

# **ABC for ancestral inference**

**Richard Wilkinson and Simon Tavaré**

ABCiL — May 5th 2011

## ABC for ancestral inference: timescales

- Tumour evolution
  - mitotic division, evolution within individual
  - agent-based models (cellular Potts model)
- Coalescents
  - human evolution (where ABC began)
  - SerialSimCoal
- Phylogeny of primates
  - species evolution

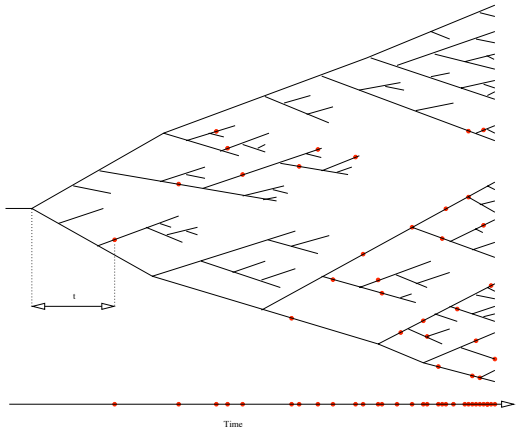
# The Primates

doi:10.1039/c9pr00001a

Posted



# Primate Evolution



## Reconciling molecular and fossil records?

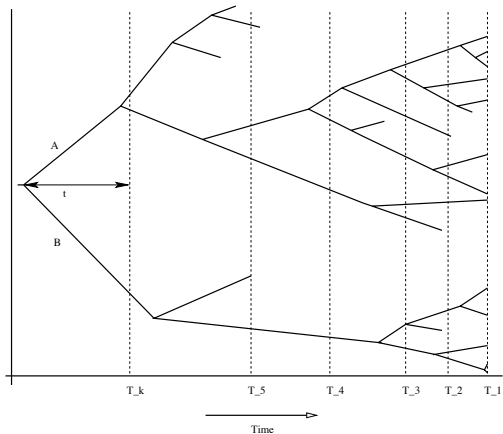
- Extant primates are strepsirrhines (lemurs and lorises) and haplorhines (tarsiers and anthropoids)
- Molecular estimate of time of divergence is approximately 90 mya
- Fossil record suggests 60-65 mya
- Fossil record is patchy

Problem: Use the fossil record to estimate the age of the last common ancestor of extant primates

## Primate Data

Epoch	$k$	$T_k$	Observed number of species ( $D_k$ )
Late Pleistocene	1	0.15	19
Middle Pleistocene	2	0.9	28
Early Pleistocene	3	1.8	22
Late Pliocene	4	3.6	47
Early Pliocene	5	5.3	11
Late Miocene	6	11.2	38
Middle Miocene	7	16.4	46
Early Miocene	8	23.8	36
Late Oligocene	9	28.5	4
Early Oligocene	10	33.7	20
Late Eocene	11	37.0	32
Middle Eocene	12	49.0	103
Early Eocene	13	54.8	68
Pre-Eocene	14		0

# The evolutionary process



## What happened?

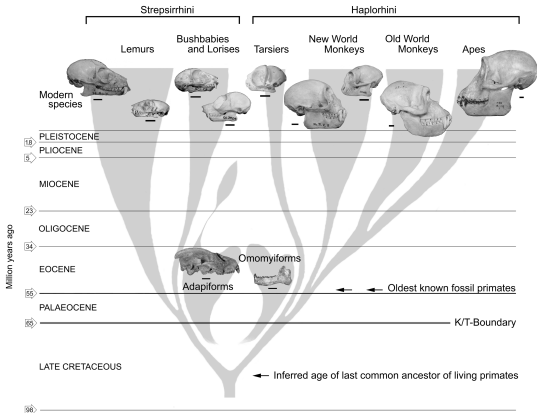
- Average sampling fraction of 5.7%
  - upper 95% limit 7.4%
- Estimated divergence time 81.5 mya
  - 95% CI (72.0, 89.6) mya

avaré, Marshall, Will, Soligo & Martin *Nature*, 2002

- Pravda, Times, BBC, . . . , assorted religious fanatics,  
...



# Primate Evolution



## Why more?

- Bayesian approach more natural
- Allows us to incorporate prior information
- Sampling fractions
  - probability of finding a fossil in bin  $i$  is  $\alpha_i$
  - $\alpha = \alpha \mathbf{p}$ ,  $\mathbf{p}$  known
  - reasonable?
- Other models for finds?
- Allowing for dinosaur extinction at K/T boundary?

## Fossil record: ABC approach

Data can be thought of in two parts:

- (a) the observed number of fossils  $F_{\text{obs}}$  found
- (b) the proportions  $p_{j,\text{obs}}$  found in  $j$ th bin

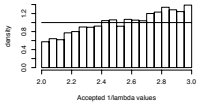
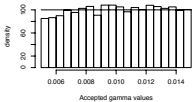
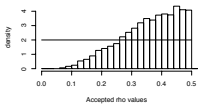
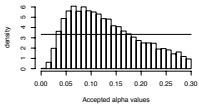
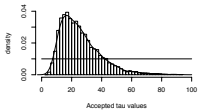
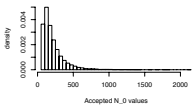
Note: no data summaries used here

A suitable metric might be

$$\left| \frac{F}{F_{\text{obs}}} - 1 \right| + \sum_{j=1}^{k+1} |p_j - p_{j,\text{obs}}|$$

# Results

$$\epsilon = 0.1$$



## Sensitivity: Exploring Other Models

dings : doi:10.1038/npre.2011.5963.1

One advantage of ABC – it is easy to change the input ...

- Choice of  $\rho$
- Demography
- Sampling fractions
- K/T crash 65 mya
  - the time of origin of primates is even further back in the Cretaceous
- Poisson sampling scheme: length in bin matters
- Dating other split points

## Tweak metric

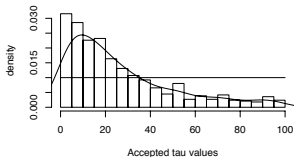
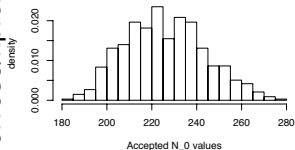
- The observed  $N_0$  values are too small
  - require  $N_0 > 235$
  - change the metric

$$\rho(\mathcal{D}, \mathcal{D}') = \sum_{i=1}^k \left| \frac{D_i}{D_+} - \frac{D'_i}{D'_+} \right| + \left| \frac{D'_+}{D_+} - 1 \right| + \left| \frac{N'_0}{N_0} - 1 \right|$$

- Penalises trees with  $N_0$  values far from 235

# Results: $\epsilon = 0.3$

	min	LQ	Median	mean	UQ	Max
$N_0$	184	212	224	226	238	279
$\tau$	0.0	8.0	18.6	26.3	36.8	99.5

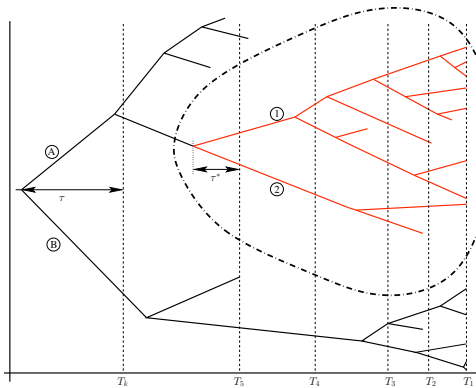


# Old World/New World Split

Epoch	$k$	$T_k$	Hap/Strep number of species ( $D_k$ )	Plat/Cat number of species ( $D_k^*$ )
Late Pleistocene	1	0.15	19	19
Middle Pleistocene	2	0.9	28	28
Early Pleistocene	3	1.8	22	22
Late Pliocene	4	3.6	47	44
Early Pliocene	5	5.3	11	10
Late Miocene	6	11.2	38	33
Middle Miocene	7	16.4	46	43
Early Miocene	8	23.8	36	30
Late Oligocene	9	28.5	4	3
Early Oligocene	10	33.7	20	6
Late Eocene	11	37.0	32	2
Middle Eocene	12	49.0	103	0
Early Eocene	13	54.8	68	
Pre-Eocene	14		0	



## Dating Two Splits



## Details

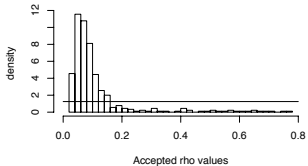
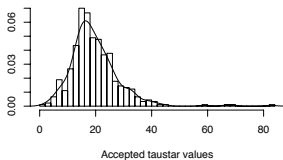
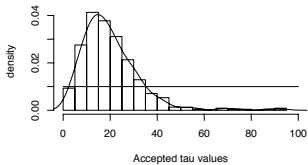
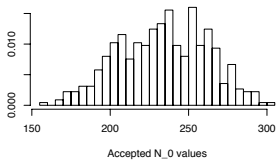
- $N_0 = 235$  species for the Strep/Hap,
- $\epsilon = 0.4$  for both metrics

	min	LQ	Median	mean	UQ	Max
$N_0$	159	212	234	233	254	303
$\tau$	0.9	12.1	17.6	20.1	25.3	94.5
$\tau^*$	1.6	14.5	18.2	19.6	23.5	82.9

The median posterior sampling fractions ( $\times 100$ )

$\alpha_1$	$\alpha_2$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_8$	$\alpha_9$	$\alpha_{10}$	$\alpha_{11}$	$\alpha_{12}$	$\alpha_{13}$	$\alpha_{14}$
8	10	12	3	6	7	1	8	22	41	80	1
8	8	8	4	4	4	1	4	8	8	8	1

# Posteriors

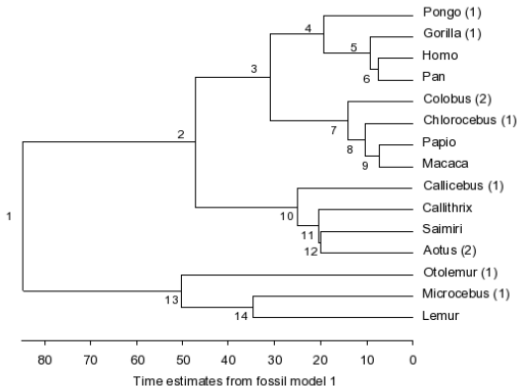


## Combining fossil record with molecular data

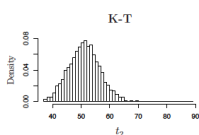
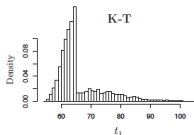
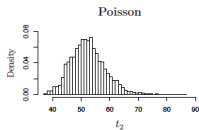
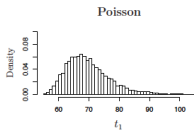
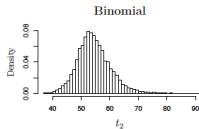
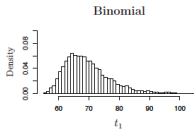
Yesterday's posterior is tomorrow's prior ...

- Estimate posterior for two primate divergence times
- Use as prior for dating nodes from molecular data  
(*mcmctree*)
- Data are updated from earlier analysis

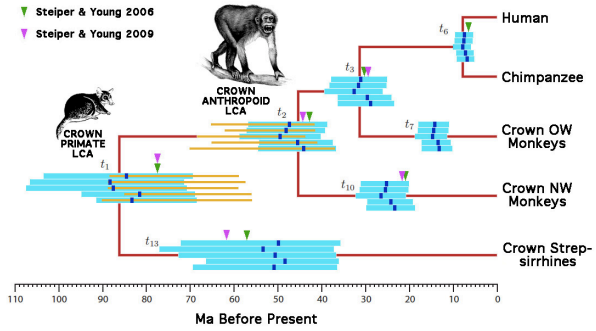
# The phylogeny of the species



# The priors



# The molecular data



# The drawbacks

ScienceDaily®

Your source for the latest research news

[Did eyes evolve?](#) Look clearly at all the evidence. You can't ignore th

[Buy Trilobites in stock](#) Free Authentication Certificate 100s of Kids

[The DNA Ancestry Project](#) Discover Your Ancestry with DNA. Find E



News

Articles

Videos

Images

Books

Health & Medicine

Mind & Brain

Plants & Animals

Earth & Climate

Space & Time

Matter & E

## Science News

[Share](#) [Blog](#) [Cite](#)

### New Statistical Model Moves Human Evolution Back Three Million Years

*ScienceDaily* (Nov. 5, 2010) — Evolutionary divergence of humans and chimpanzees likely occurred some 8 million years ago rather than the 5 million year estimate widely accepted by scientists, a new statistical model suggests.

#### See Also:

##### Plants & Animals

- [Evolutionary Biology](#)
- [Nature](#)

##### Computers & Math

- [Statistics](#)
- [Computer Modeling](#)

##### Fossils & Ruins

- [Fossils](#)

The revised estimate of when the human species parted ways from its closest primate relatives should enable scientists to better interpret the history of human evolution, said Robert D. Martin, curator of biological anthropology at the Field Museum, and a co-author of the new study appearing in the journal *Systematic Biology*.

Working with mathematicians, anthropologists and molecular



*A new statistical model suggests that evolutionary divergence of humans from chimpanzees likely occurred some 8 million years ago, rather than the 5 million year estimate widely accepted by scientists. (Credit: iStockphoto/Eric Gevaert)*



## Collaborators

Richard Wilkinson	[University of Nottingham]
Bob Martin	[Field Museum, Chicago]
Christophe Soligo	[University College London]
Ziheng Yang	[UCL]
Michael Steiper	[Hunter College, New York]

The dangers of ABC **H. L. Mencken:**

*For every complex problem, there is an answer that is short, simple and wrong*

Why use ABC? **J. Galsworthy:**

*idealism increases in direct proportion to one's distance from the problem*



## References

- ST, Marshall C, Will O, Soligo C & Martin R (2002) Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*, **416**, 726–729
- Wilkinson R & ST (2009) Estimating primate divergence times by using conditioned birth-and-death processes. *Theor Popul Biol*, **75**, 278–295
- Wilkinson R, Steiper M, Soligo C, Martin R, Yang Z & ST (2010) Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology*, **60**, 16–31, 2011.