6

Kladisztika

(A "veszekedősek" tudománya)

Az előző két fejezettel korántsem zárhatjuk le a biológiai osztályozás tematikáját. Az eddig ismertetett módszerek a biológia objektumain kívül akár cserépedények, rajzszegek, gépjárművek vagy nagyvárosok osztályozására is felhasználhatók. (Csupán arra van szükség, hogy a sok objektum számos változóval legyen leírható.) Ugyanakkor a taxonómiában központi fontosságú leszármazási (evolúciós) viszonyokról nem adnak tájékoztatást. (Pontosabban fogalmazva: adhatnak, de alkalmazásuk során nem elsődleges ezek feltárása.) Ebben a fejezetben viszont olyan módszerekről lesz szó, amelyek kifejezetten az *evolúciós utak rekonstrukcióját* célozzák, hogy ezáltal alapot szolgáltassanak az élővilág filogenetikai osztályozásához. Ennek érdekében részben fel kell áldozniuk az "objektivitás"-t, a személyes döntésektől való függetlenséget is.

A leszármazási mintázatokat kereső metodológiát – némi általánosítással – *kladisztika* néven foglalhatjuk össze. Számomra nem kétséges, hogy a kladisztika is besorolható a többváltozós módszerek közé, hiszen sok objektum sok tulajdonsággal szerepel a vizsgálatban. A kladisztikai módszerek egy része azonban jóval specializáltabb a többinél, s a matematikai alapvetésen kívül igen fontosak, ha nem a legfontosabbak, a kutatónak az evolúcióról alkotott elképzelései is. Az egyes karakterek például, az eddigi módszereknél elfogadott alapelvekkel ellentétben, már nem egyformán lényegesek, s csupán azokra van szükségünk, amelyek evolúciós információt hordoznak, míg a többiek csak "zavarják" az összképet. Azonban még egy ilyen lényeges tulajdonság különböző állapotai sem teljesen egyenrangúak evolúciós szempontból. Kimondhatjuk, továbbá, hogy ha valamely csoport evolúciója során egy tulajdonság eltűnik, akkor az nagyon kis eséllyel vagy egyáltalán nem jelenhet meg újra. "Különleges" tulajdonságokról feltételezzük, hogy nem alakulhattak ki egymástól függetlenül két evolúciós vonalon, és így tovább. A felsorolásból talán már látható: rengeteg minden múlik az elemző biológuson, az evolúcióról alkotott általános elvein és a tanulmányozott csoportra alkalmazott speciális elképzelésein. Számos, nemcsak technikai döntést kell hoznia, mielőtt a formális elemző procedúráit bevetné. A kladisztika semmiképpen sem fekete cső, melynek egyik végén beletöltjük az adatokat, s a másik végén kijön a kész törzsfa (az efféle "csőhatás" veszélye sajnos fennáll a többváltozós módszereknél, általában). Igen alapvető sajátosság,

hogy a kladisztika a vizsgált objektumok mai állapotát figyelembe véve szándékozik egy valójában sohasem igazolható múltbéli eseménysorozatot felderíteni[1]. Így azután nem meglepő, hogy a kladisztikában számos, egymással nehezen összeegyeztethető, s olykor nehezen megbékélő irányzat létezik, amelyek nem csak vitatkoznak egymással (Gould, 1990, a kladisták egyes képviselőit a legveszekedősebb kutatók közé sorolja). Nem is könnyű a kladisztikáról egy fejezetben szólni, de úgy érzem, a legfontosabb alapelvekről s néhány módszerről mindenképpen említést kell tenni. A kladisztika témájának akár egy külön kötetet (mégpedig vastagat) is lehetne szentelni, feltéve ha valaki képes ebben a szerteágazó, nehéz filozófiai érvelésekkel alaposan "megspékelt" tudományágban tájékozódni (Stuessy, 1990, kifejezetten kétli, hogy ma ez egyáltalán lehetséges). Az alábbiakban megadott hivatkozások segítségével azonban minden Olvasó elindulhat az őt érdeklő irányban.

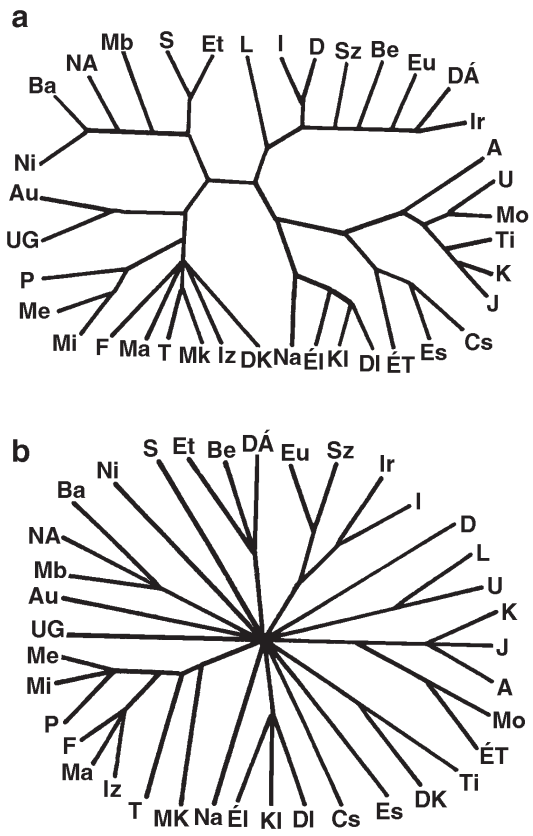
E fejezet, a fent elmondottak ellenére, nemcsak biológusoknak lehet érdekes. A kladisztika módszerei más tudományban is számításba jöhetnek, ha a történetiségnek, a leszármazás elemzésének szerepe van. Ilyen terület például a nyelvészet, amely meg is próbálkozott a nyelvek törzsfájának az elkészítésével (pl. Cavalli-Sforza et al. 1988). E fának és a humán populációk közötti genetikai különbségekből származó fának (6.1 ábra) az összevetéséből a népcsoportok nyelvi és genetikai koevolúciójára lehet következtetni (Penny et al. 1993). Mindezt csak kedvcsinálóul, a kladisztikus módszer alkalmazási lehetőségeinek illusztrálásaként bocsátjuk előre.

6.1 Alapelvek és alapfogalmak

A kladisztika egyes számú alapelve az, hogy a leszármazási viszonyok *fa-gráfok* formájában ábrázolhatók. Ebben önmagában persze nem sok újdonság van, hiszen már Darwin (1955) alapművének is egy leszármazási fa volt az egyetlen illusztrációja. A lényeg az, hogy mindvégig elágazások rendszerében, azaz "fában kell gondolkodnunk" ("*tree thinking*", O'Hara 1988), ha bárminemű evolúciós vizsgálatot végzünk (pl. akár a gének szintjén, egy populáción belül is). A fák általánosan elfogadott elnevezése a kladogram (gör. *kladosz* = ág szóból, vö. Camin & Sokal 1965), s a leggyakoribb ábrázolásmódját már az 5.1c ábrán bemutattuk. A kladogram más formában (rendszerint "lombozatával" felfelé), s akár még kör alakban is felrajzolható. A kladisták általában ügyelnek arra, hogy az ábrázolásmód ne legyen összekeverhető – az előző fejezetben sokszor látott – dendrogramos illusztrációval, ezzel is kiemelve a módszer- és szemléletbeli különbségeket.

A kladogram levelei, azaz végső szögpontjai, a tanulmányozott taxonok, a numerikus taxonómia definíciójával: OTU-k ("operational taxonomic unit", Sneath & Sokal 1973) vagy, a kladisztika célját jobban kiemelendő: EU-k ("evolutionary unit", Estabrook 1972). A gráf belső szögpontjai olyan "kihalt" evolúciós egységeket reprezentálnak, amelyek egykori létét csupán feltételezzük (HTU-k, "hypothetical taxonomic unit", Farris 1970), kivéve ha jó okunk van egy megfigyelt taxont ilyen belső szögpontként feltüntetni. A legelső (a 6.2 ábrán a legalsó) belső szögpont a *gyökér* helyét mutatja, ez az összes taxon *legfiatalabb közös őse.* A

1 Viszonylag újabb irányzat a sztratokladisztika (Fisher 1992), amely a tulajdonságokról rendelkezésünkre álló rétegtani információkat (vagyis az időbeli sorrendiséget) is figyelembe veszi.



6.1 ábra. Humán populációk kladogramjai genetikus (**a**) és nyelvészeti (**b**) információk alapján, Cavalli-Sforza et al. (1988) után, Penny et al. (1993) módosításaival. A gyökér pozíciója ismeretlen, bár mindkét esetben hajlamosak volnánk azt középre helyezni. Rövidítések: Mb: Mbuti (pigmeus), NA: Ny.-afrikai, Ba: Bantu, Ni: Nílusi-szaharai, S: Szan (busman), Et: Etiópiai, Be: Berber, Ir: Iráni, DÁ: DNy.-ázsiai, Eu: Európai, Sz: Szardíniai, I: Indus, D: Dravida (D.-indiai), L: Lapp, U: Uráli, Mo: Mongol, Ti: Tibeti, K: Koreai, J: Japán, A: Ainu (kis nép Japánban), ÉT: É. türk, Es: Eszkimó, Cs: Csukcs, DI: D.-amerikai indián, KI: Közép-amerikai indián, ÉI: É.-amerikai indián, Na: Na-Dene (amerikai indián népcsoport), DK: D. kínai, MK: Mon és Khmer (indokínai), T: Thai, Iz: Indonéz, Ma: Maláj, F: Filipino, P: Polinéz, Mi: Mikronéziai, Me: Melanéziai, UG: Új-guineai, Au: Ausztrál őslakó.

kladogram azonban, a dendrogrammal ellentétben, nem mindig rendelkezik gyökérrel, mert a közös ős helyzetének megállapítása általában a vizsgálat legbizonytalanabb része (de erről majd később). A gyökeres gráf egyes élei a leszármazási utakat jelölik, azaz az "ős → leszármazott" relációk kifejezői (a gráf tehát irányított, de a nyilakat rendszerint "elfelejtjük" feltüntetni).

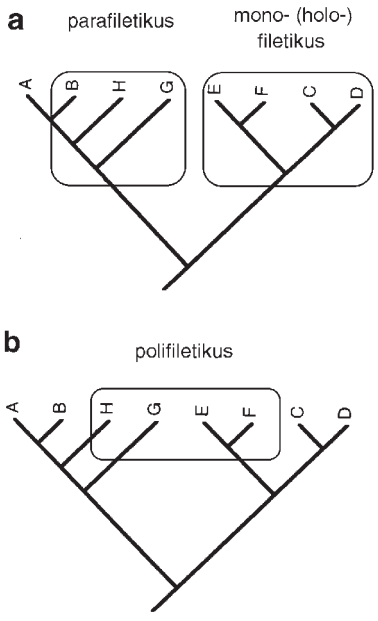
A kladogram sok esetben dichotomikus, azaz minden ősből két leszármazott jön létre. Számos kladisztikus irányzat szinte kötelezően előírja a dichotomikus elágazódást, és "feloldatlannak" (unresolved) tekinti azokat az eseteket, amikor háromszoros vagy többszörös elágazások is jelen vannak a fában (mint pl. a 6.1b ábrán). (Ez utóbbiak a politomikus fák, Wiley 1981.) Miután szigorúan csak fákról lehet szó, azaz kör nélküli gráfokkal dolgozunk, ki kell zárnunk a már elágazott részek újra kapcsolódását (anasztomózisok, vagyis "retikuláris evolúció"), bár tudjuk, hogy ez a valóságban egyáltalán nem lehetetlen (gondoljunk a hibridizációra és más génáramlási lehetőségekre, részletesebben Sneath & Sokal 1973: 352-356). Következésképpen a kladisztika módszerei főleg olyan típusú evolúciós utak rekonstruálására alkalmasak, amelyekben az egyes elágazások már stabilizálódtak, de nem igazán használhatók populációs vagy annál alig magasabb szinten, ahol még minden "kialakulatlan, kiforratlan". Azaz, a kladisztika inkább a makro-, s kevésbé a mikroevolúciós folyamatok kutatásának

eszköze – az utóbbi esetben az a bizonyos "fában gondolkodás" nem is olyan biztos, hogy a legmegfelelőbb.

A kladogram bármely rész-fáját *ág*nak, vagy *klád*nak nevezzük, az ide tartozó OTU-k egy ún. *monofiletikus* (egy közös őstől származó) csoportot alkotnak (6.2a ábra). A közös őst közvetlenül megelőző őstől származó másik klád az előbbiek *testvér*-csoportja (*sister* group, mint például az új-guineai – ausztrál a többi ázsiai és pacifikus népcsoportnak a 6.1b ábrán). A kladisták monofiletikusnak kizárólag olyan csoportot tekintenek, amelyek a közös őstől származó *összes* taxont tartalmazzák, s ha csak egy ilyet is kihagyunk (6.2a ábra), akkor a csoport már csak a *parafiletikus* jelzővel illethető (Hennig 1966). Ez a szóhasználat kétségtelenül ellentmondásban van a monofiletikus eredet korábbi (még nem kladista, hanem – mondjuk úgy: filogenetikai) definíciójával, amely nem írta elő "kötelezően" az összes leszármazott jelenlétét (Mayr 1942). A probléma Ashlock (1984) szerint úgy oldható meg, hogy az adott őstől származó összes leszármazottra a *holofiletikus* kifejezést alkalmazzuk, s ezt a taxonómusok jelentős része el is fogadja (vö. Stuessy 1990).

Hogy a dolgok egy kicsit tovább komplikálódjanak, beszélnünk kell még a *polifiletikus* csoportokról is, mivel a kladisztika szakirodalma ezeket is sűrűn emlegeti, gyakorta nem ugyanazt értve alatta. A 6.2b ábra polifiletikus csoportját a ma általánosan elfogadott definíció (Farris 1974) szerint jelöltük meg. A rajz alapján azonban első látásra azt mondhatnánk, hogy a polifiletikusnak mondott csoport valójában parafiletikus, hiszen megvan a közös ős, és a tőle származók közül nem vettük bele mindegyiket a csoportba. Van azonban egy döntő különbség: a polifiletikus csoport olyan kisebb csoportokat egyesít, amelyek saját, közvetlen ősei kimaradnak – míg a parafiletikus csoportok nem ilyenek.

Felmerülhet bennünk a kérdés, hogy mi az értelme a fenti okoskodásoknak és definícióknak? Mindez megválaszolható, ha egy csoport leszármazását a *taxonómia* szempontjából értékeljük. Elfogadva azt az alapelvet, hogy az élővilág osztályozása a leszármazási viszonyokon alapuljon (ezzel egyébként a biológusok zöme egyetért), akkor nyilvánvalóvá válik



6.2 ábra. Taxonok leszármazási relációi. A kladisztikai értelemben vett monofiletikus csoportban (**a**: jobb oldalon) minden leszármazott szerepel. A parafiletikus csoport (**a:** bal oldalon) nem tartalmazza a közös ős minden leszármazottját. A polifiletikus csoportból a közvetlen ősök kimaradnak (**b**).

számunkra, hogy a holofiletikus taxonok a legegyértelműbbek, ezután következnek a parafiletikusak, míg a polifiletikus csoportok tűnnek legproblematikusabbnak. Az általunk ismert és elfogadott klasszifikáció viszont a kladisztika megközelítésében sok ponton bizonyul paravagy polifiletikusnak. Mindezt Gould (1990) rendkívül szellemesen illusztrálja a halak példáján. A hagyományos értelemben vett halak csoportja polifiletikus, hiszen a bojtosúszósok a négylábú szárazföldi gerincesekhez kladisztikailag közelebb állanak, mint a többi halhoz. Az "ellentmondás" elsősorban abból ered, hogy a hal fogalma felületes makromorfológiai egyezéseket tükröz csupán. De még alacsonyabb rendszertani szinten, mondjuk a zebra definíciójával is zavarba jöhetünk, mert egyes csonttani bélyegek alapján a közönséges ló bizony a zebrafajok közé "ékelődik" a kladogramon – vagyis a zebra kifejezés sem biztos, hogy holofiletikus taxont takar.

A kladisztikára általánosan érvényes alapelvek közül még egyet kell itt megemlítetünk: általánosan elfogadott a biológiában (és más tudományokban is), hogy egy jelenség magyarázatában az egyszerűbb hipotéziseket kell megtartanunk a komplikáltabbakkal szemben. Itt ez a *minimális evolúciós utak elvében* fejeződik ki, melynek értelmében olyan fát tartunk optimálisnak, amely a taxonok legkisebb mértékű megváltozásával jár. Ez könnyen megérthető a távolság-alapú kladisztikában, s – mint később látni fogjuk – központi jelentőségű a karakter-alapon működő módszerek esetében. Az utóbbi esetben *parsimony* néven hivatkozunk rá, amelyet "takarékossági elv"-ként fordíthatunk magyarra (pl. Rédei 1987, p. 719). A parszimónia fogalma félreértések forrása lehet, hiszen az evolúció semmilyen értelemben sem "takarékos", és bizonyos, hogy az evolúció nem "ilyen irányban" zajlik. Csupán arról van szó, hogy nekünk *nincs más érdemi lehetőségünk* a jelenben fennálló eltérések kialakulását magyarázó munkahipotézis felállítására[2]. Ezért a továbbiakban a túl erősnek hangzó magyar fordítás helyett a semlegesebbnek tűnő latin megfelelőt használjuk. A parszimónia-elv filozófiai-biológiai vonatkozásairól egyébként Sober (1983, 1988) adja a legjobb áttekintést (lásd még Kluge 1984).

6.2 Kladisztika távolságok alapján

E módszereket részletezzük először, hiszen közvetlen kapcsolatba hozhatók az 5. fejezet hierarchikus osztályozó módszereivel. Mindjárt az elején felmerülhet bennünk a kérdés: körültekintéssel kiválasztott tulajdonságokra, egy jól megválasztott genetikai- vagy szekvenciákra alkalmazott távolságformula mellett használhatjuk-e a hierarchikus osztályozás módszereit kladogramok megbecslésére? Nos, sok kladista számára ez a kérdés fel sem merül, mások azonban – véleményem szerint helyesen – azt vallják, hogy kiindulásképpen érdemes pl. a csoportátlag módszert is kipróbálnunk a többi elemzéssel párhuzamosan (e módszer pl. Felsenstein **PHYLIP** programcsomagjában is benne van). A hierarchikus klasszifikációval szemben felhozható legfontosabb kladisztikai ellenérv az, hogy a kapott dendrogramon minden OTU *egyforma távolságra* van a gyökértől (ami az objektumokra "ráerőltetett", "túl szigorú" ultrametrikus sajátságok következménye). Rendszerint valószí-

2 Pl. nukleotid-szekvenciák esetén, ha mondjuk a 10. pozícióban A van az ősi szervezetben, és G a leszármazottban, az nem jelenti azt, hogy időközben nem következett be valamilyen más nukleotid-csere ezen a ponton. A végeredményben persze csak egy csere látszik. Ez a bizonytalanság persze a fa minden ágán megvan, így a parszimónia elvet nyugodtan alkalmazhatjuk még ebben az esetben is (részletesebben lásd a fejezet további részeit).

nűtlennek tartjuk azonban, hogy az evolúció során minden irányban pontosan egyforma mértékű lenne a változás a közös őstől számítva (mégha az eltelt idő azonos is)[3]. Tehát olyan fákat kell szerkesztenünk, amelyek már nem feltétlenül ultrametrikusak, hanem más *optimalitási feltételnek* felelnek meg (pl. legjobb közelítés az *additív* fához). Az alábbiakban áttekintünk néhány ilyen eljárást, sokszor csak az alapelvek ismertetésével, máskor pedig az algoritmust is részletezve. A fa rekonstrukciója során két dolgot kell figyelembe vennünk: a fa *elágazásainak a mintázatát* (vagyis a fa topológiai viszonyait) illetve az egyes ágakhoz (élekhez) súlyként rendelhető *távolságokat*. E két dolog változtatása az optimalizációs algoritmus során nem éppen könnyű feladat, s az egyes módszerek elsősorban ebben térnek el egymástól. Számos módszer gyökér nélküli fát eredményez először, amelyből azután gyökérrel rendelkező kladogram is előállítható.

Az additív fáknak kiemelt jelentősége van az evolúciós utak rekonstrukciójában. Ha a genetikai változásokat teljes mértékben ismernénk, akkor ezek összesítése egy tökéletesen additív fát eredményezne: a *valódi* törzsfa additív. A változásokat azonban nem ismerjük-ismerhetjük teljesen, csak a "végeredményül" kapott taxonokat. A közöttük felmérhető távolságok csupán *közelítései* (vagyis becslései) az evolúciós távolságoknak, s ezekből kell a fa rekonstrukcióját elvégeznünk a maximális additivitás szem előtt tartásával.

6.2.1 Élek összhosszának minimalizálása

Az egyik legrégebben ismert javaslat evolúciós távolságok fákkal történő reprezentálására Cavalli-Sforza & Edwards (1967) nevéhez fűződik. E szerzők azt tekintik optimális fának, amelyben az élek összhosszúsága a lehető legkisebb. Ez tulajdonképpen egy olyan minimális feszítőfa (5.4.3 rész) amelyben az m OTU-n kívül HTU-k is vannak. Gyökér nélküli fa esetében (vagyis m–2 HTU-val), a feladat 2m–3 db él összhosszának a minimalizálása. A szerzők eredeti algoritmusa meglehetősen bonyolult és nehezen követhető, Saitou & Imanishi (1989) azonban egy sokkal gyorsabb számításmenetet fejlesztett ki. Az eljárást Nei (1991) *"minimális evolúció módszere*" néven ismerteti. Ha a kiinduló távolságmátrix megfelel a négy-pont metrika feltételeinek (5.11 egyenlőtlenség), akkor tökéletesen additív fát kapunk eredményül.

A dendrogramok "hibáját", vagyis a konstans evolúciós változást, a Saitou & Nei (1987) által javasolt *szomszéd-összevonó* módszer (*neighbor joining*) eredeti módon próbálja meg kiküszöbölni. A fában egymás mellé kerülő objektumok kiválasztásánál nemcsak a távolságmátrix d_{ij} értékeit, hanem az objektumoknak az összes többivel alkotott távolságait is figyelembe vesszük. A módosított "távolság" annál kisebb lesz az eredetinél, minél nagyobb a két objektum átlagos távolsága a többitől, hiszen a nagy átlag (="nagy evolúciós sebesség") voltaképpen megnöveli kettejük relatív közelségét. A fa teljes felépítése egy agglomeratív osztályozáshoz hasonló módszerrel történik, melynek során a **D** távolságmátrix fokozatosan redukálódik. Végeredményben a fa éleinek az összhosszúságát optimalizáljuk, s e tekintetben az eljárás megfelel a minimális evolúció módszerének, csak annál sokkal gyorsabb és egysze-

3 Ez a megállapítás természetesen objektum-függő. Nagy számban találunk a molekuláris genetikai irodalomban olyan vizsgálatokat, ahol feltételezhető az azonos mérvű mutációs változás minden ágon (az evolúciós óra "ketyeg"). Ekkor a csoportátlag módszer hatékonyan alkalmazható (Degens 1983, Nei et al. 1983, konkrét példákat lásd Miyahara et al. 1992 és Adegoke et al. 1993 cikkeiben).

rűbb (Nei 1991). Az egyes lépéseket Swofford & Olsen (1990) leírását véve alapul a következőképpen összegezhetjük:

1) A $\mathbf{D}_{m,m}$ távolságmátrixhoz meghatározzuk a $\mathbf{v}_m$ vektort, melynek j-edik eleme a j objektum többitől vett távolságainak összege:

$$v_j = \sum_{k=1}^{m} d_{jk} \tag{6.1}$$

2) Megkeressük azt az objektumpárt, melyre nézve a

$$t_{jk} = d_{jk} - \frac{v_j + v_k}{m-2} \tag{6.2}$$

mennyiség minimális. t_{jk} voltaképpen nem távolság, s többnyire negatív az értéke. Ez csupán egy döntésfüggvény, amely megadja, hogy mely objektumpárt kell összekötnünk egy újonnan definiálandó u belső szögponttal. Legyen ez az objektumpár mondjuk h és i.

3) A h és i objektumoktól az u belső szögponthoz húzott élekhez rendeljük az alábbi távolságokat:

$$e_{hu} = d_{hi}/2 + \frac{v_h - v_i}{m-2} \tag{6.3a}$$

$$e_{hu} = d_{hi} - e_{hu} \tag{6.3b}$$

ami azt jelenti, hogy az u szögpont ahhoz esik majd közelebb, amelyiknek kisebb a távolságátlaga a többi taxonnal. Ha például $v_h<v_i$ akkor $e_{hu}<e_{iu}$. Ha a két távolságösszeg között jelentős a különbség, az élhosszra negatív érték is adódhat. Ez a jelenség analóg egyes hierarchikus osztályozások visszafordulásaival, és zavarhatja az eredmény interpretációját ill. ábrázolását, de szerencsére ritkán fordul elő.

4) Most következik **D** átszámolása. A h és i kiesik, helyettük az u-hoz tartozó sor ill. oszlop kerül be a mátrixba. **D** sorainak ill. oszlopainak száma tehát 1-gyel csökken. A következőkben most már u-nak a többi szögponttól vett távolságát kell használnunk, amelyek a következőképpen kaphatók meg:

$$d_{uk} = \frac{d_{hk} + d_{ik} - d_{hi}}{2} \tag{6.4}$$

Ha visszalapozunk az 5.1 táblázathoz, akkor felismerhetjük, hogy a fenti formula az egyszerű lánc (legközelebbi szomszéd) osztályozó módszer átszámolási kritériumának felel meg (bár az indexelés eltérő).

5) Amennyiben **D** mérete nagyobb, mint 2×2, akkor visszatérünk az 1. lépéshez. Ha már csak két szögpontunk van, akkor csupán a közöttük húzódó él hosszát kell megadnunk, ami nem más, mint $e_{hi} = d_{hi}$.

Ha a kiinduló távolságok tisztán additívak (amint az 5.10 mátrixban), akkor az optimális fát mindenképpen megkapjuk ezzel a módszerrel is. Egyéb esetekben a fa csupán közelítése lehet egy additív fának. A gráf egy gyökér nélküli fa, ami jól mutatja az evolúciós távol-

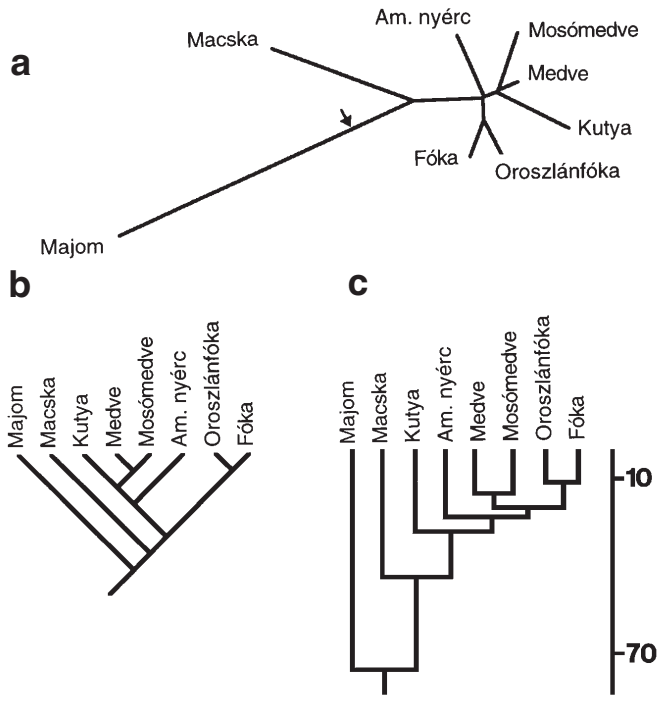
ságviszonyokat, de nem ad hipotézist a leszármazás irányairól, hiszen a közös ős pozíciója ismeretlen. A fát tehát le kell "gyökereztetnünk", ami kétféleképpen történhet:

1) Feltételezzük, hogy az egymástól legtávolabbi objektumpár azonos evolúciós sebességgel távolodott a közös őstől. Vagyis a gyökér a két objektum közötti út felezőpontjára kerül (*midpoint* módszer). Ezzel a feltételezéssel azonban – legalábbis részlegesen – visszatérünk ahhoz, amit eredetileg elkerülni szándékoztunk.

2) Még a számítások előtt leszögezzük, hogy a vizsgált taxonok egy holofiletikus csoportot alkotnak (nevezzük ezeket *ingroup*nak, "*belcsoport*"-nak). Keressünk hozzájuk egy vagy több olyan taxont, amelyek rendszertanilag kapcsolódnak a belcsoporthoz, de bizonyosan távolabb állanak tőlük evolúciósan, mint a belcsoport tagjai egymástól (*outgroup*, amelyet pedig "*külcsoport*"-nak fordíthatunk). Az elemzésbe mindkét csoport tagjait bevonjuk. A kapott fán azt az élt, ahol a külcsoport a vizsgált taxonokhoz kapcsolódik, joggal használhatjuk fel a belcsoportra nézve közös ős kijelölésére. Ha az elemzésből kiderül, hogy a két csoport mégis keveredik egymással, akkor felül kell vizsgálnunk a külcsoportot, és esetleg az egész vizsgálatmenetet új alapokra kell helyeznünk. A külcsoport bevonása azonban némi önkényességet visz az elemzésbe. Sok esetben egyébként nincs is lehetőség a külcsoport alkalmazására, így a 6.1 ábra példáján sem (ui. nincs olyan emberi populáció, amelyet akár genetikai, akár nyelvi alapon bizonyosan külcsoportként foghatnánk fel).

A szomszéd-összevonó módszert és a gyökér pozíciójának meghatározását egyes ragadozó emlősök immunológiai távolságmátrixának (A5 táblázat, Sarich 1969) felhasználásával illusztráljuk. A külcsoportot a majom képviseli az elemzésben.

A gyökér nélküli fát a 6.3a ábrán láthatjuk, amelyen az élek hosszúsága arányos a megfelelő távolságokkal. Miután ebben a példában a gyökér helyzetére mindkét módszer hasonló javaslatot tesz, a gyökeret reprezentáló szögpontot a leghosszabb út felezőjére helyeztük (6.3b



6.3 ábra. Ragadozók és a majom immunológiai mátrixából (A5 táblázat) végzett elemzés Saitou és Nei (1987) szomszéd-összevonó módszerével. **a**: gyökér nélküli fa, amelyen nyíl jelöli a gyökér jövendő helyét, **b**: kladogram, **c:** a csoportátlag módszer eredménye, amely topológiailag eltér a kladogramtól. A **c** ábra dendrogramján az eredetileg kiszámított ("különbözőségi") szinteket 2-vel osztva tüntetjük fel, ilymódon a patrisztikus távolságok megközelítik az evolúciós távolságokat.

ábra). A kladogramon a távolságokat már csupán súlyként fogjuk fel, az élek hossza nem arányos velük, mert az *elágazások mintázatát kívánjuk csak érzékeltetni*. Megjegyzendő, hogy azon taxonok között, amelyeket csak egy szögpont választ el, a *patrisztikus* távolság megegyezik az *immunológiai* távolsággal, míg más esetben valamivel kisebb v. nagyobb annál. Az élek összhosszúsága egyébként 274 egység.

6.2.2 Legkisebb négyzetek módszere

Cavalli-Sforza & Edwards munkásságával gyakorlatilag egyidőben Fitch & Margoliash (1967) az alábbi kritériumot javasolta:

$$FM = \sum_{i<j} \frac{(d_{ij} - e_{ij})^2}{d_{ij}^c} \tag{6.5}$$

amelyben d_{ij} a megfigyelt távolság, e_{ij} pedig a fában felmért patrisztikus távolság az i és a j taxon között, és $c = 2$. A feladat egy olyan fa megszerkesztése, amelyben FM a legkisebb értéket veszi fel (az irodalomban sokszor a fenti függvény különféle változatai szerepelnek). A fa megtalálása lényegében véve két fő lépés kombinációjából áll: 1) egy adott topológiájú fa esetében azokat az élhosszakat kell kiszámítatnunk, melyekre a fenti formula a legjobb illeszkedést adja, és 2) meg kell keresni azt a topológiát, amelyre nézve majd az FM mennyiség minimális lesz. A művelet tehát többlépcsős, és Fitch & Margoliash eredeti módszere nem is volt alkalmas az optimális fa megtalálására.

Az eljárás sok tekintetben a csoportátlag módszerre emlékeztet. Weir (1990) ismertetését követve az alábbi tömör leírást adhatjuk:

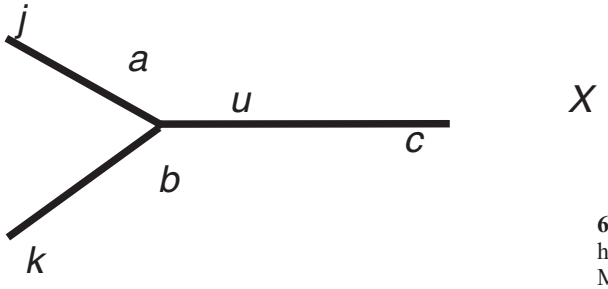
1) **D** alapján kiválasztjuk a két, egymáshoz legközelebbi taxont; legyen ez j és k. Közöttük felveszünk egy új HTU-t, amelyet jelöljünk u-val. Az összes többi taxont egy csoportnak tekintjük (ennek jele legyen X, elemszáma pedig n_X). A j és k taxonok távolságát X-től az összes távolság átlagával definiáljuk:

$$d_{jX} = \sum_{i \neq j,k} d_{ij} / n_X = a + c \tag{6.6a}$$

$$d_{kX} = \sum_{i \neq j,k} d_{ik} / n_X = b + c \tag{6.6b}$$

$$d_{jk} = a + b \tag{6.6c}$$

Amit keresünk, az éppen az a, b és c szakaszok hossza (6.4 ábra), amelyek a 6.6a-c összefüggésekből könnyen kiszámíthatók.



6.4 ábra. Illusztráció az élek hosszának kiszámítására a Fitch-Margoliash módszer egy lépésében (l. a szöveget).

2) A kapott új szögpont távolságát a taxonoktól az $s = (a+b)/2$ képlet alapján számítjuk ki. *j*-t és *k*-t ezután *u* képviseli **D**-ben, ennek méretei tehát eggyel csökkennek.

3) A csoportátlag módszerből (5.2.1 rész) ismert módon átszámítjuk *u* távolságát az *X* halmaz mindegyikétől, külön-külön. Más szóval, az *u*-val reprezentált taxonok és egy másik, *h* szögpont távolságainak az átlagát határozzuk meg. Ezeket **D** megfelelő helyére írjuk be.

a) Ha csak egy távolságérték marad **D**-ben, akkor *s*-t kivonva megkapjuk az utolsó él hosszát s a gráf elkészült.

b) Minden más esetben a legkisebb távolságérték keresendő meg **D**-ben, és az 1) lépésben vázolt ill. a 6.4 ábrán illusztrált módon újabb HTU-t definiálunk a megfelelő élhosszakkal. Visszatérünk a 2. lépéshez.

A kapott dendrogram topológiája a szerzők "beismerése" szerint sem feltétlenül optimális. Az ágak átrendezésével, amolyan próbálgatós taktikával igyekeztek tehát a legjobb topológiájú fát megtalálni. A fa éleihez rendelt "távolságok" pedig negatívak is lehettek, akárcsak a szomszéd-összevonó módszer esetében. (A negatív hosszúságú élek kezelésére Swofford & Olsen, 1990: 449, ad ötleteket, pl. ezeket 0 értékűnek tekinthetjük). Ma már az eredeti módszer elvesztette jelentőségét, mert jóval hatékonyabb eljárásokat ismerünk. Arra azonban továbbra sincs mód, hogy viszonylag nagy számú taxonra (>20) minden lehetséges topológiát kipróbáljunk. Azt tekintjük a legjobb algoritmusnak, amely adott idő alatt a legtöbb lehetőséget tudja számításba venni illetve kizárni.[4]

Cavalli-Sforza & Edwards (1967) a $c = 1$ behelyettesítést javasolta, tehát a tiszta eltérésnégyzetösszeget minimalizálta. (Implicit módon ezt optimalizálja a szomszéd-összevonó módszer is amellett, hogy az élek összhosszát a lehető legkisebbre veszi.) Ezen kívül *c* más értékei is elképzelhetőek, s így akár egy teljes módszersorozatot is generálhatunk. Felsenstein (1993) szerint *c* értékének megválasztása a távolságok becslési hibájától függ. Ha jó okunk van feltételezni, hogy a távolságértékeket azonos hibával becsüljük, bármekkora is az értékük, akkor a c=0 opció a megfelelő. Amennyiben a hibavariancia nő a távolság növekedésével, a c=2 értéket kell választanunk. A közbülső c=1 érték valójában annak az esetnek felel meg, amikor a hibavariancia a távolság négyzetgyökével arányos.

A Sarich-féle immunológiai távolságmátrixra a Fitch & Margoliash módszer (a **PHYLIP** programcsomag **FITCH** programja, Felsenstein 1993) gyakorlatilag ugyanazt az eredményt adta (több száz fát kipróbálva), mint a szomszéd-összevonó eljárás, csupán az élek hosszúságában mutatkozik némi eltérés. Ilyen szoros egyezésre persze nem mindig számíthatunk, különösképpen akkor, ha a taxonok száma jóval nagyobb 8-nál.

6.2.3 A négy-pont feltétel teljesülésének maximalizálása

Az 5. fejezetben már láttuk, hogy ha az 5.11 egyenlőtlenség fennáll, akkor a távolságmátrix egyértelműen ábrázolható egy additív fa formájában. A valóságban ez a feltétel ritkán teljesül, mert a távolságviszonyok kisebb-nagyobb mértékben "torzulnak". Sattath & Tversky (1977) javaslata alapján (lásd még Fitch [1982]) a mátrixra legjobban illeszthető additív fát a topológia optimalizálásával kell kezdenünk. Ilymódon a feladat egy olyan fa meghatározása, amelyben a lehető legkisebb számú objektum-négyesre sérül meg a négy-pont metrika feltétele. Ha ezt megtaláltuk, akkor valójában már a legkisebb négyzetek elvének figyelembe-

4 Így talán érthető, hogy a kladisztikai cikkek bírálói megkövetelik a korrekt utalást mind az alkalmazott módszerre mind pedig a felhasznált programcsomagra.

vételével számoljuk ki az élek hosszúságait (6.5 formula, $c = 1$). Az esetlegesen előadódó negatív értékeket 0-val kell behelyettesítenünk. Amennyiben a kiinduló mátrix távolságai additívak, a módszer pontosan reprodukálja az additív fát.

6.2.4 A Wagner-távolság módszere

A fentiekben tárgyalt módszerek közös tulajdonsága, hogy a fa éleihez rendelt értékek, vagyis a becsült távolságok alapján a patrisztikus távolságok kisebbek is és nagyobbak is lehetnek, mint a kiinduló **D** mátrix értékei. Az optimalizáló stratégia ugyanis az eltérés irányára teljesen érzéketlen (s ezért adódhatnak negatív élhosszak is). Ha azonban kimondjuk, hogy a kiinduló távolságok a fában felvehető távolságoknak az alsó határát jelentik, akkor a negatív értékeket biztosan kiküszöböljük. Optimális tehát az a fa, amelyben az *élek összhosszúsága minimális, és egyetlen patrisztikus távolság sem haladja meg a megfelelő kezdeti távolságot.* Ezt a fát Farris (1970) nyomán Wagner fának nevezzük[5]. Az eljárás megértéséhez újra idézzük fel a minimális feszítőfát (5.4.3 rész). Ebben minden pont megfelel egy OTU-nak, m–1 élünk van, s az élek összhosszúsága minimális (a Sarich-féle immunológiai távolságmátrixra ez 365). További szögpontok hozzáadásával azonban az élek összhosszúsága csökkenthető. (Gondoljunk vissza a Saitou & Nei-féle megoldásra, ami 274 volt! A 365 egység még akkor is sok, ha nem engedjük meg a távolságok lefelé ingadozását.) Ezek a hozzáadott pontok lesznek a HTU-k. A Farris javasolta módszer egy heurisztikus közelítése a lehetséges optimumnak, és több változata is ismeretes (Farris 1972, Swofford 1981, Tateno et al. 1982, Faith 1985), és csak a Manhattan távolságokra (3.48 formula) alkalmazható. Az elemzés a két legközelebbi OTU összekapcsolásával kezdődik. A további lépésekben egy-egy OTU adódik a fához olymódon, hogy a hozzá legközelebbi élen egy új belső szögpontot (HTU-t) létesítünk, s ahhoz csatlakoztatjuk.

> Az algoritmus részletes bemutatásától eltekinthetünk. Az immunológiai távolságok fenti példájában ugyanis Fitch (1984) a Wagner-távolság módszerrel egy 291-es összhosszúságú fát talált, amely mindenképpen "rosszabb", mint a szomszéd-összevonó és a Fitch - Margoliash módszerrel kapott fák esetében. Mint Felsenstein (1993) is rámutat, a Wagner távolság módszere inkább csak történeti jelentőségű, mert más algoritmusok rendszerint jobb eredményre vezetnek.

6.3 Evolúciós fák rekonstruálása karakterek alapján

Eltekintve azoktól az esetektől, amikor eleve távolság-adataink vannak (pl. DNS hibridizáció [Krajewski & Dickerman 1990], immunológia), a kladisták szemében a távolság-alapú módszerek fő hátránya az, hogy alkalmazásukkal rengeteg információt veszítünk. Véleményük szerint, ha van egy OTU × tulajdonság adatmátrixunk, akkor a távolságmátrixba alakítás során éppen az egyes karakterek megváltozásáról (*character evolution*, Maddison & Maddison 1992) nem tudunk következtetéseket levonni, vagyis a távolság-módszerek teljesen használhatatlanok. Anélkül, hogy ebbe a vitába (ami időként meglehetősen elmérgesedett)

5 Azért Wagner fa, mert Farris módszere egy karakter-alapozású (6.3 rész) kladisztikai eljárás általánosítása folytonos típusú változókra, s annak kidolgozása és első alkalmazása Wagner nevéhez fűződik. Ha a "távolság"-ra nincs utalás, amikor Wagner-fákat emlegetnek, akkor nem erről, hanem a 6.3 részben ismertetendő módszerről van szó az illető cikkben.

belemennénk, meg kell jegyezni: a karakter-alapú kladisztika elsősorban csak diszkrét típusú változókat képes kezelni, használata tehát korlátozott. (Vannak ugyan folytonos változókat diszkretizáló módszerek, de érezzük, hogy ekkor az átalakítással vész el információ, tehát amit nyerünk a réven, azt könnyen elveszíthetjük a vámon.) A megoldás csak az lehet, hogy lehetőleg mindkét eljárástípust kipróbáljuk (sőt a numerikus klasszifikációt is, amint azt Duncan et al. 1980 javasolják). Ebben a fejezetben természetesen nem lesz már szó távolságokról, mert figyelmünket a tulajdonságok közvetlen kladisztikai hasznosítására fordítjuk.

Egyébként most az "igazi", a szűkebb értelemben vett kladisztika "vadászmezejére" érkeztünk. Bár a karakter-kladisztika elméleti előfutárának, eszmei atyjának egyértelműen a német Hennig tekinthető (1950, 1966), a további fejlemények szinte kizárólag az angolszász nyelvterületre korlátozódtak. Kialakult egy, a kívülállók számára ezoterikusnak ható szakzsargon, amely nagymértékben okozója annak, hogy a karakter-kladisztika még ma sem tört át bizonyos gátakat. A 6.1 részben ismertetett alapelveken túlmenően tehát meg kell ismerkednünk számos, csak erre a területre érvényes fogalommal is.

Hennig eredeti érvelései azon alapulnak, hogy az evolúció során a tulajdonságok (karakterek, bélyegek) változnak, az ősi (primitív v. a kladisztika nyelvén: *pleziomorf*) állapotból az újabb (leszármaztatott vagy *apomorf*) állapotba jutnak. Egy ősi állapotból természetesen sok leszármaztatott állapot alakulhat ki[6], és adott monofiletikus rendszertani csoportban csak egy pleziomorf állapot képzelhető el. Az evolúciós leszármazási mintázat rekonstruálásában arra kell törekednünk, hogy egy taxon tagjai *minél több leszármaztatott karakterben egyezzenek meg* (az ilyen egyezést *szünapomorfiának* nevezte el Hennig), míg az ősi állapotokban való megegyezéseknek (*szümpleziomorfia*) nincs különösebb jelentőségük, hiszen ezek semmilyen evolúciós információt nem hordoznak. Ha egy leszármaztatott bélyeg egyetlen egy ágon jelentkezik csupán, akkor *autapomorf* karakterről beszélünk.

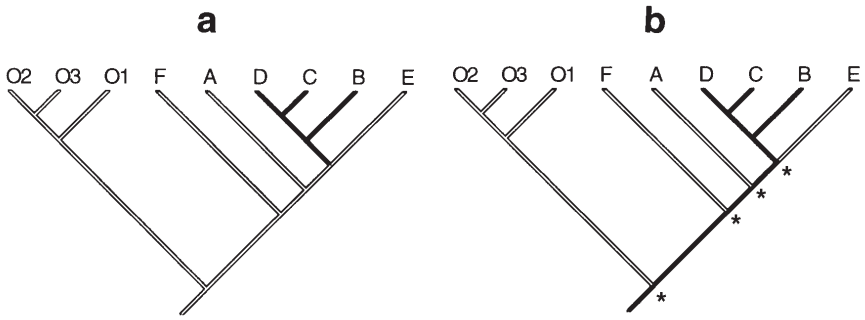
Annak eldöntésére, hogy adott tulajdonságnak melyik az ősi és melyek a leszármaztatott állapotai, más szóval a *karakterek polaritásának* vizsgálatára, a kladisztika számos "trükköt" ismer, amelyeket az alábbiakban foglalhatunk össze:

1) Eredeti és ötletes kladista eljárás a már említett *külcsoport*-módszer. A rokon taxon bevonása az elemzésbe nemcsak a leszármazási fa gyökerének megállapításában, hanem a polaritás kimutatásában is hasznos lehet (Watrous & Wheeler 1981). Tételezzük fel, hogy az elemzett (a bel-) csoportban egy adott tulajdonság "fehér" és "fekete" állapota létezik. Logikusnak tetszik ekkor a gondolat, hogy nézzük meg ezt a tulajdonságot a külcsoportban is, és ha ott a "fehér" állapot fordul csupán elő, akkor ez tekinthető az ősi állapotnak (6.5a ábra). Elképzelhető persze az is, hogy tévedtünk, és mégis a "fekete" az ősi, de ehhez sokkal több karakterváltásra van szükség az evolúció során, azaz a kapott kladogram kevésbé valószínű (6.5b ábra). Ha a külcsoportban is előfordul mindkét állapot, akkor azt tekintjük pleziomorfnak, amelyik gyakrabban fordult ott elő.

2) A "gyakoribb az ősibb" elv alapján akár mellőzhető is a külcsoport, és ősinek a belcsoportban leggyakoribb állapotot tekintjük (pl. Kluge 1967, Stuessy 1990). *Általában* ugyanis sokkal kevesebb változást kell feltételeznünk egy fában, ha a gyakoribb állapotot tekintjük ősinek. (Ez a megfontolás a 6.5 ábra esetében azonban nem segít.)

3) A biológus számára a fenti érveléseknél elsődlegesebbnek hat a fosszilis bizonyítékok alkalmazása: a régebbi rétegekből ismert állapotot nagy valószínűséggel tekinthetjük ősibbnek,

6 Ezek egy nominális típusú változó különféle értékeivel feleltethetők meg, 1.4.1 rész.



6.5 ábra. A külcsoport (outgroup) módszer karakterállapotok polaritásának megállapítására. Tételezzük fel, hogy a fa topológiája a tényleges evolúciós mintázatot tükrözi, vagyis az apomorf állapot ("fekete") az B, C és D taxonok közös ősében jelent meg (**a**). A fát még nem ismerve tehát joggal tarthatjuk a külcsoport taxonjaiban (O1-O3) is meglevő "fehér" karakterállapotot a belcsoportban ősinek. Ha nem ezt tennénk, akkor a valódi fa csak négy karakterváltással lenne megmagyarázható (**b**).

mint az új rétegekből ismertet (pl. Gingerich 1979). Ehhez persze elegendően nagy, és lehetőleg "hiánytalan" leletsorokra van szükség.

4) Az egyedfejlődési kritériumok közül a legfontosabb az ontogenezis során legelőször megjelenő állapot (gondoljunk itt a – már elavultnak tekintett – Haeckel féle "alaptörvényre": az egyedfejlődés a filogenezis rekapitulációja). A fenyőféléknél így pl. a tűlevél az ősi állapot, mert a pikkelyes levelűeknél is megjelenik az egyedfejlődés korai szakaszában. Ez kétségtelen empirikus tény, még akkor is, ha a fenti "alaptörvény" érvényessége nem mindig egyértelmű.

5) A szervfejlődés kisebb eltérései is bizonyítékot szolgáltathatnak az ősibb állapot eldöntésére: egy abnormális állapot felbukkanása a populációban utalhat a tulajdonság korábbi állapotára ("atavizmus").

6) A csökevényes szervek jelenléte is utalhat az ősi állapotra, amennyiben meggyőződtünk arról, hogy az illető szerv semmiféle funkciót nem lát el.

7) Evolúcióbiológusok és taxonómusok tapasztalata szerint primitívnek tekintett bélyegek nagy eséllyel jelennek meg együtt, egy csoportban (azaz: korrelálnak). Ennek alapján egy még nem vizsgált tulajdonság csoport-beli állapotáról eldönthetjük, hogy ősi-e vagy sem (Crisci & Stuessy 1980, Sporne 1976).

8) Egy másik, hasonlóképpen másodlagos szempontot jelent a csoportokon belüli párhuzamos evolúciós trendek összehasonlító értékelése. Példa a fejecskevirágzat másodlagos aggregációja 40 virágos génuszban, miszerint tehát a szimpla fejecske az ősibb állapot.

9) A legvitatottabb, de sokszor mégis jól használható kritérium a taxonok elterjedésének figyelembevételén alapszik. Miután a régebbi taxonoknak több idejük volt az evolúció során az elterjedésre, a legelterjedtebb taxonban megfigyelt bélyeg tekinthető ősinek a szűk elterjedésű taxont jellemző bélyeghez képest.

A fentiekben csupán egy rövid felsorolást adhattunk; a nehézségek és korlátok, a relatív előnyök és hátrányok megemlítésére, pláne alapos megvizsgálására nem volt módunk, hiszen ez is megtöltene egy külön fejezetet (részletesebben lásd Stuessy 1990:106-113, Quicke 1993: 16-22, Mayr & Ashlock 1991: 212-214). Amint azonban az elsőként idézett szerző is rámutat: nincs kitüntetett módszer, amely mindig helyes megoldásra vezetne.

A polaritás feltárásán túlmenően vigyáznunk kell a *homológiára*, vagyis arra, hogy a karakterállapotok egyezése a közös leszármazás eredménye legyen. (Itt bizonyos értelemben egy ördögi körbe kerülünk, hiszen a homológia eldöntéséhez először ismernünk kellene a leszármazási viszonyokat, amelyeket éppen most keresünk. Az esetek egy részében azonban a biológus könnyedén felfedi a homológ állapotokat, mert valamilyen külső információt is felhasznál. A homológia egyébként nemcsak a kladisztika, hanem a numerikus taxonómia fontos alapelve is, hiszen ott sincs értelme a legyet meg a madarat hasonlónak tekinteni csupán azért, mert mindegyiknek szárnya van.) A kladista "ellensége" a *homoplázia,* a homológiával ellentétes eset, amikor is egy adott karakterállapotbeli megegyezés nem a közös leszármazás bizonyítéka. A parallel vagy konvergens evolúció révén, egymástól függetlenül kialakult tulajdonságokban tapasztalt megegyezés ugyanis félrevezető a valódi evolúciós út feltárásában, és így megnehezíti a kladista dolgát[7]. A homoplázia megnyilvánulásának tekintjük azt is, ha az evolúció során egy apomorf jelleg *visszaalakul* az ősi állapotba. Ez tehát a pleziomorf jelleget "utánozza" és attól bizony igen nehéz megkülönböztetni.

A bélyegek visszafordulása azonban számos esetben kizárható, hiszen sok tulajdonságról semmiképpen sem képzelhető el, hogy újra az ősi állapotba kerüljön. Nem elegendő tehát önmagában a polaritás eldöntése, mert a lehetséges irányokról is határoznunk kell, mielőtt egy részletesebb elemzésbe fogunk. Itt következik tehát a kladisztika egy másik kritikus területe, az állapotok között általunk megengedett átmenetek vizsgálata. Az alábbiakban egy rövid összefoglalását adjuk ennek a szerteágazó, és sok vitával "terhelt" témának. Láthatjuk majd, hogy a kladisztikában nem elegendő az adattípusok 1. fejezetbeli csoportosítása, mert további finomításra van szükség. Ugyanakkor a kladisztikai terminológia több ponton átfed az általunk már megismert definíciókkal, s ezek tisztázása csak növelheti a tájékozottságunkat e kérdéskörben.

1) A nominális változóknak felelnek meg egyértelműen a kladisztika *rendezetlen* ("unordered" vagy Fitch [1971] -féle) karakterei. A polaritás itt még nem is érdekes, hiszen egy tulajdonság állapotai az evolúció során bármelyik másikba át- ill. visszaalakulhatnak (6.6a ábra). Minden ilyen átalakulást egyenlő súllyal veszünk figyelembe, s egy változás éppen egységnyi értékkel járul a patrisztikus távolsághoz a kladogramon. Tipikus példa a DNS molekula egy adott pozíciójában található nukleotid minősége. Itt a mutáció minden párosításban végbemehet (bár tudjuk, az átalakulás nem egyformán valószínű minden irányban, ezért a transzverziókat nagyobb súllyal vehetjük figyelembe, lásd a 6.3.1.3 részt).

2) A kladisztika összes többi karaktere már tartalmaz ordinális információt. A Wagner-féle karakterek (Farris 1970) állapotai *sorba rendezettek,* az átalakulás pedig mindkét irányban végbemehet ("ordered and reversible"). A rendezettség miatt egy adott A állapotból csak a szomszédos állapotba juthatunk el közvetlenül (6.6b ábra). A kladisztikában egy ilyen ugrást egységnyi változásként kezelnek, vagyis a karaktert nem csupán ordinális, hanem intervallum típusúnak fogják fel. "Jó" példa a levélkék száma egy összetett levélben, hiszen ez statisztikai értelemben is intervallum (sőt: arány-) skálán mért változó. Gyakori azonban, hogy a "kicsi-

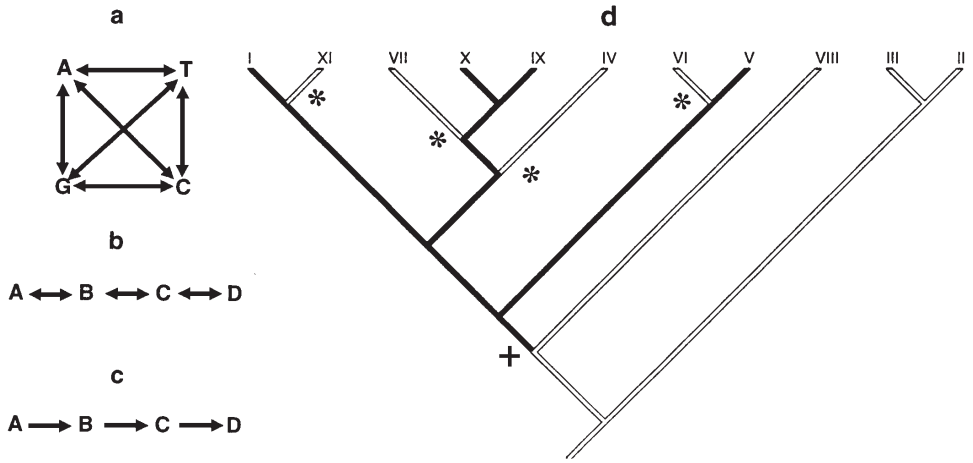
7 A parallelizmus és a konvergencia megkülönböztetése számunkra nem feltétlenül szükséges, és nem is mindig egyszerű feladat. A légy és madár szárnya vagy pedig a kaktuszok és egyes kutyatejfélék pozsgás szára igen távoli taxonok konvergens evolúciójának eredményei és ezek ritkán okoznak problémát a kladisztikai elemzésben, mert könnyen felismerhetők és a további vizsgálatból kizárhatók. A parallelizmus viszont azt jelenti, hogy adott taxonok teljesen hasonló feltételekkel "indulnak", s azonos változásoknak vannak a későbbiek során kitéve (Gosliner & Ghiselin 1984, Harvey & Pagel 1990). Példa erre az énekesmadarak körében megfigyelhető nagyszámú morfológiai párhuzamosság a különböző kontinenseken.

közepes-nagy-legnagyobb" és hasonló jellegű állapotsorozatokat is ilyeténképpen kezelik, pedig ezek csupán "tisztán" ordinális típusú változók.

3) Amennyiben a rendezett karakter egyes állapotai csak egy irányban alakulhatnak át egymásba, akkor a kladisztika *irreverzibilis* tulajdonságairól van szó (amit első alkalmazóikról Camin - Sokal-féle [1965] karaktereknek neveznek, bár ez utóbbi terminus sok szerzőnél csak a bináris típusú irreverzibilis tulajdonságokra vonatkozik). Az ilyen tulajdonságok tehát nem "veszhetnek el" az evolúció során, csupán egy még újabb állapotba alakulhatnak át (6.6c ábra). Az irreverzibilis tulajdonságok meglehetősen ritkán fordulnak elő, és megfordíthatatlan jellegük állandó vita tárgya (pl. poliploidia).

4) Bizonyos értelemben a reverzibilis és irreverzibilis tulajdonságok közötti átmenetet jelentik a Dollo-karakterek (LeQuesne 1974, Farris 1977). Itt is létezik egy pleziomorf kiindulópont, amelyből legegyszerűbb esetben csak *egy* új állapot jön létre (6.6d ábra), de új állapotok egy *sorozata* is elképzelhető. Az új állapot a törzsfa különféle ágain egymástól függetlenül többször is elveszhet (bármelyik előbbibe, vagy a legősibbe visszaalakulhat). Lényeges továbbá az a feltétel, hogy minden leszármaztatott karakterállapot egyszer és csak egyszer alakulhat ki a törzsfejlődés során ("uniquely derived"), vagyis a parallelizmus és a konvergencia kizárt. Ez igen szigorú feltétel, s leginkább csak a restrikciós enzimek esetén tekintik érvényesnek (Swofford & Olsen 1990). Ide sorolhatók azonban egyes kemotaxonómiai bélyegek is, hiszen egy bonyolult szekunder anyagcseretermék szintetizálásának a képessége igen valószínűen csak egyszer következik be az evolúció során, míg ez a képesség könnyen elveszhet, ha bármelyik intermedier előállításának a lehetősége valami oknál fogva kiesik.

5) A fenti típusok azt feltételezik, hogy adott taxon minden egyede megegyezik a kérdéses tulajdonságban. Ha egy populációban adott gén több allélje is jelen van, akkor ez a tulajdonság a fenti módokon már nem írható le, s be kell vezetnünk a *polimorf* karakter fogalmát. Elemzésük viszonylag nehézkes vagy egyáltalán nem megoldható a jelen fejezet módszereivel, s inkább az allélgyakoriságokat is figyelembe vevő genetikai távolságokból célszerű kiindulnunk. A téma legújabb áttekintését Wiens (1995) adja.



6.6 ábra. Karakterállapotok közötti lehetséges átmenetek egyes kladisztikai változók esetén. **a:** rendezetlen, **b:** rendezett és reverzibilis, **c**: irreverzibilis, **d**: a Dollo karakter csak egyszer alakulhat ki (+) az evolúció során, de többször is visszaalakulhat az ősi állapotba (*).

6) Végezetül megemlítjük az ún. *sztratigráfiai* karaktereket is, amelyek fosszilis leletanyagból származó sorrendi (időbeli) információt hordoznak, és Fisher (1992) munkássága nyomán nyertek alkalmazást a kladisztikában. A rétegtani karakterek voltaképpen irreverzibilisek, mivel a leszármazottak nyilván nem lehetnek az ősöknél idősebbek. A legrégebbi réteg 0-val, a következő 1-gyel kódolható, és így tovább.

A kladisztikai karakterek alaptípusainak ismeretében most már hozzáláthatunk a hipotetikus törzsfa megszerkesztéséhez. Két módszercsaládot említhetünk, amelyek közvetlenül a karaktereken alapulnak, a nagyobbik – és fontosabb – csoportot a parszimónia elvet alkalmazó módszerek jelentik, a kisebbikbe pedig a karakter-kompatibilitást értékelő eljárások tartoznak.

6.3.1. Parszimónia módszerek

Általánosságban a parszimónia módszerek az evolúciós fa ágainak összhosszúságát minimalizálják. Más szóval: olyan fát keresnek, amely a lehető legkisebb számú karakterállapotváltozást (evolúciós lépést) teszi szükségessé a leszármazási viszonyok megmagyarázásához. Mielőtt a matematikai részletek ismertetésébe belefognánk, a történetiség kedvéért nézzünk meg egy példát arra, hogy Hennig eredeti "kézi" módszere miképpen működött. Ilymódon a modern eljárásokkal való összehasonlításra is lehetőség nyílik.

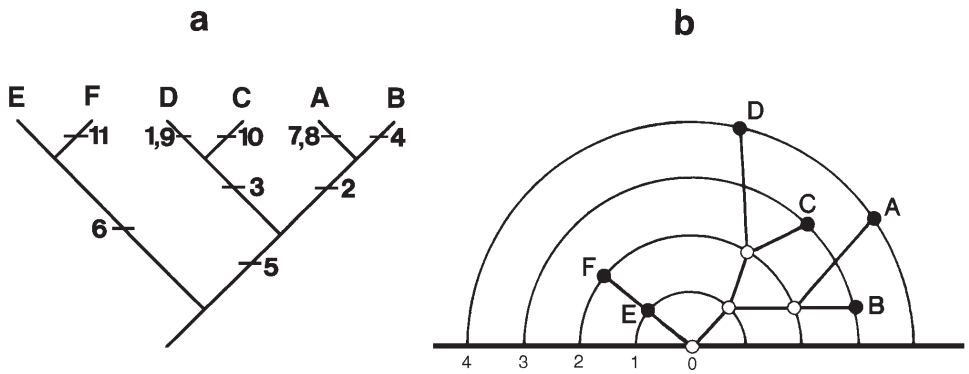
Tételezzük fel, hogy 6 taxonunk van, s ezeket 11 tulajdonság jellemez. Minden karakternek két állapota van, 0 jelöli az ősi, 1 pedig a leszármaztatott állapotot, és az átalakulás irreverzibilis (6.1 táblázat). Az adatokból első látásra megállapítható, hogy sok az autapomorf karakter (1,4,7-11), a megmaradó négy karakterre pedig a következő szünapomorfiát adó csoportok írhatók fel: 2: {A, B}, 3: {C, D}, 5: {A, B, C, D}, és 6: {E, F}. E megoszlásból arra következtethetünk, hogy az első dichotómia az {A,B,C,D} és az {E,F} csoportok között jelentkezett az evolúció során. Ez utóbbiak állanak legközelebb a hipotetikus, "tiszta" 0-val leírható közös őstől, hiszen attól csak 1 ill. 2 karakterben térnek el. A 2. és 3. karakter pedig egyértelműen jelzi, hogy a következő elválás az {A,B} és a {C,D} csoportok között következett be. Ezek után már nem nehéz a három kéttagú csoportot is tovább bontani, hogy megkapjuk a 6.7a ábra kladogramját. A fa egyes ágain az ott megváltozott karakterek sorszámai vannak feltüntetve. A változásokat összeadva megkapjuk a fa éleinek összhosszát, ami éppen a karakterek száma, azaz 11. Némi próbálgatással beláthatjuk, hogy a fa topológiájának bármilyen átalakítása ennél több állapotváltozást tenne szükségessé.

A fentivel teljesen egyenértékű a Wagner-féle (1961) alaprajz/eltérés ("groundplan/divergence") módszer, csupán az ábrázolásmód változik. A koncentrikus félkörök középpontja felel meg a hipotetikus közös ősnek, minden egyes ugrás egy karakter megváltozását jelzi, az üres körök a HTU-k, a telt körök pedig az OTU-k. A közös őstől való távolodás mértéke itt jobban kifejeződik mint a kladogramon, s ugyanakkor az is jól látszik, hogy az E taxon az F-től való elválás után nem is változott, tehát az F ősének tekinthető.

Ez a példa szándékosan olyan egyszerű, hogy a fa megszerkesztése nem okozhatott problémát. Könnyedén megtaláltuk azt a fát, amely homoplázia nélkül, a minimális számú állapotváltozás révén magyarázza meg az evolúciós viszonyokat. A gyakorlatban azonban ritkán van ilyen egyszerű dolgunk, hiszen sokkal több karakterrel ill. OTU-val kell számolnunk, és rendszerint nem létezik olyan fa sem, amelyben ne lenne homoplázia. Ha például a 6.1 táblázatban az A OTU 1. karakterállapotát 1-re módosítjuk, máris gondban vagyunk: az A és D taxonok külön ágon vannak a 6.7a ábra kladogramján, s eszerint a topológia szerint az első karakter apomorf állapota kétszer kellett, hogy kialakuljon az evolúció során, mégpedig egymástól függetlenül. Ez pedig tipikus homoplázia. Ha más topológiájú fát keresünk, amely-

6.1 táblázat. Mesterséges adatmátrix a Hennig módszer illusztrálására. Az utolsó előtti oszlop a leszármaztatott állapotok számát, az utolsó pedig az autapomorfiák számát összesíti az egyes taxonokra.

	Karakterek												
OTUk	1	2	3	4	5	6	7	8	9	10	11	Σ_1	Σ_1
A	0	1	0	0	1	0	1	1	0	0	0	4	2
B	0	1	0	1	1	0	0	0	0	0	0	3	1
C	0	0	1	0	1	0	0	0	0	1	0	3	1
D	1	0	1	0	1	0	0	0	1	0	0	4	2
E	0	0	0	0	0	1	0	0	0	0	0	1	0
F	0	0	0	0	0	1	0	0	0	0	1	2	1



6.7 ábra. A 6.1 táblázat adataiból szerkesztett kladogram a Hennig módszerrel (**a**) ill. a megfelelő Wagner-féle "groundplan/divergence" diagram (**b**).

ben A és D közelebb kerülnek, megszüntetve ezt a "rendellenességet", akkor viszont a 2. és a 3. karakterek fognak homopláziát okozni. Egy lehetséges "megoldás" az 1. karakter kihagyása az elemzésből, de ezt nem igazán tekinti senki sem követendőnek (s ez végeredményben a 6.3.2-ben ismertetendő módszerekhez vezet). Nem kell azonban semmit sem kiiktatnunk, ha a homopláziák megtűrésével a fa ágainak összhosszúságát minimalizáljuk, azaz a *parszimónia* módszert alkalmazzuk. Erre Hennignek és Wagnernek még nem volt lehetősége, hiszen gyors számítógépekről az ő idejükben legfeljebb csak álmodni lehetett. Ma már rendelkezésünkre áll jó néhány számítógépes eljárás, amivel nagy – ha nem is száz százalékos – biztonsággal kikereshetők a legrövidebb ágrendszerű evolúciós fák.

Swofford & Olsen (1990) szerint a parszimónia módszerek célja az összes lehetőség közül megkeresni azt a τ-val jelölt fát, amelyre az alábbi általános optimalitási kritérium értéke minimális:

$$L(\tau) = \sum_{k=1}^{N_B} \sum_{j=1}^{n} w_j \cdot \Delta(x_{k1j}, x_{k2j}) \qquad (6.7)$$

ahol N_B jelöli az ágak számát, n a változók száma, x_{k1j} és x_{k2j} a k-adik ág két végéhez tartozó szögpontok állapota a j-edik karakterre nézve, w_j a j karakter fontosságát kifejező súlyérték (rendszerint 1), $\Delta(x_{k1j}, x_{k2j})$ pedig a két karakterállapot közötti átmenet "költsége". Eme két

karakterállapot vagy közvetlenül az adatmátrix egy konkrét értékének felel meg (az ág megfelelő végén egy OTU van), vagy pedig a fa belső szögpontjaihoz (HTU-k) rendelt állapotról van szó. Az $L(\tau)$ mennyiséget a fa "hosszának" ("*tree length*") nevezzük. Az optimális fa[8] hossza és szerkezete attól függ, hogy milyen állapot-átmeneteket engedünk meg és miként értelmezzük a költségfüggvényt. A feladat – hasonlóan a távolság-alapú módszerekhez – kettős: 1) az adott topológiához legmegfelelőbb (legkisebb hosszúságot eredményező) állapotokat kell rendelnünk a belső szögpontokhoz, és 2) a fa topológiáját kell optimalizálnunk. A topológia változtatása minden karaktertípus esetén ugyanúgy történhet, a belső szögpontokhoz rendelendő állapotok kikeresése azonban már más és más algoritmust igényel. Ezért kell tehát már a vizsgálat legelején tisztáznunk, milyen karaktertípusok szerepelnek az adatmátrixban.

6.3.1.1 Adott fa hosszának optimalizálása

A feladat tehát az, hogy a *h* karakterre a fa végágain elhelyezkedő OTU-k ismeretében meghatározzuk a belső szögpontok (HTU-k) állapotait amelyek minimális hosszt eredményeznek (ez a fa *rekonstrukció*ja). A rendezetlen és a Wagner-féle karaktertípusok esetében – miután a karakterállapotok reverzibilisek – a gyökér helyzete nem befolyásolja az eredményt, s ezt majd ki is használjuk az elemzés során. Az optimalizációs algoritmust, Swofford & Maddison (1987) után, erősen leegyszerűsítve mutatjuk be a rendezetlen (Fitch-féle) karaktertípusra és szigorúan dichotomikus fákra. Az eljárás lényege, hogy az egyik OTU-t gyökérnek tekintve kétszer végigpásztázzuk a fát, először a többi taxontól a gyökérig, majd visszafelé. Ha van olyan OTU, amely önmagában külcsoportot képvisel, akkor célszerűen ezt tekintjük gyökérnek. Az első pásztázás során a belső szögpontokon kijelöljük a szóba jöhető állapotok kombinációit, majd a második fő stádiumban, immár visszafelé haladva a fán, eldöntjük, hogy ezek közül melyiket tartjuk meg.

1) Az OTU-kra nyilván csak egyféle karakterállapotunk lehet, míg a HTU-kra kiindulásképpen nincs megadva karakterállapot, de ezek száma – mint említettük – később ideiglenesen egynél több is lehet. Legyen a gyökérnek tekintett szögpont jele *g*. A fa hossza a *h* karakterre, L_h, legyen kezdetben 0.

2) Válasszunk ki egy *k* belső szögpontot, amelynek mindkét közvetlen leszármazottja ismert állapottal rendelkezik. Jelölje ezeket *i* és *j*. Ekkor az alábbiak között kell döntenünk:

2a) ha van(nak) olyan állapot(ok) mely(ek)re nézve *i* és *j* megegyezik, akkor az összes ilyen állapotot hozzárendeljük *k*-hoz;

2b) ha nincs egy ilyen állapot sem, akkor *k*-hoz az *i* és *j* állapotainak az összességét rendeljük és L_h értéke 1-gyel nő.

3) Ha *k* éppen a *g* közvetlen leszármazottja, továbbmegyünk a 4. lépésre. Egyébként visszatérünk a 2. lépéshez.

4) Ha *g* állapota nem egyezik meg közvetlen leszármazottjának egyik állapotával sem, akkor L_h értéke 1-gyel nő. Az első stádium ezzel befejeződött, s L_h értéke már meg is adja a fa

8 Az is igen gyakran előfordul, hogy a 6.8 optimalitási feltételnek kettő vagy több fa is eleget tesz. Ezek között jelentős eltérések is adódhatnak, s az egyetlen megoldás a 9.4.2 részben részletesen tárgyalt konszenzus elemzés.

hosszúságát a *h*-adik karakterre. Ezután megkezdjük a HTU-k karakterállapotainak kiválasztását, a gyökértől visszafelé haladva.

5) Válasszunk ki egy olyan *k* belső szögpontot, melynek állapotát még nem véglegesítettük, de közvetlen őséét, melyet *o* jelöl, már igen (először tehát a *g*-hez legközelebbi belső szögpontról van szó).

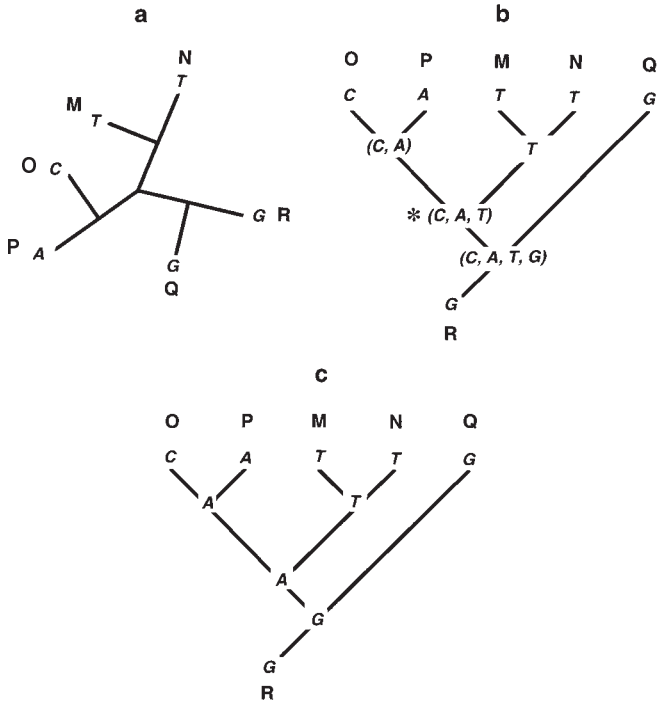
6) Ha az *o* állapota a *k*-hoz rendeltek között is megvan, akkor *k* végső állapota is ez legyen. Egyéb esetben *k* állapotai közül kiválasztjuk az egyiket, s azt tartjuk meg.

7) Ha minden belső szögpontot megvizsgáltunk, akkor a keresés ezennel véget ért. Egyéb esetben visszatérünk az 5. lépéshez.

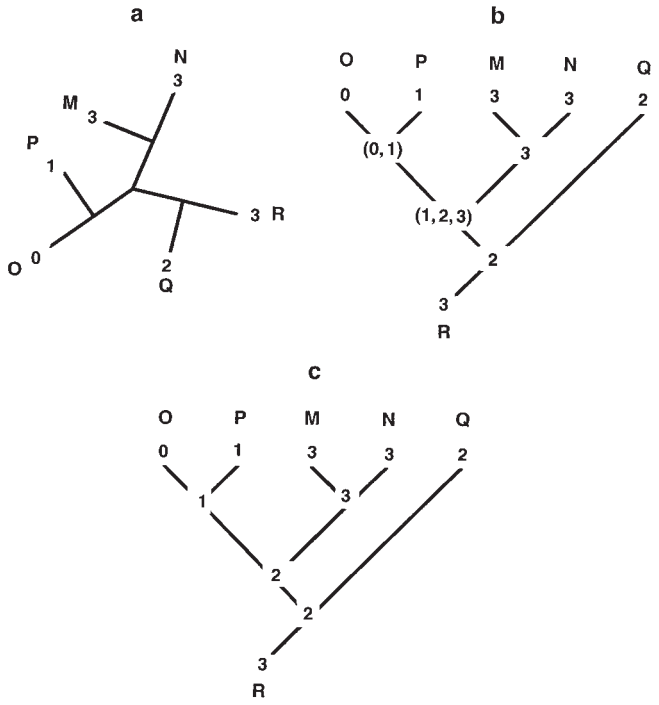
Az algoritmust a legáltalánosabban ismert rendezetlen karakter, valamely nukleinsav molekula egy adott pozíciójában lévő nukleotid milyenségének (mondjuk A, T, G, C) a példáján mutatjuk be (6.8 ábra). A kiinduló fában kiválasztjuk az R taxont, mert ezt tekintjük külcsoportnak (bár az optimalizáció szempontjából ez most nem lényeges) és a 2-4. lépések szerint meghatározzuk a fa hosszát, ill. a belső szögpontok lehetséges karakterállapotait (6.8b ábra). Az elemzés szerint három él mentén kell változásnak bekövetkeznie, azaz L=3. Az utolsó feladat a belső szögpontok állapotainak a kijelölése, amelyet a 6.8c ábra illusztrál. A *-gal jelölt pozícióban önkényesen döntöttünk, de könnyen meggyőződhetünk arról, hogy minden más választásra ugyanúgy 3 lenne a fa hossza. Az önkényes döntés miatt azonban a fának több lehetséges rekonstrukciója is lehetséges (lásd ACCTRAN és DELTRAN: D függelék).

A Wagner-karakterekre, mivel sorrendiséget és különbséget is értelmezünk, a fenti algoritmus 2a, 2b, 4. és 6. lépését kell módosítani a következőképpen:

2a) ha *i* és *j* állapotai átfednek egymással, akkor az átfedést adók legyenek a *k* állapotai (pl. ha *i*-t 1,2,3 ill. *j*-t 2,3,4 jellemzi, akkor *k* állapota 2,3 lesz).



6.8 ábra. A fa hosszának és a belső szögpontok állapotainak a meghatározása egy Fitch-típusú (rendezetlen) karakter esetén (nukleotidok egy adott pozícióban) az M-R taxonokra. **a:** Kiinduló fa önkényesen kiválasztott gyökérrel, **b:** fa az első pásztázást követően, a lehetséges állapotok kombinációival, **c:** végeredményül kapott fa a belső szögpontok optimális állapotaival.



6.9 ábra. A belső szögpontok állapotának meghatározása Wagner karakterek esetén. **a-c:** mint a 6.8 ábrán.

2b) ha nincs átfedés, akkor a két legközelebbit és a közöttük lévő többi állapotot rendeljük *k*-hoz , *L* pedig a két legközelebbi állapot különbségével nő (pl. ha *i*-t 1,2,3 ill. *j*-t 5,6 jellemzi, akkor *k* ideiglenes állapota 3,4,5 lesz, L_h értéke pedig 2-vel növekszik)

4) Ha *g* állapota nem egyezik meg a közvetlen leszármazottjának egyik állapotával sem, akkor L_h új értéke L_h+ | *g* állapota – a legközelebbi állapot a leszármazottban | .

6) *k* állapotai közül kiválasztjuk azt, amelyik *o* állapotához a legközelebb van (vagy azzal egyenlő) s azt tartjuk meg.

Mindez érthetőbbé válik a 6.9 ábra példáján. Tegyük fel, hogy hat taxont most egy négy állapotú rendezett reverzibilis karakter jellemez, amelyet a 0, 1, 2, és 3 értékekkel kódolunk (6.9a ábra). Az R taxont gyökérnek választva megint elindulunk felülről, s ideiglenes kombinációkat rendelünk a belső szögpontokhoz (6.9b ábra). A 2a) lépést alkalmazzuk a 3 illetve a 2 állapot, a 2b) lépést pedig a (0,1) és az (1,2,3) kombinációk megválasztásakor. A fán visszafelé haladva meghatározzuk a végső értékeket. A fa hossza egyébként 4 egység.

Az optimalizációt a többi karakterre is végrehajtjuk, és végül ΣL_h lesz a fa teljes hossza. Az összegzés természetesen eltérő típusú karaktereket is megenged.

A többi kladisztikai karakterre alkalmas, illetve a többszörös elágazást is megengedő parszimónia algoritmusok meglehetősen bonyolultak, ismertetésüket ezért mellőzzük. Alkalmazásuk amúgy sem megy a megfelelő programcsomag nélkül, így a részletekért a felhasználói kézikönyvet kell fellapoznunk (pl. Maddison & Maddison 1992, Felsenstein 1993).

6.3.1.2 Evolúciós fák topológiájának optimalizálása

Ha egy adott fa minden belső szögpontjára megtaláltuk a legmegfelelőbb karakterállapotokat, akkor a probléma kisebbik részét oldottuk csak meg. A 6.7 optimalitási kritérium ugyanis jóval nagyobb mértékben függ az elágazások topológiájától, mint a karakterállapotok elosztásától. A legjobb topológia kikeresése azonban további nehézségeket támaszt, amint az alábbi rövid ismertetésből is kiderül.

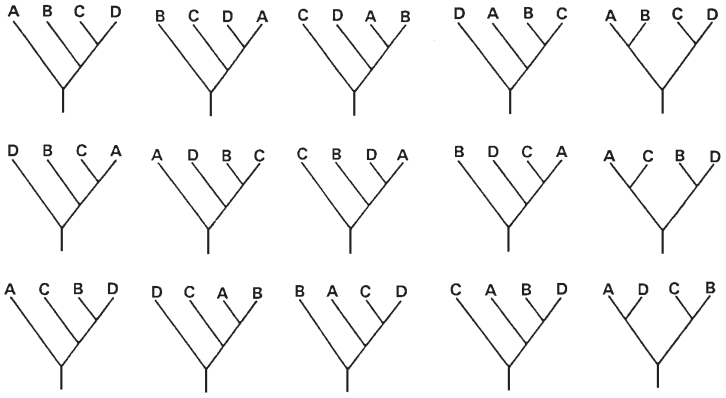
Teljes enumeráció. Elsőként az a megoldás juthat eszünkbe, hogy az összes lehetséges fát "legyártjuk", és mindegyiket megvizsgáljuk az előző részben ismertetett módon. Ekkor biztosak lehetünk abban, hogy a 6.7 kritériumra minimális értéket adó fa a legmegfelelőbb (a parszimónia elv alapján, legalábbis). Az összes lehetőség megvizsgálása azonban nem is olyan egyszerű feladat, amint első pillantásra látszik. Már említettük az 5. fejezetben, hogy milyen irdatlan nagyszámú különböző dendrogram írható fel már 10 objektumra is (5.16 formula), s ez a szám megegyezik a gyökérrel rendelkező kladogramok lehetséges számával *(m*=10 esetén, mint láttuk, több, mint 34 millió). Ha a gyökeret kiiktatjuk, akkor a következő összefüggés adja meg a lehetőségek számát:

$$\prod_{i=3}^{m}(2i-5)=\frac{(2m-5)!}{2^{m-3}(m-3)} \qquad (6.8)$$

(Felsenstein 1978). Még ez is igen nagy szám lehet, hiszen *m*=10-re meghaladja a kétmilliót. A valóságban rendszerint jóval nagyobb számú taxonnal dolgozunk, amelyre már csillagászati számok jönnének ki, így az összes lehetőség számbavétele gyakorlatilag lehetetlenné válik.

A teljes enumeráció egyébként a gyökér nélküli fákra a 3 objektumra felrajzolható egyetlen egy lehetséges fából indul ki, amelyben 3 él van. A következő taxont e 3 él bármelyikére helyezhetjük, vagyis *m*=4-re három különböző elrendezés adódik. Ezen a fán már öt él lesz, ami az 5. objektum elhelyezési lehetőségeinek a száma, és ez szorzódik az *m*=4-re kapott fák számával: 3×5=15 (6.10 ábra). Minden egyes taxon hozzáadásával az előző lépésben előállított fák száma 2*i* –5-tel szorzódik (*i* a taxonok száma az adott lépésben), s így már jobban érthető a fenti formula jelentése.

Exakt módszerek. Felmerül tehát az igény, hogy olyan algoritmust keressünk, amely nem vizsgál meg minden lehetőséget, de relatíve rövid idő alatt mégis eljut a legkedvezőbb megoldásig.



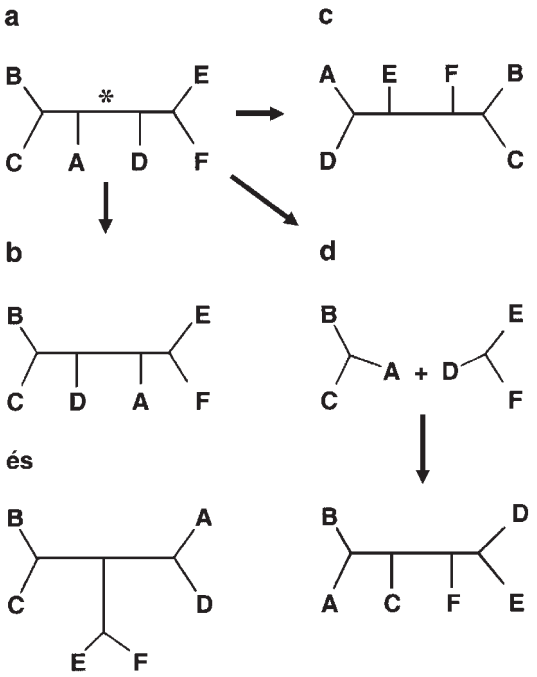
6.10 ábra. A négy OTU-ra felrajzolható összes lehetséges dichotomikus kladogram.

Ezt célozza az 5.3.1 részben már említett "branch and bound" módszer, melynek első kladisztikai alkalmazása Hendy és Penny (1982) nevéhez fűződik. Kezdő összehasonlítási alapként egy olyan fát választunk, amelyet – mondjuk – az alábbiakban ismertetett heurisztikus módszerekkel kaptunk, így az viszonylag közel áll az optimálishoz. Legyen ennek hossza L_{min} (a "bound"). Ezután "0-ról" indulunk, mintha teljes enumerációt akarnánk véghezvinni a fent leírt módon. A fa hosszát azonban menet közben minden "rész"-fára kiértékeljük, és ha az túllépi L_{min} értékét, akkor a kereséssel ebben az irányban ("branch") már nem próbálkozunk tovább. Voltaképpen minden olyan fa, amelynek ez a részfa alkotóeleme, egyszer és mindenkorra kiesik, hiszen a továbbépítés során ezen fa hossza már csak növekedhet. Ha azonban felépül egy teljes fa, amelynek hossza kisebb, mint L_{min}, akkor már javítottunk is a kiinduló eredményen. A további keresés során természetesen ez az új L_{min} lesz a viszonyítási alap. Ebből a pár mondatos jellemzésből – amely persze nagyon távol áll az algoritmus pontos ismertetésétől – talán belátható, hogy a módszer a legrosszabb esetben éppen a teljes enumerációval egyezik meg, de ha a kezdő L_{min} igen közel áll az abszolút optimumhoz, akkor sokszorta hatékonyabb annál. A módszer legjobb számítógépes implementációi sem képesek azonban több, mint 20-25 taxon értékelésére, hiszen a gépidő rendkívül gyorsan növekszik m növekedésével.

> Nincs tehát garancia arra, hogy a "branch and bound" módszer akármilyen kiindulásból belátható időn belül eredményre vezet mondjuk 100 taxonra. Ilyen módszert voltaképpen még nem ismerünk. A legoptimálisabb topológia megkeresése ugyanis egy, a matematikában már régen vizsgált témakörbe, az *NP-teljes* problémák körébe tartozik (Graham & Foulds 1982). Arról van lényegében szó – persze matematikailag elnagyoltan –, hogy egy adott számítási feladat megoldásához szükséges idő hogyan változik m növekedésével. Átlagos többváltozós elemzések során az idő négyzetesen vagy köbösen növekszik (pl. hierarchikus klasszifikáció, stb.) és ez a mai számítógépek gyorsaságát ismerve még könnyen elviselhető. Az optimális fa megtalálására azonban, ha m egy bizonyos határt elér, az idő növekedése hirtelen kezelhetetlenné válik, *nem-polinomiális* összefüggés szerint változik (innen: NP). Kimutatták, hogy ha bármely NP-teljes problémára sikerülne egy gyors algoritmust találni, akkor az összes NP-teljes probléma megoldható lenne vele (Lewis & Papadimitriou 1978).

Heurisztikus eljárások. Nagyszámú taxon esetén el kell fogadnunk tehát azt a tényt, hogy nem ismeretes olyan módszer, amely biztosan megtalálja a legjobb topológiájú fát (Day 1983). Csak abban bízhatunk, hogy a heurisztikus, keresgélős/iterációs stratégia relatíve gyorsan kellő közelségbe juttat minket az abszolút optimumhoz. E módszerek sokban hasonlatosak a nem-hierarchikus osztályozás k-közép módszeréhez (és még más, a későbbi fejezetekben sorra kerülő eljárásokhoz): valamilyen kiinduló eredményt javítgatunk bizonyos átalakítások segítségével, és ha már további javulás nem érhető el, leállunk az elemzéssel. Célszerű azonban többféle kiindulást is kipróbálni, mert a végeredmény erősen függhet a kezdő konfigurációtól. A sok lokális optimumból kiválaszthatjuk a legjobbat, tudva persze, hogy ez sem feltétlenül az abszolút optimális eredmény.

Kladogramok esetében kétféle iterációs stratégia között dönthetünk. Ez egyik lehetőség a fa *fokozatos felépítése* egy-egy taxon hozzáadásával. Kiindulásképpen véletlenszerűen (vagy a fa hosszát minimalizálva) kiválasztunk három taxont. Az első lépésben minden egyes további taxont végigpróbálgatunk az összes lehetséges helyen, s megvizsgáljuk, hogy menynyivel növekedett a fa hossza. Azt az esetet tartjuk meg, amelyre minimális volt a növekedés. A következő lépésben újabb taxont "ragasztunk" a fához, s ezt a fa teljes felépüléséig folytat-

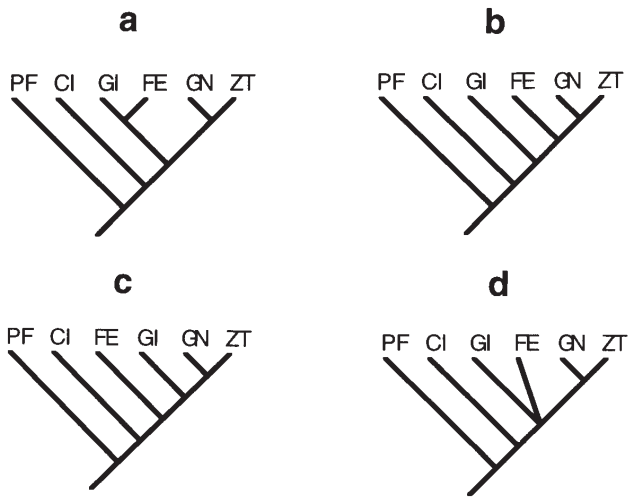


6.11 ábra. Az **a** kladogram átrendezésének lehetőségei. **b:** szomszéd ágak felcserélése (a *-gal jelölt élre nézve), **c:** egy ág átoltása (a B-C részfát tettük át az F-hez futó élre), **d:** a fa elvágása két részfára és összekapcsolása új éllel (a *-gal jelölt élt megszüntetjük, majd a C-hez és F-hez futó két élt összeköt-jük).

juk. E módszereknél (akárcsak az agglomeratív osztályozásnál) az a gond, hogy egy adott taxon pozíciója a későbbiek során már nem változtatható meg. Erre azonban jó megoldást ad a fa *iteratív átrendezése*, amely alapvetően háromféle stratégiát követhet:

- *Legközelebbi szomszéd felcserélése.* A fa egy-egy belső éléhez tartozó rész-fákat egymással felcserélve (6.11a-b ábra) kis lépésekben érhetünk el javulást. Minden ilyen élhez négy részfa csatlakozik, s miután ezek háromféleképpen rendezhetők el, a kipróbálandó új lehetőségek száma kettő.
- *Ágak "átoltása".* A fa összes lehetséges rész-fáját áthelyezzük az összes lehetséges helyre minden egyes lépésben (egy ilyen áthelyezést mutat be a 6.11c ábra).
- *"Metszés" és újraegyesítés.* A fát minden lehetséges helyen kettévágjuk, az elvágott élt megszüntetjük, s a kapott részfákat minden lehetséges módon újra összekötjük (pl. 6.11d ábra). E két utóbbi procedúra hirtelen nagy javulást is eredményezhet egy-egy lépésben.

Példaképpen először a 6.1 táblázat adatait vizsgáljuk meg. A **PHYLIP** programcsomag **MIX** programja (Felsenstein 1993) egyértelműen megerősítette a 6.7a ábrán látható kladogramot. Ennél a 11-es hosszúságú fánál jobbat, vagy akár azzal megegyező hosszúságú, de más topológiájú fát sem talált. Rendszerint azonban nem ilyen egyértelmű a helyzet, amint azt az A6 táblázat adatainak elemzése is igazolja. A táblázat alapján 5 magvas taxon leszármazási viszonyait próbáljuk rekonstruálni a páfrányok (mint külcsoport) bevonásával. Az összes tulajdonság bináris típusú, tehát mindegy, hogy Fitch- vagy Wagner-karakternek fogjuk fel őket. A **MIX** program 50 random kiindulásból három optimális hosszúságú fát adott eredményül (6.12a-c ábra). A felhasznált információk alapján a fenyők és a *Ginkgo* helyzete nem egyértelmű, felcserélhetők egymással s akár egy külön csoportot is alkothatnak. Általános tapasztalat, hogy minél nagyobb a vizsgálatba bevont taxonok száma, annál több egyformán op-



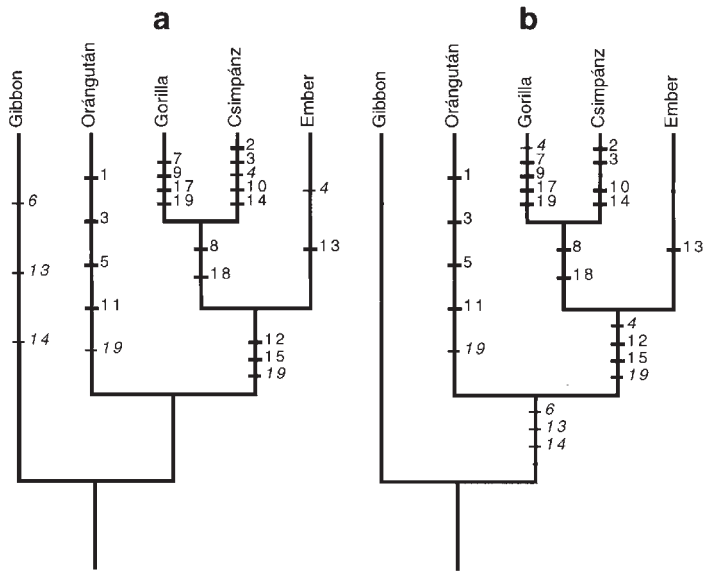
6.12 ábra. A magvas növények csoportjainak három optimális hosszúságú ("equally parsimonious") kladogramja az A6 táblázat adataiból kiindulva (**a-c**) és ezek szoros konszenzus kladogramja (**d**). PF: páfrányok (külcsoport), CI: cikászok, GI: *Ginkgo*, FE: fenyők, GN: *Gnetum*, ZT: Zárvatermők.

timális hosszúságú, de egymástól eltérő topológiájú fa adódik eredményül. E fák az ún. *konszenzus* módszerek segítségével (9.4.2 rész) egy újabb kladogram formájában összegezhetők, s ezt a konszenzus kladogramot fogadjuk el végeredményül. A 6.12d ábra – egy helyen politomikus – kladogramja adja a másik három fa egy lehetséges (ún. "strict consensus") szintézisét. Az evolúciós viszonyok értelmezését az Olvasóra bízzuk.

A következő példa a molekuláris információn alapuló törzsfa-keresést illusztrálja. Az alábbi táblázatban az ember és négy főemlős két mitokondriális tRNS génjének az eltéréseit összesítjük, az első öt oszlop a LEU tRNS-re, a többi pedig a SER tRNS-re vontakozik (Brown et al., 1982, adatai alapján). A két RNS szakasz összhossza 131 nukleotid. A nukleotid pozíciók túlnyomó többségében a fajok megegyeznek, ezeket az egyszerűség kedvéért be sem mutatjuk, hiszen egyáltalán nem befolyásolják az eredményt, a pozíciók számozása ezért teljesen önkényes (Megjegyzendő, hogy az orángutánnál bekövetkezett nukleotid-kiesés ("gap") sem számít majd bele a fa hosszába). A számunkra lényeges információk az alábbiak:

	Pozíció																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Ember	A	T	A	C	C	T	A	C	A	C	A	T	G	C	C	C	A	T	C
Csimpánz	A	C	G	C	C	T	A	T	A	T	A	T	A	T	C	C	A	C	C
Gorilla	A	T	A	A	C	T	G	T	G	C	A	T	A	C	C	C	G	C	T
Orángután	G	T	C	A	T	T	A	C	A	C	T	C	A	C	T	.	A	T	G
Gibbon	A	T	A	A	C	C	A	C	A	C	A	C	T	A	T	C	A	T	A

A **PHYLIP** programcsomag **DNAPARS** programja (Felsenstein 1993) ill. a **MacClade** program (Maddison & Maddison 1992) is egyetlen egy, 24-es hosszúságú fát talált a legoptimálisabbnak. A gyökér pozícióját külső információ figyelembevételével állapítottuk meg, hiszen a gibbon számos szempontból a többitől eléggé távoli taxonnak tekinthető (6.13 ábra). Az ábrázolás most szándékoltan dendrogram-szerű, hogy megkönnyítsük a karakterváltozások jelölését. Az orángutánhoz futó élen pl. az 1-es jel azt indikálja, hogy a többihez képest e fajnál következett be változás az 1. pozícióban (A helyett G), a csimpánz neve alatt a 2-es pedig a második pozícióbeli váltásra utal (T helyett C), és így tovább. A nukleotid váltások "múltja" a pozíciók többségében egyértelműen kijelölhető, de a 4., a 6., a 13., a 14. és a 19. esetében voltaképpen önkényesen kell döntenünk (6.3.1.1). A 6.13a és b ábrák két ilyen döntési alter-



6.13 ábra. Az ember és a főemlősök evolúciós kapcsolatának rekonstrukciója RNS parszimónia módszerrel a mitokondriális LEU tRNS és SER tRNS gének nukleotidszekvenciái alapján. A két kladogram egyes, önkényesen kijelölhető nukleotidváltásokban különbözik csupán.

natívát mutatnak be, mindkét esetben az ágakon feltüntetett változások száma azonos (24). Messzemenő következtetéseket persze nem szabad levonnunk ebből a kladogramból, hiszen az elemzést egy relatíve rövid RNS-szakaszra alapoztuk csupán (a HIS tRNS gén alapján egyébként a csimpánz az emberhez áll közelebb, l. Weir 1990). Meg kell azt is jegyeznünk, hogy a nukleotidcserék során egyformán fontosnak vettük a *tranzíció*kat (A-G, ill. C-T cserék, azaz hasonló szerkezetű nuklotidok cseréjét), mint a *transzverzió*kat (azaz amikor egy purinvázas nukleotid pirimidinvázasra cserélődik, vagy fordítva). A valóságban azonban, bár az utóbbi esetben a lehetőségek száma kétszer akkora, kémiai okokból a tranzíciók sokkal gyakoribbak a transzverzióknál. (Példánkban a fa 24-es összhosszúságából mindössze 6 eltérés magyarázható transzverzióval.) Ezt súlyozással lehet kiegyenlíteni (pl. Williams & Fitch 1990, Williams 1992).

6.3.1.3 Kladogramok értékelése

A – nem Dollo-típusú – karakter-alapon számított kladogramokat néhány egyszerű index segítségével értékelhetjük ki. Kluge és Farris (1969) javaslata szerint például minden egyes karakterre érdemes megvizsgálni, hogy a változások száma hogyan aránylik az elméletileg elképzelhető minimumhoz. Ha az adott fában a *j*-edik karakter éppen s_j változást mutat, holott az adatokra felírható egy olyan fa is, amelyben a minimális számú m_j változás következik csak be, akkor a

$$CI_j = \frac{m_j}{s_j} \tag{6.9}$$

hányados (*konzisztencia-index*) fejezi ki a keresett arányt. *CI* értéke 1, ha a fában a lehetséges minimum fordult elő, melynek jelentése: a karakter nem utal homopláziára. Minden egyéb érték homopláziát jelez, a CI_j=0,5 érték pl. azt mutatja, hogy éppen kétszer annyi változás következett be, mint amennyire minimálisan szükség van. A konstans karakterekre, a 0/0 miatt, az index nem értelmezhető.

A 6.13 ábra kladogramján mindössze egy karaktert találunk, amire *CI* értéke nem 1, ez pedig a 4. nukleotidpozíció (CI_4=0,5). A 6.13a kladogram szerint itt az A-t az embernél és a csimpánznál egymástól függetlenül C váltotta fel, a 6.13b fa szerint pedig az A (az ősi állapot) visszafordulással jelent meg újra a gorillánál, mert időközben már C került arra a helyre. (Mindkét eset elképzelhető a szerkesztés során óhatatlanul felmerült önkényes döntések miatt.) A 4. karakter csak egy változást mutatna, ha a gibbon, az orángután és a gorilla ugyanazon az ágon, a csimpánz és az ember pedig egy másik ágon lenne, de mivel ez nem így van, 2 lépésre volt szükség. Mindez persze csak az index *alkalmazásának illusztrációja*, hiszen a szekvencia adatok esetében távolról sem olyan valószínűtlen a visszafordulás, vagy a parallel előfordulás, mint mondjuk morfológiai bélyegek esetében, vagyis a homoplázia jelensége itt másként értelmezendő.

Az összes karakterre kiszámítható az *átlagos konzisztencia index* is:

$$CI(\tau) = \sum_{j=1}^{n} m_j \Big/ \sum_{j=1}^{n} s_j \qquad (6.10)$$

amely a példánkban 0,96 (a 16. pozíciót nem vettük figyelembe a kiesés miatt). Maddison & Maddison (1992) szerint azokat a karaktereket is mellőznünk kell, amelyek autapomorfiát mutatnak, hiszen esetükben *CI* értéke eleve nem lehet más, mint 1,0. Ezek bevonása csalóka módon felfelé torzítaná az átlagos konzisztencia-indexet.

Más szemszögből értékeli a karakter "viselkedését" a Farris-féle (1989) *összetartási index*, mert ez figyelembe veszi a lehetséges megváltozások maximális számát is, amit M_j jelöl:

$$RI_j = \frac{M_j - s_j}{M_j - m_j} \qquad (6.11)$$

Ennek értéke annál magasabb, minél kevesebb a homoplázia részesedése a szünapomorfiák kialakulásában. RI_j=1, ha egyáltalán nincs homoplázia, és RI_j=0, ha az összes szünapomorfiát homoplázia okozza. E függvény már nemcsak a konstans karakterekre ad 0/0-t, hanem az autapomorfiát mutató bélyegekre is. Így csak olyan esetekben van értelme kiszámítani, amikor homoplázia egyáltalán kialakulhat, vagyis a minimum és a maximum nem egyezik meg (a nevező nem nulla).

A szekvencia-példánkban 5 karakter jöhet szóba *RI* kiszámítására. A 4. karakter a maximálisan lehetséges két változást adta (RI_4 = (2–2) / (2–1) = 0). A 8., 12., 15. és a 18. pozíciók esetén lehetett volna még elképzelni homopláziát, de mindegyik esetben ettől mentes, "valódi" szünapomorfia alakult ki (a leszármaztatott állapot egy-egy kládon "összetartott"), s mindegyikükre RI_j = (2–1) / (2–1) = 1.

Az alábbi, ún. *együttes összetartási index*be is értelemszerűen csak azok a karakterek vonhatók be, amelyekre $M_j > m_j$:

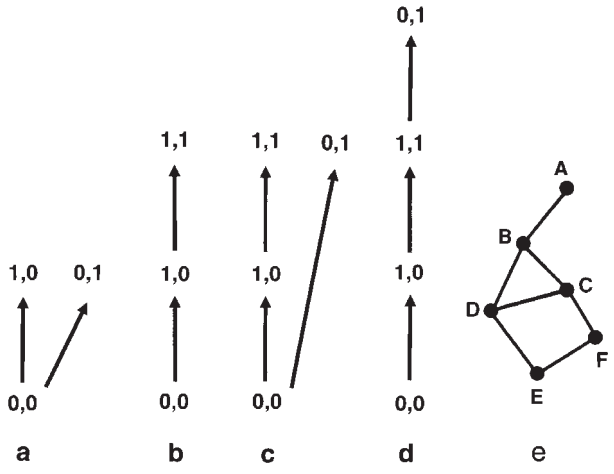
$$RI(\tau) = \frac{\sum_{j=1}^{n} M_j - s_j}{\sum_{j=1}^{n} M_j - m_j} \qquad (6.12)$$

Ennek értelmezése hasonló a karakterenkénti *RI* értelmezéséhez. A 6.13 ábra kladogramjára $RI(\tau)$ értéke 4/5 = 0,8.

6.3.2 A karakter-kompatibilitás elemzése

A parszimónia módszerek alternatívája a karakter-alapú kladisztikában a LeQuesne (1969, 1972), Estabrook et al. (1976) és mások által kidolgozott *kompatibilitás analízis.* A feladat itt is az evolúciós fa rekonstruálása, azzal a döntő különbséggel, hogy a homopláziát okozó bélyegeket ("hamis" karakterek) teljes mértékben kiszűrjük a vizsgálatból, s csak azokat tartjuk meg, amelyek nem mondanak ellent egymásnak (kompatibilisek). A feladat központi része az ilyen karakterek lehető legnagyobb részhalmazának a kikeresése, amelyen a kladogram szerkesztés alapszik.

A kompatibilitás alapelvét az alábbi egyszerű illusztrációval igyekszünk világosabbá tenni. Tételezzük fel, hogy az A és B karakterek összeférhetőségét szeretnénk megállapítani; mindegyiküknek két állapota lehetséges, 0 az ősi és 1 a leszármaztatott. Kezdetben csak a (0,0) kombináció fordult elő a vizsgált taxonok körében (amelyek természetesen más karakterekben többé-kevésbé különböztek egymástól). Az evolúció során először az A karakter változott meg a 0→1 módon, s így megjelent az (1,0) kombináció. A későbbiek során a B karakter is evolválódik, mégpedig vagy a (0,0) ősi kombinációból kialakítva a (0,1) kombinációt, vagy pedig az (1,0)-ból tovább "fejlődve" az (1,1) kombinációt (6.14a-b ábra). Az a lényeg, hogy vagy csak a (0,1) vagy az (1,1) kombináció fordulhat elő a vizsgált taxonok között, mindkét kombináció jelenlétéhez ui. az kell, hogy a B karakter 1-es állapota kétszer, egymástól függetlenül jelenjen meg parallel evolúció révén (6.14c ábra), vagy pedig az A karakter visszafordulást mutasson (6.14d ábra). Ha ezeket, az evolúciós utak feltárásában zavaró jelenségeket – vagyis a homopláziát – kizárjuk, akkor a négy lehetséges kombináció közül legfeljebb három fordulhat csak elő a vizsgált csoportban. A két karaktert tehát akkor tekinthetjük kompatibilisnek, hogyha nem találtuk meg az összes kombinációt. A kombinációk értékelését minden párosításban elvégezzük, s az eredményt egy kompatibilitási gráfban (6.14e ábra) összesítjük. Ebben a gráfban a szögpontok a karaktereknek felelnek meg, s az egymással kompatibilis karaktereket él köti össze. A gráfból kikeressük a legnagyobb teljes részgráfot ("clique",



6.14 ábra. Két bináris karakter kompatibilis egymással, ha állapotaiknak legfeljebb csak három kombinációja fordul elő a vizsgált taxonok között (**a-b**), mert a negyedik kombináció megjelenése csak homoplázia révén lehetséges (**c-d**). A további elemzésre alkalmas karaktereket a kompatibilitási gráf maximálisan összekötött részgráfjának a kikeresésével választjuk ki (**e:** B,C,D).

amiben minden pont össze van kötve az összes többivel), s ezeket a karaktereket vesszük csak tekintetbe a fa szerkesztésénél (a 6.14e ábrán a B, C és D). Ide vehetjük még az – eddigi páros összehasonlításokból nyugodtan kihagyható – autapomorfiát mutató karaktereket, amelyek minden ilyen részgráfnak elemei, mert eleve nem adhatnak négy kombinációt.

A parszimónia algoritmusokkal ellentétben könnyen előfordulhat, hogy a tulajdonságok jelentős részétől meg kell "szabadulnunk", ami a taxonómusok legfőbb ellenérve a kompatibilitás elemzésével szemben. Meacham & Estabrook (1985) úgy találták, hogy az addig publikált kutatások során általában a tulajdonságok felét kellett kihagyni, de volt, amikor majdnem 90 %-át! További problémát jelent az, hogy a módszer csak bináris karakterekre alkalmas, és a kapott fa rendszerint politomikus. Alkalmazásainak száma – a vonzó elméleti megalapozás ellenére – szinte elhanyagolható a parszimónia módszerekéhez képest. A módszer további részletezésétől ezért eltekinthetünk. Az érdeklődők pl. Mayr & Ashlock (1991: 307-313) könyvében találnak példát a teljes számításmenetre.

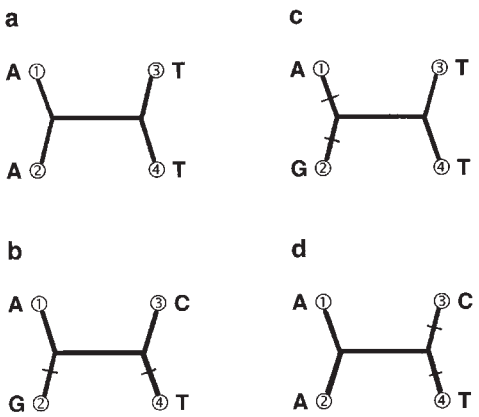
6.4 Nukleinsav-szekvenciák elemzésének egyéb lehetőségei

Mint láttuk, nukleinsav-szekvencia adatokra a távolság- és a karakter alapon működő módszerek is alkalmasak, és ezzel még távolról sem zártuk le a kipróbálható lehetőségek körét. A teljesség kedvéért röviden megemlítünk két olyan eljárást is, amelyek – a speciális alkalmazási terület és más okok miatt – nem illeszkednek az előző fejezetek tematikájába. A hangsúly is eltolódik: a fa topológiáját optimalizáló algoritmusok helyett a nukleotidátmenetek megfelelő interpretálása ill. modellezése kerül előtérbe.

6.4.1 Az invariánsok módszere

A transzverziók és tranzíciók problémáját már röviden említettük a 6.3.1.2 rész végén is, a mitokondriális tRNS gének példájával kapcsolatban. Míg a parszimónia-algoritmusok nem tesznek különbséget az egyes átmenetek között[9], az invariánsok módszere (Lake 1987) a leszármazási viszonyok feltárásában kizárólag a transzverziókra épít. Sőt, még ennél is tovább megy: egyidejűleg csupán *négy* szekvenciát tud értékelni, s csak azokat a pozíciókat veszi tekintetbe, amelyeken két szekvenciában *purin*-, a másik kettőben pedig *pirimidin*-vázú nukleinsav található. A fa végágaira összpontosít, s a rajtuk végbemenő transzverziókat negatív előjellel veszi figyelembe. A négy szekvenciára felírható három lehetséges gyökér nélküli fát alaposan megvizsgálja minden egyes értékelhető pozícióra és egy speciális pontrendszer segítségével választja ki közülük a legmegfelelőbbet. Anélkül, hogy a módszer teljes bemutatására törekednénk, érdemes a pontozási szisztémát röviden illusztrálni. Tegyük fel, hogy az éppen értékelt fában az 1. és a 2. taxon van egy ágon, ill. a 3. és a 4. pedig a másikon. Ha az első két taxonnak azonos purinbázisa van, s emellett a 3. és 4. taxonnak pedig azonos a pirimidinbázisa, akkor ez a pozíció támogatja a kérdéses fát (6.15a ábra). Az 1. és 2. taxon közös leszármazása ugyanis igen valószínű, mert bármely más topológiára két végágon azonos jellegű transzverziót kellene feltételeznünk, s ez már jóval valószínűtlenebb (tévedés forrása,

9 Megtehetjük persze, hogy a tranzíciókat egyszerűen kihagyjuk a parszimónia elemzés során. A "global parsimony" elemzéssel szemben, a "transversion parsimony" módszere csak a transzverziókra összpontosít (pl. Cracraft & Helm-Bychowski 1990).



6.15 ábra. A Lake-féle módszer annak eldöntésére, hogy egy adott pozíció támogatja-e (**a-b**) vagy ellenzi (**c-d**) az 1-4 szekvenciák bemutatott leszármazási viszonyait.

ha mégis ez történt, de ez a hibalehetőség elkerülhetetlen). Hasonló a helyzet akkor is, ha 1. és 2. szekvencia különböző purinnal, a 3. és 4. pedig különböző pirimidinnel rendelkezik (ui. csak tranzíciókat kell két végágon feltételeznünk, 6.15b ábra). Ha azonban az első két szekvenciában eltérő purinbázis van, míg a második két szekvenciában azonos pirimidinbázis található, akkor az illető pozíció "ellene van" a kérdéses topológiának, hiszen ennek kialakulásához két, parallel (bár nem azonos) transzverzióra lenne szükség (6.15c). A fordított eset (6.15d ábra) ugyanez okból szintén ellenszavazatnak tekinthető. A negatív és pozitív "szavazatok" pozíciók szerinti összegzése után kiderül, hogy melyik fát támogatja a pozíciók többsége. Jelentős hátrány azonban, hogy négynél több szekvenciára nincs még megfelelő algoritmus (Swofford & Olsen 1990:474).

6.4.2 A maximum-likelihood módszer

Ennek az eljárásnak az alkalmazása már egy konkrét evolúciós modell alkalmazását igényli: az evolúciós mintázat feltárásához pontosan meg kell adnunk, hogy miképpen alakulhat át az egyik szekvencia a másikba (morfológiai karakterekre ilyen célra használható általános modellről még nem tudunk). A maximum likelihood módszer a modell ismeretében megadja, hogy a sok lehetőség közül melyik fa kialakulása a leginkább valószínű (a fa megváltoztatása nem része a modellnek, ez a 6.3.1.2 részben ismertetett módokon történhet). A legegyszerűbb az ún. Jukes & Cantor modell (Felsenstein 1981), miszerint a bázisok gyakorisága azonos és minden nukleotidcsere egyformán valószínű. A Kimura-féle kétparaméteres modell a k tranzíció/transzverzió hányados bevezetésével már különbséget tesz a behelyettesítések kétféle alaptípusa között. Ennek általánosított változata a nukleotidgyakoriságok eltérését is megengedi (Kishino & Hasegawa 1989). A számítások során a teljes szekvenciát figyelembe kell vennünk, nemcsak az eltéréseket okozó pozíciókat (ahogy a parszimónia esetben tettük). A gyakoriságokból és k-ból meghatározható a modell "szíve", egy 4×4-es mátrix, amely a *nukleotidcserék* rátáit tartalmazza az evolúciós időegységre vonatkoztatva. A mutációs ráták segítségével kiszámítható annak az eseménynek a valószínűsége, hogy t idő elteltével mondjuk az A bázis helyére a G bázis kerül (a részleteket lásd pl. Swofford & Olsen 1990:477-478). Jelöljük ezt a valószínűséget $P_{AG}(t)$-vel. Annak az esélye (L=likelihood!), hogy adott szekvencia valamely pozíciójában az A nukleotid van, s ezt t idő elteltével G váltja fel, a következő:

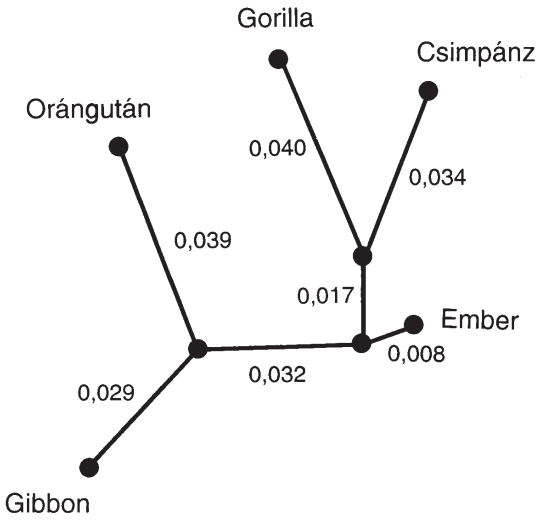
$$L_{AG}(t) = f_A P_{AG}(t) \tag{6.13}$$

ahol f_A az A nukleotid relatív gyakorisága a kezdeti szekvenciában. Ha feltételezzük, hogy a szekvencia minden egyes pozíciója függetlenül változik a többitől az evolúció során (bár ez a valóságban nem így van, vö. Weir 1990), akkor annak az esélye, hogy az X szekvenciából t idő elteltével éppen az Y szekvenciát kapjuk, a következő likelihood-függvénnyel kapható meg:

$$L_{XY}(t) = \prod_{i=1}^{s} f_{x_i} P_{x_i, y_i}(t) \tag{6.14}$$

ahol s a két lánc hosszúsága (ezek tehát egyformák; helyesebben: az esetleges nukleotid-kieséseket a modell nem kezeli), x_i és y_i pedig az i-edik pozícióban található nukleotid az X ill. az Y szekvenciában (vagyis *A, G, C* vagy *T* (*U*)). Miután ez rendszerint igen kicsiny szám, célszerű az ln $L_{XY}(t)$ átalakítás, így a számítások is jelentékenyen leegyszerűsödnek.

A 6.13 függvény voltaképpen az X és Y molekulák *hasonlóságának* tekinthető, minél nagyobb a likelihood, annál közelebb áll a két szekvencia egymáshoz. Most már "csupán" az a kérdés, hogy miképpen térünk át a kettőnél több szekvencia rokonságát kifejező kladogram megvalósulási esélyének a kiszámítására. Anélkül, hogy a komplikált számításmenetet részleteznénk, megemlítjük, hogy a fa egy-egy taxon hozzáadásával épül fel, minden pozícióra külön-külön ki kell számítani az átmenet valószínűségét a már meglévő részfák között, s az utolsó szorzat adja teljes fa likelihood értékét. A feladat egy olyan fa megtalálása, amelyre a szorzat maximális. Ez a fa mutatja a legvalószínűbb leszármazási mintázatot, feltéve, hogy a modell kiindulási feltételei helyesek voltak. A belső szögpontok meghatározását és a részletes számításmenetet lásd pl. Felsenstein (1981), Weir (1990:276-286) és Swofford & Olsen (1990:478-482) munkáiban.



6.16 ábra. Az ember és egyes főemlősök evolúciós kapcsolatának rekonstrukciója a maximum likelihood módszerrel a mitokondriális LEU tRNS és SER tRNS gének teljes nukleotid-szekvenciái alapján.

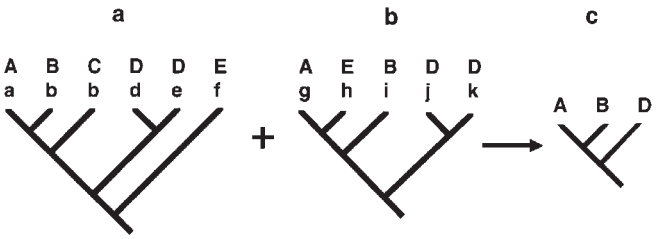
A LEU és SER tRNS gének teljes szekvenciáira (vö. 6.3.1.2 rész) végrehajtott maximum likelihood elemzés eredményét mutatja a 6.16 ábra. A számításokat a **PHYLIP** programcsomag **DNAML** rutinja végezte, az adatokból számított nukleotidgyakoriságok, és az általunk becsült k=3,0 paraméter (várható tranzíció/transzverzió hányados) figyelembevételével. Mivel az összes lehetséges – gyökér nélküli – fák száma 5 taxonra csupán 15, bizonyosak lehetünk abban, hogy megtaláltuk a legoptimálisabb fát. Az élek hossza az egy pozícióban átlagosan várható változások száma a két valós vagy hipotetikus szekvencia között (kizárva persze egy bázis önmagával történő helyettesítését, ami nem számít mutációnak). Ez nem jelenti azt, hogy a 0,05-ös élhossz esetén éppen 5 %-ban különböznek a kérdéses szekvenciák, hiszen mindig van olyan pozíció, ahol több mutáció is előfordul, s ez a "végeredményen" nem látszik. A tényleges eltérések tehát mindig kisebb mérvűek az élhosszaknál. Az ábrán bemutatott fa gyökér nélküli, de a gibbont külcsoportnak véve a topológia megegyezik a parszimónia módszerrel kapott kladográméval. A hasonlatosság nem véletlen, mivel a karakter alapon ill. a maximum likelihood alapján működő eljárások sok szempontból analógnak tekinthetők egymással (Swofford & Olsen 1990).

6.5 Kladisztikus biogeográfia

A mikrovilágból most egy hirtelen ugrással a kladisztika legnagyobb léptékű alkalmazási területére érkezünk. Az állat- és növényföldrajz egyik ága, a *történeti biogeográfia* kifejezetten azt célozza, hogy múltbéli események rekonstruálásával magyarázza meg az élővilág mostani elterjedését. Miután elsősorban a mai állapotról vannak ismereteink, magától értetődőnek tűnik, hogy a probléma a kladisztika módszereivel is megközelíthető. Az irányzat Nelson (1975), Nelson & Rosen (1981) és Parenti (1981) munkásságával kezdődött (ichtiológiai témában) és *kladisztikus* vagy *vikariancia* biogeográfia néven ismert. Bár az ilyen kutatások léptéke bizonyosan nagyobb, mint amit Magyarországon belül egyáltalán megtehetünk, érdemes legalább három oldalnyit szánni erre a témára is.

A biogeográfiai mintázat feltárásának az alapja számos, erőteljes *endemizmust* mutató rendszertani csoport kladisztikus elemzése. Feltételezzük, hogy az egyes csoportokon belüli leszármazási viszonyok egyúttal az előfordulási helyek közötti kapcsolatrendszerről is informálnak bennünket. Logikusnak tetszik, hogy két, közeli rokonságban álló taxon biogeográfiailag is közel áll egymáshoz, míg a nagyobb mérvű rendszertani eltérés már jelentősebb földrajzi távolságra utal. Mindez csak akkor igaz persze, ha a vikarianciát tekintjük minden eltérés magyarázatának a migrációval szemben: vagyis a közös ős mindenütt jelen volt a speciációt megelőzően, s a fajok nem vándorlással sugároztak szét. (Ez bizony nem általános érvényű, mutatva a kladisztikus biogeográfia korlátait.) A módszer lényege röviden az, hogy a – kettő vagy több monofiletikus rendszertani csoportra vonatkozó – kladogramokon a taxonok helyére az egyes területeket írjuk be, és az így kapott *área-kladogramok* összevetéséből vonjuk le a biogeográfiai következtetésket. A taxon-kladogramokat a fentebb leírt kladisztikai módszerek valamelyikével hozzuk létre, módszertani újdonság tehát az alternatív kladogramok értékelésében van. "Tökéletes" egyezés ugyanis ritkán áll fenn az área-kladogramok között: az egyes rendszertani csoportok múltja *nem feltétlenül* utal hasonló biogeográfiai kapcsolatra. Az egyes taxonok vándorlása vagy a kihalás csupán két lehetőség az eltérések magyarázatául.

Rosen (1978) módszerének lényege, hogy az área-kladogramokból csak az egyezéseket hangsúlyozzuk, ami sokszor a kladogram méretének csökkenésével jár ("reduced area cladograms"). Példaként vizsgáljuk meg a 6.17 ábrát, amely két rendszertani csoport kladogramját

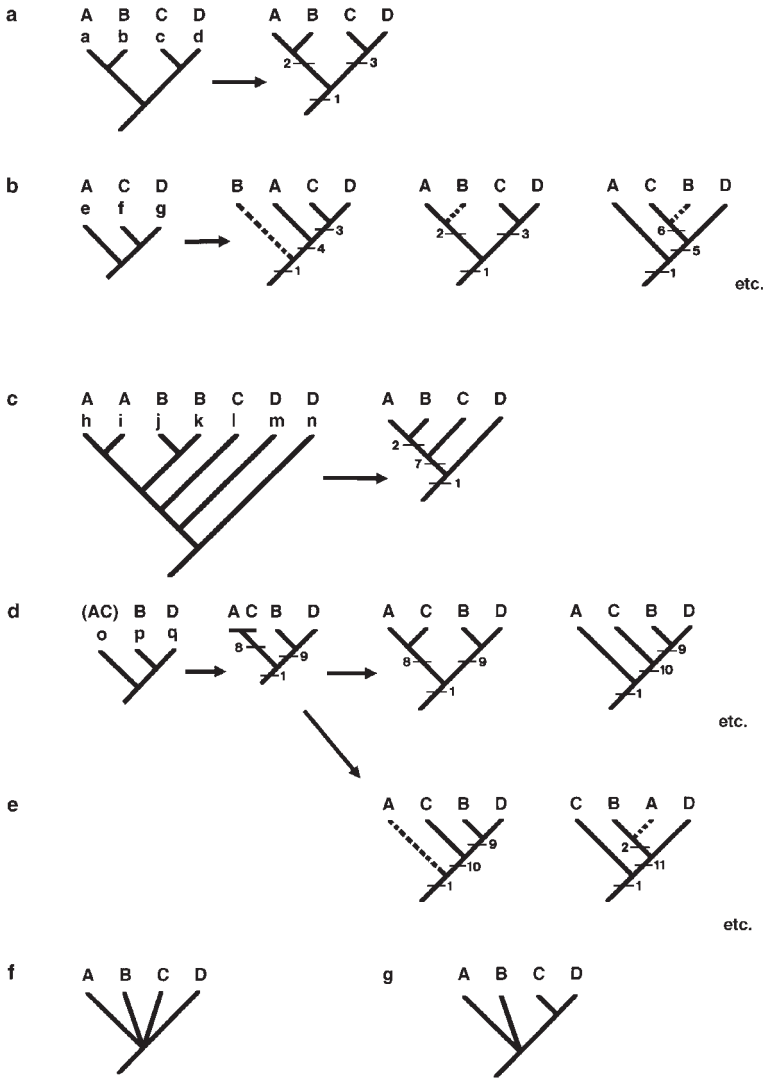


6.17 ábra. A Rosen-féle redukált área kladogram (**c**) mint a kiinduló kladogramok (**a-b**) egy lehetséges konszenzusa.

tünteti fel, a taxonok fölött bejelölve az előfordulási helyeket is. Az a-f taxonok alkotta rendszertani csoport mind az öt területről informál bennünket, a másik csoport egyetlen tagja sem fordult elő viszont a C-n, így ez az área eleve kiesik. Az E területre nézve a két kladogram rendkívül eltérő interpretációt sugall, ezért ezt is mellőznünk kell. Marad az A, B és D areákra vonatkozó viszonylag jelentős egybeesés, így a 6.17c redukált konszenzus kladogram lesz az, ami maximálisan adódhat az elemzésből: az A és B régiók biogeográfiailag hasonló múlttal rendelkeznek, s a D régió történetileg távolabb áll tőlük. A konszenzus elv tehát már kezdettől fogva lényeges alkotóeleme a kladisztikus biogeográfiának.

Amennyiben több área-kladogram is rendelkezésünkre áll, de ezek között ugyanúgy eltérések, sőt: ellentmondások vannak, mint a 6.17 ábra példáján, akkor Nelson & Platnick (1981) eljárása alkalmazható a közös információ "kihámozására". Előnye, hogy nem kell egyetlen areát sem kihagyni az elemzésből, bár a hiányos információ megmutatkozik a végeredményben. A szerzők az área kladogram rész-fáit komponenseknek nevezik, magát a módszert pedig *komponens elemzés*nek (ami persze nem tévesztendő össze a főkomponens analízissel, 7. fejezet). A komponenseket minden egyes kladogramon meghatározzuk, meg is számozzuk, majd ezek összesítő értékelése adja a keresett végeredményt. A komponensek azonosításának alapesetei a következők:

- Minden areának egy taxon felel meg a kérdéses rendszertani csoportban (6.18a ábra). Ez a legegyértelműbb eset, az ábra példáján az 1-3 komponensekre (az 1. triviális).
- A taxonok száma kevesebb, mint az areáké, ezért a hiányzó areák helyzete ismeretlen a kladogramon, megengedve számtalan alternatív lehetőséget (6.18b ábrán bemutatunk hármat). Ezek a 2-3. komponenseket megerősíthetik, de újakat is eredményezhetnek.
- Az areák száma meghaladja a taxonokét. A redundáns információ ekkor egy kisebb fába sűríthető, s csak ebben kell komponenseket keresni (6.18c ábra). A példában a 7. új komponens jelentkezik.
- Egy vagy több taxon több területen is elterjedt, ami az área kladogramon feloldatlanságot idéz elő. Ebben az esetben Nelson & Platnick (1981) után vagy azt feltételezzük, hogy 1) a több helyen élő faj valójában mindenütt jelen volt, csak a többi területről kihalt, a feloldatlan areák tehát monofiletikus ill. parafiletikus kapcsolatban vannak egymással (ezek láthatók a 6.18d ábrán) vagy pedig 2) a kladogram csak az egyik ilyen areára nézve informatív, míg a másikra nézve nem, mert a faj elterjedését is megengedjük, ami a kladisztika nyelvén a feloldatlan areák polifiletikus eredetét is jelentheti (mint a 6.18e ábra kladogramjain, ahol az A pozícióját vesszük bizonytalannak).



6.18 ábra. Komponensek azonosítása különböző típusú área-kladogramokon (**a-e**). **f:** Az área-kladogramok összesítése minden komponens alapján egy triviális politómiára vezet, **g**: Az área-kladogramok összesítése a 2., 3., 4. és 10. komponensek figyelembevételével.

A komponensek listája attól függ tehát, hogy a nagy elterjedésű taxonok esetében melyik feltételezést alkalmazzuk. A listában minden bizonnyal lesznek egymásnak szögesen ellentmondó komponensek is (a 6.18 ábra – szándékosan – ezt illusztrálja). Ha ezeket egyidejűleg figyelembe akarjuk venni, akkor könnyen kaphatunk egy igencsak triviális politómikus konszenzus kladogramot (6.18f ábra), ami nem biztos, hogy továbblépés a Rosen-féle redukált kladogramhoz képest (legfeljebb annyiban, hogy minden área benne van). Azt is megtehetjük, hogy a komponensek egy részét, mint "hamisat" elvetjük, s a megmaradtak segítségével állít-

juk elő a végeredményt. (Például a 2., 3., 4. és 10. komponensek alapján kapott, részben dichotomikus kladogram a 6.18g ábrán.)

A fenti módszer nem nélkülözi a szubjektív elemeket, s ezen egy "trükkel" segíthetünk Brooks (1981) nyomán. Az egyes komponensek bináris adatvektorok formájában is felírhatók (x_{ij}=1 ha a *j* área benne van az *i* komponensben, x_{ij}=0, ha nincs). Ezek a vektorok egy taxon × komponens adatmárixban összesíthetők, s a szokásos karakter-alapú parszimónia elemzéssel vizsgálhatók (vö. Humphries et al. 1988). Az így kapott maximálisan "takarékos" kladogram már mentes a konszenzus keresés problémáitól. Persze "cseberből vederbe" is eshetünk, hiszen számos, egyformán optimális parszimónia kladogramot is kaphatunk eredményül.

6.6 Irodalmi áttekintés

A kladisztika irodalma eléggé bőséges, és – valljuk be – nehezen áttekinthető, különösen a témával éppen csak ismerkedni szándékozók számára. Mi sem jellemzi jobban a helyzetet, mint Hull (1984) szarkasztikus megjegyzése: "Ha valaki megkérdezne, hogy miből kezdje megtanulni a kladisztika alapelveit, nem ajánlanám Hennig (1966) alapművét". Valóban, még egy rovarász is jobban jár, ha máshol kezdi az ismerkedést, s az ugyancsak rovarász Hennig könyvének böngészését inkább a tudománytörténészekre hagyja. A zoológusoknak talán Mayr & Ashlock (1991), a botanikusoknak Stuessy (1990), a molekuláris alapra helyezkedőknek pedig Swofford & Olsen (1990) ajánlható a viszonylag friss irodalomból. E művekre mi is számtalanszor hivatkoztunk ebben a fejezetben. Sokkal régebbi irodalmat nem nagyon érdemes elővenni, kivéve ha valaki részletesebben akar elmélyedni valamely résztéma múltjában. A módszertani vonatkozásokat illetően elég gyorsan elavulnak az ismeretek. A változásokat, és a főbb trendeket természetesen inkább a folyóiratokból érzékelhetjük. A kladisztika saját folyóirata a *Cladistics* (mi más is lehetne a neve) természetesen nem az egyedüli a biológiai irodalomban, amire figyelnünk kell. Lagalább ennyire jelentős orgánum a *Systematic Biology* (korábban *Systematic Zoology*), a *Systematic Botany*, a *Taxon* és a *Plant Systematics and Evolution* is. A legújabb biogeográfiai alkalmazásokról a *Journal of Biogeography* tájékoztat bennünket első kézből. A molekuláris kladisztika iránt érdeklődőknek az *Evolution* és a *Journal of Molecular Evolution* ajánlható elsősorban, de ezzel semmiképpen sem teljes a felsorolás. Ma már szinte minden taxonómiai és genetikai folyóirat közöl kladisztikai alapon végzett vizsgálatokat, ami egyben a téma növekvő fontosságára (és népszerűségére) is utal. Jelentős a speciális cikkgyűjtemények (pl. Duncan & Stuessy 1985) és konferencia-kiadványok (pl. Duncan & Stuessy 1984, Funk & Brooks 1981) száma is, hogy csak néhányat említsünk közülük.

Az általános művek sorában megemlítendő még Forey et al. (1992), amely kifejezetten egy bevezető tanfolyam anyagának szánja a leírtakat. Ebből a könyvből a DNS szekvenciák elemzésétől a kladisztika biogeográfiai alkalmazásáig sok mindenről informálódhatunk. Hasonlóképpen ajánlható Quicke (1993) könyve is, amely általános rendszertani alapozást ad, kiemelve a kladisztika módszereit.

6.6.1 Számítógépes programok

A kladisztikai programcsomagok "piacát" négy program uralja, mert ezek adják a legtöbb lehetőséget a kladisztikai adatelemzésre. A 6.2 táblázat segít bennünket abban, hogy a jelen kötetben is szereplő módszerekre megfelelő programot találjunk (az általunk nem említett módszereket nem tüntettük fel a táblázatban). Az összeállítás csak részben támaszkodik saját tapasztalatainkra, mert sokat merítettünk Sanderson (1990) összefoglaló értékeléséből is.

A táblázatbeli információn túlmenően meg kell jegyezni, hogy a **MacClade** (Maddison & Maddison 1992) csak Macintosh gépeken futtatható. Mellette szól viszont a könnyű használ-

6.2 táblázat. A fejezetben tárgyalt módszerek előfordulása a négy legfontosabb kladisztikai programcsomagban (mindegyik tartalmaz mást is, pl. bootstrap, konszenzus, stb.).

Módszer	PHYLIP	PAUP	HENNIG	MacClade
Saitou - Nei szomszéd összevonó	+			
Fitch - Margoliash	+			
Wagner távolság			+	
Rendezetlen kar. parszimónia (pl. DNS)	+	+	+	+
Wagner parszimónia	+	+	+	+
Dollo parszimónia	+	+		+
Sztratigráfiai parszimónia				+
Camin-Sokal parszimónia		+		
"Branch and bound"	+	+	+	
Invariánsok módszere	+	+		
Maximum likelihood	+			
Karakter-kompatibilitás	+			

hatóság, a kiemelkedő grafikus és nyomtatási lehetőségek (a jelen fejezet ábráinak egy része is ezzel a programmal készült). A **MacClade** egyedülálló a karakterek változásának nyomon követésében, amely egyformán fontos lehet a morfológiai és molekuláris kladisztikában is. Nem igazán hatékony viszont az optimális hosszúságú fák megtalálásában.

A PHYLIP programcsomag (Felsenstein 1993) mellett szól a sok opció, és a forrásnyelvi (!) valamint futtatható változatok ingyenessége és sokfélesége (l. B függelék), míg ellene a viszonylagos lassúságot hozzák fel a legtöbbször (Sanderson 1990). A parszimónia módszerekre általános vélemény szerint a **PAUP** (Swofford 1990) vagy a **HENNIG** (Farris 1988) inkább alkalmas.

Nem soroltunk fel a táblázatban sok, speciálisabb célra készült programot. A karakter-kompatibilitás elemzésére a legismertebb és legjobb kidolgozású a **CLINCH** program (K. Fiala), míg a biogeográfiai kladisztika vezető szoftvere a **COMPONENT** (Page 1989). Felsenstein (1993), a **PHYLIP** dokumentációjában még vagy 15 további programot (pl. Lake programja az invariánsok módszerére) sorol fel, a hozzáférés megjelölésével, úgyhogy már csak ezért is érdemes a **PHYLIP**-et beszereznünk.

6.7 Kérdezz – Válaszolok!

K: *Ahogy a fejezet végére jutottam (okoztál néhány "kellemetlen" órát), végül is nem igazán látom be: miért is olyan veszekedősek e tudományág művelői? Láthatóan itt is sokféle módszer leledzik, de ez a többi témakörben is így van, s ez még nem lenne önmagában ok a civódásra.*

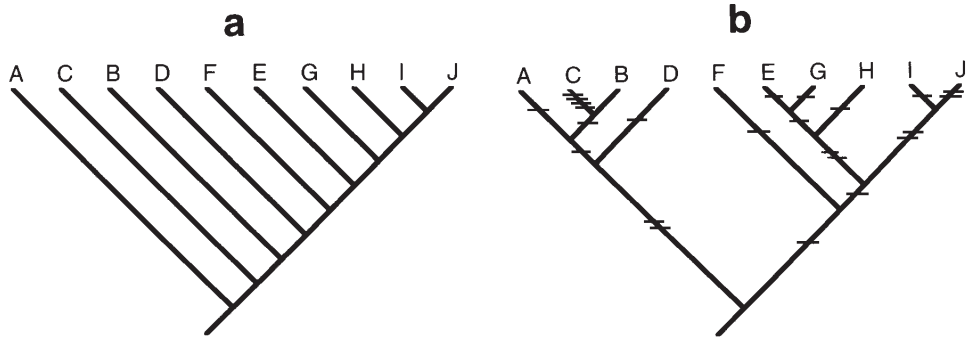
V: Igen, a technikai részleteken a kladisták éppen úgy vitatkoznak, mint a többiek a saját problémáikon. Viszonylag kevés helyem jutott viszont túllépni a módszertani aszpektusokon; a biológiai vonatkozásokról, pláne a filozófiaiakról már tényleg kevesebb szó esett, holott

vitára ezek lennének inkább alkalmasak. Csak egy példát: a transzformált kladisták ("pattern cladists") szerint a kladisztikus módszer alkalmazásához nem feltétlenül szükséges az evolúcióra gondolnunk. A kladisztika számukra mint egy hierarchikus mintázatot feltáró technika jön számításba. Rettentő nagy vihart kavartak ezzel a kijelentésükkel, bár egyikük sem tagadta az evolúció létét (elsősorban a kladisztikus biogeográfusokról van szó, pl. Nelson, Platnick és Rosen). A továbbolvasáshoz melegen ajánlhatom Gould (1990) utolsó fejezetét, és Dawkins (1994) 10. fejezetét.

K: *Gondolom, akkor itt beszélhetnénk a kladisztika és az osztályozás kapcsolatáról, mert ezt is éppen csak megemlítetted a 6.1 rész vége felé.*

V: Rátapintottál a lényegre: a kladisták és a – tradicionális – taxonómusok rengeteget vitatkoznak arról, hogy a kladogramok mennyire alkalmasak formális osztályozások létrehozására. Ha mondjuk a kladista egy, a 6.19a ábrán látható – vagy ahhoz hasonló, "fésűszerű" – eredményre jut, akkor a taxonómus csak legyint: "jó-jó, hogy ez a legoptimálisabb leszármazási mintázat, de akkor hogyan definiálsz különböző szintű taxonokat? Még ha valóban így is zajlott le az evolúció, csak nem gondolod, hogy ugyanannyi rendszertani kategóriát fogok bevezetni, amennyi hierarchikus szinted van? Mert akkor elvesznénk az alalcsaládok és a főfőalrendek dzsungelében!" De még egy ideálisnak látszó topológia (6.19b ábra) sem mentes a problémáktól! Egy végső kládon ugyanis sok autapomorfia jelenhet meg, de ezek kialakulása (az anagenezis) nem igazán érdekli a kladistát! •t csak az elágazásrendszer izgatja, tehát számára a B és a C taxonok tartoznak együvé, mondjuk egy génuszba. A rendszertanos viszont – az évszázados gyakorlat alapján – némi joggal tenné a B-t inkább az A-val egy génuszba hiszen csak két karakterben különböznek egymástól, míg a C a B-től négyben! Ez azonban a kladista számára egy parafiletikus csoport lenne. A ma általában alkalmazott rendszerek sok ponton parafiletikusaknak bizonyulnának egy alaposabb kladisztikus vizsgálódás során, s ez a nomenklatúrát is alaposan felborítaná. Elég, ha elolvasod De Queiroz & Gauthier (1990) és Bryant (1994) cikkeit, melyekben teljesen világos kladisztikai álláspont fejeződik ki: az elnevezéseknek holofiletikus csoportokon ("korona kládokon") kell alapulniuk.

K: *Ez valóban érdekes vitatéma. De akkor hol van az a terület, ahol úgy tűnik, harmonikusabb az egyetértés a kladista és a taxonómus között?*



6.19 ábra. Fésűs elágazási mintázat (**a**), amely "megnehezíti" a kladisztikai eredmények alkalmazását a formális osztályozásokban. A **b** kladogram a taxonómia és a kladisztika egy másik lehetséges konfliktusát illusztrálja: az autapomorfiák nagy száma miatt a C taxon klasszifikációs helyzete vitatott.

V: Nos, a makrotaxonómia területén, az élővilág regnum szintű osztályozásában és azon belül nem túl nagy mélységig már – vagy még? – kevesebb a konfliktus. A kladisztika módszereinek alkalmazásával sok érdekességre derült fény pl. a szárazföldi növények főbb csoportjainak kapcsolatáról (lásd Mishler & Churchill 1984, Bremer at al. 1987), vagy a zárvatermőkön belüli evolúciós viszonyokról (lásd Stuessy 1990 összefoglalóját). A kladisztika eredményei azonban – meg kell vallanunk – még nemigen jelentkeznek a rendszertanban.

K: *Sok mindent megtudtam a nukleinsav-szekvenciák értékeléséről, de támadt egy hiányérzetem: miért mellőzted az aminosavláncok, azaz a fehérjék kladisztikai alkalmazhatóságának bemutatását?*

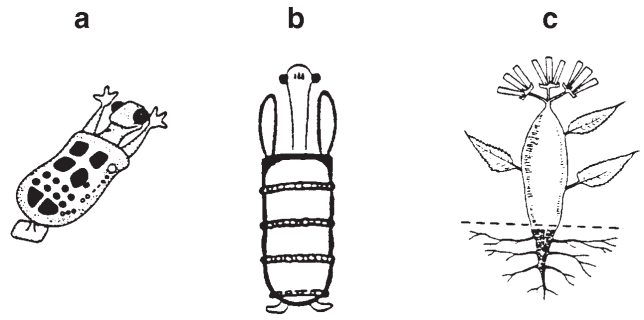
V: Ha már megkérdezted, akkor tényleg kell néhány szót szólnom erről. Kladisztikai következtetésekre ugyanis az aminosav szekvenciák is alkalmasak. Swofford & Olsen (1990) szerint proteinekre alapozva három főbb lehetőség merül fel: 1) Az aminosav behelyettesítések számának minimalizálása (vagyis a Fitch-féle rendezetlen karakterekre vezetjük vissza az elemzést, akár a nukleinsavak esetében). Itt a fő problémát az jelenti, hogy az egyes aminosavcserék mögött eltérő számú báziscsere áll. 2) mRNS szintre visszavezetve az aminosav láncot, az átalakításhoz szükséges nukleotid-szubsztitúciók számát minimalizáljuk (vö. Goodman 1981), figyelembe véve tehát a genetikai kód degenerált jellegét. Itt viszont a "csendes" szubsztitúciók túlhangsúlyozásának a veszélye áll fenn, ugyanis bizonyos báziscserék (a kodonok harmadik pozíciójában) nem változtatják meg az aminosavat. 3) A Felsenstein (1993) féle **PROTPARS** programban a csendes szubsztitúciók is kiküszöbölődnek. Láthatod tehát, hogy végeredményben a fehérjék nem önmagukban, hanem a mögöttük álló genetikai kód figyelembevételével csak közvetve alkalmasak igazán a vizsgálódásra.

K: *Kiderült számomra, hogy a kladisztika eredményét legalább annyira befolyásolhatja a kutató egyéni ízlése, mint mondjuk a hierarchikus osztályozását. Abban a fejezetben be is mutattad a módszerek közötti választás jelentőségét. A kladisztika esetében azonban mintha jóval kevesebb ilyen összehasonlítást tettél volna...*

V: Erre már – őszintén megvallom – nem jutott se helyem, se időm, se energiám. De az irodalomban bőven találsz olyan cikkeket, amelyek ezt már megtették helyettem, sokkal alaposabban, mint amire itt egyáltalán mód nyílna, például Duncan et al. (1980) és Astolfi et al. (1981). A nukleotid-szekvenciákra alkalmas módszerek összehasonlító értékelését Saitou & Imanishi (1989) és Nei (1991) végezte el. Az is előfordul, hogy ugyanazon módszerre írt különböző programokat értékelnek, mint például Luckow & Pimentel (1985) a Wagner parszimónia módszerek esetében. Ennek okát már sejted: a fa keresgélés egy NP-teljes probléma, és nagyon sok múlik a programokon.

K: *A maximum likelihood módszernél úgy tűnt számomra, hogy a modellrendszer változtatásával végül is felmérhetjük: miként változik a kapott eredmény. A modell tovább csiszolásával várhatóan még pontosabb eredményeket kaphatunk. De, mint írod, morfológiai bélyegekre ilyen modellek nincsenek. Valóban nincs semmi esélyünk arra, hogy karakter alapú módszereknél is megnézzük bizonyos változtatások hatását az eredményre?*

V: Vannak próbálkozások ilyen irányban is. Figyeld meg a 6.20 ábra "élőlényeit", amelyek voltaképpen mesterséges organizmusok, és éppen egy megfelelő elméleti modell híján születtek. A mesterséges organizmusok esetében az evolúciót maga az ember szabja meg, ismeri az



6.20 ábra. Egy-egy "példány" a mesterséges organizmusok *Caminalcules* (**a**, J. H. Camin, vö. Sokal 1983), *Didaktozoa* (**b**, Wirth 1995) és *Dendrogrammaceae* (**c**, W. H. Wagner, vö. Duncan et al. 1980) csoportjaiból.

egyes lépéseket, és úgy változtatgatja a feltételeket, ahogy akarja. E taxonok segítségével azután összehasonlíthatóak a kladisztika – és a fenetika – különféle módszerei is, ahogy azt Sokal tette négyrészes nagy cikksorozatában (Sokal 1983). A Caminalcules "csoport" elemzése azzal az eredménnyel járt, hogy kiderült: az összes karakter esetében a kladisztika módszerei jobban "eltalálták" az igazi törzsfát, mint a numerikus klasszifikáció. Érdekes módon azonban a karakterek számának csökkentése már az utóbbi módszereknek kedvezett, ami azt támasztja alá, hogy a kladisztikai következtetéseket is nagyszámú tulajdonságra kell építeni...

K: *Tényleg, van valami kikötés a tulajdonságok számát illetően?*

V: Általános szabály nincs, de nyilvánvalóan annál nagyobb esélyünk van egy teljesen feloldott kladogram előállítására, minél több tulajdonság szerepel az adatokban. És a fent említett Sokal-féle vizsgálat is a karakterek számának növelése mellett szól. No de visszatérve az előző kérdésedre, van azért lehetőség az evolúciós folyamatok számítógépes "lejátszására" is, melynek során összehasonlítható az egyes kladisztikus módszerek hatékonysága. Fiala & Sokal (1985) és Rohlf et al. (1990) nominális karakterek véletlen megváltozásával szimulálták a speciációs folyamatokat ("random walk"), vagyis közvetlenül a tulajdonságok szintjén vizsgálódtak. Persze evolúciós távolságok mátrixa is előállítható egy alkalmas modell felhasználásával (pl. Lynch 1989). Javítanom kell tehát magamat, több modell is van, amely alkalmas a kladisztika karakter-alapú módszereinek szimulációs értékelésére, de még sok új eredmény várható ezen a területen.