

Phylogenetische Systematik

kleinste Untersuchungseinheit in der phylogenetischen Systematik ist die **Art**. Artgrenzen werden definiert durch das **Biologische Artkonzept**: "Species are groups of interbreeding natural populations that are reproductively isolated from other such groups" (E. Mayr 1969, 1975) (= "Arten sind Gruppen sich miteinander kreuzender natürlicher Populationen, die hinsichtlich ihrer Fortpflanzung von anderen derartigen Gruppen isoliert sind" aus Ax 1984)). Ein wesentliches Problem bei dieser Definition ist die fehlende **Artabgrenzung in der Zeit**. Nach Hennig, Ax, Willmann sind für Beginn und Ende der Existenz einer Art folgende Zeitpunkte anzunehmen: Die Lebensspanne neuer Arten beginnt mit der Aufspaltung einer existierenden Art in zwei reproductiv voneinander isolierte Tochterarten (! **dichotome** Verzweigungen !) (Im Verhältnis zueinander sind dies **Schwesterarten**!). Das zeitliche Ende einer Art ist gegeben, entweder durch das Aussterben ohne Folgeart, oder durch die Auflösung in zwei Folgearten. Die sich auflösende Art wird somit zur **Stammart** der beiden **Tochterarten**.

Ausgangspunkt:

Frage:

Hilfsmittel zur Klärung dieser Frage:

Was ist ein Merkmal?

mehr oder weniger große Anzahl von Organismen

wie sind diese untereinander Verwandt

"**Merkmale**"

für Morphologie keine allgemeingültige Definition, generell: "alles was der Bearbeiter / die Bearbeiterin als Merkmal erkennt (z.B. Fehlen / Vorhandensein von Borsten oder sonstigen Strukturen, Verhaltensweisen, Proteinen, ...)

für DNA / RNA - Sequenzen i.d.R.: eine Base (ein Basenpaar) = ein Merkmal

Was sagen diese Merkmale aus?

Man kann Unterschiede und/oder Gemeinsamkeiten zwischen Organismen erkennen (-> rein typologisch!)

Wichtig für phylogenetische Analyse:

(Interpretation der erkannten Merkmale in Hinsicht auf ihre Entwicklung / Evolution)

(-> homologe Merkmale -> Homologiekriterien nach Remane (1962, 1955): 1: Kriterium der Lage, 2. Kriterium der speziellen Qualität, 3. Kriterium der Stetigkeit/Kontinuität !mit Vorsicht handhaben!)

(**Leserichtung**: welches ist der ursprüngliche, welches der abgeleitete Zustand -> **Außengruppenvergleich** s.u.)

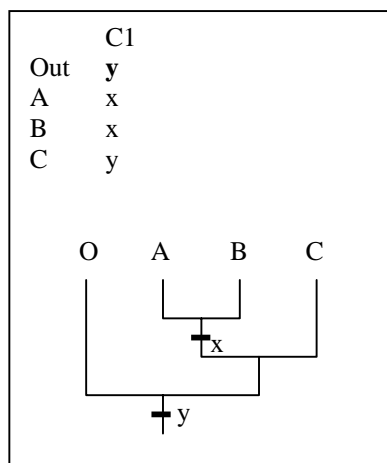
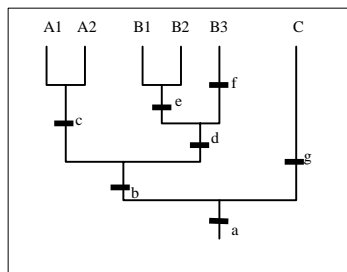
-> gemeinsame Merkmale, genauer solche gemeinsamen Merkmale, die bei dem letzten gemeinsamen Vorfahren entstanden sind und von diesem ererbt wurden (= **Synapomorphien**).

Gemeinsame Merkmale, die älter sind, z.B. vom Vorfahren des letzten gemeinsamen Vorfahren stammen, werden als **Symplesiomorphien** oder einfach Plesiomorphien bezeichnet.

(**Autapomorphien** sind abgeleitete (= neue) Merkmale, die erst in der Stammlinie des betrachteten Taxon (z.B. der Art) entstanden sind. Diese Merkmale helfen uns Mitglieder einer Art (eines Taxon) als ebensolche zu erkennen, liefern aber keine Informationen über Verwandtschaftsbeziehungen.)

Im Idealfall sind solche apomorphen Merkmale / Merkmalsausprägungen (z.B. f, g) einmalig - das würde heißen, das ein Individuum, das ein solches Merkmal besitzt eindeutig als einem bestimmten Taxon zugehörig gekennzeichnet wäre.

Leider ist das sehr oft nicht der Fall, d.h. das gleiche Merkmal, die gleiche Merkmalsausprägung tritt in nicht näher verwandten Gruppen auf -> konvergente (od. analoge) Merkmale. Auch kann es zu Merkmalsumkehr kommen, d.h. ein abgeleitetes Merkmal nimmt in einer späteren Phase der Evolution wieder eine - scheinbar - plesiomorphe Ausprägung an. Diese beiden Fälle, **Konvergenzen** und **Merkmalsumkehr**, werden in der (Computer-) Kladistik oft unter dem Begriff **Homoplasi** zusammengefasst, da sie in der computergestützten Analyse den gleichen (Zähl-) Wert haben.



Es ist hier offensichtlich sehr wichtig zu wissen, welches die ursprünglichen (d.h. die älteren) und welches die abgeleiteten Merkmale sind. Also in welcher Richtung die Evolution der Merkmale vonstatten gegangen ist (Leserichtung). Die Bestimmung der **Leserichtung** (Polarität) kann mit genügend hoher Wahrscheinlichkeit einer zutreffenden Erkennen nur nach einer Methode erfolgen, nämlich mit dem Außengruppenvergleich. Dabei wird zusätzlich zu den eigentlich

interessierenden Taxa eine weitere Gruppe untersucht, die nach Möglichkeit nah mit den zu analysierenden Taxa verwandt ist aber auf gar keinen Fall in eines dieser Taxa gehört oder Schwestergruppe eines dieser Taxa ist. Eine solche Gruppe wird daher als **Außengruppe** bezeichnet, die Taxa für die man die Verwandtschaftsbeziehungen wissen möchte als Innengruppe.

Treten in der Innengruppe für ein Merkmal die Ausprägungen x und y auf so ist allein aufgrund dieser Feststellung nicht entscheidbar welche Ausprägung die plesiomorphe ist. Wird nun in der Außengruppe das gleiche Merkmal untersucht und festgestellt, daß hier die Ausprägung y vorhanden ist, so kann mit hoher Wahrscheinlichkeit davon ausgegangen werden, daß y der plesiomorphe Zustand ist.

Diese Information ist wesentlich für die Rekonstruktion des Stammbaumes. Bei einer Merkmalsverteilung wie im nebenstehenden Beispiel ist nur dann eine Aussage über die Verwandtschaftsverhältnisse in der Innengruppe möglich, wenn y der plesiomorphe Zustand ist. In diesem Fall wird x zur Synapomorphie von A und B. Ist aber x der plesiomorphe Zustand so ist y Autapomorphie von C, was uns keinerlei Information über die Verwandtschaftsverhältnisse liefert.

Wie findet man das zutreffende Verwandtschaftsdiagramm? Drei-Taxon-Problem

- 2 Taxa -> eine Lösung
- 3 Taxa -> drei Lösungen
- 4 Taxa -> 15 Lösungen

Mentale Methode:

Individuen betrachten = untersuchen - Merkmale erkennen - Gemeinsamkeiten (Synapomorphien) zwischen je zwei Taxa finden - Taxa zu Kladogramm arrangieren

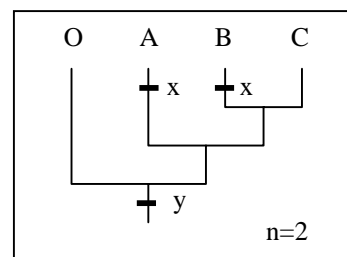
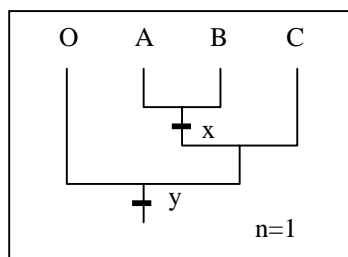
Computerkladistische Methode:

Individuen untersuchen - Merkmale erkennen - Merkmale kodieren und in Datenmatrix eintragen - Kladogramme vom Programm berechnen lassen - Kladogramme analysieren

Welches Kladogramm ist das "bessere"?

Das am häufigsten benutzte Kriterium um das wahrscheinlichste Kladogramm zu finden ist das **Parsimonie-Prinzip**. Danach ist jene Hypothese über die Verwandtschaftsbeziehungen die wahrscheinlichste, die die geringste Zahl an Merkmalsänderungen erfordert.

	C1
Out	y
A	x
B	x
C	y



Wie findet das Programm die sparsamsten Verwandtschaftsdiagramme?

Datenmatrix:

Um die Bearbeitung der Merkmalsdaten mit dem Computer überhaupt zu ermöglichen, müssen die Merkmale zunächst so kodiert werden, daß ein Computerprogramm damit umgehen kann. D.h. es muß eine Matrix (eine Tabelle) erstellt werden, die alle Taxa mit den zu berücksichtigenden Merkmalen enthält. Dabei werden i.d.R. die Taxa in den Zeilen und die Merkmale in den Spalten der Matrix eingetragen. Die Merkmale werden traditionell so kodiert, daß sie in der Matrix nur als 0 / 1 erscheinen. Aktuelle Programme wie PAUP 4.x u.a. sind aber auch in der Lage andere Darstellungen zu verarbeiten (z.B. bei PAUP: alle Zeichen von 0 bis Z). Bei der Analyse von DNA / RNA Sequenzen werden die Basen in der allgemein üblichen Darstellung (A G T C U) in die Datenmatrix eingetragen.

Merkmalskodierung:

Die Umkodierung der Merkmale ist relativ einfach, solange es sich um ja/nein-Aussagen handelt, z.B.: Postorbitale vorhanden - ja = 0 / nein = 1. Es spielt keine Rolle ob z.B. ja als = 0 oder 1 kodiert wird; der Computer weiß nicht was ja/nein bedeutet!

Komplexere Merkmale können natürlich auch so simpel kodiert werden, z.B.: "Flügel vorhanden - 0 / 1" oder "Vorderextremitäten zu Flügeln umgebildet - 0 / 1". Hier besteht aber die große Gefahr, daß Strukturen vom Computer als Synapomorphien gedeutet werden, die nicht einmal homolog sind (Vögel <-> Pterosaurier). Derart komplexe Merkmale werden i.d.R. in mehrere Einzelmerkmale zerlegt. Bei den Flügeln der Vertebraten könnte man z.B. die Ausbildung der einzelnen Knochen genauer darstellen (z.B.: Vorhandensein einzelner Knochen, Verschmelzungen zwischen zwei Knochen, etc.). Auf diese Art ist es möglich auch sehr komplexe Strukturen in reiner 0/1-Darstellung wiederzugeben. Allerdings erzeugt man so schnell große Matrizen mit sehr vielen Einzelmerkmalen (-> Rechenzeit).

Gewichtung etc.:

Eine andere Möglichkeit besteht darin, mehr als nur zwei Merkmalszustände zu benutzen, z.B.: 0 = Laufextremität, 1 = Flügel mit einem stark verlängerten Finger, 2 = Flügel mit drei leicht verlängerten Fingern, 3 = Flügel mit vier deutlich verlängerten Fingern, 4 = Greifhand, nicht/selten zum Laufen eingesetzt, 5 = fehlt. Auf diese Weise kann die Gleiche Informationsmenge Dargestellt werden wie durch die Auflösung in Einzelmerkmale, die Matrix wird aber nicht in gleichem Maße aufgebläht. Wie die Merkmale beim Anpassen auf ein Kladogramm behandelt werden, läßt sich durch eine Reihe von Einstellungen beeinflussen: In der Grundeinstellung zählt eine Veränderung in einem Merkmal immer als ein Schritt (= Gewichtung 1). Dies kann man ändern, indem man Merkmale unterschiedlich gewichtet. Als Grund für solche Gewichtungen wird i.d.R. die Komplexität einzelner Merkmale angeführt. Basis für diese Begründung ist die Annahme, daß "einfache" Strukturen mit größerer Wahrscheinlichkeit mehrfach unabhängig (konvergent) entstehen als kompliziert gebaute Strukturen (mit Vorsicht einsetzen!). So könnte man das Merkmal "Existent von Flügeln" mit der Gewichtung 100 belegen. Eine Veränderung in diesem Merkmal würde dann als 100 Schritte gezählt.

ordnen:

Normalerweise zählen Veränderungen in Merkmalen in jeder Richtung gleich viel, d.h. eine Änderung von 0 nach 1 wird genauso als ein Schritt gezählt wie eine Änderung von 1 nach 0. Bei Merkmalen mit mehreren Zuständen zählen die Veränderungen 0->3, 2->1, 0->1 alle gleich als jeweils ein Schritt (ungeordnet = **Fitch-Parsimonie** (Fitch 1971)).

Auch hier kann man Einfluß nehmen, indem man Merkmale als "geordnet" definiert (**Wagner-Parsimonie** (Kluge & Farris 1969, Farris 1970)). Bei einem "geordneten" Merkmal würde eine Veränderung von 0 nach 3 drei Schritte zählen, von 1 nach 3 zwei Schritte usw.. Geordnete Merkmale machen dann Sinn, wenn man eindeutige Hinweise darauf hat (z.B. aus dem Fossilbericht), daß die Evolution in einer bestimmten Richtung abgelaufen ist.

Weitere Varianten der Merkmalsinterpretation sind die **Dollo-Parsimonie** (Farris 1977) und die **Camin-Sokal-Parsimonie**. Bei der Dollo-Parsimonie kann in einem Merkmal zwar einmal eine Merkmalsumkehr auftreten, danach ist es aber nicht mehr möglich, daß der abgeleitete Zustand erneut auftritt. Dies hat zur Folge, daß Synapomorphien in einem Kladogramm nur einmal auftreten können, d.h. das gleiche Merkmal kann nicht zweimal parallel entstehen.

Bei der **Camin-Sokal-Parsimonie** sind Veränderungen generell nur in eine Richtung zulässig, d.h. es gibt keine Merkmalsumkehr (selten benutzt).

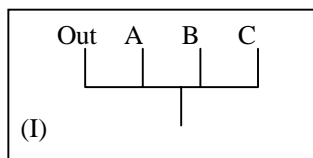
Ermittlung des sparsamsten Kladogramms: Es gibt im wesentlichen zwei Verfahrensweisen (Algorithmen) um ausgehend von einer Datenmatrix die sparsamsten Kladogramme zu finden.

1. Hennig-Argumentation:

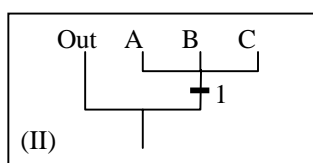
Wie der Name schon ahnen läßt wurde diese Methode von Hennig vorgeschlagen, was gleichzeitig auch bedeutet, daß sie nicht explizit für Computeranwendungen erdacht wurde sondern allgemein zur Ermittlung von Verwandtschaftsverhältnissen.

Man geht so vor, daß die in der Datenmatrix gegebenen Informationen Merkmal für Merkmal abgearbeitet werden und dabei der Stammbaum jeweils dem Kenntnisstand angepaßt wird.

Die Vorgehensweise soll an der Matrix 1 verdeutlicht werden:



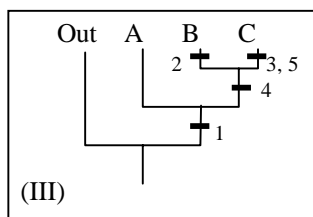
- Ausgangspunkt ist ein sog. "Busch", d.h. es werden alle Taxa in einer Polytomie zusammengefaßt (I).



- Nimmt man das Merkmal Nr. 1 hinzu, so ist es möglich die Taxa A, B und C zusammen zu gruppieren (II).

- Merkmal 2 ist eine Autapomorphie von Taxon B und liefert somit keine Informationen über Verwandtschaftsverhältnisse (III).

- Merkmal 3 ist eine Autapomorphie von Taxon C (III).



- Merkmal 4 tritt im Zustand 1 nur bei den Taxa B und C auf und kann daher als Synapomorphie dieser beiden Taxa gedeutet werden (III).

- Merkmal 5 ist Autapomorphie von Taxon C (III).

Bei realen Datensätzen treten sehr oft Konflikte zwischen zwei oder mehr Merkmalen auf, die die Rekonstruktion erschweren. Matrix 2 kommt der realen Situation etwas näher und ermöglicht es, das Hennig-Verfahren an einem etwas komplexeren Beispiel zu üben. (In dem Ergebnis-Kladogramm treten 11 Merkmalsänderungen auf!)

Matrix 1:

Taxa\Chars.	1	2	3	4	5
Outgroup	0	0	0	0	0
A	1	0	0	0	0
B	1	1	0	1	0
C	1	0	1	1	1

Matrix 2:

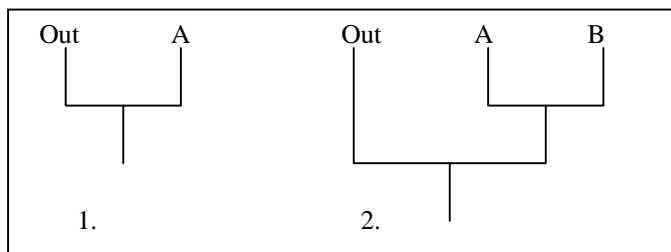
Taxa\Chars.	1	2	3	4	5	6	7	8	9	10
Outgroup	0	0	0	0	0	0	0	0	0	0
Alpha	0	0	0	1	1	1	1	0	1	0
Beta	0	0	0	1	0	1	1	0	1	0
Gamma	0	0	0	0	0	0	0	0	1	0
Delta	1	1	1	0	0	0	1	1	1	0
Epsilon	1	1	1	0	0	0	1	0	1	1
Zeta	1	1	0	0	0	0	1	0	1	1
Theta	1	0	0	0	0	0	1	0	1	1

2. Wagner-Verfahren:

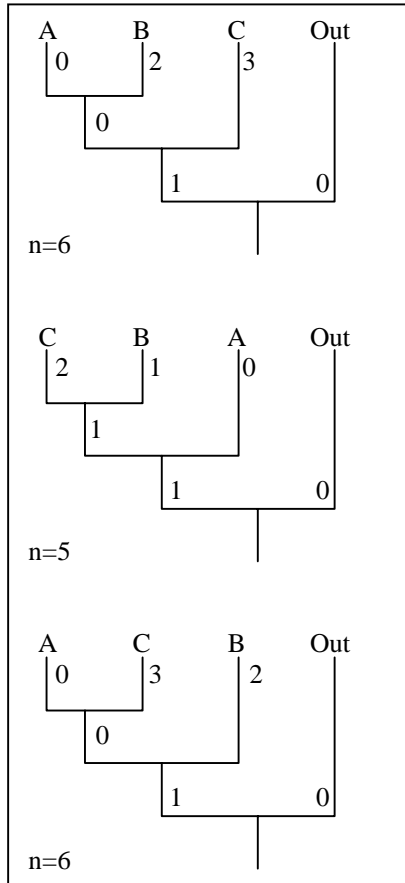
Bei dieser Methode werden die einzelnen Taxa nach und nach dem Kladogramm hinzugefügt. Dies geschieht so, daß die Anzahl der Merkmalsänderungen auf dem resultierenden Kladogramm minimal gehalten wird.

Als Beispiel dient wieder Matrix 1. Um die Reihenfolge zu bestimmen in der die Taxa zum Kladogramm hinzugefügt werden, ermittelt man zunächst die Anzahl der abgeleiteten Merkmale für jedes Taxon (#abg).

Matrix 1:						
Taxa\Chars.	1	2	3	4	5	#abg
Outgroup	0	0	0	0	0	0
A	1	0	0	0	0	1
B	1	1	0	1	0	3
C	1	0	1	1	1	4



Knotenpunkte					
AxB	1	0	0	0	0
AxC	1	0	0	0	0
BxC	1	0	0	1	0
(AB)xC	1	0	0	0	0



	Out	A	B	C	AxB	AxC	BxC	AxBxC
Out	-	1	3	4	1	1	2	1
A		-	2	3	0	0	1	0
B			-	1	2	2	1	2
C				-	3	3	2	3
AxB					-	-	-	0
AxC						-	-	0
BxC							-	1

-Als erstes wird nun das Taxon mit der geringsten Zahl an abgeleiteten Merkmalen (A) mit der Außengruppe verbunden.

- Es folgt das Taxon mit der nächsthöheren Zahl abgeleiteter Merkmale, also B. Es wird mit A verbunden und am Knotenpunkt von A und B wird die am weitesten abgeleitete Merkmalsausprägung notiert, die diese beiden Taxa gemeinsam haben. Im Beispiel also 10000. Dies kann so interpretiert werden, daß, wenn A und B Schwestertaxa sind, die Stammart von A & B die Merkmalsausprägung 10000 besessen hat.

- Als nächstes wird nun wieder das Taxon mit der nächsthöheren Zahl abgeleiteter Merkmale (C) in das Kladogramm eingefügt. Dies kann an drei verschiedenen Positionen geschehen: 1. als Schwestertaxon zum Taxon A+B, 2. als Schwestertaxon zu B oder 3. als Schwestertaxon zu A. Taxon C muß so eingefügt werden, daß nur eine minimale Zahl von Merkmalsänderungen erforderlich ist. Die entsprechende Stelle im Kladogramm wird durch Differenzbildung (XOR der Bool'schen Arithmetik) zwischen den Merkmalen des einzufügenden und des benachbarten Taxon ermittelt:

C 10111
AxB 10000
Differenz 00111 -> d.h. würde C als Schwesterguppe zu A in das Kladogramm eingefügt, so würden 3 weitere Merkmalsänderungen zur "Länge" des Baumes hinzukommen.

C 10111
B 11010
Differenz 01101 -> 3 Änderungen.

C besitzt den Merkmalsatz 10111
A 10000
Differenz 00111-> 3 Änderungen

C 10111
Out 00000
Differenz 11111 -> 5 Änderungen
D.h. es stehen drei Verknüpfungspunkte mit gleicher Wertigkeit zur Verfügung. Man muß also alle drei Varianten ausprobieren und jedesmal die Gesamtlänge des Baumes, d.h. die Gesamtzahl der

Merkmalsänderungen ermitteln. Die "Schrittlängen" für jeden Ast der drei möglichen Kladogramme sind in den nebenstehenden Diagrammen angegeben. Danach ist das mittlere Kladogramm mit einer Länge von 5 Schritten, d.h. mit 5 Merkmalsänderungen, zu bevorzugen. Der eine zusätzliche Schritt in den beiden anderen Bäumen kommt dadurch zustande, daß hier für Merkmal 4 jeweils eine zweimalige unabhängige Änderung von 0 nach 1 angenommen werden muß.

Auch dieses Verfahren sollte an der Komplexeren Matrix 2 nocheinmal geübt werden.

Diese beiden Verfahren bilden die Basis für alle zur Zeit verfügbaren Computerprogramme. Beide Verfahren liefern ein exaktes Ergebnis. D.h. die resultierenden kürzesten Bäume sind mit Sicherheit die, nach dem Parsimonie-Prinzip kürzesten möglichen Lösungen für die Ausgangsmatrix.

Tabelle 1	
Taxa	# dichot. Klad.
3	3
4	15
5	105
6	945
7	10395
8	135135
9	2027025
10	34459425
20	8,2·10 ²¹

Eine weitere Möglichkeit ein exaktes Ergebnis zu erzielen ist alle für die in der Matrix vorgegebene Anzahl von Taxa möglichen Bäume zu konstruieren und für jeden dieser Bäume die Länge zu berechnen. Diese Methode ist aber nur bei relativ kleinen Datensätzen sinnvoll anwendbar. Das liegt daran, daß die Anzahl möglicher dichotomer Kladogramme überproportional mit der Anzahl der Taxa in der Datenmatrix zunimmt. So sind bei 3 Taxa nur 3 Kladogramme zu untersuchen, bei 4 Taxa 15 Kladogramme und bei 7 Taxa sind es schon 10395 mögliche dichotome

Kladogramme (siehe Tabelle). Dementsprechend nimmt natürlich auch die erforderliche Rechenzeit zu. Bei PAUP 4.x kann zudem für Matrizen mit mehr als 12 Taxa keine vollständige Suche (Befehl: "AllTrees") durchgeführt werden, da ab 13 Taxa mehr Bäume zu verwalten sind als in einer 32 bit integer Zahl gezählt werden können (= 4.294.967.295). Bei komplexen Matrizen, die eine größere Zahl von Widersprüchen (Homoplasien) enthalten können auch bei den heutigen Computern schon bei 10 Taxa Rechenzeiten von bis zu einer Stunde (oder mehr) entstehen.

Die beiden oben gezeigten Verfahren sind hier schneller, da sie nur die Kladogramme untersuchen die sozusagen auf dem Weg zum kürzesten Kladogramm liegen. Mit diesen Methoden können auch deutlich größere Matrizen analysiert werden, da die 32 bit Grenze hier kaum je erreicht werden wird (und wenn, dann ist die Datenmatrix sehr wahrscheinlich unbrauchbar). In PAUP ist mit "BandB" (= branch and bound) ein Verfahren verfügbar, daß sich an das Wagner-Verfahren anlehnt und für mittelgroße Datensätze relativ schnell Ergebnisse liefert.

Für größere und/oder komplexere Datensätze werden i.d.R. sogenannte **heuristische Suchverfahren** angewandt (bei PAUP "Hsearch"). Dabei wird zunächst ein Baum konstruiert, i.d.R. mit einem Wagner-Verfahren. Dieser Baum wird anschließend modifiziert und für jede Variante wird die Länge bestimmt. Dieses Verfahren wird solange wiederholt bis keine der Modifikationen des jeweiligen Ausgangsbaumes kürzer ist als dieser. Das Modifikationsverfahren wird als "branch-swapping" bezeichnet. Es existieren drei Varianten: "nearest neighbour interchange" (NNI), "subtree pruning and regrafting" (SPR) und "tree bisection and reconnection" (TBR). TBR ist die Variante, die die meisten möglichen Stamm bäume abdeckt. Somit ist die Wahrscheinlichkeit die absolut kürzesten Bäume zu finden bei dieser Variante am besten. SPR kann als Teilmenge der von TBR ausgeführten Operationen betrachtet werden und NNI ist dann Teilmenge von SPR. Beide Algorithmen sind schneller als TBR aber die Gefahr den oder die kürzesten Bäume nicht zu finden ist deutlich größer.

Inseln:

Ein zentrales Problem bei den heuristischen Suchverfahren sind lokale Minima. Es kann vorkommen, daß das Programm ein Kladogramm erreicht, das im Vergleich zu allen Kladogrammen, die mit dem nächsten Berechnungsschritt erreichbar sind, die geringste Länge besitzt, obwohl es in der Menge aller möglichen Kladogramme noch sparsamere gibt. Beim "normalen" Branch-Swapping ist es dem Programm nicht möglich den absolut kürzesten Baum zu finden, da die Algorithmen ihre Arbeit beenden, sobald mit den ihnen möglichen Veränderungen alle neuen Bäume länger sind als der Ausgangsbaum.

Boote:

Dieses Problem kann man auf zweierlei Wegen umgehen: Zum einen besteht die Möglichkeit die Reihenfolge in der die Taxa bei der Erstellung des Anfangsbaumes hinzugefügt werden zufällig zu variieren. Dies hat zur Folge, daß man sich dem absoluten Minimum auf unterschiedlichen Wegen nähert und so den lokalen Minima ausweichen kann. Außerdem kann man das Programm veranlassen auch nicht minimale Bäume in die Suche einzubeziehen, so daß relativ "flache" lokale Minima noch verlassen werden können.

Viele Kladogramme

Es kann vorkommen, daß man als Ergebnis der Analyse einer Datenmatrix sehr viele gleichlange Bäume erhält. Um sich schnell einen Überblick die Auflösung (Qualität) dieser Bäume zu verschaffen besteht die Möglichkeit Konsens-Bäume berechnen zu lassen. Zwei Methoden werden dabei häufig genutzt:

"Strict consensus" - in den Konsens-Baum werden nur die Verzweigungen übernommen, die in allen Ergebnisbäumen auftreten.

"Majority rule" - im Konsens-Baum treten die Verzweigungen auf die in einem bestimmten Prozentsatz der Ergebnisbäume vorhanden sind. Als Standard werden hier i.d.R. 50% angenommen, dieser Wert kann aber verändert werden.

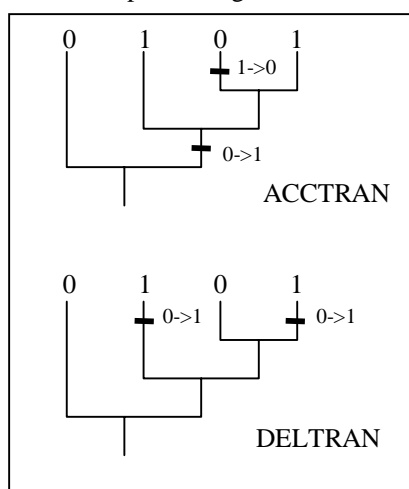
Topologie

Läßt man PAUP mit den Standardeinstellungen eine Matrix analysieren, so erhält man als Ergebnis zunächst keine Kladogramme sondern Topologie. In der Darstellung am Bildschirm unterscheiden sich Topologie und Kladogramm leider nur minimal (-> TreeView), inhaltlich ist der Unterschied aber gravierend. Anders als bei unserer "Von-Hand-Analyse" einer Datenmatrix, bei der die Außengruppe als Ausgangspunkt genommen wird und darüber auch gleich die Polarität der Merkmale bestimmt ist, wird vom Programm die Leserichtung zunächst ignoriert. Das hat zur Folge, daß die resultierende Topologie nur darüber Auskunft gibt, welche Taxa sich besonders ähnlich sind (rein typologisch). Die Topologie ist nur eine Möglichkeit die gegebenen Taxa auf möglichst kurzem Wege dichotom miteinander zu verbinden. Sie enthält keine Informationen über den Ablauf der Evolution.

Kladogramm

Ein Kladogramm entsteht aus der Topologie erst, wenn dem Programm die Polarität der Merkmale bekanntgegeben wird. Dies geschieht entweder über eine Außengruppe oder durch Angabe eines hypothetischen Vorfahren. Mit Hilfe dieser Außengruppe/des Vorfahren wird die Topologie gewurzelt, d.h. es wird die Richtung der Entwicklung festgelegt.

Merkmalsoptimierung



Für die Entwicklung der Merkmale in einem gegebenen Kladogramm gibt es zwei Optimierungsvarianten:

ACCTRAN = ACCelerated character TRANSformation: bevorzugt Merkmalsumkehr (reversals). Eine Merkmalsänderung geschieht so früh, d.h. so nahe an der Wurzel wie möglich.

DELTRAN = DELayed character TRANSformation: bevorzugt konvergente (mehrfach unabhängige) Entwicklung von Merkmalen. Merkmalsänderungen geschehen sehr spät, d.h. so weit entfernt von der Wurzel wie möglich.

Die Länge des Kladogramms wird durch diese Verschiebung des Zeitpunktes der Merkmalsänderung nicht beeinflusst!

Wie "gut" ist ein Kladogramm?

Zur Bewertung von Kladogrammen und zum Vergleich verschiedener Kladogramme sind mehrere Indices in Gebrauch. Der erste, der eingeführt wurde ist der "consistency index" (ci, CI) (Kluge & Farris 1969). Er gibt wieder wie gut einzelne Merkmale oder eine ganze Datenmatrix auf ein gegebenes Kladogramm "paßt". Je höher der Wert ist (max. 1,0) um so besser ist die Passung, d.h. um so weniger Homoplasien treten auf. Berechnet der ci für ein einzelnes Merkmal nach der Formel:

Consistency index

ci

$$ci = \frac{m}{s} \text{ mit}$$

m = minimale Anzahl von Änderungen des Merkmals und
s = Anzahl der Änderungen des Merkmals auf dem betreffenden Kladogramm.

Der Gesamt-CI (CI) für ein Kladogramm ist die Summe der ci's aller Merkmale.

CI

$$CI = \frac{\sum_{i=1}^n ci_i}{n}$$

Rescaled consistency index

Der consistency index ist nicht in der Lage beim Auftreten von Homoplasien wiederzugeben, ob diese überhaupt einen Beitrag zur Bestimmung der Topologie des Kladogramms leisten oder nicht (sie tun dies nicht, wenn alle Merkmalsänderungen in Stammlinien terminaler Taxa auftreten - Autapomorphien). Um diesen Nachteil auszugleichen wurde von Farris (1989) der rescaled consistency index (rc, RC) eingeführt. Der rc berechnet sich als Produkt des ci und des retention index (r, R). Der

retention index wird berechnet aus m, s, und g, wobei g die minimale Anzahl von Veränderungen ist, die ein Merkmal im "schlechtesten" Kladogramm erfahren kann. Dieses "schlechtesten" Kladogramm ist immer ein Busch, d.h. eine Polytomie der alle Taxa der Matrix entspringen.

Retention index

$$r = \frac{g - s}{g - m} \qquad R = \frac{\sum_{i=1}^n r_i}{n}$$

$$rc = ci \cdot r \qquad RC = \frac{\sum_{i=1}^n rc_i}{n}$$

Retention index (R) und rescaled consistency index (RC) für das gesamte Kladogramm werden wiederum als Summe der Einzelwerte aller Merkmale gebildet.

Bootstrap & Jack-knife

Um die Stabilität der Verzweigungspunkte in den resultierenden Kladogrammen zu überprüfen sind zwei Algorithmen in Gebrauch, die die Datenmatrix zufällig modifizieren und mit dieser modifizierten Matrix Kladogramme berechnen. Diese Prozedur wird wenigstens 100 Mal wiederholt. Es gelten dann die Verzweigungen als am besten unterstützt, die in möglichst vielen Neuberechnungen auftreten.

Beim **bootstrap**-Verfahren wird durch zufällige Auswahl aus den Taxa der Originalmatrix bei jedem Durchgang eine neue Matrix von gleicher Größe erstellt, mit der dann eine Baumsuche durchgeführt wird.

Beim **jack-knife**-Verfahren werden die Matrizen durch zufällige Auswahl von Merkmalen aus der Originalmatrix aufgebaut. Die Größe der Matrix wird dabei i.d.R. um 50% reduziert, d.h. bei jedem Durchgang wird nur die Hälfte der Merkmale berücksichtigt.