

# Raten und Muster der Evolution von DNA- und Protein-Sequenzen

## Das Konzept der Molekularen Uhr

Review: Nature Genetics 2003, 4:216



Thomas Hankeln, Institut für Molekulargenetik

SS 2008

JOHANNES  
GUTENBERG  
UNIVERSITÄT  
MAINZ



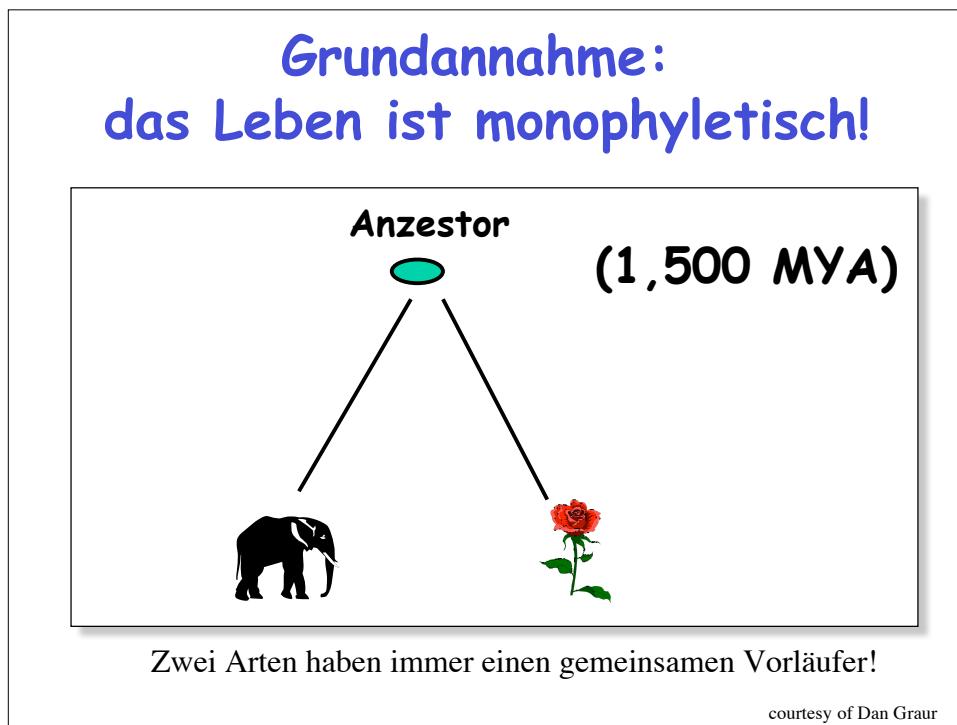
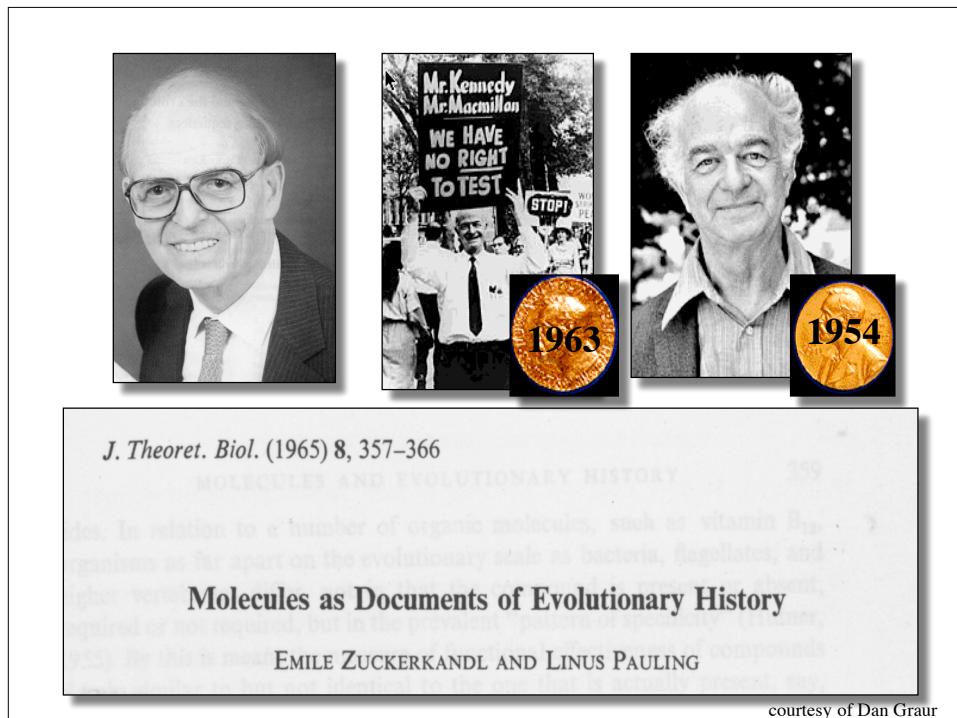
Sydney Brenner

**"Searching for an objective reconstruction of the vanished past is surely the most challenging task in biology."**

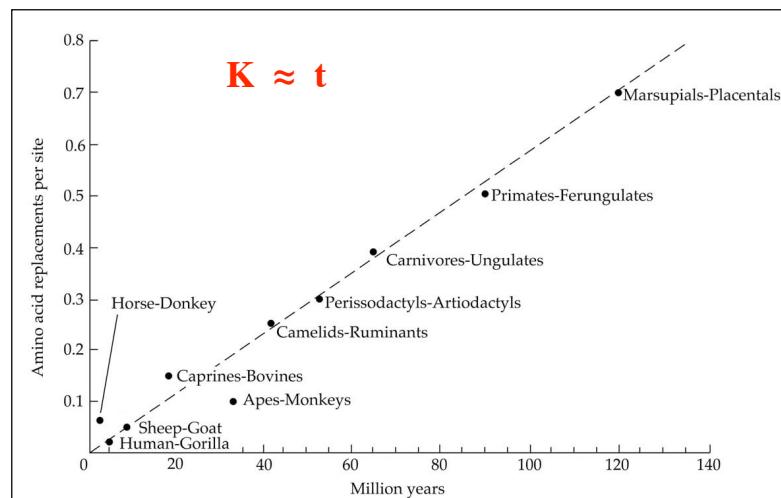
**"In one sense, everything in biology has already been 'published' in the form of DNA sequences of genomes."**

**1991**

courtesy of Dan Graur



## „Molekulare Uhr“ bei Säger-Proteinen



## „Molekulare Uhr“ bei Säger-Proteinen

Wenn Proteine mit konstanter Geschwindigkeit evolvieren, kann man anhand ihrer Divergenz den Zeitpunkt der evolutionären Trennung berechnen!

Analogie:  
geologische Datierung durch radioaktiven Zerfall

## Die Hypothese einer „Molekularen Uhr“

- Zuckerkandl und Pauling 1962, 1965  
Margoliash 1963
- zunächst Sequenzdaten von Globinen und Cytochrom C
- Hypothese:  $K \approx t$
- Anwendung: Datierung von Speziations- oder Genduplikations-Ereignissen anhand beobachteter Divergenzwerte und einer zeitlich geeichten Evolutionsrate

$$R = K / 2T \quad > \quad T = K / 2R$$

## Berechne die Artentrennung mit einer „Molekularen Uhr“!

- $T = K / 2R$
- Nicht-synonyme Rate ( $\alpha$ -Globin) =  $0,56 \times 10^{-9}$  Subst/site/yr
- Gemessene Distanz „Ratte-Mensch“  $K_A = 0,093$  Subst/site
- Ergibt  $T = 80$  Mio. Jahre für die Artentrennung

# Die Zuverlässigkeit einer „Molekularen Uhr“ hängt ab von...

... der Genauigkeit, mit der genetische Divergenz gemessen wird („Evolutionsmodelle“)

... der Genauigkeit der zeitlichen Kalibrierung

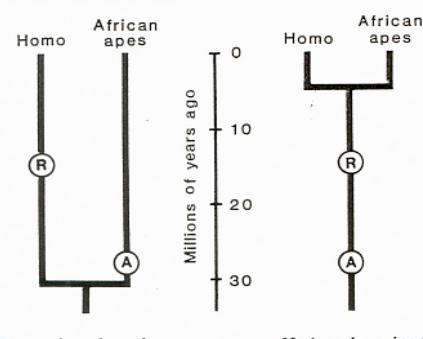


Fig. 3. Changing views about the evolutionary history of the human lineage. Emphasis is given to the question of how to link two well-dated fossils, Ramapithecus (R) and Aegyptopithecus (A), to the lineages leading to humans and African apes<sup>23</sup>. There has been a dramatic shift in views about the phylogenetic position of Aegyptopithecus, which lived about 30 million years ago. Formerly assigned on the basis of tooth characters to the ape lineage<sup>24</sup>, Aegyptopithecus is now considered, on the basis of bones from diverse parts of the body, not to be an ape, or a hominoid, or even a catarrhine primate; indeed, it may lie on the lineage leading to the common ancestors of New World monkeys, Old World monkeys and hominoids<sup>25</sup>. A lesser shift, but a highly significant one, has occurred in views about the affinities of Ramapithecus, which lived 14 million years ago; once regarded as being related to the human lineage, Ramapithecus is now linked to the common ancestral lineage for humans and African apes<sup>26</sup>.

Die Molekulare Uhr revolutionierte die Sichtweise der menschlichen Evolution

## Timing the Ancestor of the HIV-1 Pandemic Strains

B. Korber,<sup>1,2\*</sup>† M. Muldoon,<sup>2,3</sup> J. Theiler,<sup>1</sup> F. Gao,<sup>4</sup> R. Gupta,<sup>1</sup>  
A. Lapedes,<sup>1,2</sup> B. H. Hahn,<sup>4</sup> S. Wolinsky,<sup>5</sup> T. Bhattacharya<sup>1†</sup>

HIV-1 sequences were analyzed to estimate the timing of the ancestral sequence of the main group of HIV-1, the strains responsible for the AIDS pandemic. Using parallel supercomputers and assuming a constant rate of evolution, we applied maximum-likelihood phylogenetic methods to unprecedented amounts of data for this calculation. We validated our approach by correctly estimating the timing of two historically documented points. Using a comprehensive full-length envelope sequence alignment, we estimated the date of the last common ancestor of the main group of HIV-1 to be 1931 (1915–41). Analysis of a gag gene alignment, subregions of envelope including additional sequences, and a method that relaxed the assumption of a strict molecular clock also supported these results.

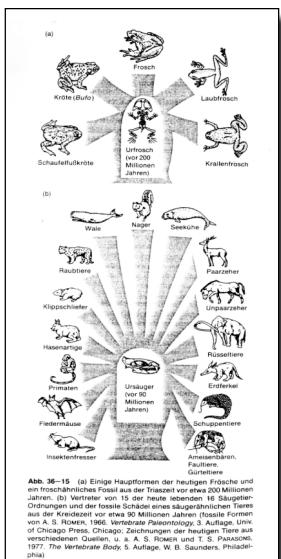
Science 2000, 288:1789

The big divorce...



Es gibt keine Beziehung zwischen der Geschwindigkeit der molekularen Evolution und der morphologischen Evolution!

## „Uhr-ähnliche“ Evolution...



...im Widerspruch zur sprunghaften morphologischen Evolution?

Nicht unbedingt!!!

Die morphologische Evolution wird primär vermutlich von Mutationen in genregulatorischen Sequenzen bestimmt, während z. B. Mutationen in Strukturgenen sich durchaus „clock-like“ anhäufen können.

## Die „neutral theory“ als Grundlage der Molekularen Uhr

*„For each protein, the rate of evolution in terms of amino acid substitutions is approximately constant per year per site for various lines, as long as the function and tertiary structure of the molecule remain essentially unaltered“*

Motoo Kimura 1983

## neutral theory & molecular clock

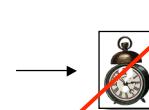
$$K = 2N \cdot \mu \cdot u$$

K Substitutionsrate (pro site•yr)  
 2N diploide Population  
 $\mu$  Mutationsrate  
 u Probab. Fixierung

bei **neutraler** Mutation gilt  $u = 1/2N$ , d. h.  
 $K = \mu$  !!

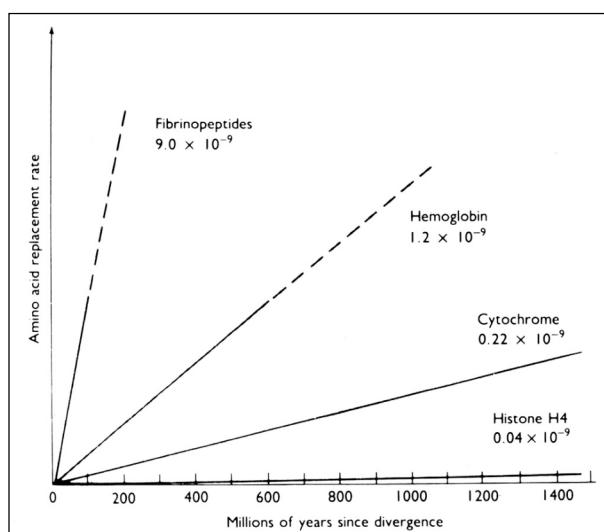


bei **selektiv vorteilhafter** Mutation gilt  
 $u = 2sN_e/N$ , d. h.  
 $K = 4N_e s\mu$



- drei Variablen  $N_e$ ,  $s$  und  $\mu$  sind einzubeziehen, um konstante Evolutionsrate und damit Molekulare Uhr zu produzieren.
- Uhr-ähnliche Sequenzervolution schwierig unter Selektion zu erklären

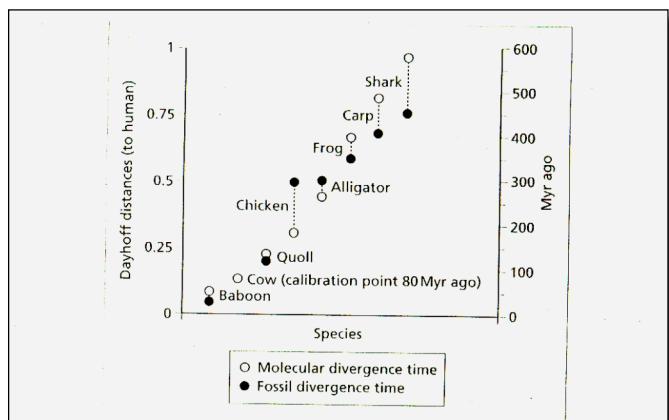
## „Molekulare Uhr“ bei Säuger-Proteinen



Warum tickt die Uhr unterschiedlich schnell?

Je mehr neutrale Positionen im Protein, desto schnellere Uhr!

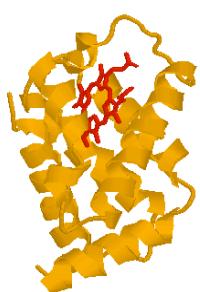
## Die „ $\alpha$ -Hämoglobin-Uhr“ ...



...funktioniert  
passabel in  
Vertebraten!

**Fig. 7.1** The  $\alpha$ -globin molecular clock. The distances between the amino acid sequences of humans and a number of other vertebrates are used to estimate species divergence times, using the split between humans (primates) and cows (artiodactyls) at approximately 80 Myr ago as a calibration point. These molecular estimates (open circles) are then compared with those taken from the fossil record (closed circles). The difference between the two estimates is indicated by the dotted line. The Quoll is a marsupial cat.

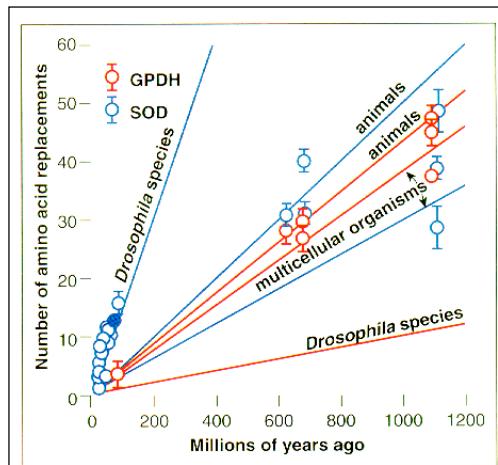
## Warum verhalten sich gerade die Hämoglobine oft „clock-like“?



Die Aminosäuren an der Außenseite des „globin fold“ müssen lediglich hydrophil, die As in der Hämtasche dagegen hydrophob sein

> erlaubt relativ weitgehende, quasi **neutrale** Austausche

## Keine „universelle“ Uhr für alle Gene/Proteine & Organismen

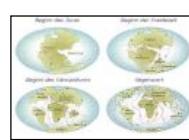


Science 283  
(1999), p. 1435

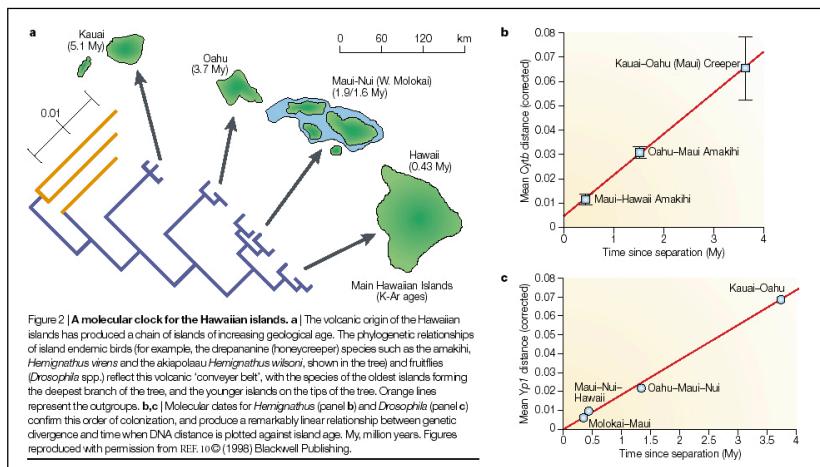
Rate spread. Not only do the enzymes GPDH and SOD have different rates of evolution, the rates vary in different groups of organisms.

## Zeitliche Kalibrierung

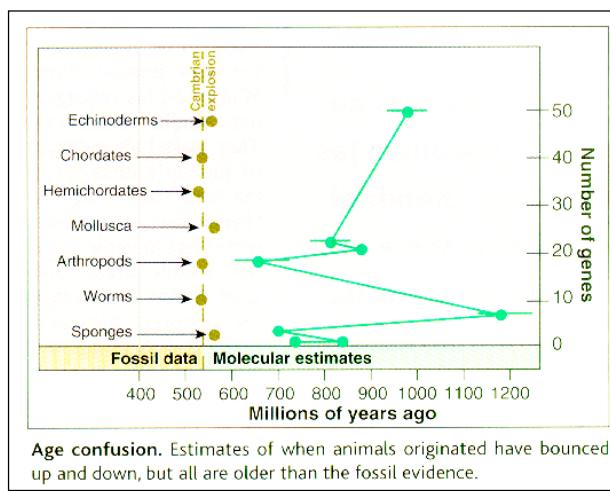
- Physiko-chemisch datierte fossile Funde
- Biogeographie (z. B. See-Entstehung, Kontinentaldrift)
- Pollenanalyse
- Zeitpunkt der Probensammlung (Mikrobio)



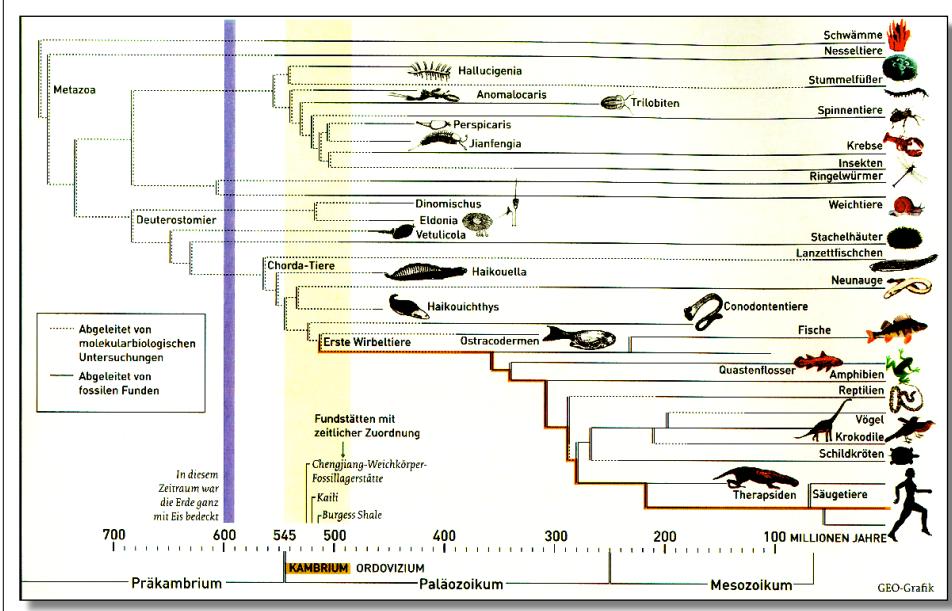
## Zeitliche Kalibrierung



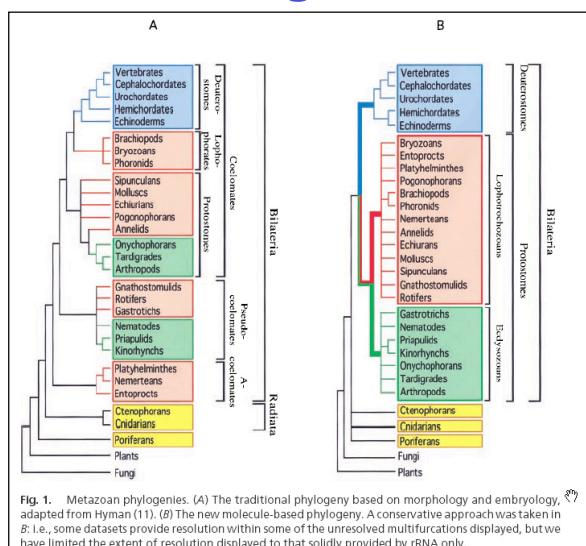
## Problem: Datierung früher Ereignisse



## Problem: Korrekte Datierung



## Die „deep metazoan phylogeny“ ist völlig unklar!



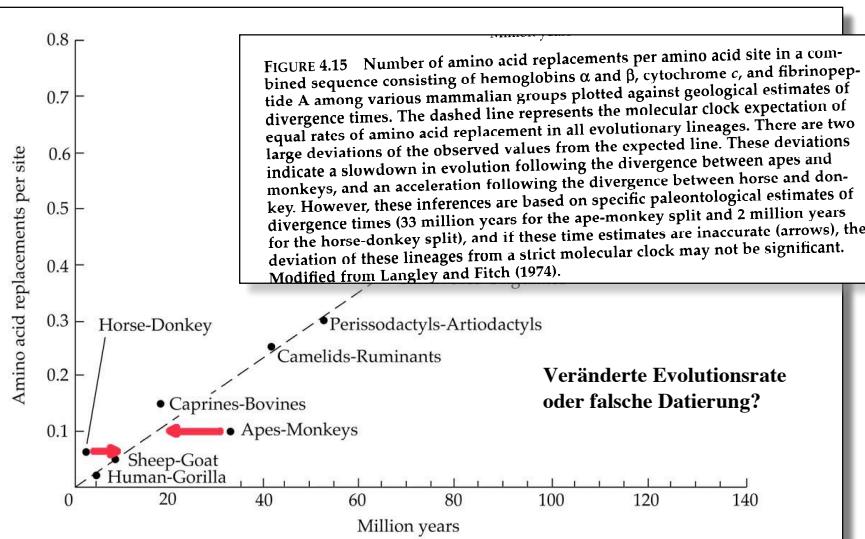
Adoutte et al. 2000

## Problem: Korrekte Datierung

Viele Argumente (gegen wie für die Molekulare Uhr) beruhen auf einer zeitlichen Eichung durch fossile Funde!

Wünschenswert wäre also eine Methode, mit der man ohne konkrete zeitliche Daten Moleküle auf eine Ratenkonstanz/Molekulare Uhr hin überprüfen kann...

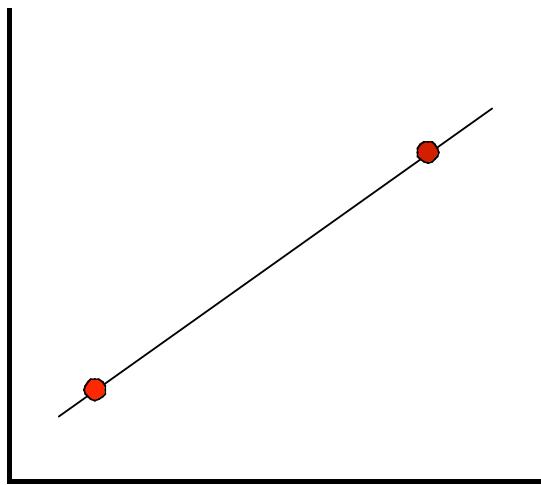
## Was bedeuten Abweichungen?



courtesy of Dan Graur

**Q: How to draw a straight line?**

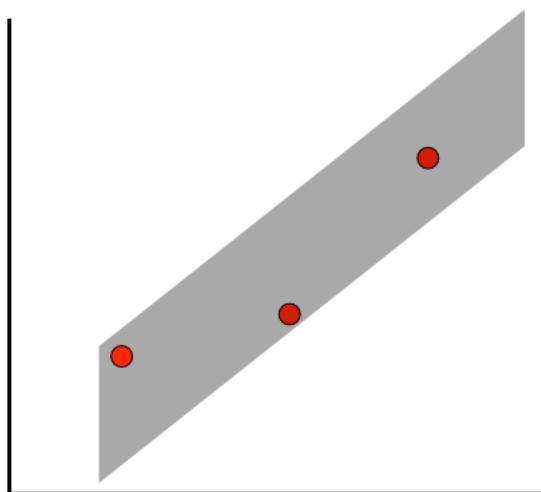
**A1: Have no more than two observation points.**



courtesy of Dan Graur

**Q: How to draw a straight line?**

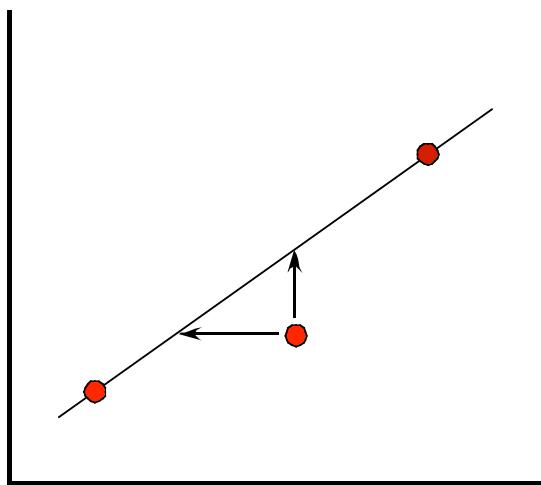
**A2: With more than two observation points, use a very thick line.**



courtesy of Dan Graur

**Q: How to draw a straight line?**

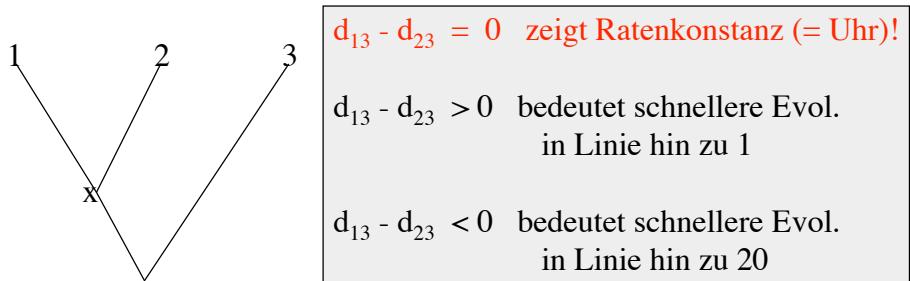
**A3: With more than two observation points, deny the accuracy of the measurements on one or both axes.**



courtesy of Dan Graur

## Der „relative rate test“ (Sarich and Wilson 1973)

- ermöglicht Vergleich von Evolutionsraten ohne absolute Zeitangaben
- Vorgehensweise: Vergleich der Raten in den Linien ( $x>1$ ) und ( $x>2$ ) durch Einbeziehung der Referenzspezies (outgroup) 3



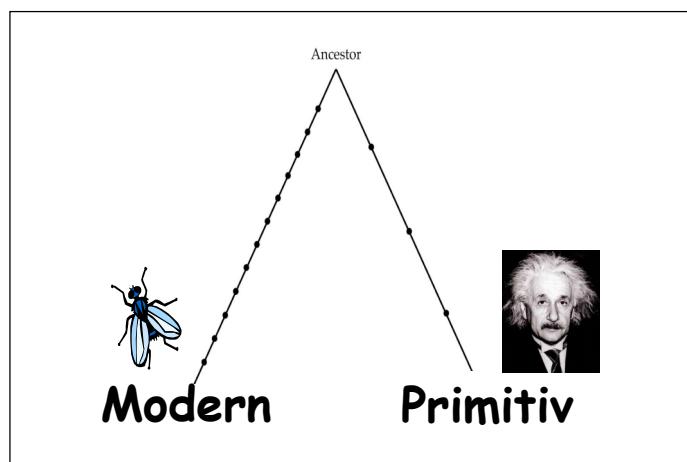
## Exkurs: Primitiv vs. Modern

- Giardia - ein primitiver Einzeller?
- Schwämme - primitive Metazoa?
- Farne - primitive Gefäßpflanzen?
- der Mensch – „advanced“?



Der „relative-rate“-Test gibt uns eine Möglichkeit, die primitive (= langsamer evolvierende) Linie objektiv festzulegen.

## Exkurs: Primitiv vs. Modern Nur eine Frage der Evolutionsrate!



courtesy of Dan Graur

## Exkurs: Sind „living fossils“ auch molekulare Fossilien?



- Substitutionsraten in Haien, Krokodilen oder gar in Limulus sind „weitgehend normal“
- Krokodile evolvieren sogar schneller als Vögel (mtDNA)

Der relative-rate-test zeigt:  
**Es gibt „local clocks“ für bestimmte Gene/Proteine in bestimmten phylogenetischen Linien**

z. B. in Ratte vs. Maus (ca. 30 Mio. Jahre getrennt)

**TABLE 4.10** Differences in the number of substitutions per 100 sites between mice (species 1) and rats (species 2), with hamsters (species 3) as a reference

Type of substitution	Number of sites compared	Number of substitutions <sup>a</sup>			
		$K_{12}^b$	$K_{13}$	$K_{23}$	$K_{13} - K_{23}$
Synonymous	4,855	$19.9 \pm 0.7$	$31.1 \pm 0.9$	$32.4 \pm 1.0$	$-1.3 \pm 7.9$
Nonsynonymous	17,440	$1.9 \pm 0.1$	$2.9 \pm 0.1$	$2.7 \pm 0.1$	$0.3 \pm 1.3$

Modified from O'hUigin and Li (1992).

<sup>a</sup>Mean  $\pm$  standard error.  $K_{ij}$  = number of substitutions per 100 sites between species  $i$  and  $j$ .

## Relative-rate-Tests zeigen verlangsamte Evolution in Primaten

TABLE 4.12 Differences in the number of nucleotide substitutions per 100 sites in noncoding regions and the relative rates of substitution between the African monkey (species 1) and the human (species 2) lineages, with a New World monkey (species 3) as reference

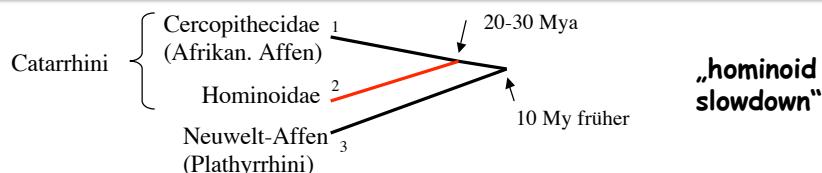
Type of sequence	Sequence length	$K_{12}^a$	$K_{13}$	$K_{23}$	$K_{13} - K_{12}$	Ratio <sup>b</sup>
Pseudogene	8,781	6.7	11.8	10.7	1.1 ± 0.3**	1.4
Introns	8,478	7.1	14.7	13.9	0.8 ± 0.3**	1.3
Flanking and untranslated regions	936	7.9	14.9	11.7	3.1 ± 1.1**	2.3

Data from Bailey et al. (1991), Porter et al. (1995), and Ellsworth et al. (1993).

\*\*Significant at the 1% level.

<sup>a</sup> $K_{ij}$  = number of substitutions per 100 sites between species *i* and *j*.

<sup>b</sup>The ratio of the rate in the African monkey region to that in the human lineage.



## Rodentia evolvieren schneller als Primaten

Table 7. Rates of synonymous substitution per site per year in primates and rodents.<sup>a</sup>

Species pair	Number of sites	Percent divergence	Substitution rate ( $\times 10^9$ )
Human vs. chimpanzee	921	1.9	1.3 (0.9–1.9) <sup>b</sup>
Human vs. Old World monkeys	998	11.0	2.2 (1.8–2.8)
Mouse vs. rat	3,886	23.7	7.9 (3.9–11.8)

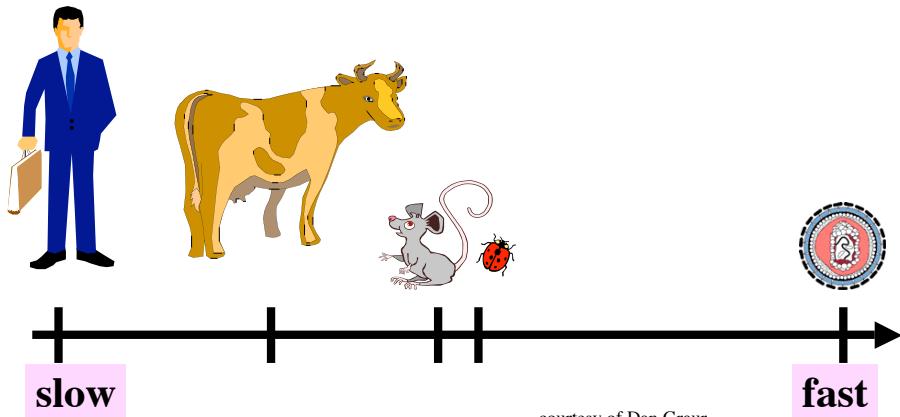
From Li et al. (1987a).

<sup>a</sup> The divergence times used are 7 (5–10) million years ago for the human-chimpanzee split, 25 (20–30) million years ago for the human–Old World monkey split, and 15 (10–30) million years ago for the mouse–rat split.

> sowohl zwischen, als auch innerhalb der verschiedenen Säuger-Ordnungen gibt es Unterschiede in der Nukleotid-Substitutionsrate

> „local clocks“ können jedoch durchaus verwendet werden

Es gibt keine „universelle“  
Molekulare Uhr, wohl aber gut  
funktionierende „lokale Uhren“!

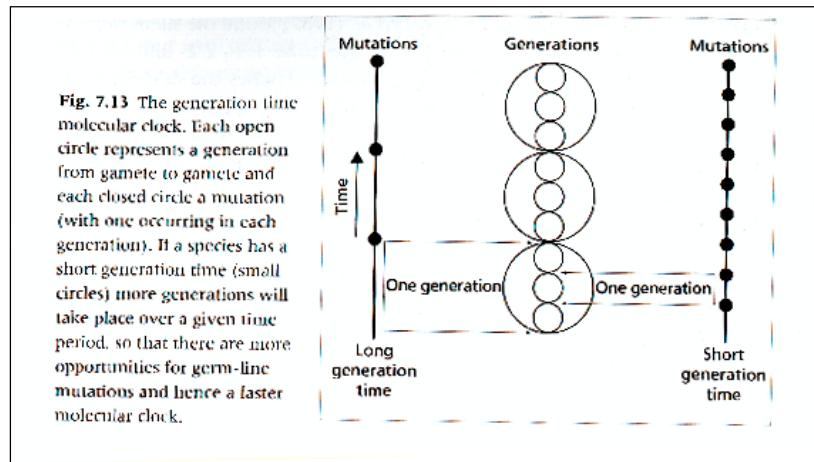


courtesy of Dan Graur

Warum gibt es „lineage effects“  
bei den Substitutionsraten?

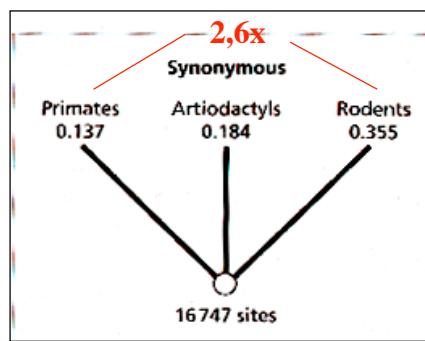
- „generation time effect“ (Laird et al. 1969; Kohne 1970)  
    > Replikationsfehler sind verantwortlich
- „metabolic rate hypothesis“ (Martin, Palumbi 1993)  
(body size, temperature)  
    > mutagene Radikale sind verantwortlich
- DNA-Reparatur-Hypothese (Britten 1986)

## Der Generationszeit-Effekt



- Spezies mit **kurzer** Generationszeit > **hohe** Substitutionsrate pro Zeiteinheit
- Spezies mit **langer** Generationszeit > **niedrige** Substitutionsrate pro Zeiteinheit

## Nicht nur die Generationszeit!



Auch die Anzahl der Replikationsrunden in der Keimbahn bestimmt die „synonyme Uhr“!

- Nager haben 40x kürzere Generationszeit als Primaten
- Keimbahnreplikationen im Mann jedoch 205 vs. 57 in Ratte  
-> **ergibt insgesamt 2-4 fach schnellere Ks in Rodentia**

## Aber...

- Wale haben kürzere G-Zeit als Primaten, evolvieren aber langsamer
- in Haien (kurze G-Zeit) ist Ks der mtDNA 5-7x niedriger als in Säugern
- Seeigel (lange G-Zeit, aber auch viele Keimbahn-Replikationen) evolvieren sehr schnell

## „metabolic rate hypothesis“

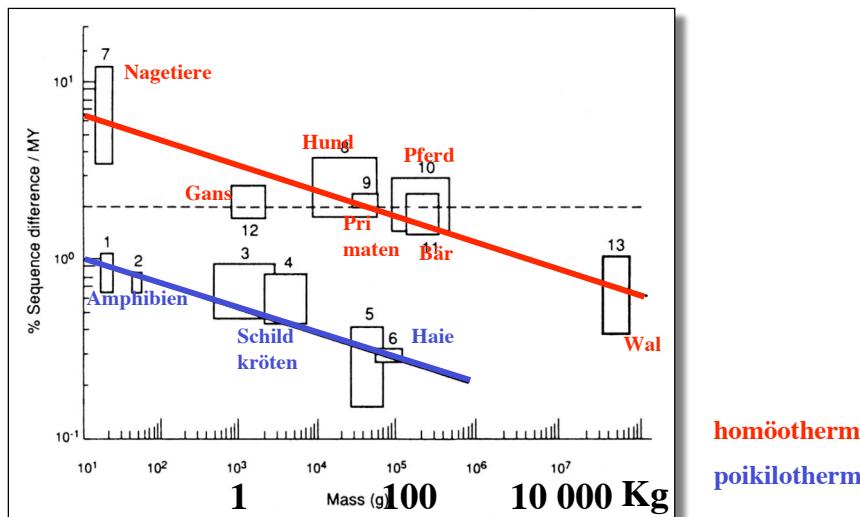
### Voraussage:

Organismen mit intensivem Stoffwechsel erzeugen mehr DNA-schädigende Radikale

> höhere Mutationsrate!

Kleine	> Große
Homöotherme	> Poikilotherme

## „metabolic rate hypothesis“



## Nochmal: die „neutral theory“...

...erklärt gut, dass an **synonymen Kodonpositionen** in **Genen oder neutralen Aminosäurepositionen**

a) manchmal eine Ratenkonstanz vorliegt (clock)

b) manchmal variable Evolutionsraten (no clock)

zu beobachten sind. Der Prozeß ist mutationsgetrieben und hängt daher von Faktoren wie G-Zeit oder metabol. Rate ab.



Wie aber ist eine „molekulare Uhr“ der unter Selektion stehenden **nicht-synonymen Kodonpositionen** zu erklären, die sich oftmals Generationszeit-unabhängig an der real vergangenen Evolutionszeit ausrichten lässt?

## Die „nearly neutral theory“ (Ohta 1992)

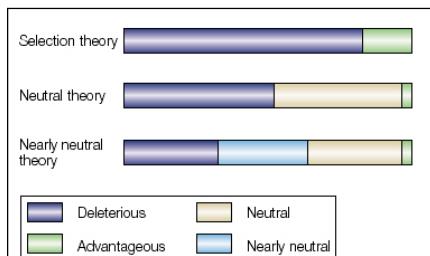


Figure 1 | Selectionist, neutral and nearly neutral theories.  
a | Selectionist theory: early neo-Darwinian theories assumed that all mutations would affect fitness and, therefore, would be advantageous or deleterious, but not neutral. b | Neutral theory: the neutral theory considered that, for most proteins, neutral mutations exceeded those that were advantageous, but that differences in the relative proportions of neutral sites would influence the rate of molecular evolution (that is, more neutral sites would produce a faster overall rate of change) (BOX 3). c | Nearly neutral theory: the fate of mutations with only slight positive or negative effect on fitness will depend on how population size affects the outcome (BOX 4). Figure modified with permission from REF. 22.

- die meisten nichtsynonymen Substitutionen sind nicht perfekt neutral, sondern sind schwach vorteilhaft oder nachteilig
- Ihre Fixierung hängt demnach von schwacher Selektion und Genetischer Drift ab
- Selektion ist wichtiger in großen Populationen, da in kleinen Populationen die Fixierung durch den Zufall überwiegt

## Kräfte in der „nearly neutral theory“

### Variable Mutationsraten

(G-Zeit, metabol. Rate, Reparatur)



### Nukleotid-Substitutionsraten



### „Nearly Neutral Mutations“

(schwache Selektion, Populationsgröße)

## Die „nearly neutral theory“ erklärt Molekulare Uhr an nicht-synonymen Positionen

Lange G-Zeit  
(niedrige Mutationsrate)

Kurze G-Zeit  
(hohe Mutationsrate)



Realzeit-abhängige  
Molekulare Uhr



Kleine Population  
(höhere Fix-Wahrscheinl.)

Große Population  
(geringere Fix-Wahrscheinl.)

- Mutationen in Spezies mit kleinen Populationen haben höhere Fix-Wahrscheinlichkeit. Da jedoch diese Arten oft längere G-Zeiten haben, ist die Rate neu entstehender Mutationen gering

## Anwendung der Molekularen Uhr

- Gibt es validierte paläontologische/biogeographische Befunde, die den Zeitrahmen vorgeben (Eichung der Uhr)?
- Verhält sich das Gen/Protein „clock-like“? > relative rate test
- Welche „Uhr“ soll man nehmen?

$K_s$  oder  $K_{s(4)}$  → - gute Auflösung für nahe verwandte Taxa (= schnelle Uhr)  
- Problem: schnelle Sättigung  
- Achtung bei codon usage bias oder unterschiedlichem GC/AT-Kontext

$K_A$  oder  $A_S$  → für entfernt verwandte Taxa

## Welche Gene/Proteine für welche Fragestellung?

<u>Marker</u>	<u>aufzulösende Taxa</u>	<u>Evol.geschwindigkeit</u>
Globine	z.B. Säuger	mittel
rRNA	Pro/Eukaryoten	sehr langsam
rRNA- od. Histon-Spacer	Spezies/ Subspezies	schnell
mtDNA	Spezies	schnell
mtDNA- Kontrollregion	Populationen	sehr schnell

## Störfaktoren für die „Molekulare Uhr“



- Stark unterschiedliche Generationszeiten bzw. Anzahl der Replikationsrunden in der Keimbahn in verschiedenen phylogenetischen Linien
- artspezifische Unterschiede in der DNA-Reparatur
- Wechsel des Genomkontexts orthologer Gene z.B. durch eine Translokation
- Funktionswechsel/-verlust des Gens in einer Linie
- Genkonversion zwischen paralogen Genkopien
- Horizontaler Gentransfer
- Unkenntnis, ob die orthologen Allele zweier Taxa verglichen worden sind

Diversity and Distributions, (Diversity Distrib.) (2006) 12, 35–48



### Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times

Frank Rutschmann

Institute of Systematic Botany, University of Zürich, Zollikonstrasse 107, CH-8008 Zürich, Switzerland

Correspondence: Frank Rutschmann, Institute of Systematic Botany, University of Zurich, Zollikonstrasse 107, CH-8008 Zürich, Switzerland. E-mail: frank@plant.ch

#### ABSTRACT

This article reviews the most common methods used today for estimating divergence times and rates of molecular evolution. The methods are grouped into three main classes: (1) methods that use a molecular clock and one global rate of substitution, (2) methods that correct for rate heterogeneity, and (3) methods that try to incorporate rate heterogeneity. Additionally, links to the most important literature on molecular dating are given, including articles comparing the performance of different methods, papers that investigate problems related to taxon, gene and partition sampling, and literature discussing highly debated issues like calibration strategies and uncertainties, dating precision and the calculation of error estimates.

#### Keywords

Divergence time estimation, molecular dating methods, rate heterogeneity, review.