

Biosüsteematika teooria ja meetodid

Prof. Erast Parmasto loengukursuse konspekt

Esimene versioon, okt. - nov. 1994

0. Sissejuhatus.

0.1. Biosüsteematika on teadus eluslooduse mitmekesisusest: selle vormidest, põhjustest, tekkest; liikide ja teiste süsteematika ühikute piiritlemisest, nimetamisest ja teaduslikult põhjendatud klassifitseerimisest.

Kasutatakse ka lühemat nimetust *süsteematika* (alates Linné tööst 1737. a.), kuid süstematiseeritakse ka elutu looduse nähtusi. Camp ja Gilly (1943) kasutasid biosüsteematika mõistet erinevas tähenduses: eelkõige tsütogeneetiliste uurimiste kohta. See nimetusviis on aga juba ammu ajalukku vajunud. Et käesolevas kursuses segimineku ohtu pole, kasutatakse mõlemat terminit siin samatähenduslikena. - Kasutusel on ka sõna *taksonoomia* (esmakordselt 1813. a. De Candolle poolt); nende tähenduserinevus pole aga ühtselt käsitletav. Euroopas peetakse süsteematikat peamiselt organismide mitmekesisust uurivaks teaduseks, taksonoomia alla käivad klassifitseerimise teooria ja praktika. Põhja-Ameerika bioloogidel on enamasti vastupidi: taksonoomia hooleks on teooria, süsteematika alla käib praktiline klassifitseerimistöö. Nähtavasti on õigem loobuda nendele kahele terminile erineva tähendusvarjundi andmisest.

(Märkigem, et paljud bioloogia üldised terminid on küllalt noored. Isegi termin *bioloogia* on kasutusel alles alates 1800. aastast - Burdach, Lamarck; *ökoloogia* termini tõi kasutusele 1866. a. Haeckel. *Evolutsioon* tuli kasutusele mitte Darwini, vaid Spenceri töödega.)

Süsteematika **ehk** taksonoomia alla käib organismide klassifitseerimine, identifitseerimine ja nomenklatuur (nimetuste andmine). Nende ülesannete täitmiseks on vajalik ka evolutsiooniprotsessi uurimine (varieeruvus, liigiteke, hübriidide moodustumine jm.) ning fülogeneesi (viisid, aeg, koht) selgitamine.

Süsteematika rakendusalaadeks on identifitseerimiseks ja klassifitseerimiseks vastavate käsiraamatute (ka "määrajate") ning arvutiprogrammide (ka dialoogrezhiimis identifitseerimisprogrammide) koostamine; mingi ala (või kogu maailma) elusolendite inventariseerimine ja vastavate andmepankade pidamine, andmestiku koondamine bioloogilise mitmekesisuse kaitseks.

Eestis on mõnikord süsteematikaks nimetatud identifitseerimist ja floora või fauna inventariseerimist; see on ainult väike ja marginaalne osa süsteematikast.

Tänapäevane biosüsteematika ei piirdu kirjeldamisega ja inventariseerimisega; see on süsteematiku töös küll tähtis, kuid ainult teadustöö esimeseks etapiks. Pelgalt andmestiku kogumisega piirdujaid hüütakse *margikorjajaks* (stamp-collectors); eelkõige amatöörteadlasi iseloomustav enesepiirang, kui sellega kaasneb täpsus, põhjalikkus ja andmestiku teadlastele kättesaadavaks tegemine, on siiski heakskiitmist pälviv tegevus kui teadustöö ajalisel esimene etapp.

0.2. Süstemaatika kui teaduse olemasolu aluseks on tõsiasi, et elusolendid ei moodusta oma varieeruvuses kontiinuumi, vaid üksteisest enam või vähem selgelt erinevaid rühmitusi (hulki; elusolendite kui hulga osahulki). Süstemaatika põhiühikuks loetakse **liiki**; nende kirjeldamisega tegeleb nn. α -taksonoomia. Liigid ühendatakse süsteemis perekondadeks, sugukondadeks jt. liigist kõrgemateks rühmitusteks (β -taksonoomia); neid kõiki koos nimetatakse **taksoniteks**. Tuleb eristada taksonit (näit. perekonda või liiki) kui süstemaatika **kategooriat**, hierarhilise liigitamise klassi (viimase sõna loogika-alases tähenduses) ja konkreetset taksonit (näit. perekonda *Anemone*, ülane; liiki *Anemone nemorosa*, võsaülane). Liikide kõrgemateks taksoniteks ühendamine (või elusolendite kogumi jagamine liikideks) on *klassifitseerimine* ehk süstematiseerimine. Seda võib teha mitmeti sõltuvalt klassifitseerimise eesmärgist, aluseks olevast vaadete süsteemist eluslooduse mitmekesisuse päritolule, rakendatavast filosoofilisest alusest ja loogikast. Oluline on süstemaatikas ka liikide identifitseerimine ja sellega seotud nimetamine; viimane on kokkuleppeliselt rangelt reguleeritud **nomenklatuurikoodeksitega**, mis taimede (incl. seemed), loomade ja bakterite jaoks on mõneti erinevad.

Tänapäeval on süstemaatikas toimumas revolutsioonilised muutused: tänu uutest uurimismeetoditest tulenevatele suurematele võimalustele on alles nüüd, 100-aastase hilinemisega süstemaatika kujunemas evolutsiooniliseks, s.t. ajalooliseks teaduseks. On isegi väidetud, et (uus) süstemaatika on esmakordselt muutunud kunstist teaduseks.

1. Liik (*species*).

1.1. Liigi olemus.

1.1.1. Liigi eelteaduslik käsitlus. - Liike on inimesed eristanud juba aastatuhandeid; selle aluseks on olnud vajadus eristada söödavaid või muul viisil kasulikke (ravimeiks, nõidumiseks vm. kasutatavaid), kahjulikke (ohtlikke, mürgiseid jne.) liike. Loodusrahvastel on liikide eristamise oskus olnud silmatorkavalt hea; näiteks oli Paapua Uus-Guineas 138 linnuliigi eristamiseks kasutusel 137 nimetust; ka Eestis tunti sajand tagasi paljude (eriti nn. tarkade või nõidade poolt) poolt 150-200 taimeliiki; jahimehed - kõiki ulukeid, jne. See arv on suurem kui paljude praeguste diplomeeritud bioloogide tuntu. Eristatud ühikud vastavad sageli liigile selle termini teaduslikus mõttes, paljudel juhtudel perekonnale, mõnikord eluvormile. Kui üksikliikide eristamine oli ebaoluline, oli omaette nimetus ainult perekonnal (näit. 80 tarnaliigi jaoks oli Eestis rahva seas veel hiljuti käibel ainult paar nimetust).

Aristotelesel (384-322 a.e.m.a.) oli liik loogika mõiste ja ebakonkreetse sisuga; *sort*, *tõug*, *liik* olid tollal samatähenduslikud. **Süstemaatika** põhiühikuks sai liik alles John Ray töödega (1686-1704) ja muutus alles siis loogika terminist bioloogiliseks.

1.1.2. Mittedimensionaalne (dimensioonideta) liik. - Liikide eristamine ei tekita kohalikku, lokaalset floorat, seenestikku või faunat uurivale teadlasele enamasti erilisi raskusi: ta näeb mistahes liigi varieeruvusest

ainult väikest osa ning kitsamal alal puuduvad enamasti ka lähedased sugulasliigid - eriti, kui need on tekkinud geograafilise isolatsiooniga seotult (allopatriliselt). Nn. *liigi probleem* on hoopis raskem uurijale, kes vaatleb liike kogu nende levila (areaali) piires ja hõlmab oma uurimisega kogu ligidaste liikide kompleksi. Need võivad tunduda äärmiselt sarnastena; kui nad on ka tõepoolest üksteisega lähedastes sugulussidemetes, kasutatakse terminit *sõsarliigid* (sibling species; eesti keeles on kasutatud ka vähemsobivaid termineid *kaksikliigid*, *teisikliigid*). Sageli on need tegelikult teatud meetodeid (mikroskoopilist, etoloogilist vm.) kasutades eristatavad, ja nad ei kujuta endast mingit liikide erisorti. Eesti botaanikute poolt kasutatud kvaasi-termin *kriitiline liik* näitab ainult uurija emotsionaalset suhtumist; *pisiliik* on botaanikute juures olnud kasutusel üldterminina nii sõsarliikide kui ka mikroliikide (vt. allpool) kohta ja seega ebatäpne.

1.1.3. Liikide tekkimine on uurimisobjektiks evolutsiooniõpetusele selle mitmetes eri suundades. Süstemaatika kursuses pole eriti oluline, millist liigitekke viisi peetakse õigeimaks; kogu tänapäevane süstemaatika lähtub siiski eeldusest, et liigid tekivad, muutuvad (ja kaovad); et liikidel on ajalugu. Mõõndes liikide loomist valmiskujul mistahes jumala poolt välis-taks vajaduse süstemaatikat (ja üldse bioloogiat) arendada ja õppida.

1.1.3.1. Käesolevas kursuses eeldame, et õppijal on ettekujutus liikide **tekke** tunnustatud **põhiviisidest**:

1) Allopatriline liigiteke, mis seisneb ühe, algse liigi jagunemisel vähemalt kaheks geograafilise isoleerituse tõttu. Isolatsiooni on põhjustanud algse levila jaotumine geograafiliste tegurite mõjul (mäeahelike teke, mandrite triiv, jne.), või uue osa-areaali teke eelmisest geograafiliselt eraldatud alal (seemnete sattumine kaugele ookeanisaarele vms.).

2) Sümpatriline liigiteke, mille puhul liik jaguneb üksteisest ikka enam eristunud allühikuteks (rassideks) spetsialiseerumisega erinevale ökoloogilisele nishile (kaasa arvatud parasiitide spetsialiseerumine erinevatele peremeesorganismidele), ristumist vältivate piirangute tekke tõttu (etoloogiline sobimatus; erinev õitseaeg taimedel, jne.) vms. Seejuures võivad tekkinud liikide areaalid olla üldjoontes või tunduvas ulatuses kokkulangevad.

3) Kvant-liigiteke, mille aluseks on põhiareaalist eemalasuvas populatsioonis talletatud geneetilise informatsiooni piiratus ja infovahetuse lakkamine põhiareaali isenditega.

4) Hübrid-liigiteke, mille puhul eri liikide hübriidid on muutunud iseseisvalt paljunevaks taksoniks.

5) Polüploidisuse tekkega seotud liigiteke (peamiselt taimedel).

1.1.3.2. Ristumise puudumisel (**ühevanemalise pärilikkuse korral**, s.t. **uniparentaalsetel** liikidel) jääb ära geneetilise informatsiooni vahetus isendite vahel; koos päriliku muutlikkusega (mutatsioonid) tekivad üksteisest ikka rohkem erinevad kloonid, mis loodusliku (eelkõige stabiliseeriva) valiku tõttu moodustavad üksteisest vähe, mõnevõrra või silmatorkavalt erinevaid **mikroliike**. (Neid võiks eesti keeles nimetada ka täpses tõlkes

pisiliikideks, kuid see termin on Eestis paraku kujunenud mitmetähendusliikuks.) Taimedel esineb nn. *agaamseid komplekse*, kuhu kuuluvad apomiktilised mikroliigid, neile päritolult väga lähedased seksuaalsed liigid, hübriidid ning lisaks sellised mikroliigid, mis - kuigi harva - käituvad tavaliste, seksuaalsete liikidena. Selliseid komplekse erineb taimeperekondades või-lill (*Taraxacum*), hunditubakas (*Hieracium*) jt.

1.1.3.3. Liikide tekkimise ja olemuse kohta tuleks õppijal lugeda Henni Kallaku raamatuid *Adaptatsiogenees ja liigiteke* (Tartu, 1986; eriti lk. 47-86) ja *Bioevolutsioon* (Tallinn, 1990; eriti lk. 66-78).

1.1.4. Küsimusele "mis on liik?" püütakse vastust anda liigikontseptsioonide, liigi defineeringute ja liigi praktilise standardi esitamise kaudu. Suuremas osas teoreetilisestki kirjandusest on need kolm mõistet sageli segi aetud; kõneldes liigikontseptsioonist mingis perekonnas peetakse silmas hoopis praktilist standardit.

1.1.4.1. Liigikontseptsioon on teoreetiline lähtekoht, mille alusel liike üldse (põhimõtteliselt) eristada saab. Erinevate kontseptsioonide piirid pole alati selged, kuid eristada saab nelja põhilist:

1.1.4.1.1. Nominalistlik kontseptsioon lähtub seisukohast, et liike polegi reaalselt olemas, et need eksisteerivad ainult inimese jaoks, inimese mõttetegevuse viljana. Olevat ainult isendid või nende rühmad, mille vahel pole selgeid piire. Niisiis on liik kokkuleppeliselt eristatav isendite kogum. Levini (1979) järgi on liigi mõiste vaid tööriist eluslooduse mitmekesisuse iseloomustamiseks ja analüüsiks. Liigne tähelepanu vahendile tõmbavat ära tähelepanu organismidelt enestelt. - Vastavalt Stebbinsile (1969) olevat olemas vaid populatsioonide ja populatsioonirühmade organiseeritud süsteemid, mis moodustavad ebakorrapärase varieeruvusmusteri. Seda mustrit iseloomustavat sarnaste variantide kobarad, mida eraldavat üksteisest suuremad või väiksemad mittepidevused.

1.1.4.1.2. Essentsialistlik ehk morfoloogiline kontseptsioon väidab, et liigid erinevad üksteisest morfoloogiliste (laias mõttes - siia kuuluvad ka näit. biokeemilised, füsioloogilised jt.) tunnuste poolest; need olevat nii vajalikud kui ka piisavad liigi eristamiseks (kuigi ei tarvitse kõigil ontogeneesi astmeil esineda). Oma äärmuslikul kujul esineb see kontseptsioon **tüpoloogilise** liigikontseptsioonina: isendid peavad vastama liigi "tüübile"; kõrvalekalded sellest kujutavad *ebatüüpilisi* isendeid. See vastab essentsialistliku maailmapildi seisukohtadele: looduslikel objektidel olevat reaalne **essents**, olemus; kõrvalekalde puhul sellest ei väljendu essents mingil põhjusel täielikult, on häiritud, tegemist on mingi veaga. Essentsialistlik liigikontseptsioon oli valitsev ajal, millal liike peeti muutumatuiks ja/või jumala poolt looduks. Selle liikide varieeruvust eirava vaatekoha järgi on näit. ka Eestis botaanilises kirjanduses kasutatud väljendit *liik ja tema vormid / varieteetid / teisendid* - nagu ei kuulukski need liigi koosseisu...

Morfoloogiline (tüpoloogiline) kontseptsioon tundub olevat kõige operatiivsem, kuid annab suured võimalused suvaliseks liikide piiritlemiseks vastavalt süstemaatiku soovile.

1.1.4.1.3. Isolatsioonistliku ehk **bioloogilise liigikontseptsiooni** koha-

selt koosneb liik reaalselt või potentsiaalselt ristuvate isendite populatsioonidest, mis (E. Mayri hilisema täpsustuse kohaselt) esinevad spetsiifilises ökoloogilises nishis. Isendite kuuluvuse ühte liiki ei määra inimene, vaid isendid ise üksteist "ära tundes". Selle kontseptsiooni järgi pole (rohked) ühevanemalise pärilikkusega ehk mitteseksuaalsed taksonid tõelised liigid.

1.1.4.1.4. Fülogeneetiline kontseptsioon ühendab ühte liiki monofüleetilise isendite rühma, s.t. sama esivanema(te) järglased. Kontseptsioon on praegu teoreetilise seisukohana üldtunnustatud, kuid on väheoperatiivne. Täpsustatud formuleeringus on öeldud nii (Mishler, Donoghue jt., 1982, 1987): Liik on väikseim klassifikatsiooniühik, millesse organismid on gruppeerunud tänu nende monofüülsele päritolule (mistõttu neil esinevad enamasti sünapomorfseid tunnused) ja mis seetõttu pälvib formaalset tunnustamist liigina. (Termini *sünapomorf* tähendust käsitletakse siin edaspidi.) - Selle kontseptsiooni nõrkuseks on populatsiooni, liigi ja alamliigi eristamise võimatus, liigi reaalsuse ebaselgus. Tõepoolest, mitmed autorid (näit. Raven, 1986) väidavad, et mingit liikide evolutsiooni ei toimugi, selle asemel on õigem rääkida populatsioonide evolutsioonist.

Et ainult osa organismidest on seksuaalsed, võib ülalkirjeldatud bioloogilist kontseptsiooni pidada ühte *erijuhtu* käsitlevaks seisukohaks.

Praktilises süstemaatika töös on tänapäeval veel valitsev morfoloogiline kontseptsioon - hoolimata sellest, et kõik möönavad selle põhjendamatust ja anakronistlikku iseloomu.

1.1.4.1.5. Sellega ei piirdu veel kontseptsioonide mitmekesisus. Märkimisväärsem Templetoni (1989) **kohesioonikontseptsiooni**. Selle järgi hoiab isendeid liigina koos mingi tegur: piiratud geenisiire; stabiliseeriv valik; ajaloolised, arengu- või ökoloogilised piirangud. Selline kontseptsioon sobib ka aoseksuaalsete organismide rühmadele. Sarnaseid vaateid (stabiliseerivat valikut olulisimaks pidades) on avaldanud ka Parmasto (1986).

1.1.4.1.6. Märkimisväärsem veel, et **Ch. Darwin** oma teoses *liikide tekkimisest* ei esitanud liigikontseptsiooni: see pole raamat liigist, vaid liikide muutumisest. Darwini jaoks oli liik vaid muutumise hetkeseis, s.t. peaaegu et mittereaalne.

1.1.4.2. Liigi **definiitsioone** on ainuüksi taimede jaoks enam kui kolmsada. Selles esitatakse lisaks üldistele printsiipidele ka liigi **omadused** ning üksikasjalikult need üldised kriteeriumid, mis võimaldavad liike eristada. Paljudes definiitsioonides on esile tõstetud: liik on ajalooliselt (evolutsiooniliselt) kujunenud ühise päritoluga (monofüleetiline) isendite hulk; liigid on varieeruvad ja ajas muutuvad, samal ajal ka suhteliselt stabiilsed (püsivad); liigi isendid on omavahel ristuvad ja annavad elujõulisi järglasi, ei ristuta aga teiste liikide esindajatega; isendid on omavahel sarnased, teistest liikidest eristab neid *hiatus* - tunnuste selge, üleminekuteta erinevus [ideaaljuhul, mida võib reaalselt täheldada peamiselt mittedimensionaalse liigieristuse puhul, s.t. vähese arvu isendite puhul]; liigile on iseloomulik teatud areaal (levila), mis võib koosneda osa-areaalidest, ning teatud ökoloogiliste nõudmiste kompleks. Liike iseloomustab ka teatud minimaalne arvukus: üks eksemplar pole veel liik (are-

aal puudub!); loomade puhul on arvatud, et alates umbes 500 isendist võiks rääkida juba liigist. Muuseas, selle nõude vastu on palju patustatud, ja see on ka üheks põhjuseks, miks paljusid liike pole hiljem leitud, või need on osutunud varieeruva liigi sünonüümideks. - Toodud liigi omaduste loetelu pole ammendav.

1.1.4.3. Liigi praktiline standard on nende tunnuste loetelu ja iseloomustus, mis on kasutusel mingis rühmas (tavaliselt perekonnas või sugukonnas) liikide eristamiseks. See tugineb uurijate senistel kogemustel ja on tavaliselt empiiriline, s.t. teoriavaba, sisaldab vaieldavaid subjektiivseid arvamusi ning piirdub sageli lihtsalt kasutatud tunnuste ja nende varieeruvuse iseloomustamisega ja suhtelise hinnangu andmisega (mis on antud rühmas 'hea', mis 'halb' tunnus). Paraku on siin sageli jõutud tautoloogiani: hea tunnus olevat see, mis võimaldab hästi liike eristada; hea liik on see, mis kasutatud tunnuste poolest üksteisest erinevad...

1.1.5. Liikide eristamine on paljudel juhtudel küllalt raske; eri teadlased võivad samas rühmas (näit. perekonnas) eristada ühe, mõne või palju liike. Anekdootliku (nominalistlikku seisukohta väljendava) aforismi kohaselt on liik see, mida hea süstemaatik liigiks peab. Eriti zooloogide seas varem laialt levinud bioloogilise liigikontseptsiooni kohaselt ei saa liikideks pidada isetolmlevaid taimi, homotallistlikke (ristumiseta paljunevaid) seeni jt. arvukaid ühevanemalise (*uniparentaalse*) pärilikkusega organisme - kuigi neil on siiani liike eristatud. Ristumisbarjäär on sageli ainult osaline, tuntakse ka perekondadevahelisi hübriide. Seentel esineb nn. A-B-C -süsteemi: näit. võivad ühel mandril esinevad taksonid A ja B olla omavahel mitteristuvad, kuid on seda teisel mandril esineva väga ligidase taksoniga B. Seentel esineb ka nn. intersteriilseid grupe: sama liigi (?) eri populatsioonid erinevad omavahel ainult mitteristumise poolest; kõigi teiste tunnuste (ka molekulaarsete) poolest on nad praktiliselt samad, erinevused on palju väiksemad kui sama liigi omavahel ristuvatel isenditel. Liikidevahelise erinevuse aste on erineva mastaabiga eri organismirühmades (taimed, linnud, teisseened, jne.).

Nii võimegi **liigi reaalsuse probleemi** kohta väita: ühise päritoluga (monofüleetiliste) isendirühmade esinemine looduses on reaalsus, nende nimetamine liikideks on aga **mudel**, mille on loonud inimene. Enamusel juhtudest on see mudel ilmselt **vastavuses** reaalsusega (kuid pole mitte reaalsuse). Varasemad (näit. Linné) seisukohad liigi ühemõttelisest reaalsusest tuginesid eelkõige napile infole liikide varieeruvuse kohta, samuti filosoofilis-ideoloogilistest eelarvamustest.

1.1.6. Elusolendite esinemine liikidena on ilmselt seoses nende ehituse ja elutegevuse spetsialiseerumisega teatud elu- (kasvu-)tingimustele, nende seosega nn. 'oma' ökoloogilise nishshiga. Oletatavasti on see **kujunenud evolutsiooni** vältel, s.t. evolutsioon (looduslik valik?) on soodustanud liigilist esinemisvormi. Sellest tulenevalt ei saagi liik ühes organismirühmas olla samasugune kui mõnes teises. Veel enam: isegi samas perekonnas võib esineda erinevaid liigitüüpe (näit. hetero- ja homotallistlikud, s.t. ühevanemalise pärilikkusega liigid mõnes seeneperekonnas). Järelikult ei saa mistahes liigikontseptsioon olla nende eristamise praktiliseks juhi-

seks, vaid jääb enam või vähem teoreetiliseks üldistuseks. - Liik on sama-aegselt nii looduses reaalselt eksisteeriv eluslooduse organiseerituse vorm kui ka inimese poolt formaliseeritud üldistus.

1.2. Liigisisised süstemaatika ühikud (alamtaksonid) on formaalse, kokkuleppelise ja nomenklatuurikoodeksites fikseeritud käsitluse järgi:

1.2.1. Alamliik (*subspecies*) - varem ja paljudel juhtudel ka praegu peetakse selleks liigi indiviidide kogumit, mis erineb (peam. morfoloogiliselt) teistest tunduvalt. Õigem on järgida peam. zooloogide käsitlust: liik on (peam. geograafiline) rass, millel on erinev (geograafiline või ökoloogiline) levik, pärilikud erisused morfoloogias ja ökoloogias, kuid mis on (seksuaalselt paljunevate puhul) ristuvad sama liigi teis(t)e alamliigi esindajatega.

1.2.2. Varieteet (*varietas*) on morfoloogiliselt (pärilikult) erinevate isendite rühm, millel pole oma areaali. Zooloogias varieteete ei eristata.

1.2.3. Vorm (*forma*) on veelgi väiksemate, tõenäoliselt mittepärilike erinevustega teisend.

Botaaniline nomenklatuur lubab eristada ka **alamvarieteeti** ja **alamvormi**. Paraku on nii varieteet kui ka vorm kindla sisuta mõisted, mida käsitatakse eri teadlaste poolt üsna subjektiivselt. Nende eristamine ja kirjeldamine on ainult üks ja mitte parim võimalik viis liigisisese varieeruvuse kirjeldamiseks.

1.2.4. Varieteet ja vorm on tüpoloogilisest mõtteviisist ja vastavast liigikontseptsioonist tulenevad ühikud, sageli ebakindlalt eristatavad. Linné ja tema järglased tähistasid liigisiseseid ühikuid mitte nime, vaid tähe (A; B; a; b; α ; β jne.) või numbri abil (tänapäeval nimetame neid taimede ja seente puhul tinglikult varieteetideks). Eristamine ja nimetamine oli lausa moeasjaks möödunud sajandivahetusest peale botaanikas; praegu on see kombeks näiteks lihheniseerunud seente puhul. Bioloogiliselt mõtestatud on ainult alamliigi ja **populatsiooni** eristamine; sisukas on ka näit. seentel intersteriilsete rühmade ja (parasiitseentel teatud peremeestaimedele) spetsialiseerunud vormide eristamine. - Populatsiooni mõistet käsitletakse geneetika ja evolutsiooniõpetuse kursustes, siin ei korrata.

1.2.5. On kasutatud ka **poolliigi** mõistet: see on kujunev liik, mis pole veel iseseisva liigini jõudnud. Eriti ornitoloogias on kasutusel olnud ka **ringliigi** termin: tsirkumboreaalne levikuga liigi esindajad võivad tänu klinaalsele muutlikkusele olla areaali "otste" kokkupuutepunktis olla mitte-ristuvad ja ka morfoloogiliselt erinevad.

1.2.6. Alamliik, varieteet, vorm on hierarhilised kategooriad; kasutatud on ka mitteformaalseid ühikuid, millega tähistatakse teatud ühiste tunnustega indiviidide rühmi. Gilmour ja Gregor ning Heslop-Harrison tõid sisse nn. deemi- (*deme*) terminoloogia. *Topodeem* ja *ökodeem* tähistavad ühes paigas või samades ökoloogilistes tingimustes esinevaid grupeeringuid; *gamodeem* on omavahel ristuvate isendite kogum; *autodeem* on autogaamsete või iseviljastuvate isendite ühendus, jne. Sagedamini on kasutatud esimest kolme mõistet; süstemaatik saab hästi hakkama ka ilma deemi-terminoloogiat kasutamata.

2. Liigist kõrgemad süstemaatika põhiühikud olid möödunud sajandi keskpaigani vähearvukad: vajadus nende arvu suurendamiseks tekkis alles paljude uute liikide (ja perekondade) juurdekirjeldamisega. Hõlpsaks orienteerumiseks on sobivaim hierarhiliste tasemete arv nii umbes 8... Põhiühikuiks loetakse praegu (hierarhilises reas) perekond, sugukond, selts, klass, hõimkond ja riik. Monograafiliste uurimustes eristatakse neil alljaotusi (seksioon, alamperekond, triibus, alamsugukond, ülemsugukond (zooloogias), alamklass...). Linné järgi olid liik ja perekond jumala loodud, kõrgemad ühikud aga inimese mõttetegevuse produkt. Tänapäeval leidub mõningaid süstemaatikuid (feneetilise kooli esindajad), kes peavad perekonda reaalselt eksisteerivaks; selle aluseks on küllaltki häguse ja liigagi üldise mõiste *reaalsus* lai tõlgitsus. Kuna evolutsioon toimub tänapäevaste teadmiste järgi ainult liikide, mitte aga kõrgemate süstemaatika ühikute muutumise teel, ei saa ka perekonda (nagu sugukondagi jne.) pidada muuks kui ühise fülogeneesiga, s.t. monofüleetiliseks liikide ajalooliselt kujunenud rühmituseks. Kujutades liikide teket piltlikult dihhotoomselt haruneva arengupuuna, on süstemaatikute kokkuleppe küsimus, millise harunemisjärgu "oksa" pidada perekonnaks, millist sugukonnaks jne. Olulisem on siin teatud ühtluse taotlus. Koos ikka uute liikide kirjeldamisega on näit. perekonna maht käesoleva sajandi jooksul muutunud ikka väiksemaks ja väiksemaks. Et sellega koos on vähenenud süsteemi ülevaatlikkus, on praegu täheldatav mõningane vastupidine tendents.

Vaadeldes liike (ja teisi taksoneid) mitte nende ajaloolise kujunemise produktidena, vaid kui klassifitseerimist vajavaid sõltumatuid objekte, saame nende suhteid **modelleerida** kui punkte hulgamõõtmelises tunnusruumis (sellest tuleb juttu ühes edasises peatükis). Sobiva mudeli puhul võime näha, et need mudeli punktid (mitte liigid ise!) esineva parvedena, mis omakorda moodustavad suuremaid parvesid. Sellise käsituse puhul on liigist suuremail süstemaatika ühikuil mudeli kontekstis ka teatud "reaalsus".

3. Enne järgnevat tuleb defineerida mõni tänapäevase süstemaatika põhi-mõiste. Mõõndes liikide ajaloolist (evolutsioonilist) kujunemist tuleb eristada *monofüleetilisi*, *parafüleetilisi* ja *polüfüleetilisi* taksoneid (liigirühmitusi).

3.1. Monofüleetiline on liikide rühm, kuhu kuulub nende ühine esivanem (liik) ja kõik sellest kujunenud taksonid (vt. joonis 1, A).

3.2. Parafüleetiline on selline 'monofüleetiline' rühm, millest osa taksoneid (liike) on välja arvatud (joonis 1, B). Selle eristamise aluseks on tavaliselt üht või mitut liiki (taksonit) teistest hästi eristava tunnuse esinemine, mida süstemaatik oma subjektiivse hinnangu alusel (või feneetilisi kriteeriume kasutades) peab oluliseks.

3.3. Polüfüleetiline on liikide rühm, kuhu on ühendatud erinevaist esivanemaist pärinevad organismid (joon. 7, C). Polüfüleetiline oleks näiteks rühm, mis ühendaks kalad ja vaala, või linnud ja kärbsed.

3.4. Polüfüleetilistest rühmitustest püüavad hoiduda kõik süstemaatikud; fülogeneetilise süstemaatika viljelejad eitavad ka parafüleetiliste rühmade

lubatavust süsteemis. Nn. evolutsiooniline süstemaatika möönab parafüleetiliste taksonite lubatavust, s.t. võimalust evolutsiooni käiku kujutavat süsteemi "parandada" subjektiivselt hinnatud sarnasuse arvestamisega. Sellest seisukohast kinni pidades on *mistahes* meie loodud taksoneid suvaliselt välja jättev rühm "monofüleetiline" - näiteks perekond, millesse on ühendatud ainult jännes, inimene ja lehm.

4. Tunnused ehk *atribuudid* (filosoofia termin, tähistab *olulisi* tunnuseid) ehk *deskriptorid* ehk (feneetilise klassifitseerimise algoritmides) *muutujad* võimaldavad isendeid ja taksoneid iseloomustada, kirjeldada ja klassifitseerida. Kuni kaunis hiljutise ajani liigitati tunnuseid süstemaatika õpikuis selle järele, millisest valdkonnast või millist uurimismetodit kasutades need on saadud (morfoloogilised, anatoomilised, ultrastruktuuri, biokeemilised jne. tunnused). Tunnuste hindamisel on eelduseks nende **pärilikkus**. Kõrgemalt on hinnatud taksoni- (liigi-)sisesele vähevarieeruvaid tunnuseid; tegelikult on näit. meetrilised ja ka paljud meristilised tunnused **alati** üsnagi muutlikud. (Lineaarsete meetriliste tunnuste liigisisese variatsiooni koefitsient CV on enamasti $100 \frac{s}{x} = 5-10$; pindala puhul on see umbes 2 ja ruumala ning kaalu puhul 3 korda suurem). - Jättes kõrvale juhuslikud tunnused (s.t. mitte-atribuudid - näit. mõne eksemplari kogumise kellaaeg või koguja vanus), on tunnused *mistahes* süstemaatika analüüsi eel võrdtähenduslikud, nende "kasutamiskõlblikkus" selgub töö käigus, tulemustega. Lähtudes analüüsi teostamisest, eristame n.ö. tehnilisest seisukohast siiski järgmisi tunnuste tüüpe.

4.1.1. Kvantitatiivsed tunnused - sellised arvtunnused, mille väärtust võime mõõta või loendada (saba pikkus, kroonlehtede arv, munade koorumise aeg päevades pärast viljastamist, jne.). Eristatakse **meetrilisi** (möödetavaid) tunnuseid (näit. pikkus sentimeetrites) ja **meristilisi** ehk diskreetseid (neid väljendatakse enamasti ikka täisarvuliselt või kindla intervalliga). Diskreetsed tunnused on omakorda loendatavad (näit. tedremarana kroonlehtede või karu jalgade arv) või astakulised ehk intervall-tunnused (näit. taime õitseage juuni 1.; 2.; 3. dekaadil).

4.1.2. Kvalitatiivsed tunnused - sellised, mida me (muundamata) ei saa arvudes väljendada. Need võivad olla **ordineeritud** ehk järjestatavad (näit. kerajas--ellipsoidne--silinderjas; väike--keskmise--suur) või **ordineerimata** ehk järjestamata (näit. sinine, kollane, pruun, purpurne - neid värvusi ei saa iseloomustada spektri lainepikkustega; looma liikumisviis võib olla lendav, ujuv, roomav...).

4.1.3. Loogilised tunnused on erivorm kvalitatiivseist tunnustest: mingi tunnuse olemasolu organismil saab olla kas *tõene* või *väär*, s.t. mingi tunnus esineb või puudub. Mitmed meetodid samastavad selle tunnuste tüübi 2-astmelise (binaarse) ordineeritud või ordineerimata tunnusega - algoritmides saab neid kasutada ühtmoodi).

4.1.4. Tunnuseks võib olla näiteks *kollane*, või hoopis *värvus* ja selle **tunnuse seisundiks** *kollane*. Inglise keeles on vahe selgem: esimene 'tunnus' on *character*, teine (tunnuse seisund) on *characteristics* ehk *character state*. Tarvitades terminit 'tunnuse seisund' eeldame, et ühe tunnuse kõik

seisundid on omavahel **homoloogsed** (sellel peatume pikemalt edaspidi). Kasutades tänapäevast, kladistilist metodoloogiat kasutatakse tunnuse seisundite tähistamiseks sageli ka terminit **transformatsiooniseeria** (*transformation series*). Viimane on seotud ideega, et tunnuse seisundid on kujunenud ajalooliselt, s.t. ühe tunnuse (= tunnuse seisundi, kaasa arvatud selle puudumine) muutumisel ehk transformeerumisel teiseks.

4.2. Mingi taime-, looma- või seenetaksoni kirjeldus (näit. uue taksoni kirjeldamisel, või mingis käsiraamatus) on selle tunnuste ja tunnuste seisundite **jutustav** loetelu. Jutustuses lisandub küll täiendavat informatsiooni (näit. tunnuse seisundi täpsustav, kommenteeritud kirjeldus), kuid ka informatsioonilist müra. Sageli väljendatakse sama tunnuse seisundi esinemist eri taksonite juures pisut erinevas sõnastuses (näit.: õis on verev / veripunane; leht on karvane / karvkattega, jne.). Sageli "ununeb" mõni antud taksoni jaoks autori poolt vähemoluliseks peetud tunnus hoopiski märkimata; tunnuste järjestus kirjeldustes on sageli ebajärjekindel. Seda kõike aitab vältida tunnuste kodeerimine ja arvutiprogrammide (näit. DELTA - seda õpime praktikumis) kasutamine. Veelgi tähtsam on kodeerimine klassifitseerimisel.

4.2.1. Tunnuste kodeerimine võimaldab nende hõlpsat väljendamist ja kasutamist arvutiprogrammides. Kui meie loetelus on kolmandaks tunnuseks värvus ja sellel on neli tunnuse seisundit (1 - valge, 2 - kollane, 3 - sinine, 4 - punane), võime punase värvuse kui tunnuse kodeerida numbritega **3,4**. Kui tunnus on varieeruv, saab ka seda kodeerida (näit.: esineb tunnuse seisund 5 või 6; 5 kuni 6; 5 ja 6 samaaegselt; konkreetne tähistusviis sõltub kasutatavast programmist). Saab ka tähistada (- või ? või näit. **999** abil) tunnuse puudumist (näit. tunnust "saba pikkus" sabata loomal), samuti andmete puudumist.

4.3. Tunnuste transformeerimine on sageli vajalik ühe või teise klassifitseerimisprogrammi kasutamisel, mis nõuab ainult kvalitatiivsete (või isegi ainult binaarsete), või ainult kvantitatiivsete tunnuste kasutamist. See on puht-tehniline küsimus: näiteks võime värvust kirjeldada kas ordineerimata **kvalitatiivse** tunnusega (sinine; kollane...) või mitme **binaarse**, kaheastmelise tunnusega (sinine värvus on olemas või puudub; kollane on olemas või puudub...). Ordineeritud kvalitatiivsed tunnused võib transformeerida nii binaarseks kui ka **aditiivseiks**: näiteks võime ühe neljaseisundilise tunnuse kodeerida (transformeerida) kolme kvaasi-tunnuse abil:

	Aditiivne kodeerimine	Binaarne kodeerimine
Seisund	Kolm tunnust	Neli binaarset tunnust
0	0 0 0	1 0 0 0
1	1 0 0	0 1 0 0
2	1 1 0	0 0 1 0
3	1 1 1	0 0 0 1

Binaarne kodeerimine sobib ka ordineerimata tunnuste jaoks.

Kvantitatiivsed tunnused võime transformeerida kvalitatiivseiks nende klassideks jagamise teel (näiteks lehe laius 2-4 cm on esimene tunnuse seisund, 4,1-6 cm teine, 6,1-8 on kolmas kvalitatiivse tunnuse "laius" sei-

sund, jne.); saame ordineeritud kvalitatiivsed tunnused. Iga lehe suuruse klassi võime käsitada ka hoopis omaette loogilise (binaarse) tunnuseks.

Meristilisi tunnuseid saab väljendada ordineeritud või ordineerimata kvalitatiivsete tunnustena; näiteks võime tunnuse *kroonlehtede arv* (4; 5; või 6) väljendada seisunditena 1; 2; 3, või siis kolme binaarse tunnuseks.

4.4. Tunnuste informatiivsus oleneb tunnuse seisundite jaotumusest uuritud OTUdel ja on määratav Shannoni valemi abil. Arusaadavalt on kõigi uuritavate OTUde juures samas seisundis esineva tunnuse informatiivsus null; kui igal objektil on tunnuse seisund erinev, on see väga informatiivne tunnus. Informatiivsuse määr on oluline näit. "määrajate" koostamisel - informatiivsete tunnuste eeliskasutamine võimaldab koostada lühikesi ja seega mugavaid määramistabeleid.

Informatsiooni mõõt on sama mis entroopia puhul: korratuse määr suletud süsteemis (bittides) $H_a = -\sum p_k \cdot \log_2 p_k$ ($\log_2^x = \log_{10} x / \log_{10} 2$)

kus **a** on mingi tunnus, millel on mitmed erinevad seisundid **k**; **p_k** on iga seisundi tõenäosus (s.t. meie juhul - esinemissagedus uuritud OTUdel).

Valemist selgub, et tunnus on informatiivsem, kui selle eri seisundid on ühtlaselt jagunenud OTUde vahel; kui tunnusel on rohkem eri seisundeid. See ei tähenda, nagu oleks tegelikult klassifitseerimises kasulik kasutada paljude seisunditega tunnuseid: see aitab küll eristada OTU-sid, kuid mitte neid klassifitseerida rühmadesse.

4.5. Fülogeneetilise (kladistilise) analüüsi puhul on suhtumine tunnustesse ja nende eristamine pisut teistsugune ja rangem. Tunnuse seisundeid vaadeldakse ajalooliselt kujunenutena. Varem esinenud, ürgset, evolutsiooniliselt vanemat tunnust (tunnuse seisundit) nimetame **plesiomorfseks** (tunnuseks), sellest kujunenud homologset, hilisemat, nooremaga **apomorfseks**. Praegu kasutatavate algoritmide kohaselt võivad tunnused olla ainult kvalitatiivsed, mitte aga kvantitatiivsed; viimased tuleb transformeerida kvalitatiivseks. Analüüsi alguses ei tarvitse me veel teada, milline tunnus on just apomorfne; see selgub analüüsi tulemusena. Olulisim on seejuures, et tunnused (tunnuse seisundid) oleks tõepoolest homologsed, omavahel võrreldavad.

5. Homoloogia on algselt võrdlev-anatoomiline ja -morfoloogiline mõiste, samal ajal ka "morfoloogia keskne kontseptsioon" (J. Huxley, 1928), mis on äärmiselt oluline nii organismide evolutsiooni käsitleluses kui ka süstemaatikas (eriti fülogeneetilises süstemaatikas). Terminit kasutatakse alates R. Oweni tööst selgrootute kohta (1843), käsitus on muidugi vanem ja esineb näiteks ka Goethe uurimuses metamorfooside kohta taimedel (1790). Owen defineeris homollogi kui "sama organ erinevatel loomadest mistahes vormi ja funktsiooni teisendis". Kõneldes tunnuste (tunnuse seisundite, transformatsioonireala liikmete) homologsusest peame silmas nende evolutsiooniliselt sama päritolu, nende kujunemist üksteisest või tunnusest, mis esines nende eellastel. Süstemaatikas mingi organismirühma uurimisel on üks esimesi ja tähtsamaid ülesandeid kasutatavate **tunnuste (ja tunnuse seisundite) loetelu** koostamine. Pealiskaudse suhtumise korral (või ükskõiksuse puhul - et olgu

me süsteem kasvõi kunstlik) on see lihtne ja seisneb peamiselt teiste teadlaste poolt taksoneid kirjeldades kasutatud tunnuste koondamises. Korrektese uurimistöö puhul on tegemist **teaduslike hüpoteeside** püstitamiselega selle kohta, millised eri organismide juures esinevad tunnused on homoloogsed ja millised mitte (vaid näiteks **analoogsed**, konvergentsi teel tekkinud).

Lisaks konvergentstile võib mittehomoloogsete tunnuste sarnasus olla tingitud **parallelismist** - sama tunnuse tekkest tavaliselt suguluses olevates rühmades (Vavilovi paralleelse muutlikkuse fenomen!), **mimikrist**, ka **juhuslikust** sarnasusest. Analoogilised tunnused võivad endast kujutada ka **näivat sarnasust**, mis täpsemal uurimisel 'kaob'.

5.1. Homoloogia kriteeriume pole kuigi palju ja nad on sageli ebakindlad. Esimeseks on muidugi morfoloogiline (anatoomiline, jne.) **sarnasus ehituses ja struktuuris**; see esineb ka analoogia puhul, kuid struktuuri hoolosal uurimisel võib neid sageli eristada. Oluline on homoloogse organi või struktuuri erinevatel organismidel eri kujul esinemisel **üleminekuvormide** olemasolul. Teiseks on homoloogsete organite sarnane (suhteline) **asend** eri organismide juures. Reeglina ei esine ühel organismil mitut homoloogset organit (näiteks ei saaks inglitel korruga esineda tiibu ja käsi - juhul, kui tiib on esijäseme homoloog). Erandiks on metameersed organismid (taimelehed!), aga ka näit. hümeeniumielementide esinemine 'kõrgematel' seentel. Homoloogsed organid on sama või sarnase **ontogeneesiga**. Tuleb aga arvestada, et evolutsioonis võib ontogenees muutuda mitmel viisil: mitte ainult tunnuste lisandumise teel (Haeckel), vaid ka mitteterminaalse lisandumisega, teatud ontogeneesi etappide kadumisega, asendumisega. Homoloogia kaudseks, kuid heaks kriteeriumiks on **konvergenstest**: kui kahel võrreldaval taksonil esineb mitmeid homoloogseid tunnuseid, võime ka kahtluseluse homoloogia puhul eeldada homoloogia tõesust.

Tõsiseks komistuskiviks süstemaatikule on **latentne homoloogia**: mõni tunnus võib olla fülogeneesis mitte kadunud (*reversioon*), vaid olla mingil geneetilisel põhjusel maha surutud ja võib hiljem kergesti taastuda.

5.2. Tunnuste homoloogia pole nii oluline siis, kui meie ülesanne piirub "määrarajate" koostamisega: näiteks on taime õie värvus hea diagnostiline tunnus sõltumata sellest, et see võib olla põhjustatud erinevatest pigmentidest. Ka feneetikas (üldisest sarnasusest lähtuvas klassifitseerimises) pole homoloogiate leidmine nii otsustava tähtsusega kui fülogeneetilises süstemaatikas, kus eksimus tunnuste homologiseerimisel annab suure vea.

5.3. Tavalisemaks veaks tunnuste nimekirja koostamisel süstemaatiku poolt on mitme tunnuse n.ö. segamini kasutamine: ühe tunnusena püütakse näidata mitut erineva tunnust või nende kompleksi.

6. Klassifitseerimine on üks loogika osa: ühikute või nähtuste rühmitamine ja rühmadele nime andmine. Klassifitseerimine lähtub *suhte* olemasolust objektide vahel. Klassifitseerimine on kommunikatsiooni-, suhtlemisvahend inimeste vahel: tohutu arvu üksiknähtuste liigitamine väiksemaks arvuks klassideks võimaldab nendest ülevaadet saada, neis orienteeruda. Nii on klassifitseerimise aluseks teatud paradigma olemasolu ja inimeste praktiline vajadus.

6.1. (Aristotelese; 384-322 e.m.a.) **loogika** järgi on **liigitamine** mõiste (= esemete või nähtuste klassi) mahu avamine nende jaotamise teel (alam)klassideks. Liigitamise aluseks on *tunnus* ehk atribuut, liigitamise tulemuseks on *liikmed*). Loogiliselt korrektse liigitamise puhul peab sellel olema üks alus; mõiste maht peab hõlmama kõik liikmete mahud; liikmed peavad üksteist välistama.

Loogikas on klassifitseerimine liigitamise erijuht: klassid jagatakse alamklassideks, need omakorda alam-alamklassideks jne. Igal liikmel on oma üks ja kindel koht selles **hierarhilises süsteemis**. Selline klassifikatsioon on informatsioonirikkam kui mittehierarhiline, kus samaväärseina eristatakse suur hulk klasse.

Märkigem, et loogiline klassifitseerimine on hulga **jagamine** alamhulka-deks, niisiis üks *divisiivne* meetod. Bioloogilises klassifitseerimises on tänapäeval esikohal liikide ja teiste taksonite **ühendamine** (loogika) klas-sideks selliselt, et tulemusena on kõigil elusolenditel oma koht süsteemis (klassifikatsioonis): meetod on **aglomeratiivne** ehk kuhjav.

Loomulikus süsteemis on aluseks oluline tunnus (tunnused). **Kunstlikus** süsteemis on liigitamise (klassifitseerimise) alus kokkuleppeline - näiteks tähestikuline, või formaadijärgne raamatute paigutus raamatukogus (nagu seda on raamatud ELUS-is või Morton Arboretum'i raamatukogus USA-s).

6.2. Elusolendite süstematiseerimine algas (teaduslikul alusel) ilmselt Aristotelesega, kes jagas loomad verrega ja vereta loomadeks, need edasi 'perekondadeks' - näit. kalad, linnud, imetajad. Theophrastos (371-268 e.m.a.) eristas puud, pöösad, rohttaimed ja nendel 'loomulikud sugukonnad', mis näiteks vastavad tänapäevastele liilialistele, sariköielistele. Clusius eristas 1698. a. kaks rühma seeni: söödavad ja mürkseened. See on tüüpiline kunstliku süsteemi näide: klassifitseerimise aluseks pole organismide eneste olulised tunnused, vaid inimese suhtumine seente kasutamisse. Tänapäevasele süstemaatikale pani aluse Linné, kes lähtus hulgast läbimõeldud printsiipidest ja võttis kasutusele **binaarse nomenklatuuri** - igal liigil on kahest sõnast koosnev ladinakeelne nimetus, millest esimene on perekonna nimi. Nii loetaksegi taimede (incl. seente) süstemaatika formaalseks alguspunktiks tema tööd 1753. aastast ja loomade osas 1758. aastast. - Linné süsteem oli tüüpiline kunstlik süsteem; hiljem jõudis ta loomulike süsteemi kujundamise alguseni. Tänapäeval on bioloogias ainuvaldav taotlus luua loomulikku süsteemi. Selleks on mitu teed sõltuvalt süstemaatika poolt rakendatavast paradigmast (teooriate ja vaadete, tõekspidamiste kogumist). Vaatleme neist nelja.

6.2.1. Empiiriline (artistlik-intuitiivne) ehk kogemuslik süstemaatika, mis on enam kunst kui teadus, oli veel hiljuti ainuvalitsev. See tugineb enam või vähem subjektiivselt hinnatud ja tõlgendatud *sarnasusel* ja (osal süstemaatikuist) *oletataval evolutsioonil*. Tegelikult on enam või vähem ebateadlikult kasutatud *feneetilist* metodoloogiat. Tulemuseks on püüde puhul ka võimalikku evolutsiooni näidata *narratiivne*-jutustav hüpotees antud rühma oletatava evolutsiooni kohta, tavaliselt parafüleetilised, osalt ka monofüleetilised ja polüfüleetilised alljaotused (taksonid), ja võimatus hüpoteesi reaalsusele vastavust kontrollida (falsifitseerida).

6.2.2. Feneetiline süstemaatika põhineb oletusel, nagu kajastaks üldine sarnasus küllalt hästi uuritavate organismide seoseid. See tähendab konvergentsi mitteamvestamist. See suund on leidnud arendamist alates nn. numbrilise taksonoomia meetodite väljatöötamisest käesoleva sajandi kuuekümnendail aastail, õitseage oli kuni kaheksakümnendate aastate alguseni.

Metoodiliselt seisneb feneetiline klassifitseerimine liikide (või teiste taksonite - lühidalt nimetatakse neid **OTU**-deks lühendina ingliskeelsest terminist *operational taxonomic unit*) **summaarse** erinevuse mõõtmises kasutatud tunnuste erinevuste alusel; kõigi OTUde omavahelise erinevuse (kauguse) alusel on neid võimalik rühmitada - moodustada **klastreid**. OTUsid võib kujutada punktidenä ja punktivarvedena hulgamõõtmelises tunnusruumis, mille vektoreiks on kasutatud tunnused; võib kasutada ka teisi **ordinatsioonimeetodeid**. OTUde rühmitamise aluseks on **üldine** sarnasus (või erinevus - need on sama mündi kaks poolt), mitte sarnasus või erinevus mingi konkreetse tunnuse poolest.

Saadud feneetiline klassifikatsioon pole ei tõene ega väär; ta ei ole teaduslik hüpotees, mida oleks võimalik kinnitada või ümber lükata. Tegemist on ainult algandmete (OTUde tunnuste ja tunnuse seisundite) ülevaatliku korrastamisega.

6.2.3. Fülogeneetiline süstemaatika (kladistika) eesmärgiks on klassifitseerida taksoneid nende ajaloolise kujunemise, nende fülogeneesi alusel. See põhineb ainult monofüleetiliste (mitte aga parafüleetiliste) rühmade tunnustamisel; tunnuste polariseerimisel (plesiomorfsete ja apomorfsete tunnuste eristamisvõimaluse mõõtmisel); võimalike evolutsioonipuude (kladogrammide) "joonistamisel" ja nendest säästuprintsiibi (*parsimony*) alusel kõige lühema, kõige väiksema tunnuste muutumise arvuga puu väljavalimisel. Saadud fülogeneesipuu on teaduslik hüpotees, mida on võimalik falsifitseerida ja/või kinnitada. Iga OTUde rühma (fülogeneesipuu haru ehk oksa) eristamise aluseks on teatud kindla tunnuse (tunnuste) kindlasuunaline muutus.

6.2.4. Evolutsiooniline süstemaatika taotleb kahe teineteist välistava, feneetilise ja fülogeneetilise metodoloogia sünteesi. Aluseks on püüd selgita uuritava rühma liikide (ja teiste taksonite) fülogenees ja saadud pilti "parandada" sarnasuse arvestamise teel, s.t. lähtudes feneetilisest metodoloogiast. Vastavalt sellele aktsepteeritakse parafüleetilisi rühmi; mõndakse võimalust rühmi eristada plesiomorfsete tunnuste alusel. Tulemusena kujutatakse fülogeneesi narratiivselt; et sarnasuse hindamiseks puudub objektiivne mõõdupuu, sõltuvad saadud klassifikatsioonid süstemaatiku suvast.

7. Feneetiline süstemaatika.

7.1. OTUde (näit. liikide) vahelise sarnasuse (erinevuse, distantssi) mõõtmine. - Sarnasus ja erinevus on teineteisest sõltuvad parameetrid. Kui sarnasust ja erinevust väljendada mingi koefitsiendiga, mis täieliku sarnasuse puhul saavutab väärtuse 1, on on mingi uuritud OTUpaari erinevus üks miinus sarnasus; analoogiliselt on sarnasus üks miinus erinevus.

7.1.1. Binaarsete kvalitatiivsete tunnuste kasutamine. - Tähistagem mingi tunnuse esinemist mõlemal võrreldaval OTU-l (*i*; *j*) tähega **a**, puudu-

mist **d**-ga, esinemist ainult esimesel OTU **b**-ga ja ainult teisel **c**-ga. Olgu võrreldud tunnuste üldarv **p**. Kasutades andmeid kõigi tunnuste kohta, võime arvutada liht-sarnasuskoefitsiendi (Sokali ja Micheneri, 1958 järele):

$$s_{ij} = (a + d) / p.$$

Jaccard'i sarnasuskoefitsient on $s_{ij} = a / (a + b + c)$. On ka teisi, vähem kasutatavaid koefitsiente; ainuüksi personaalarvuti-programm NTSYS-pc pakub neid üksteist. Kõigil puhkudel on kahe OTU sarnasus väljendatud arvuga, mille väärtus on alla ühe.

7.1.2. Mitme tunnuse seisundiga kvalitatiivsete tunnuste puhul võib need muundada binaarseteks tunnusteks (teatud seisund puudub või esineb) või arvestada ainult ühise tunnuse seisundi puudumist või esinemist.

7.1.3. Kvantitatiivsete tunnuste puhul on otstarbekas OTUde erinevust väljendada kaugusena üksteisest hüpoteetilises hulgamõõtmelises ehk hüper-ruumis (tunnusruumis), mille üksteisega mudelis risti olevateks vektoriteks on eri tunnused. Nagu juba märkisime, moodustavad OTUd selles punktide parvi.

7.1.3.1. Eukleidiline kaugus arvutatakse Pythagorase teoreemi alusel: liidetakse iga tunnuse kohta käiva kahe OTU erinevuse ruudud: $\sum (x_{ki} - x_{kj})^2$ (kus *k* tähistab mistahes tunnust, *x* selle mõõtmistulemust, *i* ja *j* aga võrreldavaid taksoneid), ja võetakse saadud summast ruutjuur. See algoritm on sama nii kahe tunnuse kui ka kuitahes paljude tunnuste puhul.

Kasutatakse ka eukleidilise kauguse asemel selle **ruutu**.

7.1.3.2. Kui enne juurimist jagame summa kasutatud tunnuste arvuga *n*, saame **keskmise taksonoomilise kauguse**.

7.1.3.3. Manhattani kvartalikaugus arvutatakse valemiga

$$M_{ij} = \sum |x_{ki} - x_{kj}| / n$$

Lähtudes oletusest, et muutuste aluseks on iga tunnuse osas eraldi toimunud mutatsioon(id), on Manhattani kvartalikaugus bioloogiliselt ja loogiliselt mõneti ehk põhjendatum kui teised mõõtmisviisid. Tegelikult on selline käsitlus muidugi üsnagi vedel: enamus tunnustest on ju **polügeensed**, s.t. nende ilmumine on tingitud mitme geeni koosmõjust.

7.1.3.4. Kasutusel on veel vähemalt 18-20 muudki koefitsienti, kaasa arvatud vektoritevahelise nurga (koosinuse) kasutamisega. Kaks neist väärivad märkimist:

Smirnoffi koefitsient **kaalub** tunnuseid: tunnus(e) seisund), mis on harvaesinev, muutub väga oluliseks - seda enam, mida OTU- (liigi-)rohkem on uuritav kogum. Kaalumine on süstemaatikas väga vaieldava õigustusega; Smirnoffi koefitsient võimaldab aga leida liikide kogumist teistest väga erinevaid, omapäraseid liike, mis väärivad erilist tähelepanu. - See meetod leidub programmpaketis TYTAN.

Canberra meetrikat kasutav algoritm võimaldab 'head' rühmade eristamist uuritavas kogumis: üksteisele 'liiga' ligidal asuvad OTUd eristatakse paremini, teistest 'liiga' kauged (erinevad) tuuakse ligemale:

$$d_{ij} = 1/n \sum (|x_{ki} - x_{kj}| / \sum (x_{ki} - x_{kj}))$$

7.1.3.5. Kasutades OTUde vahelise erinevuse (distanti) määramiseks erinevaid kvantitatiivseid tunnuseid koos, tuleb neid kuidagi ühtlustada: ei

saa ju võrreldavana kokku panna elevandi pikkust millimeetrites ja raku läbimõõtu samas mõõtühikus. Kooskõlastamiseks on tavaliselt kasutusel kaks viisi: tunnuse väärtuste **standardiseerimine** ja **ranzheerimine**.

7.1.3.5.1. Standardiseerimisel muundame distantssi (erinevuse) iga tunnuse puhul eraldi, arvestades antud tunnuse varieerumist kõigi OTUde juures. Selleks arvutame iga tunnuse kohta selle aritmeetilise keskmise ja standardhälbe s . Uus erinevuse väärtus $x' = (x - \bar{x}) / s$

7.1.3.5.2. Mastaapimisel ehk ranzheerimisel (Goweri meetod) muundatakse distantssi väärtus arvestades antud tunnuse piirväärtusi uuritud OTUde juures:

$x' = (x - x_{\min}) / (x_{\max} - x_{\min})$ kus x' on distantssi uus, transformeeritud väärtus.

7.1.3.6. Mistahes OTUde paaride erinevust (distantssi) mõõtvat meetodit kasutades saame andmed, mis võimaldab koostada andmete tabeli (vt. näide: joon. 1) ja **andmematriksi** (näide: joon. 2); selle tulpades ja ridades on eri OTUd (näit. liigid), nende ristumiskohal OTUde vaheline kaugus (erinevus, distantss). Seda saame kasutada kas OTUde klasterdamiseks või paljude ordinatsioonimeetodite kasutamiseks.

7.2. Teades iga kahe OTU-paari omavahelist kaugust (= **summaarset** erinevust kõiki kasutatud tunnuseid arvestades) saame modellerides luua uuritud OTUde (taksonite) geomeetrilise (graafilise) mudeli. [Meenutagem, et **graaf** on punktide hulk, milles vähemalt osa punkte on ühendatud joontega.] Selleks on mitu võimalust, mida võime realiseerida **ordinatsioonimeetodeid** või **klasteranalüüsi** kasutades. Allpool vaatleme klasteranalüüsi teostamist.

7.2.1. Enim kasutatud vorm on **klasteranalüüs**, mille tulemuseks on harunev, sarnasust kujutav **fenogramm**. Sarnased OTUd on selles ühendatud hierarhiliselt harunenud ühise "juurega" puu harudeks, kobaraiks, **klastreiks**. Klasterdamismeetodeid on rohkelt, neist levinumad on lähimnaabri meetod (*simple-linkage clustering*), aritmeetilisi keskmisi kasutav kaalumata paaride meetod (UPGMA - *unweighted pair-group method using averages*) ja täissideme-klasterdus (*complete-linkage clustering*).

7.2.1.1. Lähimnaabri meetodi puhul leitakse kaks OTUt, mille erinevus on minimaalne, ja ühendatakse need arvestades OTUde distantssi teineteisest (joon. 2). Edasi arvutatakse distantss selle klastri ja ülejäänud OTUde vahel; järgmisena ühendatakse klastrisse OTUd (või üks OTU ja juba olemasolev klaster), mille distantss on väikseim. Nii jätkates saadakse fenogramm, milles üksteisele kõige sarnasemad taksonid on rühmitunud. Teisiti öeldes: OTU kuulub ühte klastrisse siis, kui tal on teatud läve ületav sarnasus vähemalt ühe antud klastri liikmega (Estabrook, 1986). (Kuidas seda teha, selle kohta on näide toodud konspekti lisas nr. 1).

7.2.1.2. Täissideme-klasterduse puhul ühendatakse OTU olemasoleva klastri juurde mitte juba klastris olevat kõige sarnasemat, vaid kõige vähemsarnast OTUt erinevuse mõõtmisel arvestades.

7.2.1.3. UPGMA-klasterduse puhul ühendatakse osaklastrid omavahel mitte arvestades kumbagi klastri kõige sarnasemaid liike, vaid sarnasuse (erinevuse) keskmisi. Lisatakse OTU (või teine juba konstrueeritud klaster), mille keskmine kaugus kõigist klastri OTUdest on minimaalne.

7.3. Eri klasterdusmeetodid annavad tavaliselt isegi samu distantsiandmeid kasutades erineva fenogrammi, s.t. erineva klassifikatsiooniskeemi. Kui arvestada ka taksonite erinevuse / sarnasuse mõõtmise erinevaid meetodeid, võime vastavat arvutiprogrammi kasutades samadest algandmetest saada vähem kui tunni jooksul 60-100 erinevat fenogrammi. Neist **ükski pole õige ega ükski pole väär**: oleme lihtsalt oma algandmeid mitmesugusel viisil koristanud, ordineerinud. Iga fenogramm on nagu uuritavate taksonite tunnuste erineva vaatenurga alt nähtud pilt.

Ometi on süstemaatika praktikas mõned meetodid eelistatumad. Lähimnaabri meetod võimaldab eelkõige näha seost, "üleminekuid" erinevate taksonite vahel. UPGMA on kasutatavaim, n.ö. standardne meetod, ja võimaldab seetõttu eri teadustöodes ilmunu võrdlemist. Kui kasutatakse seejuures, või täissidemeklasterduse puhul ka Canberra meetrikas avaldatud distantse, saame enamasti uuritud OTUde selge rühmitumise ilusateks klastriteks. See kõlbab siis, kui tahame näiteks mõnda perekonda jagada alamperekondadeks või sektioonideks.

7.3.1. Erinevatel meetoditel saadud fenogramme võime hinnata nende vastavuse alusel algandmestikule. Leiame korrelatsiooni kõigi OTUpaaride vaheliste kauguste vahel fenogrammis ja algses andmemaatriksis; vastav kofenetiilise korrelatsiooni koefitsient (*cophenetic correlation coefficient*) on seda suurem (teoreetiliselt kuni 1,0), mida vähem on klasterdamine moonutanud liikidevahelisi sarnasussuhteid.

7.3.2. Erinevatel meetoditel samade OTUde kohta saadud erineva informatsiooniga fenogramme võime liita üheks, **konsensuspuuks** (vt. joon. 3). Selles on omaette klastritena näidatud need harud (klastrid), mis on ühised kõigile eri fenogrammidele. Võime genereerida ka sellise konsensuspuu, kus on esitatud 50 %, 90 % jne. fenogrammidest esinevad klastrid. Selline konsensus näitab sarnasust OTUde vahel, mis ei olene kasutatud meetodite iseärasustest. Väga sageli kaob aga konsensuspuus suur hulk klastreid: **resolutsioon** on sellisel juhul madal.

7.3.3. Fenogramm on nn. **juurega puu** (suunatud ehk orienteeritud graaf) - ta aheneb alusel ainsaks tüveks. Tüve asend ja harunemise viis ei sisalda aga informatsiooni kasvõi hüpoteetilise fülogeneesi kohta uuritavas rühmas: puu näitab selgelt arusaadaval viisil üleüldist sarnasust, mitte midagi muud.

7.3.4. Fenogrammide kasutamine uuritud organismide süsteemi (klassifikatsiooni) **loomiseks** on arutusel olnud peamiselt 1950-1960-ndail aastail. Väideti, nagu võiks fenogrammi harunemise tasemete (sarnasuse / erinevuse nivoode) järgi eristada alamperekondi, perekondi jne. Veel 1990 arvas evolutsioonilise süstemaatika pooldaja (ja tuntud taimesüstemaatika õpiku autor) T.S. Stuessy, et selline hierarhia standardiseerimine pole võimalik eri organismirühmade puhul, küll aga saaks tulevikus sellise võtte kasutamist kokku leppida ligidaste rühmade uurimisel. - Et fenogrammis ei avaldu kindlate tunnuste muutumine, vaid sarnasus paljude tunnuste "segu" puhul, oleks nii eristatavad taksonid sageli küll halvasti iseloomustatavad.

7.4. Ordinatsioonimeetodeist kitsamas mõttes (s.o. välja jättes klasterdamise) on süstemaatikas levinuim **peakomponentanalüüs**.

7.4.1. Peakkomponentanalüüs (*principal component analysis*) seisneb hulga-mõõtmelises tunnusruumis endiste vektorite (kasutatud tunnustega määratud telgedes) asendamises uutega, mis täielikumalt hõlmavad koguvarieeruvust. Esimese telje (vektori, komponendi) asend valitakse selline, et ta (oma pikkuses) haaraks maksimaalse osa üldvarieeruvusest (oleks OTUde ellipsoidse pilve pikiteljeks). Selleks kasutatakse vähimruutude meetodit: OTUpunktide teljest kauguse ruutude summa peab olema minimaalne. Teine, eelmisega risti asetsev telg hõlmab jällegi maksimaalselt võimaliku varieeruvuste summa, jne. (Vt. joon. 4.) Iga uus telg väljendab nüüd mitte varieeruvust ühe tunnuse piires, vaid kombineerib mitme (paljude) tunnuste varieeruvuse. Kui meil hästi läheb, on esimese kolme vektoriga hõlvatud näit. 75 või isegi 85 % koguvarieeruvusest.

Edasi on juba lihtne kujutada OTUsid punktidenä tasapinnal, millel koordinaatideks (telgedeks) on 1. ja 2. komponent, 2. ja 3. komponent, 1. ja 3. komponent. Näeme (kui näeme), et OTUd on grupeerunud (vt. joon. 5); punktid ei asetu ühtlaselt-juhuslikult, vaid rühmadena. Nende rühmade alusel võime teha järeldusi taksonite omavaheliste *suhete* kohta.

Peakomponentanalüüsi saab teha vaid kvantitatiivseid tunnuseid kasutades, soovitatavalt eukleidilist OTUdevahelist erinevust kasutades. Saadud graafilisel kujutisel on OTUpunktide vaheline kaugus (teatud) vastavuses selle erinevusega (vt. joon. 3). Kasutatud on ka binaarsete tunnuste kasutamisele tuginevat peakomponentanalüüsi, kuid selle "interpreteerimine võib tekitada raskusi" (Dunn & Everitt, 1982).

7.4.2. Peakoordinaatanalüüs (*principal coordinates analysis*) võimaldab kasutada ka mitte-eukleidilisi distantse (erinevusi), kusjuures hulga-mõõtmelise ruumi distantside moonutamist püütakse minimeerida. Ka siin saame tasapinnalise, OTUpunktide rühmadega pildi.

Ordinatsioonimeetodeid on teisigi, kuid tänapäevases fülogeneetilises süstemaatikas nad eriti kasutamist ei leia. Vajadusel saab neid hõlpsalt genereerida arvutiprogrammidega NTSYS-pc ja SYSTAT; kasutatavate algoritmide lühikirjelduse võib leida näit. raamatus *G. Dunn & B.S. Everitt, An introduction to mathematical taxonomy. Cambridge University Press, Cambridge et al., 1982.*

7.4.2.1. Minimaalkauguste puu (*minimum spanning tree*) on niisugune graaf, milles OTUpunktid (juba olemasoleval, näit. peakoordinaatanalüüsiga saadud tasapinnalisel kujutisel) on omavahel joontega ühendatud vastavalt nende (algsele, mitte kujutisel mõneti moonunud) omavahelistele distantsidele (erinevustele). See on tüüpiline matemaatilises topoloogias tuntud *rändkaupmehe ülesande* lahendus: joontega seotakse OTUpunktid nii, et joonte kogupikkus oleks minimaalne. Piiravateks tingimusteks on, et 1) ei või esineda suletud, kinniseid "silmuseid"; 2) igasse punkti viiks vähemalt üks joon; 3) kõik punktid peavad olema joon(t)ega ühendatud. Sellise puu näide on toodud joonisel 6. Saadud graaf (koos peakoordinaat- või muu analüüsi tulemuste kujutisega) esitab meile informatsiooni OTUde sarnasusseoste kohta. Kuivõrd see klassifitseerimises kasulik on, see on iseküsimus; igal juhul jätab (jättis) ta teadustöö illustratsioonina mulje sügavast teaduslikust analüüsist ja matemaatiliste meetodite valdamisest.

7.4. Feneetiliste meetodite kasutamist saab õigeks lugeda siis, kui süstemaatik täie arusaamisega ja läbimõeldult soovib lähtuda ainult sarnasusest, ja ei taotle seejuures või peab millegipoolt ebaõigeks fülogeneetilist lähenemist. See on sama võimalik kui teadlikult kunstliku süsteemi loomine.

Teatud erandjuhtudel on feneetiliste meetodite kasutamine siiski hädaabinõuna mõistatav. Väga sarnaste liikide puhul, kui uurijal pole võimalik leida kladistiliseks analüüsiks sobivaid tunnuseid (vaid need on näit. peamiselt arv-tunnused), on mõeldav esimese lähendusena teostada feneetiline klassifikatsioon, mis võimaldaks liigirikkas perekonnas esialgselt orienteeruda. Teatud molekulaarsüsteematiliste tunnuste puhul on samuti feneetiline meetodika praktiliselt mõõdapääsmatu; et sel juhul tunnused on genotüüpi iseloomustavad, on tulemus palju enam geneoloogilisi seoseid väljendav kui fenotüüpi iseloomustavate "morfoloogiliste" tunnuste puhul.

8. Fülogeneetilise süstemaatika eesmärgiks on klassifitseerimine taksonite *genealoogia*, nende ajaloolise kujunemise, s.t. nende evolutsiooni alusel. Lähtutakse järgmistest põhiprintsiipidest:

8.1.1. Klassifitseerimisühikuteks, s.t. taksoniiks saavad ja tohivad olla ainult **monofüleetilised** rühmad selle termini täpses (kitsamas tähenduses). Parafüleetiliste rühmade mõistet vt. **3.2.** ja joon. 7.

8.1.2. Tunnustel tuleb eristada sellised tunnuse seisundid (tunnused transformatsiooniseerias), mis on omavahel diskreetselt erinevad (kvalitatiivsed ordineeritud või ordineerimata tunnused), ja mille seas saaks eristada **plesiomorfset** (-eid) ja **apomorfset** (-eid) (vt. **4.5**).

8.1.3. Taksonite rühmitamise aluseks saavad olla ainult neile ühised **apomorfset** tunnused (**sünapomorfid**), mitte aga ühised plesiomorfset tunnused (sümplesiomorfid).

8.1.3. Süsteemi (klassifikatsiooni) aluseks on **hüpoteetiline fülogeneesipuu** (kladogramm), mis näitab tunnuste muutumist oletatavas evolutsioonis, mille tulemusena kujunes uuritavatel taksonitel praegu olemasolev tunnuste komplekt.

8.1.4. Paljudest võimalikest hüpoteetilistest fülogeneesipuudest tuleb välja valida - vastavalt **säästuprintsiibile** - kõige lühem, s.t. selline, kus on kõige väiksem arv tunnuste muutumise samme. Kui selliseid puud saadakse mitu, on tegemist võrdvõimalike hüpoteesidega, mille vahel on küll võimalik valikut teha muudest printsiipidest lähtudes.

Lühima fülogeneesipuu eelistamine ei tähenda, nagu eeldaks me evolutsiooni käiku lihtsana ja väheste muutussammude produktina. Säästuprintsiipi rakendatakse **hüpoteeside** võrdleval hindamisel.

8.2. Monofüleetilisus on kriitiline nõue mingi rühma kladistiliseks uurimiseks. Selle kohta otsustuse tegemine on mõnikord küllalt raske ja nõuab kindlasti küllaldasi eelteadmisi rühma liikide kohta, nende senise klassifitseerimise ajaloo tundmist. Paraku eristatakse uuritav rühm sageli feneetiliste (eelkõige sarnasust arvestavate) kriteeriumide alusel.

Kahtluse korral (ja hea analüüsi jaoks üleüldse) on soovitatav uurimise alla võtta mitte ainult põhiobjekt, vaid sellega koos ka oletatavasti kõige

lähedasemad naaberrühmad.

8.3. Tunnused. Et kladistikas vaadeldakse tunnuse seisundeid kui homoloogseid, muutumisseriesana seotud **transformatsiooniseeria (TS)** liikmeid (tunnuseid), loobume edasises *tunnuse seisundi* terminist.

TS võib olla kahetunnuseline (binaarne) või paljutunnuseline. Viimane võib olla ordineerimata või ordineeritud. Ordineeritud tunnuste puhul eeldame, et muutumine (transformatsioon) on toimunud üle vaheastmete: seeria puhul **A--B--C** saab **A** muutumine **C**-ks (või vastupidi) toimuda ainult kahe muutussammuna, s.t. üle **B**. Apomorfsus ja plesiomorfsus on siin suhtelised: **kui** areng on läinud suunas **A**-st **C** poole, on **A** plesiomorfne tunnus; **B** on **A** suhtes apomorfne, kuid **C** suhtes plesiomorfne; **C** on igal juhul apomorfne.

8.3.1. Tunnuste polariseerimine. Selle kindlakstegemist, milline tunnus on plesiomorfne ja milline apomorfne, nimetatakse tunnuste polariseerimiseks ehk tunnuse *argumenteerimiseks*. See on otsustava tähtsusega kladistilises analüüsis.

Mingi tunnuse (organi, selle osa, vms.) puudumist võiks pidada plesiomorfseks, olemasolu - apomorfseks. Kuid me teame paljude organite ja struktuuride taandarengut või kadumist evolutsiooni käigus; seda nimetame tunnuse **reversiooniks**. Polariseerimist võime teostada n.ö. oletuse korras, lähtudes varem esitatud hüpoteesidest, oletustest; see on kahtlase väärtusega argumenteerimine.

On oletatud, et sageli esinev tunnus peaks olema plesiomorfne, haruldasem aga apomorfne. Teatakse aga hulgimalt vastupidiseid näiteid.

Paleontoloogilise andmestiku olemasolul võib polariseerimine olla küllalt lihtne; paljude organismi(rühma)de kohta aga see puudub.

Kasulikuks on peetud **ontogeneetilist** meetodit, mis lähtub Haeckel'likust rekapitulatsioonireeglist: evolutsiooniliselt vanemad tunnused ilmnevad varem kui hilistekkelised (s.t. apomorfsed). Tegelikult on see rekapitulatsioon ainult üks erijuhus ontogeneesis. Ontogeneesi muutumises evolutsiooni käigus võivad esineda ka *deletsioon* (mõne vahepealse arenguastme elimineerimine), ontogeneesis mõne vahelmise etapi lisandumine, jne.

8.2.1.1. Välisrühma meetod. Alates 1980-ndate aastate algusest on ainsaks objektiivseks polariseerimise meetodiks peetud seda.

Siserühmaks (*ingroup*) nimetatakse uuritavate taksonite (näit. liikide) hulka. **Välisrühm** (*outgroup*) on sellele evolutsiooniliselt lähedane sugulorganismide rühm (näiteks ligidaseks peetava teise perekonna liigid, või see perekond ühise tunnuste kirjeldusega iseloomustatuna). Kui välisrühm kujutab endast kõige lähedasemat taksonit (välis- ja siserühm on kujunenud otseselt ühiseellasest), nimetame seda **sõsarrühmaks** (*sister group*).

Kui mingi tunnus esineb nii välis- kui ka siserühmas, võime oletada selle pärimist nende ühiselt (hüpoteetiliselt) eellaselt. Kui homoloogne tunnus, sama TS liige esineb ainult siserühmas, on see ilmselt hilisema tekkega ja on järelikult apomorfne (ning välis- ja siserühmale ühine tunnus on plesiomorfne).

Kasutades ainult ühte välisrühma võib meetodi kasutamisel tekkida eksitus: võib-olla oli tunnus ka ühiseellasel olemas, kuid välisrühma kujunemisel taandarenes? Järelikult on parem kasutada kahte või kolme välisrühma,

et jälgida TS tunnuste võimalikku muutumist pikema evolutsiooniaja jooksul.

8.3. Vastavalt kladistika metodoloogiale ei iseloomusta monofüleetilist taksonit mitte nende liikidele ühised plesiomorfseid, vaid sünapomorfseid tunnused. Tõepoolest, me võime leida väga suure hulga tunnuseid, mis puuduvad nii inimesel kui ka küülikul. Kui suur see arv ka ei oleks, nende liikide kuulumist ühisesse taksonisse (ühete perekonda) need kinnitada ei saa. Teiselt poolt on inimesel ja shimpansil hulk ühiseid tunnuseid, mis õigustavad nende asetamise samasse (kõrgemasse kui perekond) taksonisse.

Ühised plesiomorfseid tunnused ühendavad varem süstemaatikas kasutatud taksoneid *selgrootud*, *taimed* (kaasa arvatud kõik vetikad), *seeded* (kus nende hulka loeti ka munasseened, Oomycetes ja sageli ka limasseened). Need on parafüleetilised rühmitused.

Tuleb eristada **autapomorfseid** tunnuseid, mis iseloomustavad ainult ühte taksonit, s.t. ei ühenda mitut taksonit neil kõigil esinemisega. Autapomorfne tunnus võimaldab küll ühte taksonit tema söstartaksonist *eristada*, kuid mitte neid süstematiseerida, s.t. sugulustaksoneid *ühendada*. Järelikult on autapomorfseid tunnused kladistilisel analüüsil kasutatud, ja kui me programmile vastava käsu anname, neid lihtsalt ignoreeritakse. See lihtsus tab ja ka kiirendab analüüsi.

8.4. Kladogrammide koostamisel "joonistatakse" tunnuste ajas muutumist rekonstrueerides selline puu, mille harutippudeks on uuritavad taksonid ja mille tüve alusel (juuritud puu puhul) on hüpoteetiline ühiseellane - liik. Ühise päritoluga (sünapomorfidega iseloomustatud) liigirühmad moodustavad puu ühe "oksa" ehk **klaadi**. Iga tunnusemuutust (üleminekut plesiomorfsest tunnusest apomorfseks, **reversiooni** ehk taandarengut tagasi plesiomorfseks, sama tunnuse paralleelset teket arengupuul eri harudes (**parallelismi**) loendame ühe **sammuna**; sammude koguarv näitab kladogrammi (puu) **pikkust**.

8.4.1. Olenevalt uuritavate taksonite (harutippude) arvust võime konstrueerida **suure hulga erineva harunemisega puid**. Kui kasutame binaarseid (kaheastmelisi) tunnuseid, on võimalike puude arv:

taksonite arv - puude arv

2	2
3	9
5	62
20	$5.24 \cdot 10^{24}$
100	$1.0 \cdot 10^{198}$

Märkigem siinjuures, et tänapäevase arvutustehnika kasutamisega nõuaks 100-taksoniliste kõigi võimalike puude kuju väljaarvutamine aega $6 \cdot 10^{179}$ aastat, universumi vanus on aga alles ligikaudu $2 \cdot 10^{10}$ aastat.

Ilmselt vajab lahendamist kaks praktilist küsimust: kuidas paljude puude seast välja valida üks või piiratud hulk, mis oleksid mingite objektiivsete kriteeriumide alusel sobivad fülogeneesihüpoteesidena; kuidas piirata kõikvõimalike puude arvu juba nende genereerimisel mingite piirangute panemisega.

Esimesele küsimusele annab lahenduse **säästuprintsiip**, millel juba peatusime: kõikvõimalikest puudest valime välja ainult selle (need), mis on

kõige lühem(ad) - kõige väiksema tunnuste muutumissammude arvuga. Teise küsimuse lahendus tuleneb esimesest: tuleb kasutada **heuristilisi** algoritme, mis kõikvõimalike puude konstrueerimisel väldiksid juba leitud puust (puudest) pikemate genereerimise - ja seda soovitatavalt ennetavalt.

8.5. Puude genereerimiseks (vt. joon. 8) on paar head personaalarvuti programmi; neist tuntumad, kättesaadavamad ja küllalt kiired suhteliselt suuremate andmemaatriksite (taksonite ja tunnuste hulga) jaoks on kaks.

PAUP versioon **3.1.1** on määratud kasutamiseks *Apple Macintosh*-tüüpi personaalarvutitel (mida hinna tõttu on Eestis veel suhteliselt vähe). Programmil on hulk võimalusi kasutada erinevaid meetodeid, kaasa arvatud molekulaarsüsteemaatika tunnuseid kasutavad. Versioon **2.4** on määratud kasutamiseks meil tavalistel IBM PC-tüüpi, matemaatilise koprotsessoriga varustatud personaalarvutitel, kuid on juba moraalselt vananenud ja palju väiksemate võimalustega. Selle uus versioon on tulekul, kuid saabumine võib võtta veel palju aega.

HENNIG86 on määratud kasutamiseks IBM PC tüüpi arvutitele ja ei nõua koprotsessori olemasolu. Programmi on palju kasutatud Skandinaaviamaades; piiranguks on väiksem kasutamisisvariantide hulk (**PAUP** 3.1.1-ga võrreldes), ka eelistatakse **PAUP**-i seetõttu, et selles kasutatavad algoritmid on vähemalt põhijoontes avaldatud, **HENNIG86** omad aga teadmata.

Lisaks pakub palju erinevaid meetodeid programm **PHYLIP**, mis pealegi on erinevalt eelmistest, kuigi üsna odavatest, kätte saadav tasuta. Vähegi suuremate andmemaatriksite jaoks on see programm aga (praeguses versioonis) väga aeglane.

Kui oleme (näiteks **PAUP**-i abil) saanud mingi kladogrammi - hüpoteetilise fülogeneesipuu, võime programmi **MacClade** abil otse ekraanil jälgida tunnuste muutumist puu harudes; võime oma suva järgi ümber tõsta puu harusid või üksikuid taksoneid teise kohta ja vaadata, kuidas see mõjustab puu kogupikkust.

8.6. Puude "ehitamisel" võime iga TS (tunnuste) kohta ette määrata, kuidas see tohib muutuda. Tähistame järgnevas seletuseks (binaarse transformatsiooniseeria) tunnused kui **0** ja **1**.

Dollo parsimoonia (kasutusel aastast 1974) puhul tähistame plesiomorfse seisundi **0**-ga. Võimalik on tunnuse ainult ühekordne muutumine **0** → **1** ja mingit paralleelset sedapidi muutumist rohkem ei lubata. Tunnus võib tagasi pöörduda (→ **0**), kuid selliste muutuste arv on minimeeritud.

Camin-Sokali parsimoonia (1965) puhul lähtutakse evolutsiooni pöördumata printsiibist: tunnuse reversion (**0** ← **1**) on lubamatu; tunnuse paralleelne teke puu eri harudes on lubatav, kuid selle hulk on minimeeritud. Sellist tunnuse iseloomustamist saab kasutada ainult siis, kui ehitame juurega puud.

Fitchi parsimoonia (1971) puhul võib (alati ordineerimata!) tunnus igas suunas vabalt muutuda (**0** → **1** ja **1** → **0**); minimeeritud on muutuste koguarv. See mudel on universaalne ja üldkasutatav.

Wagneri parsimoonia (selle põhimõtte formuleeriti 1969) on eelmisega sarnane, kuid seda kasutatakse, kui esineb ka ordineeritud tunnuseid, mis

tuleb aditiivselt kodeerida. (Mis on aditiivne kodeerimine, sada vt. 4.3.)

8.7. Tundub olevat ükskõik, kas me oma analüüsis kasutame sama tunnuste-komplekti (ühe tunnuse eri seisundeid) ordineerimata või ordineeritud tunnustena. Veel enam; võiks arvata, et ordineeritud tunnused peavad oma muutumises kindlasti läbima teatud vaheastmeid.

Tegelikult on iga tunnusteseeria ordineerimine (tunnuste järjestamine) üks täiendav, uuriija poolt juurde toodud lisahüpotees: et muutused just nimelt nii, üle selliste vaheastmete peaks käima. Seda võib ja isegi peab rakendama ainult siis, kui see on kladistilisele analüüsile eelnevate uuri-mistega kinnitatud. Kui see nii pole, on parem ordineerimisest hoiduda. Veel enam: kui me analüüsi alustades ordineerimisega väidame, et näit. tun-nus **A** peab tunnuseneni **C** muutumiseks kindlasti läbima vaheastme **B**, siis on sellise muutumise "hind" kaks muutumissammu. Programm hoidub sellist puud kahe sammu võrra pikendava muutuse kasutamisest. Kui **A**-st **C**-ni on ainult üks samm, on selline muutus võrdväärne mõne teise muutusega (näiteks **F** → **G**), mida programm esimesel juhul eelistaks.

8.8. Ometi on juhuseid, millal me päris hästi teame, kuidas tunnuste transformatsiooniseerias tunnuste muutumine peab olema toimunud; mingi tun-nuse **muutuste puu** (*character state tree*) võib isegi olla harunenud; me võime mõnda muutust (või selle suunda) pidada tõenäolisemaks kui mingit teist. Sellisel juhul on meil võimalik arvutiprogrammi sisestada meie poolt ette antud piirangud - meie poolt õigeks peetavate muutuste mudeli, mida programm on sunnitud järgima. (Vt. joon. 9.) Näiteks teame, et molekulaar-süsteemaatikas kasutatavate nukleotiidide sekvensi tunnuste puhul transver-sioon toimub harvemini kui transitsioon. Võime analüüsi eel ette anda seda arvestava tunnuse muudatuste puu (vt. joon. 10).

8.9. Puude ehitamisele asudes võime lähtuda meie poolt **ette antud hüpoteetilisest eellasest**, mille puhul kõigile tunnustele on meie poolt juba antud nende seisund (näit. **0**). Nii oli see kombeks veel kümme-neliteist aastat tagasi, kui ei tuntud välisrühma-meetodit. Selline lähenemine sisal-dab suurt subjektiivse vea ohtu ja seda välditakse. Mõeldav on küll anda hüpoteetilise ühiseellase selline iseloomustus, milles osa tunnuste puhul on antud põhjendatult eeldatavseisund, teiste puhul aga ainult küsimärk.

Enamuse teiste parsimooniameetodite puhul ehitab programm alguses **juu-reta puu** ("kerakujulise" kladogrammi, nn. Wagneri võrgu - *Wagner network*), ja alles lõpuks juuritakse see meie poolt näidatud välisrühma ja siserühma vahelt. Kui me kuidagi ei oska välisrühma analüüsi jaoks eristada, võib puu lõpuks juurida keskpunktist; sellest mistahes taksonini (harutipuni) on sama palju muutussamme.

8.10. Juurega puu puhul võime PAUPi kasutades anda käsud, mis võrdselt lühimate puude saamisel paneb kladogrammide genereerimisel programmi eelis-tama tunnuste paralleelset teket reversioonile (DELTRAN), või tagasipöördu-mist (reversiooni) paralleeltekkele (ALLTRAN).

8.11. Arvutiprogrammile võib "ette kirjutada" ka mitmesuguseid piiran-guid ja "sundkäike". Võime nõuda, et teatud liigid või liigirühmad **peavad** olema ühes harus (klaadis) koos; et teatud rühmad peavad olema kõrvutistes sõsarokstes (harudes). Sellised "sundused" (*constraints*) on tavaliselt õi-

gustatud ainult siis, kui need tuginevad mõnel eelnenud kladistilisel analüüsil uuritava liigikogumi osahulga kohta.

8.12. Puu ehitamise mistahes algoritm (programm) püüab minimeerida puu pikkust, s.t. vältida **homoplaasiat** - paralleelsust ja tagasipöördumist. Iga saadud puu puhul arvutab programm iga tunnuse kohta selle **püsivusindeksi** (*consistency index*); kui tunnus on muutunud ainult üks kord, on selle väärtus 1. Mida rohkem on reversioone ja paralleelsust, seda väiksem on indeksi väärtus; seda pikem sammudes tuleb ka puu. Kõigi tunnuste peale kokku ise-loomustatakse kogu puud selle püsivusindeksiga.

Üldise seaduspärasusena sõltub püsivusindeks taksonite ja (vähem) tunnuste arvust, ja väheneb koos nende suurenemisega.

8.13. Kui meil on head tunnused (s.t. sellised, mille püsivusindeks on 1), saame ühe või väikese arvu kladogramme. "Halbade" tunnuste puhul saame neid sageli mitu või isegi palju. Programmil PAUP 2.4 on maksimaalseks puude arvuks, mida programm suudab oma mällu jätta, 100; PAUP 3.1 puhul on see üle 32 000. Sellise kladogrammide hulga puhul tekib uurijal kiusatus mõned "halvad" tunnused välja jätta või tunnuseid **kaaluda**: anda parematele tunnustele suurem kaal, suurem sisuline väärtus.

Igasugune tunnuste kaalumine on subjektiivse hinnangu andmine. Sisuliselt võiks see ju igati põhjendatud olla: kõik tunnused ilmselt pole süstemaatika jaoks samaväärsed. Puudub aga objektiivne kriteerium, mõõdupuu, mille abil "õiglaselt" kaaluda.

Kladistikas on kasutusel Farrise poolt pakutud kaalumisviis. Vastavalt esimesele kladistilisele analüüsile leitakse iga tunnuse püsivusindeks; tunnusele antakse kaal, mis vastab selle väärtusele. Nüüd teostatakse teine analüüs ja uus kaalumine, kolmas, vajadusel neljaski taasanalüüs, kuni püsivusindeksi väärtus (ja vastavalt ka saadud puude pikkus) jääb püsima. Tavaliselt väheneb sellise ülekaalumise tagajärjel leitud ühepikkuste puude arv umbes kümme korda või enamgi. Paraku on saadud "kaalutud" puud mõnigi kord algsetest puudest märgatavalt pikemad. (Selle nägemiseks tuleb kasutada programmis erivõtteid!) Ka võib tunnuste kaalumiste järel saadud kladogrammides mõni takson "nihkuda" täiesti ootamatusse ja ilmselt ebaõigesse paika.

8.14. Suure andmemaatriksi (palju liike, palju tunnuseid) ja tunnuste madala püsivusindeksi puhul võime kladistilise analüüsiga saada mitu või terve hulga sama pikki puid, s.t. võrdväärseid hüpoteese. Enne nende vahel valiku tegemist (või tegemata jätmist) on otstarbekas programmi abil teha üksikkladogramme võrdlev ja üldistav **konsensuspuu**.

8.14.1. Ranges konsensuspuus (*strict consensus tree*) on näidatud kõik need klaadid (harud, harunemiskohad), mis on ühised **kõigile** saadud üksikkladogrammidele. Sageli juhtub, et erinevused eri kladogrammide vahel puudutavad ainult üksikuid selle osi, ainult detaile. Need konsensuspuu osad, kus eri kladogrammides on fülogenees kujutatud erinevalt, on **polütoomse** harunemisega, s.t. lahendamata.

Konsensuspuu **pole** mingi fülogeneesi hüpotees, selle järgi ei tohi kujutada võimalikku, hüpoteetilist evolutsiooni - välja arvatud neil haruldastel juhtudel, kui mõni üksik kladogramm on täpselt sama ehitusega (topoloogia-

giaga). Viimasel juhul leiab see üks kladogramm konsensuspuu poolt kinnitust (mitte vastupidi). Konsensuspuu näitab ainult uuritava taksonite hulga vastastikuste suhete **topoloogiat**; sama topoloogia võib olla saavutatud mitmete erinevate tunnuste muutumise skeemidega. Muuseas, kui käsitleme konsensuspuud kladogrammiga ja mõõdame ta pikkuse, osutub ta tavaliselt aluseks olnud kladogrammidest pikemaks; niisiis on ta säästuprintsiibist lähtudes nendest "halvem".

Ühte ja väga olulist näitab range konsensuspuu ometi: kõik konsensuspuu harud (oksad) ühendavad taksoneid (liike), mis kõigi kladogrammide põhjal moodustavad monofüleetilise rühma. Niisiis on konsensuspuu hea abinõu monofüleetiliste rühmade väljaselgitamisel.

8.14.2. Lisaks rangele konsensuspuule kasutatakse ka teisi. Kujutleme, et paljudes kladogrammides on mingi klaad (haru, seega oletatavalt monofüleetilise rühma sisese fülogeneesi hüpotees) sama topoloogiaga, ainult ühes kladogrammis tuhandest on see detail erinevana kujutatud. Ranges konsensuspuus loetakse see klaad lahendamata alljaotuseks. Ilmselt oleks õigem arvestada ka kladogrammidevaheliste erinevuste esinemissagedust.

PAUP pakub võimalust konstrueerida 50 %-enamusreegli konsensuspuud. Selles näidatakse neid klaade (harusid), mis esinevad vähemalt 50 % puudes, ja iga klaadiharu juures on märgitud esinemissageduse protsent. Programmile võib anda ka käsu 50 % asemel piiriks võtta 95 %, mis tundub mõistlikult range piiranguna.

8.15. Kladiistilise analüüsi tulemusena saame parimal juhul üheainsa kladogrammi, enamasti aga mitu võrdse pikkusega puud ja konsensuspuu(d). Paljud süstemaatikud on sel juhul piirdunud konsensuspuu kasutamisega. On ju klassifikatsiooni jaoks oluline saada ülevaade monofüleetiliste rühmade ja allrühmade olemasolust, ja ainuüksi seegi tulemus õigustab analüüsi teostamist.

Teistel juhtudel soovime ikkagi esitada taksonite tõenäolise evolutsiooni skeemi. Selleks tuleb läbi uurida **kõik** saadud kladogrammid ja jälgida iga tunnuse muutumist nendes (kas uurides arvuti poolt koostatud muutuste tabelit või kasutades programmi MacClade). Lähtudes kõigist varasematest teadmistest antud organismirühma kohta ja püüdes rakendada loogilist mõtlemist võime välja praakida need puud (kladogrammid), mis on selgelt vastuolus seniste teadmistega. Näiteks saab vähetõenäolisteks pidada arengupuid, milles mingi komplitseeritud struktuuriga organ tekib paralleelselt mitmes või isegi paljudes klaadides.

8.16. Leides või valides mingi juuritud kladogrammi (puu, fülogeneesihüpoteesi), ei tohi piirduda ainult puu **topoloogia**, selle struktuuri teadmiseks võtmisega. Iga kladogrammi **sõlmevahe** (kahe harunemispunkti või harunemispunkti ja *terminaalse taksoni*, s.t. harutipu vahe, mida kujutab sirgjoon) tähendab teatud tunnuste teatud muutusi, transformatsioone. Me ütleme, et need muutused **toetavad** (*support*) muutuskohast ülalpool olevat klaadi ehk haru. Need tunnuste muutused kantakse kladogrammidele (enamasti) nii: sõlmevahet märkivale joonele tõmmatakse lühike ristkriips ja selle juurde märgitakse muutunud tunnuse number. Kui tegemist on sama tunnuse paralleelse tekkega mitmes kladogrammi harus, on tunnuseks kaks kõrvuti-

olevat kriipsukest (võrdusmärk üle sõlmevahejoone). Reversiooni (tagasi-pöördumine plesiomorfse tunnuseni) puhul kasutatakse joonele paigutatud ristikest. (Vt. joon. 11.)

8.17. Uurides iga klaadi toetavaid tunnuseid, võime mõnikord näha, et klaad on tugevasti toetatud mitme "olulise" tunnusega. Teistel juhtudel võime üllatusega leida, et suurt klaadi (näiteks uuritava liigikogumiku kaheks jaotumisel) toetab ainult üks tunnus, seegi mõnikord suhteliselt väheveenev (näiteks mingi organi mõõtme suurenemine). Paljude kladogrammide võrdleval uurimisel neist "kahtlaste" väljapraakimiseks on selline *klaadi-toetuste* erinevus meile oluline.

8.18. Kladistikaprogrammid võimaldavad kladide toetuse kvantitatiivset hindamist. Laialt kasutatakse nn. **bootstrap-meetodit** (mõistlikku eestikeelset nime sellel veel pole). Teostame sama andmemaatriksit kasutades (näiteks) sada kladistilist analüüsi, jättes iga kord välja ühe juhuslikult võetud tunnuse. Saadud kladogrammide konsensuspuu moodustades võime näha, mitmes protsendis kladogrammidest on selle või teise klaadi olemasolu toetatud. Paraku on selline analüüs väga aeganõudev (kestab mõnikord nädalaid).

8.19. Kladistilise analüüsi edukuse peamiseks takistajaks on vasturääkivused erinevate tunnuste muutumises. Tunnuste muutumise suund tundub olevat vastupidine; nad ei toeta üksteist, vaid "segavad": üksiktunnuste muutumise puud (kõiki taksoneid arvestades) ei anna ennast kokku panna üheks. Selle tulemusena tekib mitmeid võrdse pikkusega lühimaid puid; selle tulemusena on tunnustel madal püsivusindeks (s.t. esineb rohkelt paralleelismi ja reversiooni).

Nendest vasturääkivustest on püütud lahti saada **klikimeetodi** (*clique method*) abil (G.F. Estabrook jt.). Selles kasutatakse ainult neid tunnuseid, mis moodustavad *kliki*, mis on oma muutumiste poolest omavahel kooskõlas; teised jäetakse analüüsis kasutamata. Sellisel viisil saadakse analüüsi tulemusena küll ainult üks kladogramm, kuid suur hulk informatsiooni jääb kasutamata, ja iga väljajätmise põhjenduseks on ju *ad hoc* hüpotees, mis on vastuolus säästuprintsiibiga. Et meetod on liiga jäik, on analüüsi tehtud mitmes osas. Esialgse analüüsil saadud klaade on edasi analüüsitud ükshaaval, alustades jälle kõigi tunnuste hõlmamisega ja neist *antud klaadile* kohase kliki väljaselgitamisega.

Meetodit arendas J. Felsenstein (1978), kasutades **lävemeetodit** (*threshold*): mistahes tunnusel võib sellises analüüsis olla teatud, üldisena kõigi tunnuste jaoks ette antud arv homoplaasiat (paralleelsust ja reversiooni); alles selle ületamisel jäetakse tunnus kasutamisest välja.

Klikimeetodid mitmetes oma teisendustes on rakendatavad juba mainitud programmiga PHYLIP; nagu märgitud, on need kasutatavad eelkõige väikese taksonite arvu puhul, ja analüüs on aeganõudev.

8.20. Nagu kõigi meetodite puhul, nii ka kladistikas tuleb nende kasutamisel arvestada (hinnata) meetodi **robustsust**. Robustne meetod on selline, mis on vähetundlik võimalike üksikute vigade, häirete, andmestiku mittekülaldase hulga või ebapiisava täpsuse suhtes.

Kladistilise analüüsi tulemus sõltub tunnuste arvust; liiga väike arv

annab küll vähem homoplaasiat, kuid tulemus ei tarvitse olla küllalt usaldatav. Üldreeglina on soovitatud kasutada vähemalt sama palju tunnuseid, kui on taksoneid. Sellel soovitusel on aga täiesti empiiriline iseloom, selle toetuseks pole mingeid teoreetilisi põhjendusi. On edukalt teostatud analüüse, kus taksonite arv on tunnuste omast 2 - 3 korda suurem.

Analüüsi tulemus sõltub hõlmatud taksonite arvust. Uurides ühte monofüleetilist rühma tuleks analüüsi võtta kõik teadaolevad liigid. Kui jagaksime nende kogumi pooleks ja teostaksime kaks analüüsi, saame kaks mõneti erinevat kladogrammi. Kasutades näit. ainult Eestis esinevaid liike, saame erineva kladogrammi, kui kogu maakera liike hõlmates. Püüdes analüüsida mingi sugukonna evolutsiooni ja võttes selleks igast perekonnast ühe liigi, sõltub tulemus mõnevõrra sellest, millise liigi just valime. Paraku on praktilises töös piiravaiks tegureiks ebapiisav info paljude taksonite kohta, mida sooviksime analüüsida, ja arvutite (ning programmide) võimsus. Maksimaalseks taksonite arvuks, mille puhul analüüs enam-vähem korralikult välja tuleb, on praegu (1994) umbes 100.

Paraku pole ka teised teadaolevad meetodid robustsemad. Kõnelemata metodoloogilisest nõrkusest, on feneetilised meetodid sama tundlikud, väherobustsed. Veel enam, fenogramm on veelgi tundlikum tunnuste arvu suhtes; täiendavate tunnuste lisamine analüüsile võib selle tulemusi ühel juhul parandada, teisel puhul aga moonutada.

8.21. Fülogeneetilises süstemaatikas on kladistilise analüüsi tulemuseks (hüpoteetiline) fülogeneesipuu, mille harud dihhotoomselt harunevad. Kuidas sellist skeemi **üle viia hierarhiliseks klassifikatsiooniks**, mis meie kõigi juures kujunenud veendumuste kohaselt peaks olema kooskõlas Linné poolt algatatud jaotusega (liigid - perekonnad - sugukonnad - seltsid - klassid, jne.)?

Kladistilise metodoloogia kohaselt annab iga fülogeneetilise puu harunemiskoht kaks sōsarkklaadi - kaks monofüleetilist rühmitust, mis peavad järelikult kuuluma samasse süstemaatika kategooriasse (mõlemad on sugukonnad, alamsugukonnad, triibused, perekonnad või midagi muud). Üksteise järel paigutatunud harunemispunkte on aga nii palju, et linneeliku hierarhia kategooriaid lihtsalt ei jätku. Algul püüti sellest üle saada kategoorianimetustest hoopiski loobumisega, asendades need numbritega. See süsteem ei saanudki aga pooldamist leida. Praegu kasutatakse teist, linneeliku süsteemiga kompromissi taotlevat võimalust. Kujutleme ühte fülogeneesipuu haru (klaadi), mida peame vastavaks ühele seni sugukonnaks peetud taksonile. Selle klaadi kammipiidena üksteise järel asuvad külgharud võime kõik tinglikult nimetada alamsugukondadeks, loetledes neid klaadi alusest tipu poole minevas järjestuses (vt. joon. 4; sellel kujutatud sugukonna alamsugukonnad **A - E** tuleks loendades esitada nii:

```
Sugukond X
  Alamsugukond A
  Alamsugukond B
  ...
  Alamsugukond E.
```

Raskused fülogeneetilistel puudel kujutatatu üleviimisel tavapärasesse klassifikatsiooni ei tulene kladistilise meetodi puudustest, vaid meil

kasutusel olevast, "kivistunud" linneelikust mõtteviisist, õigemini selle mittevastavusest eluslooduse kujunemise mehhanismidele. Väheste kategooriate arvuga nimetamismoodus on abinõu looduse mitmekesisuses orienteerumiseks, mitte aga looduses esinev reaalsus. Seda näitab ka kategooriate arvu kasv koos ikka uute liikide kirjeldamisega viimase 250 aasta jooksul; Linné sai taimede juures hakkama 4 - 5 kategooriaga. Piirav on ka inimese mälu ja mõtteviisi iseärasus: me ei suuda hästi orienteeruda hierarhilises klassifikatsioonis, kus kategooriaid tuleb meeles pidada üle 7-8.

8.22. Käesolevas kursuses ei käsitleta kladistilist analüüsi **molekulaarsüsteematiliste tunnuste** kasutamisel. Põhilised printsiibid on muidugi samad. Tuleb aga mainida raskusi ja hüvesid sama uuritava liigikogumi juures "morfoloogilisi" ja molekulaarseid tunnuseid kasutades saadud kladogrammide ja fülogeneetiliste puude võrdlemisel.

Katsed ühendada ühte andmematriksisse mõlemat tüüpi tunnused on seotud ületamatute meetodiliste raskustega ja pole ennast õigustanud, samuti mitte nende alusel konsensuspuu koostamine. See nõuaks ühete või teiste tunnuste kaalumist, millele aga objektiivset alust pole. Ilmselt on kõige perspektiivikam kasutada eri andmestiku alusel saadud fülogeneesipuid nende kui evolutsioonihüpoteeside vastastikuseks kinnitamiseks / falsifitseerimiseks. Senised kogemused on näidanud, et paljudel juhtudel on molekulaarse ja "mittemolekulaarse" analüüsi tulemused omavahel üldjoontes kooskõlas - kuid mitte alati. - Molekulaarseid tunnuseid kasutatav süstemaatika areneb tänapäeval äärmiselt kiiresti ja on muutumas lähiaastail põhimiseks.

8.23. Kladistiline analüüs võib sageli anda mitu või palju võrdlühikesi puid, kusjuures neid uurides ei leia me põhjendust neist ühegi kõrvalejätmiseks. See tähendab, et meil (praegu) olemasolev **andmestik ei võimaldagi ainsa, ainuõige fülogeneesihüpoteesi eelistamist.**

Fülogeneetiline süstemaatika taotleb klassifitseerimiseks fülogeneesi rekonstrueerimist alati ebapiisavate andmete alusel. Nagu nägime, on meetodiliselt piiravaks teguriks asjaolu, et paljud esinevad liigid on veel leidmata ja kirjeldamata; arvata äle 95 % kui mitte enam on välja surnud, jätmata meile fossiile. Süsteemi loomine isegi parimate meetodite alusel on aproksimatsioon, lähendus; meie võimuses on ainult suurendada selle tõenäolist vastavust reaalsele eluslooduse ajaloole. Olulisim on seda teha elusolendite süsteemi suurjaotuste osas. Konspekti autori arvates on seetõttu lubatav süsteemi detailide puhul hädavajaduse sunnil **ajutiselt** kasutada ka feneetilisi meetodeid või isegi kunstlikku süsteemi, kui see on süstemaatika poolt selgelt koos põhjustega esile toodud.

9. Kladistilise metodoloogia seisukohti tunnustades tuleb meil ümber hinnata mitmeid justkui endastmõistetavaks peetud termineid ja tava"tõdesid".

Süstemaatikute kõnekeeles, aga ka õpikuis on sageli räägitud sellest, nagu oleks üks **organismirühm kujunenud, põlvnenud mõnest teisest**, praegu veel eksisteerivast. Vastavalt kladistikale (ja ka meie teadmistele makroevolutsioonist) saavad liigist kõrgemad taksonid tekkida ainult liigitekke teel. Edasise divergentsi teel võib lähteliigist kujuneda perekond või suu-

remgi takson. Algesivanema tekkel kujunes (piltlikult öeldes) omaette klaad, sellega koos - teine klaad, kujunenud taksoni sõsarrühm. Mõlemad sõsarrühmad on sama vanusega, sama tekkeajaga, ja mõlemad on monofüleetilised rühmad.

On võimalik, et lähedased (näit. samasse perekonda kuuluvad) liigid võivad oma genotüübi sarnasuse tõttu olla kalduvad samasuguste mutantide tekkele. Seda ei saa aga tõlgendada ühest **perekonnast** teise perekonna tekitamisenä.

Pole olemas primitiivseid, ürgseid praegu elavaid **liike ega perekondi**. Esiteks on *primitiivsuse* mõiste antud juhul ebasobiv: liigid, mis pole küllaldasel määral sobivad, kohastunud, surevad kiiresti välja. "Primitiivne" ehitus on selline ehk ainult inimese kummalisi mõõdupuid kasutades; tegelikult on see üks võimalikest ehituse või struktuuri tüüpidest, mis ilmselt on enam või vähem täiuslikult vastav organismi ökoloogilisele nishile ja ajalooliselt kujunenud elustrateegiale. Teiseks saavad ürgsed, vanad oma päritolult olla ikkagi vaid **plesiomorfsete tunnused**, mitte aga neid "kandvad" taksonid.

Süstemaatikud on kõnelenud **heterobatmiast**: igal taksonil (näit. liigil) esinevat kõrvuti koos "kõrgeltarenenud", suhteliselt uudsete tunnustega ka primitiivseid, mis näitavat tema sugulusele mingi teise organismiderühmaga. Fülogeneetilise süstemaatika keelt kasutades tähendab heterobatia nii plesiomorfsete kui ka apomorfsete tunnuste samaaegset esinemist antud organismil. See on täiesti normaalne ja tavaline nähtus, mis ei vaja omaette terminit. "Vana kooli" süstemaatikud osutavad suurt tähelepanu plesiomorfsetele tunnustele, need näitavat uuritava taksoni asendit süsteemis. Vastavalt fülogeneetilise süstemaatika seisukohtadele näitavad ainult apomorfsete tunnused evolutsiooni toimumist, plesiomorfsete tunnused on selleks kõlbmatud.

10. Kladiстика ühe suhteliselt uue vooluna on mitmete süstemaatikute poolt arendatud nn. **transformeeritud kladiismi** ehk **mustrikladiismi** (*transformed cladistics, pattern cladistics*). Selle pooldajad on püüdnud luua *teooriavaba* kladiistikku metodoloogiat, milles eitatakse evolutsiooni olemasolu, või õigemini väidetakse evolutsiooniideede ebaolulisust kladiistolise analüüsi teostamisel. Ei eristata plesiomorfseid ega apomorfseid tunnuseid (tunnuse seisundeid), vaid lihtsalt tunnuse seisundeid (või tunnuseid). Transformeeritud kladiismi võib käsitada kui kladiismi ilam ajalise dimensioonita. Tegelik süstemaatikute juures see suund märgatavat poolehoidu võitnud pole.

11. Evolutsiooniline süstemaatika. Feneetiline ja fülogeneetiline süstemaatika taotlevad klassifitseerimist (süsteemi loomist) lähtudes kahest erinevast ja mitteühilduvast seisukohast: (üldisest) sarnasusest ja taksonite genealoogiast (põlvnemisest). Süstemaatikute üldiseks ideaaliks on ikka olnud ühe, kõikearvestava ja universaalse süsteemi loomine, mis talletaks endas informatsiooni nii sarnasuse kui ka suguluse kohta. Sageli tsiteeriti varem Gilmoure'i ja Waltersi poolt ideaalile viitavat väljendit

«üldotstarbeline klassifikatsioon» (*general-purpose classification*), mille aluseks temal oli küll feneetiline lähenemisviis. Sageli on viidatud ka tuntud evolutsionisti ja zooloogi-süstemaatiku Simpsoni kladistikaeesel ajal öeldud aforismile: süsteem ei saa kujutada fülogeneesi, küll aga arvestada seda.

Teoreetiliselt oleks võimalik luua evolutsioonilise süstemaatika nõuetele vastav süsteem, mis sisaldab informatsiooni summeerituna nii sarnasuse kui ka fülogeneesi kohta. Klassifikatsiooni on nimetatud informatsiooni talletamise ja kättesaamise süsteemiks, sellega tuleb nõustuda. Igasugune informatsiooni talletamine on aga mõttekas ainult siis, kui seda on võimalik ka taas kätte saada. Mingist "summaarse informatsiooniga" süsteemist välja lugeda eraldi infot sarnasuste ja fülogeneesi kohta on nii põhimõtteliselt kui ka praktiliselt võimatu.

Evolutsioonilist süstemaatikat pole reaalselt arendatud seetõttu, et sellel **puuduvad oma meetodid**. Üksikud katsed mingeid algoritme luua pole viinud praktiliselt rakendatavate tulemusteni. Nii piirdubki evolutsiooniline süstemaatika siiani teoreetiliste arutlustega, praktikas aga kas feneetilise süsteemi juurde narratiivse (jutustava) ebakonkreetse evolutsiooniteemalise kommentaariumi lisamisega, või (sagedamini) parafüleetiliste rühmade loomisega. Viimane toimub fülogeneetilise süsteemi "parandamise" teel, eristades autapomorfsete tunnuste poolest teistest erinevaid taksonid. Kui joonistaksime sellise eristamise tulemuse "kladistika graafiliste vahenditega", võiksime näha näiteks klaadi (endist monofüleetilist perekonda), mille üheks terminaalseks taksoniks (harutipuks, niisiis eeldatavasti liigiks) on teine perekond või isegi sugukond.

Evolutsiooniline süstemaatika peab klassifitseerimisel, taksonite ühendamisel tähtsaks ka sümpleiomorfseid tunnuseid (vrd. joon. 12).

Lisaks metodoloogilistele vasturääkivustele ja eksimustele loogika vastu (sugukond mingi perekonna liikmena) on evolutsioonilise süstemaatika põhihädaks *subjektiivsus* - ainult uurija suva otsustab, kui suur erinevus on küllaldane, et mõnest monofüleetilisest rühmast (klaadist) üks või teine liige välja eraldada.

Evolutsioonilise süstemaatika pooldajate peamiseks teeneks on kladistiliste meetodite nõrkade külgede leidmine ja kritiseerimine, eriti fülogeneesipuudest hierarhilise klassifikatsiooni tuletamise osas.

12. Identifitseerimine. Süstemaatika üheks allosaks on identifitseerimine - üksikisendite mõnda juba teadaolevasse (kirjeldatud) liiki kuuluvuse kindlaksmääramine. Seda teostatakse võrdlemise teel, kusjuures meil peab (peaks) eelnevalt teada olema oletatavalt sarnaste liikide varieeruvus kõigi identifitseerimisel kasutatavate tunnuste osas.

Feneetilisi meetodeid kasutades on meil kahtlastel juhtudel võimalik kasutada diskriminantanalüüsi (mida siin ei käsitleta). Tavapraktikas kasutame nn. määramistabeleid ja arvutiprogrammide kasutamisega töötavat dialoogrezhiimis (*on-line*) identifitseerimist.

Bioloogide üheks ülesandeks on selliste määrajate koostamine. Määramistabelite koostamise idee on pea muutumatult pärit Lamarckilt, kes koostas

esimese (1778). [Esimene arvutil dialoogrezhiimis kasutatav valmis Borghey jt. poolt 1968; mustvalged joonised lisati sellistele 1985, värvilised illustratsioonid - 1988.]

Hea määramistabeli puhul, mis on dihhotoomse ülesehitusega (igas punktis on üks küsimus ehk tees ja antitees), on küsimuste arv üldjuhul $Q = T - 1$, kus T on taksonite arv. Parimad on need tunnused, mis on kõige informatiivsemad (selle määramist vt. punkt 4.4); need jagavad käsitleda jäänud liikide hulga kaheks võrdseks osaks. Ideaalses määramistabelis tuleb liigi nimeni jõudmiseks vastata küsimustele, mille arv $L \geq \log_2 T$. Arvutades mingi olemasoleva määramistabeli juures keskmise küsimuste arvu, millele tuleb vastata kõigi liikideni jõudmiseks, ja võrreldes seda ülalesitatud algoritmi järele leitud L väärtusega, saame hinnata tabeli koostamise edukust.

Määraja koostamisele asudes tuleb küsimusi kaaluda: selgeid, kindlamaid, kasutajale mugavamaid küsimusi tuleb kasutada tabeli algusosas.

Sellistele nõuetele vastavaid, Boole'i loogika operaatoreid kasutavaid määramistabeleid, milles pealegi on garanteeritud teesi ja antiteesi tegelik teineteise välistamine, on võimalik koostada kasutades programmi DELTA (seda teeme praktilistes töödes). Esimeseks tööks on seejuures tunnuste loetelu koostamine; iga liigi kodeeritult kirjeldamine tunnuste loetelu järgides, kusjuures on lubatav tunnuse kahe või mitme seisundi puhul kasutada kodeeritud sõnu *või; ja; kuni; tunnus puudub; tunnuse seisund on teadmata*. Tunnustele antakse subjektiivne kasutusheaduse hinnang (mida korduval katsetamisel on lihtne muuta); märgitakse, mitu korda sama (varieeruv) liik võib määramistabelis korduda. Programm genereerib trükivalmis määramistabeli, mida võib toimetamisega muuta, üksikuid osi uuesti teha, jne.

Sama andmebaasi abil saab genereerida dialoogrezhiimis määramistabeli, mille kasutamisel võib tunnuseid kasutada mistahes järjestuses, või programmi poolt soovitatud parimate tunnuste seast suvalise väljavahetamisega. Sellisele arvutimäärajale võib sisse skaneerida jooniseid või fotosid. Ikka suurema levikuga on maailmas (teiste programmide abil koostatud) CD ROM-idele (s.t. mahukaile diskettidele) kantud määramisvahendid.

13. Süstemaatiku töö korraldus. Siinkohal käsitletakse ainult neid süstemaatiku igapäevase töö külgi, mis on vähemtraditsioonilised ja ei olene kuigivõrd mingi kindla rühma käsitlemisest.

13.1. Välitöödel uue materjali kogumine on Eesti süstemaatikuile traditsiooniline, kuid mitte hädavajalik tegevus. Õigemini on see vajalik eelkõige selleks, et saavutada looduse kui terviku, liikide kui terviklike, kuid varieeruvate olendite ja koosluste kui liikidest erinevate reaalsuste olemasolu "tunnetamist". Et välitöödega sageli kaasneb ka teatud biogeograafilise andmestiku kogumine, peaks nende teostaja olema tuttav levikukaartide koostamise tänapäevaste meetoditega (*grid*-süsteemid. eriti UTM-võrgustik; levikukaartide koostamine arvutiprogrammide abil; oma asukoha = leiukoha koordinaatide määramine kaasaskantava suhteliselt odava ja kerge aparatuuri abil).

13.2. Kollektatsioonide kasutamine, asugu need kustahes, on süstemaatiku

töös endastmõistetav. Selleks peab omama neist mingi ülevaate (botaanikutel on vastavaks käsiraamatuks *P.K. Holmgren et al. (eds.), Index Herbariorum*; viimane, 8. trükk ilmus 1990. a.). Enamus kollektsioone saadab eksemplare laenuna tunnustatud teadusasutustele ajutiseks kasutamiseks. Nende kasutamisel tuleb piinliku täpsusega kinni pidada laenutamisel kehtestatud reegleist, mis tagavad eksemplaride säiluvuse ja õigeaegse tagastamise.

13.3. **Standarddeid** ja hea süstemaatiku poolt peaaegu sunduslikult kasutatavaid **käsiraamatuid** pole teabkuipalju. Paraku oskan siin tutvustada peamiselt vaid botaanilisi ja mükoloogilisi. Taksoneid nimetades, eriti aga uusi kirjeldades on kohustuslik jälgida nomenklatuurikoodeksites (vt. 14.) esitatud nõudeid. Taime- ja seenetaksonite autorinimede standardiseeritud lühendid (mille kasutamist nõuavad kõigi soliidsemate ajakirjade toimetused) on toodud teoses *R.K. Brummitt & C.E. Powell (ed.), Authors of plant names*. Kew, 1992. Botaanilise (aga rohkelt ka muu bioloogilise ja rakendusteaduste alase) perioodika standardlühendid on raamatus *G.D.R. Bridson (comp. & ed.), E.R. Smith (ed. ass.), B-P-H/S. Botanico-Periodicum-Huntianum/Supplementum*. Pittsburgh, 1991. Suurim botaaniline bibliograafia (autorite kaupa) koos teoste standardsete lühenditega on paljukõitelises teoses *F.A. Stafleu (& E.A. Mennega), Taxonomic literature*. Königstein. Selle 2. trükile on lisaks ilmunud 2 täiendusköidet, neist viimane 1993. Kasutusel olevad taimede, seente ja limaseente perekondade korrektsed nimed koos kommentaaridega on toodud teoses *W. Greuter et al., NCU-3. Names in current use for extant plant genera*. Königstein, 1993. Taimede geograafilise leviku regioonide, nende nimetuste ja nimelühendite standard on avaldatud broshüüris *S. Hollis & R.K. Brummitt, World geographical scheme for recording plant distributions*. Pittsburgh, 1992. Liigikirjelduste koostamise standardprogramiks on botaaniliste andmebaaside rahvusvahelise tööühma poolt kinnitatud juba mainitud *DELTA*.

13.4. **Andmebaasid** tuleb tavaliselt igal süstemaatikul endal vastavalt vajadusele luua. Personaalarvuteile sobivaimad süsteemid on praegu FoxPro ja (paremate arvutite puhul) ACCESS. Taksonoomiliste andmebaaside tegemisel tuleks kohe ette näha nende polüfunktsionaalsus: sama eksemplaride kohta koostatud andmebaas peaks võimaldama etikettide printimist, tunnuste kodeeritud või kodeerimata sisestamist, levikukaartide koostamist arvuti abil, jne.

13.5. **Internet** on praegu levinuim elektroonilise kommunikatsiooni süsteem, mis võimaldab pidada kiiret kirjavahetust teiste süstemaatikutega ja saada erialastesse listidesse registreerumisel jooksvat infot ning osaleda diskussioonides. Süstemaatikule võiks huvi pakkuda neist vähemalt kolm:

...

...

Kasutades *FTP* ja *GOPHER* -võimalusi saab oma arvutiise üle kopeerida mitmeid kasulikke programme (näit. ka *DELTA*'t) ja faile kasuliku infoga. Nende abil saab teha ka päringuid mitmetest üldvajalikest andmebaasidest. Esialgset-sissejuhatavat informatsiooni bioloogide võimaluste kohta saab, kui *FTP*-d kasutades ...

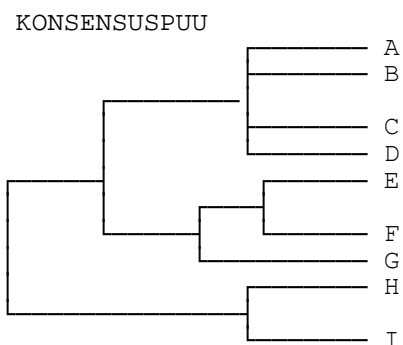
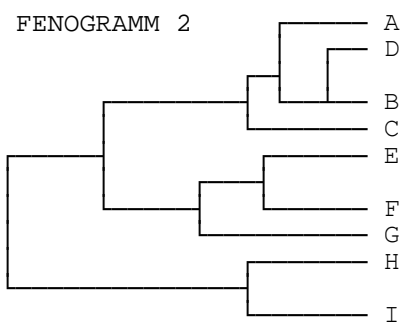
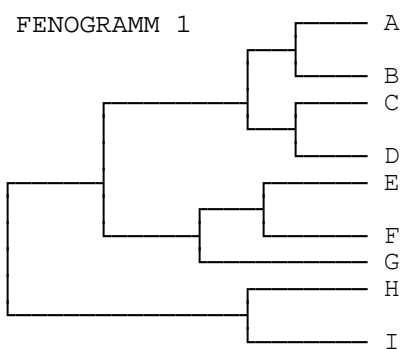
Joon. 1. Algandmete tabeli näidis. Tulpades on tunnuste (A - B) väärtused

LIIK / TUNNUS	A	B
Liik1	1.0	1.0
Liik2	1.0	2.0
Liik3	6.0	3.0
Liik4	8.0	2.0
Liik5	8.0	0.0

Joon. 2. Liikidevaheliste erinevuste maatriks (Eukleidiline kaugus)

LIIK	Liik1	Liik2	Liik3	Liik4	Liik5
Liik1	0.0				
Liik2	1.0	0.0			
Liik3	5.39	5.10	0		
Liik4	7.07	7.00	2.24	0	
Liik5	7.07	7.28	3.61	2.0	0

Joonis 3. Kaks fenogrammi ja nende konsensuspuu.



Liikide A; B; C; D klaster on esimeses ja teises puus erineva topoloogiaga; konsensuspuus pole seetõttu näidata mingeid ühisjooni, ja see klaster on lahendamata (*unresolved*).

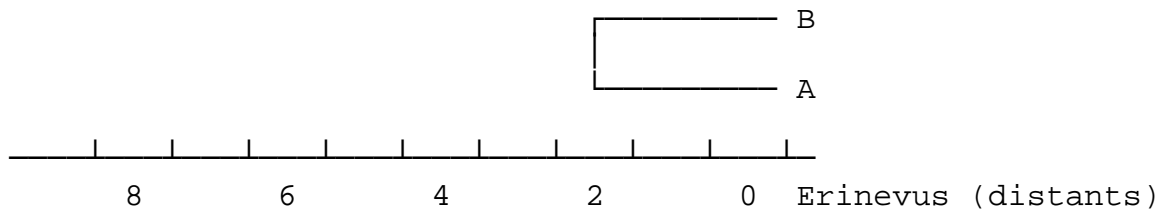
Joonis 7. Mono-, para- ja polüfüleetilised rühmad.

**Näide klasterdamise kohta: lähimnaabri- (üksiksideme-) klaster
dus**

Andmemaatriks (erinevus OTU-de vahel, standardiseerimata):

OTU	A	B	C	D	E
A	0				
B	2	0			
C	6	5	0		
D	10	9	4	0	
E	9	8	5	3	0

Kõige sarnasemad (vähemerinevad) üksteisele on A ja B. Ühendame need:



Nüüd arvutame selle 2-liikmelise klasteri erinevused ülejäänud OTU-dest:

$$d_{(AB)C} = \min \{d_{AC}, d_{BC}\} = d_{BC} = 5$$

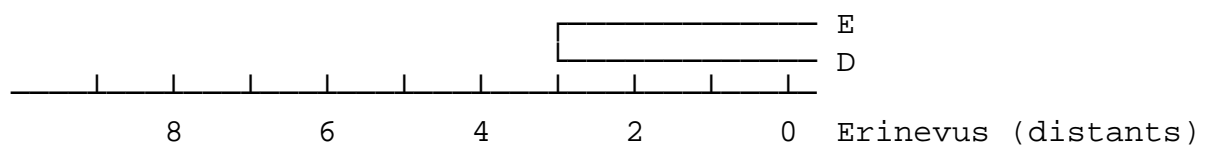
$$d_{(AB)D} = \min \{d_{AD}, d_{BD}\} = d_{BD} = 9$$

$$d_{(AB)E} = \min \{d_{AE}, d_{BE}\} = d_{BE} = 8$$

Teeme uue tabeli:

OTU	(AB)	C	D	E
(AB)	0			
C	5	0		
D	9	4	0	
E	8	5	3	0

Järgmine väikseima erinevusega paar on D ja E (distant on 3):



Millised on nüüd erinevused kahe klasteri vahel ning nende ja ülejäänud OTU C vahel?

$d_{(AB)C}$ juba on arvutatud: 5

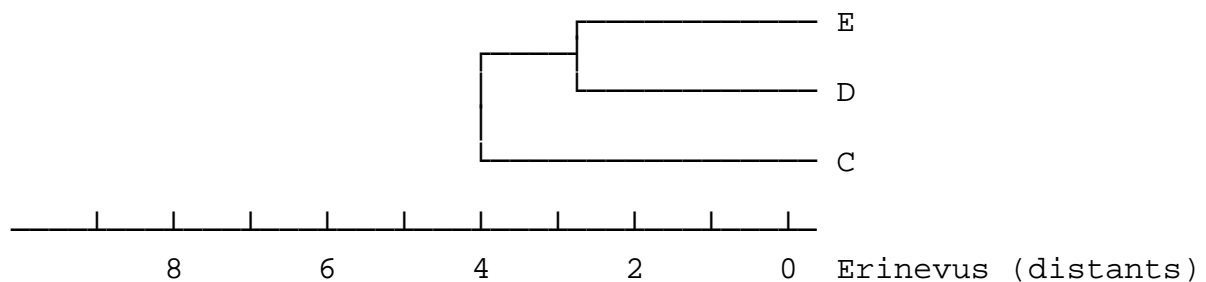
$$d_{(AB)(DE)} = \min \{d_{AD}, d_{AE}, d_{BD}, d_{BE}\} = 8$$

$$d_{(DE)C} = \min \{d_{CD}, d_{CE}\} = 4$$

Paneme need distantssid (erinevused) uude maatriksisse:

OTU	(AB)	C	(DE)
(AB)	0		
C	5	0	
(DE)	8	4	0

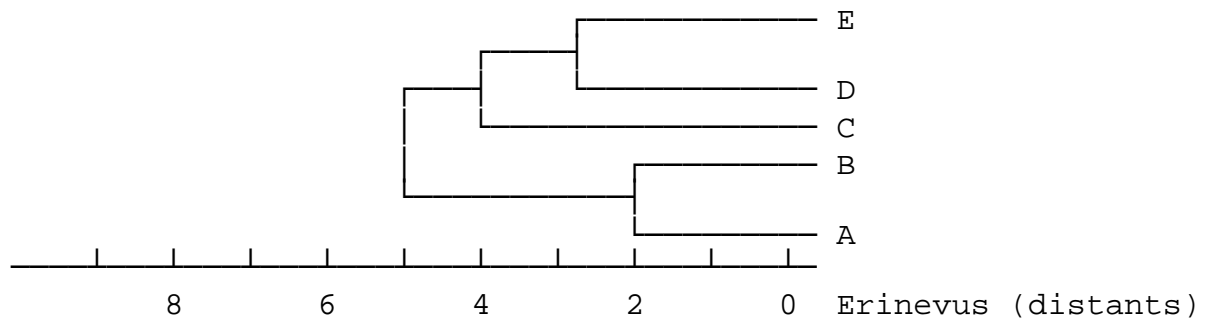
Kõige väiksem vahe (suurim sarnasus) on klastri (DE) ja C vahel, nimelt 4:



Jääb liita klastrid (AB), (CDE). Nende minimaalne distantis:

$$d_{(AB)(CDE)} = \min \{d_{AC}, d_{AD}, d_{AE}, d_{EC}, d_{BD}, d_{BE}\} = 5$$

Järelikult tuleb klastrid ühendada erinevuse tasemel 5:



See on üks lihtsamaid klasterdusmeetodeid. Neid on mitmeid.

Näiteks täissideme-klasterdusel leiame mitte minimaalse, vaid maksimaalse erinevuse iga klastri vahel. Näiteks (meie andmemaatriksi puhul):

$$d_{(AB)C} = \max \{d_{AC}, d_{BC}\} = 6 \quad \text{jne.}$$

Saadav dendrogramm on eelmisest erinev.

Üsna häid tulemusi saab grupikeskset klasterdust (UPGMA) kasutades:

$$d_{(AB)C} = 1/2 (d_{AC} + d_{BC}) = 5,5 \quad \text{jne.}$$

