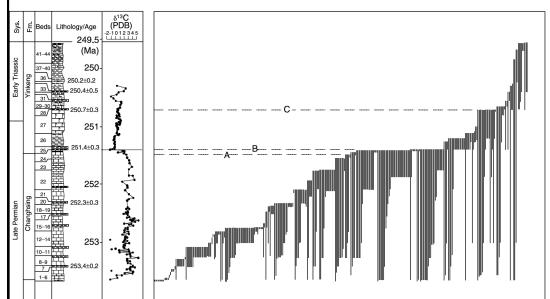
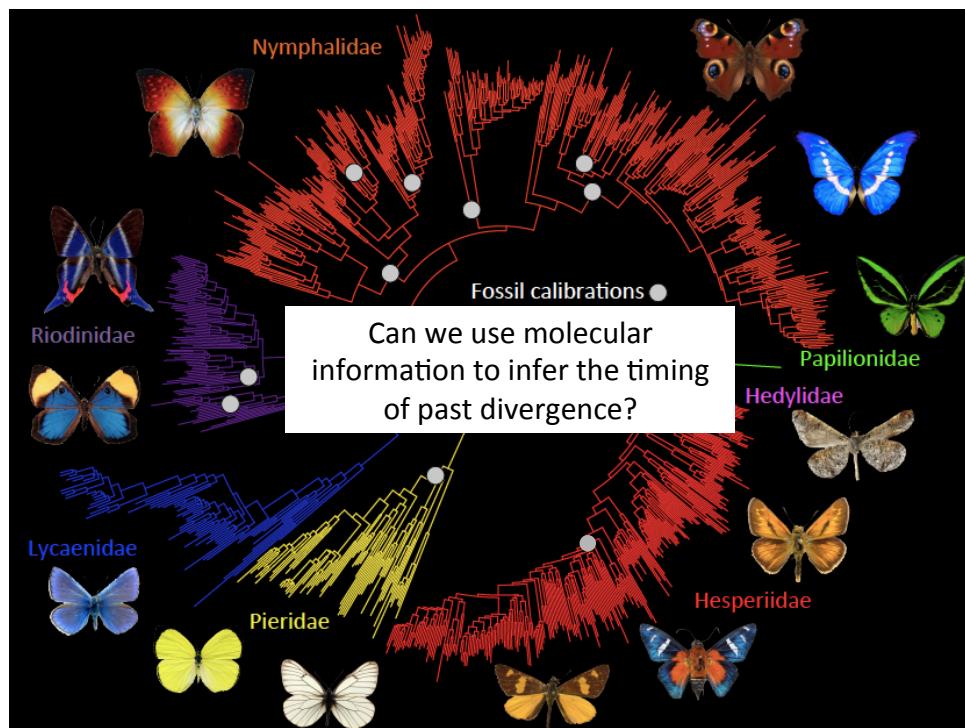
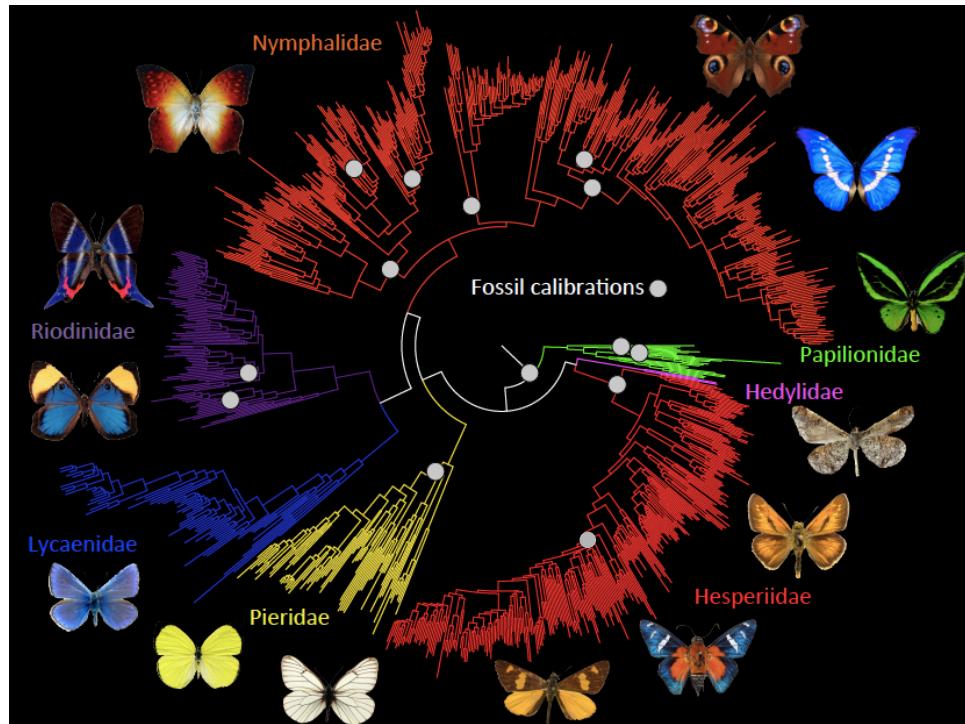


Introduction to molecular dating methods

The fossil record is the direct evidence of past events and the time at which they occurred



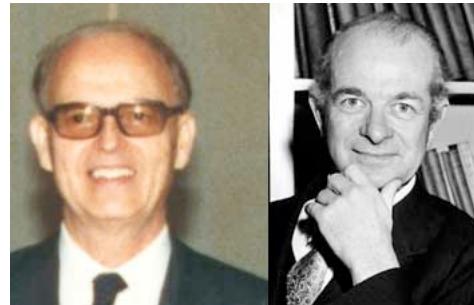


The Molecular Clock

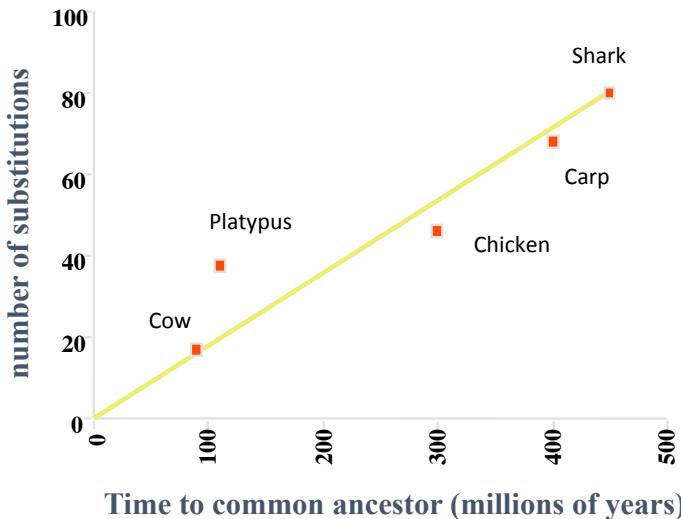
Going back to ancient times

Is there a molecular clock?

- The idea of a molecular clock was initially suggested by Zuckerkandl and Pauling in 1962 and 1965



The molecular clock for alpha-globin:
Each point represents the number of substitutions separating
each animal from humans



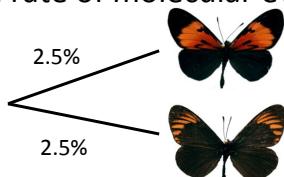
Is there a molecular clock?

- The idea of a molecular clock was initially suggested by Zuckerkandl and Pauling in 1962 and 1965
- They noted that rates of amino acid replacements in animal haemoglobins were roughly proportional to time - as judged against the fossil record

=> implies the existence of a sort of molecular clock ticking faster or slower for different genes but at a more or less constant rate for a genes among different lineages

The molecular clock hypothesis

- Assumes an equal rate of molecular evolution over time



- A 5% difference between species means they have each diverged 2.5% since their common ancestor
- If a fossil or other evidence will let us calibrate this clock we can convert % difference to years

Assumptions of a perfect clock

- Molecular change is a linear function of time with substitutions accumulating following a Poisson distribution - any variation will be stochastic [imagine 1 substitution / million yrs]
- Rate of change is equal across all sites and lineages
- The phylogeny can be estimated without error

Assumptions of a perfect clock (cont.)

- The number of substitutions along each lineage can be estimated without error
- Calibration dates for all times of divergence used to calculate the rate of the molecular clock are known without error
- A regression of time on number of substitutions can be conducted without error

Dating with a molecular clock

- “Universal Molecular Clocks”
- Calibrations proposed for various taxa / genes
- eg. mtDNA molecular clock of animals
 - ~ 2% sequence divergence per million years for vertebrates
 - ~ 1% sequence divergence per million years for invertebrates

There is no universal molecular clock

- The initial proposal saw the clock as a Poisson process with a constant rate
- Now known to be more complex - differences in rates occur for:
 - different sites in a molecule
 - different genes
 - different regions of genomes
 - different genomes in the same cell
 - different taxonomic groups for the same gene
- There is no universal molecular clock

Problems

- Saturation

Ancest GGC~~G~~CG

Seq 1 AGCG~~A~~G

Seq 2 GCGG~~A~~C

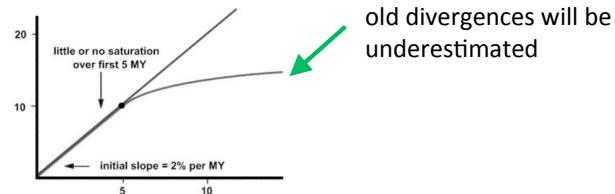
Number of changes

1 2 3

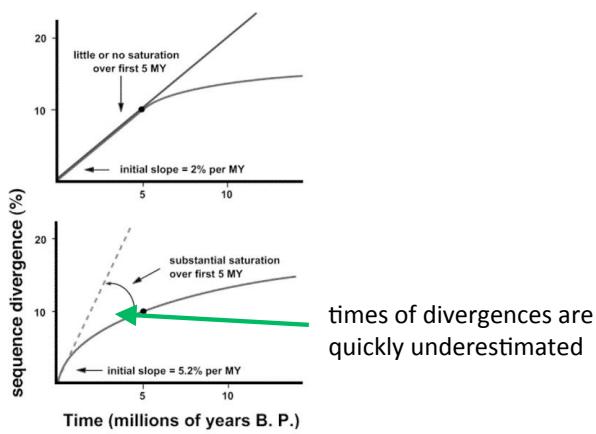
Seq 1 C → G → T → A

Seq 2 C → A
 1

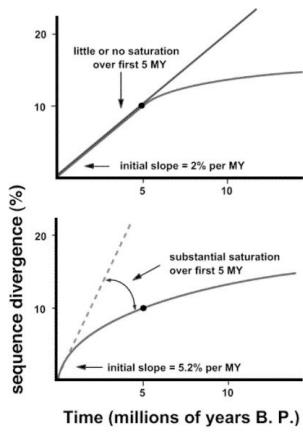
Saturation problems



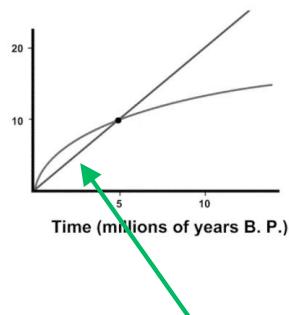
Saturation problems



Saturation problems



And if we underestimate the rate of substitution...

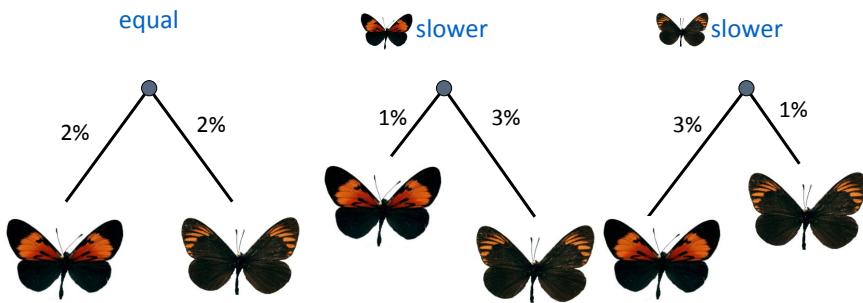


Both recent divergences will be overestimated and older divergences will be underestimated

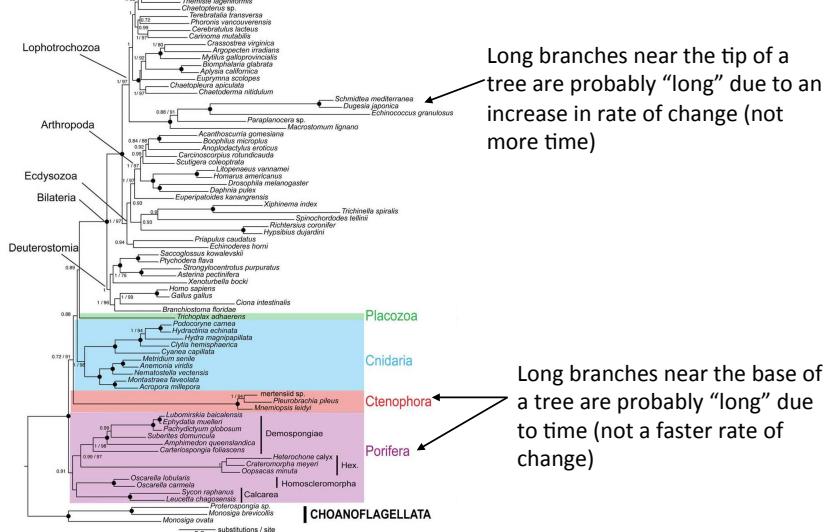
Problems

- Saturation
- Rate Heterogeneity - violation of homogeneity

No universal molecular clock



Molecular distance from to is the same in all cases



Pick et al (2010) MBE 27:1983-1987

Teasing apart RATE and TIME

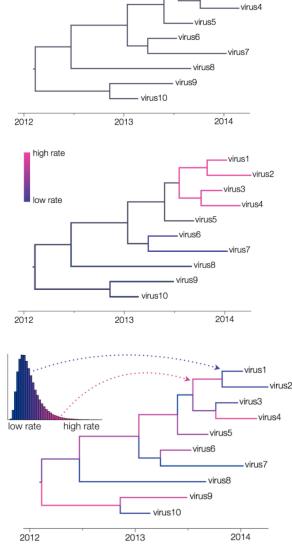
- Branch lengths are proportional to:

$$\text{RATE} * \text{TIME}$$

- If rates are constant then lengths are proportional to time
- If rates are not constant then *we have a hard time relating branch lengths to time*

Molecular clocks can be relaxed

- Strict or "global" clock
 - Many programs/methods/algorithms
- Local clocks
 - Maximum Likelihood (PAML, QDate)
 - Mean path length (Pathd8)
- Relaxed clocks
 - Non-parametric rate smoothing (r8s)
 - Penalized likelihood (r8s)
 - Bayesian, fixed tree (multidivtime, PhyBayes)
 - Bayesian, tree co-estimated (BEAST, MrBayes)



What is a relaxed clock?

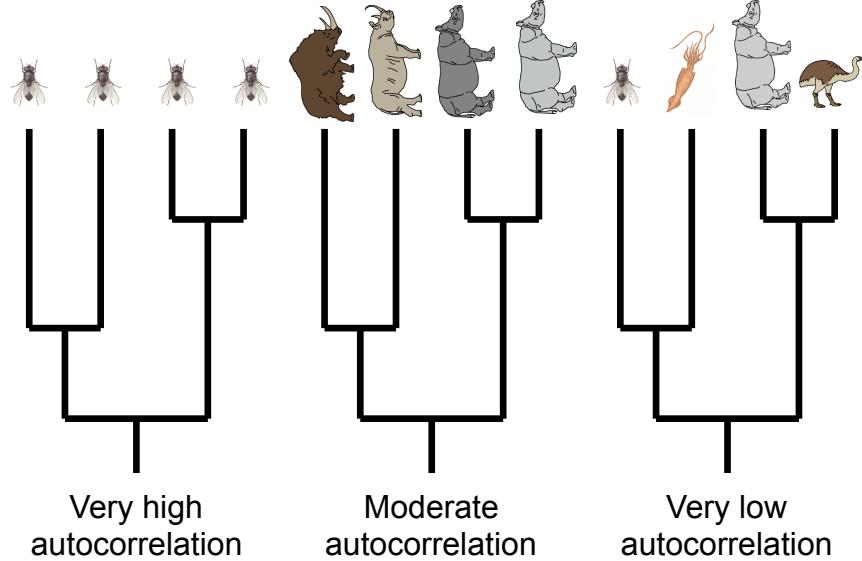
- Strict clock: rate identical in all branches
- Relaxed clock: rate allowed to vary among branches
 - Autocorrelated relaxed clock: rates in adjacent branches are related
 - Uncorrelated relaxed clock: rates identically and independently distributed among branches



Autocorrelated relaxed clocks

- Fixed topologies are input!
- Treat substitution rate as a heritable trait, so that it can 'evolve' through the tree
- Rate is assumed to be tied to:
 - Life history traits (e.g., generation time, population size, body size)
 - Cellular/biochemical environment
- Available in r8s, multidivtime, PhyBayes, BEAST, PAML

Autocorrelated relaxed clocks



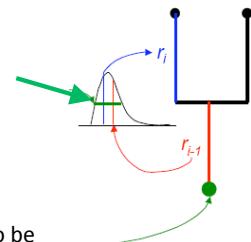
25

Modeling autocorrelation

- Model of autocorrelated rate change used to describe prior distribution of rates
- Lognormal
 - $\log(r_i) \sim N(\log(r_{i-1}), v t)$

v controls the s.d. of the distribution

Further assumption needs to be made about rate at the root



Uncorrelated relaxed clocks

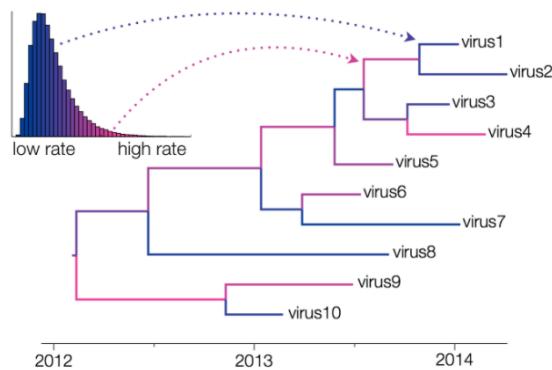
- Models available in *BEAST*

- **Lognormal distribution**

Most rates cluster around the mean

- **Exponential distribution**

Most rates are quite low



Lognormal uncorrelated relaxed clock

- In the uncorrelated lognormal relaxed clock, two statistics can be obtained:

- 1. Coefficient of variation of rates**

Measures the rate variation among branches

A value of 0 indicates clocklike evolution

- 2. Covariance of rates**

Measures autocorrelation of rates between adjacent branches

Problems

- Saturation
- Rate Heterogeneity - violation of homogeneity
- Calibration

Calibration



Separating rate and time

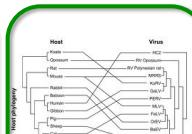
- Information about rate
 - Substitution rate obtained from an independent study
- Information about time – *prior information*:



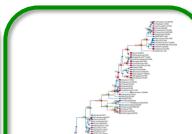
Fossil record



Biogeography



Ecology



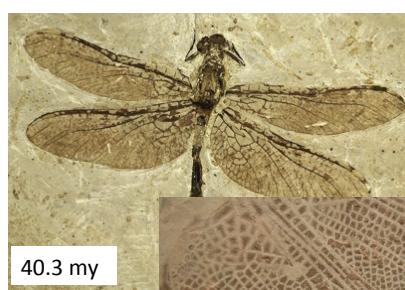
Sampling times

Calibration: Fossil record

- Fossil record provides minimum estimates of divergence times



Fossil record



Identified as belonging to the family Aeshnidae and genus *Aeshna*



informative

Calibration: Fossil record

- Fossil record provides minimum estimates of divergence times

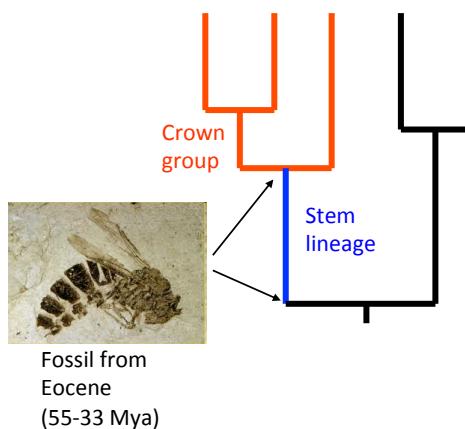


Identified as belonging to the family Phasianidae, to genus *Gallus*, to the species *Gallus gallus domesticus*

Minimum age for the birds...
BUT not informative

Problems with fossils

- Incompleteness of fossil record
- Identification
 - Species / Genus / Family?
- Position
 - Stem or crown?
- Which date?
 - Min / Mid / Max of Epoch?

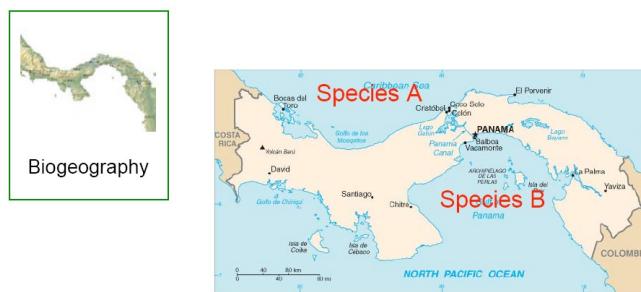


Calibration errors

- Preservational bias
 - Hard parts
 - Environment, proximity to water bodies
 - Age
 - Sampling effort
 - Taxonomic affinity
 - Fragmentary fossils
 - Extinct, stem lineages
 - Stratigraphic and isotopic dating errors

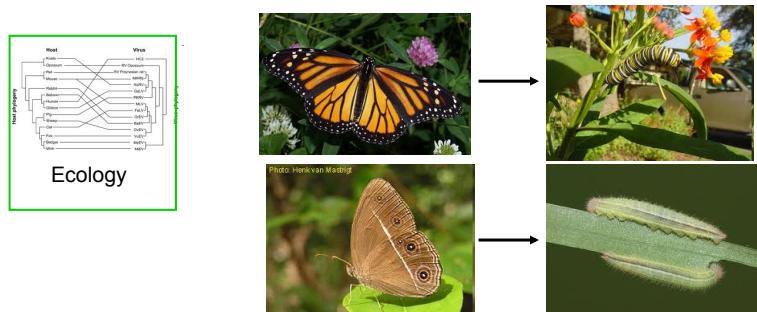
Calibration: Biogeography

- Biogeographic events can provide maximum estimates of divergence times



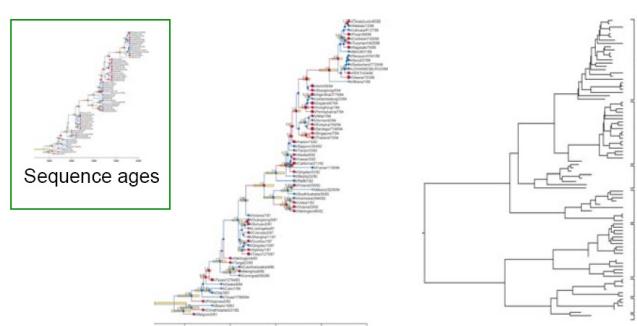
Calibration: Ecology

- Knowledge of tight ecological associations can be used to provide maximum estimates of divergence times



Calibration: Sequence ages

- Sequence ages provide sufficient age information for e.g. viruses



Calibration in Bayesian framework

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

θ : model (substitution model(s), tree, etc)

prior: prior expectation we have for parameters of the model

Calibration in Bayesian framework

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

θ : model (substitution model(s), tree, etc)

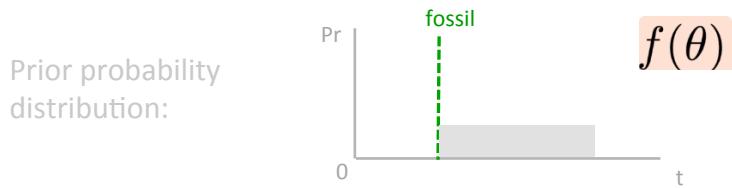
prior: prior expectation we have for parameters of the model

For example: age of nodes based on fossil information

Calibration in Bayesian framework

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

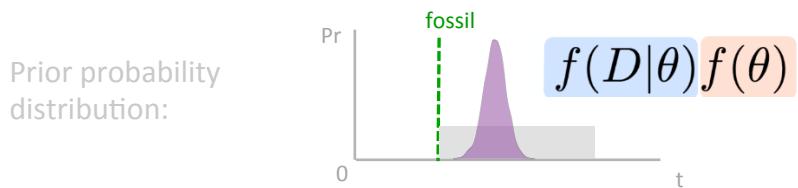
For example: age of nodes based on fossil information



Calibration in Bayesian framework

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

For example: age of nodes based on fossil information

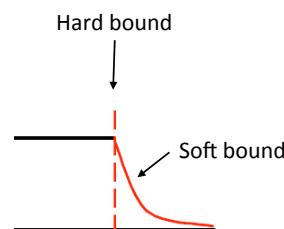


Calibration types

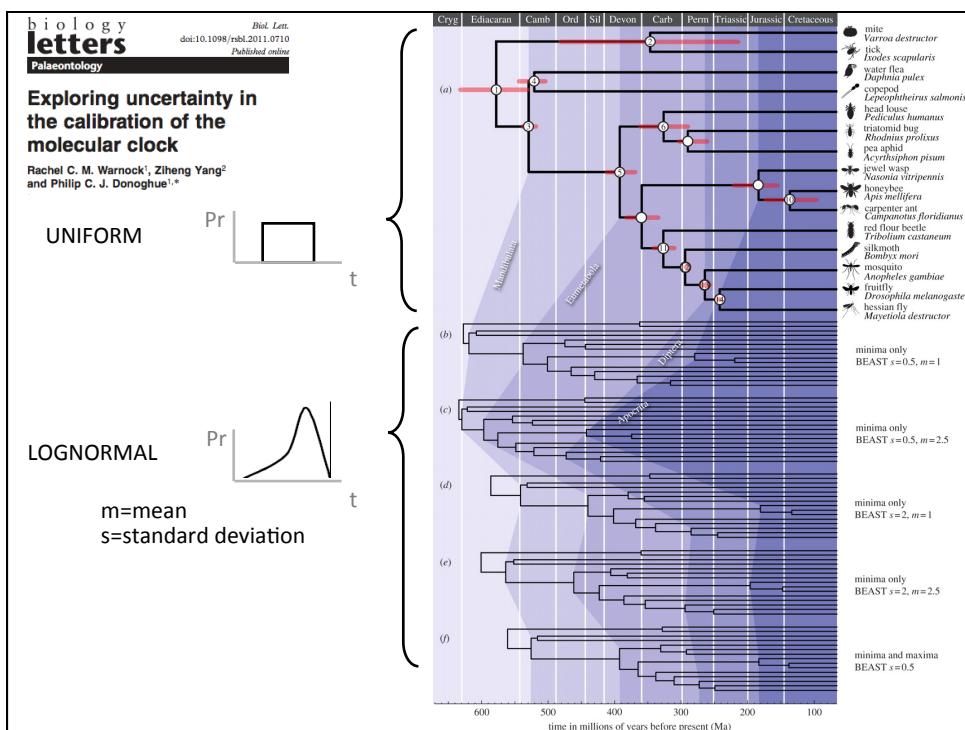
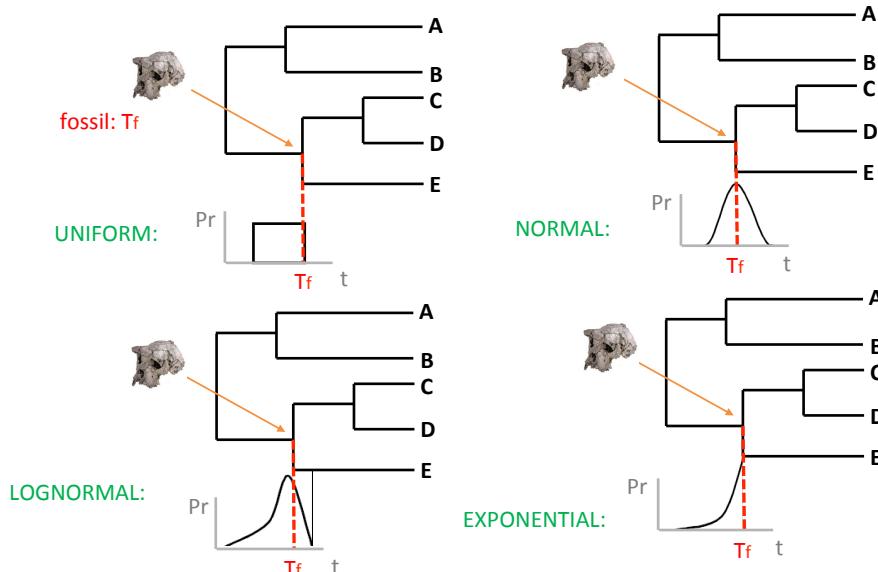
- Point calibrations
- Hard minimum/maximum bounds
- Soft minimum/maximum bounds
- Parametric prior distributions
 - Normal distribution
 - Lognormal distribution
 - Exponential distribution

Hard/Soft Bounds

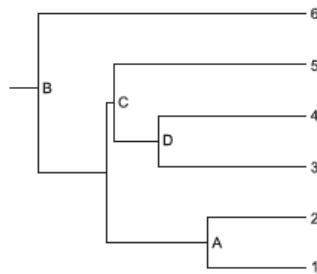
- Extension of hard bounds
- Soft:
 - Assign non-zero probability to values outside bound
 - Able to forgive calibration errors



Prior distributions



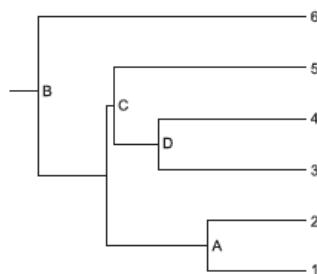
Multiple calibrations



- ▶ Molecular-clock estimates can be sensitive to the positions of the calibrations in the phylogenetic tree, especially when only a single or very few calibrations are available
- ▶ a small number of calibrations can lead to a biased estimate of the substitution rate if there is substantial among-lineage rate variation

7

Multiple calibrations

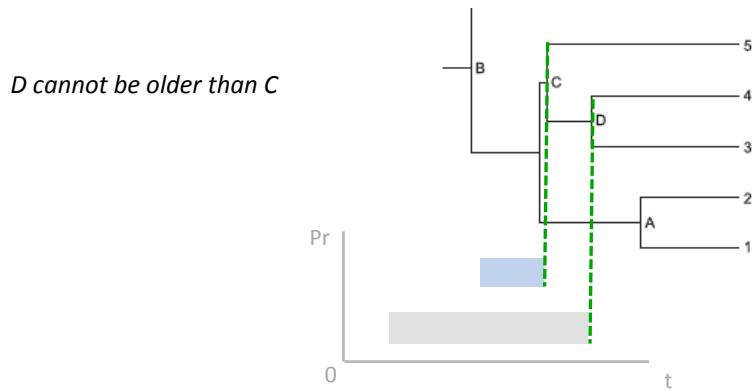


- ▶ can improve the accuracy of date estimates in the presence of taxon undersampling
 - ▶ substitution rate is primarily estimated from the branches between the calibrating nodes and the tips
- => deeper calibrations capture a larger proportion of the overall genetic variation.

8

Multiple calibrations

- ▶ Be careful: priors interact with each others
- ▶ For example, node orders



Multiple calibrations

- ▶ Be careful: priors interact with each others
- ▶ For example, node orders
- ▶ Marginal priors resulting from prior interactions can differ from the initial user prior
 - This can be visualized by removing the data and running the same analysis

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

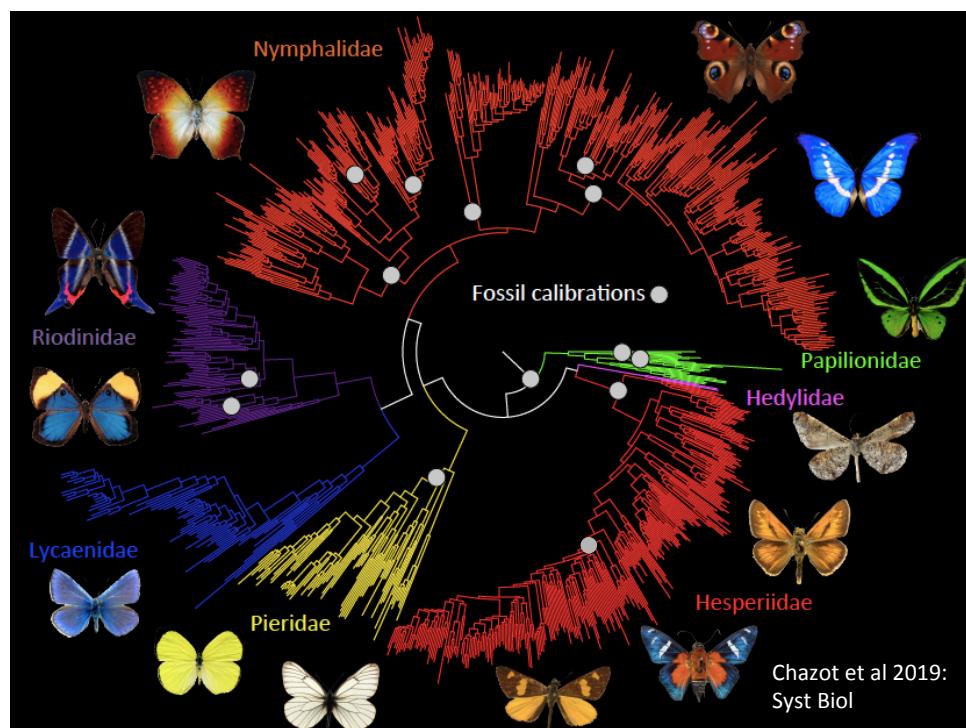
50

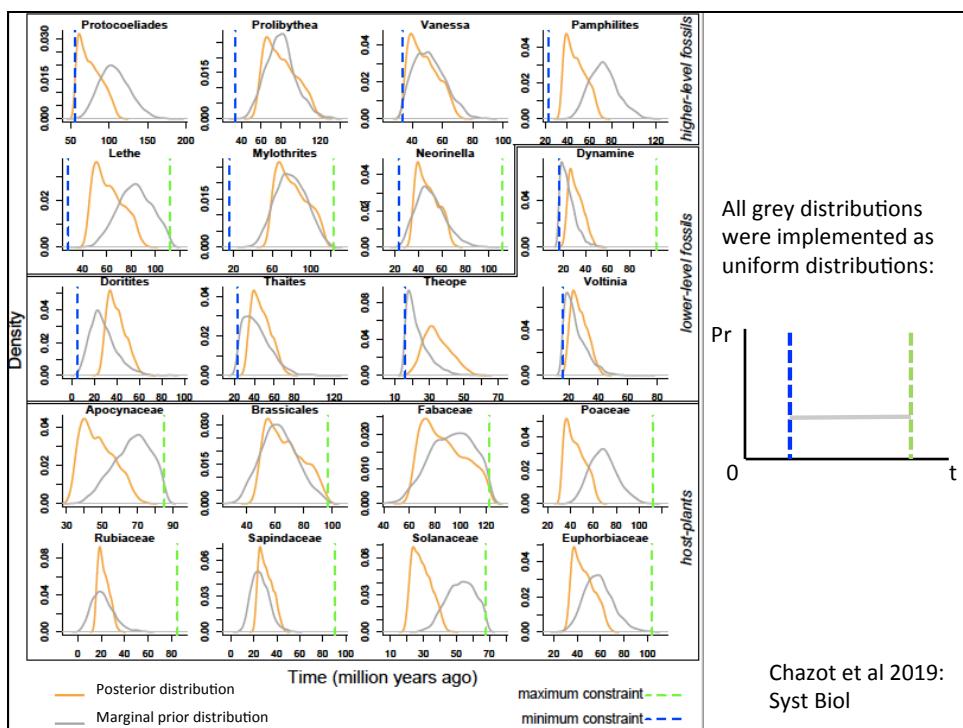
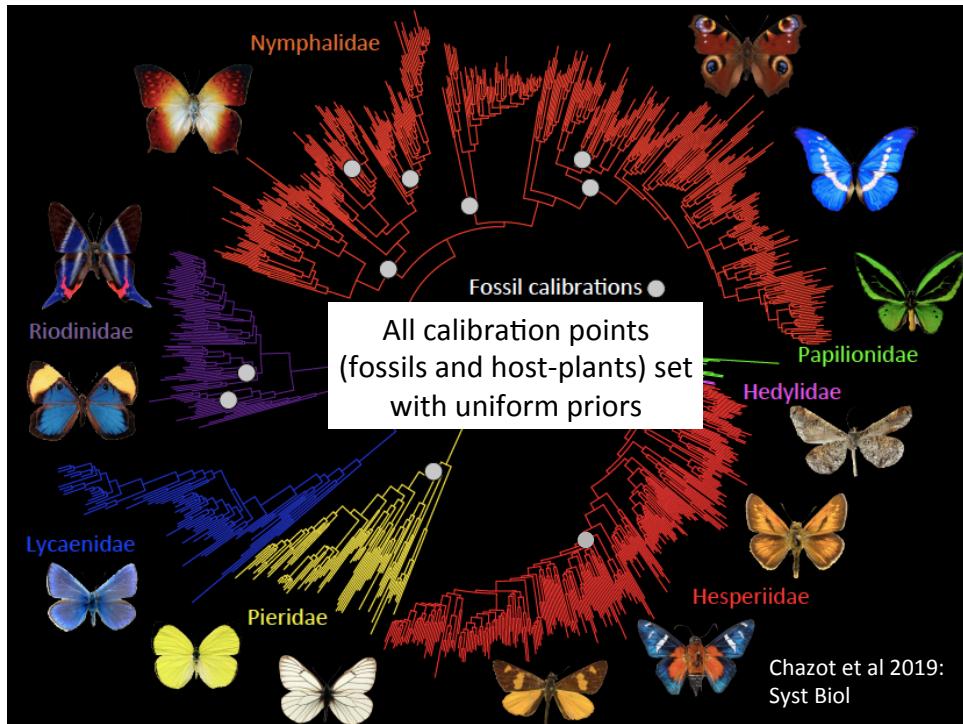
Multiple calibrations

- ▶ Be careful: priors interact with each others
- ▶ For example, node orders
- ▶ Marginal priors resulting from prior interactions can differ from the initial user prior
 - This can be visualized by removing the data and running the same analysis

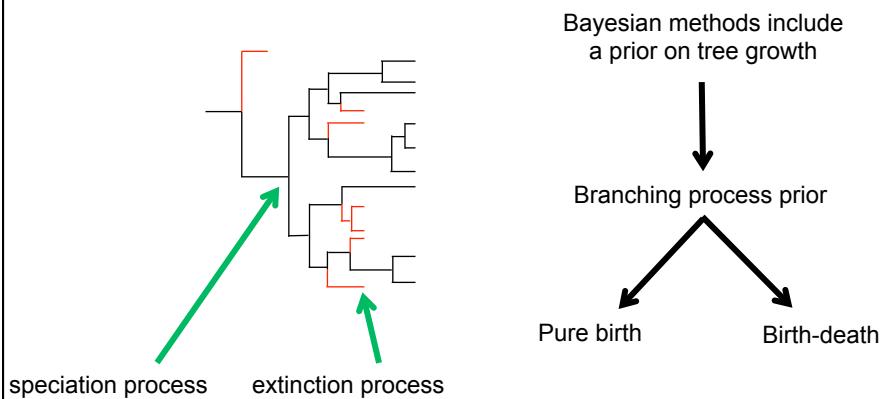
$$f(\theta|D) = \frac{f(\theta)}{\int f(D|\theta)f(\theta)d\theta}$$

51

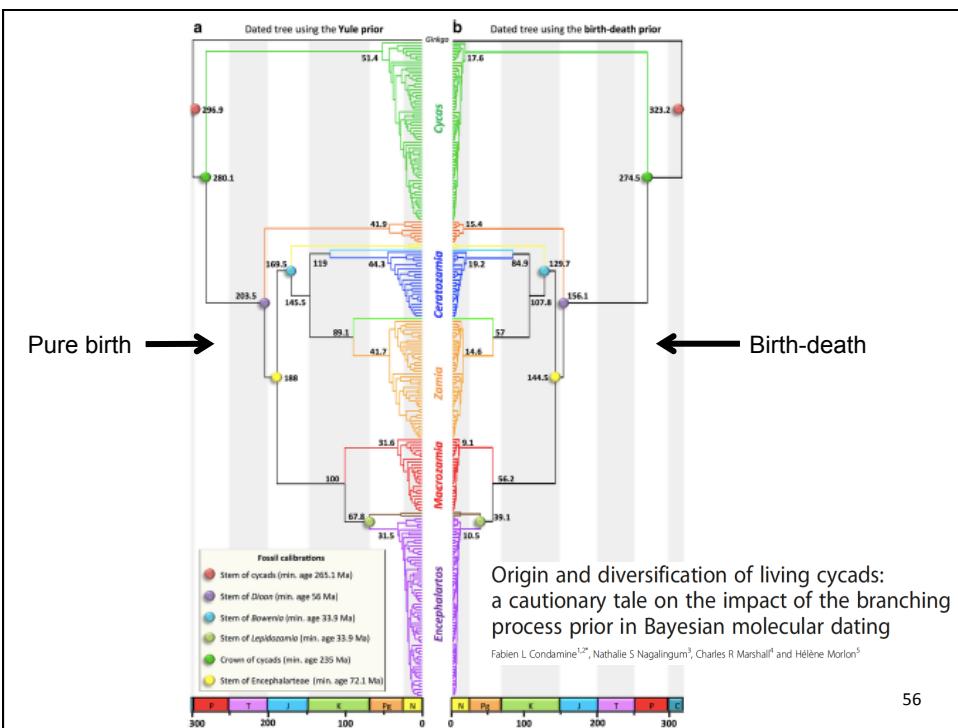


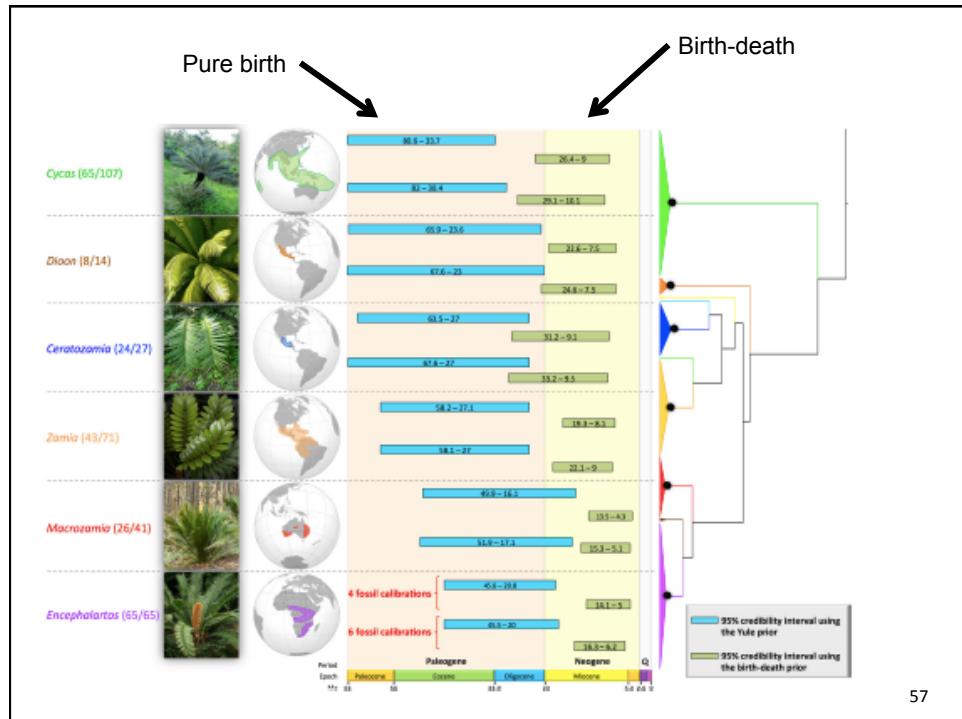


Prior sensitivity



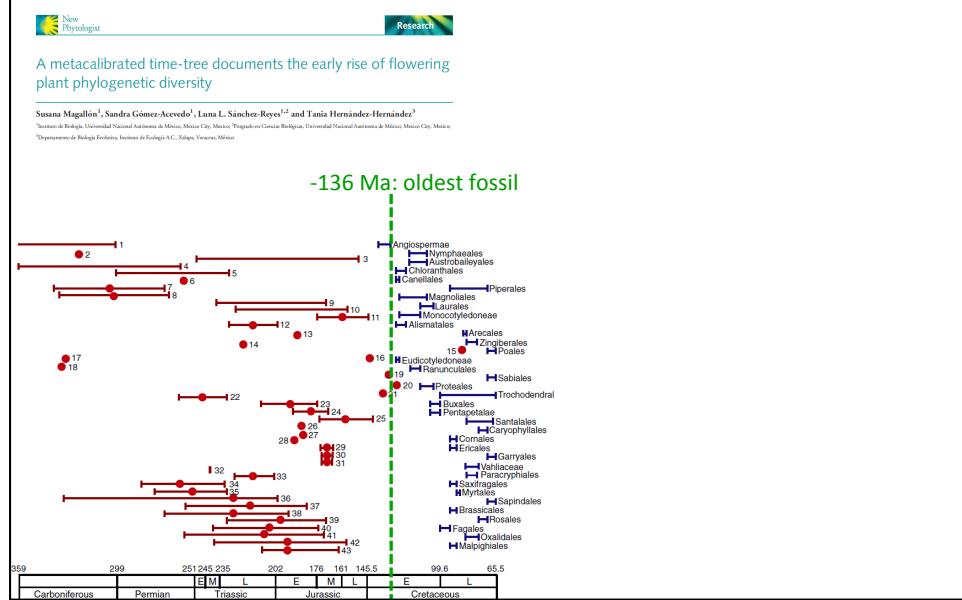
55



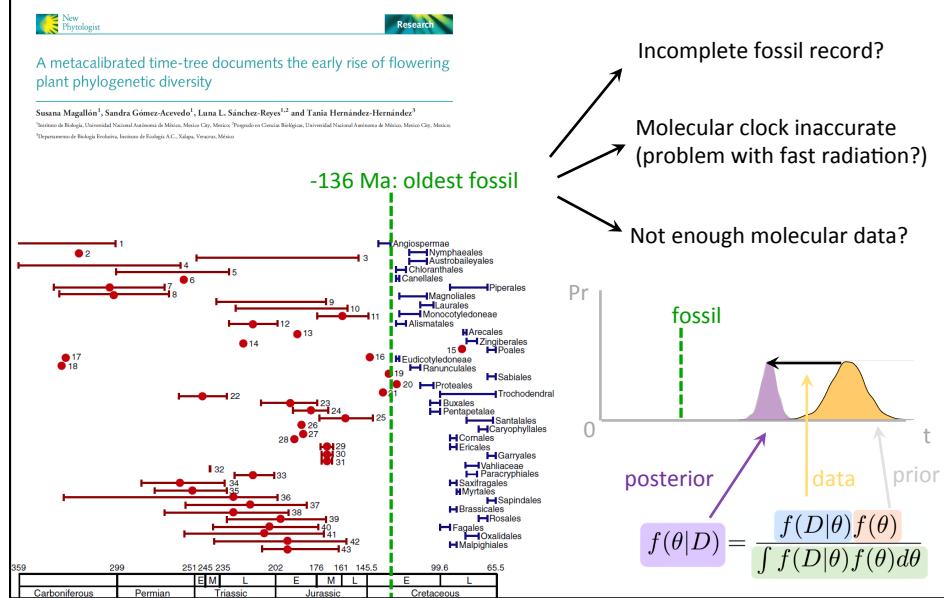


57

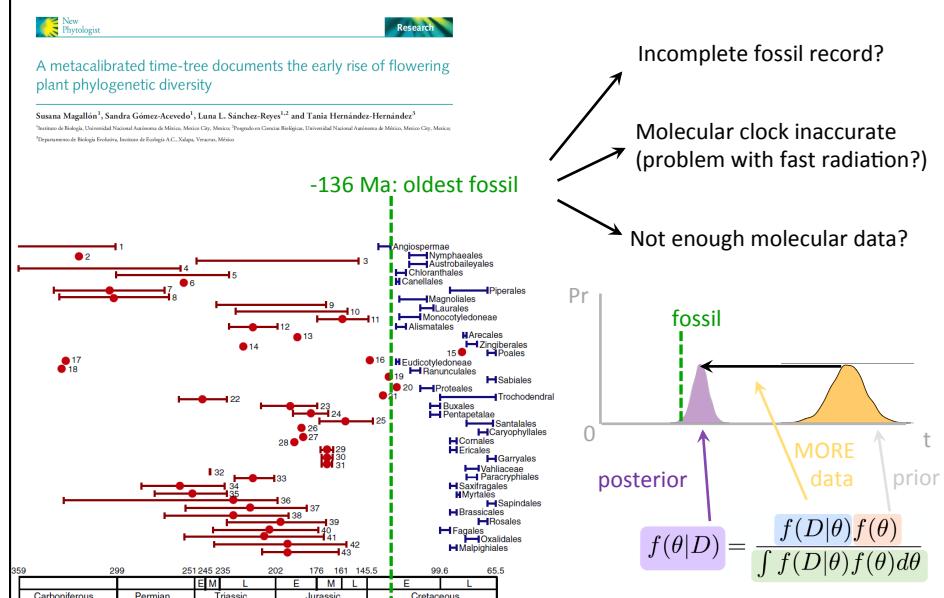
Timing the origin of Angiosperms



Timing the origin of Angiosperms



Timing the origin of Angiosperms



Timing the origin of Angiosperms

Eocene lantern fruits from Gondwanan
Patagonia and the early origins of Solanaceae

Peter Wilf^{1,*}, Mónica R. Carvalho², María A. Gandolfo², N. Rubén Cúneo³

* See all authors and affiliations

Science 06 Jun 2017;
Vol. 355, Issue 6320, pp. 71-75
DOI: 10.1126/science.aag2737



Peer Reviewed
+ see details

Physalis infinemundi
Physalis
tomatillo group - 9 to 11 My
Nightshades - 35 to 51 My



Timing the origin of Angiosperms

Eocene lantern fruits from Gondwanan
Patagonia and the early origins of Solanaceae

Peter Wilf^{1,*}, Mónica R. Carvalho², María A. Gandolfo², N. Rubén Cúneo³

* See all authors and affiliations

Science 06 Jan 2017;
Vol. 355, Issue 6320, pp. 71-75
DOI: 10.1126/science.aag2737



Peer Reviewed
+ see details

Physalis infinemundi
Physalis
tomatillo group - 9 to 11 My
Nightshades - 35 to 51 My



Timing the origin of Angiosperms

Eocene lantern fruits from Gondwanan
Patagonia and the early origins of Solanaceae

Peter Wilf^{1,*}, Mónica R. Carvalho², María A. Gandolfo², N. Rubén Cúneo³

* See all authors and affiliations

Science 06 Jun 2017;
355, issue 6292, pp. 71-75
DOI: 10.1126/science.aag2737

