

**Bayesian Inference for Phenotypic and Palaeo Data**

Michael Lee  
South Australian Museum and Flinders University

**Bayesian Inference for Phenotypic and Palaeo Data**

- Bayesian phylogenetics (especially BEAST) has a lot to offer for morphologists and palaeontologists
- Peculiarities of morphological / palaeontological data
- Some case studies
  - Cambrian trilobites (epoch clocks)
  - Crocodile evolution (detecting convergent evolution)
  - Lizard viviparity (ancestral state reconstruction)

**Bayesian Inference for Phenotypic and Palaeo Data**

- Bayesian phylogenetics (especially BEAST) has a lot to offer for morphologists and palaeontologists
- Peculiarities of morphological / palaeontological data
- Some case studies
  - Cambrian trilobites (epoch clocks)
  - Crocodile evolution (detecting convergent evolution)
  - Lizard viviparity (ancestral state reconstruction)

**Bayesian Inference for Phenotypic and Palaeo Data**

- Ph.D 1995 – Evolutionary Morphology of Pareiasaurs
- Not a programmer, just an end-user of BEAST et al.
- You don't have to be a coding geneious to get a lot out of BEAST

**Bayesian Inference for Phenotypic and Palaeo Data**

Michael Lee  
South Australian Museum and Flinders University

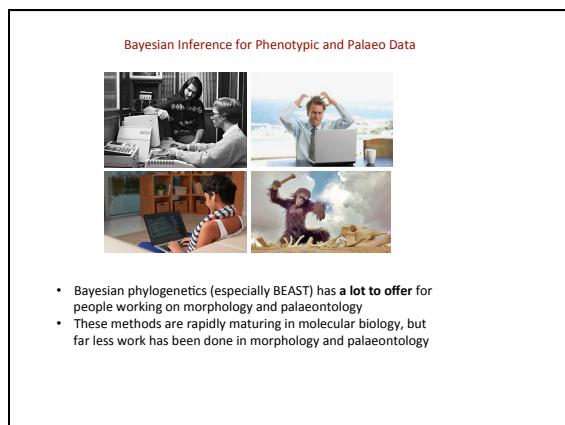
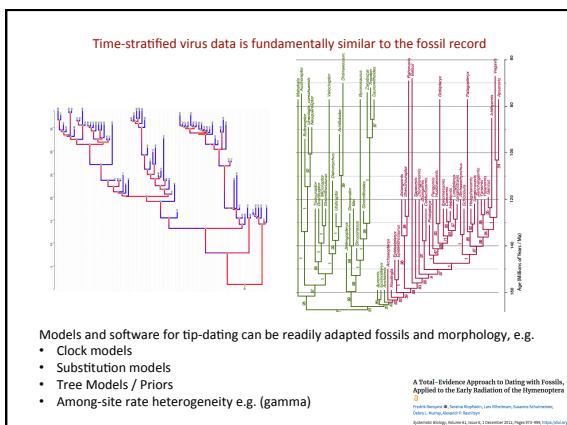
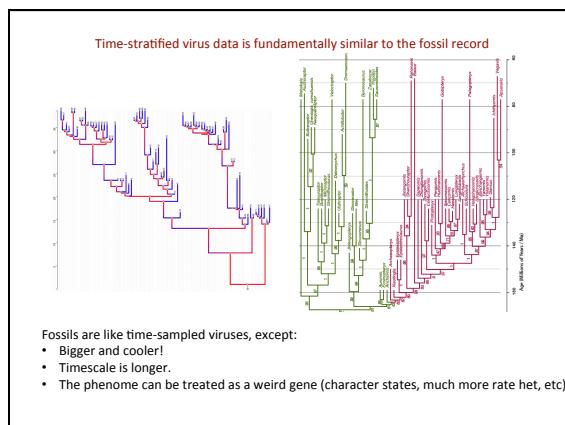
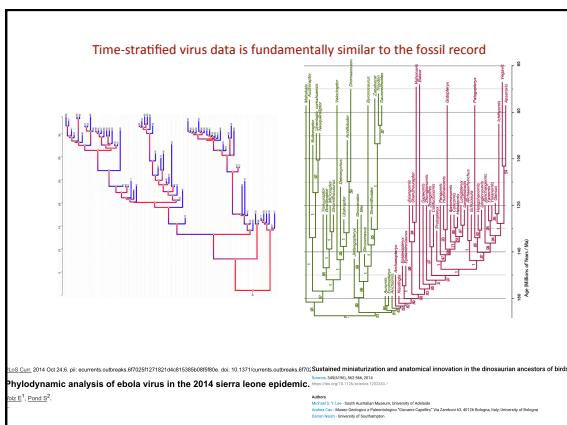
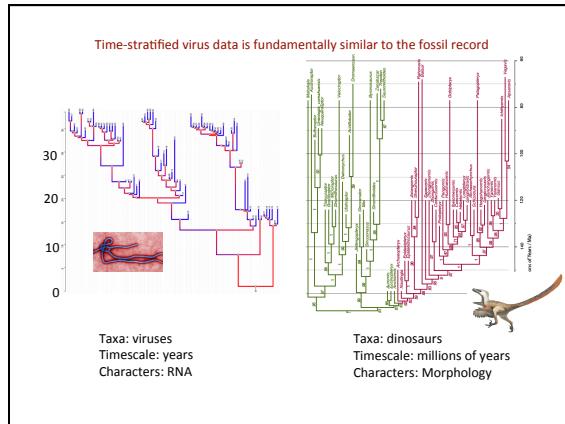
- Ph.D 1995 – Evolutionary Morphology of Pareiasaurs
- Not a programmer, just an end-user
- You don't have to be a coding Geneious to get a lot out of BEAST

**Bayesian Inference for Phenotypic and Palaeo Data**

Michael Lee  
South Australian Museum and Flinders University

- Ph.D 1995 – Evolutionary Morphology of Pareiasaurs
- Not a programmer, just an end-user
- You don't have to be a coding Geneious to get a lot out of BEAST

## Bayesian Inference for Phenotypic and Palaeo Data



## Bayesian Inference for Phenotypic and Palaeo Data

The image consists of four separate photographs arranged in a 2x2 grid. The top-left photo shows two people in a lab setting, one standing and one seated at a desk with a computer. The top-right photo shows a man sitting at a desk, looking stressed with his hands on his head, while looking at a laptop screen. The bottom-left photo shows a person from behind, sitting at a desk in what appears to be a library or study room, working on a laptop. The bottom-right photo shows a person sitting in a chair, with a computer monitor in front of them displaying a complex 3D visualization of a biological specimen.

- Bayesian phylogenetics (especially BEAST) has a **lot to offer** for people working on morphology and palaeontology
- These methods are rapidly maturing in molecular biology, but far less work has been done in morphology and palaeontology
- **Lot of opportunities to do interesting stuff!**

- Simply translating established methods in BEAST / Molec Biol onto morphological datasets can be worthwhile:

## Relaxed Phylogenetics and Dating with Confidence

Alexei J Drummond , Simon Y. W Ho, Matthew J Phillips, Andrew Rambaut  

Published: March 14, 2006 • <https://doi.org/10.1371/journal.pbio.0040088> 2006

Bayesian Inference for Phenotypic and Palaeo Data

- Simply translating established methods in BEAST / Molec Biol onto morphological datasets can be worthwhile:

2013

Geobiol. 2013, Vol 11(3), 198-206, doi:10.1089/geob.2013.0746, Back 2013 Sep 12  
Rates of phenotypic and genomic evolution during the Cambrian explosion.

- Simply translating established methods in BEAST / Molec Biol onto morphological datasets can be worthwhile:

*Syst Biol.* 2014 Jul; 63(4): 493–504.  
Published online 2014 Mar 12. doi: [10.1093/sysbio/syu015](https://doi.org/10.1093/sysbio/syu015)

PMC ID: PMC4055869  
 PMID: 24627184

### Inferring Heterogeneous Evolutionary Processes Through Time: from Sequence Substitution to Phylogeography

Philip Bliege,<sup>1</sup> Philippe Lemey,<sup>1</sup> Guy Baile,<sup>1</sup> Andrew Rambaut,<sup>2,3</sup> and Marc A. Suchard<sup>4,5</sup>

The diagram shows four nucleotides arranged in a square: A (top-left), G (top-right), C (bottom-left), and T (bottom-right). Four arrows connect them in a cycle: a red arrow from A to G labeled 'b' (Transitions), a blue arrow from G to C labeled 'd' (Transversions), a blue arrow from C to T labeled 'e' (Transitions), and a red arrow from T back to A labeled 'f' (Transversions).

Transitions

Transversions

Transitions

2014

Bayesian Inference for Phenotypic and Palaeo Data

- Simply translating established methods in BEAST / Molec Biol onto morphological datasets can be worthwhile:

Fig 3. Dynamic Biogeographic Model

Dynamic Biogeographic models and dinosaur origin  
Morten L. Jørgensen, Daniel S. Bolnick, and Michael J. Sander  
bioRxiv preprint doi: [https://doi.org/10.1101/277000](https://doi.org/10.1101/2018.12.04.277000); this version posted December 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

- Bayesian phylogenetics (especially BEAST) has a lot to offer for morphologists and palaeontologists
- Peculiarities of morphological / palaeontological data
  - Some case studies
    - Cambrian trilobites (epoch clocks)
    - Crocodile evolution (detecting convergent evolution)
    - Lizard viviparity (ancestral state reconstruction)

Peculiarities of Morphological Data					
	Genetic characters Sequence of portion of chloroplast DNA				
Japanese black pine ( <i>Pinus thunbergii</i> )					
Bhutan white pine ( <i>Pinus bhutanica</i> )					
Chiapas pine ( <i>Pinus chiapensis</i> )					
Eastern white pine ( <i>Pinus strobus</i> )					
Lacebark pine ( <i>Pinus bungeana</i> )					
Red pine ( <i>Pinus resinosa</i> )					
Single leaf pine ( <i>Pinus monophylla</i> )					
	Morphological characters				
	number of vascular bundles per needle	sheath around needle bundle	number of (1=straight, 2=bent back)	seed wing (0=absent, 1=detachable, 2=permanent)	
Japanese black pine ( <i>Pinus thunbergii</i> )	2	1	2	2	
Bhutan white pine ( <i>Pinus bhutanica</i> )	1	2	5	1	
Chiapas pine ( <i>Pinus chiapensis</i> )	1	2	5	1	
Eastern white pine ( <i>Pinus strobus</i> )	1	2	5	1	
Lacebark pine ( <i>Pinus bungeana</i> )	1	2	3	2	
Red pine ( <i>Pinus resinosa</i> )	2	1	2	2	
Single leaf pine ( <i>Pinus monophylla</i> )	1	2	1	0	
Up to millions of characters (phylogenomics)					
Typically <100s of characters (TOL mammals ~4500 chars)					

- Fewer characters
- More homoplasy

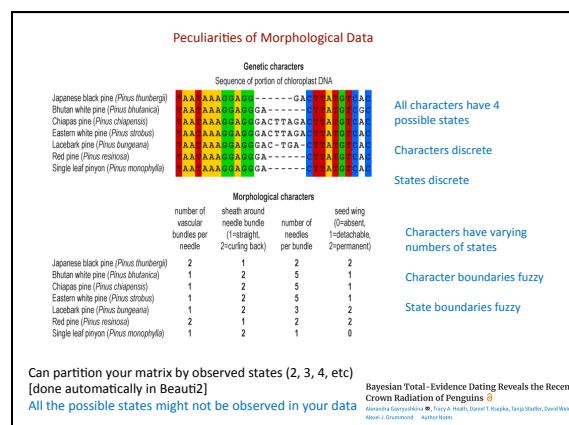
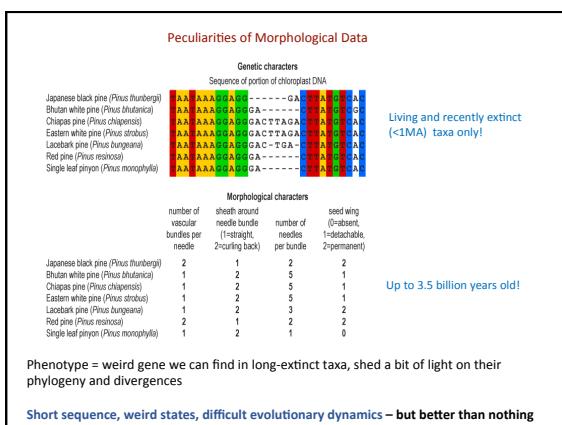
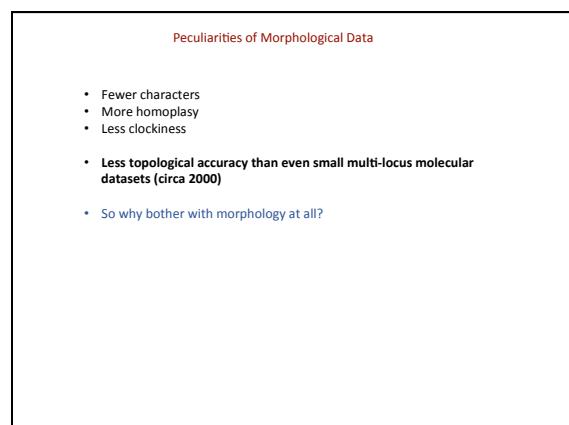
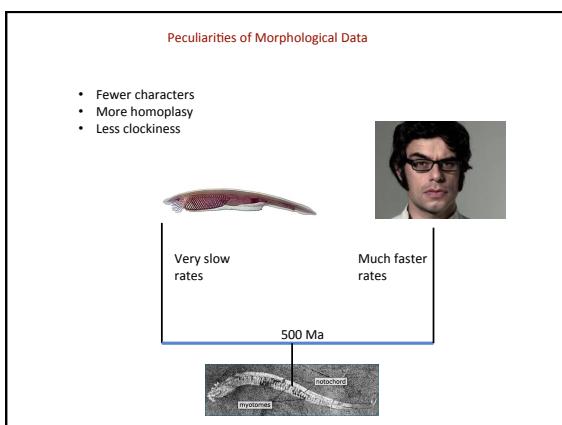
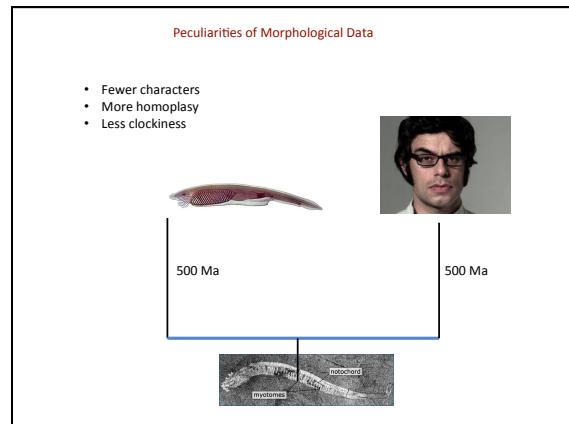
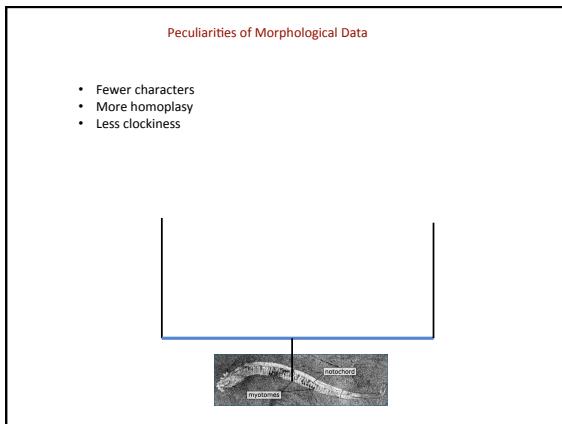
Completely disparate taxa can be close relatives

Peculiarities of Morphological Data

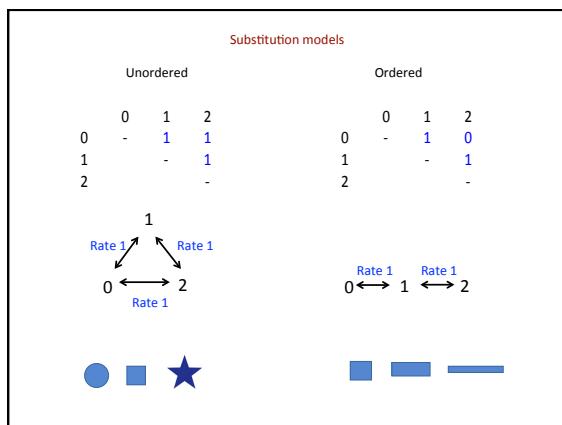
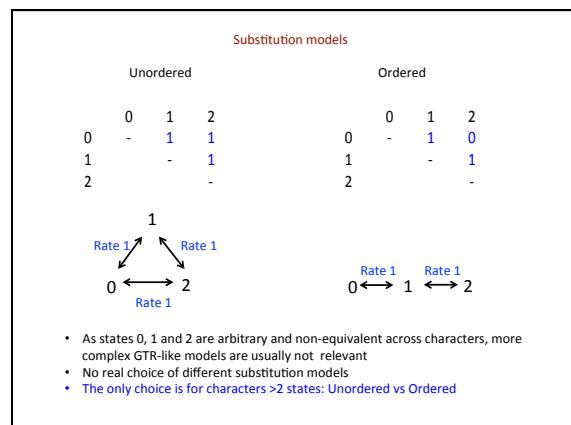
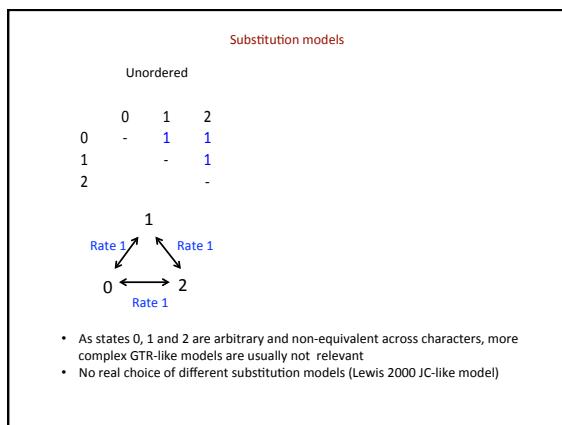
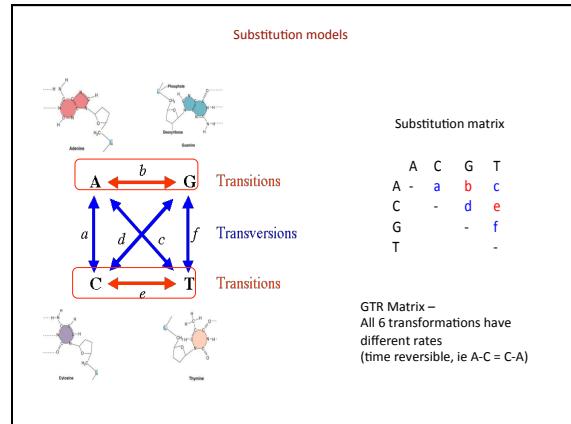
- Fewer characters
- More homoplasy

Very similar taxa can be distantly related, e.g. legless reptiles  
(modelling co-variation is challenging but worth doing)

## Peculiarities of Morphological Data



Peculiarities of Morphological Data						
Genetic characters						
Sequence of portion of chloroplast DNA						
Japanese black pine ( <i>Pinus thunbergii</i> )						
Bhutan white pine ( <i>Pinus bhutanica</i> )						
Chiper pine ( <i>Pinus chiaepensis</i> )						
Eastern white pine ( <i>Pinus strobus</i> )						
Lacebark pine ( <i>Pinus bungeana</i> )						
Red pine ( <i>Pinus resinosa</i> )						
Single leaf pine ( <i>Pinus monophylla</i> )						
Character states (ATGC) are equivalent across characters						
Morphological characters						
number of vascular bundles per needle	sheath around needle bundle		number of (1=straight, 2=curling back) needles per bundle		seed wing (0=absent, 1=detachable, 2=permanent)	
Japanese black pine ( <i>Pinus thunbergii</i> )	2	1	2	2	0	
Bhutan white pine ( <i>Pinus bhutanica</i> )	1	2	5	1		
Chiper pine ( <i>Pinus chiaepensis</i> )	1	2	5	1		
Eastern white pine ( <i>Pinus strobus</i> )	1	2	5	1		
Lacebark pine ( <i>Pinus bungeana</i> )	1	2	3	2		
Red pine ( <i>Pinus resinosa</i> )	2	1	2	2		
Single leaf pine ( <i>Pinus monophylla</i> )	1	2	1	0		
Limits substitution models you can use to JC-equivalent (Lewis 2000 model)						
Base frequency models are also problematic						



If the substitution matrix applies to a SINGLE character

- States DO mean something
- You can tailor the model in any way you want!

To

	Arid	Forest	Savannah	Temperate	Subtropical
From	.07	<b>1.68</b>	.66	1.27	
Arid	.34	.54	.18	.40	
Forest	.60	<b>1.97</b>	-.44	<b>1.79</b>	
Savannah	<b>2.41</b>	<b>1.80</b>	.29	-.48	
Temperate	.59	.56	1.46	<b>2.50</b>	
Subtropical					

- Non-sampling of invariant characters can be corrected easily (Beauti2)
- Non-sampling of parsimony-uninformative characters is harder to correct, due to variation in character-state space
- This can compromise clock analyses, by reducing the amount of change (and thus estimated duration) of terminal branches

Including autapomorphies is important for palaeontological tip-dating with clock-like data, but not with min-check data.

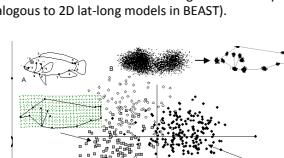
Bayesian Analyses Using a Single Likelihood Model Outperform Maximum Likelihood Phylogeny from Discrete Morphological Data

### Ascertainment bias

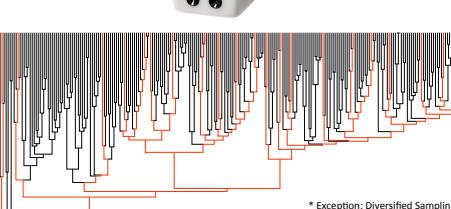
- Most morphological datasets are not very suitable for tip-dating\*
- If you are compiling new datasets, score **all** observed variation (whether parsimony-informative or not). Future-proof your matrices!
- Analogy: imagine the problems we'd have if all our molecular datasets were like SNP datasets (but with parsimony-uninformative SNPs removed!)

\*Including many we've analysed!

- BEAST can also take continuous phenotypic data (examples later)
- So you can directly use these variables as *measured*, rather than recoding them into arbitrary categories (loss of information).
- E.g. use **actual body weight**, instead of 'small', 'medium' or 'large'
- Potential in future to use 2D and 3D geometric morphometric data (analogous to 2D lat-long models in BEAST).



- Most tree priors assume taxa are evenly and randomly sampled across the entire tree\*

\* Exception: Diversified Sampling  
Total-Evidence Dating under the Fossilized Birth-Death Process [B](#)

Chi Zhang, Torsten Stadler, Serafinová Klopštejn, Tracy A. Heath, Hendrik Nautsch <https://doi.org/10.11919/jig.2024.100000>

**Taxonomic sampling**

- Most tree priors assume taxa are evenly and randomly sampled across the entire tree
- Few morphological (or molecular) datasets do this!
  - Most sample poorly from the outgroup(s)
  - Many use 'diversified' sampling within ingroup, where exemplars are chosen from each major group.

**Taxonomic sampling**

- Most tree priors assume taxa are evenly and randomly sampled across the entire tree
- Few morphological (or molecular) datasets do this!
  - Most sample poorly from the outgroup(s)
  - Many use 'diversified' sampling within ingroup, where exemplars are chosen from each major group.

Speciation events underestimated in outgroups and near-present

**Taxonomic sampling**

- Might need to add taxa (e.g. to outgroups) to make sampling more appropriate
- Can delete outgroup, but then rooting in the ingroup can be uncertain (clock rooting is not very reliable).

**Taxonomic sampling**

- Might need to add taxa (e.g. to outgroups) to make sampling more appropriate
- Can delete outgroup, but then rooting in the ingroup can be uncertain (clock rooting is not very reliable).
- Compromise "hack"
  - Run undated analysis (parsimony, MrBayes) with outgroup to work out root in ingroup.
  - Run Bayesian analysis enforcing this root position, but deleting the outgroup.

**Bayesian Inference for Phenotypic and Palaeo Data**

- Bayesian phylogenetics (especially BEAST) has a lot to offer for morphologists and palaeontologists
- Peculiarities of morphological / palaeontological data
- Some case studies
  - Cambrrian trilobites (epoch clocks)
  - Crocodile evolution (detecting convergent evolution)
  - Lizard viviparity (ancestral state reconstruction)