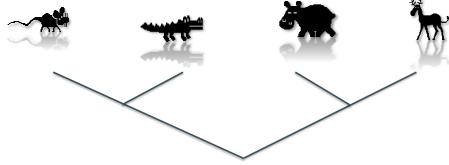
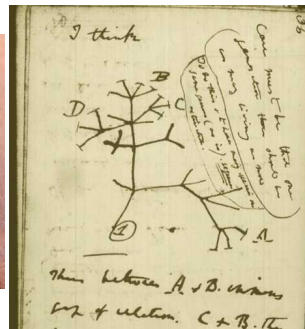


Darwins Erben - Phylogenie und Bäume



Charles Darwin und Bäume



Darwin's Notizbuch B

Charles Darwin und Bäume

- It is a truly wonderful fact ... that **all animals and all plants** throughout all time and space **should be related to each other** in groups, subordinate to groups...
- The affinities of all the beings of the same class **have sometimes been represented by a great tree**. The green and budding twigs may represent existing species; and those produced during former years may represent the long succession of extinct species.

Charles Darwin, (1872), pp. 170-171. The Origin of Species.
Sixth Edition. The Modern Library, New York.

UNIVERSITÄT Tübingen

Center for Bioinformatics

Charles Darwin und Bäume

Die einzige Abbildung in "On the Origin of Species"

Daniel Huson 2012

UNIVERSITÄT Tübingen

Center for Bioinformatics

Phylogenetischer Baum

Monophyletischer Stammbaum der Organismen aus 'Generelle Morphologie der Organismen'

Ernst Haeckel, 1866

Daniel Huson 2012

UNIVERSITÄT Tübingen

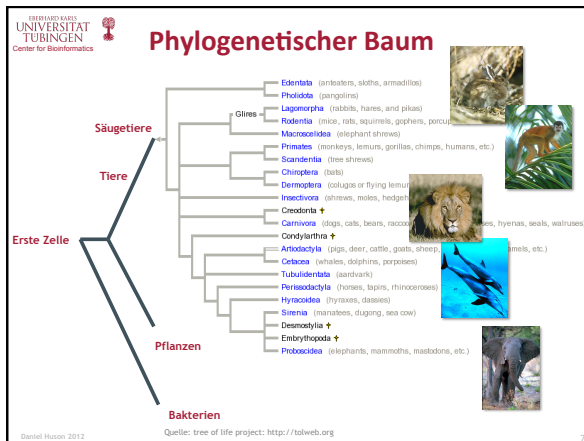
Center for Bioinformatics

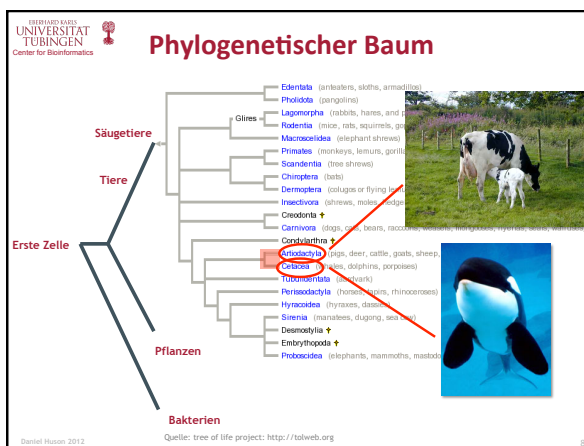
Morphologische Merkmale

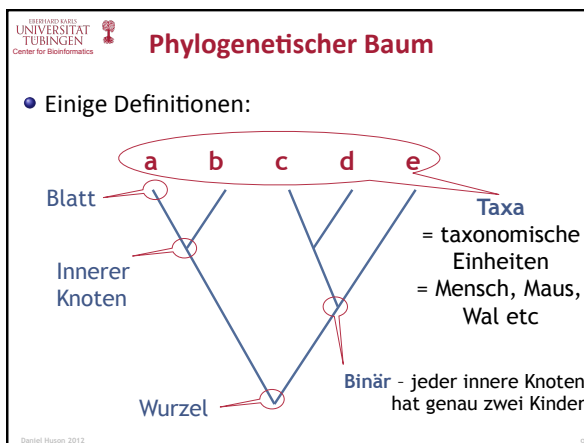
- Evolutionäre Verwandtschaftsverhältnisse können häufig mittels morphologischer Merkmale bestimmt werden
- Beispiel: Unterarmknochen

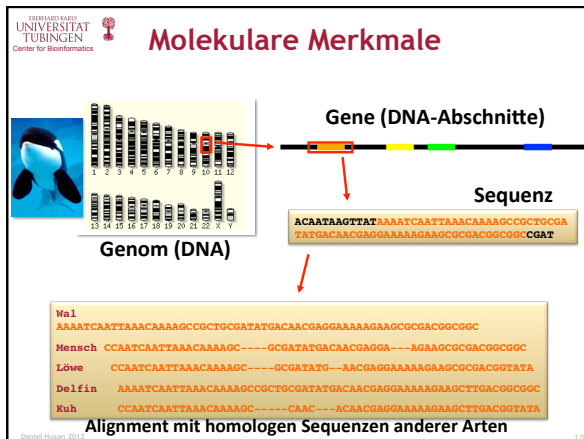
Quelle: Universität Zürich / Institut für systematische Botanik: http://www.sysbot.unizh.ch/studium/grundstudium/bot_113/kursunterlagen/Phylogenie.ppt

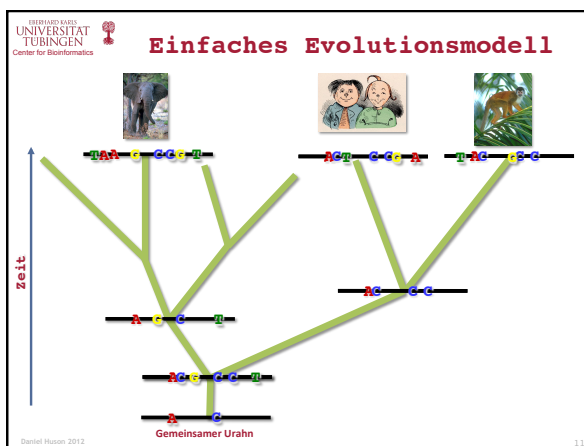
Daniel Huson 2012

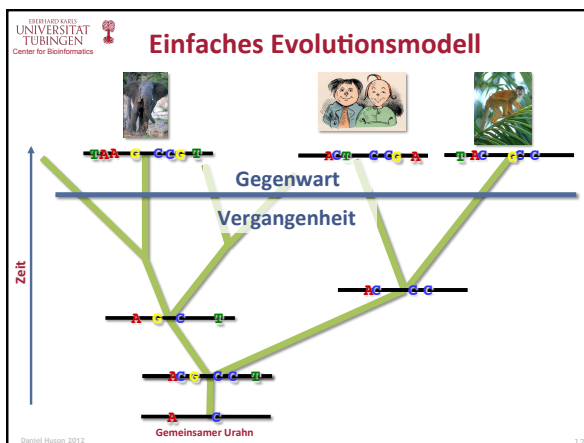




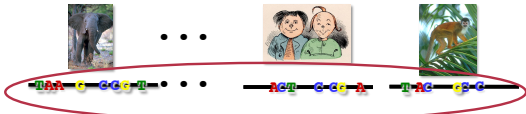









Phylogenetische Rekonstruktion



Gegeben DNA Sequenzen
 heutiger Arten, wie sah
 ihre Evolution aus?



Ziel ist die Rekonstruktion der Evolutionsgeschichte auf Grund des Vergleiches molekularer Sequenzen

Daniel Huson 2012


Phylogenetische Rekonstruktion

Wal	AAATCAATTAAACAAAGCCGCTGCGATATGACAACGAGGAAAAAGAGCGCGCGCGC
Mensch	CCATCAATTAAACAAAGC----GCGATATGACAACGAGGA---AGAAGCGCGACGCGCGC
Löwe	CCATCAATTAAACAAAGC----GCGATATG---AAGAGGAAAAAGAGCGCGCGGTATA
Delfin	AAATCAATTAAACAAAGCCGCTGCGATATGACAACGAGGAAAAAGAGCTTGACGCGCGC
Kuh	CCATCAATTAAACAAAGC----CAAC---ACAACGAGGAAAAAGAGCTTGACGGTATA

Gegeben DNA Sequenzen
 heutiger Arten, wie sah
 ihre Evolution aus?



Ziel ist die Rekonstruktion der Evolutionsgeschichte auf Grund des Vergleiches molekularer Sequenzen

Daniel Huson 2012


SSU rRNA als Standard

- Die ersten sequenzbasierten Phylogenien benutzen Proteinsequenzen
- Ca. 1970 identifizierten Carl Woese und Kollegen die SSU rRNA als ein "universelle molekulare Uhr":
 - Kommt in allen prokaryotischen und eukaryotischen Genomen vor
 - Hat langsam- und schnell-evolvierende Bereiche
 - Struktur der Sequenz ist universell konserviert
- Der heutige "Tree of Life" basiert überwiegend auf dem Vergleich der rRNA Sequenz.



Daniel Huson 2012


Klassifikation des Lebens vor 1970

- Fünf Reiche
- Zwei Domänen



Bakterien

Prokaryoten



Tiere Pflanzen Pilze Protisten

Eukaryoten

Daniel Huson 2012 16


Heutige Klassifikation des Lebens

- Analyse von 16S rRNA zeigt:
 - Archaeen unterscheiden sich von Bakterien
- Folge: Es gibt **drei** Domänen des Lebens



Bakterien

Prokaryoten



Archaeen

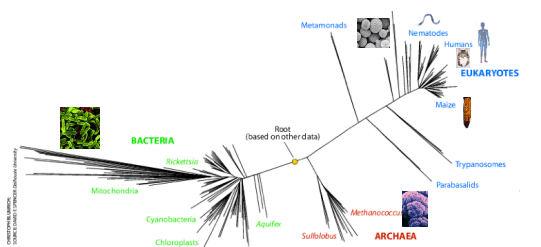


Tiere Pflanzen Pilze Protisten

Eukaryoten

Daniel Huson 2012 17


SSU rRNA "Baum des Lebens"



Daniel Huson 2012 18

Baumrekonstruktionsmethoden

- Gegeben sei ein multiples Alignment **A** von Sequenzen:
 - Es gibt:
 - Distanzbasierte Methoden
 - Maximum Parsimony Methoden
 - Maximum Likelihood Methoden
 - Bayesian Methoden
- um einen phylogenetischen Baum für **A** zu rekonstruieren

Daniel Huson 2012

Maximum Parsimony

- Eingabe:
Alignment **A** von Sequenzen, z.B:
 - Ausgabe:
Ein phylogenetischer Baum **T**, der die Sequenzen *möglichst sparsam* erklärt, d.h. mittels einer möglichst kleinen Anzahl von Mutationen
- a: A C C C**
b: A C T A
c: T C T A
d: A T A A
e: T T C T

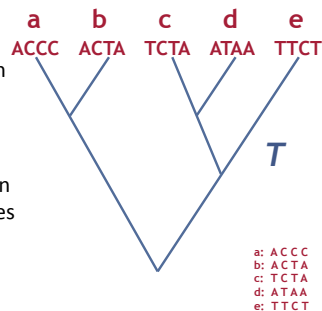
Daniel Huson 2012

20

Das kleine Parsimony Problem

Beispiel

- Wie viele Mutationen braucht dieser Baum, um die Unterschiede in den Sequenzen *a-e* zu erklären?
- Es müssen dafür optimale Sequenzen an den inneren Knoten des Baumes angebracht werden



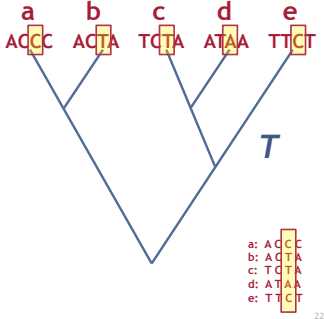
Daniel Huson 2012

21


Das kleine Parsimony Problem

Beispiel

- Betrachte dazu eine Spalte des Alignments



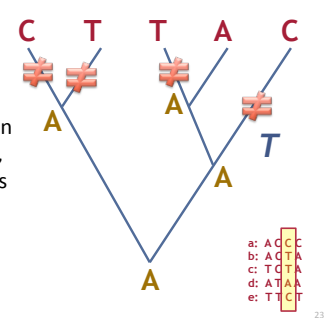
a: ACC
 b: ACT
 c: TCT
 d: ATAA
 e: TTCT

22


Das kleine Parsimony Problem

Beispiel

- Ein mögliches Labeling der inneren Knoten
- Mit diesem Baum und diesem Labeling werden **4 Mutationen** benötigt, um die dritte Spalte des Alignments zu erklären
- Geht es sparsamer?



a: ACC
 b: ACT
 c: TCT
 d: ATAA
 e: TTCT

23


Berechnung des Optimalen Scores

- Gegeben ein Alignment **A** und ein Baum **T**, ein optimales Labeling der inneren Knoten kann mit dem Fitch-Algorithmus (Fitch, 1971) berechnet werden.
- Zwei Phasen:
 - Phase I: Bestimmung der möglichen Labels
 - Phase II: Wahl eines Labels für jeden Knoten

24

Fitch Algorithmus – Phase I

- Phase I: Bestimmung der möglichen Labels „bottom up“
- Für jeden Knoten v mit Kindern p und q :
- Falls $F(p) \cap F(q) \neq \emptyset$:
 - $F(v) = F(p) \cap F(q)$
- Sonst:
 - $F(v) = F(p) \cup F(q)$
 - Score geht um 1 hoch

Score: 111

a: A C C
b: A C T
c: T C T
d: A T A
e: T T C

Fitch Algorithmus – Phase II

- Phase II: Wahl von Labels „top down“
- Wähle Label für Wurzel
- Für jeden Knoten v mit Eltern p :
- Falls $F(v)$ das für p gewählte Label $L(p)$ enthält:
 - $L(v) = L(p)$
- Sonst:
 - Wähle $L(v)$ aus $F(v)$ beliebig

Score: 111

a: A C C
b: A C T
c: T C T
d: A T A
e: T T C

Fitch Algorithmus – Ergebnis

- Die dritte Spalte benötigt drei Mutationen auf diesem Baum
- Gesamtscore:
 - Fitch Algorithmus auf jede Spalte anwenden
 - Spaltenscores aufsummieren:
 - $2 + 1 + 3 + 2 = 8$
- Löst das kleine Parsimony Problem

Score: 111

a: A C C
b: A C T
c: T C T
d: A T A
e: T T C

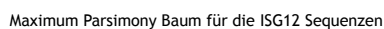
- $$T_{MP} = \operatorname{argmin} \{S(T)\}$$

28

[illegible]

Quelle: Parker and Porter BMC Genomics 2004 5:8
<http://www.biomedcentral.com/1471-2164/5/8>

29

[illegible]

Quelle: Parker and Porter BMC Genomics 2004 5:8
<http://www.biomedcentral.com/1471-2164/5/8>

30

[illegible]

Zusammenfassung

- Die Evolutionsgeschichte von Spezies wird durch einen phylogenetischen Baum beschrieben
- Phylogenetische Bäume werden mittels des Vergleiches von DNA- oder Proteinsequenzen heutiger Organismen rekonstruiert
- Die Maximum Parsimony Methode basiert auf der Suche nach einem Baum, der Unterschiede in Sequenzen möglichst sparsam erklärt.

