

**Einführung in die Angewandte Bioinformatik:  
Phylogenetik und Taxonomie  
24.06.2010**

Prof. Dr. Sven Rahmann

## Phylogenetik: Berechnung phylogenetischer Bäume

**Phylogenetik** (phylum = Stamm):

Rekonstruktion von evolutionären Stammbäumen

(= phylogenetische Bäume, **Phylogenien**)

aus unterscheidbaren **Merkmalen**, insbes. aus DNA- / Proteinsequenzen,  
und/oder aus Merkmalen berechneten **Distanzen**

- für Spezies [**Spezies-Bäume**]
- für einzelne Gen- / Protein-Familien [**Gen-Bäume**]

### Tree Of Life - Projekt

Langfristiges Ziel:

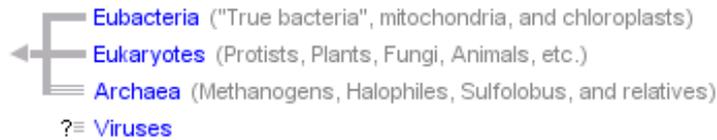
Stammbaum aller existierenden (und ausgestorbenen) Arten.

Tree Of Life Web Project: <http://www.tolweb.org>

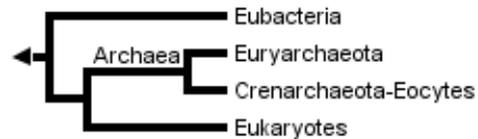
## Tree Of Life Web Project

<http://www.tolweb.org>

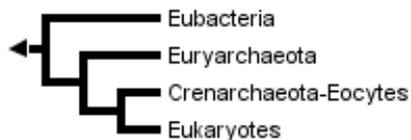
Kontroverse über Verzweigungen  
nahe der Wurzel



The "archaea tree":

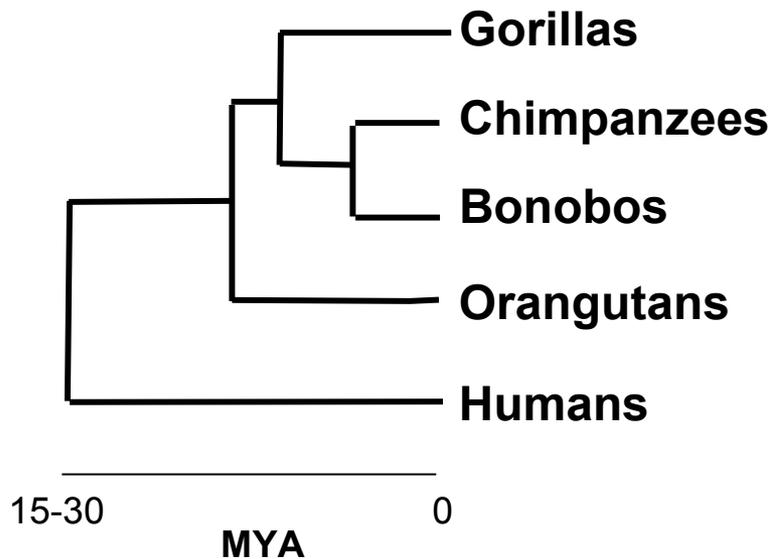


The "eocyte tree":

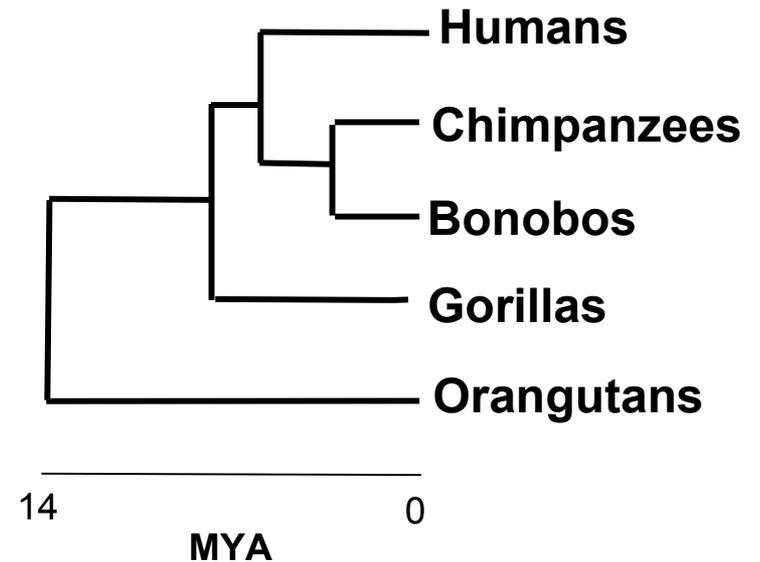


## Aufgaben der Phylogenetik (Speziesbäume)

Bestimmung der Verwandtschaftsverhältnisse zwischen Arten (Spezies) und der Ereignisse, die bei den Arten seit der Speziation stattfanden.



Klassische Sichtweise,  
basierend auf äußeren Merkmalen



Moderne Sichtweise,  
basierend auf DNA-Vergleich

## Aufgaben der Phylogenetik (Genbäume)

Bestimmung der Verwandtschaftsverhältnisse zwischen Genen  
in einer Gen- / Proteinfamilie / Domäne.

Beispiel:

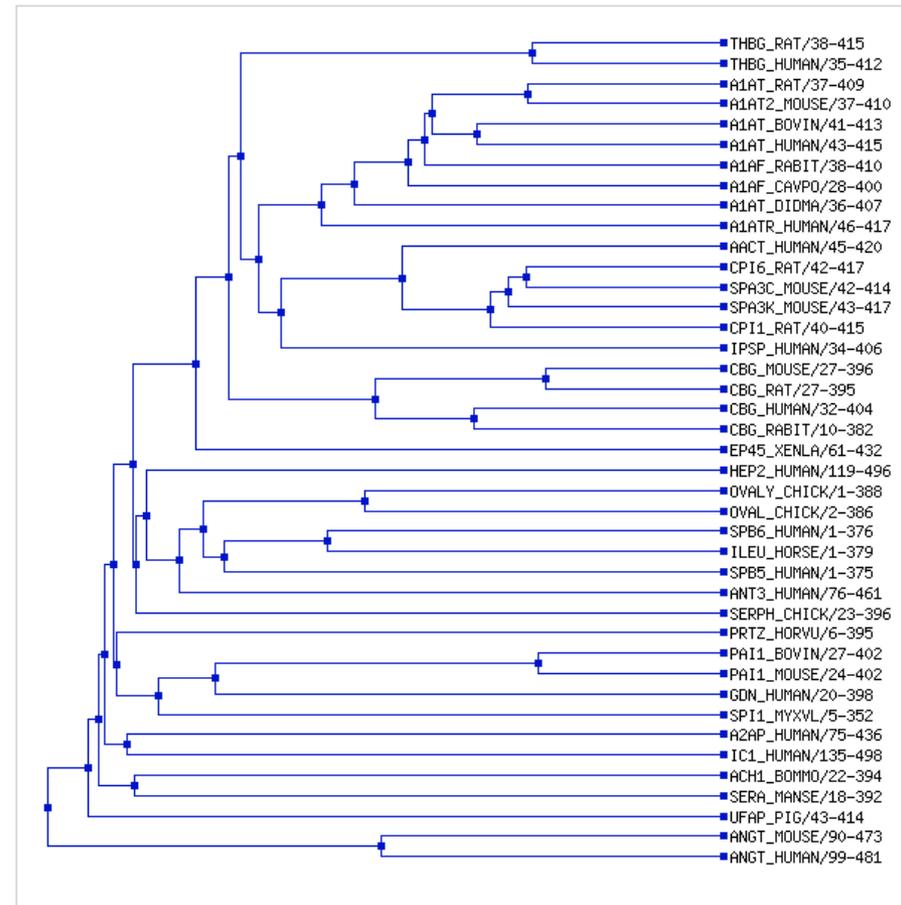
Serpin-Familie in der pfam-Datenbank

Dies ist ein Gen-Baum (gene tree),  
kein Arten-Baum (species tree):

An den Blättern (rechts) stehen  
einzelne Gene/Proteine/Sequenzen,  
keine Spezies.

Evolutionäre Ereignisse (innere Knoten):

- Genduplikation
- Speziation



## Gibt es einen DNA-Bereich (Gen?), der sich zur Rekonstruktion von Spezies-Bäumen eignet?

### Schwierig:

langsam evolvierende Gene unterscheiden nicht nah verwandte Spezies,  
dafür Gruppen auf höherem Niveau.

schnell evolvierende Gene unterscheiden gut nah verwandte Spezies,  
dafür geringe Auflösung auf höherem Niveau.

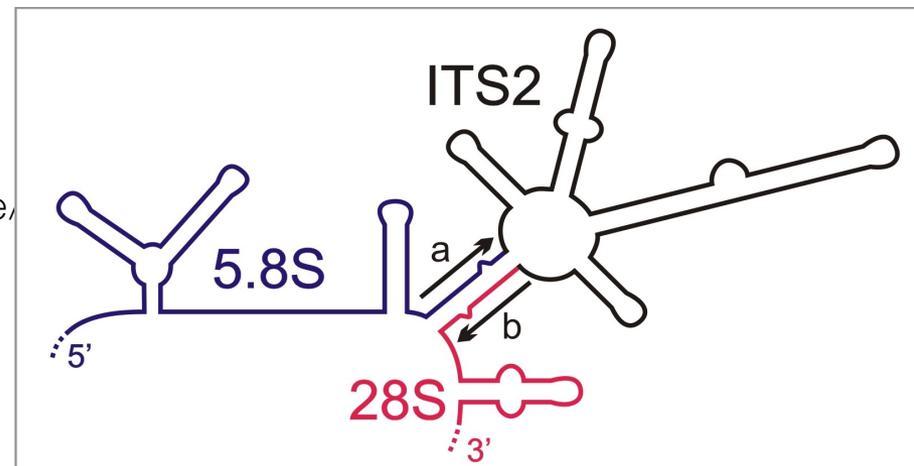
### Mögliche Lösung (Eukaryoten):

ITS2-Marker (internal transcribed spacer 2 der rRNA)

- Sequenz evolviert schnell.
- Sekundärstruktur evolviert langsam.

ITS2-Datenbank mit Strukturen:

<http://its2.bioapps.biozentrum.uni-wuerzburg.de/>



## Taxonomie (gr. *táxis*=Ordnung, *-nómos* =Gesetz)

systematische hierarchische Klassifikation  
der Verwandtschaftsbeziehungen von Lebewesen  
anhand von beobachtbaren Merkmalen.

### Unterschied zur Phylogenetik:

benutzte Hierarchie muss nicht evolutionäre Verwandtschaft sein.  
Idealerweise verlangt man aber monophyletische taxonomische Gruppen (Taxa).

Klassische Einteilung der Lebewesen in:

- Reich
- **S**tamm (Abteilung); **U**nterstamm
- **K**lasse
- **O**rdnung; **U**nterordnung
- **F**amilie; **U**nterfamilie
- **G**attung
- **A**rt; **U**nterart

## NCBI Taxonomy - Datenbank

<http://www.ncbi.nlm.nih.gov/Taxonomy/>

### Inhalt:

- taxonomische Klassifikation der Lebewesen (nicht: Tree of Life!).
- weiterführende Informationen in anderen NCBI-Datenbanken zu jeder Spezies.

NCBI Taxonomy Browser

PubMed Entrez BLAST OMIM Taxonomy Structure

Search for  As complete name  lock

**Taxonomy browser**

- Archaea
- Bacteria
- Eukaryota
- Viroids
- Viruses

Taxonomy common tree

Taxonomy information

Taxonomy resources

Taxonomic advisors

Genetic codes

Taxonomy Statistics

Taxonomy Name/Id Status Report

Taxonomy FTP site

### The NCBI Taxonomy Homepage

**These are direct links to some of the organisms commonly used in molecular research projects:**

<a href="#">Arabidopsis thaliana</a>	<a href="#">Escherichia coli</a>	<a href="#">Pneumocystis carinii</a>
<a href="#">Bos taurus</a>	<a href="#">Hepatitis C virus</a>	<a href="#">Rattus norvegicus</a>
<a href="#">Caenorhabditis elegans</a>	<a href="#">Homo sapiens</a>	<a href="#">Saccharomyces cerevisiae</a>
<a href="#">Chlamydomonas reinhardtii</a>	<a href="#">Mus musculus</a>	<a href="#">Schizosaccharomyces pombe</a>
<a href="#">Danio rerio (zebrafish)</a>	<a href="#">Mycoplasma pneumoniae</a>	<a href="#">Takifugu rubripes</a>
<a href="#">Dictyostelium discoideum</a>	<a href="#">Oryza sativa</a>	<a href="#">Xenopus laevis</a>
<a href="#">Drosophila melanogaster</a>	<a href="#">Plasmodium falciparum</a>	<a href="#">Zea mays</a>

Comments and questions to [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)

## NCBI Taxonomy – Datenbank: Beispiel

Suche nach „red fox“.

Beachte: taxonomy ID, genetische Codes, Abstammungslinie, Entrez links.

### Vulpes vulpes

*Taxonomy ID:* 9627

*Genbank common name:* **red fox**

*Inherited blast name:* **carnivores**

*Rank:* species

*Genetic code:* [Translation table 1 \(Standard\)](#)

*Mitochondrial genetic code:* [Translation table 2 \(Vertebrate Mitochondrial\)](#)

*Other names:*

synonym: **Vulpes vulpes var.**

common name: **silver fox**

[Lineage\( full \)](#)

[cellular organisms](#); [Eukaryota](#); [Fungi/Metazoa group](#); [Metazoa](#); [Eumetazoa](#);  
[Bilateria](#); [Coelomata](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#);  
[Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Tetrapoda](#); [Amniota](#);  
[Mammalia](#); [Theria](#); [Eutheria](#); [Laurasiatheria](#); [Carnivora](#); [Caniformia](#); [Canidae](#);  
[Vulpes](#)

Entrez records		
Database name	Subtree links	Direct links
<b>Nucleotide</b>	<a href="#">354</a>	<a href="#">352</a>
<b>Nucleotide GSS</b>	<a href="#">83</a>	<a href="#">83</a>
<b>Protein</b>	<a href="#">257</a>	<a href="#">257</a>
<b>Genome Sequences</b>	<a href="#">1</a>	<a href="#">1</a>
Popset	<a href="#">44</a>	<a href="#">44</a>
<b>PubMed Central</b>	<a href="#">128</a>	<a href="#">128</a>
Gene	<a href="#">37</a>	<a href="#">37</a>
<b>OMIA</b>	<a href="#">7</a>	<a href="#">7</a>
<b>Taxonomy</b>	<a href="#">2</a>	<a href="#">1</a>

## Grundlagen der Phylogenetik

- Graphen
- Bäume
- „Vokabeln“

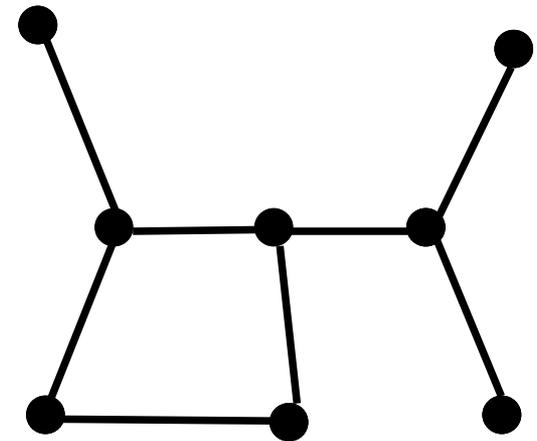
## Grundlagen: Graphen

Ein Graph besteht aus:

- **Knoten** (vertices  $V$ , nodes) und
- **Kanten** (edges  $E$ ),  
die zwischen den Knoten verlaufen.

Graphen können **gerichtet** oder **ungerichtet** sein  
(Kanten haben eine Richtung oder nicht).

Der **Grad** eines Knoten  
ist die Anzahl seiner Nachbarn.



## Grundlagen: ungewurzelte Bäume

Ungewurzelter **Baum** (unrooted tree)  $T = (V, E)$  :

Graph ist

- **zusammenhängend**  
(:= jeder Knoten wird von jedem anderen erreicht)
- **kreisfrei**  
(:= es gibt nur einen Weg zwischen je zwei Knoten)

In einem Baum unterscheidet man zwischen

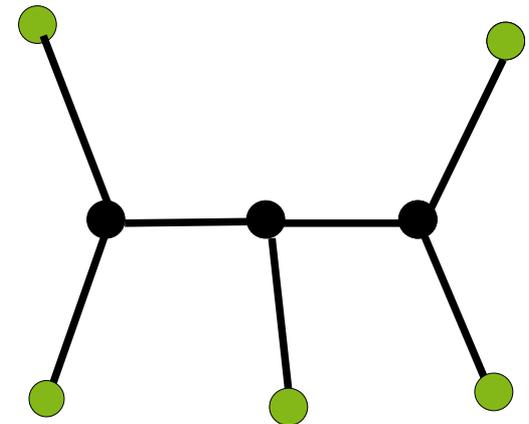
- **inneren Knoten** (inner nodes  $N$ ), und
- **Blättern** = äußeren Knoten (leaves  $L$ ).

Baum heißt **Binärbaum**,

wenn innere Knoten Grad 3 und Blätter Grad 1 haben.

Ungewurzelte binäre Bäume erfüllen

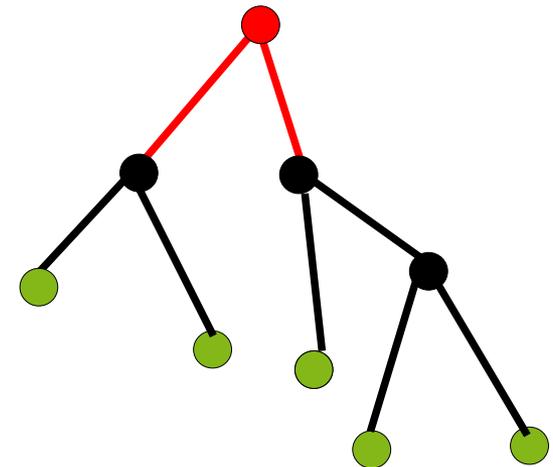
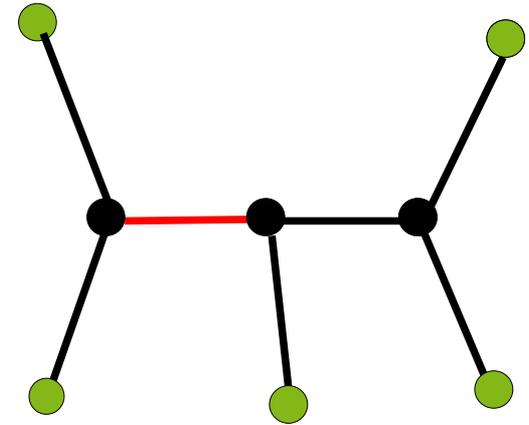
$|E| = 2|L| - 3$  und  $|N| = |L| - 2$ .



## Grundlagen: Gewurzelte Bäume

Ungewurzelter Baum kann gewurzelt werden, indem man eine Kante auswählt, in deren Mitte einen Knoten (**Wurzel**) einfügt, und den Baum daran aufhängt.

Wurzel hat Grad 2, stets ganz oben.  
Kanten werden von der Wurzel weg gerichtet  
(Kanten verlaufen von oben nach unten.)  
Blätter unten.



## Anzahl Binärbäume

### Ungewurzelte Binärbäume:

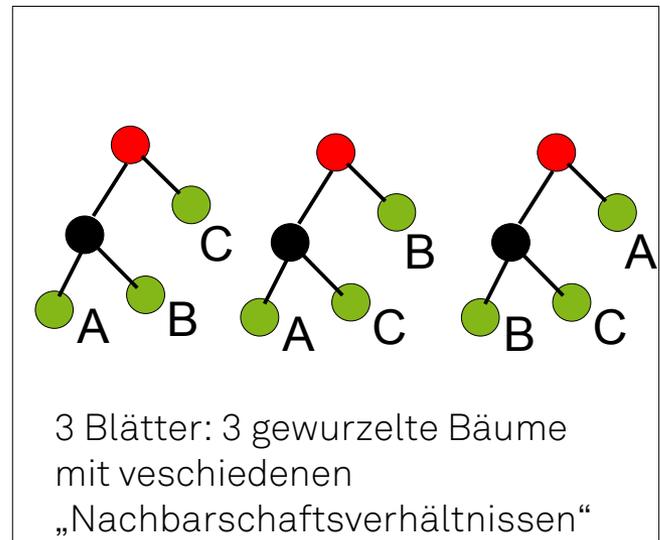
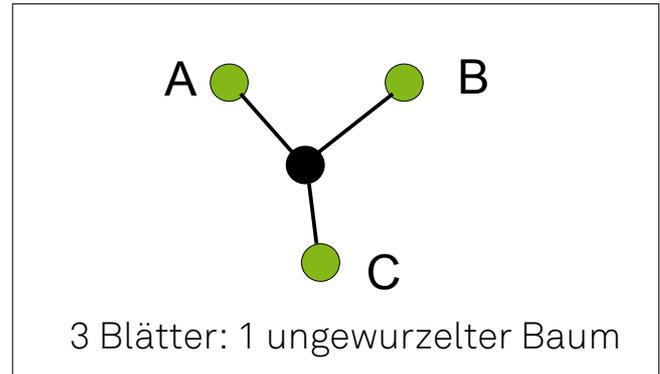
- 3 Blätter: 1 Baum mit 3 Kanten
- 4 Blätter: 3 verschiedene Bäume, je 5 Kanten
- 5 Blätter:  $3 \cdot 5 = 15$  Bäume, je 7 Kanten
- 6 Blätter:  $15 \cdot 7 = 105$  Bäume ...

### Gewurzelte Binärbäume:

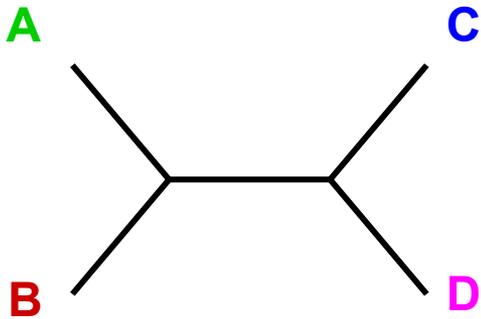
Man kann die Wurzel wie ein weiteres Blatt bei ungewurzelten Bäumen behandeln:

- 3 Blätter: 3 Bäume
- 4 Blätter:  $3 \cdot 5 = 15$  Bäume
- 5 Blätter:  $15 \cdot 7 = 105$  Bäume ...

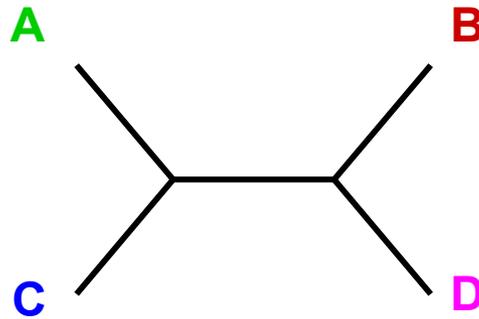
Anzahl der binären Bäume wächst super-exponentiell in der Zahl der Blätter  $n$  (schneller als  $c^n$  für jede Konstante  $c$ )



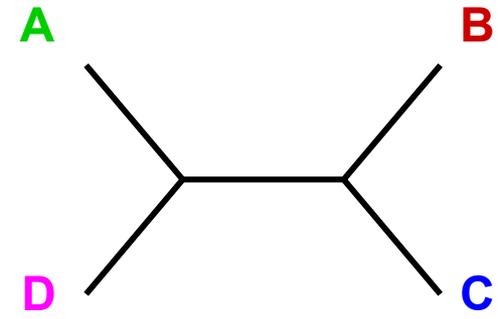
## Die drei Quartette (ungewurzelte Bäume mit 4 Knoten)



AB || CD



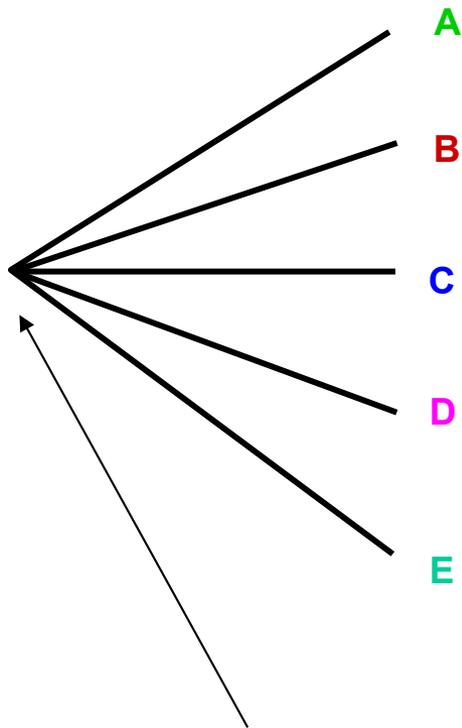
AC || BD



AD || BC

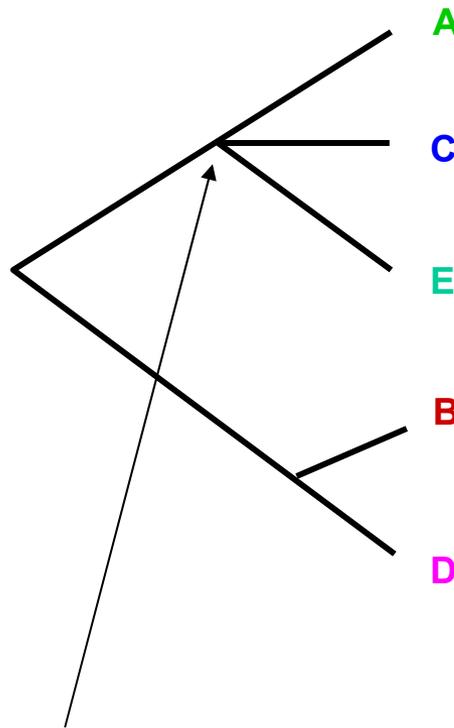
## Grad der Auflösung des Verzweigungsmusters

“Stern”-Baum, nicht aufgelöst

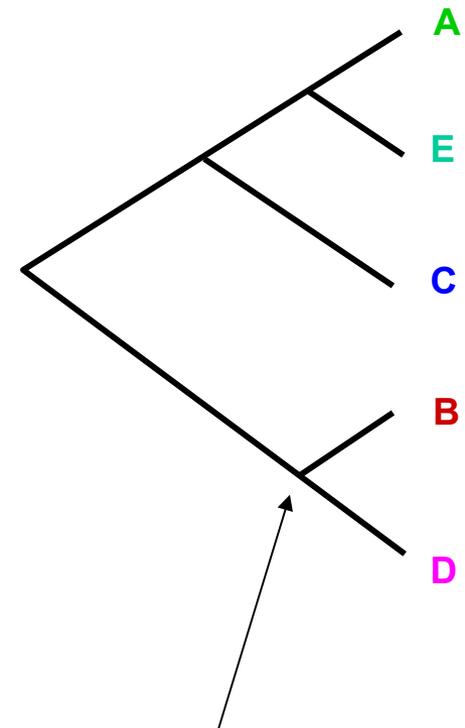


Polytomie, Multifurkation

Teilweise aufgelöst



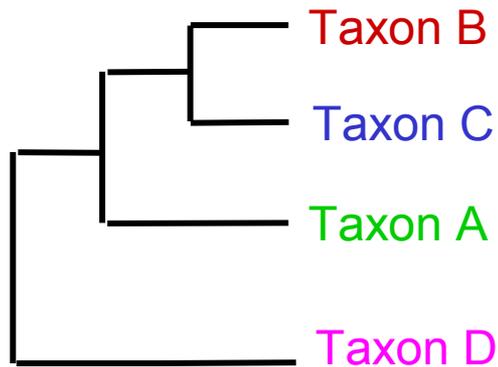
Vollständig aufgelöst  
(Binärbaum)



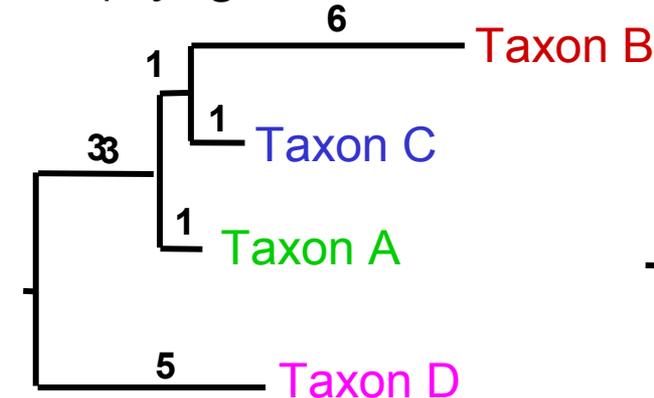
Bifurkation

## Darstellung phylogenetischer Bäume

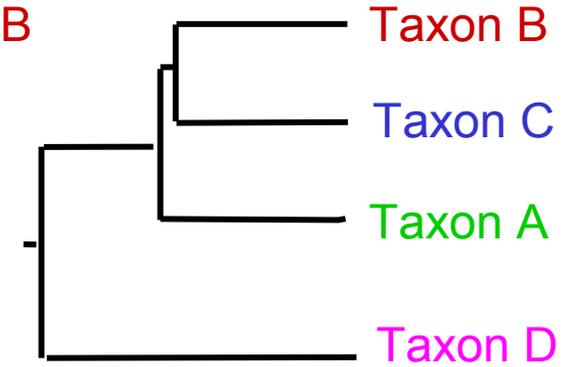
Kladogramm  
(cladogram)



Phylogramm  
(phylogram)



Ultrametrischer Baum  
(ultrametric tree)

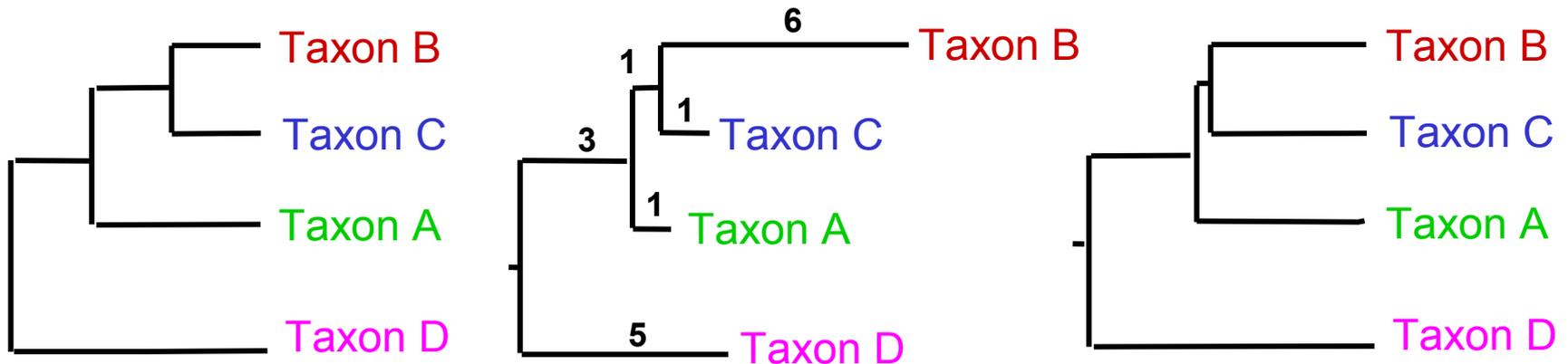


Achse hat  
keine Bedeutung

Menge an Evolution  
(z.B. PAM-Einheiten)

Realzeit: gleicher Abstand  
aller Blätter zur Wurzel

## Darstellung phylogenetischer Bäume



Alle Darstellungsformen (Kladogramm, Phylogramm, ultrametrischer Baum) zeigen die gleiche Baumtopologie (Verzweigungen zwischen den Taxa). Taxa im selben Unterbaum bilden eine **monophyletische Gruppe** (clade).

Darstellung durch geschachtelte Klammern:

- Kladogramm:  $((B,C),A),D$
- Phylogramm:  $((B:6,C:1):1,A:1):3), D:5)$
- ultrametrischer Baum: analog

## Die molekulare Uhr (molecular clock)

Evolutionäre Abstände zwischen Sequenzen werden gemessen

- nicht in Realzeit,
- sondern als "Evolutionsmenge" [PAM].

Problem: Distanz ist nicht Realzeit.

Evolutionsrate variiert idR zeitlich, genspezifisch, gattungsspezifisch.

Berechnete Bäume repräsentieren keine realen zeitlichen Abstände.

Auch die Wurzel kann nicht zuverlässig positioniert werden.

Man berechnet ungewurzelte Phylogramme.

Ausnahme: Evolutionsrate konstant.

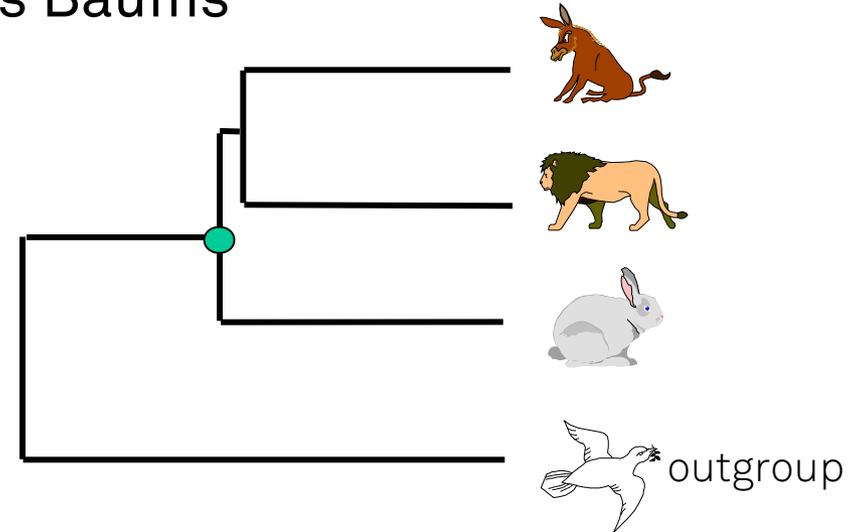
Man sagt: Es gilt die Hypothese der „molekularen Uhr“.

Man erhält einen gewurzelten ultrametrischen Baum.

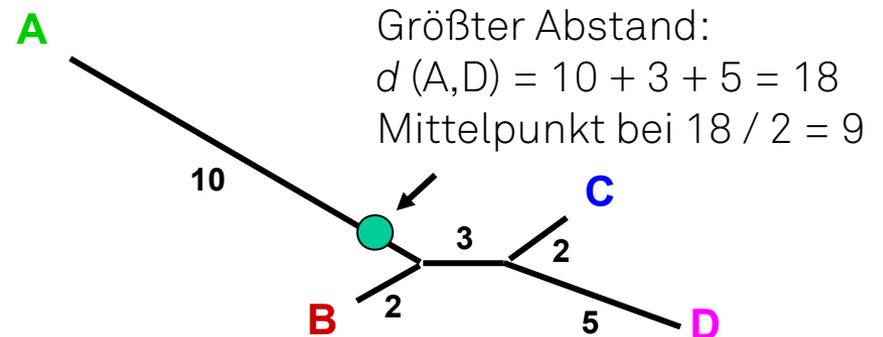
## Zwei Methoden zum Wurzeln eines Baums

(1) Mit Hilfe einer „outgroup“:  
Spezies, die bekanntermaßen außerhalb der betrachteten Gruppe liegt.

- erfordert Vorwissen
- relativ zuverlässig



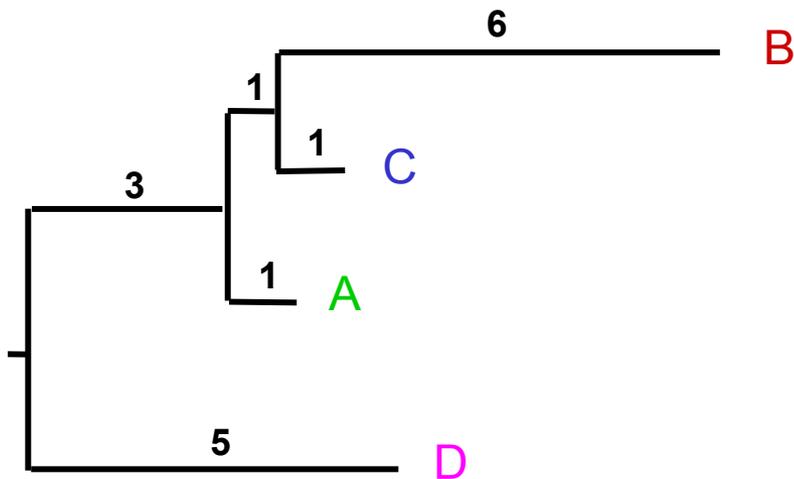
(2) Als Mittelpunkt der am weitesten voneinander entfernten Taxa (einfach, relativ unsicher)



## Ähnlichkeit ist nicht Verwandtschaft

Ähnlichkeit, Distanz: beobachtbar, messbar.

Verwandtschaft: historische Tatsache, kann nicht mehr beobachtet werden.



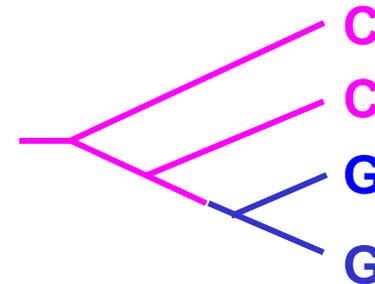
C ist ähnlicher zu A ( $d = 3$ ) als zu B ( $d = 7$ ),  
aber C und B sind am nächsten verwandt.

Das heißt, C und B hatten einen  
späteren gemeinsamen Vorfahren  
als C oder B mit A.

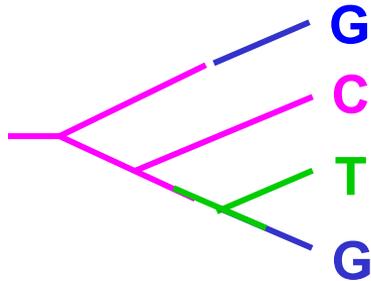
### 3 Gründe für beobachtete Ähnlichkeit

(1) Evolutionäre Verwandtschaft

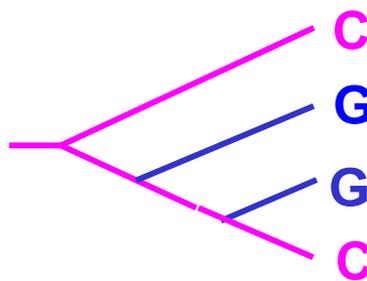
- gemeinsame Stamm-Merkmale (C)
- gemeinsame abgeleitete Merkmale (G)



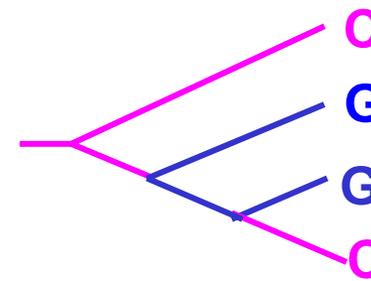
(2) parallele unabhängige Entwicklung von Merkmalen (Homoplasie)



konvergente Mutation zu G



parallele Mutation zu G



Rückmutation zu C

(3) geringe Evolutionsraten (hoher negativer selektiver Druck)

## Methoden der Phylogenetik

Unterscheidung nach:

- Art der Daten (Merkmale oder Distanzen)
- Prinzip des Algorithmus

		ALGORITHMUS BASIERT AUF	
		einem Optimalitätskriterium	Clusteringverfahren
ART DER DATEN	Merkmale	PARSIMONY (Sparsamkeit) MAXIMUM LIKELIHOOD	
	Distanzen	MINIMUM EVOLUTION LEAST SQUARES (kleinste Quadrate)	UPGMA NEIGHBOR-JOINING

## Merkmals-basierte Methoden

### Merkmale:

- Objekte, die einen von mehreren Zuständen annehmen können, z.B. Spalte in einem multiplen Alignment von DNA-Sequenzen.
- In jedem Taxon liegt eine **Merkmalsausprägung** vor.
- Merkmale (Alignments) werden direkt zur Baumkonstruktion verwendet.
- Merkmalsausprägung der Ahnen kann geschätzt werden.
- Zeitpunkte evolutionärer Ereignisse können geschätzt werden.
- Häufig schwierige (NP-schwere) resultierende Probleme

Taxa	Merkmale
Species A	ATGGCTATTCTTATAGTACG
Species B	ATCGCTAGTCTTATATTACA
Species C	TTCACTAGACCTGTGGTCCA
Species D	TTGACCAGACCTGTGGTCCG
Species E	TTGACCAGTTCTCTAGTTCG

## Distanz-basierte Methoden

Merkmale werden mit mathematischem Modell ("evolutionärer Markovprozess")  
in paarweise Distanzen zwischen Taxa umgerechnet.  
Nur Distanzen werden zur Baumkonstruktion verwendet.  
(Merkmale werden "vergessen".)

- vielseitig (auch außerhalb der Phylogenetik) verwendbare Methoden
- Evolutionäre Ereignisse und Ahnen-Merkmale können nicht geschätzt werden.

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>Species A</b>	----	<b>0.20</b>	<b>0.50</b>	<b>0.45</b>	<b>0.40</b>
<b>Species B</b>	<b>0.23</b>	----	<b>0.40</b>	<b>0.55</b>	<b>0.50</b>
<b>Species C</b>	<b>0.87</b>	<b>0.59</b>	----	<b>0.15</b>	<b>0.40</b>
<b>Species D</b>	<b>0.73</b>	<b>1.12</b>	<b>0.17</b>	----	<b>0.25</b>
<b>Species E</b>	<b>0.59</b>	<b>0.89</b>	<b>0.61</b>	<b>0.31</b>	----

Zwei berechnete Distanzmatrizen

rechts: Anteil Unterschiede, links: PAM-Schätzung / 100

## Optimierungs-Methoden

Kriterium (Eigenschaft des zu berechnenden Baumes) wird definiert, das es zu optimieren gilt.

### Beispiele:

- möglichst wenig Mutationsereignisse insgesamt im Baum (Merkmals-basiert)
- möglichst wenig Diskrepanz zwischen gegebenen Distanzen und den Distanzen, die die Baumtopologie impliziert (Distanz-basiert)

### Aufgabe des Algorithmus:

Finde einen (den) Baum, der das Kriterium optimiert.

- Unterscheide exakte Algorithmen vs. Heuristiken.
- Kann Ausgaben (Güte) verschiedener Algorithmen miteinander vergleichen:  
"Baum A ist (bezüglich des Kriteriums) besser als Baum B".

## Clustering-Methoden

Definiere Regeln,  
nach denen Taxa zu (2er-)Bäumen zusammenzufassen sind,  
diese wiederum zu größeren Bäumen, etc.

Liefern immer einen einzigen Baum;  
dieser kann nicht gegenüber Alternativen bewertet werden.  
(Es gibt kein Optimalitätskriterium.)

### **Achtung:**

Weder Optimierungs-Methoden noch Clustering-Methoden  
garantieren evolutionär korrekten Baum.

## Ausgewählte Verfahren in der Übersicht

		ALGORITHMUS BASIERT AUF	
		einem Optimalitätskriterium	Clusteringverfahren
ART DER DATEN	Merkmale	PARSIMONY (Sparsamkeit) MAXIMUM LIKELIHOOD	
	Distanzen	MINIMUM EVOLUTION LEAST SQUARES (kleinste Quadrate)	UPGMA NEIGHBOR-JOINING

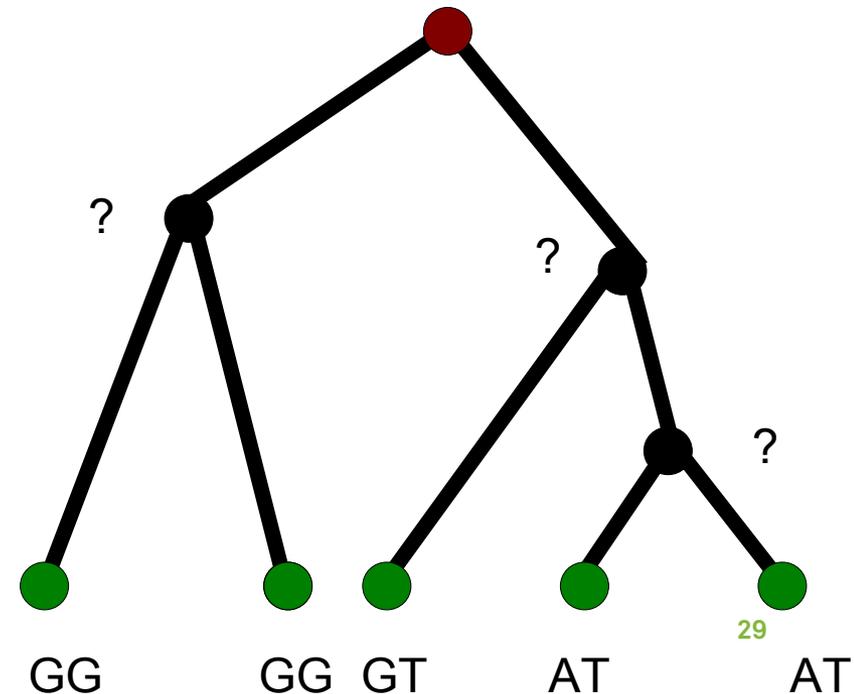
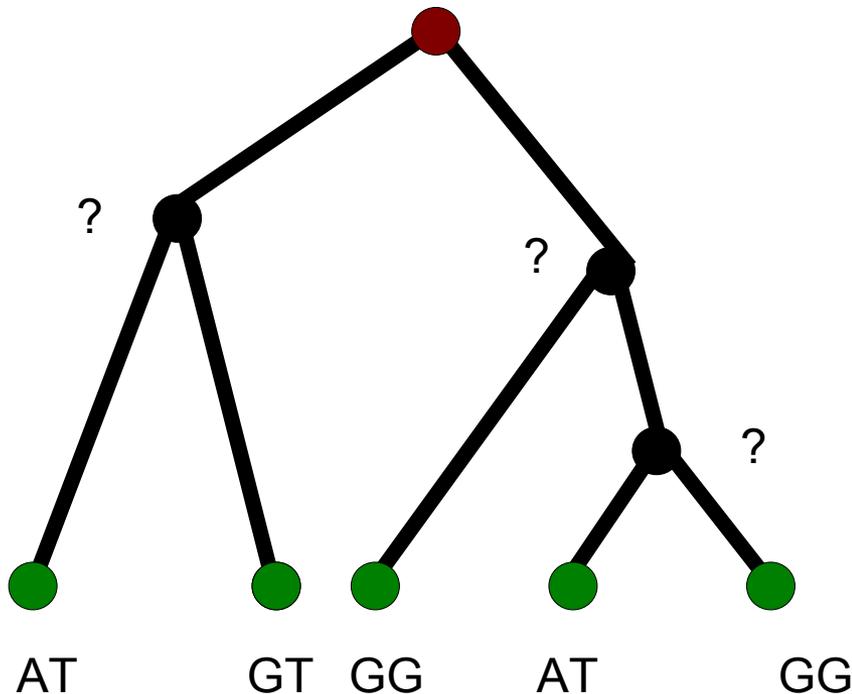
## Einfaches Beispiel zu Maximum Parsimony und ML

Gegeben sind 5 Sequenzen: AT, AT, GT, GG, GG

Zwei Baumtopologien:

Welche Sequenzen stehen jeweils an den inneren Knoten?

Welcher Baum ist besser?



## Maximum Parsimony

### Gesucht / Ausgabe:

Baumtopologie, Ahnen-Sequenzen in den inneren Knoten & Wurzel

Optimalitätskriterium: “sparsamster” Baum (“most parsimonious” tree):  
über alle Kanten summiert am wenigsten Änderungen.

### Vorteile

- Kriterium intuitiv, motiviert durch Occam's Razor:  
„die einfachste, kürzeste Erklärung ist die beste“.

### Nachteile:

- Unterschätzt die wahre Anzahl von evolutionären Ereignissen  
bei entfernt verwandten Sequenzen (z.B. wegen Homoplasien)
- liefert falsche Ergebnisse  
bei weit entfernten Sequenzen und stark unterschiedlichen Raten
- NP-schweres Problem

## Maximum Likelihood

### Gesucht / Ausgabe:

Baumtopologie, Kantenlängen (in PAM-Einheiten o.ä.).

Optimalitätskriterium:

Zu gegebenem statistischem Evolutionsmodell + Baumtopologie + Kantenlängen berechnete Wahrscheinlichkeit, dass beobachtete Sequenzen auftreten.

### Vorteile

- Modellbasiert: Alle Annahmen werden explizit gemacht
- Mehrfach- und Rücksubstitutionen sind im Modell berücksichtigt
- Konsistente Schätzung evolutionärer Distanzen (Kantenlängen)

### Nachteile:

- Ergebnisse Modell-abhängig; falsches Modell führt zu beliebigen Ergebnissen
- schwierig(er) zu verstehen (als MP)
- NP-schweres Problem

## Ausgewählte Verfahren in der Übersicht

		ALGORITHMUS BASIERT AUF	
		einem Optimalitätskriterium	Clusteringverfahren
ART DER DATEN	Merkmale	PARSIMONY (Sparsamkeit) MAXIMUM LIKELIHOOD	
	Distanzen	MINIMUM EVOLUTION LEAST SQUARES (kleinste Quadrate) Programme: fitch, kitch	UPGMA NEIGHBOR-JOINING

## Least Squares und Minimum Evolution

**Gesucht / Ausgabe:** Baumtopologie, Kantenlängen

Optimalitätskriterium: Zu jeder Baumtopologie werden Kantenlängenbestimmt, so dass die Baum-Distanzen die gegebenen Distanzen optimal approximieren (im Sinne kleinster Quadrate).

Optimal ist die Baumtopologie

- mit kleinstem quadratischen Fehler (bei least squares)
- mit kleinster Gesamtlänge (bei minimum evolution)

### Vorteile

- optimalitäts-basiert („Lösungen“ können vergleichend evaluiert werden)
- schneller als Merkmals-basierte Methoden

### Nachteile:

- langsamer als Clustering-Methoden NJ und UPGMA

## Ausgewählte Verfahren in der Übersicht

		ALGORITHMUS BASIERT AUF	
		einem Optimalitätskriterium	Clusteringverfahren
ART DER DATEN	Merkmale	PARSIMONY (Sparsamkeit) MAXIMUM LIKELIHOOD	
	Distanzen	MINIMUM EVOLUTION LEAST SQUARES (kleinste Quadrate)	UPGMA NEIGHBOR-JOINING

## UPGMA: Unweighted Pair Group Method Using Averaging

**Gesucht / Ausgabe:** Baumtopologie, Kantenlängen

**Verfahren:**

Solange mehr als ein Objekt vorhanden ist:

Finde das Paar  $(x,y)$  von Objekten mit kleinster Distanz.

Ersetze sie durch ein einziges Objekt  $z=\{x,y\}$ .

Berechne Distanzen von  $z=\{x,y\}$  zu den anderen Objekten  $a, \dots$ ,  
mittels Durchschnittsbildung zwischen  $d(x,a)$  und  $d(y,a)$ .

Baum ergibt sich aus der Hierarchie der zusammengefassten Objekte.

**Vorteile**

- einfach zu verstehen, einfach durchzuführen.
- liefert korrektes Resultat bei ultrametrischen Distanzen,  
d.h. wenn es einen ultrametrischen Baum gibt, der die gegebenen Distanzen  
als Distanzen zwischen den Blättern aufweist, wird dieser (schnell) gefunden.

**Nachteile**

- Ergebnisse schlecht interpretierbar bei nicht ultrametrischen Eingaben.

## NJ: Neighbor Joining

**Gesucht / Ausgabe:** Baumtopologie, Kantenlängen

### Verfahren:

- ähnlich wie bei UPGMA
- zu aggregierendes Objektpaars wird „sorgfältiger“ gewählt: berücksichtigt alle Distanzen, nicht nur die kleinste.
- Baum ergibt sich aus der Hierarchie der zusammengefassten Objekte.

### Vorteile

- korrektes Resultat bei additiven Distanzen, d.h. wenn es einen (ungewurzelten) Baum gibt, der die gegebenen Distanzen als Distanzen zwischen den Blättern aufweist, wird dieser gefunden.
- schnell ( $O(n^3)$ ), verbesserte Version Fast-NJ  $O(n^2)$  für  $n$  Taxa)

### Nachteile

- Ergebnisse schlecht interpretierbar bei nicht additiven Eingabe-Distanzen. 36

## Empfehlung

**Distanzbasierte Clustering-Methoden:** gut und schnell

- schnell und gut, wenn das richtige Distanz-Modell verwendet wurde.
- Es empfiehlt sich, eine NJ-Variante zu benutzen, nicht UPGMA.

**Probabilistische Methoden** (ML, Bayes'sche Methoden): häufig gut

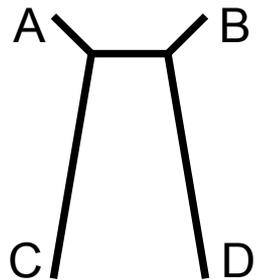
- langsam (versuchen NP-schwere Probleme zu lösen),
- können bei guter Modellwahl bessere Ergebnisse liefern als NJ.
- können alternative Bäume und Konfidenzwerte berechnen.

**Parsimony-Methoden:** langsam, fehleranfällig

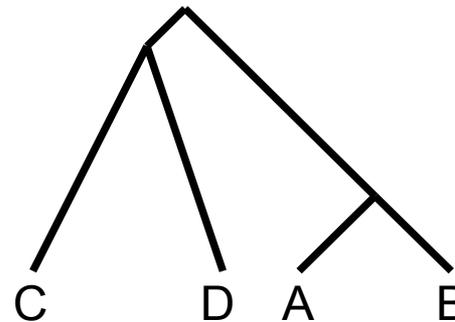
- unterschätzen den evolutionären Abstand (ok bei nah verwandten Sequenzen)
- keine Kantenlängen / Zeitschätzungen
- langsam (NP-schweres Problem)
- Problem der **long branch attraction**

## Long Branch Attraction bei Parsimony-Methoden

Wahrer Baum



von MP berechneter Baum



### Problem

lange Kanten = schnell evolvierende Ahnenlinien,  
dort Wahrscheinlichkeit konvergenter Evolution (Homoplasie) erhöht.  
MP misinterpretiert dies als Synapomorphie (gemeinsamer Ursprung),  
da die Kantenlänge bei MP nicht beachtet wird.  
=> Taxa an langen Ästen erscheinen verwandter als sie sind.

## Software

Umfangreiche Sammlung phylogenetischer Software:

<http://evolution.genetics.washington.edu/phylip/software.html>

The image shows a screenshot of the 'Phylogeny Programs' website. The page features a large grid of colorful icons representing various phylogenetic software packages. The text 'Phylogeny Programs' is prominently displayed in the center of the grid. Above the grid, there are six navigation tabs: 'Methods', 'By computer', 'Cross-referenced', 'Data types', 'New programs', and 'Submitting'. Below the grid, there are six more navigation tabs: 'Changes', 'Waiting list', 'Other lists', 'Old programs', 'Not listed', and '???'.

## Software: Phylip

- häufig benutztes Paket zum Erstellen von Phylogenien
- enthält zahlreiche Programme für verschiedene Aufgaben
  
- Projekt-Homepage: <http://evolution.genetics.washington.edu/phylip.html>
- Web-Interface zu einer Teilmenge von Phylip + mehr: <http://mobylye.pasteur.fr>