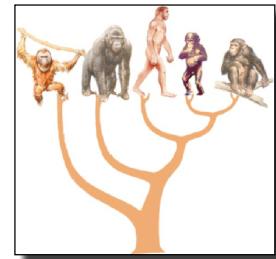


Molekulare Phylogenie



- Grundbegriffe
- Methoden der Stammbaum-Rekonstruktion

Thomas Hankeln, Institut für Molekulargenetik

SS 2010

JOHANNES
GUTENBERG
UNIVERSITÄT
MAINZ

Grundlagen der molekularen Phylogenie

- Evolution äußert sich durch **Veränderungen** auf Ebene von Nukleotid- oder Proteinsequenzen
- Die Veränderungen sind graduell und quantitativ/qualitativ **erfaßbar**
- heute existierende, verwandte Spezies leiten sich von einem **gemeinsamen Vorfahren** ab
- die Entstehung von Spezies ist ein **hierarchischer Prozeß**
- es existiert nur eine historisch korrekte, „**tatsächliche**“ **Phylogenie**

Was leistet die molekulare Phylogenie?

- **Verwandtschaft von Organismen**
(molekulare Systematik, Forensik)
 - **Verwandtschaft zwischen Genen/Proteinen**
(Genomevolution, Gen/Proteinfunktion)
 - **Wie haben sich Lebewesen ausgebreitet?**
(Anthropologie, Ökologie, Epidemiologie)

Unterschiedliche Datensätze für verschiedene Fragestellungen

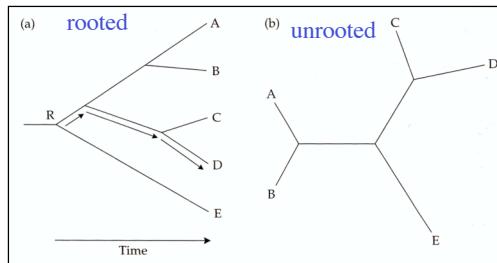
- auf Proteinebene
 - Aminosäuresequenzen
 - Allozyme
 - auf DNA-Ebene
 - **DNA-Sequenzen**
 - polymorphe Marker:
(Mikro)satelliten, SNPs
RFLPs, AFLPs, RAPDs

Molekulare Daten erleichtern die Phylogenie

- Sequenzen sind direkt vererbt; **keine Umwelteinflüsse**
- Sequenzdaten sind in großer Menge relativ **kostengünstig** und schnell zu erhalten (Dank sei der PCR!!!)
- weitgehend **frei von Interpretationseinflüssen** („reduziert“, „etwas abgeflacht“, „dünn“ etc.)
- Sequenzen **evolvieren** insgesamt **gleichförmiger** als morphologische oder physiologische Charaktere
- Sequenzen erlauben **Vergleiche über große Distanzen** (Tiere, Pilze, Pflanzen)
- „sophisticated“ **Modelle** zur mathematisch/statistischen Behandlung der Sequenzevolution existieren

Dennoch: auch molekulare Daten können zu falschen Stammbäumen führen!

rooted vs. unrooted trees



- ein verwurzelter Baum macht Annahmen über die Richtung und den historischen Verlauf der Evolution
- ein unverwurzelter Baum gibt nur Auskunft über Verwandschaft zwischen OTUs
- für das Setzen einer Wurzel benötigt man eine **Outgroup**

rooted vs. unrooted trees

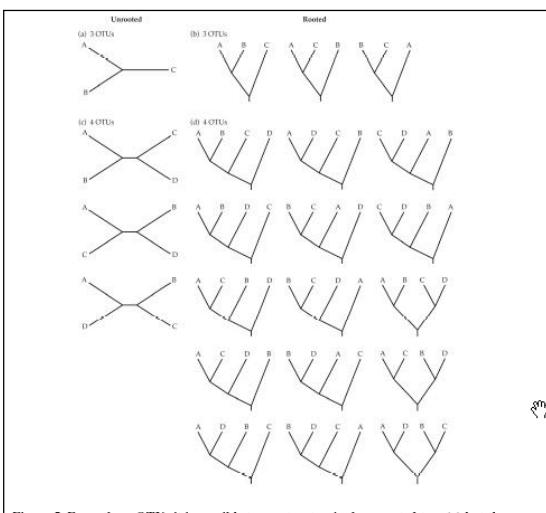


Figure 5. From three OTUs it is possible to construct a single unrooted tree (a) but three different rooted ones (b). From four OTUs it is possible to construct three unrooted trees (c) and fifteen rooted ones (d).

TABLE 5.1 Possible numbers of rooted and unrooted trees up to 10 OTUs

Number of OTUs	Number of rooted trees	Number of unrooted trees
2	1	1
3	3	1
4	15	3
5	105	15
6	954	105
7	10,395	954
8	135,135	10,395
9	2,027,025	135,135
10	34,459,425	2,027,025

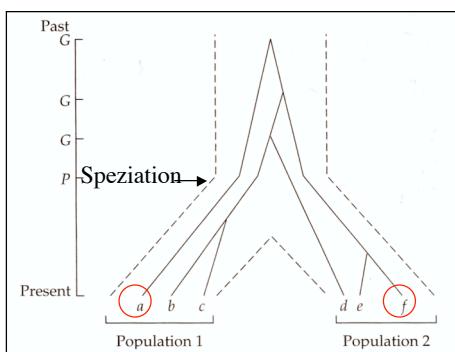
From Felsenstein (1978).

- es ist viel leichter und sicherer, einen unverzweigten Baum zu erstellen:

d. h. nur dann „rooten“, wenn die Outgroup klar definiert ist

gene tree vs. species tree

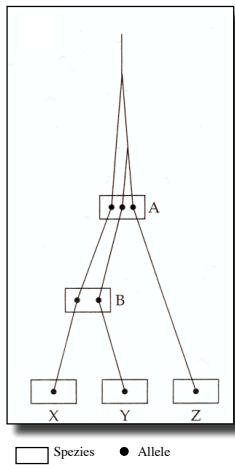
- die Divergenz zweier Gen-Allele, die per Zufall aus zwei verschiedenen Arten isoliert und verglichen wurden, kann größer sein als die Divergenz der beiden Spezies!



➤ führt zur Überschätzung des Abstandes der beiden Arten

➤ spielt vor allem eine Rolle bei geringem phylogenetischen Abstand der beiden Spezies

gene tree vs. species tree



Bei Vergleich der Gene scheinen Y u. Z Schwestertaxa zu sein.

In Wahrheit hatten die Taxa X u. Y einen gemeinsamen Vorfahren!

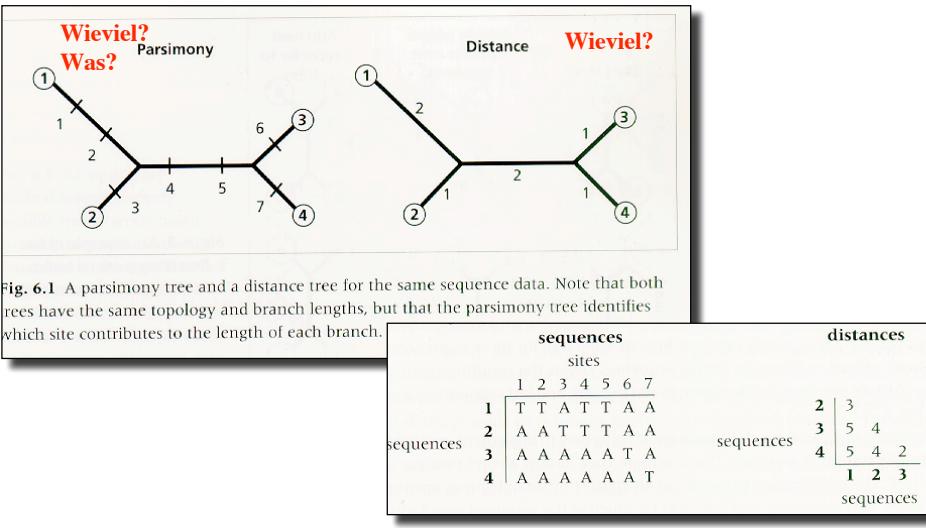
Gen-Baum kann sogar eine Topologie anzeigen, die nicht der Speziationssfolge entspricht!

Rekonstruktion von Stammbäumen aus molekularen Daten

- **Bedingung:** Homologie der Sequenzen und sinnvolles Alignment
- **Problem:** aus der Vielzahl möglicher Baum-Topologien muss das „wahrscheinlichste Szenario“ ausgewählt werden

Welche Form von Daten?
Welche Methoden?

Character state- vs. Distanz-Methoden



Zwei Kategorien von molekularen Daten

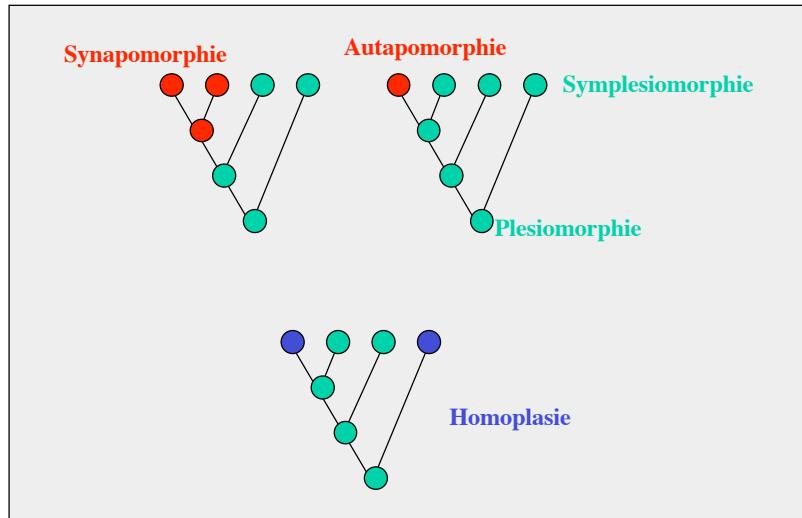
1. „Character“

geben Auskunft über eine Eigenschaft einer OTU

2. „Distanzen“

geben quantitative Auskunft über Ähnlichkeit zwischen zwei OTUs

Character & character state



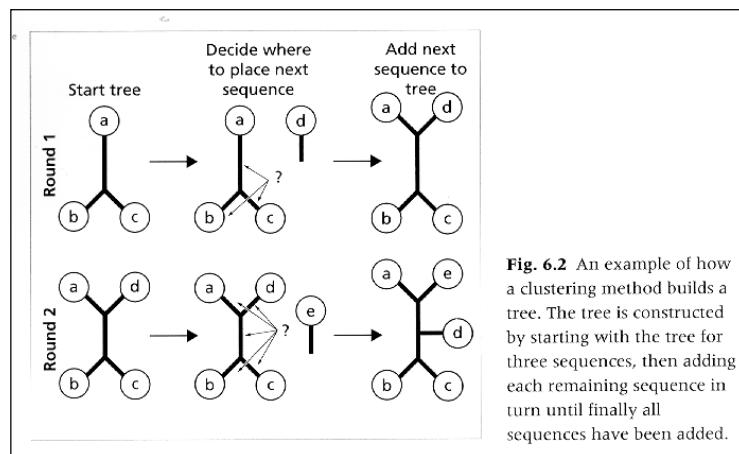
Distanz-Daten

- Distanzen treffen Aussage über zwei OTUs
- DNA/Protein-Sequenzen (character) können in Distanzdaten umgewandelt werden (zB: „94% Sequenzidentität“)
- Distanzwerte sind „einfacher erfassbar“ und liefern oft intuitiv eine Abschätzung der phylogenetischen Verwandschaft
- **Matrizes** (PAM, BLOSUM etc.) ermöglichen sinnvolle Distanzberechnung
- **Korrektur für multiple Austausche** ist absolut erforderlich!
- Kritik: Distanzwerte „verschenken“ einen Teil der phylogenetischen **Information**, nämlich den „character state“
- Distanzen nennt man „ultrametrisch“, wenn alle OTUs mit gleicher Geschwindigkeit evolvieren und damit im Baum gleichweit von Wurzel entfernt sind (= molecular clock)

Zwei methodische Ansätze der Stammbaum-Rekonstruktion

- ...ein **Clustering-Algorithmus**
konstruiert auf Grundlage eines Optimalitätskriteriums sukzessive einen Baum aus den Daten
- ...eine **Suchstrategie**
konstruiert viele, u. U. sogar alle möglichen Bäume, berechnet deren Optimalitätsgrad und sucht den oder die „besten“ heraus

Clustering-Algorithmen



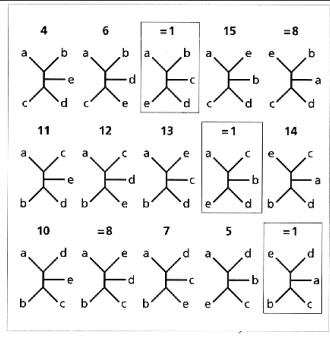
Schrittweiser Aufbau des Baumes anhand von Datenmatrix

Clustering-Algorithmen

- schnelle, einfache Programme (+)
- ein Baum (+)
- nur ein Baum (-)!!
Was, wenn zwei Bäume die Daten gleich gut erklären?
- Ergebnis hängt manchmal von der Reihenfolge der Input-Sequenzen ab (-)

Suche optimaler Bäume

Fig. 6.3 An optimality method assigns to every possible tree a 'score' based on some measure of how the data relate to the tree. Based on this score each tree is ranked in order and typically the best tree (or trees) is used as the estimate of the phylogeny. In this example, there are three optimal trees (indicated by the ranking of first equal [=1]).



- Probleme:
Welches Optimalitätskriterium ist für den Datensatz passend?
Welcher der uU Mio. von Bäumen ist der optimale?
- Auswahl zwischen mehreren Bäumen (+)
 - sehr aufwändiges Rechnen (-)
„NP-vollständig“
 - nicht gut geeignet für > 20 Sequenzen (sonst nur „heuristische“ Suche)

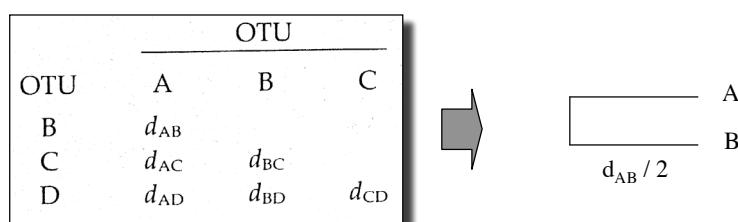
Methoden - Übersicht

		Datentyp	
		Distanzen	Character
		UPGMA	
Rekonstruktionsmethode		Neighbor joining	
		Minimum Evolution	Maximum Parsimony
			Maximum Likelihood
			Bayes
Clustering-Algorithmus			
Such-Strategie			

Distanzmethode A: UPGMA /1

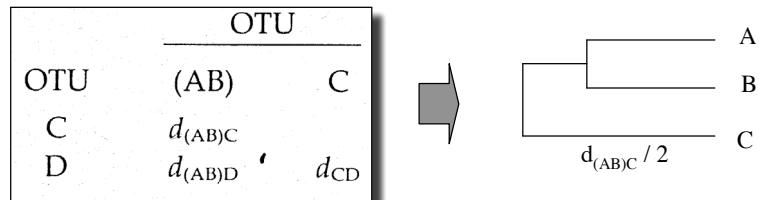
„unweighted pair group method using arithmetic mean“ Sokal and Michener 1958

• **Achtung: setzt konstante Evolutionsrate (molecular clock) voraus!!!!**



1. Identifiziere in Matrix die 2 ähnlichsten OTUs
2. Wenn z.B. d_{AB} am kleinsten ist: betrachte OTUs A und B als eine zusammengesetzte OTU. Der Verzweigungspunkt liegt bei $d_{AB} / 2$

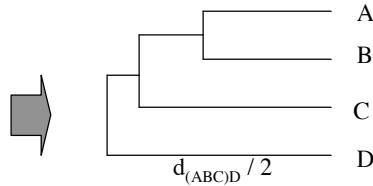
UPGMA /2



3. Erstelle neue Matrix mit (AB) als einer OTU. $d_{(AB)C} = (d_{AC} + d_{BC}) / 2$

4. Wenn z.B. $d_{(AB)C}$ am kleinsten ist: ordne C der OTU (AB) zu!
Astlänge zu C ist $d_{(AB)C} / 2$

5. Ordne OTU D zu. Die Astlänge zu D ist $d_{(ABC)D} / 2 = ((d_{AD} + d_{BD} + d_{CD}) / 3) / 2$



UPGMA kann eine völlig falsche Topologie liefern!!

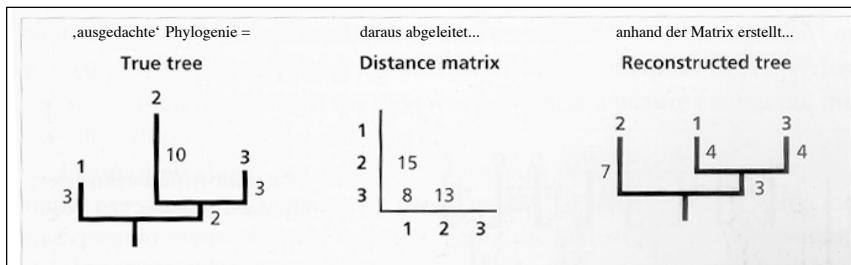


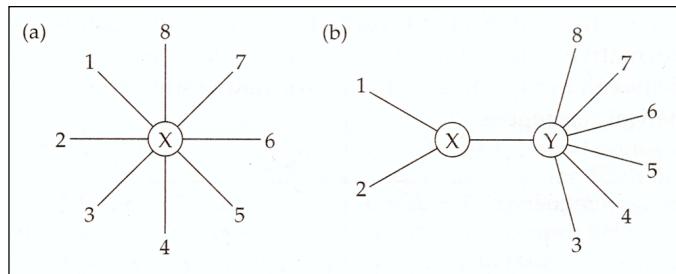
Fig. 6.27 An example where UPGMA will reconstruct the wrong tree. The edge lengths on the true tree violate the condition shown in Fig. 6.26, as $a + b = 3 + 2 < \max(c, d) = 10$. Sequence 2 has evolved more rapidly than the other two sequences, so that sequence 1 and 3 are more similar to each other than either is to 2.

Bessere Distanzmethode:

Saitou, Nei 1987

Neighbor-Joining (NJ)

- besser als UPGMA, da unterschiedliche Evolutionsraten berücksichtigt werden!!
- effizienter Clustering-Algorithmus, jedoch heuristisch



• Prinzip:

- > Suche nach dem Baum mit der kleinsten Summe an Astlängen („Minimum Evolution tree“)
- > Start mit „star-like-tree“; identifizierte sukzessive Nachbar-Taxa, die durch eine gemeinsame Verzweigung verbunden sind

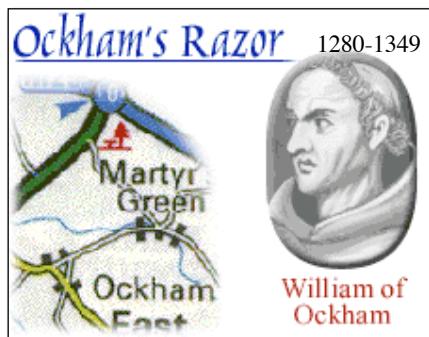
Stammbaum-Rekonstruktion mit „character state“-Methoden

- Maximum Parsimony (MP)
- Maximum Likelihood (ML)
- Bayesian MCMC

Diese Methoden analysieren das Sequenzalignment und bewerten nicht nur die Menge, sondern auch die Art der beobachteten Veränderungen.

Die Methoden arbeiten mit aufwändigen Suchstrategien, wobei im Idealfall alle möglichen Bäume (!!) konstruiert und dann auf ihre Optimalität hin überprüft werden.

Character state-Methode A: Maximum Parsimony



„Entia non sunt multiplicanda sine necessitate“

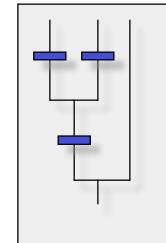
Character state-Methode A: Maximum Parsimony /1

- Annahme: „die Evolution nimmt den kürzesten Weg, d.h., derjenige Evolutionspfad ist richtig, der mit den wenigsten Veränderungen auskommt“
- gesucht wird unter allen möglichen Bäumen der (oder die!!) „maximum parsimony (MP) tree(s)“
- findet oft mehrere gleichwertige Bäume
- bietet die Möglichkeit, die ancestralen Sequenzen zu rekonstruieren

Review: Nature 361 (1993), S. 603

Maximum Parsimony /2

Sequenz	Position								
	1	2	3	4	5	6	7	8	9
1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G
					#	#	#		

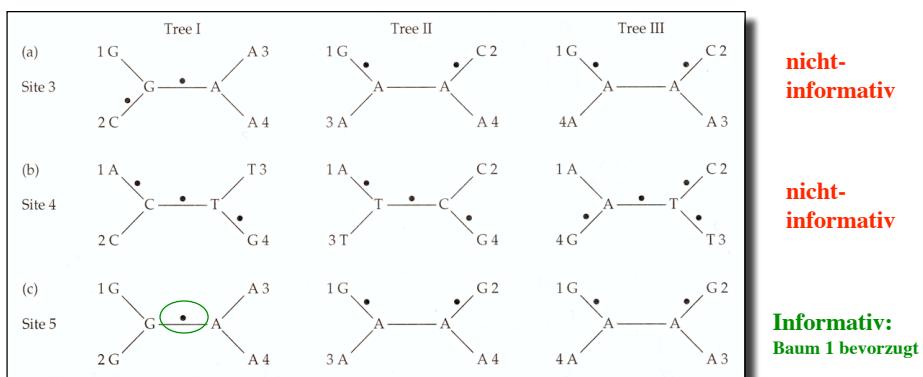


Nur variable Orte, die gleichzeitig „informativ“ sind, können ausgewertet werden!

„informativ“ = mind. 2 verschiedene character, jeweils
mind. 2x in verglichenen Sequenzen vorhanden (Synapomorphien)

Maximum Parsimony /3

- einfachstes Beispiel: es gibt drei mögliche Bäume für 4 OTUs
- Ast-Enden zeigen Basen der 4 OTUs an den Alignment-Positionen 3,4,5,9
- interne Knoten zeigen Auswahl möglicher Sequenz-Konstellationen
- Dots = notwendige Austausche, um Baum zu erklären



Maximum Parsimony /4

		Position								
		1	2	3	4	5	6	7	8	9
Sequenz		A	A	G	A	G	T	G	C	A
1										
2		A	G	C	C	G	T	G	C	G
3		A	G	A	T	A	T	C	C	A
4		A	G	A	G	A	T	C	C	G
						#	#	#		

- Positionen 5 und 7 unterstützen Baum I:

(1,2)(3,4)

- Position 9 unterstützt Baum II:

(1,3)(2,4)



Baum I ist MP-tree

MP - „heuristic search“

Oder: das Problem des blinden Bergsteigers

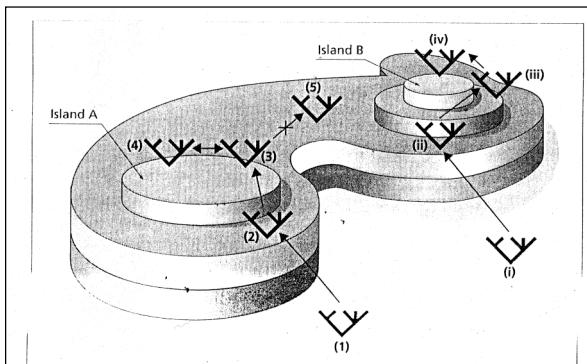


Fig. 6.4 Landscapes and the problem of islands of trees (locally optimal sets of trees). A hill climbing algorithm that started from tree 1 would succeed in finding trees 3 and 4 (which comprise island a), but would fail to discover that tree iv (island b) was even better, because to get to that tree it would have to cross a plateau of trees that were worse than trees 3 and 4. However, a search starting from tree i would succeed in finding the best tree. If the set of possible trees contains more than one island then heuristic methods may land on a suboptimal island, and the optimal island will not be discovered. After Maddison (1991).

- Ergebnis hängt davon ab, von wo die Suche nach MP startet (1 oder i?)

• keine Garantie, daß MP-Baum gefunden wird!

„long branch attraction“

- OTUs mit hoher Evolutionsrate und vielen Veränderungen („long branches“) enthalten notwendigerweise zahlreiche Homoplasien/Konvergenzen
- diese Homoplasien führen dazu, daß MP die „long branch“-OTUs im Baum fälschlicherweise zueinandergruppiert

> u. U. Taxa mit long branches entfernen!

Größtes Problem bei MP: „long branch attraction“

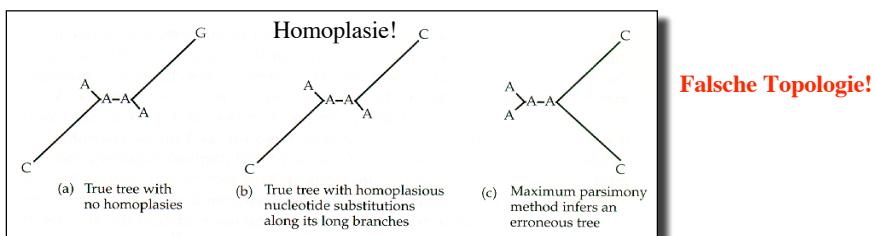


FIGURE 5.28 The long-branch attraction phenomenon. (a) The true unrooted tree has two long branches, each neighboring a short branch. The letters represent the nucleotides at the terminal and internal nodes. On the short branches, we assume that the probability of a nucleotide substitution is very small, so that the nucleotides at the tips of the short branches are likely to retain the same character state as that of the ancestral node. In contrast, on the long branches nucleotide substitutions are likely to occur with a high probability. If the nucleotide substitutions on the long branches are not homoplasious, then by using maximum parsimony we will obtain the correct tree. (b) By chance, however, a site may experience homoplasious nucleotide substitutions along the two long branches. As a consequence, the maximum parsimony method will yield an erroneous tree (c), in which the long branches are inferred to be neighbors. The reason for this error is that the correct tree (b) requires two nucleotide substitutions, whereas the erroneous tree (c) requires only a single nucleotide substitution.

„LBA“ oder
„Felsenstein zone“

Vor- und Nachteile der Maximum Parsimony

- relativ **einfache** Methode mit gut dokumentierten Programmen
- ermöglicht die **Rekonstruktion der ancestralen Sequenzen**
- häufig „**long branch attraction**“-Problem bei divergenten Sequenzen mit vielen Homoplasien
- **Astlängen werden unterschätzt**
- **große Datenmengen notwendig**, da informative Orte selten sind

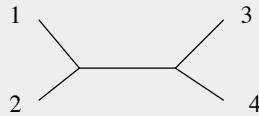
Methoden - Übersicht

Rekonstruktionsmethode	Clustering-Algorithmus	Datentyp	
		Distanzen	Character
	UPGMA		
	Neighbor joining		
	Minimum Evolution	Maximum Parsimony	
		Maximum Likelihood	
		Bayes	

Character state-Methode B: Maximum Likelihood /1

Sequenzalignment				
1	A	A	G	A
2	A	G	C	C
3	A	G	A	T
4	A	G	A	G

Ein möglicher Baum (von dreien):



L = P(data|tree)

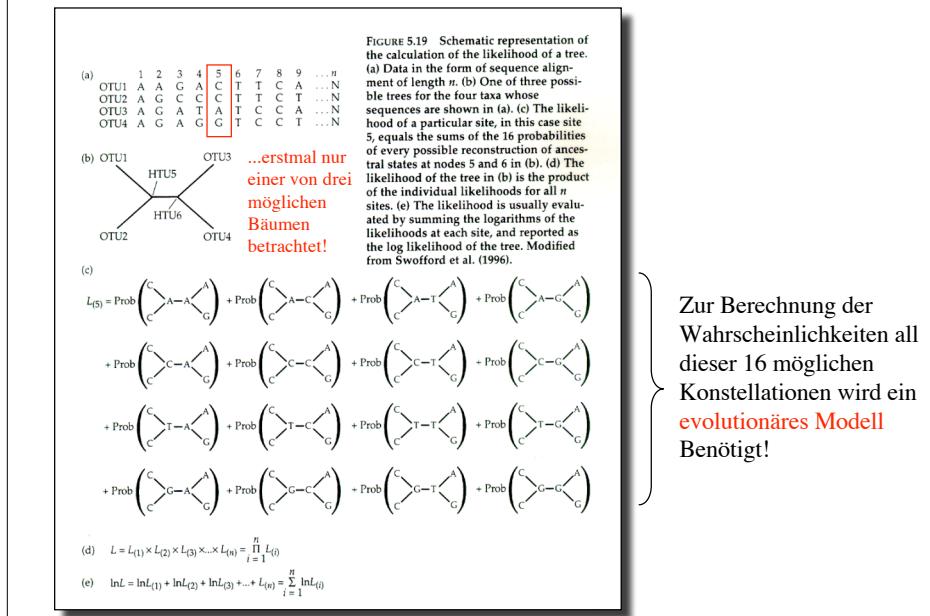
Wie hoch ist die Wahrscheinlichkeit,
dass der gezeigte Stammbaum für die
alignierten Sequenzen verantwortlich ist?

**Finde unter allen möglichen Bäumen denjenigen mit dem höchsten
Likelihood-Wert L!**

Maximum Likelihood /2

1. Alignment machen. Den ersten aller möglichen Bäume betrachten: wie hoch ist die Wahrscheinlichkeit, daß Baum 1 die Daten erklärt?
2. Annahme: alle Sequenzpositionen evolvieren unabhängig. Dann L für jede Position bestimmen. L für Pos. X ist Summe der Einzelwahrscheinlichkeiten für die möglichen „ancestral states“.
3. L-Werte aller Sequenzpositionen am Ende multiplizieren. Besser: Aus mathematischen Gründen L-Werte in Logarithmus (ln) überführen: $\ln L = \ln L_1 + \ln L_2 + \ln L_3 + \dots$
 $\ln L$ = „log likelihood“ des betrachteten Baumes
4. $\ln L$ ebenso für alle anderen möglichen Bäume berechnen.
Baum mit höchstem $\ln L$ -Wert ist „ML-Baum“.

Maximum Likelihood /3



Maximum Likelihood /4

Daten

Pos.5

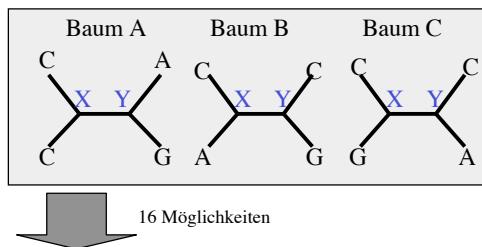
1. C
2. C
3. A
4. G

Evolutionsmodell

(nicht real!! Nur zum Rechnen!)

	A	T	G	C
A	1	0,1	0,1	0,1
T		1	0,1	0,1
G			1	0,1
C				1

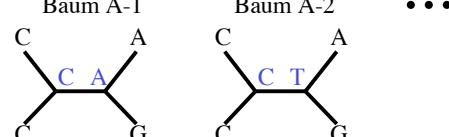
vier OTUs > drei mögliche Bäume



16 Möglichkeiten

Summe

$$A1-A16 = L \text{ (Baum A an pos. 5)}$$



$$1x1x0,1x1x0,1 = 0,01 \quad 1x1x0,1x0,1x0,1 = 0,001$$

Maximum Likelihood /5

... L-Wert zunächst an allen anderen Positionen des Alignments für den Baum A berechnen.

Die Likelihood des Baumes A ist das Produkt der Einzelwahrscheinlichkeiten an jeder Position (oder: die Summe der lnL der einzelnen Positionen).

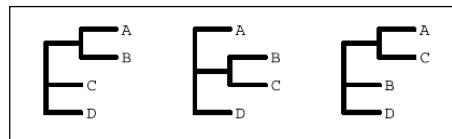
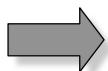
Dann lnL-Wert für ebenfalls die möglichen Bäume B und C ermitteln.

höchster lnL-Wert = ML-tree

Vieles spricht für ML !

- komplette Sequenzinfo wird genutzt; selbst nicht-Parsimony-informative Orte! Bsp:

A	acgcaa
B	acataa
C	atgtca
D	gcgtta



Keine MP-informativen Orte!

In MP sind alle Bäume gleich lang!

ML kann dennoch die vorhandene phylogenetische Information nutzen!

- wenn Evolutionsmodell stimmt, gibt es **kaum Probleme mit LBA**

Nur Weniges spricht dagegen...

- falsches Evolutionsmodell > falscher Stammbaum!
- aber: Modell muss nicht blind angenommen werden, sondern kann aus den Daten selbst berechnet werden!!

Man macht likelihood-Analysen mit verschiedenen Werten für Parameter (z. B. T_i/T_v , α -Parameter etc), vergleicht die L-Werte und stellt dann den Parameter letztendlich so ein, daß L am größten ist.

- leider ist ML extrem rechenaufwändig

...aber schnelle ML-Programme existieren schon!

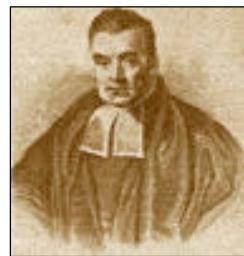


Review: Science 14.12.2001

Bayesian MCMCMC

...kombiniere Bayes-Statistik
und superschnelle Computeralgorithmen

Bayes who?



Reverend Thomas Bayes
1702-1761

Bayes, T. 1763. *An essay towards solving
A problem in the doctrine of chance.*
Philosop. Transact. Royal Soc. London

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model}) \times P(\text{model})}{P(\text{data})}$$

P(data) > bekannte Konstante
P(data|model) > Likelihood
P(model) > „Prior“, „Erwartung an Modell“



$P(\text{model}|\text{data}) \approx P(\text{data}|\text{model}) \times P(\text{model})$

Posterior probab.

Likelihood

prior probab.

Satz von Bayes: Beispiel

Berechne Wahrscheinlichkeit der Diagnose **Bronchitis** bei dem Befund „**Husten**“:

$$P(\text{Bronchitis} | \text{Husten}) = \frac{P(\text{Husten} | \text{Bronchitis}) \times P(\text{Bronchitis})}{P(\text{Husten})}$$

$$\begin{aligned} P(\text{Bronchitis}) &= 0.05 \\ P(\text{Husten}) &= 0.2 \\ P(\text{Husten} | \text{Bronchitis}) &= 0.8 \end{aligned} \quad = 0.05 \times 0.8 / 0.2 = 0.2$$

d.h., die Wahrscheinlichkeit, dass ein Patient mit Husten Bronchitis hat, ist 20%
(4 x so hoch wie die *a priori*-Wahrscheinlichkeit von Bronchitis)

Bayes und Bäume

$$\bullet \text{ Bayes Theorem: } P(\text{tree} | \text{data}) \approx P(\text{data} | \text{tree}) \times P(\text{tree})$$

$P(\text{tree} | \text{data})$ > **posterior probability** gibt die Wahrscheinlichkeit an, mit der ein Baum korrekte Topologie und Astlängen besitzt

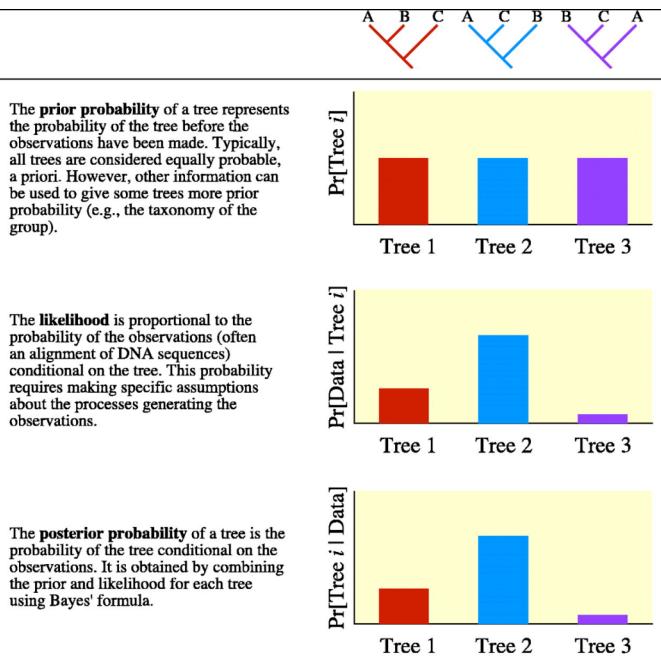
$P(\text{tree})$ > „prior probability“, daß eine Phylogenie korrekt ist (= Annahme, ohne die Daten zu kennen). (Einfachste Annahme: alle Bäume sind gleich wahrscheinlich)

$P(\text{data} | \text{tree})$ > Likelihood-Wert der Daten, auf Grundlage eines Baumes und eines Substitutionsmodells (wird berechnet)

Die Bäume mit der höchsten „posterior probability“ werden gesucht!

Dazu muss über alle Bäume summiert werden, und zwar bei allen möglichen Kombinationen von Topologien und Astlängen, sowie den verschiedenen Parametern von Substitutionsmodellen.

Bayes und Bäume



MCMC

Metropolis-coupled Markov chain Monte Carlo

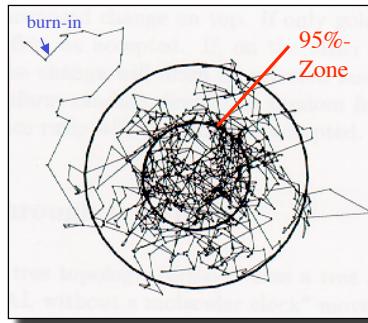
Metropolis, N. et al. (1953)
Equations of state calculations by fast computing machines.
J. Chemical Physics

Metropolis-Hastings-Algorithmus:

1. Ein neuer Baum wird durch Rearrangement existierender Bäume generiert.
2. Der neue Baum wird entweder akzeptiert oder verworfen, gemäß vordefinierter Kriterien und Wahrscheinlichkeiten
3. Zurück zu 1.

- MCMC simuliert die Verteilung der posterior probability für alle Bäume im „tree space“
- die Häufigkeit, mit der MCMC Bereiche des „tree space“ besucht und auswertet, korrespondiert mit der Höhe der probability der dort befindlichen Bäume!

MCMCMC



- „cold chain“ sammelt
- „hot chains“ als „scouts“, um die Gipfel in der Baumschaft zu lokalisieren

> der „MC Roboter“ sucht nicht DEN optimalen Baum, sondern „sammelt“ die **Bäume mit der höchsten posterior probability** („Gipfel in der Baumschaft“).

> Anhand dieser Baum-Sammlung wird ein Konsensus-Baum erstellt, dessen Verzweigungen durch die Höhe der pp-Werte gekennzeichnet und bewertet werden

Methodenübergreifende Fragen...

Wo ist die Wurzel in meinem Baum?

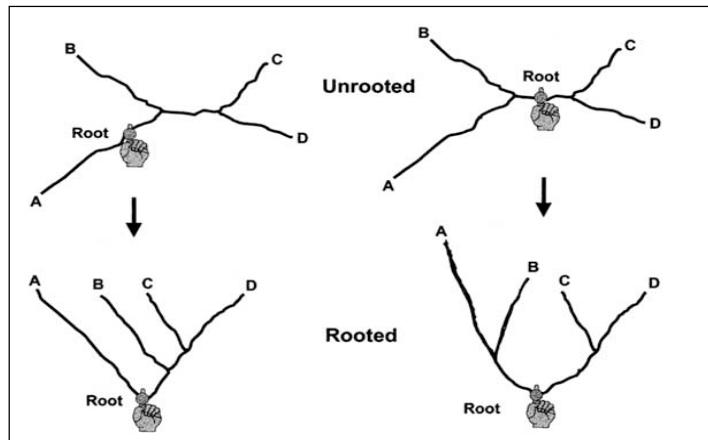
Wie vergleiche ich Bäume miteinander?

Wie bewerte ich die Verlässlichkeit von Bäumen?

Welche Programme?

Welche Methoden funktionieren am besten?

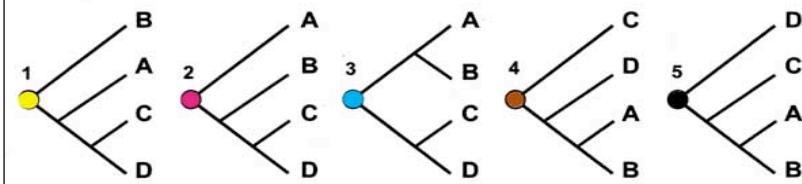
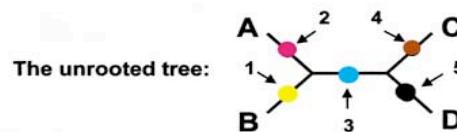
,Rooting' – Wo ist die Wurzel des Baums?



Achtung: die meisten Rekonstruktionsmethoden produzieren zunächst unrooted trees

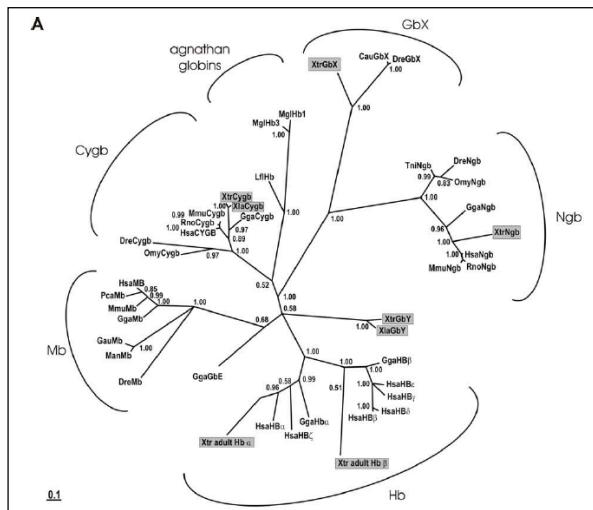
,Rooting' – Wo ist die Wurzel des Baums?

An unrooted, four-taxon tree can be rooted in five different places to produce five different rooted trees



↔ Unterschiedliche Aussage!

„Rooting“ - Ist ein unrooted tree ausreichend?



Globingen- Stammbaum (Bayes) zur Festlegung der Verwandtschaft neu gefundener Gene

Ein Baum ist eine Evolutionshypothese!

Wie wissen wir, ob der rekonstruierte Baum korrekt ist?

1. Wie verlässlich ist der Baum?
 2. Welche Verzweigungen sind verlässlich?
 3. Ist *der* Baum signifikant besser als ein anderer?

Qualitätsbewertung von Bäumen

- MP, ML, NJ u. a. Distanzmethoden:
(nicht-parametrisches) Bootstrapping
- ML- Methoden:
likelihood ratio test (LRT)
QuartetPuzzling (QP)-Werte
- Bayes MCMC:
Posteriore Wahrscheinlichkeiten
...und andere.

Bootstrapping



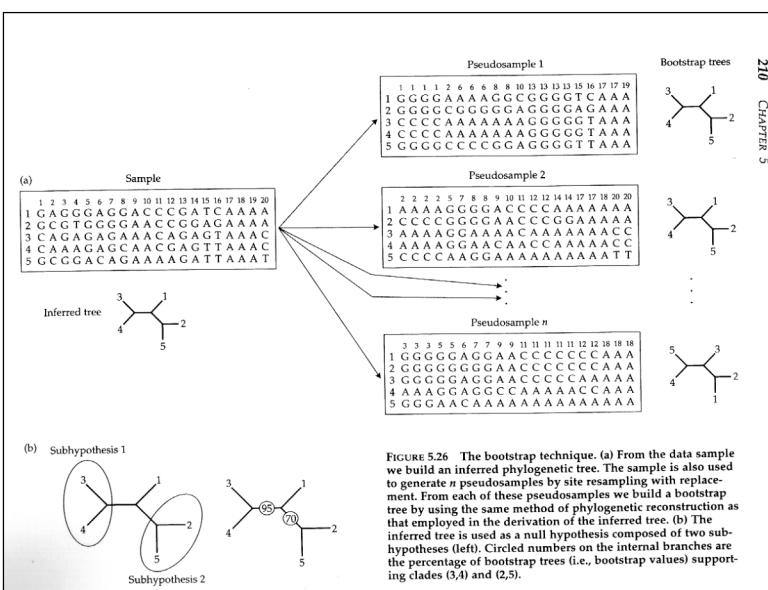
Wie gut ist die Gruppierung zweier OTUs zu einer Clade im Baum statistisch abgesichert?

- Erstellung von 500-1000 Teildatensätzen (Pseudosamples) durch „resampling with replacement“

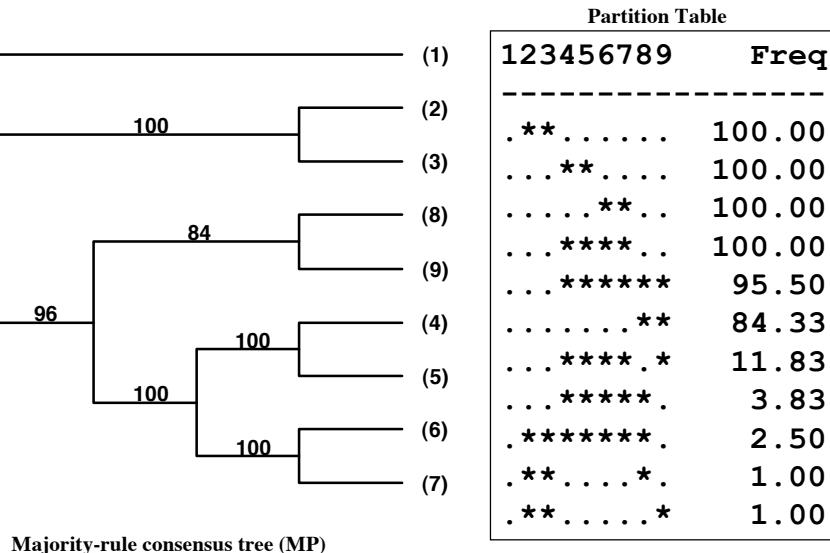
D. h., manche Positionen des Sequenzalignments werden mehrfach ausgewählt, andere dafür gar nicht!

- Baumrekonstruktion für diese Teil-Datensätze
- Bootstrap-Wert = 80% bedeutet: in 80% der Fälle werden die OTUs 1 und 2 einander zugeordnet

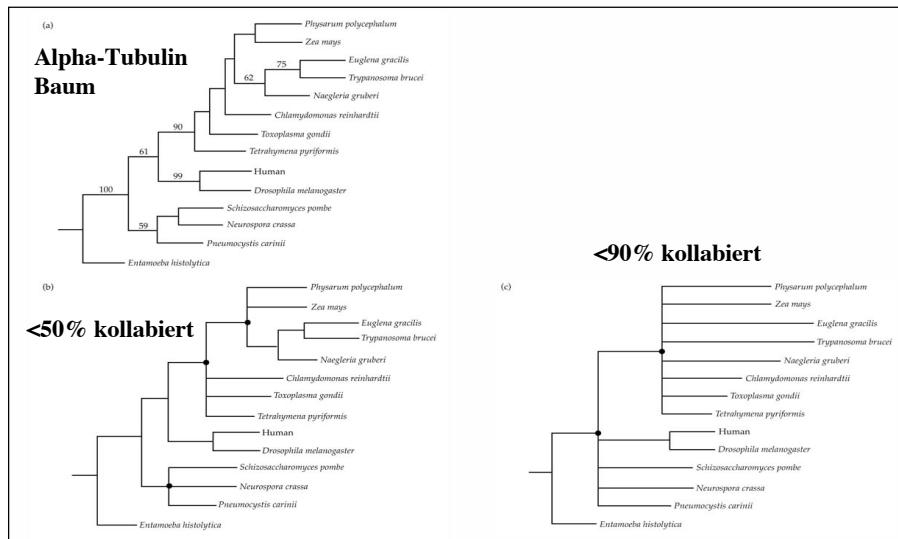
Bootstrapping



Ergebnis eines Bootstrappings



Umgang mit Bootstrap-Werten



Interpretation von Bootstrap-Werten

- hohe Bootstrap-Werte ($>70\%$) zeigen eine **gute Unterstützung** der Gruppierung durch die Daten an
- Verzweigungen mit B-Werten unter 50% sollten „auf eine gemeinsame Linie kollabiert werden“ > Polytomie
- niedrige Werte bedeuten nicht, dass die Gruppierung falsch ist! Sie ist nur von den vorliegenden Daten nicht ausreichend unterstützt.
- Bootstrapping kann als Versuch gesehen werden, die Robustheit einer phylogenetischen Rekonstruktion zu testen gegenüber Störungen in ihrer „Balance“ für und wider die Zueinandergruppierung von Taxa.
- **Wissenschaftliche Journale machen Verwendung dieser Methode zur Pflicht beim Zeigen eines Stammbaums!**

Allgemeine Hinweise



- stets mehrere Verfahren ausprobieren
- je größer der Datensatz, desto besseres Ergebnis
- bei Distanzen: adäquate Substitutionsmodelle und Korrektur für multiple Austausche
- bei ML: zu Grunde liegende Modelle relativ zueinander testen („so kompliziert wie nötig, so einfach wie möglich“)
- bei MP, aber auch bei anderen: schnell evolvierende Taxa (Homoplasien! LBA!) entfernen oder mehr Taxa hinzufügen

Darwin's letter to Thomas Huxley 1857

“The time will come I believe, though I shall not live to see it, when we shall have fairly true genealogical (*phylogenetic*) trees of each great kingdom of nature”

