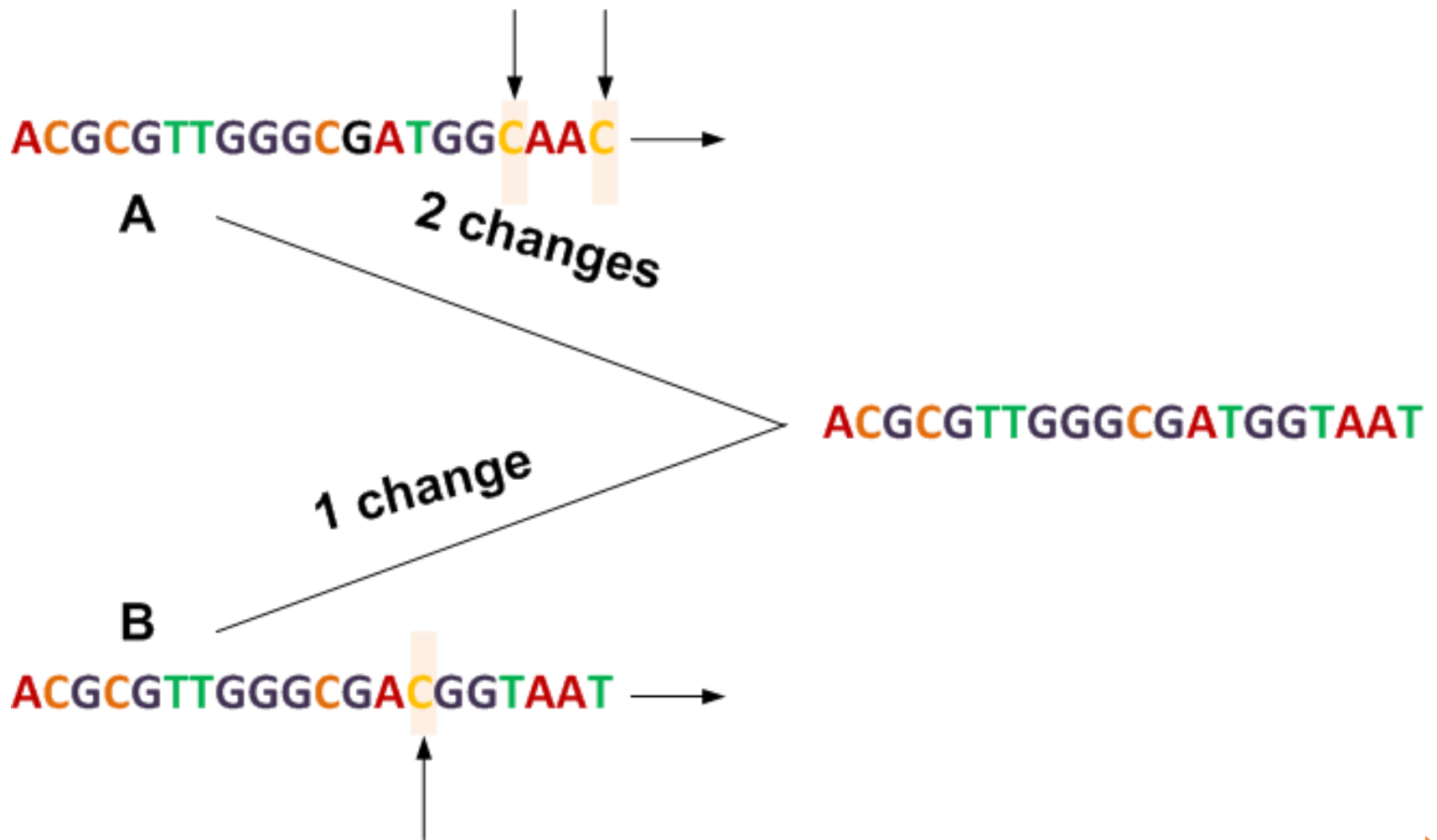
SEQUENCE C

SEQUENCE D

ACGCGTTGGGCGATGGCAAC
 ACGCGTTGGGCGACGGTAAT

SEQUENCE A

SEQUENCE B



ACGCGTTGGGCGATGGCAAC
 ACGCGTTGGGCGACGGTAAT
 ACGCATTGAATGATGATAAT
 ACACATTGAGTGATAATAAT

SEQUENCE A

SEQUENCE B

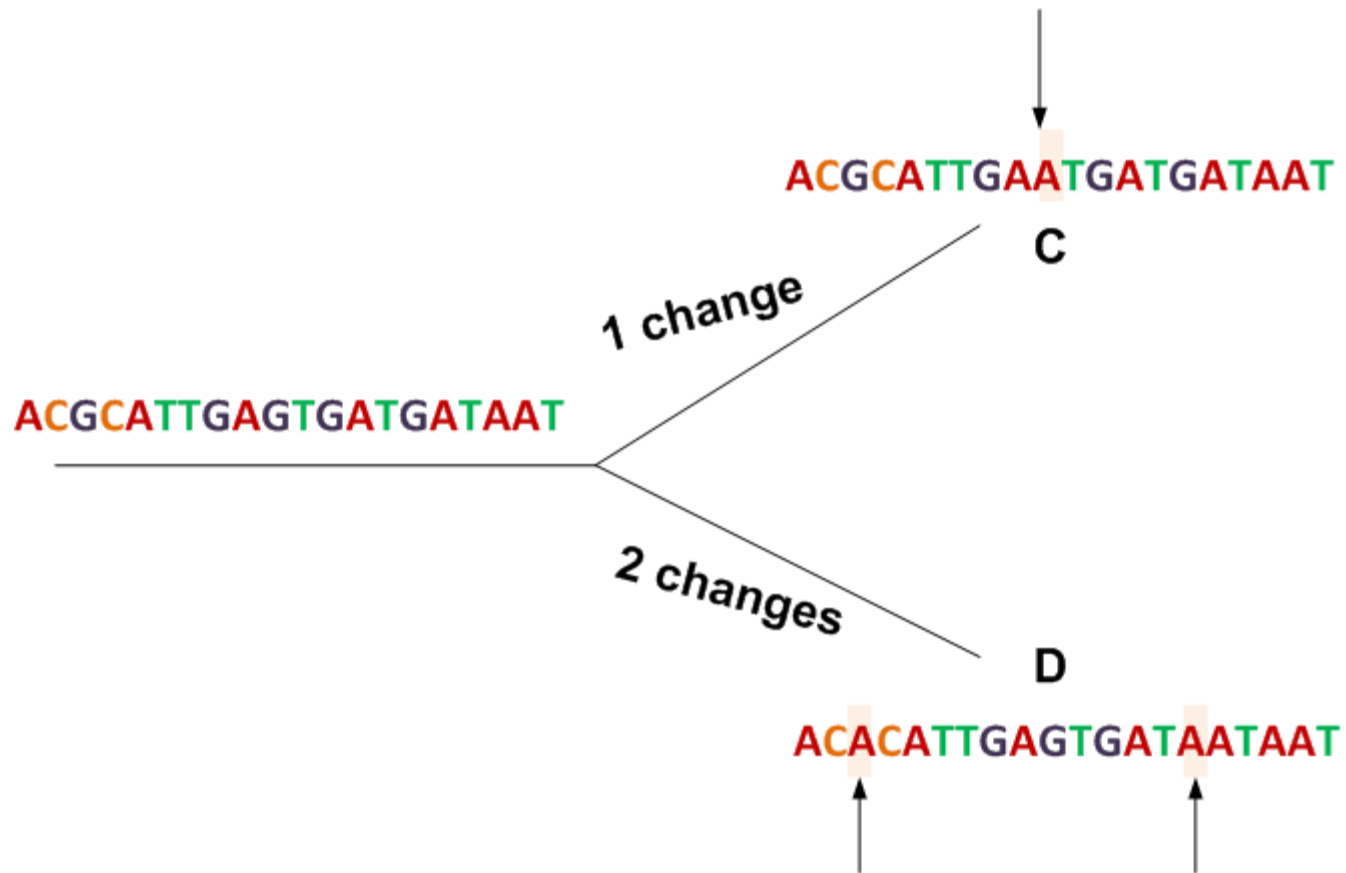
SEQUENCE C

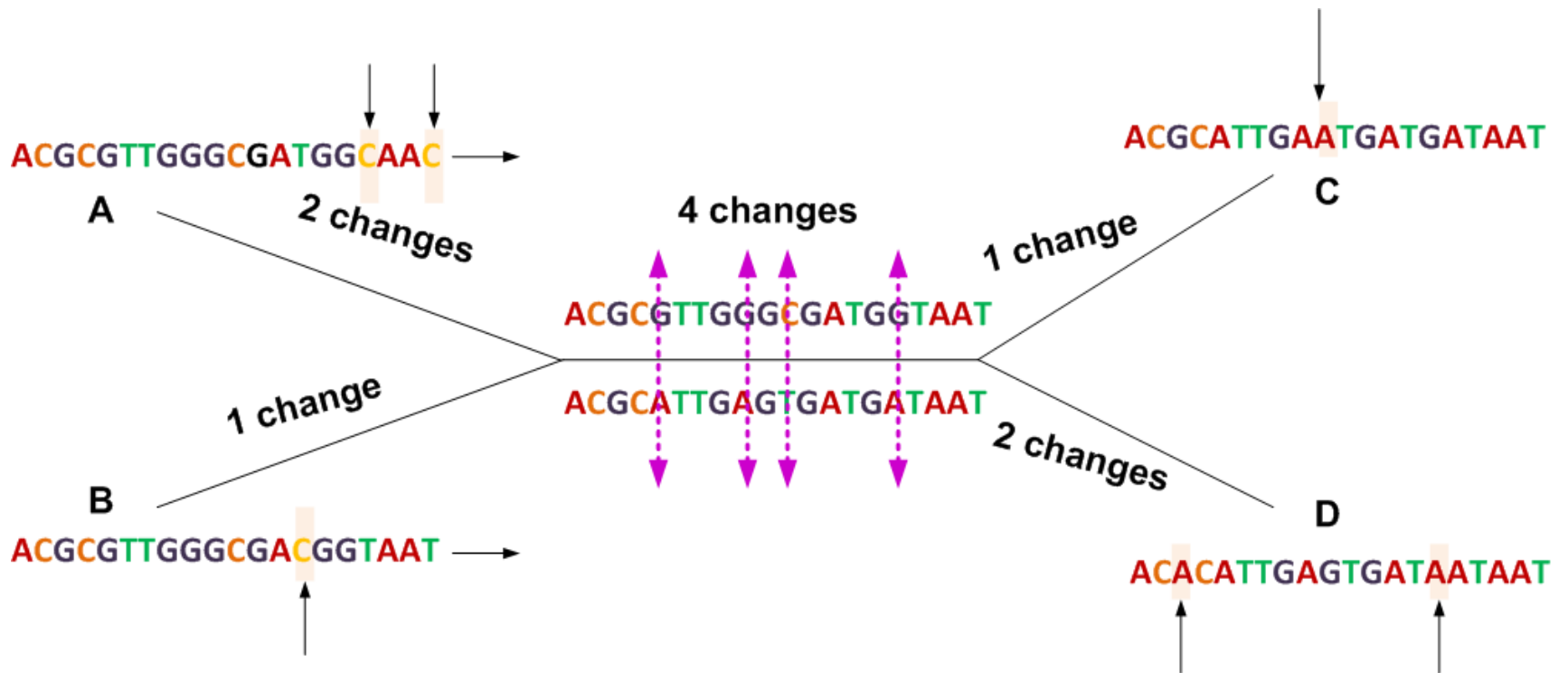
SEQUENCE D

ACGCATTGAATGATGATAAT
 ACACATTGAGTGATAATAAT

SEQUENCE C

SEQUENCE D





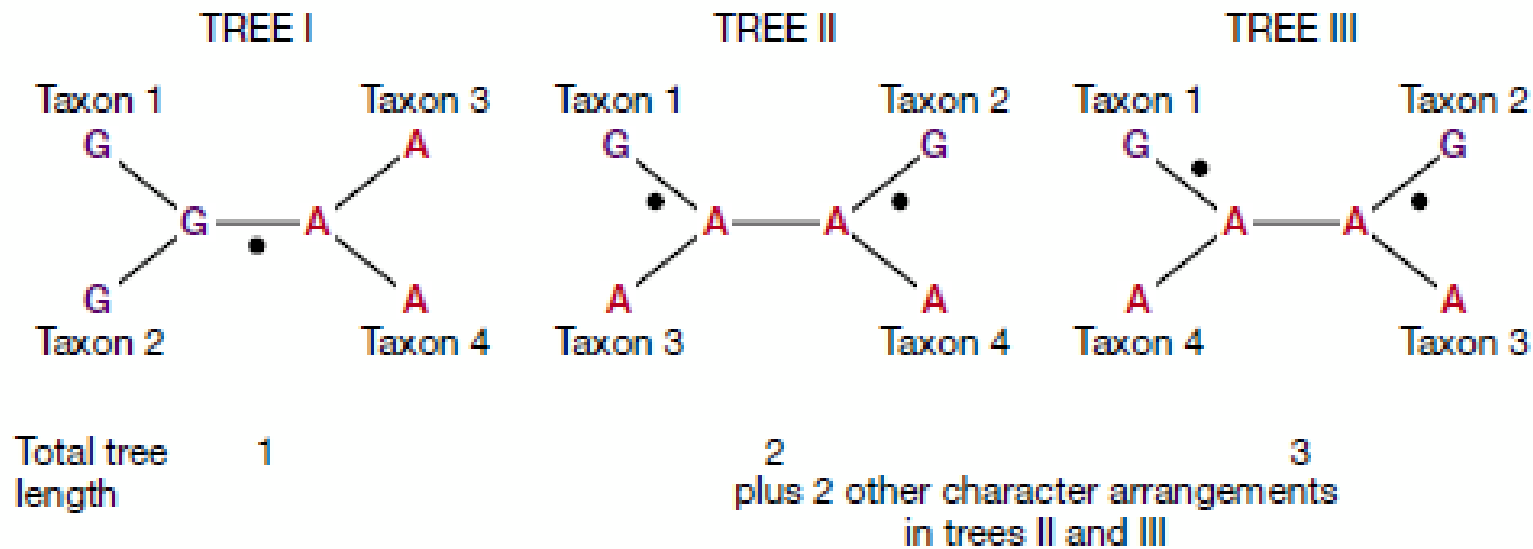
MAXIMUM PARSIMONY METHOD

МЕТОД НА МАКСИМАЛНАТА ПЕСТЕЛИВОСТ

- Този метод дава възможност да се изгради еволюционно дърво (или дървета), което минимизира броя на стъпките, необходими да се генерират необходимите промени в секвенциите
- За това, този метод се нарича *метод на минималната еволюция* (the minimum evolution method).
- Необходимо е да се извърши множествено подравняване на секвенции за предсказване на позициите в секвенциите, които има вероятност да си съответстват.
- Тези позиции се появяват във вертикалните колони при множественото подравняване.
- За всяка подредена позиция се генерират филогенетични дървета, които изискват най-малкия брой еволюционни промени.

ФИЛОГЕНЕТИЧЕН АНАЛИЗ ЗА НАМИРАНЕ НА НЕВКОРЕНЕНО ДЪРВО ОТ 4 ПОДРЕДЕНИ СЕКВЕНЦИИ ПО МЕТОДА НА МАКСИМАЛНАТА ПЕСТЕЛИВОСТ

Taxa	Sequence position (sites) and character								
	1	2	3	4	5	6	7	8	9
1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G



Трите възможни дървета

ФИЛОГЕНЕТИЧЕН АНАЛИЗ ПО МЕТОДА НА МАКСИМАЛНАТА ПЕСТЕЛИВОСТ

- Този метод намира еволюционното дърво, което променя всяка секвенция до вида на останалите с минимален брой стъпки.
- Правила за анализ по метода на максималната пестеливост:
 1. Зададени са 4 таксона, които водят до 3 възможни невкоренени дървета.
 2. Някои сайтове са информативни, i.e., те препоръчват едно дърво пред друго (сайт 5 е информативен, но сайтове 1, 6, и 8 не са).
 3. За да бъде информативен, сайтът трябва да има еднакви символи в поне два таксона (сайтове 1, 2, 3, 4, 6, и 8 не са информативни; сайтове 5, 7, и 9 са информативни).
 4. На анализ се подлагат само информативните сайтове.

ФИЛОГЕНЕТИЧЕН АНАЛИЗ ЗА НАМИРАНЕ НА НЕВКОРЕНЕНО ДЪРВО ОТ 4 ПОДРЕДЕНИ СЕКВЕНЦИИ ПО МЕТОДА НА МАКСИМАЛНАТА ПЕСТЕЛИВОСТ

- В примера разглеждаме само 4 секвенции, така че разглеждаме само 3 невякоренени дървета
- Оптималното дърво се получава, като се прибавят броя на промените на всеки информативен сайт за всяко дърво
- Избира се дървото, при което се изискват най-малък брой промени