

# Bioinformatik

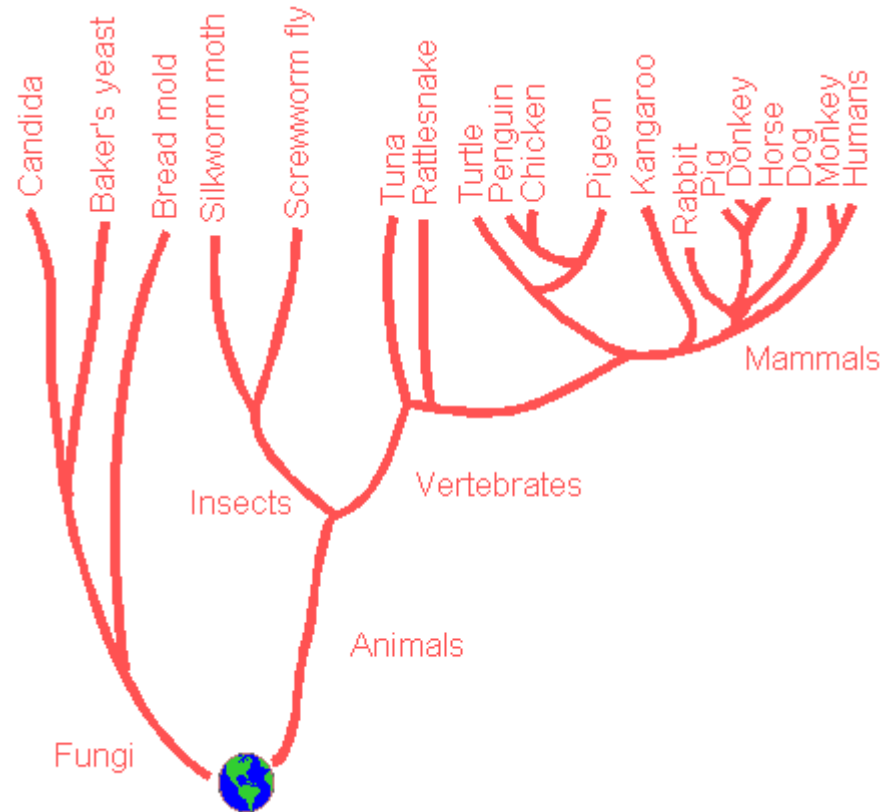
Distanzbasierte phylogenetische Algorithmen

Ulf Leser

Wissensmanagement in der  
Bioinformatik



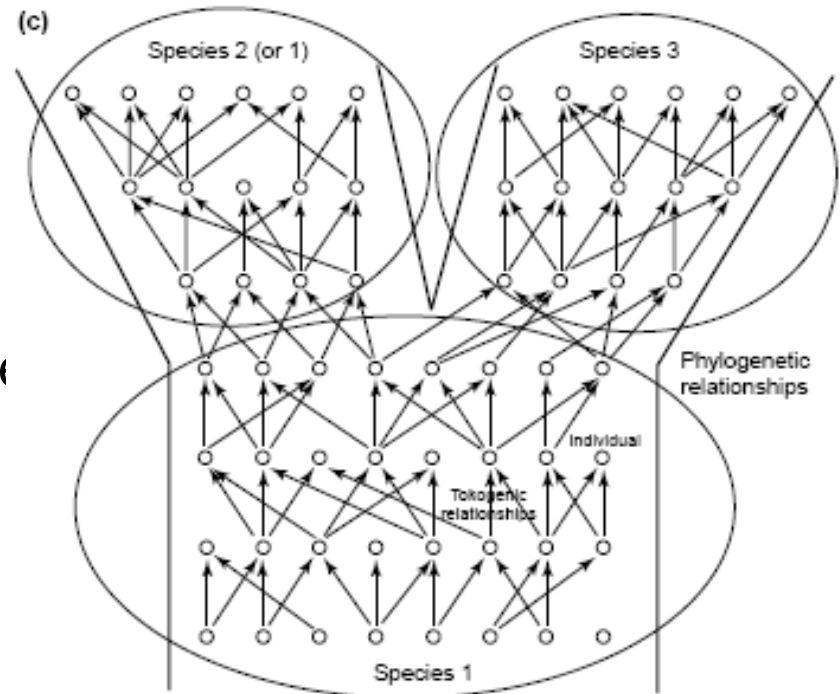
# Phylogenetische Bäume



Stammbaum (*Phylogenetic Tree*)

# Arten von Stammbäumen

- Individuelle Abstammung
  - Stammbäume, Ahnentafeln
  - Natürlich **kein Baum**: Zwei Eltern
  - Rekombination
- Speziesstammbäume
  - Ein Baum, wenn **Spezies nicht verschmelzen** können
  - Sprachen verschmelzen
- Gene Trees
  - Geschichte eines Sequenzabschnitts
    - Nicht leicht zu definieren
  - Baumförmig, wenn Gene nicht verschmelzen
  - Aber: 2 Allele jedes Gens vorhanden (Besser: **Haplotype Tree**)



TRENDS in Ecology & Evolution

# Evolutionsmodell

---

- Lebewesen vermehren sich durch Kopieren
- **Kopieren** mit kleinen Fehlern (Mutationen)
  - Führen (manchmal) zu veränderter „Funktion“
- **Selektion** – „Survival of the fittest“
  - Fatale Mutationen lassen Nachkommen gar nicht erst entstehen
  - Einschränkungen der Überlebensfähigkeit führen zur schnellen Ausrottung
- **Speziation**
  - Unterschiedliche Mutationen sind in unterschiedlichen Lebensräumen unterschiedlich vorteilhaft
    - **Anpassung** an ökologische Nischen
  - Führt zur Akkumulation von Änderungen
  - Schließlich geht die gemeinsame Fortpflanzungsfähigkeit verloren

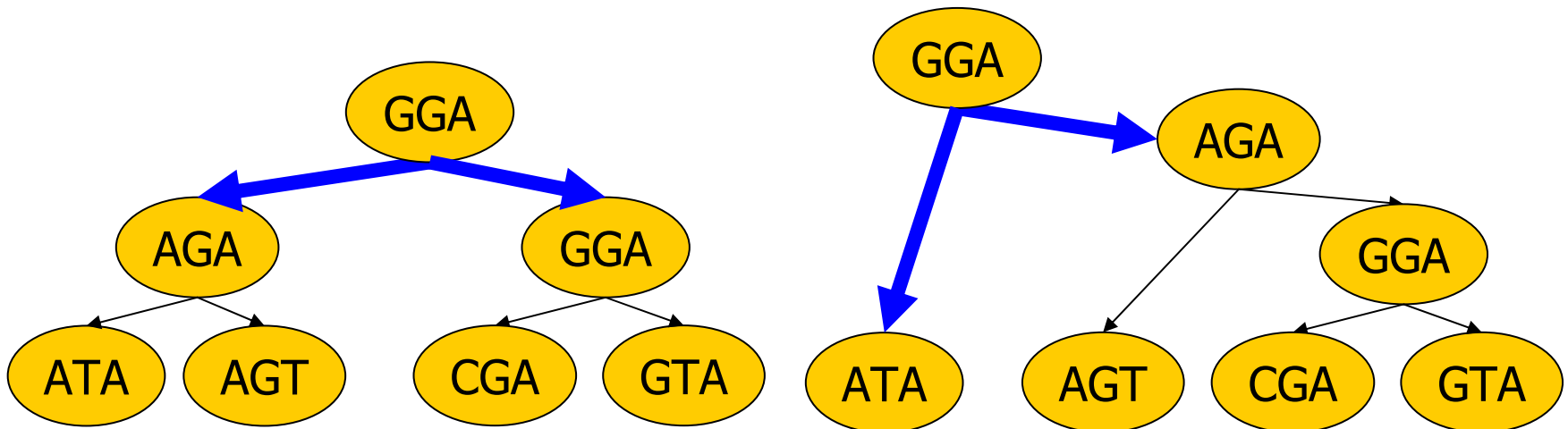
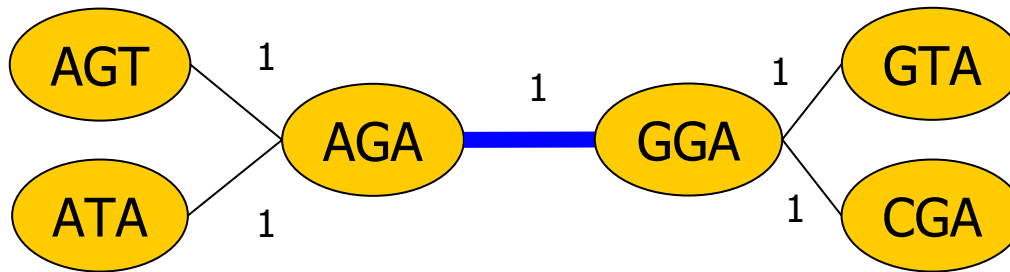
# Molecular Clock

---

- Häufige Annahme: **Molecular Clock**
  - Mutationen setzen sich bzgl. der Zeit immer mit gleicher Häufigkeit durch
  - Unabhängig von Teilbaum, Zeitpunkt, Ort und Art der Mutation
  - Anzahl der beobachteten Mutationen proportional zur vergangenen Zeit
- Die ist manchmal hilfreich, aber falsch
  - Unterschiedliche Zeiten erzeugen unterschiedlichen Selektionsdruck
  - Unterschiedliche Sequenzabschnitte unterliegen unterschiedlichem Selektionsdruck
  - Unterschiedliche Spezies unterliegen unterschiedlicher Frequenz erfolgreicher Mutationen
  - ...

# Bäume ohne Wurzeln

- Kanten symbolisieren Veränderungen
- Wir können nur Blätter vergleichen
- Damit ist idR keine Richtung der Entwicklung berechenbar



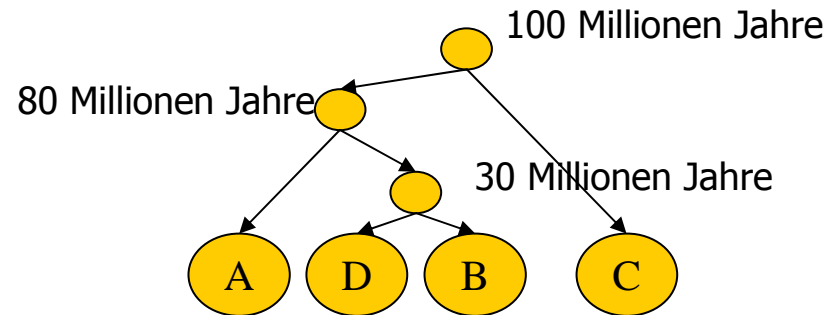
# Inhalt dieser Vorlesung

---

- Distanzbasierte Phylogenie
- Ultrametrien und UPGMA
- Additive Bäume und Neighbor Joining

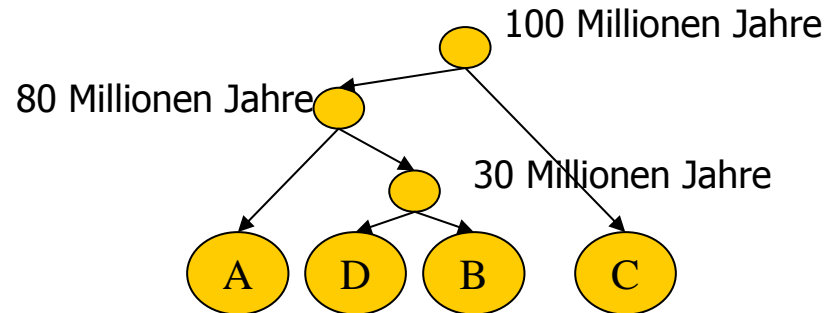
# Voraussetzungen

---



- An innere Knoten eines Stammbaum kann man den Speziationszeitpunkt schreiben („branch points“)
- Wenn die Molecular Clock Theory gilt
  - Ist die Menge an Veränderungen auf einer Kante **proportional zu der verstrichenen Zeit**
  - Damit proportional zur Länge der Kante
  - Damit ist der Editabstand zweier Knoten (=Evolution dazwischen) proportional zur Summe der **Editabstände beider Knoten zum kleinsten gemeinsamen Vater**

# Ultrametrien



- Wenn man den Baum und die Zeitpunkte weiß, dann gilt
  - Alle Zahlen auf einem Pfad von der Wurzel zu einem beliebigen Blatt nehmen strikt ab
  - Der **Zeitpunkt der Aufspaltung** ist ein **Abstandsmaß** für zwei Arten
    - Für Blätter  $X, Y$  sei  $d(X, Y)$  das Label des kleinsten gemeinsamen Vorfahren
    - Im Beispiel:  $d(A, B) = 80$ ,  $d(B, C) = 100$ ,  $d(A, D) = 80$
  - Das ist eine Metrik
    - $d(X, X) = 0$ ,  $d(x, y) > 0$ ,  $d(X, Y) = d(Y, X)$ , und  $d(X, Y) \leq d(X, Z) + d(Z, Y)$
  - Es ist sogar eine **Ultrametrik** (gleich)

# Alles ganz einfach?

---

- Wir sehen gleich
  - Schreibt man diese Abstände in eine Ähnlichkeitsmatrix, hat diese bestimmte Eigenschaften
  - Jede Matrix mit diesen Eigenschaften entspricht genau einem Baum und umgekehrt
  - Den Baum kann man aus der Matrix effizient berechnen
- Also ist das Phylogenie-Problem überhaupt kein Problem?
- Doch
  - **Molecular Clock Theory stimmt nicht**
  - Editabstand korreliert direkt nicht mit der vergangenen Zeit
  - Damit kennen wir die Speziationszeitpunkte nicht
  - Man kann das alles nur **approximieren**
- Wenn die Molecular Clock Theory aber gelten würde ...

# Ultrametrische Bäume

---

- Definition

*Sei  $T$  ein Baum und  $D$  eine symmetrische Matrix mit  $n$  Zeilen und  $n$  Spalten.  $T$  heißt **ultrametrischer Baum für  $D$**  wenn gilt:*

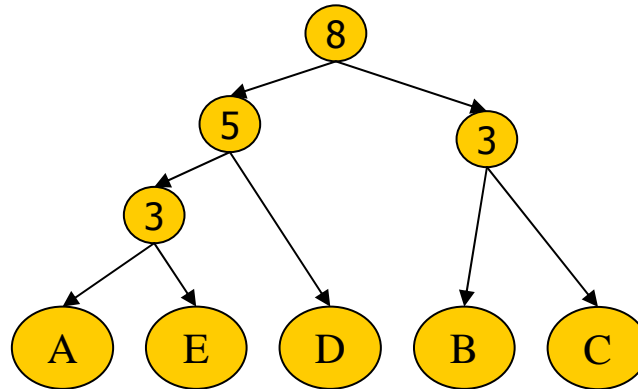
- *$T$  hat  $n$  Blätter, beschriftet mit den Zeilen von  $D$*
- *Jeder innere Knoten von  $T$  hat zwei Kinder und ist mit einem Wert aus  $D$  beschriftet*
- *Auf jedem Pfad von der Wurzel zu einem Blatt in  $T$  sind die Zahlen strikt abnehmend*
- *Für alle Blätter  $i, j$  mit  $i \neq j$  gilt: der kleinste gemeinsame Vorfahr von  $i$  und  $j$  ist mit  $D(i, j)$  beschriftet*

- Bemerkung

- Jeder Stammbaum ist ultrametrisch für die Abstandsmatrix mit den Aufsplittzeitpunkten als Abstandsmaß

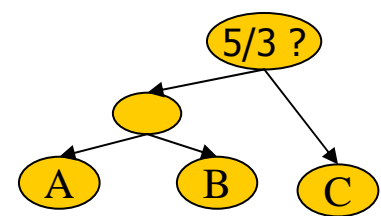
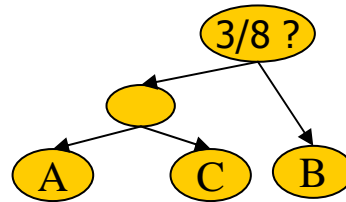
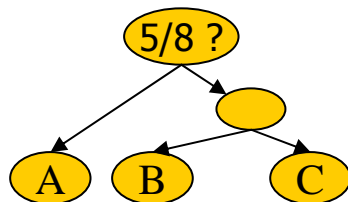
# Beispiel

	A	B	C	D	E
A		8	8	5	3
B			3	8	8
C				8	8
D					5
E					



Das geht nicht immer

	A	B	C
A		8	5
B			3
C			



# Überlegungen

---

- Das kann auch nicht immer gehen
  - Matrix hat  $(n^2-n)/2$  relevante Zellen
  - Baum hat nur  $n-1$  innere Knoten
  - Eine Matrix, zu der man einen ultrametrischen Baum konstruieren kann, muss also Duplikate enthalten
- *Definition*

*Eine symmetrische Matrix  $D$  mit  $n$  Spalten und Zeilen ist **ultrametrisch**, wenn für beliebige Zeilen  $i, j, k$  gilt, dass das Maximum von  $D(i,j)$ ,  $D(j,k)$  und  $D(i,k)$  genau zweimal vorkommt*
- *Bemerkung*
  - Also entweder
    - $D(i,j)=D(j,k)$  und  $D(i,j)>D(i,k)$
    - $D(i,j)=D(i,k)$  und  $D(i,j)>D(j,k)$
    - $D(j,k)=D(i,k)$  und  $D(j,k)>D(i,j)$

# Ultrametrik

---

- Definition

*Eine Ultrametrik ist eine Metrik für die gilt:*

$$d(a,c) \leq \max(d(a,b), d(b,c))$$

- Bemerkung

– Für Metriken muss nur gelten:  $d(a,c) \leq d(a,b) + d(b,c)$

# Von der Matrix zum Baum und zurück

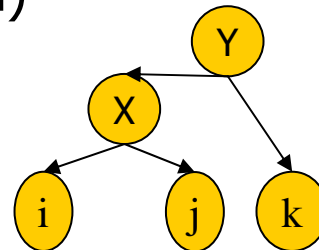
---

- Theorem

*Eine symmetrische Matrix  $D$  hat einen ultrametrischen Baum gdw.  $D$  selber ultrametrisch ist*

- Beweis

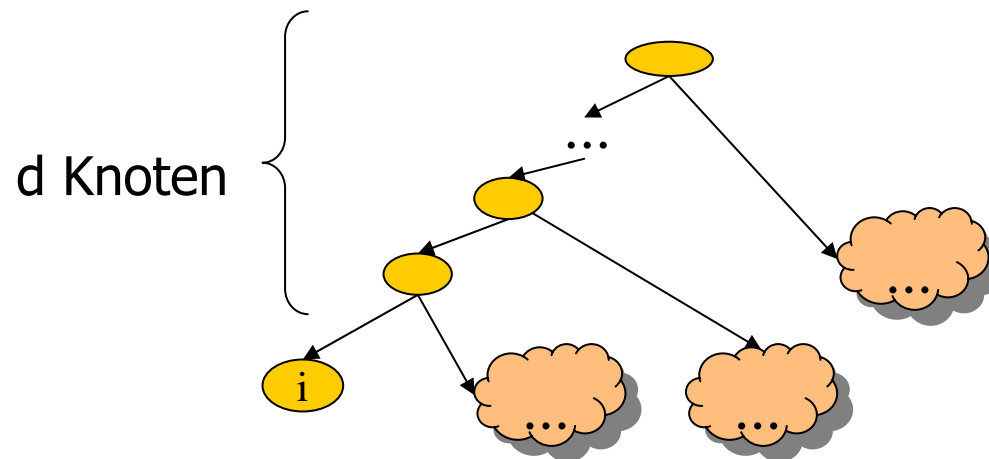
- (1) Nehmen wir erst an, dass zu  $D$  ein ultrametrischer Baum  $T$  existiert. Nehmen wir an, dass  $i, j, k$  wie folgt liegen (alle anderen Knoten können wir ignorieren)



- Dann gilt offensichtlich  $D(i,k)=D(j,k)=Y$  und  $D(i,k)>D(i,j)=X$
- Dito kann man für die zwei anderen Varianten der möglichen Lage von  $i, j, k$  verfahren
- Das gilt für alle Tripel von Knoten
- Also ist  $D$  ultrametrisch

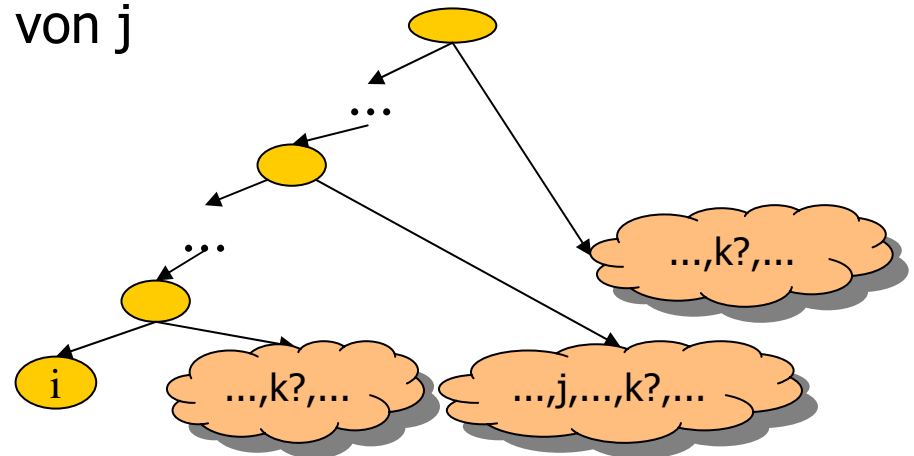
# Beweis Gegenrichtung

- (2) Nehmen wir an, dass  $D$  ultrametrisch ist. Wir konstruieren einen ultrametrischen Baum  $T$  aus  $D$ 
  - Betrachten wir eine beliebige Zeile  $i$ . Diese entspricht Blatt  $i$  in  $T$
  - $i$  hat diverse kleinste gemeinsame Vorfahren mit den  $n-1$  anderen Blättern, die beschriftet werden müssen mit den Werten  $D(i,x)$ ,  $x \neq i$ . Und das in sortierte Reihenfolge
  - Der Pfad von  $i$  zur Wurzel muss nicht  $n-1$  Knoten enthalten, denn die kleinsten gemeinsamen Vorfahren mit verschiedenen anderen Knoten sind oft identisch. Nehmen wir an, dass es auf dem Pfad  $d$  verschiedene Werte gibt ( $d \leq n-1$ )



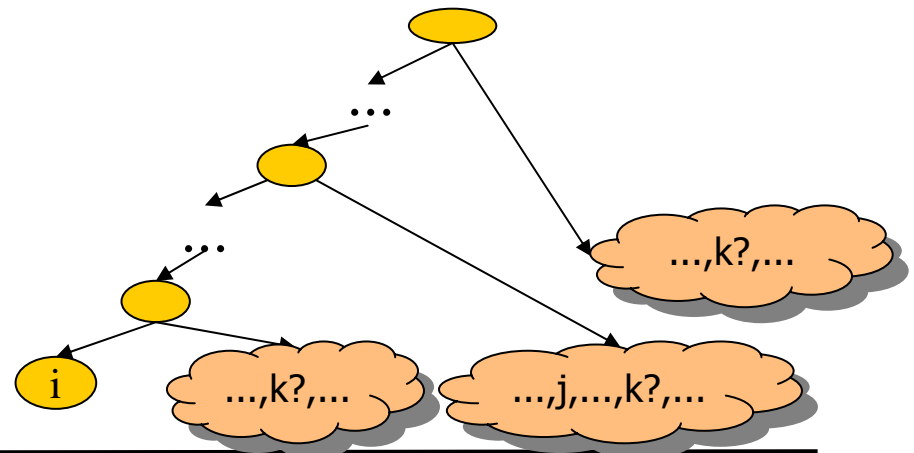
# Beweis Gegenrichtung -2-

- Die Menge aller Blätter (ohne  $i$ ) zerfällt also in  $d$  Klassen
  - Alle Knoten einer Klasse befinden sich in einem Teilunterbaum bzgl. des Pfades von  $i$  zur Wurzel
  - Alle Knoten einer Klasse haben den selben Abstand zu  $i$
- Betrachten wir ein Blatt  $j \neq i$  und ein beliebiges anderes Blatt  $k \neq i$ . Drei Möglichkeiten
  - $k$  liegt in der selben Klasse wie  $j$
  - $k$  liegt in einer Klasse „links“ von  $j$
  - $k$  liegt in einer Klasse „rechts“ von  $j$



# Beweis Gegenrichtung -3-

- Fall 1:  $j$  und  $k$  in der selben Klasse
  - Das gilt, wenn  $D(i,j)=D(i,k)$  und  $D(j,k)<D(i,j)$
- Fall 2:  $k$  liegt links von  $j$ 
  - Das gilt, wenn  $D(i,j)=D(k,j)$  und  $D(i,k)<D(i,j)$
- Fall 3:  $k$  liegt rechts von  $j$ 
  - Das gilt, wenn  $D(i,k)=D(j,k)$  und  $D(i,j)<D(j,k)$
- Damit können wir alle Knoten  $\neq i$  nach ihrem Abstand zu  $i$  in Klassen einsortieren
- Für jede der Klassen gehen wir rekursiv vor
- qed.



# Folgerung

---

- Der Beweis ist konstruktiv; man kann genau auf diese Weise einen ultrametrischen Baum bauen
- *Theorem*  
*Sei  $D$  eine ultrametrische Matrix. Dann gibt es genau einen ultrametrischen Baum  $T$  für  $D$ .*
- Beweis
  - Durch Konstruktion
- Kann in  $O(n^2)$  implementiert werden
  - Gleich

# Distanzbasierte Algorithmen

---

- Konstruktion des ultrametrischen Baumes basiert rein auf Distanzmassen
  - Einen ultrametrischen Baum gibt es nicht für alle Matrizen
  - Es gibt weniger empfindliche Verfahren (gleich)
- Die Geschichte einzelner „Sites“ wird nicht berücksichtigt
- Solche Algorithmen zur Berechnung von Stammbäumen nennt man **distanzbasiert**
- Alternative: **Merkmalsbasierte Verfahren**
  - Beachten jedes einzelne Merkmal (Sequenzposition)
  - Z.B. Perfect Phylogeny; Maximum Parsimony
  - Später

# UPGMA - Hierarchisches Clustering

---

- UPGMA
  - „Unweighted pair group method with arithmetic mean“
  - Anderer Name: [Hierarchisches Clustering](#)
- Sehr einfaches und allgemeines Verfahren, kann bei allen möglichen Problemen angewandt werden
- [Wenn eine Matrix ultrametrisch ist](#), dann findet UPGMA den dazugehörigen ultrametrischen Baum
  - UPGMA nimmt die Molecular Clock an – alle Pfade von einem Blatt zur Wurzel haben am Ende die selbe Länge
- **Achtung:** UPGMA konstruiert immer einen Baum
  - Auch wenn die Matrix nicht ultrametrisch ist

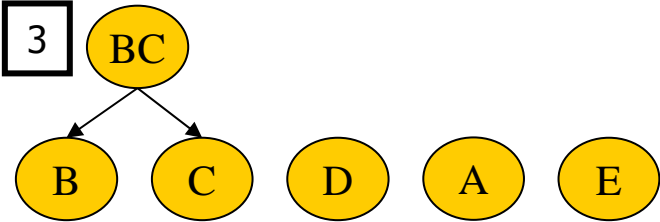
# UPGMA Verfahren

---

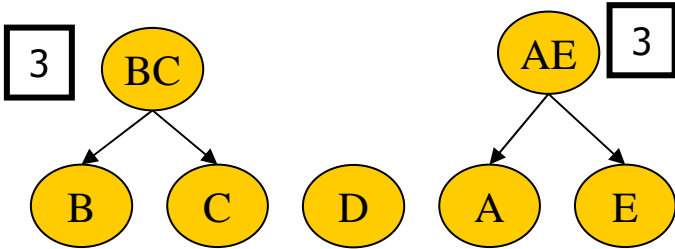
- Gegeben: Distanzmatrix  $D$
- Erzeuge ein „Baumgerüst“ mit  $n$  Blättern
- Wähle den kleinsten  $D(i,j)$  Wert der Matrix und verbinde die Knoten  $i$  und  $j$  durch einen neuen Knoten  $(ij)$  mit Beschriftung  $D(i,j)$  und Kanten zu  $i$  und zu  $j$ 
  - Anfangs sind  $i$  und  $j$  Blätter, später können es auch innere Knoten sein
- Lösche Zeilen und Spalten  $i$  und  $j$  aus  $M$
- Füge in  $D$  eine Zeile und eine Spalte  $(ij)$  hinzu mit  $D(ij,k) = (D(i,k)+D(j,k))/2$
- Wiederhole, bis  $D$  leer ist

# Beispiel

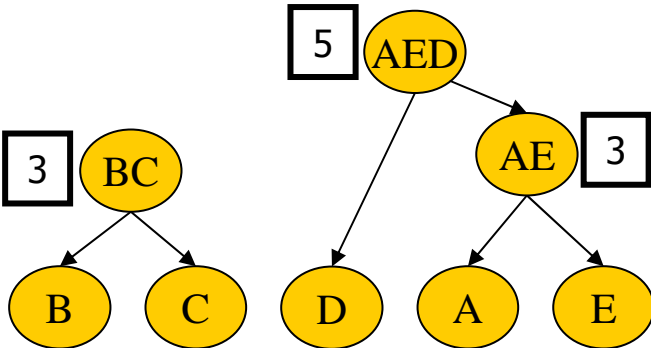
	B	C	D	E
A	8	8	5	3
B		3	8	8
C			8	8
D				5



	BC	D	E
A	8	5	3
BC		8	8
D			5

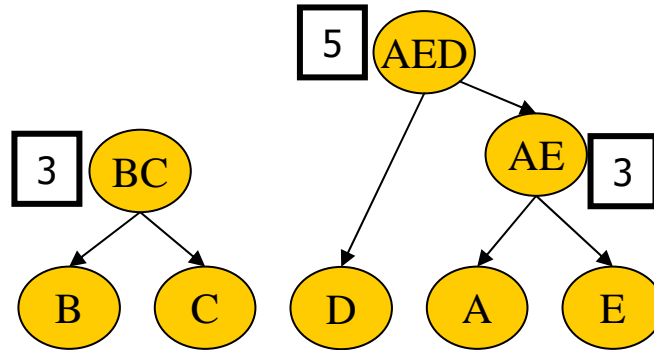


	BC	D
AE	8	5
BC		8

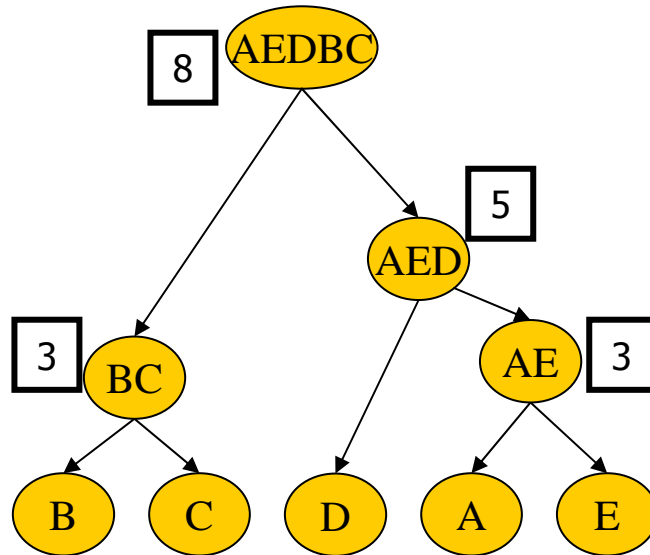


# Beispiel

	AE	BC	D
AE		8	5
BC			8



	BC
AED	8

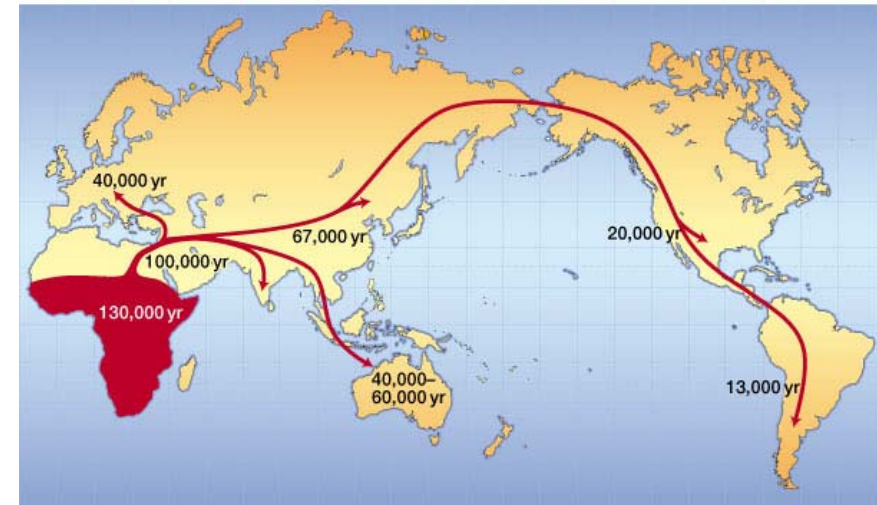
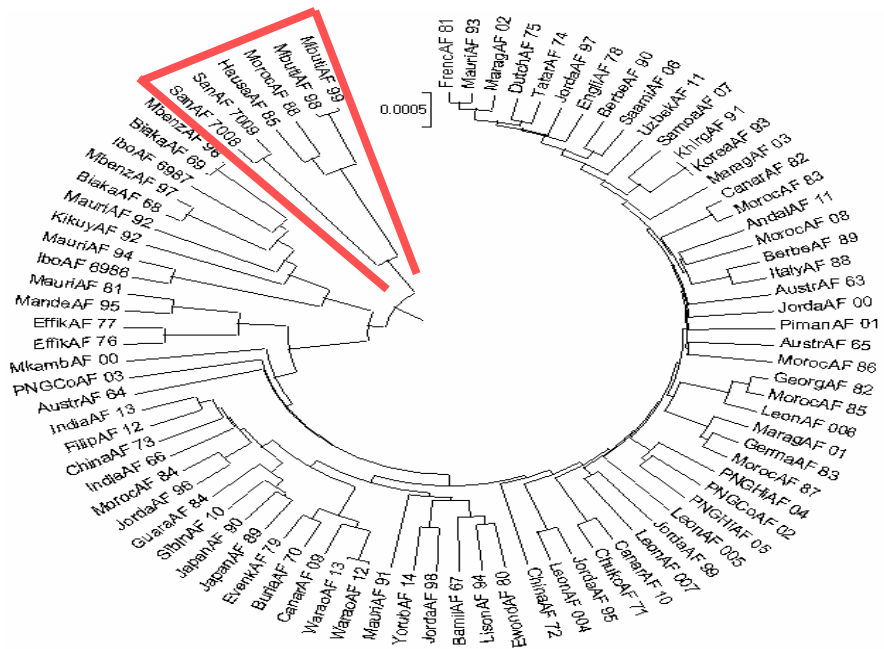


## Kontrolle

	B	C	D	E
A	8	8	5	3
B		3	8	8
C			8	8
D				5

# Anwendungsbeispiel

- Sequenzierung der mitochondrialen DNA (16 KB) von 86 geographisch verteilt lebenden Personen
- Ergebnis: Mitochondriale DNA scheint nach einer molekularen Uhr abzulaufen; Divergenz ist ca.  $1,7E-8$  pro Base und Jahr



Quelle:  
 Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. (2000)  
*Nature* 408: 708-713

Quelle:  
<http://www.genpat.uu.se/mtDB/sequences.html>

Methode: UPGMA

# Komplexität

---

- n Durchläufe
  - In jedem Durchlauf wird die Matrix um eine Zeile/Spalte kleiner
- $n+2$  Berechnungen
  - 2 Kantenlängen
  - $O(n)$  neue Einträge in der Matrix
- Also  $O(n^2)$ ?
- So nicht
  - Es bleibt das Problem, den kleinsten Eintrag in der Matrix zu finden
  - Damit sind wir bei  $O(n*(n*\log(n)+n)) = O(n^2\log(n))$
- Es gibt aber  $O(n^2)$  Algorithmen
  - Siehe Gusfield, Errata Webseite

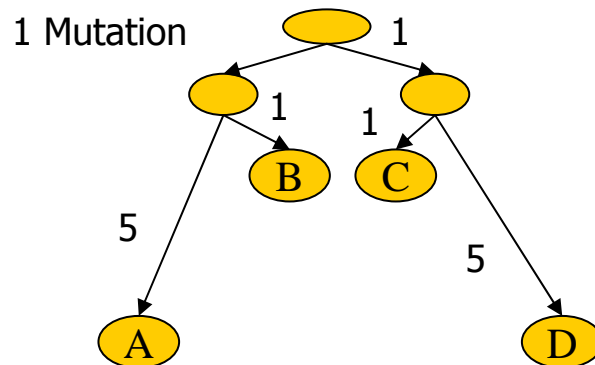
# UPGMA und Ultrametrien

---

- Theorem  
*Wenn eine Matrix ultrametrisch ist, berechnet UPGMA den dazu gehörenden ultrametrischen Baum.*
- Beweis
  - Literatur
- Aber: Wenn eine Matrix nicht ultrametrisch ist, berechnet UPGMA
  - ... auch einen Baum – aber wie gut ist der?
- Die **Molecular Clock Theory ist also Voraussetzung** für die (korrekte) Anwendung von UPGMA

# Ultrametrien und Sequenzabstände

- Realität
  - Matrix mit Editabständen der Sequenzen
- **Reale Daten** sind selten ultrametrisch
  - Spezies unterliegen spezifischem Selektionsdruck
- Wir können damit keine innere Knoten beschriften, sondern nur **Abstände von Blättern** messen
  - Aus denen wollen wir die „wahren“ Kantenlängen rückrechnen
  - Die Veränderungsanzahl ist dabei in der Regel nicht gleich verteilt
- **Beispiel: Stammbaum, aber keine Ultrametrik**



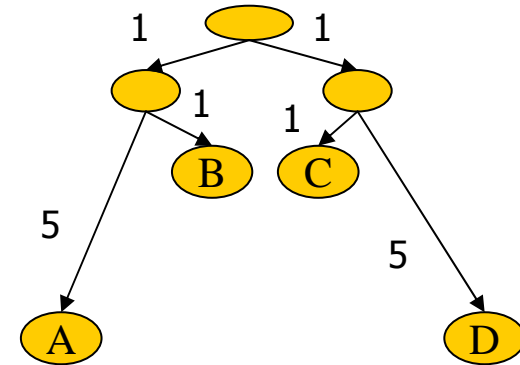
	A	B	C	D
A		6	8	12
B			4	8
C				6
D				

# Wo UPGMA irrt

Der echte Baum

	B	C	D
A	6	8	12
B		4	8
C			6
D			

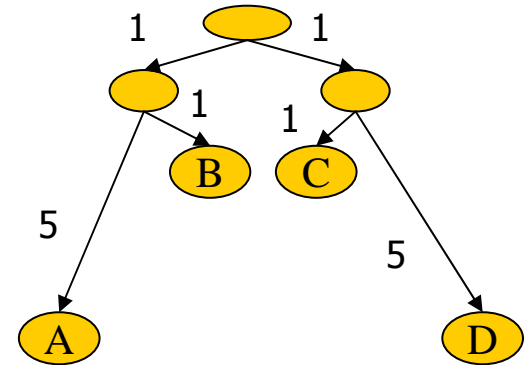
Was erzeugt UPGMA?



# Wo UPGMA irrt

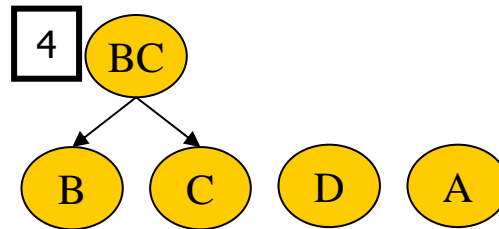
Der echte Baum

	B	C	D
A	6	8	12
B		4	8
C			6
D			



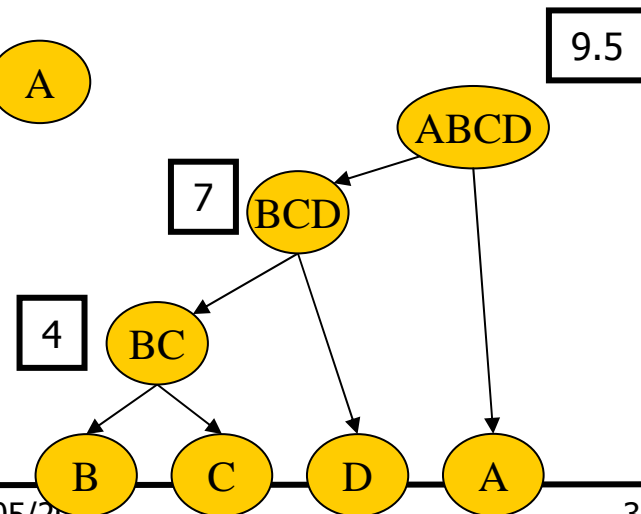
Was erzeugt UPGMA?

	B	C	D
A	6	8	12
B		4	8
C			6



	A	BC	D
A		7	12
BC			7

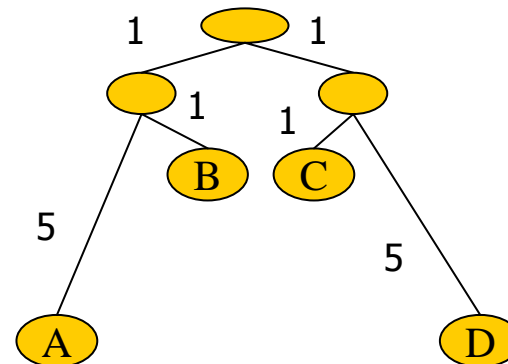
	A
BCD	9.5



# Additive Bäume

- Ultrametrien berechnen im Kern die Label der inneren Knoten
  - Und dadurch kann man Kanten nur auf bestimmte Weisen aufteilen
- Besser: Additive Bäume
  - Bäume mit **Labeln auf den Kanten**
- Eingabe ist wieder eine Ähnlichkeitsmatrix
  - Die Werte in der Matrix sind aber nicht die Kantenlabeln
  - Abstände enthalten sich gegenseitig
  - Das kann, muss aber nicht zu einem eindeutigen Baum führen
- Da es nur Abstände gibt, hat der Baum **keine Wurzel**
  - Gesucht ist eine Anordnung so, dass die Pfade im Baum die „richtigen“ Längen haben

	B	C	D
A	6	8	12
B		4	8
C			6
D			



# Additive Bäume

---

- Definition

*Sei  $D$  eine symmetrische Matrix mit  $n$  Spalten und Zeilen (und nur positiven Werten;  $D(i,i)=0$ ). Ein Baum  $T$  mit Kantenlabeln heißt **additiver Baum** für  $D$ , wenn gilt*

- *$T$  hat  $n$  Blätter, beschriftet mit den Zeilen von  $D$*
- *Für jedes Paar  $i,j$  ist  $D(i,j)$  gleich der Summe der Kantenlabel auf dem (eindeutigen) Pfad von  $i$  nach  $j$*

- Bemerkung

- Jede ultrametrische Matrix induziert einen additiven Baum aber nicht umgekehrt

# Matrizen und additive Bäume

---

- Fragen
  - Existiert zu jeder Matrix ein additiver Baum?
  - Wie findet man einen additiven Baum zu einer gegebenen Matrix?

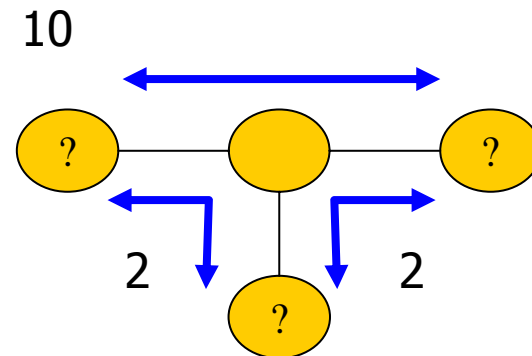
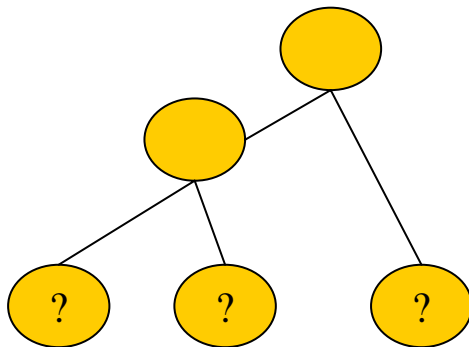
# Matrizen und additive Bäume

- Fragen

- Existiert zu jeder Matrix ein additiver Baum?
- Wie findet man einen additiven Baum zu einer gegebenen Matrix?

- Gegenbeispiel

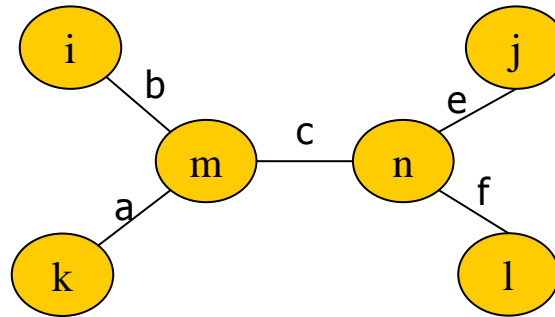
	A	B	C
A		10	2
B			2



# Beobachtung

---

- Betrachten wir beliebige vier Blätter  $i, j, k, l$ , eines Baumes mit Kantenlabeln



- In welchem Verhältnis stehen deren Abstände?
  - Die 6 Abstände setzen als Summen von 5 Kantenlabeln zusammen
- Es gilt zum Beispiel
  - $D(i,k)+D(j,l) \leq D(i,j)+D(k,l) = D(i,l)+D(k,j)$
  - Denn:  $(a+b) + (e+f) \leq (b+c+e) + (a+c+f)$
- Aber die Knoten können auch anders angeordnet sein

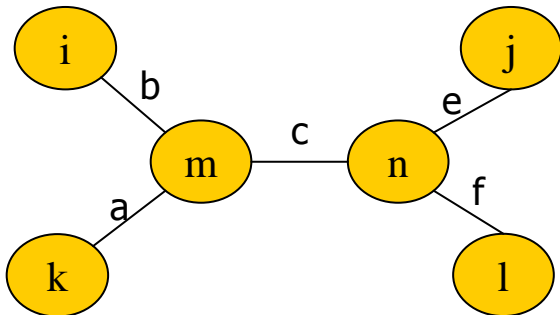
# 4-Punkt Bedingung

- Theorem

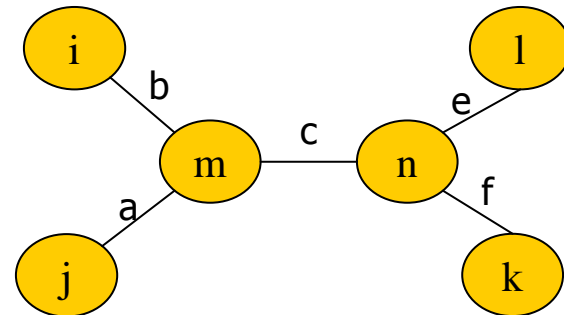
*Eine Matrix  $D$  hat einen additiven Baum gdw. für alle Zeilen  $i, j, k, l$  die 4-Punkt Bedingung gilt:*

$$D(i,k)+D(j,l) \leq \max( D(i,j)+D(k,l) , D(i,l)+D(k,j))$$

- Beweis: Literatur



$$(a+b) + (e+f) \leq \max( (b+c+e) + (a+c+f), (b+c+f) + (a+c+e) )$$

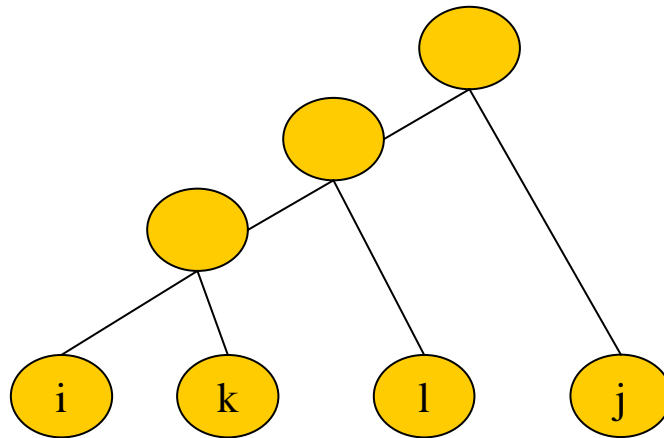


$$(b+c+f) + (a+c+e) \leq \max( (a+b) + (e+f), (b+c+e) + a+c+f )$$

# Aber ...

---

- Wenn die 4 Punkte aber so liegen?



# Neighbor-Joining

---

- Matrizen und Algorithmen
  - Ultrametrische Matrizen – UPGMA
  - Additive Matrizen – Neighbor Joining
- Hierarchisches Clusterverfahren (wie UPGMA)
  - Erzeugt einen binären Baum ohne Wurzel
  - Grundaufbau wie UPGMA
    - Beginne mit so vielen Clustern wie Blättern
    - Wähle nach bestimmtem Kriterium zwei Cluster
    - Verschmelze die zwei Cluster und verbinde Knoten im Baum
    - Iteriere, bis nur noch ein Cluster vorhanden ist
- Unterschiede
  - UPGMA wählt Cluster zur Verschmelzung nur nach Nähe zueinander
  - Neighbor Joining wählt Cluster nach der Nähe zueinander und dem Abstand zu anderen Clustern

# Verfahren

---

- Bilde aus jeder Zeile einen Cluster
- Berechne für jeden Cluster  $i$  den Abstand  $u_i$  zu allen anderen Clustern

$$u_i = \sum_{k \neq i} \frac{D(i, k)}{n - 2}$$

„nächster Nachbar“



- Suche das Clusterpaar  $(i, j)$ , für das gilt  $D(i, j) - u_i - u_j = \min_i$
- Erzeuge Cluster  $ij$  mit Kanten zu  $i$  und  $j$  mit Kantenlängen

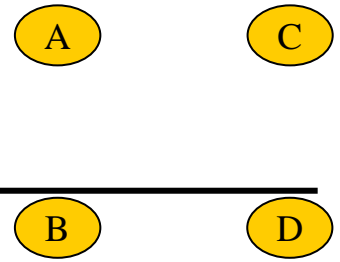
$$d(i, ij) = \frac{D(i, j) + u_i - u_j}{2} \quad d(j, ij) = \frac{D(i, j) + u_j - u_i}{2}$$

- Erzeuge neuen Clusterknoten mit Abständen zu anderen Clustern

$$D(k, ij) = \frac{D(i, k) + D(k, j) - D(i, j)}{2}$$

- Lösche Cluster  $i$  und  $j$
- Iteriere, solange mehr als ein Cluster existiert

# Beispiel (hier scheiterte UPGMA)



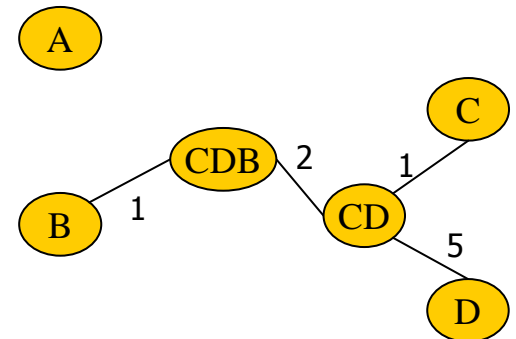
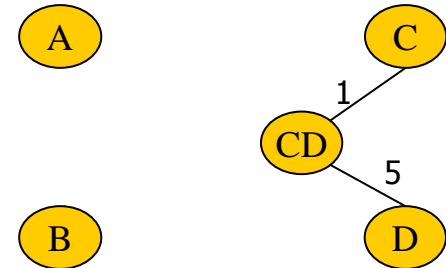
	A	B	C	D
A		6	8	12
B			4	8
C				6
$u_i$	13	9	9	13

NJ-Abstände der Clusterpaare

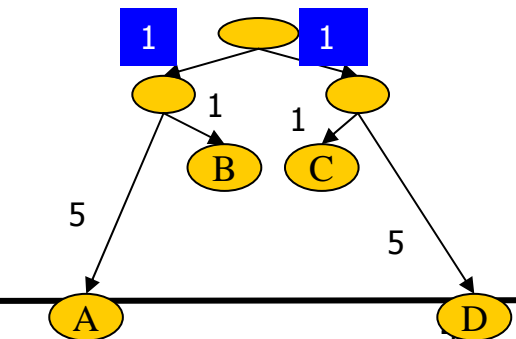
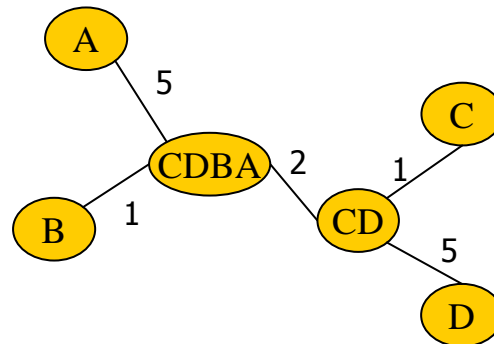
	B	C	D
A	-16	-14	-14
B		-14	-14
C			-16

	A	B	CD
A		6	7
B			3
$u_i$	13	9	10

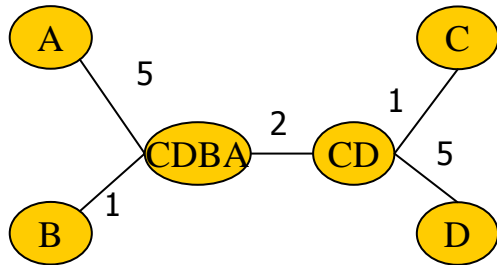
	B	CD
A	-16	-16
B		-16



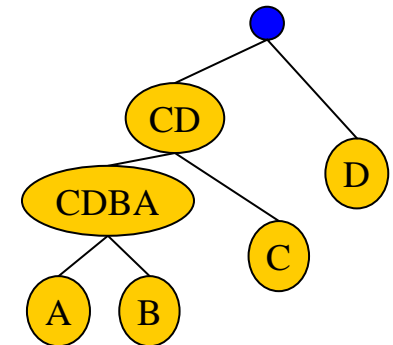
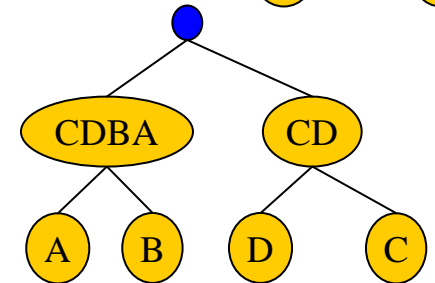
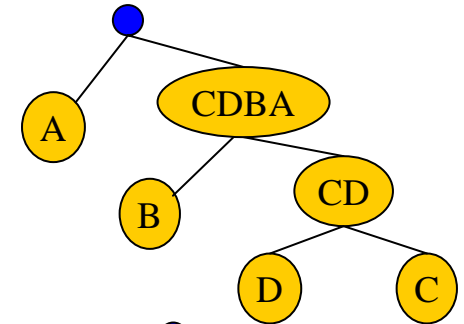
	A	BCD
A		5



# Rooting eines Baumes

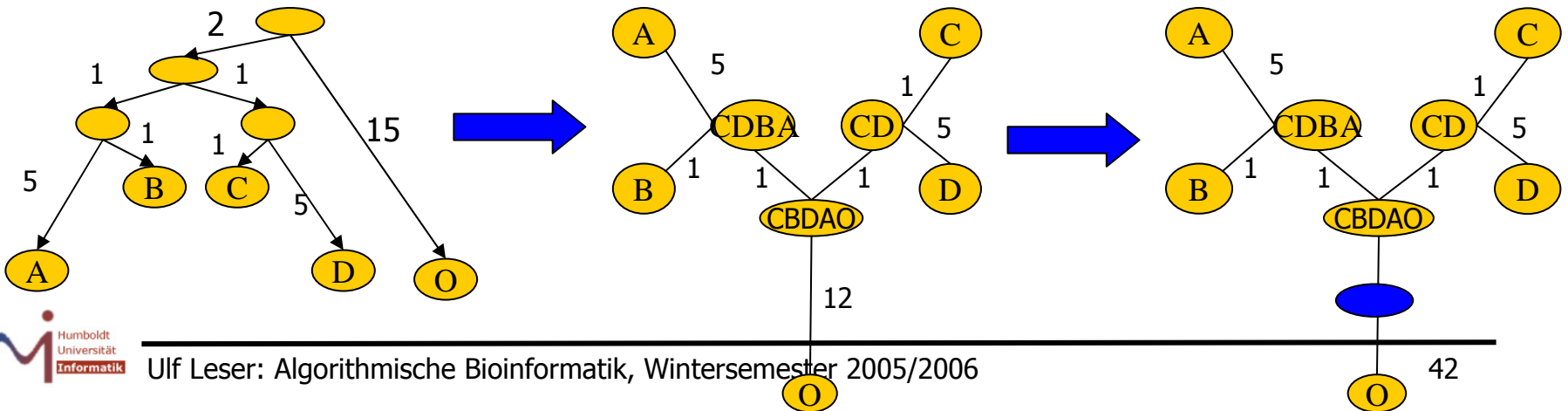


- NJ berechnet nur Abstände zueinander
- Eine zeitliche Reihenfolge der Knoten kommt nicht heraus
- Ein Wurzelknoten kann prinzipiell auf jeder Kante platziert werden
- Wie findet man die echte Wurzel?
  - Externe Datierung einzelner Knoten, z.B. durch Fossilienfunde und C14-Methode
  - Benutzung einer „**Outgroup**“



# Outgroups

- Eine **Outgroup** ist ein Taxa, das weiter von allen anderen Taxa entfernt ist also diese untereinander
  - Beispiel: Menschen, Mäuse, Ratten, Schweine – Wale
- Was passiert mit der Outgroup?
  - NJ ordnet sie im Baum ein
  - Offensichtlich muss die Kante, die zu der Outgroup führt, den zeitlich korrekten Wurzelknoten enthalten
- Beispiel



# Wenn die Daten nun ...

---

- Weder ultrametrisch noch additiv sind?
- Approximieren; z.B. durch
  - Sei  $D(i,j)$  der beobachtete Abstandswert
  - Wähle eine Baumtopologie  $T$  und berechne die tatsächlichen Abstände im Baum  $d(i,j)$  so, dass der Fehler minimiert wird

$$error(t) = \min_d \left( \sum_{i=1}^n \sum_{j \neq i} (D(i, j) - d(i, j))^2 \right)$$

- Für gegebene Topologie  $T$  ist das effizient lösbar, aber ...
- Man muss alle Topologien ausprobieren
- Damit ist das Problem NP-vollständig
- Oder man versucht Heuristiken: UPGMA, Neighbor Joining
- Oder ganz anders aufgebaute Methoden: Maximum Parsimony, Maximum Likelihood

# Literatur

---

- Eine gute Einleitung zu den Problemen (sehr praktisch orientiert)
  - Baldauf, S. L. (2003). "Phylogeny for the faint of heart: a tutorial." *Trends Genet* **19**(6): 345-51.
- Schöne, ausführliche Übersicht (weniger über die Algorithmen)
  - Morrison „Phylogenetic Tree Building“, Int J of Parasitology, 1996
- Ultrametrien und Additivität
  - Gusfield (Kapitel 17 )
- UPGMA und Neighbor Joining
  - Oft beschrieben, z.B. im Script von Ron Shamir, University Tel Aviv
  - <http://www.math.tau.ac.il/~rshamir/algmb/98/scribe/ps/lec09.ps>