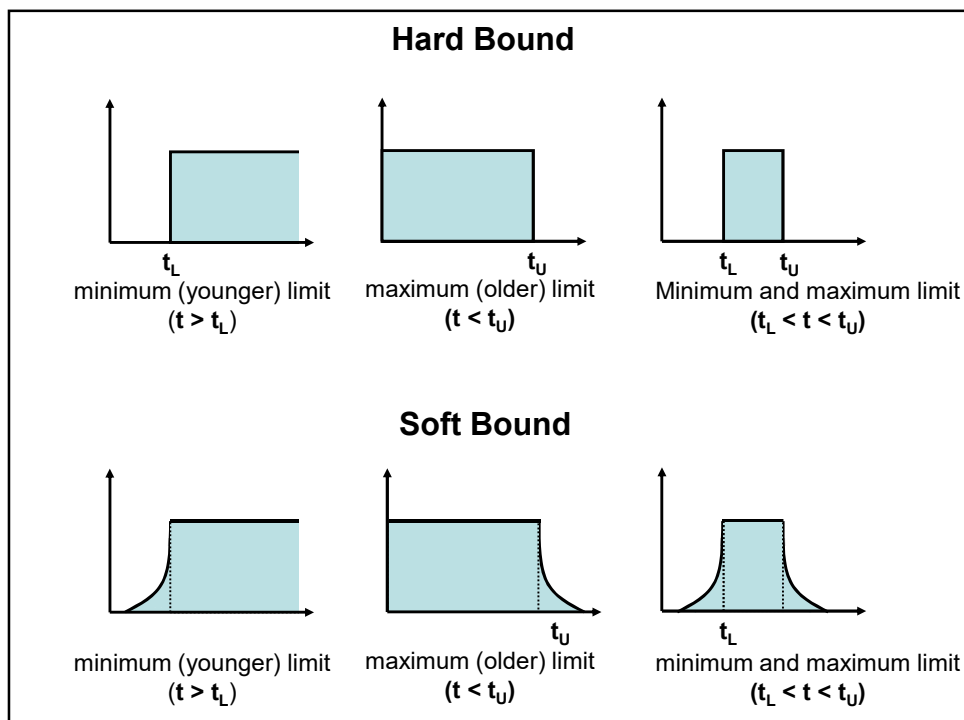


## Bayesian Method for Estimating Branching Dates

- Modeling rate variation among different lineages
- Calibration based on fossil evidence

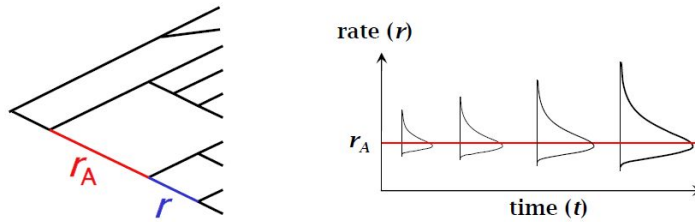
Thorne-Kishino – multidivtime

Ziheng Yang – MCMCtree

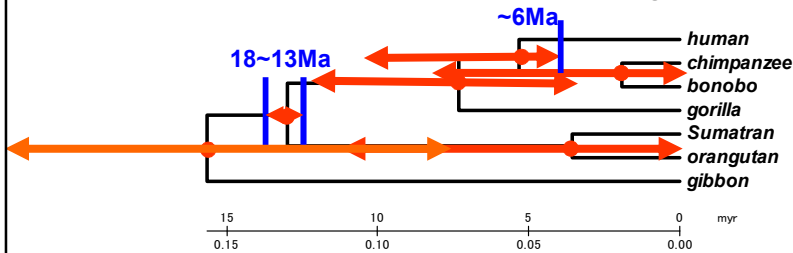


## Prior model of rate drift (geometric Brownian motion)

The rate  $r$  of a branch (node) is a random variable centred around the ancestral rate  $r_A$ . The variance  $\sigma^2$  determines how variable the rates are on the tree.

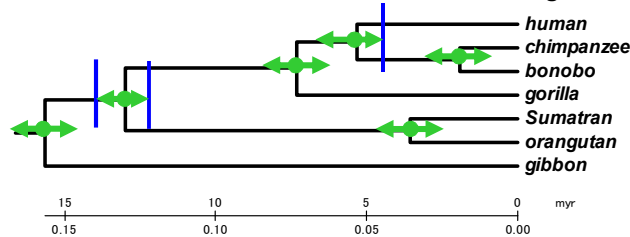


### Prior distribution of the branching dates

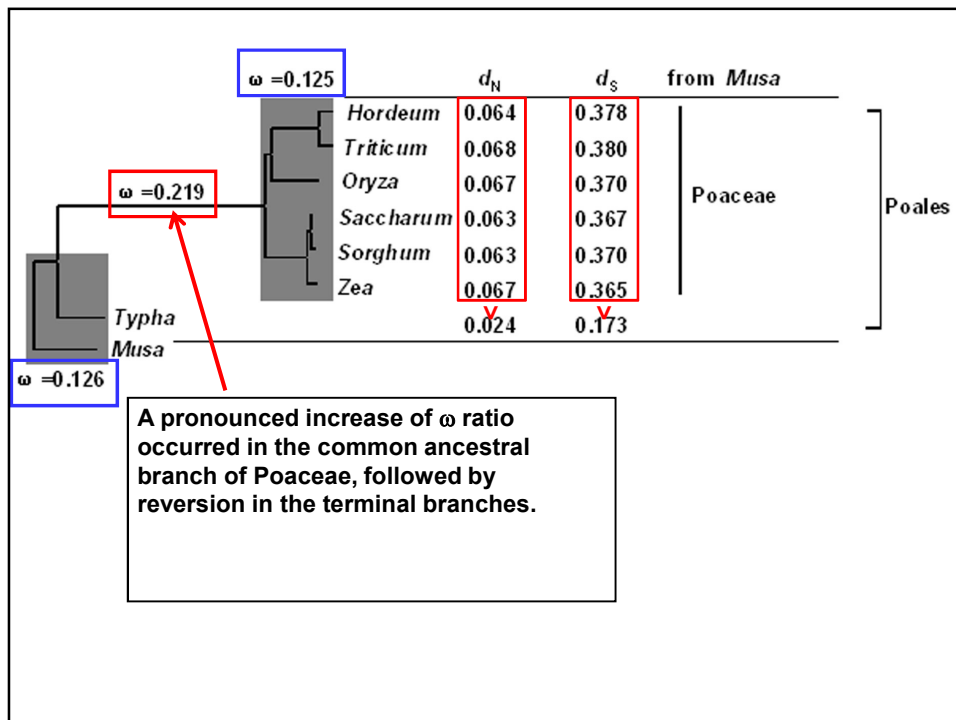
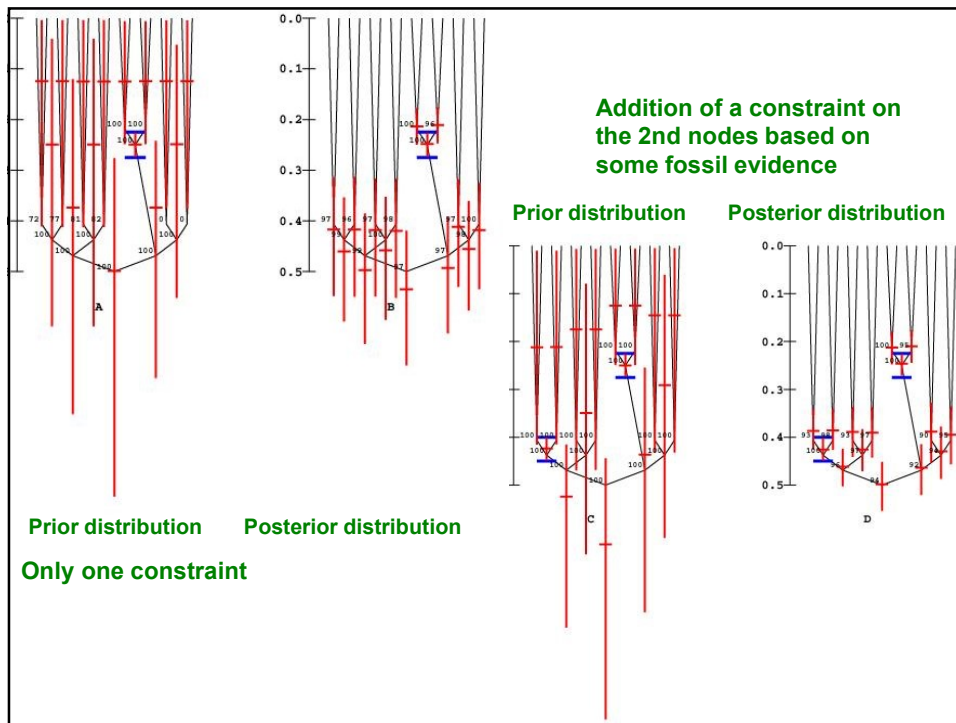


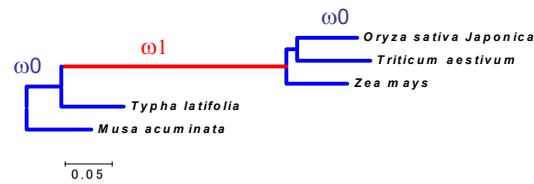
The prior distribution of the divergence times were given based on the prior distribution of “root (mean) substitution rate parameter”, “rate drift parameter” and “fossil constraint”

### Posterior distribution of the branching dates



By estimating the likelihood function from the data, the posterior distribution of the divergence times is obtained.





Is the high evolutionary rate of grasses correlated with the high  $\omega$  ratio or with the adaptive evolution? Is the rate high in grasses living today, or low similarly as the  $\omega$  ratio reverted to the low level typical of the basal lineages?

To answer this question, we must explicitly take account of the time-scale of evolution.

### Time estimation with a relaxed molecular clock, allowing variable rates among lineages

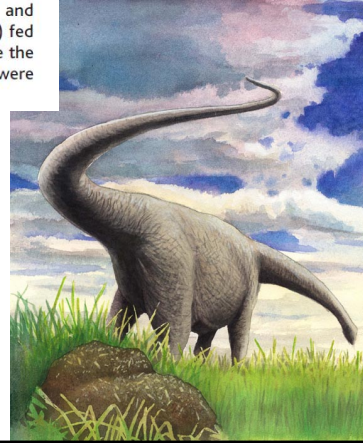
To obtain reliable time estimates, it is important to use reliable time constraints based on fossil evidence. The poor quality of the fossil record for early grasses, however, has prevented us from addressing this issue. Recent findings of grass phytoliths in Cretaceous dinosaur coprolites provided evidence that the major group of core Poaceae had already diversified before 65 Ma.

## Dinosaur Coprolites and the Early Evolution of Grasses and Grazers

Vandana Prasad,<sup>1</sup> Caroline A. E. Strömberg,<sup>2\*</sup>  
Habib Alimohammadian,<sup>3</sup> Ashok Sahni<sup>3</sup>

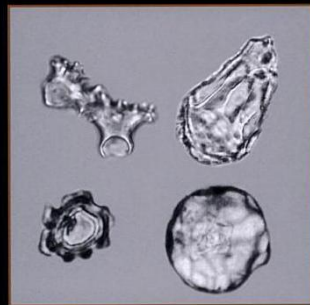
SCIENCE VOL 310 18 NOVEMBER 2005

Silicified plant tissues (phytoliths) preserved in Late Cretaceous coprolites from India show that at least five taxa from extant grass (Poaceae) subclades were present on the Indian subcontinent during the latest Cretaceous. This taxonomic diversity suggests that crown-group Poaceae had diversified and spread in Gondwana before India became geographically isolated. Other phytoliths extracted from the coprolites (from dicotyledons, conifers, and palms) suggest that the suspected dung producers (titanosaur sauropods) fed indiscriminately on a wide range of plants. These data also make plausible the hypothesis that gondwanatherian mammals with hypsodont cheek teeth were grazers.



## PHYTOLITHS

A COMPREHENSIVE GUIDE FOR  
ARCHAEOLOGISTS AND  
PALEOECOLOGISTS

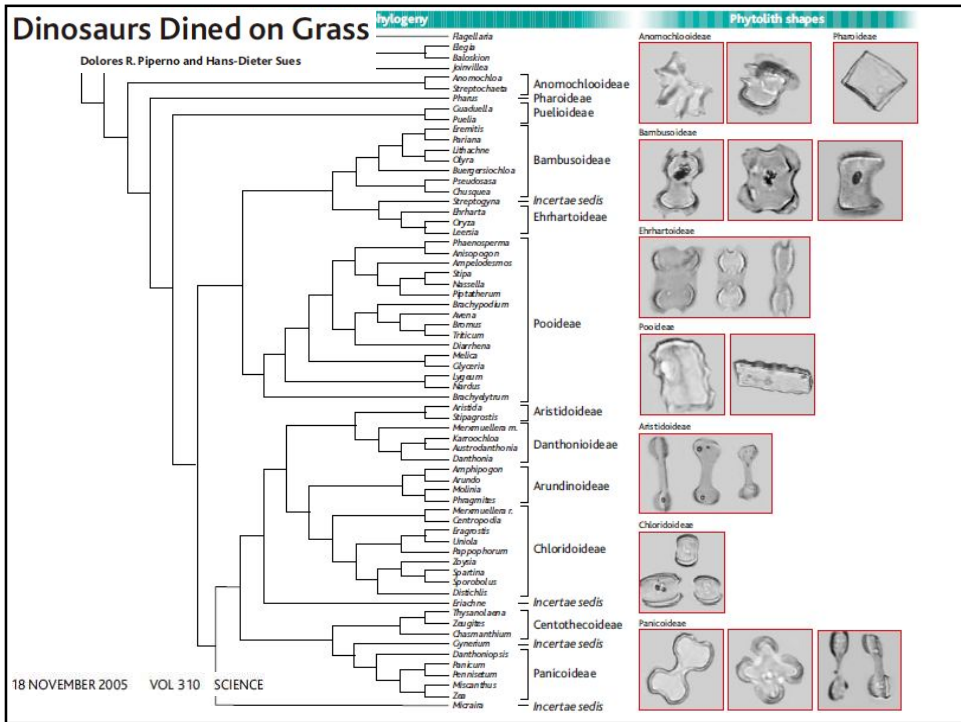
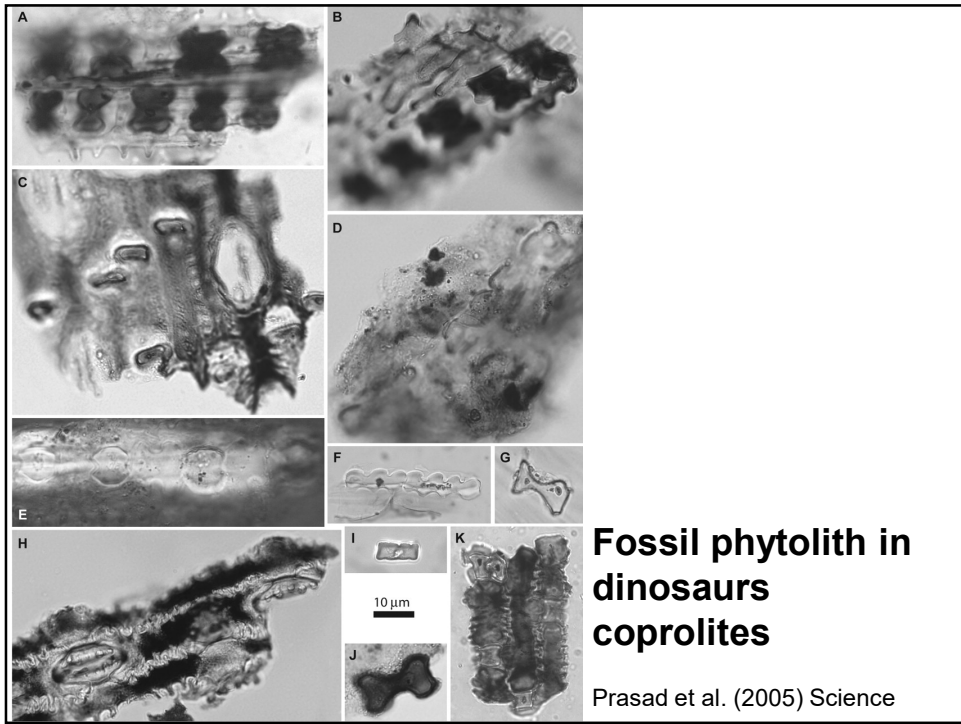


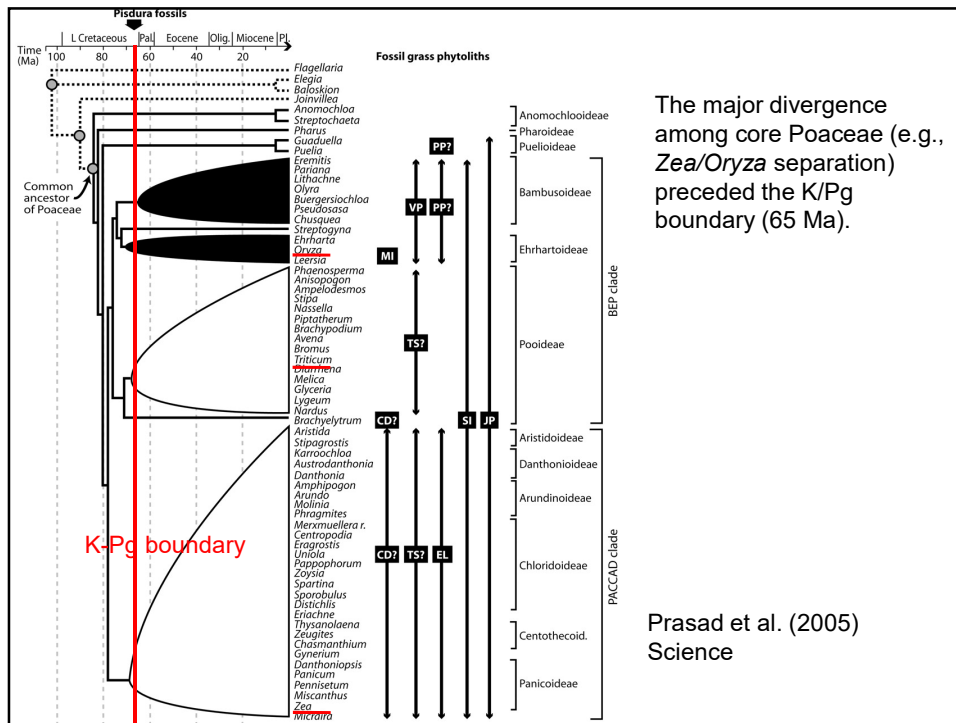
DOLORES R. PIPERNO

Phytoliths (Plant opal):  
Silica bodies, that precipitate  
in plant tissues, are especially  
abundant in the grass family  
Poaceae. This is regarded as  
a defense against herbivores.  
Since shapes of phytoliths  
differ among different  
lineages of grasses, we can  
evaluate the ancient diversity  
of grasses by analyzing  
phytoliths found in fossils.

Alta Mira Press  
Oxford, UK (2006)







**Bayesian methods for time estimation with a relaxed clock**

	Multidivtime (Jeff Thorne)	MCMCtree (Ziheng Yang)
Calibrations	Hard bounds	Soft bounds
Prior on rates	Auto-correlated rates	Auto-correlated rates + Independent rates

- Since the bounds provided by fossil evidence are not always correct, soft bounds in calibrating a clock is preferable to hard bounds. By using the soft bounds, the probability that the true divergence time is outside the bounds is small but not zero.
- In the Jeff Thorne's "Multidivtime", rates are assumed to be correlated between neighboring branches, but it might be interesting to check whether this assumption holds by using both the auto-correlated and the independent rates models implemented in "MCMCtree" by Ziheng Yang.

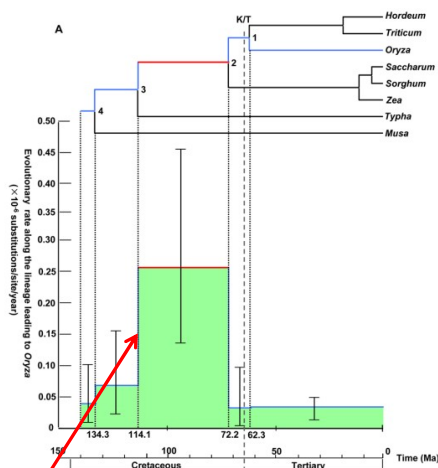


Table S3. Posterior estimates of divergence times by MCMCTREE in PAML (Yang, 2007). 95% highest posterior density (HPD) is shown in parentheses. Rate refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 2. The >65 Ma constraint is given to the *Zea/Oryza* separation

Node	Independent rates model		Correlated rates model	
	Time (Ma)	Rate	Time (Ma)	Rate
1 ( <i>Musa</i> )	115.18 (67.47-137.89)	0.059 (0.017-0.168)	88.15 (47.14-129.47)	0.109 (0.056-0.240)
2 ( <i>Typha</i> )	98.30 (59.70-122.54)	0.085 (0.032-0.236)	78.38 (42.99-119.04)	0.131 (0.061-0.318)
3 ( <i>Zea/Oryza</i> )	<b>67.76</b> (47.84-84.90)	<b>0.291</b> (0.149-0.714)	45.84 (19.66-71.66)	0.278 (0.143-0.961)
4 ( <i>Oryza/Triticum</i> )	54.44 (34.22-71.48)	0.038 (0.011-0.115)	36.51 (15.51-61.98)	0.053 (0.017-0.146)

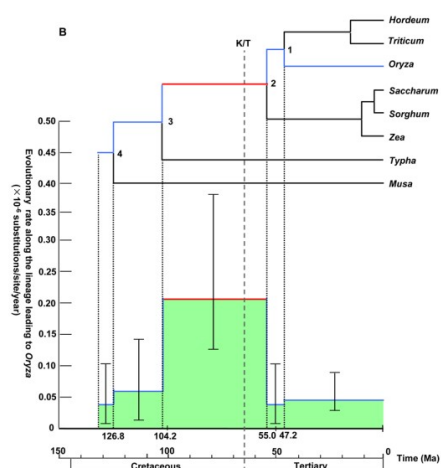
The independent-rates model gives time estimates more in accord with the fossil evidence than the correlated-rates model, suggesting that the elevation of the rate in grasses was abrupt.

>65 Ma constraint to *Zea/Oryza* separation



Rate change along the lineage from the root to *Oryza*.

No constraint to *Zea/Oryza* separation



## Conclusion

1. Elevation of the evolutionary rate and adaptive evolution in chloroplast genomes were observed in the common ancestral branch of grasses after they diverged from *Typha*.

2. Our finding highlights the need for paying attention to rates of internal branches rather than averaging along a whole lineage in addressing the rate heterogeneity problem.

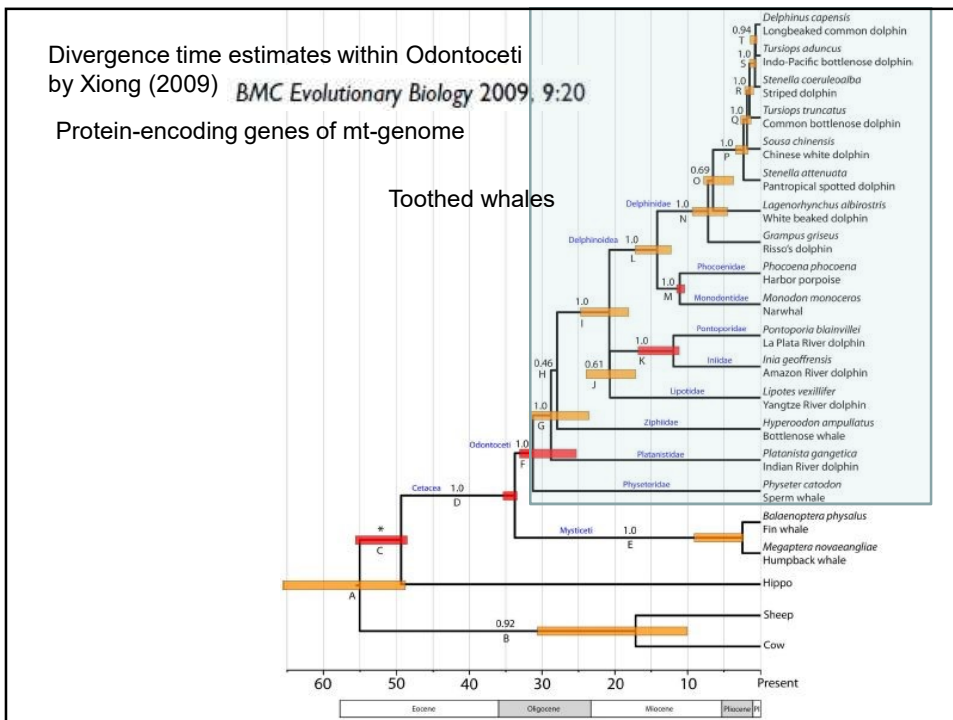
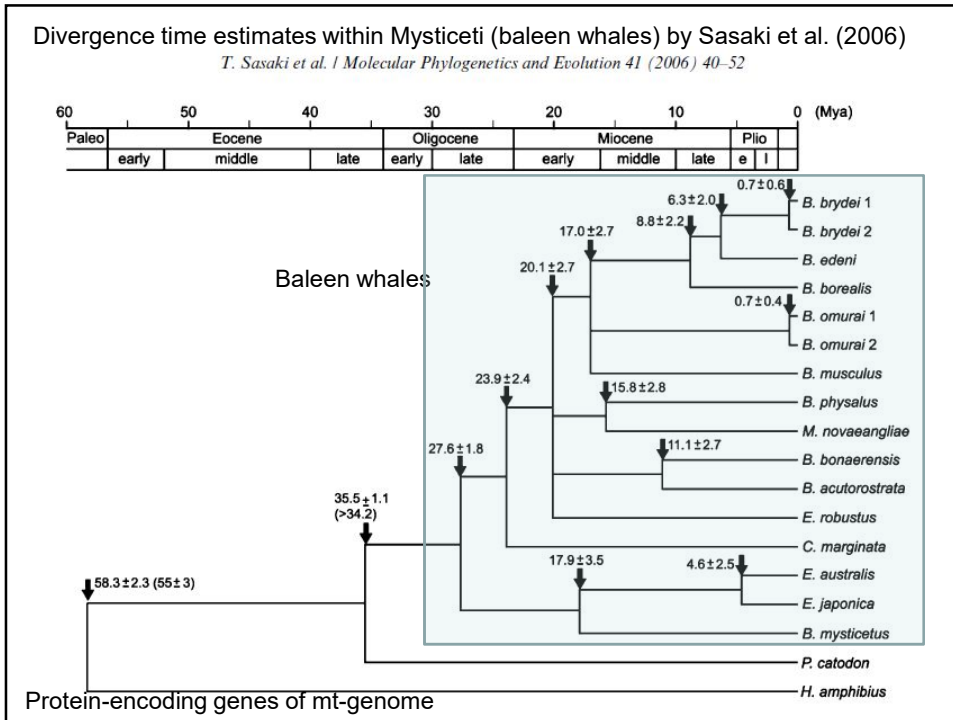
歯カジン

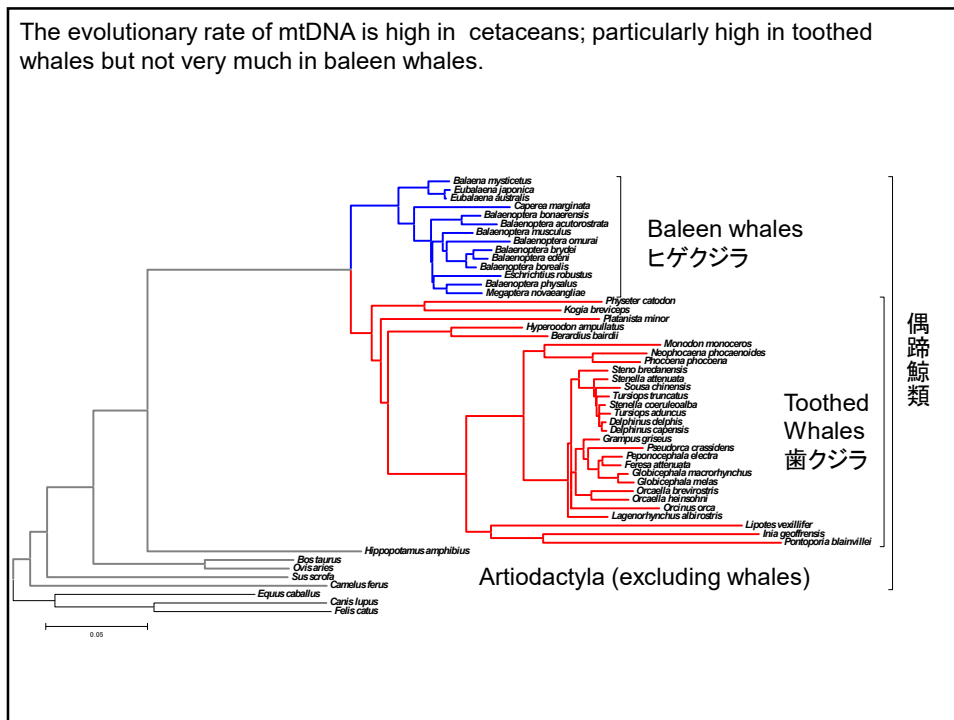
ヒゲカジン

Cetacea

- Odontoceti**
  - Delphinoidae
    - haji (*Lipotes vexillifer*)
    - Amazon river dolphin (*Inia geoffrensis*)
    - beluga (*Delphinapterus leucas*)
    - bottle-nosed dolphin (*Tursiops truncatus*)
    - killer whale (*Orcinus orca*)
    - humpback whale (*Phocoena phocaena*)
    - Stejneger's beaked whale (*Mesoplodon stejnegeri*)
  - Ganges dolphin (*Platanista gangetica*)
  - sperm whale (*Physeter catodon*)
  - bowhead whale (*Balaena mysticetus*)
  - right whale (*Balaena glacialis*)
- Mysticeti**
  - pygmy right whale (*Caperea marginata*)
  - gray whale (*Eschrichtius robustus*)
  - Brude's whale (*Balaenoptera edeni*)
  - humpback whale (*Balaenoptera musculus*)
  - fin whale (*Balaenoptera physalus*)
  - humpback whale (*Megaptera novaeangliae*)
  - minke whale (*Balaenoptera acutorostrata*)

長谷川政美(2011) 改変  
「新図説・動物の起源と進化」  
八坂書房





## Data set

Mitochondrial protein (12種類) → amino acid (mtmam+F+I model)  
Mitochondrial rRNA(2種類) → nucleotide (GTR+I model)

- **23 species data set (Sasaki2006)**

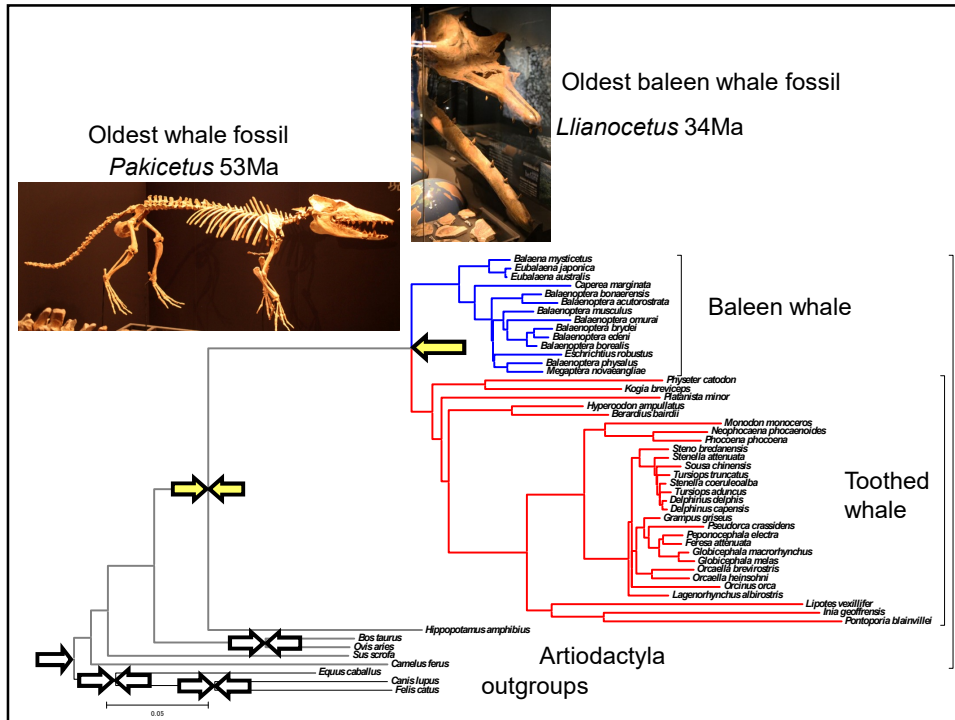
[Baleen whale 14種](#)、[Toothed whale 1種](#)、[Artiodactyla 5種](#)、[Outgroup \(Carnivora, Perissodactyla\) 3種](#)

- **39 species data (Xiong2009)**

[Baleen whale 2種](#)、[Toothed whale 29種](#)、[Artiodactyla 5種](#)、[Outgroup \(Carnivora, Perissodactyla\) 3種](#)

- **51 species data**

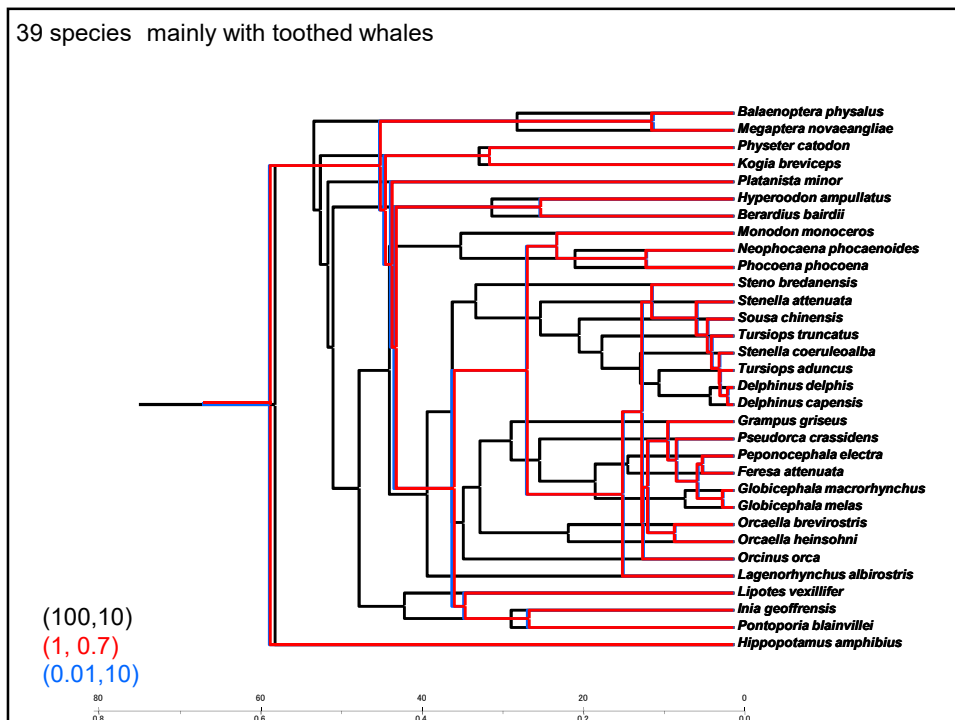
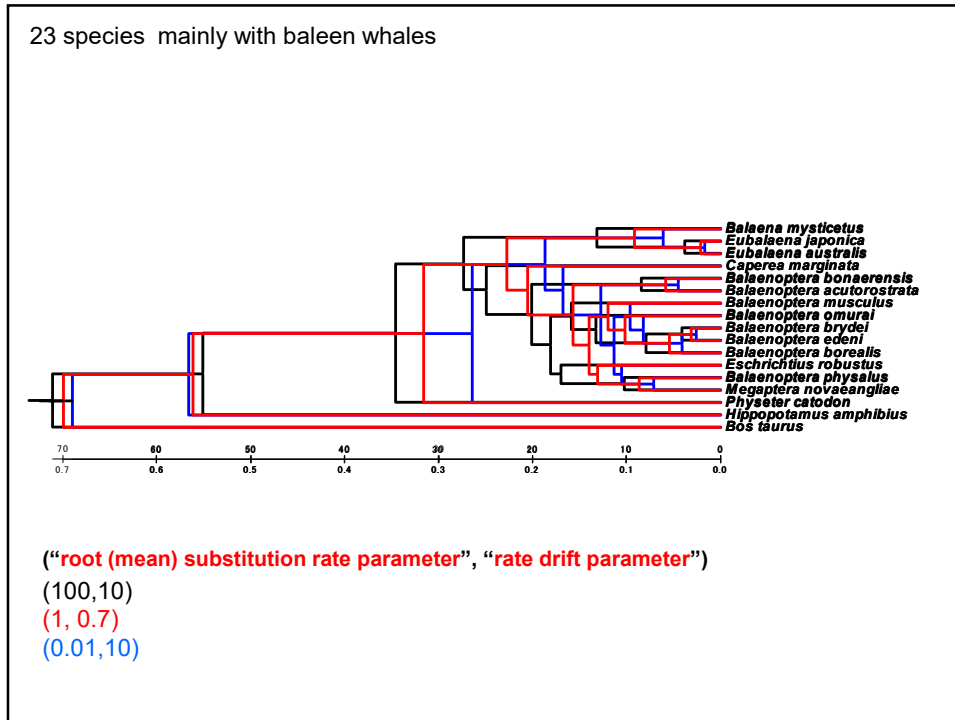
[Baleen whale 14種](#)、[Toothed whale 29種](#)、[Artiodactyla 5種](#)、[Outgroup \(Carnivora, Perissodactyla\) 3種](#)



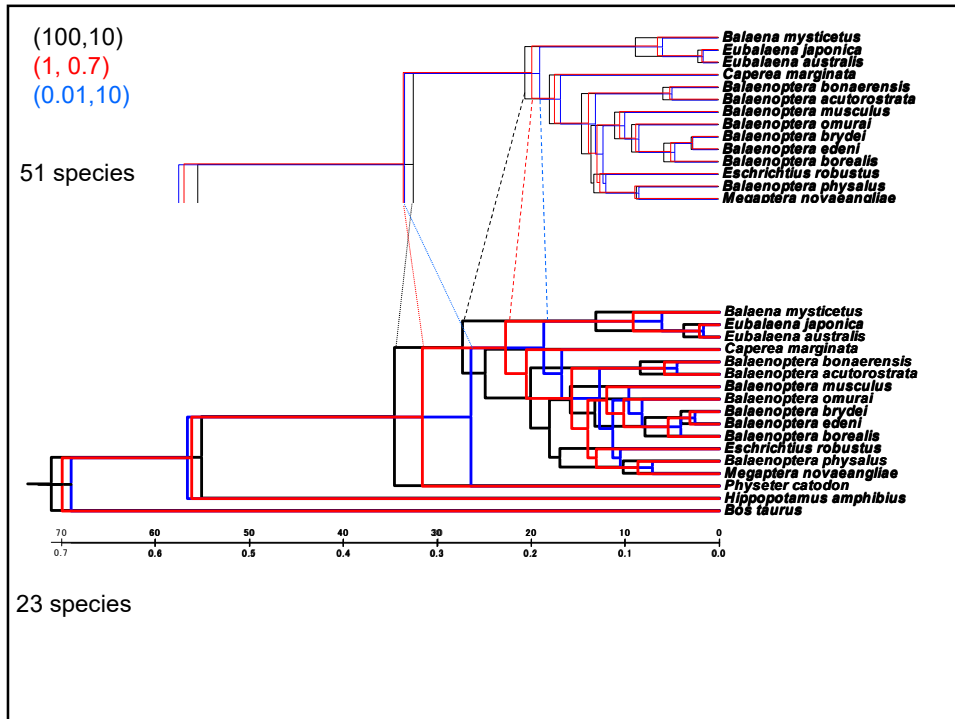
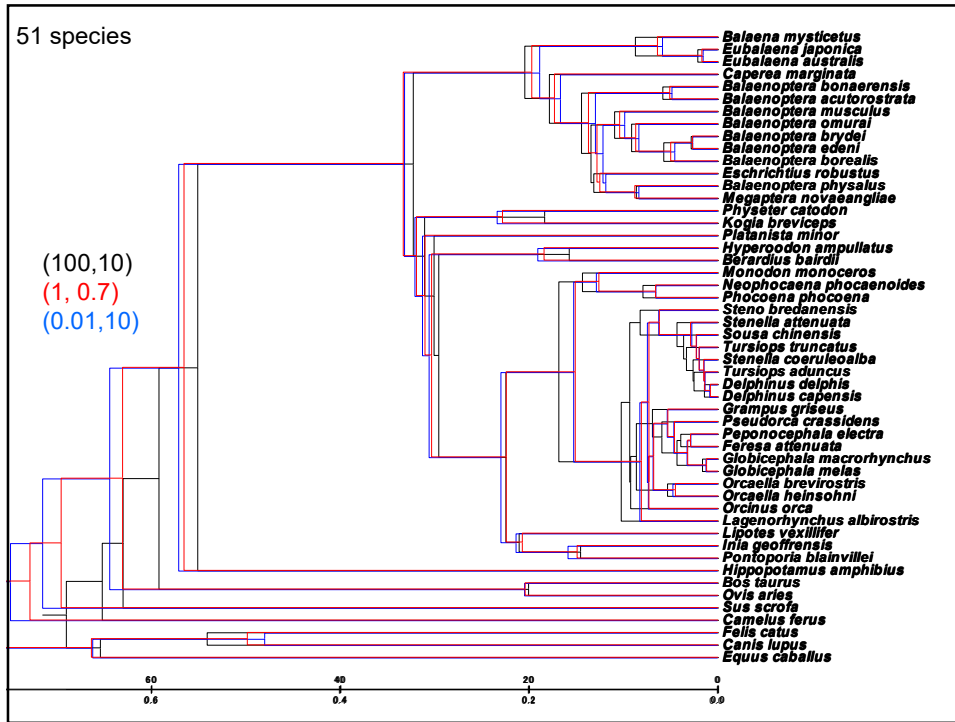
**Rgene:** Root (mean) rate parameter

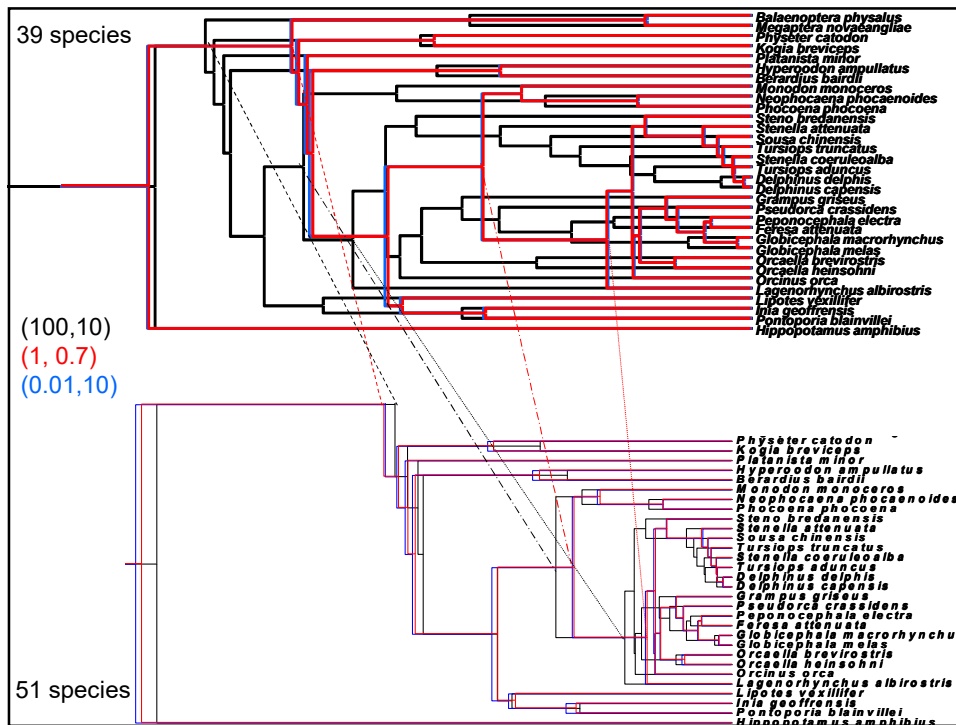
$\sigma^2$  :Rate drift parameter

Specifies how variable the rates are across branches.  
 Large  $\sigma^2$  indicates the clock is highly violated, and small  $\sigma^2$  indicates the clock roughly holds.







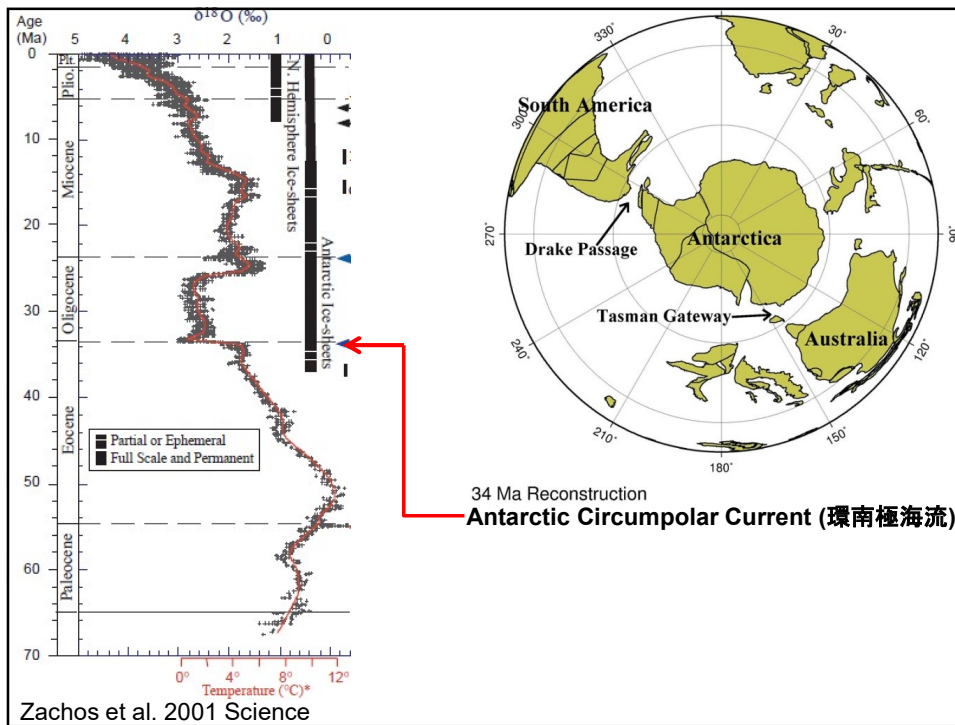
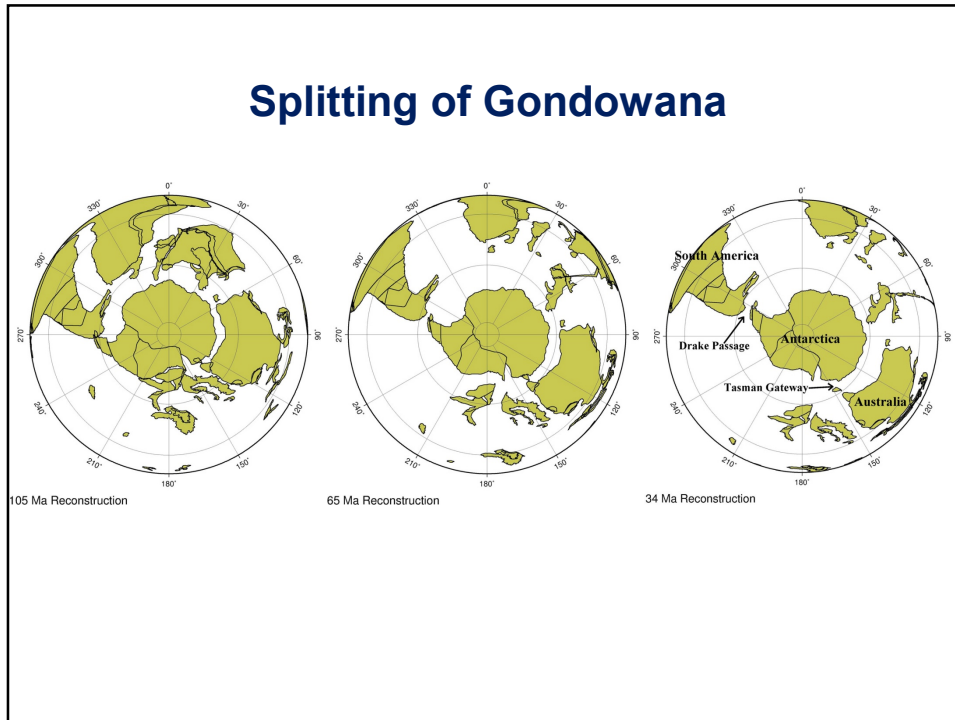


## Conclusion

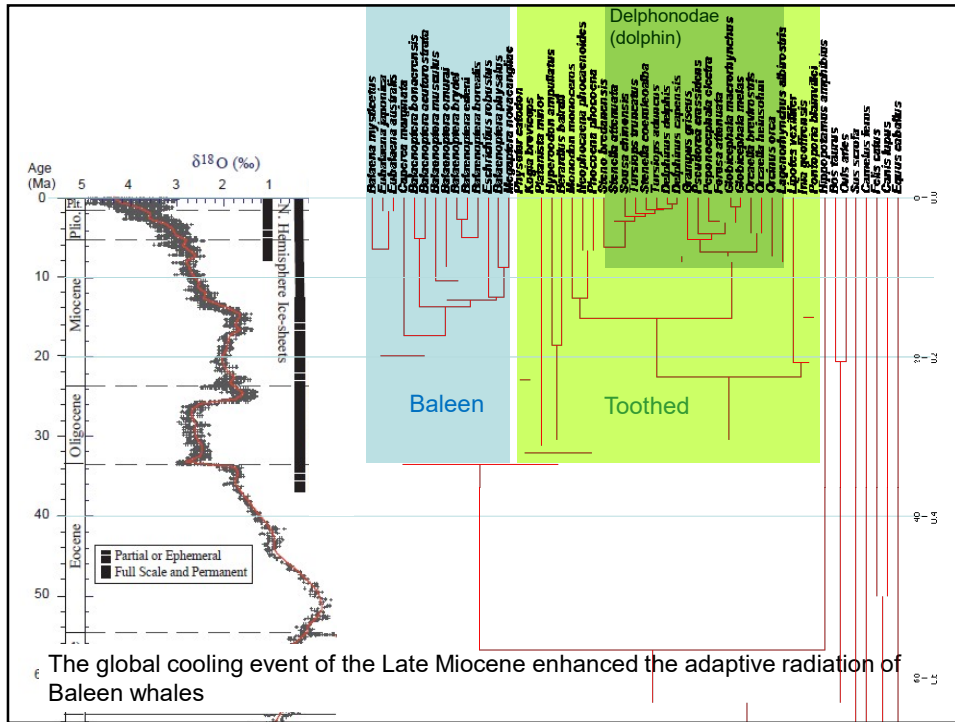
- When the rate differs among different groups, the estimates highly depends on the bias of taxon sampling.
- Dense taxon sampling could give robust estimates which do not depend on the prior distributions.

T. Yonezawa, N. Kohno, Y. Cao & M. Hasegawa (in preparation)

## Splitting of Gondwana



Zachos et al. 2001 Science



**Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny**

Mario dos Reis<sup>1</sup>, Jun Inoue<sup>1,2</sup>, Masami Hasegawa<sup>3</sup>, Robert J. Asher<sup>4</sup>, Philip C. J. Donoghue<sup>5,\*</sup> and Ziheng Yang<sup>\*,1</sup>

**Analysis of 2 stages**

- 14,632 nuclear gene (20.6x10<sup>6</sup> bp) from 36 mammalian species (33 eutherians + 2 marsupials + 1 monotreme)
- The posterior probability of the divergence times obtained in the 1<sup>st</sup> stage was used to construct the time prior for mt-genome analysis of 274 mammal species.

