Gradogramme - 'n nuwe hulpmiddel vir kladistiese taksonomie

I.C. Burger
Departement Wiskunde, Randse Afrikaanse Universiteit, Johannesburg, 2000

J. Heidema
Departement Wiskunde, Universiteit van Suid-Afrika, Pretoria,0001

W.A. Labuschagne en T.A. Meyer*
Departement Rekenaarwetenskap, Universiteit van Suid-Afrika, Pretoria, 0001

B-E. van Wyk
Departement Plantkunde, Randse Afrikaanse Universiteit, Johannesburg, 2000

Ontvang 6 Junie 1994; aanvaar 24 Augustus 1994

UITTREKSEL

Kladistiese taksonome klassifiseer organismes met behulp van kladogramme. 'n Metode word voorgestel wat daarop gemik is om aan die kladistiese taksonoom 'n hulpmiddel te verskaf om die finale mees bevestigde kladogram op te stel. Die metode is gebaseer op die konstruksie van saamgestelde voorrangrelasies.

ABSTRACT

Gradograms - a new aid to cladistic taxonomy

Cladistic taxonomists classify organisms with the aid of cladograms. A method is proposed that will serve as an aid to the cladistic taxonomist in drawing up the final most confirmed cladogram. The method is based on the construction of composite preference relations.

WAT IS KLADISTIESE TAKSONOMIE?

Die doelwit van kladistiek is om groepe organismes (wat *klades* of *taksons* genoem word) te klassifiseer op grond van hulle mees onlangse gemeenskaplike voorouers. Alle voëls het byvoorbeeld 'n gemeenskaplike voorouer met die volgende twee eienskappe: dit het nie 'n afstammeling wat 'n meer onlangse voorouer van alle voëls is nie; dit het geen afstammeling wat nie 'n voël is nie. Enige twee voëlsoorte sal dus nader aan mekaar geklassifiseer word as aan enige organisme wat nie 'n voël is nie. Die Duitse entomoloog, Willi Hennig, is die vader van kladistiek en die beginsel soos deur hom geformuleer is in 1966 gepubliseer.[1]

'n Kladistiese taksonoom stel klassifikasiesisteme vir taksons voor deur binêre bome wat kladogramme genoem word. 'n Kladogram word verkry met behulp van 'n aantal relevante eienskappe (genoem *kenmerke*) van die taksons wat geklassifiseer word. In tabel 1 stel die rye die taksons voor en die kolomme die kenmerke. Indien die taksons 'n groep plante is, kan kenmerk 1 byvoorbeeld *blaarbreedte* voorstel, waar daar tipies onderskei kan word tussen plante met breë en smal blare.

Elke kenmerkstaat ("character state") word voorgestel deur 'n unieke nienegatiewe heelgetal, na aanleiding van hoe primitief of gevorderd dit is. Hoe verder die waarde van 0 af is, hoe meer gevorderd is die kenmerkstaat. Veronderstel dat vir kenmerk 1 in tabel 1 die kenmerkstaat, breë blare, geassosieer word met 'n 0 en die kenmerkstaat, smal blare, met 'n 1. Dié toekenning van kenmerkstate sal dan geïnterpreteer

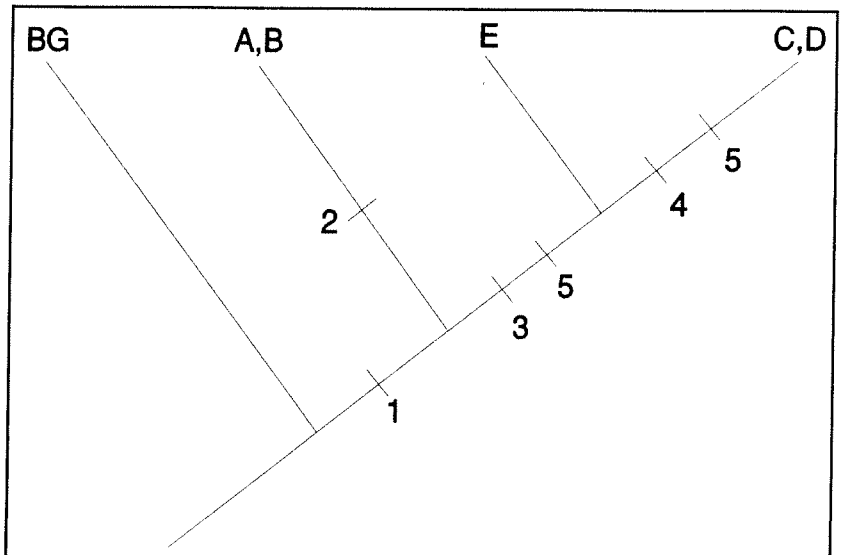
Tabel 1 'n Eenvoudige voorbeeld van 'n datamatriks met ses taksons voorgestel deur ses kenmerke

	Kenmerke					
Taksons	1	2	3	4	5	6
BG	0	0	0	0	0	0
A	1	1	0	0	0	1
B	1	1	0	0	0	0
C	1	0	1	1	2	0
D	1	0	1	1	2	1
E	1	0	1	0	1	1

word as dat 'n gemeenskaplike voorouer van al die taksons op 'n stadium breë blare gehad het, maar dat 'n meer onlangse gemeenskaplike voorouer van taksons A tot E op 'n stadium 'n smaller blaar ontwikkel het.

Die onderliggende idee tydens die opstel van 'n kladogram is dat taksons met *gedeelde ontwikkelde (d.w.s. meer gevorderde) kenmerkstate* saamgegroepeer moet word. Hierdie groeperings word gebruik om 'n binêre boom te verkry. By elke vertakking van die boom word al die taksons met 'n gedeelde ontwikkelde kenmerkstaat vir 'n spesifieke kenmerk geassosieer met een van die twee takke. Hierdie kenmerk, wat 'n

* Outeur aan wie korrespondensie gerig kan word.



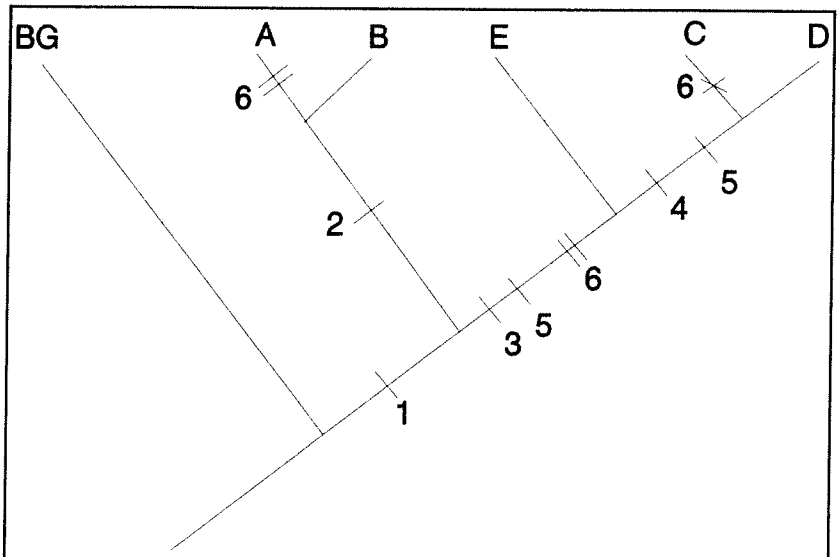
FIGUUR 1: 'n Spaarsaamste kladogram vir die ses taksons in tabel 1, opgestel uit die eerste vyf kenmerke in die tabel.

apomorfie genoem word, word dan op die tak aangedui. Die kladogram in figuur 1 is opgestel vanaf die groeperings geassosieer met die eerste vyf kenmerke in tabel 1.

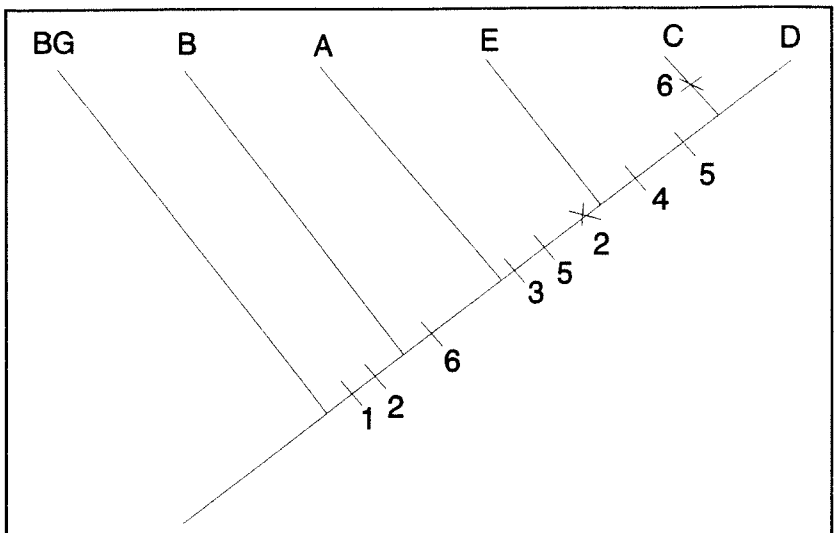
Alle taksons onder bespreking is blare van die binêre boom. Enige interne nodus *n* van 'n kladogram verteenwoordig die hipotetiese voorouer van al die nodi wat spruit uit *n*. Elke aanduiding van 'n kenmerk in die kladogram stel 'n kenmerkstaatverandering voor vanaf 'n primitiewe na 'n gevorderde kenmerkstaat. Daar is twee kenmerkstaatveranderinge vir kenmerk 5 in die kladogram in figuur 1 omdat kenmerk 5 drie kenmerkstate het. Hierdie kladogram het ses kenmerkstaatveranderinge in totaal.

Die eerste vyf kenmerke in tabel 1 het dit moontlik gemaak dat 'n suiwer binêre boom op grond daarvan opgestel kan word, maar dit is nie altyd die geval nie. Kenmerk 6 suggereer byvoorbeeld dat taksons A, D en E saamgegroepeer moet word, maar dit is in konflik met van die groeperings gebaseer op die eerste vyf kenmerke. Sulke konflikte, wat *homoplasie* genoem word, kan verduidelik word deur die *terugskakeling* en *konvergensie* van kenmerkstate. Terugskakeling verwys na 'n kenmerkstaatverandering vanaf 'n gevorderde kenmerkstaat na 'n meer primitiewe een, terwyl konvergensie verwys na die onafhanklike ontwikkeling van dieselfde kenmerkstaat deur 'n aantal taksons.

Een manier om kenmerk 6 in 'n kladogram te inkorporeer, word aangedui in die kladogram in figuur 2. 'n Terugskakeling word aangedui deur 'n "x" en 'n konvergensie deur "=". Hierdie kladogram het 'n totaal van nege kenmerkstaatveranderinge. Daar is ander maniere om kenmerk 6 te inkorporeer, waarvan sommige lei tot ander kladogramme, maar geen ander kladogram sal minder kenmerkstaatveranderinge tot gevolg hê nie. Kladistiese taksonome is geïnteresseerd in kladogramme met 'n minimum aantal kenmerkstaatveranderinge, of *spaarsaamste kladogramme*, na aanleiding van die sogenaamde "beginsel van spaarsaamheid". Hierdie benadering veronderstel nie noodwendig dat evolusie die eenvoudigste of kortste weg volg nie, maar dit beperk die aantal *ad hoc*-hipotesisse wat nodig sou wees om die waargenome patroon van kenmerkstaatveranderinge te verklaar. Let daarop dat alle informasie wat in die datamatriks bevat is, ook in die kladogram voorkom. 'n Unieke datamatriks kan dus vanaf 'n kladogram opgestel word.



FIGUUR 2: 'n Spaarsaamste kladogram vir die ses taksons in tabel 1, opgestel uit al ses kenmerke in die tabel.



FIGUUR 3: 'n Spaarsaamste kladogram vir die ses taksons in tabel 1 indien 'n gewig van 3 aan kenmerk 6 en gewigte van 1 aan al die ander kenmerke toegeken word.

Ter opsomming, 'n kladistiese taksonoom konstrueer 'n datamatriks van taksons en kenmerke en verkry 'n kladogram daaruit. Die kladogram klassifiseer die taksons op grond van hipotesisse oor die mees onlangse gemeenskaplike voorouer(s).

KOMPLIKASIES MET DIE VERKRYGING VAN KLADOGRAMME

'n Aantal probleme kan opduik wanneer kladogramme opgebou word. Eerstens kan daar meer as een binêre boom wees met 'n minimum aantal kenmerkstaatveranderinge. Beide die kladogramme in figuur 2 en 3 is byvoorbeeld spaarsaamste kladogramme wat uit die matriks verkry kan word. Dit is onmoontlik om een spaarsaamste kladogram bo 'n ander te verkies sonder om van informasie gebruik te maak wat nie bevat is in die datamatriks nie. Dit word duidelik wanneer opgemerk word dat dieselfde datamatriks opgestel word vanaf beide kladogramme. In groot datamatrikse met talle konflikterende kenmerke is daar dikwels etlike honderde spaarsaamste kladogramme wat almal ewe veel kenmerkstaatveranderinge het.

Tweedens, die aanname dat kladogramme binêre bome moet wees, is nie altyd regverdigbaar nie. Die bome hoef nie noodwendig binêr te wees nie as gevolg van faktore soos allopatriese spesiasie, dit wil sê, die verskynsel dat geografiese verwydering van groepe van dieselfde organisme tot spesiasie kan lei. Drie of meer groepe van dieselfde plantsoort wat geografies van mekaar verwyder is, mag byvoorbeeld oor 'n tydperk ontwikkel in verskillende soorte wat almal dieselfde mees onlangse voorouer het. Kladogramme hoef nie eens bome te wees nie, as gevolg van die voorkoms van hibriede. Die kultivering van 'n nuwe rooskultivar deur die kruising van twee verskillende kultivars sal byvoorbeeld tot gevolg hê dat een kultivar twee mees onlangse voorouers het.

Tabel 2 'n Datamatriks met 16 taksons en 22 kenmerke wat, met enkele uitsonderings, 'n uittreksel is uit 'n publikasie van Van Wyk[3]

Taksons	Kenmerke 1	6	11	15	21	22	23	25	31	32	33	34	35	36	37	42	43	47	48	49	50	51
hypa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
list	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0	1	1	1	0	0	0	0
digi	0	2	0	1	0	0	1	2	0	0	0	0	1	1	0	0	1	1	0	0	0	0
lept	0	2	0	1	0	0	0	1	1	0	1	0	0	2	0	0	1	1	0	0	0	0
lipo	0	2	0	0	0	0	0	1	1	0	1	0	0	2	0	0	1	1	0	0	0	0
leob	1	2	0	2	0	0	1	1	1	0	1	0	0	2	0	0	1	1	0	0	0	0
sync	1	2	0	1	0	0	1	0	1	0	1	0	0	2	0	1	1	1	0	0	0	0
euch	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	2	1	0	0	0	1	1
oxyd	0	2	1	0	0	1	1	0	0	1	0	0	2	0	0	1	1	1	0	1	1	1
loto	0	1	0	0	1	0	1	1	1	1	0	1	1	0	0	2	1	1	1	1	1	1
aula	0	2	0	0	1	0	1	2	0	1	0	0	1	0	0	2	1	1	1	1	1	1
poly	0	0	0	0	0	0	1	1	1	1	0	1	1	0	1	2	1	1	1	1	1	1
clei	0	2	0	2	0	1	1	0	0	1	0	0	1	0	0	2	1	1	1	0	1	1
mono	1	2	0	2	1	1	1	0	0	1	0	1	1	0	1	1	1	1	0	1	1	1
kreb	0	0	1	0	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1	1	1
buch	0	0	1	0	1	0	1	1	1	1	1	1	1	2	1	1	0	0	1	1	1	1
Gewigte	7	7	7	7	7	7	5	6	6	7	6	6	6	6	7	5	5	6	1	2	3	4

'n Derde probleem het te doen met die doelmatigheid van algoritmes om spaarsaamste kladogramme te verkry. Graham & Foulds het aangetoon dat die geassosieerde beslissingsprobleem om spaarsaamste binêre bome te verkry NP-volledig is.[2] Dit beteken dat, in die geval van probleme met 'n groot aantal kenmerke en taksons, al bestaan daar 'n unieke spaarsaamste boom, geen algoritme gewaarborg kan wees om dit in 'n redelike tyd te vind nie.

'n Vierde probleem mag wees dat, hoewel daar biologiese regverdiging vir die kriterium van spaarsaamheid bestaan, die "korrekte" boom nie altyd noodwendig een van die spaarsaamste bome hoef te wees nie.

HUIDIGE OPLOSSINGS

Die probleem van die keuse tussen verskillende spaarsaamste kladogramme word gewoonlik opgelos deur van verskillende vorms van intuïsie gebruik te maak. Die een vorm het te doen met die betroubaarheid van kenmerke. Numeriese gewigte wat die betroubaarheid van kenmerke verteenwoordig, word aan elke kenmerk toegeken. Die gewigte is positiewe heelgetalle, waar die betroubaarste aangedui word deur 'n gewig van 1. Die spaarsaamheid van 'n kladogram word dan gemeet deur elke kenmerkstaatverandering in die kladogram te vermenigvuldig met sy geassosieerde gewig en hierdie waardes te sommeer. Veronderstel 'n gewig van 1 word aan kenmerke 1 tot 5 in tabel 1 toegeken en 'n gewig van 3 aan kenmerk 6. Die waarde van spaarsaamheid van die kladogram in figuur 2 sal dan 15 wees en dié van die kladogram in figuur 3 sal 13 wees. In baie gevalle sal weging van kenmerke die aantal spaarsaamste kladogramme verminder, maar dit is nie 'n waarborg vir 'n unieke spaarsaamste kladogram nie.

Die ander vorm van intuïsie wat ter sprake is, behels die saamgroepering van taksons wat blyk 'n verwantskap met mekaar te behou in 'n groot aantal van die spaarsaamste kladogramme. Dit word gekombineer met die bioloog se kennis oor die taksons.

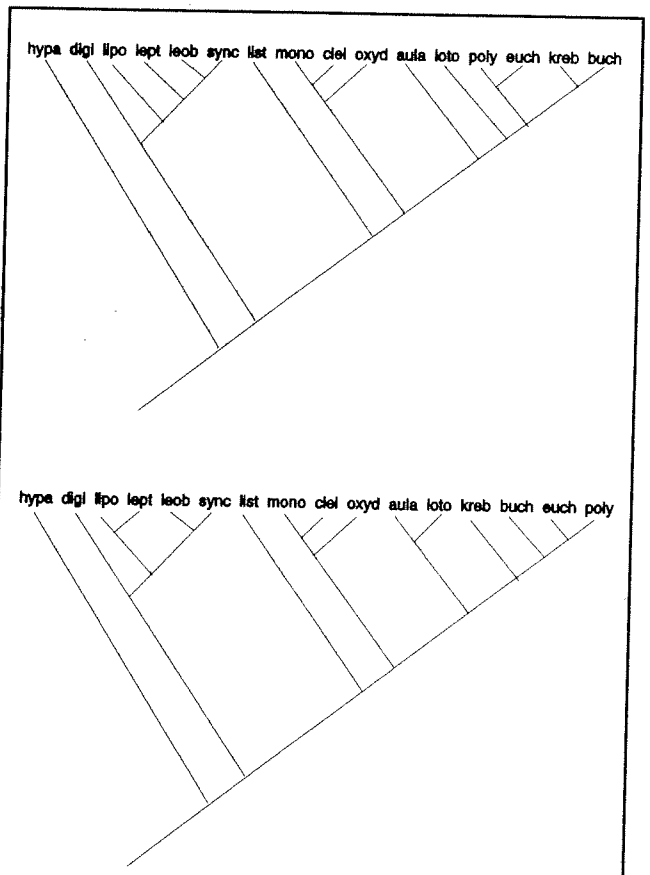
Die feit dat spaarsaamste kladogramme vir groot probleme nie in 'n redelike tyd deur enige algoritme verkry kan word nie, dwing die bioloog om dikwels nie op spaarsaamste kladogramme aan te dring nie. Daar word dan gewoonlik gebruik gemaak van 'n heuristiese metode wat kladogramme produseer met 'n waarde van spaarsaamheid wat hopelik baie naby aan die spaarsaamste is. Die aanname dat kladogramme binêre bome moet wees, word gewoonlik nie bevraagteken nie.

'n Voorbeeld word in tabel 2 gevind, wat met enkele veranderinge 'n uittreksel is uit 'n publikasie van Van Wyk.[3] Die tabel bevat 'n datamatriks met 16 taksons en 22 kenmerke. Die ry onder die laaste takson dui die gewigte aan wat aan die kenmerke toegeken is. Met die gebruik van hierdie gewigte is daar nog steeds 14 spaarsaamste kladogramme met 'n "lengte" van 374. Twee van hierdie 14 kladogramme word in figuur 4 aangedui. In die praktyk sal 'n kladistiese taksonoom heelwat tyd nodig hê om vertroud te raak met die informasie bevat in hierdie 14 kladogramme voordat hy sy intuïsie sal kan gebruik om 'n finale klassifikasie op te stel. Hierdie intuïsie word tipes gegrond op die bioloog se gevoel dat sekere taksons meer primitief as ander is, maar word gewoonlik nie eksplisiet gemaak nie. Wat nodig is, is 'n manier om sistematies op grond van die datamatriks te besluit of een takson meer primitief as 'n ander is, en dus om die basis waarop 'n kladogram gekies word, eksplisiet te maak.

GRADOGRAMME

Die primêre probleem met die huidige metode vir die verkryging van 'n finale kladogram blyk te wees dat heelwat implisiete informasie gebruik word wat die resultaat affekteer, maar wat nie duidelik onderskei word van die eksplisiete informasie nie. Die blote insluiting of weglating van bepaalde kenmerke (doelbewus, of meer dikwels die gevolg van tegniese beperkings) kom reeds neer op implisiete weging wat die uiteindelike resultaat grootliks mag beïnvloed. 'n Metode word voorgestel wat die eksplisiet formuleerbare informasie ten volle benut om 'n voorlopige ordening van die taksons te lewer, waarna die bioloog sy implisiete agtergrondkennis en intuïsie kan gebruik om die finale klassifikasie te konstrueer. Vanaf 'n gegewe datamatriks, ordenings op die kenmerkstate van elke kenmerk (op grond van hoe gevorderd 'n kenmerkstaat is), en 'n ordening op die kenmerke (op grond van byvoorbeeld betroubaarheid), word 'n grafiek aan die bioloog verskaf wat 'n *gradogram* genoem word. Die woord gradogram is afgelei van die Latynse woord "gradus" wat "trap" of "helling" beteken en die Griekse woord "gramma" wat beteken "geskrewe ding". 'n Gradogram word voorgestel as 'n hulpmiddel om die bioloog in staat te stel om die finale kladogram op te stel. In 'n gradogram word taksons gerangskik in grade van relatiewe spesialisasie, wat die taksonoom help om waarskynlike klades te visualiseer.

Die onderliggende idee is dat die ordenings op die kenmerkstate van elke kenmerk, tesame met die ordening op die kenmerke, 'n ordening op die taksons (wat as



FIGUUR 4: Twee van die veertien spaarsaamste kladogramme vir die taksons in tabel 2, met die gewigte op kenmerke soos aangedui in tabel 2.

vektore van kenmerkstate gesien kan word) sal induseer, wat dan gebruik kan word om die gradogram op te stel. In hierdie artikel word voorgestel dat 'n takson b gesien moet word as ten minste net so gespesialiseerd soos 'n takson a as en slegs as, vir élke kenmerk, takson b ten minste net so gevorderd is soos takson a t.o.v. dié kenmerk, of andersins moet daar 'n kenmerk hoër op in die ordening op kenmerke wees (byvoorbeeld 'n meer betroubare kenmerk), waarvoor takson b meer gevorderd is as takson a. 'n *Saamgestelde* voorrangrelasie word dus opgebou uit die elementêre gevorderdheidsordenings op die verskillende kenmerke, tesame met die ordening op die versameling van kenmerke.

Hierdie voorrangrelasie is geformaliseer deur Burger & Heidema op die volgend wyse:[4] Gegee 'n versameling van m taksons en n kenmerke, laat $\mathbf{N} = \{1,2,...,n\}$. Elke kenmerk i, waar $i \in \mathbf{N}$, word verteenwoordig deur $(\mathbf{A}_i, \leq_i)$, waar $\mathbf{A}_i$ die versameling van alle moontlike kenmerkstate van kenmerk i is en $\leq_i$ die ordening op die kenmerkstate van karakter i volgens relatiewe gevorderdheid is. $W=\{(a_i)_{i\in N} \mid a_i \in \mathbf{A}_i \text{ vir alle } i \in \mathbf{N}\}$ is die produkversameling wat alle moontlike taksons relatief tot die n kenmerke verteenwoordig. Laat $\mathbf{T} \subseteq \mathbf{W}$ die deelversameling van $\mathbf{W}$ wees wat die m taksons onder bespreking verteenwoordig. Vir tabel 1 is $n=6$, $m=6$, $A_1=A_2=A_3=A_4=A_6=\{0,1\}$ en $\mathbf{A}_5=\{0,1,2\}$. Verder is $\leq_i$, $i=1,2,3,4,6$ die normale lineêre ordening "kleiner as of gelyk aan" op die versameling $\{0,1\}$ en $\leq_5$ is dieselfde ordening, maar op die versameling $\{0,1,2\}$. Die takson BG in tabel 1 word verteenwoordig deur die vektor 000000, die takson A deur 110001, ensovoorts.

Die ordening op kenmerke volgens, byvoorbeeld, betroubaarheid word aangedui deur $(\mathbf{N}, \subseteq)$. Die uitdrukking $j \subseteq k$ beteken dat k ten minste net so betroubaar as j is, terwyl $j \subset k$ beteken dat k meer betroubaar as j is. Die ordening op die elemente van $\mathbf{T}$ (d.w.s die taksons) wat gebruik word om die gradogram te verkry, word as volg gedefinieer: Vir elke twee elemente $a=(a_i)_{i\in N}$ en $b=(b_i)_{i\in N}$ van $\mathbf{T}$,

$$a \leq b \Leftrightarrow (\forall j \in \mathbf{N})[a_j \leq_j b_j \text{ of } (\exists k \in \mathbf{N})(j \subset k \text{ en } a_k <_k b_k)].$$

As $a \leq b$ word gesê dat b ten minste net so gespesialiseerd as a is, terwyl $a<b$ aandui dat b beslis meer gespesialiseerd as a is, (wat uitsluit dat $b \leq a$ ook kan geld).

Beskou byvoorbeeld die taksons A,B en C in tabel 1 en $(A_i, \leq_i)$ vir $i \in \{1,2,3,4,5,6\}$, soos hierbo gegee. Die takson A word aangedui as $a=a_1a_2a_3a_4a_5a_6$, takson B as $b=b_1b_2b_3b_4b_5b_6$ en takson C as $c=c_1c_2c_3c_4c_5c_6$. Veronderstel nou daar is geen voorrang tussen kenmerke in die datamatriks nie, wat beteken dat die ordening op kenmerke as die identiteitsrelasie geneem word. Dit wil sê, vir alle $i,j \in \{1,2,3,4,5,6,\}$ is $i \subseteq j$ as slegs as $i=j$. Volgens hierdie ordening $\leq$ op taksons is A ten minste net so gespesialiseerd soos B, dit wil sê $b \leq a$, want vir elk van die ses kenmerke is a ten minste net so gevorderd soos b, dit wil sê $b_i \leq_i a_i$ vir $i=1,2,3,4,5,6$. Dit is egter nie die geval dat C ten minste so gespesialiseerd is soos B nie, want C is nie ten minste so gevorderd op kenmerk 2 soos B nie en daar is ook geen kenmerk wat voorrang het oor kenmerk 2 nie, dit wil sê $b_2 \not\leq_2 C_2$ en daar is geen k waarvoor $2 \subset k$ nie. Veronderstel nou die ordening op kenmerke word verander deur te bepaal dat daar geen voorrang tussen alle kenmerke, uitgesluit kenmerk 3, is nie, maar dat kenmerk 3 voorrang bo al die ander kenmerke het. Dit wil sê, vir alle $i,j \in \{1,2,3,4,5,6\}$ is $i \subseteq j$ as en slegs as $i=j$ of $j=3$ (of beide). Met hierdie nuwe ordening $\subseteq$ op taksons is C wel ten minste net so gespesialiseerd soos B, dit wil sê $b \leq c$, want vir vyf van die ses kenmerke, kenmerk 2 uitgesluit, is c ten minste net so gevorderd soos b, dit wil sê $b_i \leq_i c_i$ vir $i=1,3,4,5,6$, en vir kenmerk 2 is daar 'n kenmerk, kenmerk 3, met groter voorrang, waarvoor c meer gevorderd is as b, dit wil sê $2 \subset 3$ en $b_3 <_3 c_3$.

'n *Gradogram* is nou 'n diagrammatiese voorstelling van die relasie met die volgende konvensies:

- as $a \leq b$ en $b \leq a$ (d.w.s. $a_i = b_i$ vir alle $i \in \mathbf{N}$) word a en b as dieselfde nodus beskou en daar word dus verder net verwys na gevalle waar $a<b$;
- as $a<b$ dan word b bokant a geplaas met 'n *pad* wat a en b aan mekaar verbind;
- as $a<b$ dan word hulle met 'n enkele *lyn* verbind indien daar geen c is sodanig dat $a<c<b$ nie.

Die gradogram in figuur 5 word byvoorbeeld verkry vanaf tabel 1 indien daar geen voorrang tussen die kenmerke in die datamatriks is nie (d.w.s. $\subseteq$ is die identiteitsrelasie).

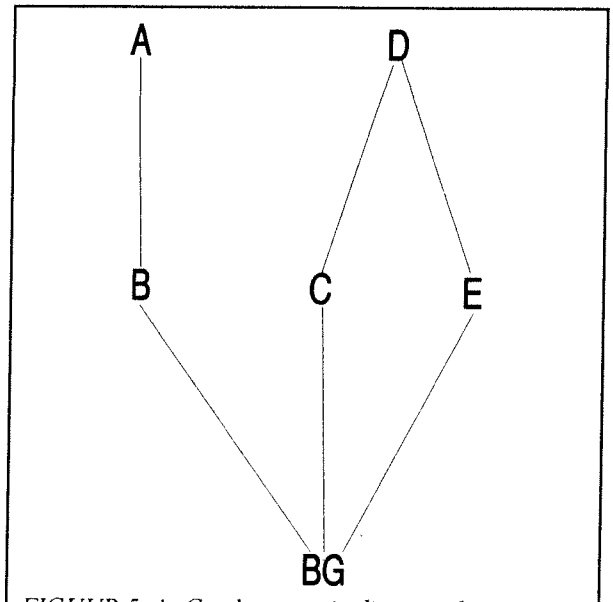
Dit is belangrik om te besef dat gradogramme nie dieselfde informasie oordra as kladogramme nie. 'n Gradogram reflekteer die relatiewe spesialisasie van taksons met betrekking tot die ordenings op kenmerkstate en die ordening op die kenmerke, terwyl 'n kladogram die taksons klassifiseer volgens die mees onlangse gemeenskaplike voorouers.

Die feit dat taksons D in die gradogram in figuur 5 bo taksons C en E is, beteken dus nie dat C en E voorouers van D is nie, maar slegs dat D meer gespesialiseerd is as sowel C en E, relatief tot kenmerke 1 tot 6.

VOORDELE VAN GRADOGRAMME

Die probleem om 'n keuse te maak tussen verskillende spaarsaamste kladogramme kom nie voor wanneer gradogramme opgestel word nie. Vir elke datamatriks, elke stel ordenings op die kenmerkstate en elke ordening op die kenmerke, word 'n unieke gradogram verkry.

Die feit dat spaarsaamste kladogramme vir groot



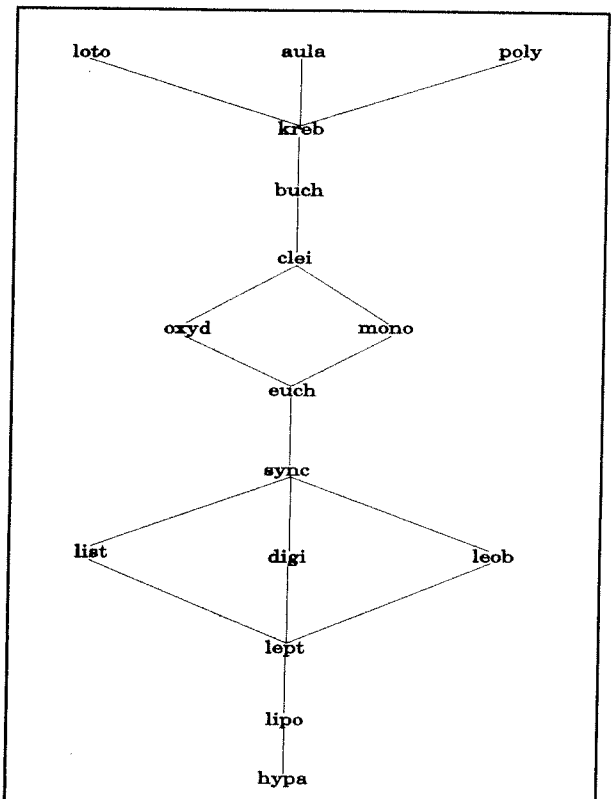
FIGUUR 5: 'n Gradogram vir die ses taksons in tabel 1, opgestel uit al ses kenmerke in die tabel, met die ordening op kenmerke geneem as die identiteitsrelasie.

probleme nie in 'n redelike tyd deur enige algoritme verkry kan word nie, speel geen rol in die opstel van gradogramme nie. Dit is duidelik dat 'n polinomiese algoritme, een in die orde van m^2n^2 berekenings, oftewel een in $O(m^2n^2)$, gevind kan word om 'n gradogram op te stel. Om te sien hoekom dit so is, let daarop dat elk van die m taksons met m-1 taksons vergelyk moet word. Dit gee $m(m\text{-}1)$ vergelykings. Elke vergelyking van twee taksons behels weer die vergelyking van hoogstens n kenmerke van die twee taksons. Verder behels die vergelyking van 'n kenmerk van twee taksons die vergelyking van hoogstens n-1 kenmerke van die twee taksons. Daar word dus hoogstens $m(m\text{-}1)n(n\text{-}1)$ vergelykings gedoen, wat beteken die aantal vergelykings is in $O(m^2n^2)$.

Die bruikbaarheid van gradogramme is tweeledig. Eerstens word die kladistiese taksonoom gevra om sy intuïsie oor die betroubaarheid van kenmerke eksplisiet te maak en word dié inligting op 'n ander wyse benut as by kladogramme. Tweedens verskaf 'n gradogram informasie oor watter taksons kragtens spesialisasie saamgegroepeer moet word. Dit is informasie wat die taksonoom gewoonlik met heelwat moeite verkry deur rond te speel met verskeie kladogramme.

Die gradogram in figuur 6 is opgestel vanaf tabel 2, met die normale "kleiner as of gelyk aan" lineêre ordening op die kenmerkstate van elke kenmerk. Die ordening op kenmerke is verkry deur van die gewigte in tabel 2 gebruik te maak. Dit blyk makliker en meer betekenisvol vir 'n kladistiese taksonoom te wees om gebruik te maak van 'n gradogram om 'n finale kladogram te visualiseer, in plaas van om sin te probeer maak uit 14 spaarsaamste kladogramme.

Ter opsomming, die vind van 'n kladogram wat taksons klassifiseer op grond van hulle mees onlangse voorouers word bemoeilik deurdat meer as een spaarsaamste kladogram uit dieselfde datamatriks verkry kan word en daar geen algoritme bestaan wat gewaarborg is om spaarsaamste kladogramme in 'n redelike tyd te vind nie. Die bioloog verkry dus die finale kladogram deur van kennis gebruik te maak wat nie in die datamatriks bevat is nie. Daarenteen is daar 'n unieke gradogram vir 'n gegewe datamatriks, daar bestaan 'n doelmatige algoritme vir die verkryging van gradogramme en alle informasie wat gebruik word om 'n gradogram op te stel, is bevat in die datamatriks, die ordenings op kenmerkstate en die ordening op die kenmerke. Gradogramme kan dus gebruik word om die basis waarop die finale kladogram gekies word, eksplisiet te maak.



FIGUUR 6: 'n Gradogram vir die taksons in tabel 2, opgestel uit al die kenmerke in die tabel. Die ordening op kenmerke is gebaseer op die gewigte in tabel 2.

SUMMARY

WHAT IS CLADISTIC TAXONOMY?

A cladistic taxonomist uses binary trees called *cladograms* to represent classifications of groups of organisms (called *taxons*) according to their most recent common ancestors. A cladogram is obtained by using a data matrix containing *character states* for a number of *characters* of the taxons to be classified, and grouping together taxons with *shared derived character states*. At every branching of a cladogram, all taxons with a shared derived character state for a specific character are associated with one of the two branches. The taxons under consideration are all leaves of the binary tree. Every internal node n represents the hypothetical ancestor of all the nodes in the subtree rooted at n. An indication of a character in a cladogram represents a character state change from a primitive to an advanced character state.

It is not always possible to obtain binary trees from a given data matrix. This can be explained by *reversal*, a character state change from an advanced character state to a more primitive one, and *convergence*, the independent development of the same character state by a number of taxons. Cladistic taxonomists are interested in *most parsimonious cladograms*, that is cladograms with a minimum number of character state changes.

COMPLICATIONS IN OBTAINING CLADOGRAMS AND CURRENT SOLUTIONS

A problem in obtaining cladograms is that there may be more than one most parsimonious cladogram. This is addressed by the *weighting* of characters, which may reduce the number of most parsimonious cladograms, and by using knowledge the biologist has about the taxons; knowledge that is usually not made explicit. Another problem is that there is no algorithm that is guaranteed to find a most parsimonious tree in a reasonable time. This problem is addressed by not insisting on most parsimonious cladograms.

Some other problems are that cladograms need to be binary trees because of the possible geographic separation of groups of organisms of the same species, that cladograms need not be trees because of the occurence of hybrids, and that the "correct" cladogram need not

necessarily be a most parsimonious cladogram. These problems are usually not addressed.

GRADOGRAMS

A method is proposed that will produce a graph called a *gradogram* from a data matrix, orderings on the character states of every character according to how advanced the character states are, and an ordering on the characters (based on, for instance, reliability). Gradograms arrange taxons in degrees of relative specialisation, which the biologist can use to construct a final classification. A *composite preference relation* is constructed from the elementary orderings on character states and the ordering on the set of characters, which is formalised as follows:

Given a set of m taxons and n characters, let $\mathbf{N} = \{1,2,\ldots,n\}$. Every character i, where $i \in \mathbf{N}$, is represented by $(\mathbf{A}_i, \leq_i)$, where $\mathbf{A}_i$ is the set of possible character states of character i and $\leq_i$ is the ordering on the character states of character i. $\mathbf{W}=\{(a_i)_{i\in \mathbf{N}} \mid a_i \in \mathbf{A}_i \text{ for all } i \in \mathbf{N}\}$ is the product set representing all possible taxons relative to the n characters. The ordering on characters is denoted by $(\mathbf{N},\subseteq)$. The expression $j \subset k$ means $j \subseteq k$ and $k \not\subseteq j$. The ordering on the elements of $\mathbf{T}$ (i.e. the taxons) that is used to construct the gradogram is defined as follows: for every two elements $a=(a_i)_{i\in \mathbf{N}}$ and $b=(b_i)_{i\in \mathbf{N}}$ of $\mathbf{T}$,

$$a \leq b \Leftrightarrow (\forall j \in \mathbf{N})[a_j \leq_j b_j \text{ or } (\exists k \in \mathbf{N})(j \subset k \text{ and } a_k <_k b_k)].$$

If $a \leq b$ it is said that b is at least as specialised as a, while $a < b$ indicates that b is more specialised than a (which excludes $b \leq a$).

A *gradogram* is a diagrammatic representation of the relation with the following conventions:

- if $a \leq b$ and $b \leq a$ (i.e. $a_i = b_i$ for all $i \in \mathbf{N}$), a and b are viewed as the same node;
- if $a < b$, b is placed above a with a *path* connecting a and b;
- if $a < b$, a and b are connected with a single *line* if there are no c such that $a < c < b$.

ADVANTAGES OF GRADOGRAMS

There is a unique gradogram obtainable from every given data matrix, there is an efficient algorithm, one in $O(m^2n^2)$, to construct a gradogram from a given data matrix and all the information contained in a gradogram is contained in the data matrix, the orderings on the character states and the ordering on the characters. Gradograms provide the biologist with a systematic way of deciding whether one taxon is more primitive than another and therefore make explicit the basis on which the final cladogram is chosen.

LITERATUURVERWYSINGS

1. Henning, W. (1966). *Phylogenetic Systematics. Translated by D.D. Davies and R. Zangerl.* (University of Illinois Press, Urbana).
2. Graham, R.L. & Foulds, L.R. (1982). Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time. *Mathematical Biosciences*, 60, 133-142.
3. Van Wyk, B-E. (1991). *A synopsis of the genus Lotononis (Fabaceae: Crotalarieae), Contributions from the Bolus Herbarium* 14 (University of Cape Town), p. 51.
4. Burger, I.C. & Heidema, J. (1993). Qualitative models of composite preference relations. In *Philosophy and the cognitive sciences. Papers of the 16th international Wittgenstein symposium*, Casati, R. & White, G. eds. (Kirchberg am Wechsel: The Austrian Ludwig Wittgenstein Society) pp. 71-75. (Voordruk beskikbaar op aanvraag.)