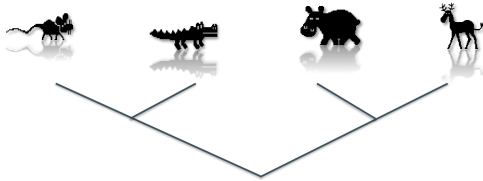
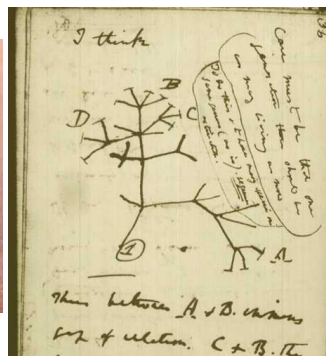


Darwins Erben - Phylogenie und Bäume



Charles Darwin und Bäume

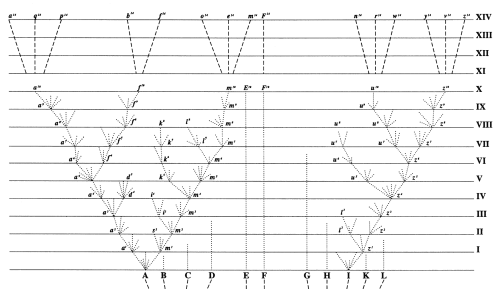


Darwin's Notizbuch B

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.



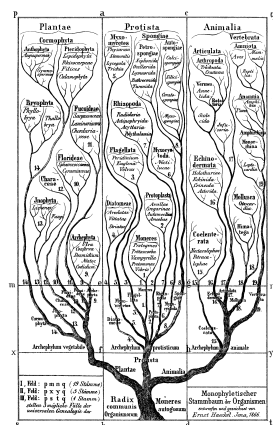
Die einzige Abbildung in "On the Origin of Species"

A

[illegible]

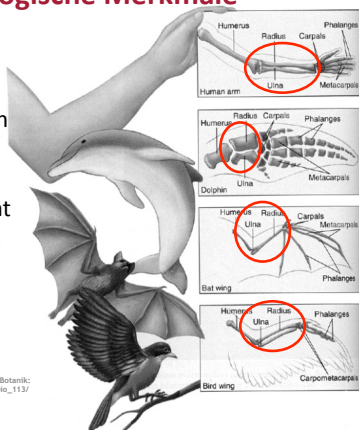
**Monophyletischer
Stammbaum der
Organismen** aus
'Generelle Morphologie
der Organismen'

Ernst Haeckel, 1866

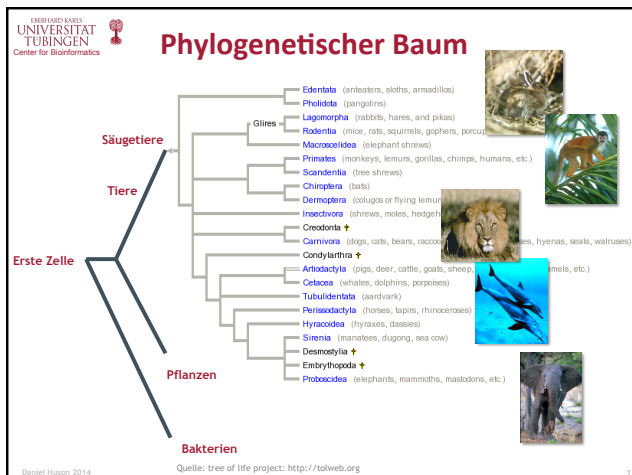


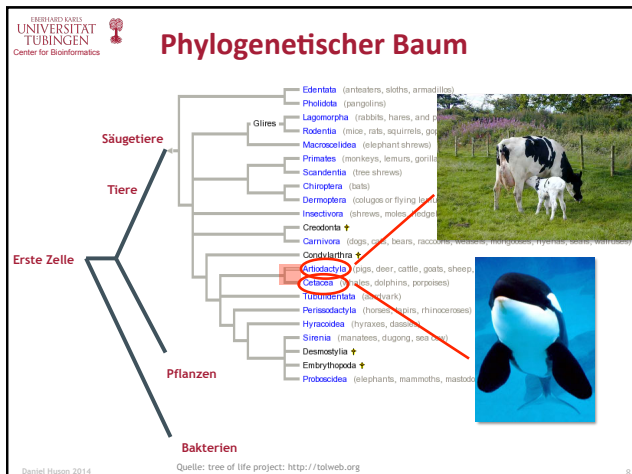
- Evolutionäre Verwandtschaftsverhältnisse können häufig mittels morphologischer Merkmale bestimmt werden

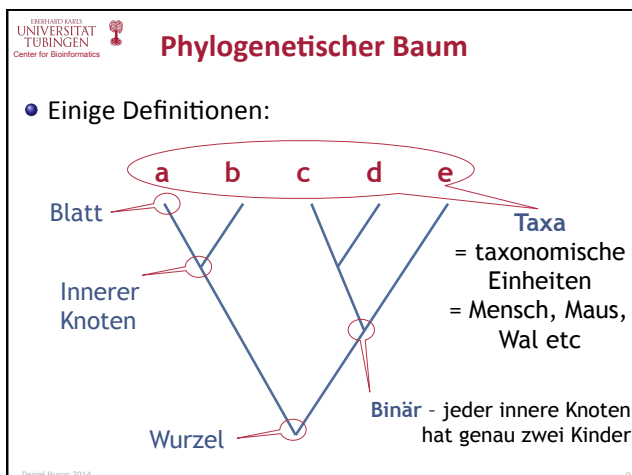
- Beispiel:
Unterarmknochen

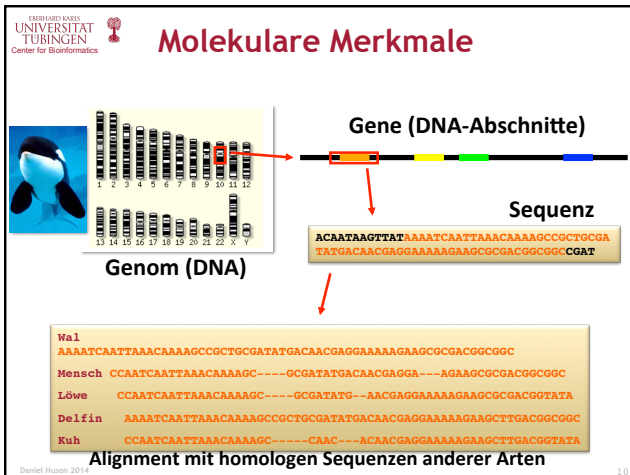


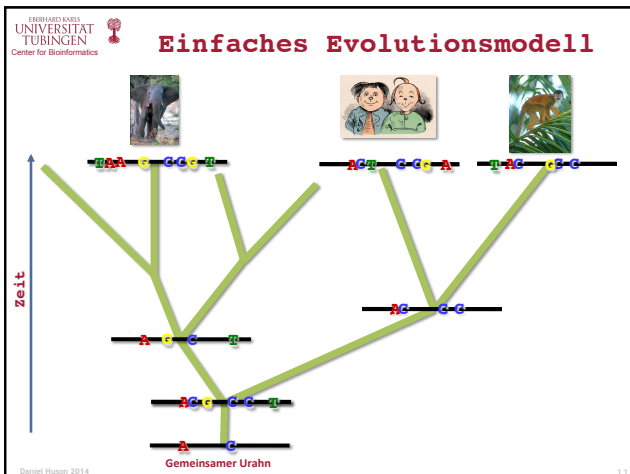
6

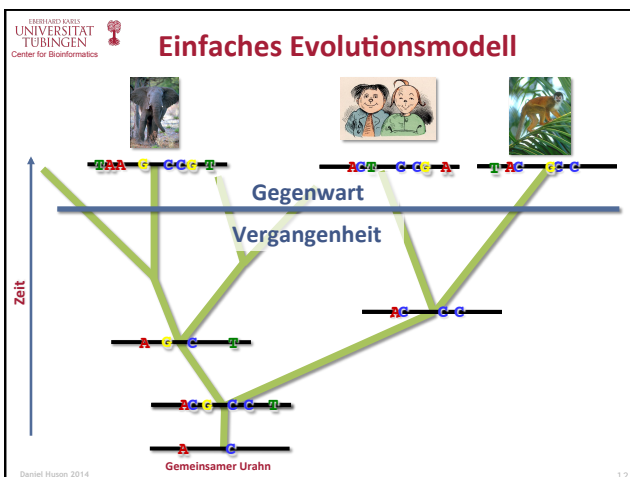




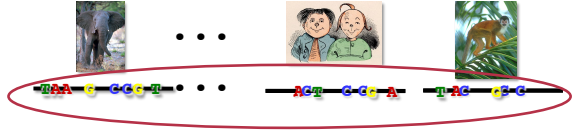




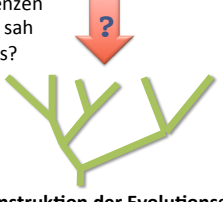





Phylogenetische Rekonstruktion



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



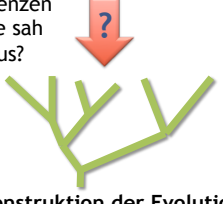
**Ziel ist die Rekonstruktion der Evolutionsgeschichte auf
 Grund des Vergleiches molekularer Sequenzen**

Daniel Huson 2014


Phylogenetische Rekonstruktion

Wal	AAATCAATTAACAAAAGCCGCTGCGATATGACAAACGAGGAAAAAGAAGCGCGACGCGCGC
Mensch	CCAATCAATTAACAAAAGC----GCGATATGACAAACGAGGA---AGAAGCGCGACGCGCGC
Löwe	CCAATCAATTAACAAAAGC----GCGATATG--AACGAGGAAAAAGAAGCGCGACGCTATA
Delfin	AAATCAATTAACAAAAGCCGCTGCGATATGACAAACGAGGAAAAAGAAGCTTGACGCGCGC
Kuh	CCAATCAATTAACAAAAGC----CAAC---ACAACGAGGAAAAAGAAGCTTGACGCTATA

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

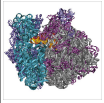


**Ziel ist die Rekonstruktion der Evolutionsgeschichte
 auf Grund des Vergleiches molekularer Sequenzen**

Daniel Huson 2014


SSU rRNA als Standard

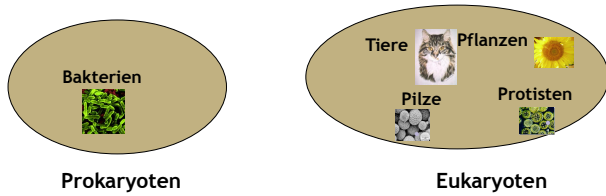
- Die ersten sequenzbasierten Phylogenien benutzen Proteinsequenzen
- Ca. 1970 identifizierten Carl Woese und Kollegen die SSU rRNA als ein "universelle molekulare Uhr":
 - Kommt in allen prokaryotischen und eukaryotischen Genomen vor
 - Hat langsam- und schnell-evolvierende Bereiche
 - Struktur der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



Daniel Huson 2014

Klassifikation des Lebens vor 1970

- Fünf Reiche
- Zwei Domänen

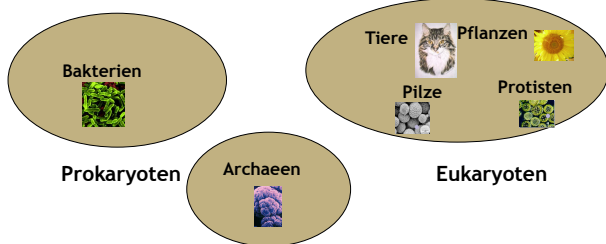


Daniel Huson 2014

[illegible]

Heutige Klassifikation des Lebens

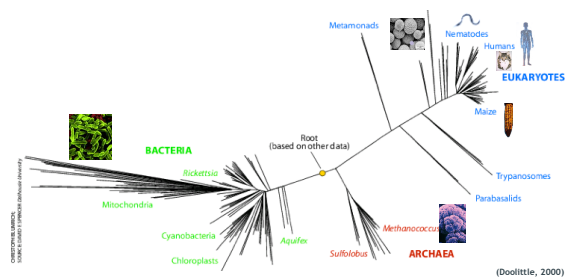
- Analyse von 16S rRNA zeigt:
 - Archaeen unterscheiden sich von Bakterien
- Folge: Es gibt **drei** Domänen des Lebens



Daniel Huson 2014

[illegible]

SSU rRNA "Baum des Lebens"



Daniel Huson 2014

[illegible]

Baumrekonstruktionsmethoden

- Gegeben sei ein multiples Alignment **A** von Sequenzen:

Wal AAAATCAATTAAACAAAGCGCGTGCATATGACACAGAGAAAAGAGCGCGACGCGCG
Memsch CCAATCAATTAAACAAAGG----CGCATATGACACAGAG---AGAGAGCGCGCGCGCGCG
Löwe CCAATCAATTAAACAAAGG----CGCATATG---AACGAGAGAAAAGAGCGCGCGGTATA
Delphin AAAATCAATTAAACAAAGCGCGTGCATATGACACAGAGAAAAGAGCTTAAAGCGCGCG
Kuh CCAATCAATTAAACAAAGG----CAAC---ACACAGAGAAAAGAGCTTACGCGTATA

- Es gibt:

- Distanzbasierte Methoden
- Maximum Parsimony Methoden
- Maximum Likelihood Methoden
- Bayesian Methoden

um einen phylogenetischen Baum für **A** zu rekonstruieren 

[illegible]

Maximum Parsimony

- Eingabe:

Alignment **A** von
Sequenzen, z.B:

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

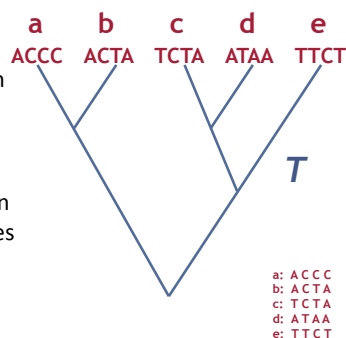
- Ausgabe:

Ein phylogenetischer Baum **T**, der die Sequenzen *möglichst sparsam* erklärt, d.h. mittels einer möglichst kleinen Anzahl von Mutationen

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



Das kleine Parsimony Problem

Beispiel

- Betrachte dazu eine Spalte des Alignments

a: ACCC
b: ACTA
c: TCTA
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: ACCC
b: ACTA
c: TCTA
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

F

-

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

-

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

F

-

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

Das große Parsimony Problem

- Eingabe:
Alignment **A** von Sequenzen, z.B:

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

- Ausgabe:
Aus der Menge aller möglichen Bäume, finde einen Baum **T**, der **A** mit möglichst wenigen Mutationen erklärt, also für den der Fitch Algorithmus den kleinsten Score berechnet.

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

Daniel Huson 2014

28

Beispiel

Multiples Alignment für Sequenzen, die das Protein Motif ISG12 enthalten

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

Daniel Huson 2014

29

Beispiel

Maximum Parsimony Baum für die ISG12 Sequenzen

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

Daniel Huson 2014

30



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.

