



# FILOGENETIKA BEVEZETŐ

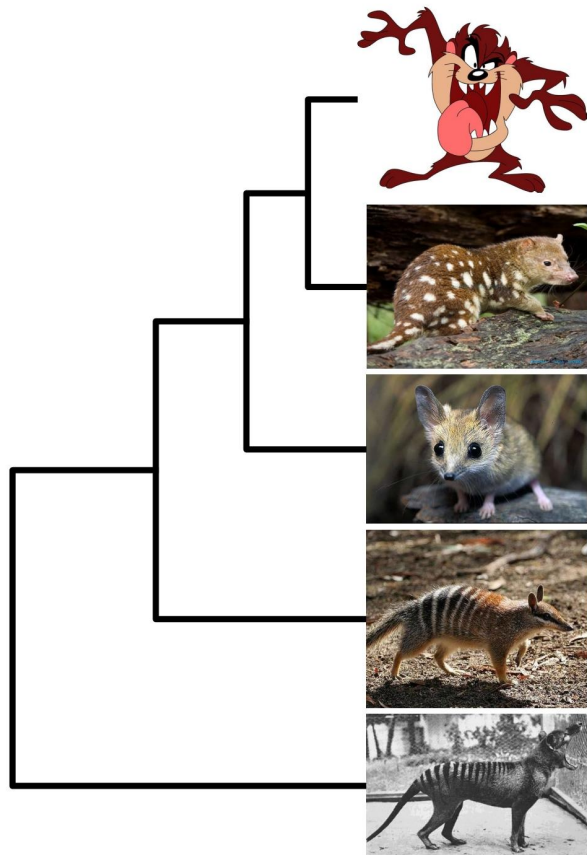
---

ELŐADÓ: GERBER DÁNIEL  
gerberd1990@gmail.com

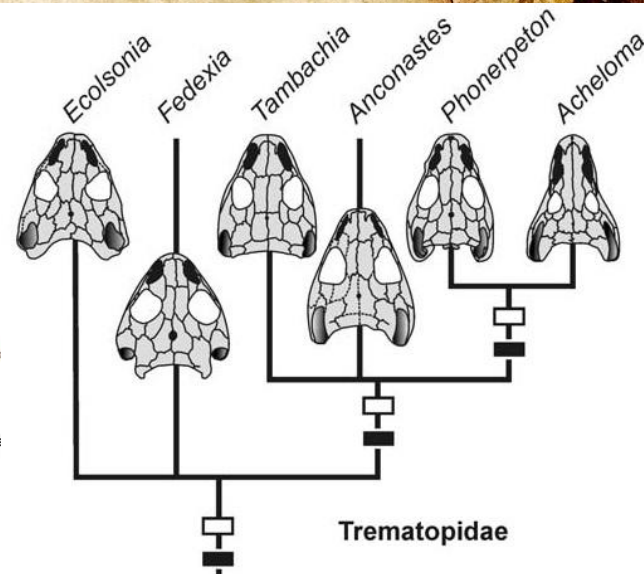
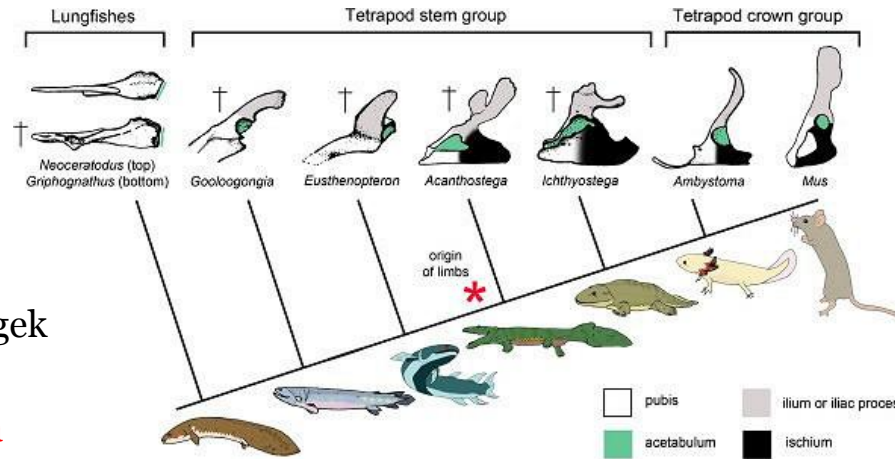


# MI AZ A FILOGENETIKA?

- Evolúciós leszármazási (alias rokonsági) kapcsolatok feltárását, vizsgálatát célzó tudományterület
- A vizsgálatot klasszikusan egymással közelebbi-távolabbi rokonságban lévő fajok között szokták elvégezni, de fajon belüli kapcsolatok tisztázása sem ritka (csak sokkal nehezebb)
- Egyszerű(bb) és bonyolult matematikai modelleket alkalmaznak a fajok közötti rokonsági és/vagy evolúciós különbségek és hasonlóságok feltérképezésére
- A számításoknál használt változók a vizsgált fajok/egyedek lényegében bármely mérhető tulajdonsága közül kikerülhetnek, a lényeg, hogy össze lehessen ezeket hasonlítani (ergo minden vizsgált mintában legyenek közös tulajdonságok\*)
- Számos alkalmazási területe van a rendszertanon felül, geográfiai elterjedés modellezésétől evolúciós trendek feltérképezésén keresztül igazságügyi felhasználásig



# MILYEN VÁLTOZÓKAT HASZNÁLUNK?



Klasszikusan:

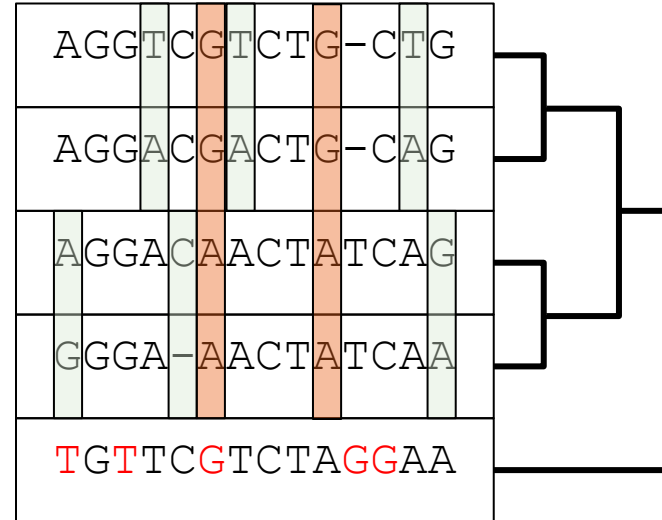
- 1) Anatómiai jellegek (kladisztika)
- 2) DNS szekvencia
- 3) Fehérje szekvencia

DAN	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	A5ASC3.1	14	SIKLWPPSQTRLLLVERMANNLST..PSIFTRK..YGLSKEEAARENAKQIEEVACSTANQ....HYEKEPDGDGSSAVQLYAKECSKLIIEV
BRAHNI	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	B4F917.1	13	SIKLWPPSESTRIMLVDRMTNNSLST..ESIFSRK..YRLGKQEAHENAKTIEELCFALADE....HFREEPDGDGSSAVQLYAKESKMLLEV
BAL0CHI01	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	A9S1V2.1	23	VFKLWPPSQGTREAVRQKMKALKLSS..ACFESSQ..FARIELADAQEHARAIIEEVAFGAQE....ADSGDKTGSAVMMVYAKHASKLMLET
BAL0CHI02	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	B9GSN7.1	13	SVKLWPPGSTRMLVERMTKNFIT..PSFISRK..YGLSKEEAEDAKKIEEVAFAAANQ....HYEKQPDGDGSSAVQIYAKESSRLMLEV
MAKRANI02	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q8H056.1	30	SFSIWPPPTQRTDRAVVRRLVDTLGG..DTLILCKR..YGAVPADAEPAARGIEAEAFDAAAA..SGEAAPATASVEEIKALQLYSKEVSRRLDF
MAKRANI01	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q0D4Z3.2	44	SLSIWPPSQTRDAVVRRLVQTLVA..PSILSKR..YGAVPEAEAGRAAAVEAEAYAAVTES..SSAAAAPASVEDGIEVLQAYSKEVSRRLLEL
UYG01	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	B9MW8.1	56	SFSIWPPPTQRTDRAIISRLIETLST..TSVLSKR..YGTIPKEEASEASRIIEEAFSGAST....VASSEKDGLEVLQLYSKEISKRMLLET
PAK	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q0IYC5.1	29	SFAVWPPTRTRDRAVVRRLVAVLGGDTTIALRKYRY..YGAVPADAEARAARAVEAQAFAAASA....SSSSSSVSDGIEITLQLYSREVSNRLLAF
CHIN	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	A9NW46.1	13	SIKLWPPSESTRMLVERMTDNLSLSS..SVFFSRK..YGLLSKEEAARENAKRIIEEAFLAAND....HEAKEPNLDOSVVOFYAREASKLMLEA
GEORG	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q9C500.1	57	SLRIWPPPTQKTRDAVNLRLIETLST..ESILSKR..YGTLLKSDDATTVAKLIEEAYGVASN....AVSDDDDGKILELYSKEISKRMLLES
UYG02	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q2HRI7.1	25	NYSIWPPKQRTDRAVKRNLRIETLST..PSVLTKR..YGTMSADEASAAAIQIEDEAFSVANA....SSSTNDNVITILEVYSKEISKRMIET
UYG03	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q9M7N3.1	28	SFKIWPPPTQRTREAVRRLVETLTS..QSVLSKR..YGVIPEDDATSAARIIEEAFSVASV..ASAASTGGRPEDEJIEVLHIYSQEIQRVYES
UYG04	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q9M7N6.1	25	SFSIWPPPTQRTDRAVINRLIESLST..PSILSKR..YGTLPQDEASETARLIEEAFAAAGS....TASDADGIEILQVYSKEISKRMIET
IRN_NE001	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q9LE82.1	14	SVKMWPPSKSTRMLVERMTKNITTT..PSIFSRK..YGLLSVEEAEDQAKRIEDLAFATANK....HFQNEPDGDTSAHVYAKESSKLMIDV
IRN_NE002	NATCANAGGT	NTATCACCT	ATTAACCACT	CACGGGAGCT	CCTCCATGCAT	Q9M651.2	13	SIKLWPPSLPTKRALIERITNNFSS..KTIFTEK..YGLSLTKDQATENAKRIEDIAFSTANQ....QFEREPDGDGSSAVQLYAKECSKLIIEV
						B9R748.1	48	SIKLWPPPTQRTDRAVITRLIETLSS..PSVLSKR..YGTISHDEAESAARRIIEDEAFVANT....ATSAEDDQIEITLQLYSKEISRRMLD



# MOLEKULÁRIS FILOGENETIKA

- (Többszörösen) Illesztett szekvenciákkal indul (ld. előző óra)
- Alapvetően a szekvenciák közötti különbségek és hasonlóságok adják meg két minta egymástól való filogenetikai távolságát, ezzel együtt a köztük lévő evolúciós (genetikai) távolságot, és a vizsgálatba bevont minták egymáshoz viszonyított rokonsági kapcsolatait
- Minél nagyobb eltérések vannak a genetikai (v. aminosav) szekvenciák között, annál nagyobb a két minta közötti evolúciós távolság. Mondhatnánk... de a helyzet ennél sokkal bonyolultabb.
- Okosok kitalálták és a gép kiszámolja nekünk, akkor miért kell tudnunk mi folyik a háttérben?



**A gép azzal számol amit adunk neki, és ha nem tudjuk, jó-e, amit neki adunk, vagy nem tudjuk, mit és hogyan kell a gépnek számolnia (mert a gép se tudja magától, miért tudná?), akkor rosszul számol, rossz eredményt kapunk, jön a világvége.**

# FILOGENETIKAI ELEMZÉST BEFOLYÁSOLÓ TÉNYEZŐK I

Főbb szempontok (a teljesség igénye nélkül):

## **Mutációs ráta:**

Sokszor előfordul (még fajon belül is akár!), hogy bizonyos genetikai elemek, genomi szakaszok, vagy akár a teljes genom egy evolúciós szétválás után az egyik fajban gyorsabban változik (mutálódik), mint a másokban (vagy a másokban lassabban, mint az egyikben, vagy mindkettő sebessége ellentétesen változik). Ennek közvetlen következménye lehet pl. rossz rokonítása a vizsgált fajoknak (pl. lelassult mutációs rátájú fajok a sok hasonló, ún. konzervált genetikai elem miatt rokonként tűnhetnek fel).

## **Mutációk típusai:**

### Pontmutációk I:

**Synonymous** (aminosav (továbbiakban AA) sorrendet nem befolyásolja), **non-synonymous** (AA sorrendet befolyásolja, szelekciós erők hatnak rá), kettő metszetén pedig a **neutrális mutáció**, ami ugyan változtathatja az AA sorrendet, de különösebben nem befolyásolja az élőlény fitneszét (bár hosszú távon ált. ezek felelősek egy csomó evolúciós újításért). Ezek összességében komolyan befolyásolják pl. a mutációs rátát, vagy különböző vizsgált gének fajok között megfigyelhető eltéréseit: **nem mindegy, milyen genetikai elemet v. fehérjét választunk fajok összehasonlítására!** Sok szempontot figyelembe kell venni, de a szelekciós nyomás megléte/hiánya az egyik legfőbb tényező, mert ez befolyásolja a filogenetikai elemzés felbontását (röviden: pl. közel rokon fajoknál nem érdemes 400 millió éve változatlan gént választani)

# FILOGENETIKAI ELEMZÉST BEFOLYÁSOLÓ TÉNYEZŐK II

## Pontmutációk II:

**Back mutáció**knak nevezzük azokat a szekvenciális változásokat, melynek során egy karakter (AA vagy nukleotid) eredeti ősi állapotára mutálódik vissza. Ennek nyilvánvaló következménye az adatsorokban létrejövő zaj vagy falszpozitív hasonlóság két szekvencia között. Szintén szükséges figyelembe vennünk, amennyiben ismerünk a vizsgálni kívánt szekvenciáinkban ilyet, az ún. **hotspotokat**, melyek olyan nukleotid (ritkábban AA) pozíciók, melyek olyan gyakran változnak, hogy nem hordoznak valós filogenetikai szignált, csak zajt generálnak.

## Inszerciók és deléciók:

Mind egyes nukleotidok szintjén, mind kromoszóma szinten előfordulnak inszerciók és deléciók, közös nevükön **indelek**. Ezek, és más kromoszómamutációk (inverziók, transzlokációk, stb.) azok, melyek igen változatos módon tudják befolyásolni közvetve (pl. kromoszóma töréspontokon vagy frameshift mutáció miatti felgyorsult mutációs ráta) és közvetlenül (szekvenciahossz változás) a filogenetikai elemzésünket.

## Hogyan tudunk ezek ellen a hatások ellen védekezni?

**Indelek:** illesztésekkel (vö. előző óra) könnyedén kiküszöbölhetőek (ugyanakkor további közvetett hatásaikra bonyolult matematikai-statisztikai modelleket és speciális vizsgálati stratégiákat kell kidolgozni, de ez advanced level, akit érdekel, keressen engem, adok irodalmat szívesen)

**Pontmutációk:** különböző matematikai-statisztikai módszerekkel modellezhetjük az evolúciós változásokat térben és időben, melyekhez a korábban említett változókra, mint inputra van szükség. Ezek segítségével jobban megérthetjük a fajaink (tulajdonságai) közötti különbségeket, és biztosabb eredményre juthatunk.



# FILOGENETIKAI MÓDSZEREK ÁTTEKINTÉSE

Alapvetően kétféleképpen állhatunk hozzá a (molekuláris) filogenetikai elemzésnek:

- Nem vesszük figyelembe a mutációk és egyéb biológiai jelenségek által generált befolyásoló tényezőket, és csak a direkt eltéréseket számoljuk le a szekvenciáink között. Ez bevett módszer nagyon közel rokon, vagy fajon belül végzendő filogenetikai elemzés esetén, ahol a back mutációk, indelek, stb. nem játszanak még akkora szerepet, hogy komolyabb zajt generáljanak, illetve szelekciós tényezők még nem jelennek meg olyan nagy mértékben, amit figyelembe kéne venni. Előfordul még, hogy filogenetikai módszerek jóságának tesztelésekor alkalmazzák. Ezt a megközelítést nevezzük **Hamming-távolságszámításnak** a filogenetikában.
- Figyelembe vesszük a befolyásoló tényezőket, és súlyozzuk az eltéréseket különböző kísérletesen megállapított **pontozómátrixok** vagy **evolúciós (ún. szubsztitúciós) modellek** segítségével. Ilyenek elérhetőek DNS és fehérjeszekvenciákra, sőt, indelekre is vannak már sikeres próbálkozások. Ezeknek a lényege, hogy egyes AA-k vagy nukleotidok eltérő valószínűséggel és gyakorisággal alakulnak át egymásba. Számos ilyen szubsztitúciós modell létezik, ezek megfelelő kiválasztása kritikus lépés, de programok is segítenek már ebben.

A modellválasztásunkkal együtt aztán alkalmazunk valamilyen filogenetikai módszert (ld. később), mellyel megkapjuk a filogenetikai elemzés kimenetelét: a filogenetikai fát.

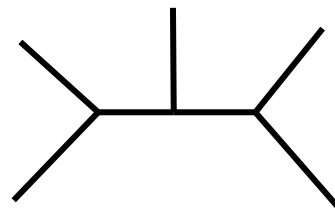
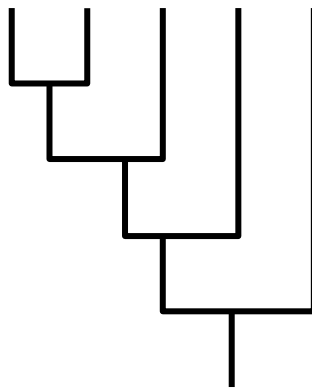
# A FILOGENETIKAI FA I

A filogenetikai fa mutatja meg nekünk a vizsgált taxonok közötti kapcsolatrendszer

Több módon ábrázolhatjuk:

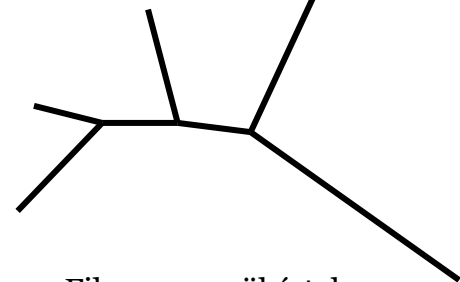
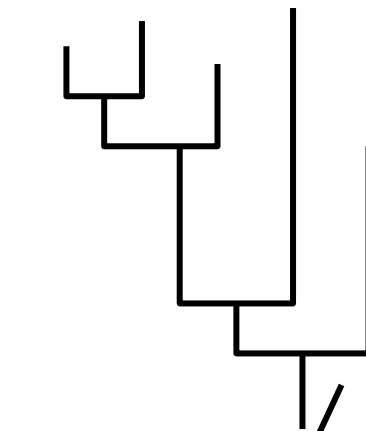
- kladogram (sima kapcsolatrendszer)
- filogram (evolúciós távolságok is jelölve vannak rajta)
- gyökértelen (unrooted) (nem ismert az evolúció iránya)
- gyökereztetett (rooted) (ismert az evolúció iránya)

Kladogram, gyökereztetett



Kladogram, gyökértelen

Filogram, gyökereztetett



Filogram, gyökértelen



# A FILOGENETIKAI FA II

A filogenetikai fán rendszertani csoportokat alakíthatunk ki/ismerhetünk fel, illetve sok esetben meglévőt módosítunk új molekuláris adatok bevonásával.

## Néhány alapfogalom:

### Monofiletikus csoport (Monophyly):

Olyan csoportja fajoknak (OTU-knak), melyek egy közös őstől erednek, és annak minden (ismert) leszármazottját magában foglalja (pl. madarak).

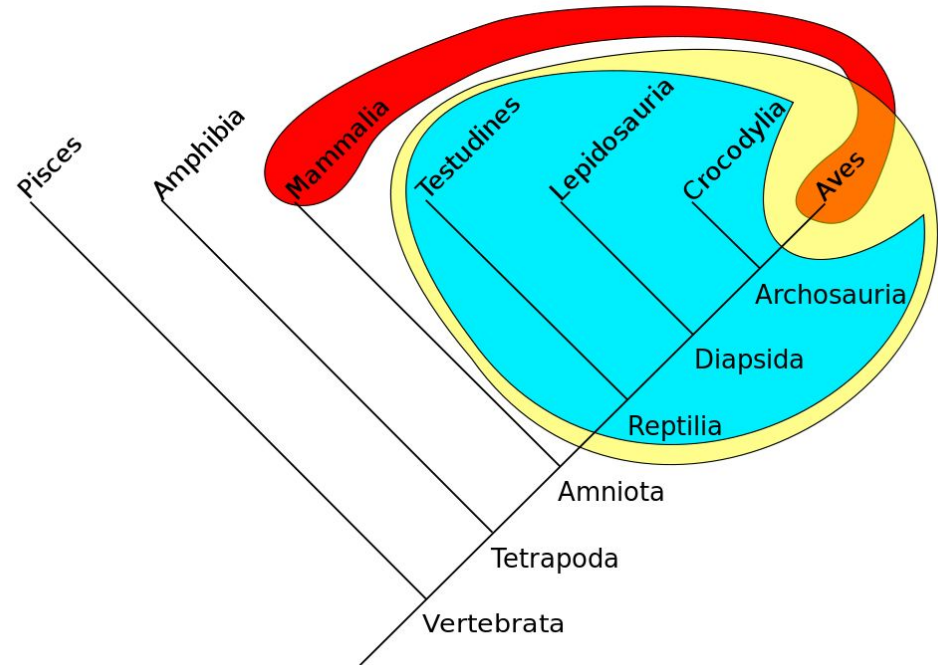
### Parafiletikus csoport (Paraphyly):

Olyan csoportja az OTU-knak, melyeknek közös rendszertani őse van, de abba nem tartozik bele minden leszármazott (pl. halak, hüllők).

### Polifiletikus csoport (Polyphyly):

Olyan csoportja az OTU-knak, melyeket önkényesen egy-egy meghatározott tulajdonságuk alapján sorolunk egy csoportba, de ezen tulajdonságaikra nézve nincs közös ősök, azok konvergens evolúcióval jöttek létre (pl. melegvérűek).

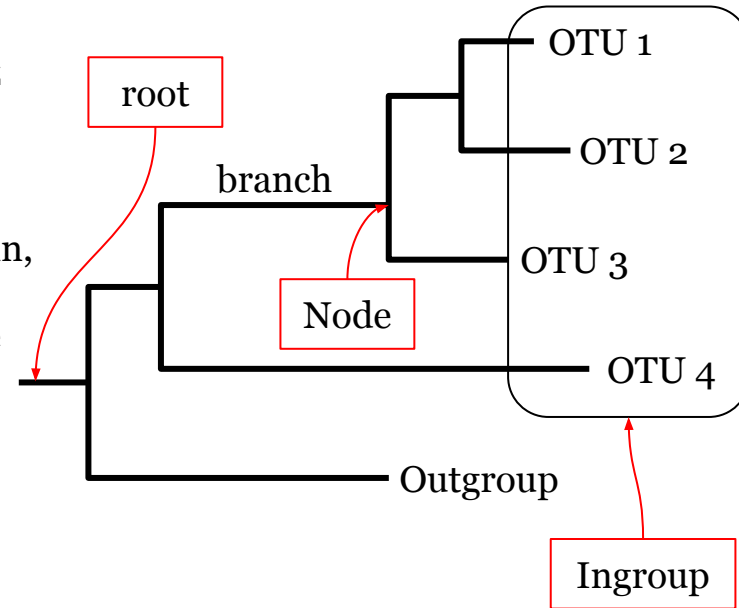
- Monophyly
- Paraphyly
- Polyphyly



# A FILOGENETIKAI FA III

Főbb részei:

- **Ág (branch):** kladogram esetén egyen hossz, filogram esetén az **ághossz (branch length)** reflektív a genetikai/evolúciós távolságra (minél hosszabb, annál nagyobb a távolság)
- **OTU (Operational Taxonomic Unit)** az ágak végén levő fajok, egyedek, minták, lényegében bármi, ami éppen vizsgálat alatt van, az ő rokonsági viszonyait nézzük a fán
- **Elágazás (node):** két (vagy több) OTU szétválását jelöljük vele (és a közös őst). Általában bifurkális (egy nódusból két OTU származik le), de előfordul többes elágazás is (trifurkális - 3, politómia,  $N > 3$ ; oka lehet valós biológiai esemény, hibás/zajos adatsor, számolási hiba)
- **Belcsoport (ingroup):** az összes vizsgált OTU-t magában foglaló ágak összessége
- **Külcsoport (outgroup):** egy vagy több olyan OTU, ami evolúciósan kívül esik a vizsgálni kívánt OTU-k csoportján, így tudunk vele gyökereztetni, azaz meg tudjuk vele adni az evolúció irányát.

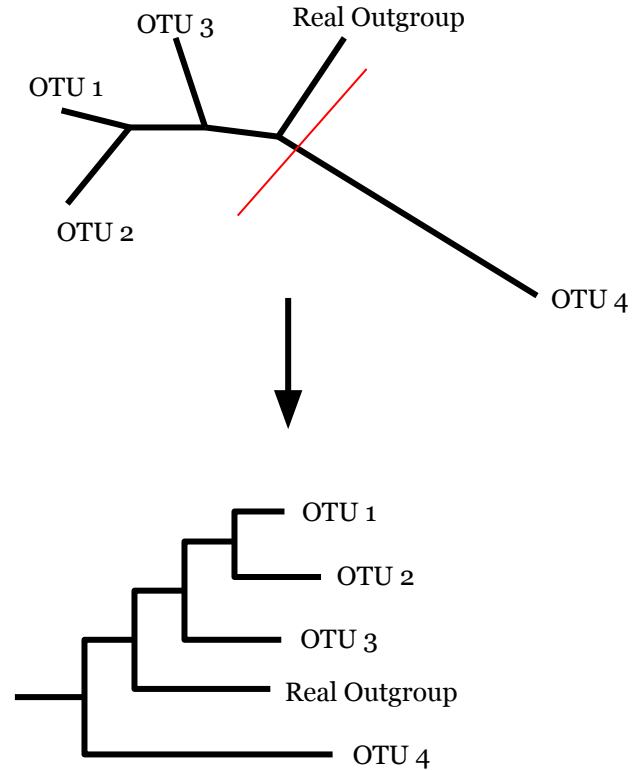




# A GYÖKEREZTETÉS

Közel sem annyira egyértelmű gyökereztetni egy filogenetikai fát, mint elsőre gondolná az ember. Külcsoportot választani legalább olyan körültekintéssel kell, mint evolúciós modellt: ha evolúciósan túl távoli, akkor lecsökkenti a vizsgálandó taxonok közötti variációt, ha túl közeli, fennáll a veszélye, hogy véletlen belcsoporthoz tartozó OTU-t jelölünk ki, ami miatt értelmetlenné válnak az elágazások.

Olyan is előfordul, hogy nem ismerünk külcsoportot, ekkor különböző alternatív megoldásokkal operálhatunk, pl. két egymástól legnagyobb távolságra lévő OTU között félúton jelöljük ki a gyökeret egy gyökértelen fán, vagy a legnagyobb ághosszú OTU-t jelöljük ki külcsoportként (de leginkább ezek kerülendőek).





# FILOGENETIKAI MÓDSZEREK

**Számos filogenetikai módszer létezik, ami szekvenciákkal és/vagy anatómiai jellegekkel dolgozik. Ma már kifinomult szoftverpark áll a felhasználó rendelkezésére, mellyel nem csak rokonságot és evolúciós távolságot lehet kiszámoltatni, hanem szétválási időt, hibridizációs eseményeket, palacknyak meglétét, idejét és mértékét, populációméretet, stb. is, akár egyetlen egyed mintájából is. Viszont éppen a nagy választék és a bonyolult, de cserébe rendkívül változatos és kifinomult matematikai-statisztikai módszertan miatt érdemes tisztában lenni azzal, hogy milyen lehetőségei, korlátai és hibaforrásai vannak egy ilyen vizsgálatnak.**



# FILOGENETIKAI MÓDSZEREK I

## Távolság alapú módszerek:

Viszonylag egyszerű felépítésű módszerek, lényegük, hogy illesztett szekvenciák közötti távolságot (Hamming vagy szubsztitúciós modellel együtt) kiszámítjuk, az eredményeket egy pontozómátrixban felvezetjük, amiből aztán kilépegetjük az OTU-k egymáshoz viszonyított helyzetét/távolságát.

Előnyök: egyszerűek; gyorsak; kis számítási kapacitást igényelnek, ezért nagy méretű adatsorokon sem kell években számolni az eredmény megszületését

Hátrányok: egyszerűek, vagyis pontatlanok, bonyolult filogenetikai problémákat nem igazán lehet vele megoldani.

Alkalmazási területek: nagy számú és nagy méretű adatsorok; közel rokon vagy fajon belüli diverzitás/leszármazás vizsgálata

Módszerek: **Neighbour Joining (NJ)**, UPGMA, Median Joining Network (MJN), Minimum Spanning Network (MSN), stb.

	OTU 1	OTU 2	OTU 3	OTU 4
OTU 1	0			
OTU 2	1	0		
OTU 3	2,2	1,4	0	
OTU 4	3,1	2,6	1,9	0

# FILOGENETIKAI MÓDSZEREK II

## Karakter alapú módszerek:

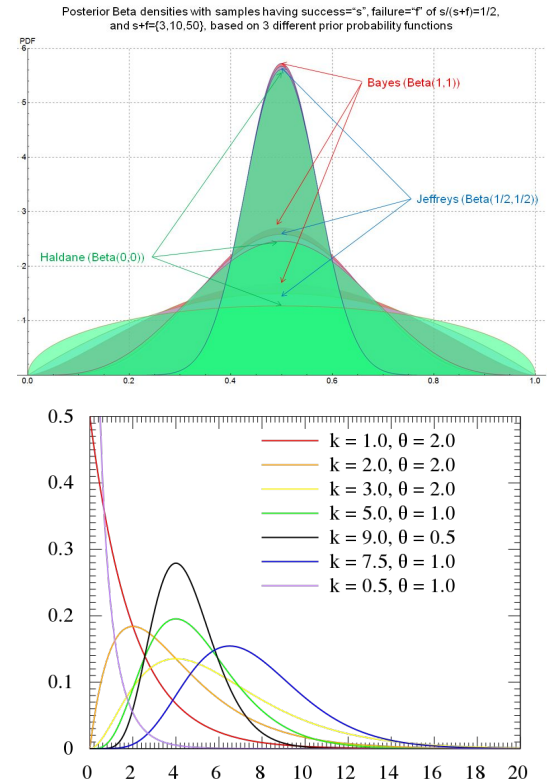
Bonyolult, komplex matematikai modelleket alkalmazó filogenetikai módszerek, karakterek (AA, nukleotid) közötti állapotváltozásokkal statisztikáznak kizárólag szubsztitúciós modellek alkalmazása mellett. Sok fajtája van, elég nagy különbségek vannak a számolási módok között, összefoglalva amit érdemes tudni:

**Előnyök:** a legtöbbje nagyon pontos, nehezen feloldható kapcsolatokra is találnak megoldásokat, precízen és megbízhatóan számolnak

**Hátrányok:** csodát egyik se tud tenni, csak mert karakter alapú módszer, még adhat rossz eredményt; iszonyatosan nagy számítási igényűek, sok esetben szerverhátter szükséges a működtetésükhöz, lassúak

**Alkalmazási területek:** genomika, komplex evolúcióbiológiai és populációgenetikai problémák feloldása, nagy adatsori zaj/evolúciós távolság leküzdése

**Módszerek:** **Bayesi filogenetika**, **Maximum Likelihood (ML)**, **Maximum Praszimónia (MP)**







**KÖSZÖNÖM A  
FIGYELMET!**