

LEZEN EN SCHRIJVEN MET BOMEN

Duur K. Aanen

Laboratorium voor erfelijkheidsleer,

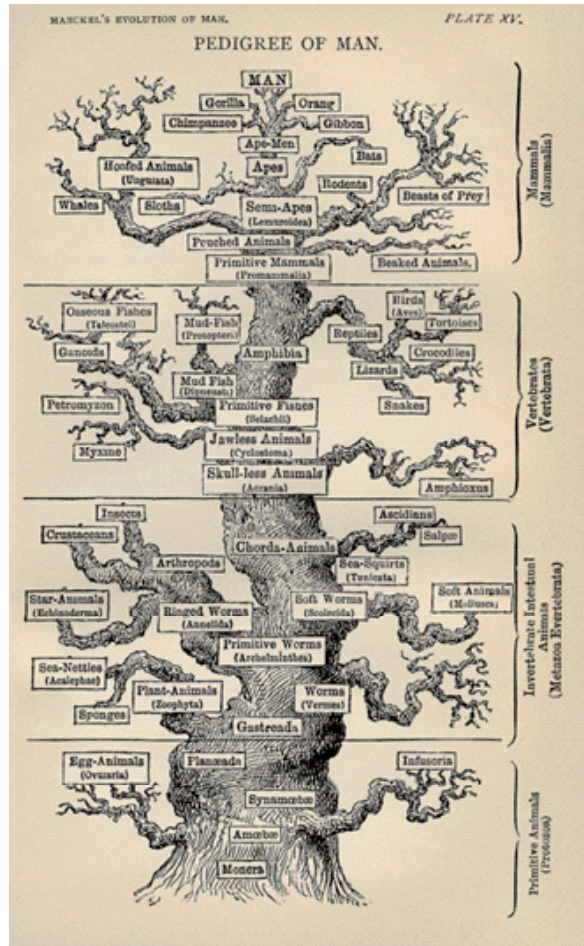
Wageningen Universiteit, Droevendaalsesteeg 1, 6708 PB, Wageningen

Aanen, D.K. 2012. Reading and writing with trees. *Coolia* 55(2): 87–94.

This paper is about the interpretation of phylogenetic trees and some persistent and widespread misunderstandings. Since species with living representatives all descend from the same single ancestor, labelling species as ‘lower’, ‘primitive’, ‘ancestral’ or ‘basal’ or ‘higher’, ‘advanced’, ‘derived’ or ‘apical’ is not justified. A precise and careful description of phylogenies is proposed, with the realisation that taxon sampling is always limited and usually biased towards the group of interest. Furthermore, phylogenies are reconstructions, meaning that phylogenies are hypotheses with some uncertainty, and not absolute truths.

De reconstructie van evolutionaire stambomen (fylogenieën) op basis van DNA-sequenties heeft de laatste decennia een grote vlucht genomen en een ware revolutie tewegegebracht in de biologie. Om enkele uiteenlopende – soms bizarre – voorbeelden te noemen: Het is mogelijk gebleken om van uitgestorven kikkers de roep te reconstrueren op basis van de roep van nu levende kikkersoorten (Ryan & Rand, 1995). Ook kunnen we dankzij stamboomreconstructies schatten hoe vaak lamellen zijn ontstaan in de evolutie van paddenstoelen (Hibbett et al., 1997). Verder weten we nu dat parasolmieren in hun evolutionaire geschiedenis meerdere malen schimmels hebben gedomesticeerd (Mueller et al., 1998), terwijl hun tegenhangers in de oude wereld, schimmelkwekende termieten, dit slechts eenmaal hebben gedaan (Aanen et al., 2002). Ook kunnen we op basis van gereconstrueerde fylogenieën taxonomische indelingen herzien – soms drastisch.

Figuur 1. Een vroege fylogenetische boom, gemaakt door Ernst Haeckel. Deze zienswijze van de diversiteit van het leven is duidelijk beïnvloed door de idee van vooruitgang. (Bron: Haeckel, 1879)

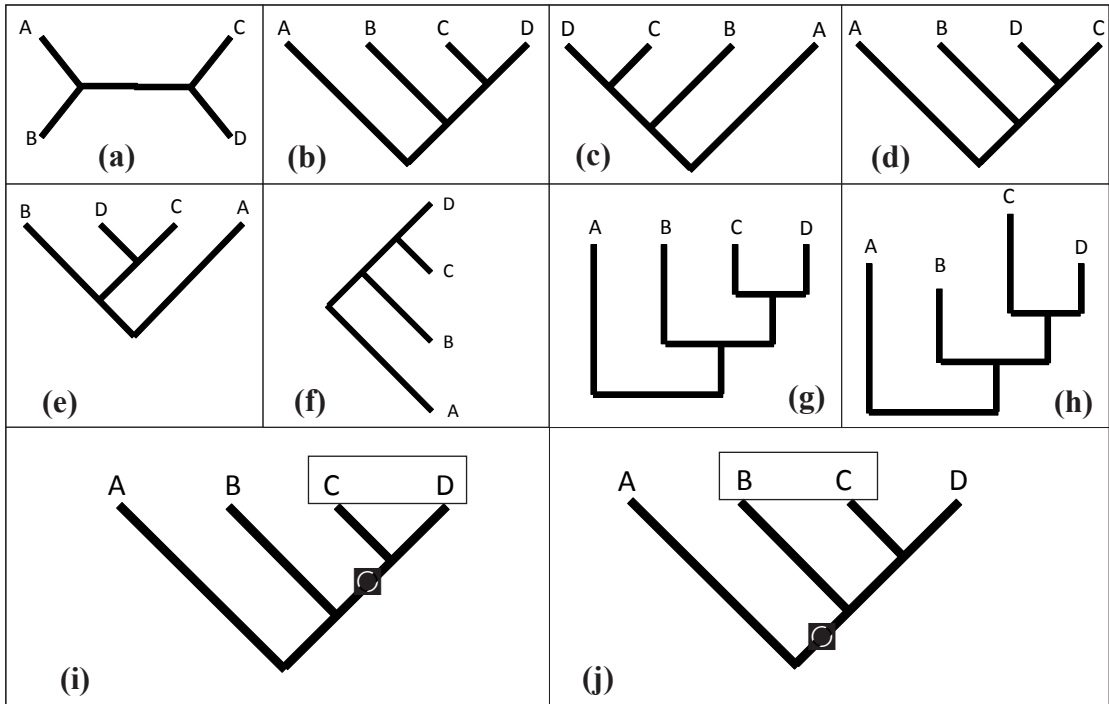


Helaas worden stambomen niet altijd even nauwkeurig geïnterpreteerd. De wijdverbreide idee dat evolutie een proces is van vooruitgang en soms ook met een doel vinden we soms weerspiegeld in de weergave en interpretatie van stambomen. Dit is bijvoorbeeld duidelijk te zien aan een van de eerste fylogenetische reconstructies van het dierenrijk, door Ernst Haeckel (Figuur 1). De mens is het voorlopige sluitstuk van evolutie en staat dus boven in de boom, terwijl andere soorten lager in de ‘boom des levens’ staan afgebeeld. Dit beeld van *the ladder of progress* is, ten onrechte, nog steeds wijdverbreid, zowel in de wetenschappelijke als de populair-wetenschappelijke literatuur, en we kunnen regelmatig beschrijvingen van stambomen vinden die voortkomen uit dit beeld (Omland et al., 2008).

In dit artikel wil ik ingaan op de vraag hoe we fylogenieën dan wel kunnen interpreteren. Wat zijn veel voorkomende misverstanden? Hoe kunnen we een boom ‘lezen’, zonder in de valkuil van *the ladder of progress* te stappen? Wat betekent de lengte van de takken? Kunnen we primitieve en geavanceerde soorten onderscheiden? Hoe kunnen we een boom vertalen naar een taxonomie?

Wat is een fylogenie of evolutionaire stamboom?

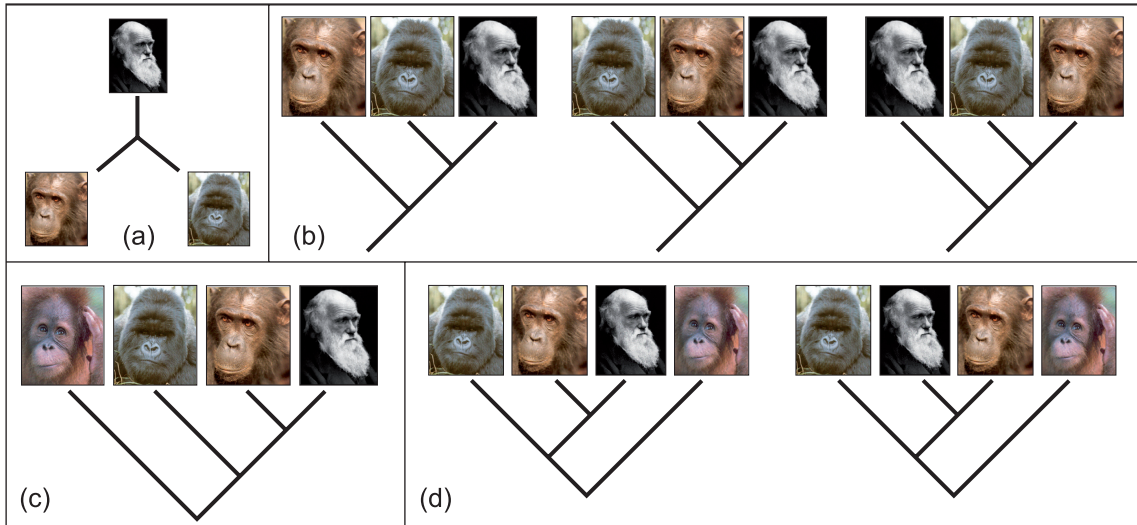
Een evolutionaire stamboom, of fylogenie, is een reconstructie van evolutionaire verwantschappen tussen een groep taxonomische eenheden. Voor deze reconstructie worden overeenkomsten in kenmerktoestanden gebruikt, tegenwoordig meestal DNA-sequenties, en een criterium om de boom te reconstrueren, zoals *maximum parsimony*, of *maximum likelihood*. In de meeste gevallen zijn de taxonomische eenheden van de stamboomreconstructie, de eindtakken dus, soorten. Maar we kunnen ook fylogenieën reconstrueren voor andere eenheden die in een hiërarchisch patroon met elkaar verwant zijn. Zo is er een fylogenie gereconstrueerd van 58 van de 80 overgebleven manuscripten van *The Canterbury Tales* (Barbrook et al., 1998). Evenals soorten van gemeenschappelijke voorouders afstammen, stammen de overgebleven manuscripten af van gemeenschappelijke ‘vooroudermanuscripten’, die op hun beurt weer afstammen van nog oudere versies, die uiteindelijk zijn terug te voeren tot het oorspronkelijke ‘oer-manuscript’. Tijdens het overschrijven zijn spelfouten gemaakt, en deze spelfouten zijn vervolgens weer gekopieerd in volgende manuscripten. Gelijkenis tussen manuscripten is dus een aanwijzing voor gemeenschappelijke afstamming van een ‘vooroudermanuscript’. Maar in de meeste analyses zijn de eindtakken van de boom dus soorten. Alle tot nu toe verzamelde evidentie duidt erop dat het leven op aarde één oorsprong heeft, dus alle nu levende organismen stammen uiteindelijk af van één gemeenschappelijke voorouder. Dit betekent dat alle organismen met elkaar verwant zijn en in een hiërarchisch vertakkingspatroon, ‘the tree of life’, kunnen worden weergegeven. Uiteraard worden in de praktijk meestal de evolutionaire verwantschappen gereconstrueerd van een beperkt aantal soorten (Figuur 2a-h). Een fylogenie is dus een hiërarchisch vertakkingspatroon dat naar hypothetische voorouders leidt. Hoe dichter bij de wortel van de boom, hoe minder voorouders, tot er uiteindelijk één gemeenschappelijke voorouder overblijft bij de wortel van de boom. Binnen een stamboom onderscheiden we monofyletische groepen (Figuur 2 i). Een monofyletische groep, of clade, is een groep soorten met een unieke gemeenschappelijke voorouder. Uniek is hier belangrijk, want zoals hiervoor al gezegd, alle soorten, dus ook soorten die geen monofyletische groep vormen, hebben een gemeenschappelijke voorouder (Figuur 2j). Echter, als deze voorouder niet uniek is, spreken we niet van een monofyletische, maar van een polyfyletische of parafyletische groep.



Figuur 2. a. De ongewortelde boom met vier soorten (A-D). b tot en met h: Verschillende weergaven van een en dezelfde gewortelde boom, waarin C en D zustersoorten zijn, en deze samen de zustergroep van B, met A als buitengroep. De bomen a tot en met g zijn alle cladogrammen, dat wil zeggen, deze bomen geven het relatieve vertakkingspatroon weer, zonder een betekenis voor de taklengte. Figuur h, daarentegen, is een fylogram: in deze figuur geven de verschillende taklengtes verschillende aantallen gereconstrueerde veranderingen weer. i. C en D vormen samen een monofyletische groep, omdat zij samen een unieke meest recente gemeenschappelijke voorouder hebben, weergegeven op de tak die naar C en D leidt. j. B en C daarentegen vormen geen monofyletische groep, want de gemeenschappelijke voorouder van hen is niet uniek voor B en C, maar ook de voorouder van D.

Het is belangrijk dat we ons realiseren dat een fylogenie een reconstructie of een hypothese is (dus geen constructie, zoals je soms leest). Dit betekent dat fylogenetische bomen altijd een bepaalde onzekerheid hebben. Dit maakt het vaak lastig fylogenetische verwantschappen direct te vertalen naar nieuwe taxonomische indelingen. De mate van onzekerheid kan worden geschat met verschillende parameters, zoals bootstrapwaarden of *a posteriori* waarschijnlijkheid.

Meestal wordt een boom afgebeeld met een *wortel* dat wil zeggen met een richting (Figuur 2). Om een boom 'te wortelen', dus van een richting te voorzien, wordt meestal een zogenaamde buitengroep aan de analyse toegevoegd. Dit is een soort, of een aantal soorten, waarvan vooraf al is vastgesteld of bepaald dat deze buiten de bestudeerde groep valt. Bij voorbeeld, als we verwantschappen tussen verschillende families basidiomyceten willen bepalen, kunnen we een ascomyceet als buitengroep gebruiken. Of als we mensapen bestuderen, kunnen we halfapen als buitengroep gebruiken.



Figuur 3. *Fylogenetische relaties tussen mensapen (mens, chimpansee en gorilla). Voor drie soorten is er één ongewortelde boom mogelijk (a), en drie verschillende gewortelde bomen (b). De fylogenie met de meeste steun (op basis van DNA-gegevens) is weergegeven in c, waarbij de orang-oetan als buitengroep is genomen. In d zijn verschillende weergaves van dezelfde boom getekend die een identieke fylogenie representeren.*

Welke betekenis kunnen we toekennen aan een vertakkingspatroon?

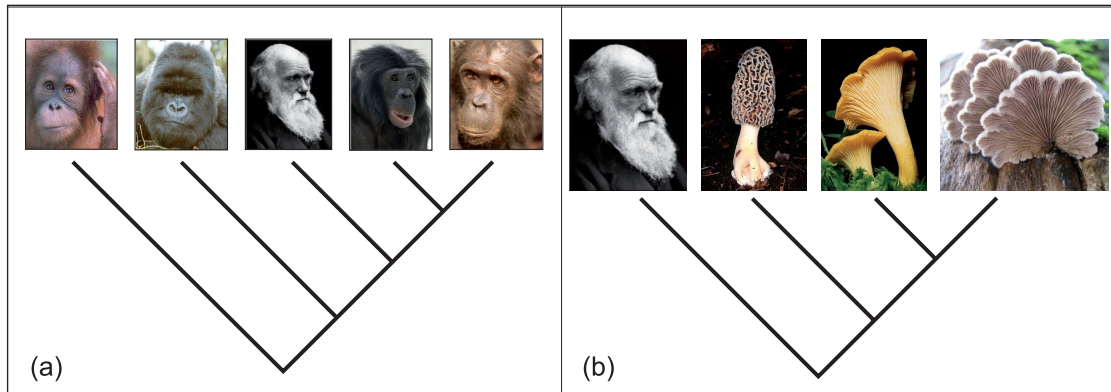
In de beschrijving van fylogenieën worden eindtakken met nu levende soorten vaak beschreven als voorouder van, of basaal aan, andere takken. Reptielen zijn voorouder van zoogdieren en amfibieën voorouder van reptielen. Het benoemen van een ‘eindtak’ in een stamboom, een nu levende soort dus, als voorouder van een andere soort is in strijd met een fundamenteel inzicht, waar alle tot nu toe verzamelde evidentie op duidt: dat het leven op aarde één oorsprong heeft. Dit betekent dat alle nu levende organismen van één gemeenschappelijke voorouder afstammen, en dus alle even oud zijn. Zoogdieren bestaan naast nu levende reptielen, en reptielen naast nu levende amfibieën. Basidiomyceten bestaan naast ascomyceten, en ascomyceten naast zygomyceten. Al deze soorten zijn terug te voeren naar één gemeenschappelijke voorouder.

Beschouw het volgende voorbeeld, de fylogenetische relaties tussen enkele mensapen, inclusief de mens. Voor drie soorten, in dit voorbeeld mens, gorilla en chimpansee, is er slechts één ongewortelde boom mogelijk, en drie verschillende gewortelde bomen (Figuur 3a,b). Vergelijking van DNA-sequenties van deze soorten en het toevoegen van een vierde soort, de orang-oetan als zogenaamde buitengroep, geeft de meeste steun aan een fylogenie waarin de nauwste verwant van de mens de chimpansee is, en dat deze samen de gorilla als nauwste verwant hebben (Figuur 3c).

Hoe kunnen we deze boom interpreteren? Waarschijnlijk zullen velen op basis van deze boom concluderen dat de gorilla primitiever is dan de chimpansee. Of, als we woorden als primitief of geavanceerd willen vermijden, op zijn minst dat de tak die naar de gorilla leidt meer basaal is dan de tak naar de chimpansee.

Maar nu voegen we de bonobo, die een zustergroep is van de chimpansee, toe aan de boom (Figuur 4a). Als we nu dezelfde criteria hanteren als eerst, moeten we concluderen dat

de mens primitiever is dan de chimpansee. Laat ons nu een -voor mycologen- aansprekender voorbeeld nemen. In Figuur 4b zijn de fylogenetische relaties weergegeven tussen het waaiertje, de cantharel en de morielje, met de mens als buitengroep. Als we dezelfde redenering als die van daarnet volgen, moeten we nu concluderen dat de mens de primitieve soort is, of op zijn minst een basale afsplitsing! Verder zou de morielje primitiever zijn dan basidiomyceten. Nu zouden we daar tegen kunnen in brengen dat deze fylogenie niet compleet is. Dat is inderdaad het geval, maar dit bezwaar kun je voor *iedere* willekeurige fylogenie maken! Sterker nog: het is uiterst onwaarschijnlijk dat we ooit alle soorten zullen kennen, laat staan dat we ooit de complete ‘tree of life’ zullen kunnen reconstrueren...



Figuur 4. Voorbeelden met stambomen waarin de mens een van de soorten is. a. *Fylogenetische relatie tussen mensapen. Bonobo en chimpansee zijn zuster groepen, en samen vormen zij de zuster groep van de mens.* b. *Fylogenetische relatie tussen twee basidiomyceten, een ascomycet en de mens als buitengroep. Het waaiertje en de cantharel zijn zuster groepen en samen zijn zij de zuster groep van de morielje.*

Bovenstaande voorbeelden laten zien dat termen als ‘primitief’, ‘lager’ of ‘basaal’ en hun tegenpolen ‘geavanceerd’, ‘hoger’ of ‘apicaal’, niet kunnen worden gebruikt voor soorten, althans nooit in absolute zin. Hooguit om een bepaalde fylogenie, met een bepaalde focus, en dus een bepaalde selectie van soorten, te beschrijven. De reden is simpel: alle nu levende organismen zijn even oud. Iedere uitspraak over een soort op basis van een fylogenie hangt dus af van het perspectief en de focus van een boom.

Taklengte

Er zijn verschillende manieren om dezelfde stamboom weer te geven (Figuur 2). Een belangrijk onderscheid is dat tussen *cladogram* en *fylogram* (Figuur 2h). In een cladogram worden alleen relatieve vertakkingspatronen weergegeven, zonder een betekenis aan de taklengtes toe te kennen. De bomen die we tot nog toe gezien hebben behoren daarom tot deze categorie. De eindtakken met de nu levende soorten staan dan ook op één (denkbeeldige) lijn afgebeeld (hetzij verticaal of horizontaal). Bij een fylogram daarentegen, worden de taklengtes weergegeven, die het gereconstrueerde aantal kenmerkveranderingen representeren. Dit betekent dat de eindtakken niet alle op eenzelfde (denkbeeldige) lijn hoeven te eindigen.

Dit verschil in taklengtes suggereert dat er verschillen zijn in evolutiesnelheid tussen soorten. Dit is inderdaad één mogelijke verklaring: de veranderingssnelheid in het gebruikte kenmerk (DNA-mutatiesnelheid bij DNA-sequenties) kan variëren tussen soorten. De belangrijkste oorzaak van verschillen in taklengte tussen terminale soorten is echter een artefact van de reconstructiemethode, die weer te maken heeft met taxon-sampling. Een typisch voorbeeld hiervan zien we in de figuur hieronder, waarin een evolutionaire stamboom is weergegeven van boleten en hun verwanten (Binder en Hibbett, 2006). Zoals we kunnen zien, lopen de eindtakken bovenin de boom in het algemeen verder naar rechts dan takken onderin de boom.

Dit betekent dat er meer veranderingen zijn gereconstrueerd dan in soorten lager in de boom. Dit is een typisch patroon: de groep die de focus is van een bepaalde studie, en waarvan dus de meeste soorten zijn bemonsterd, gaat verder naar rechts. De verklaring hiervoor is dat er bij soortenarme groepen, of groepen waarvan weinig soorten voor de analyse zijn geselecteerd, meer veranderingen ‘verborgen’ blijven bij reconstructie, omdat een kenmerk ook weer kan terug veranderen naar de oorspronkelijke toestand (terugmutaties). Dus terugmutaties kunnen het aantal opgetreden veranderingen onderschatten en dit speelt een grotere rol bij groepen met weinig (geselecteerde) soorten, zoals meestal het geval is bij buitengroepen.

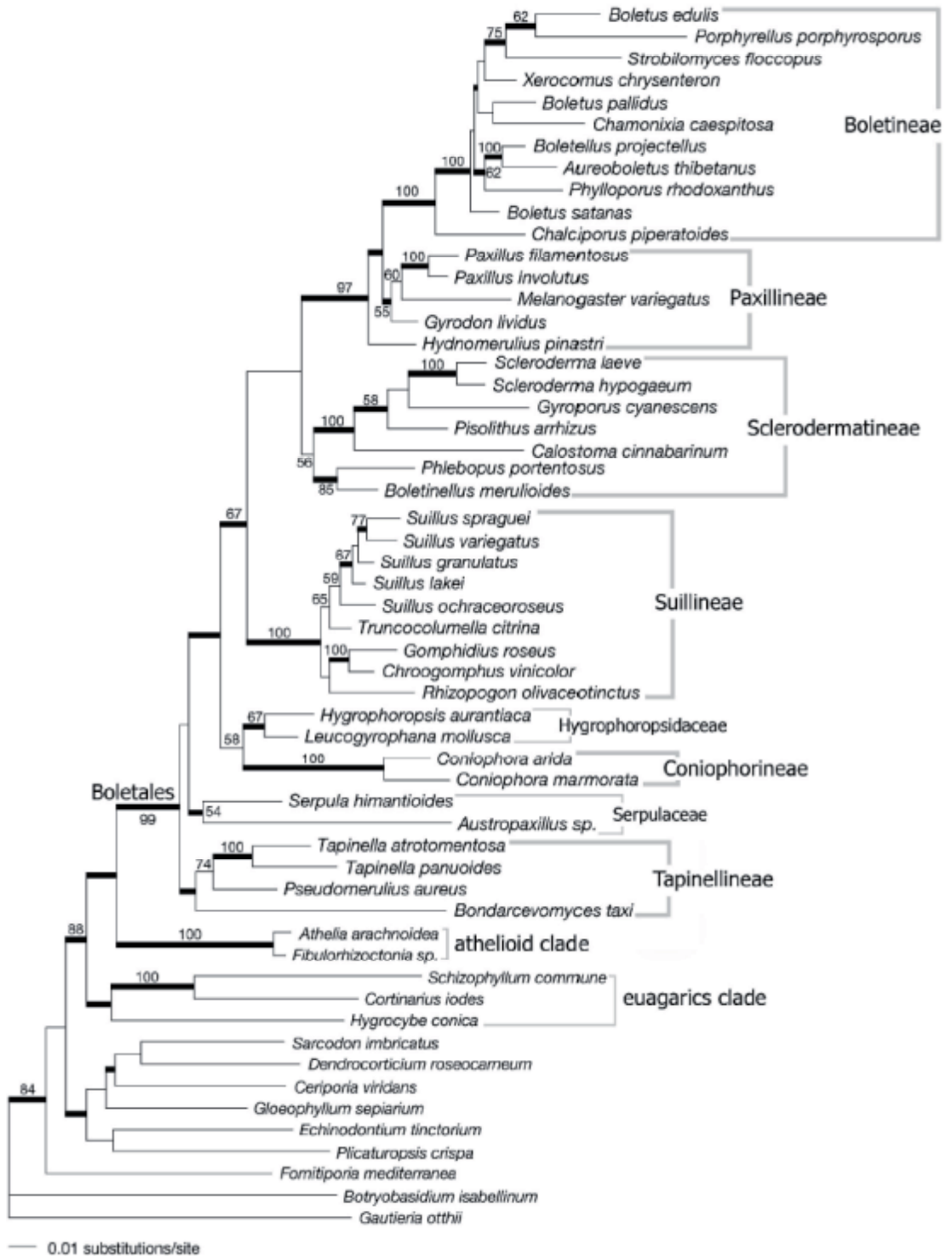
Hoe kunnen we de boletenboom lezen?

Binder en Hibbett (2006) hebben monofyletische groepen onderscheiden en benoemd in de fylogenie van Figuur 5 (rechts; Boletineae, Paxillinieae, Sclerodermatineae, Suillineae, en zo voort; overigens vormt een van deze groepen volgens de fylogenie geen monofyletische groep, ziet u welke?). De getallen op de takken geven bootstrapwaardes van die takken weer; hoe hoger de waarde, hoe waarschijnlijker het is dat die tak correct is. De takken in deze boom zijn met een verschillende dikte getekend, om de takken met een relatief goede ondersteuning gemakkelijk te kunnen onderscheiden van die met een zwakke ondersteuning. Zoals we kunnen zien, zijn vele takken dun en dus niet zeker. Dit geldt vooral voor de verwantschapsrelaties tussen de onderscheiden monofyletische groepen.

Op basis van deze stamboom kunnen we de evolutie van verschillende kenmerktoestanden reconstrueren. Een voorbeeld is de evolutie van buisjes en lamellen in deze groep. Binnen deze groep komen zowel soorten met buisjes of lamellen als soorten met resupinate of ondergrondse vruchtlichamen voor. Buisjes komen zowel in Boletineae, Paxillinieae, Sclerodermatineae als Suillineae voor, maar binnen de Paxillinieae en Sclerodermatineae komen ook soorten met lamellen voor, en voor *Phylloporus* is de kenmerktoestand intermediair tussen lamellen en buisjes. Zoals we kunnen zien vormen de Suillineae en Boletineae samen geen monofyletische groep, omdat de gemeenschappelijke voorouder van deze twee groepen ook de voorouder is van Paxillinieae en Sclerodermatineae. De vier groepen samen zijn wel



Figuur 5. *Fylogenetische verwantschappen tussen boleetachtigen en hun verwanten. De takken bovenin de boom zijn gemiddeld langer dan de takken onder in de boom. Dit is een artefact ten gevolge van verschillen in het aantal soorten per groep.*



weer monofyletisch. Om de evolutie van buisjes en lamellen in deze groep te kunnen verklaren gaat de meest parsimone ('zuinige') hypothese er van uit dat hun gemeenschappelijke voorouder buisjes had. Die zijn dan 2 keer (of 3 keer, afhankelijk van hoe je over *Phylloporus* denkt) geëvolueerd naar lamellen, namelijk in de tak die leidt naar de krulzomen (*Paxillus*, in de Paxillineae) en in de tak die leidt naar de spijkerzwammen (*Gomphidius/Chroogomphus* in de Suillineae). Als je uit zou gaan van een voorouder met lamellen, dan zouden er minimaal 3 (of 4) veranderingen nodig zijn om volgens de boom van Figuur 5 kenmerkt toestanden van de verschillende groepen te kunnen verklaren.

Conclusie

Helaas komen we bij de beschrijving van fylogenieën nog vaak interpretaties tegen die indruisen tegen het fundamentele inzicht dat het leven op aarde een oorsprong heeft en dat alle soorten dus even oud zijn. We kunnen fylogenieën beschrijven door zustergroep-relaties tussen soorten en successievelijke combinaties van soorten te beschrijven. De chimpansee en de bonobo zijn zustersoorten en samen vormen zij de zustergroep van de mens. Chimpansee, bonobo en mens samen vormen de zustergroep van de gorilla. Onterechte kwalificaties als basaal of primitief komen vrijwel altijd voort uit de focus van een studie, met de hoogste aantallen soorten uit de groep waarop de nadruk ligt in een studie.

Hoewel ik heb beargumenteerd dat termen als basaal of afgeleid geen betekenis hebben voor een nu levende soort of groep soorten, kunnen we wel *kenmerkt toestanden* reconstrueren met behulp van fylogenieën. Het benoemen van nu levende vooroudersoorten in de top van bomen, moet daarom vervangen worden door het zoeken naar oorspronkelijke kenmerkt toestanden (Omland et al., 2008). Op die manier is het duidelijk dat *iedere* nu levende soort een mix is van oorspronkelijke en afgeleide kenmerkt toestanden.

Ik dank Nico Dam voor de uitnodiging dit stukje te schrijven en enkele goede suggesties, en Eric Bastiaans voor commentaar op een eerdere versie.

Literatuur

- Aanen, D.K., P. Eggleton, Rouland-Lefèvre, C., Guldborg-Frøsvlev, T., Rosendahl, S. and Boomsma, J.J. 2002. The evolution of fungus-growing termites and their mutualistic fungal symbionts. *Proceedings of the National Academy of Sciences of the USA*, 99: 14887–14992.
- Barbrook, AC, Howe, C.J., Blake, N. & Robinson, P. 1995. The phylogeny of *The Canterbury Tales*. *Nature* 394: 839.
- Binder, M., and Hibbett, D. S. 2006. Molecular systematics and biological diversification of Boletales. *Mycologia* 98: 971–981.
- Haeckel, E. 1879. The evolution of man.
- Hibbett, D.S., Pine, E.M., Langer, E., Langer, G., and Donoghue, M.J. 1997. Evolution of gilled mushrooms and puffballs inferred from ribosomal DNA sequences. *Proceedings of the National Academy of Sciences of the USA* 94: 12002–12006.
- Mueller, U.G., Rehner, S.A. and Schultz, T.R. 1998. The evolution of agriculture in ants- *Science* 281: 2034–2038.
- Omland, K.E., Cook, L. G. and Crisp, M. D. 2008. Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. *BioEssays*, 30: 854–867.
- Ryan, M.J. and Rand, A.S. 1995. Female responses to ancestral advertisement calls in the tungara frog. *Science* 269: 390–392.