## Bayesian Inference for Phenotypic and Palaeo Data

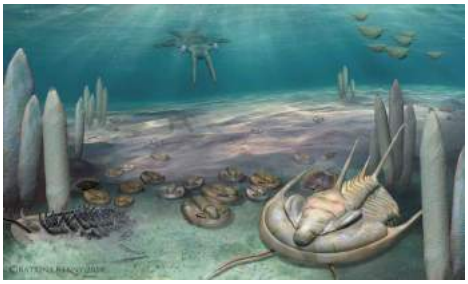
- Bayesian phylogenetics (especially BEAST) has a lot to offer for morphologists and palaeontologists
- Peculiarities of morphological / palaeontological data
- Some case studies
  - Cambrian trilobites (evolutionary rates)
  - Crocodile evolution (detecting convergent evolution)
  - Lizard viviparity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

## 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 (evolutionary rates)
  - Crocodile evolution (detecting convergent evolution)
  - Lizard viviparity (ancestral state reconstruction)

## Trilobites and the Cambrian explosion

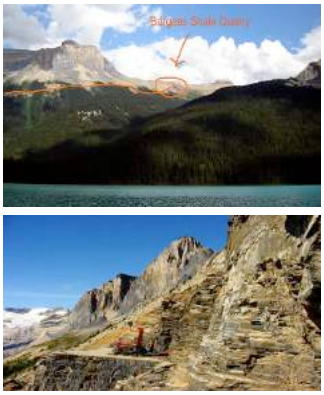
## The Cambrian Explosion



## Cambrian Explosion

| Late Precambrian ~560mya | Cambrian onwards ~540 mya |
|---|---|
| Low diversity | High diversity (nearly all fossilizable phyla) |
| Simple organisms, no complex senses | Complex organisms (eyes, appendages etc) |
| No predation (simple food webs) | Predator-prey interactions (complex food webs) |
| 2d ecosystems: No nekton (active swimmers) | 3D ecosystems: Nekton (active swimmers) |
| No infauna (burrowers) | Infauna (burrowers) |

Burgess shale (Canada)



Burgess shale (Canada)

Emu Bay shale (Australia)

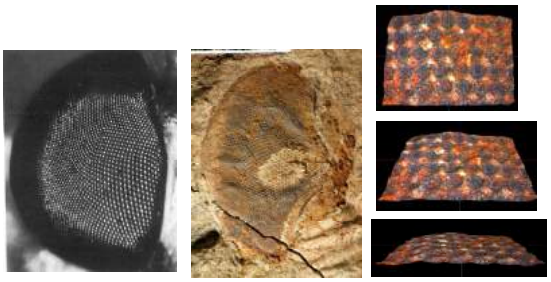
Trilobites and the Cambrian explosion

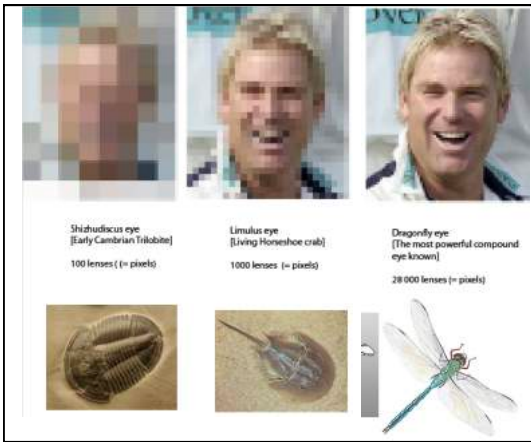
Gould's oddballs are not so odd. EBS includes:

*Hallucigenia* is an onychophoran (velvet worm)

*Anomalocaris* is a stem-arthropod

Fossil eyes of *Anomalocaris* and other taxa






Trilobites and the Cambrian explosion

Causes ?

- Predators
- Vision
- Infauna (burrowing)
- Genetic (hox genes)

← These are potentially interlinked. Explain increase in evolutionary rate (but not the timing)

Trilobites and the Cambrian explosion

Causes ?

- Predators
- Vision
- Infauna (burrowing)
- Genetic (hox genes)

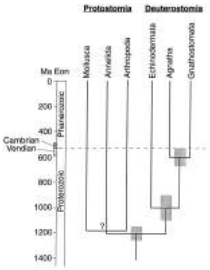
← These are potentially interlinked explain increase in evolutionary rate (but not why timing)

Interesting new idea – which explains Timing:

- Oyxgen levels increased around 540Ma
- Predatory invertebrates have higher metabolisms (and oxygen demands) than herbivores/scavengers
- Hence an increase in global oxygen allowed all the above to occur

## Timing – when did the Cambrian explosion begin ?

Molecular divergences suggests earlier diversification, well within the Precambrian



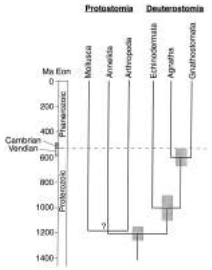

- (In)Famous 1996 paper dating the radiation of animals
- To twice the age of the Cambrian explosion!
- Used a strict clock and extrapolated slow vertebrate rates across the rest of the invertebrate tree!

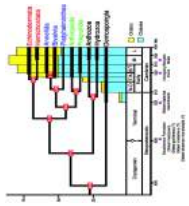
- Great demonstration of the need for relaxed clocks

Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla

## Timing – when did the Cambrian explosion begin ?

Molecular divergences suggests earlier diversification, well within the Precambrian



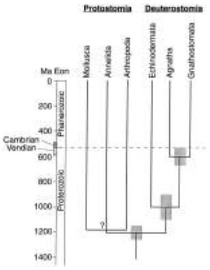
Relaxed clock (r8s) analyses halves the age!

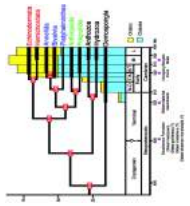
Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla

Estimating metazoan divergence times with a molecular clock

## Timing – when did the Cambrian explosion begin ?

Molecular divergences suggests earlier diversification, well within the Precambrian
If rates vary over time, then molecular clock analyses are very challenging



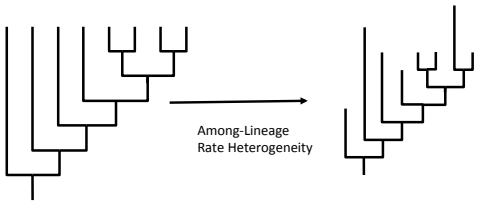



Relaxed clock (r8s) analyses halves the age!

Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla

Estimating metazoan divergence times with a molecular clock

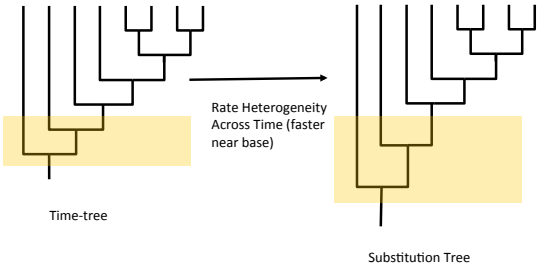
## Timing – when did the Cambrian explosion begin ?




Time-tree

Substitution Tree

## Timing – when did the Cambrian explosion begin ?

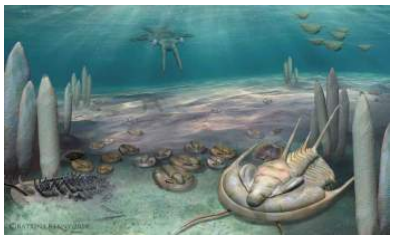



Time-tree

Substitution Tree

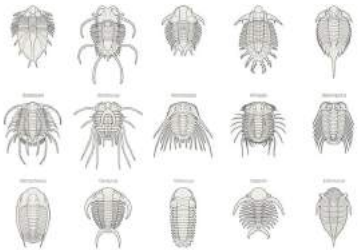
## Timing – when did the Cambrian explosion END?

- When did all the (marine) niches on our planet start to get saturated, with evolution slowing down to Phanerozoic norms ?
- No-one has ever attempted a precise answer to that question



## Trilobites – the best group for Cambrian macrovolutionary analyses

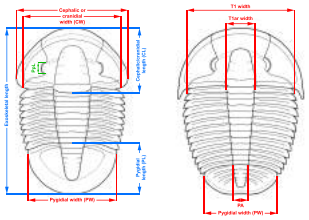
- Extremely abundant & diverse (~12% of species in the EBS); the beetles of the Cambrian
- Lots of hard body parts = lots of characters
- They appear as soon as the Cambrian fossil record starts to get good.



Trilobite evolutionary rates constrain the duration of the Cambrian explosion

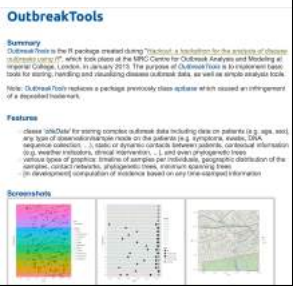
## Trilobites – the best group for Cambrian macrovolutionary analyses

- Taxon sampling: >100 species from across most of the Cambrian
- Character sampling:
  - discrete traits
  - continuous traits
  - meristic traits



## Trilobites – the best group for Cambrian macrovolutionary analyses

- Taxon sampling: >100 species from across most of the Cambrian
- Character sampling:
  - discrete traits
  - continuous traits
  - meristic traits
- Tip-dated analyses of combined data
  - BEAST and BEAST2
  - Outbreak Tools



OutbreakTools

Summary

Features

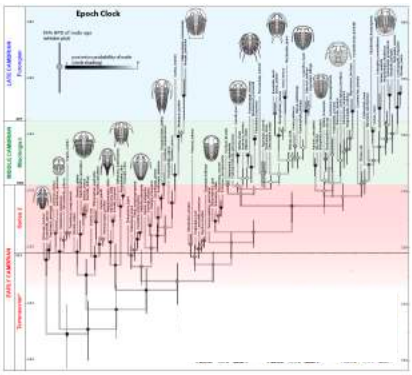
Screenshots

## Trilobites – the best group for Cambrian macrovolutionary analyses

- Taxon sampling: >100 species from across most of the Cambrian
- Character sampling:
  - discrete traits
  - continuous traits
  - meristic traits
- Tip-dated analyses of combined data
  - BEAST and BEAST2
  - Outbreak Tools
- Clock models
  - Strict clock
  - UCLN relaxed clock
  - **Epoch clock**

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

## Trilobites – the best group for Cambrian macrovolutionary analyses

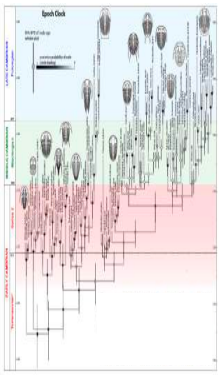



Epoch 3
Late Cambrian)

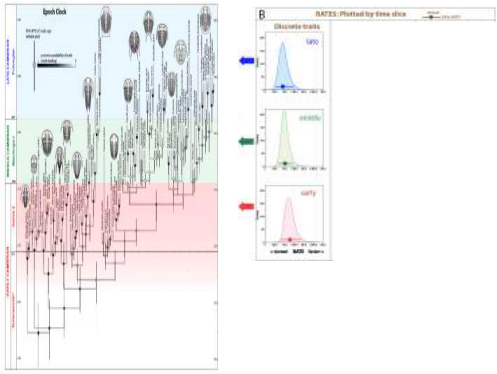
Epoch 2
(Mid Cambrian)

Epoch 1
(Early Cambrian)

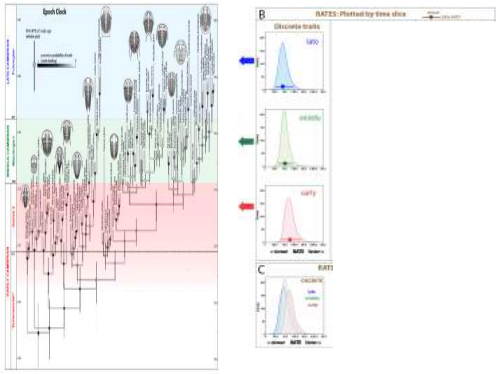
## Trilobites – the best group for Cambrian macrovolutionary analyses



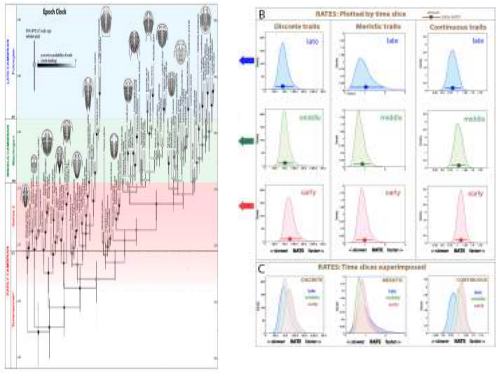
Trilobites – the best group for Cambrian macrovolutionary analyses



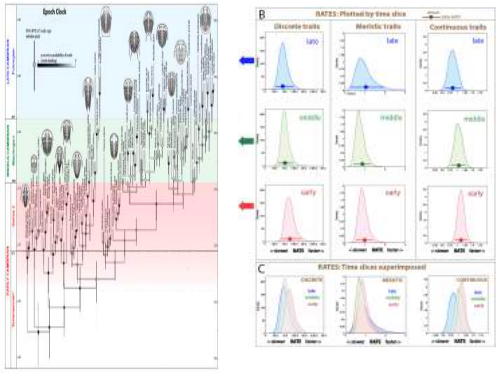
Trilobites – the best group for Cambrian macrovolutionary analyses



Trilobites – the best group for Cambrian macrovolutionary analyses



Trilobites – the best group for Cambrian macrovolutionary analyses



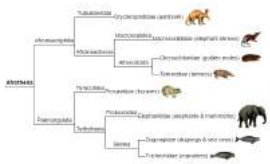
Conclusions

- The Cambrian explosion was already over by the time the first trilobites appear 520 M
- By the time the 'typical' Cambrian fossil record commences, it was evolutionary 'business as usual'
- The Cambrian explosion was briefer, and ended more abruptly, than previously believed
- The Burgess Shale, EBS and other famous Cambrian fossil sites are not windows into the greatest adaptive radiation of all time (*contra* Gould et al).

Trilobite evolutionary rates constrain the duration of the Cambrian explosion

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 (evolutionary rates)
  - Crocodile evolution (detecting convergent evolution)
  - Lizard viviparity (ancestral state reconstruction)



## Crocodile Evolution

## Crocodylians: Diversity and Anatomy

Diversity: ~24 living species

- ~14 species of crocodiles (lower canines exposed)

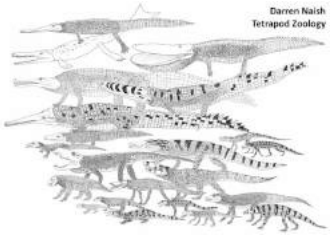
- ~ 8 species of alligators including caimans (lower canines hidden)

- 2 species of gavials/gharials (extremely narrow snouts for fish-eating)

## Crocodylians: Diversity and Anatomy

Diversity:

- All living crocs are sluggish, amphibious predators
- But fossil crocs were much more diverse
  - Active terrestrial forms (cursorial)
  - Small tree-climbing forms (arboreal)
  - Ocean-going forms



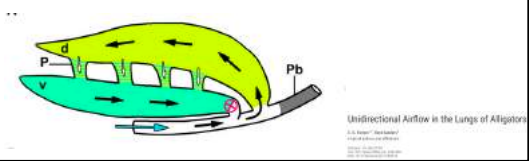

## Crocodylians: Diversity and Anatomy

Diversity:

- All living crocs are sluggish, amphibious predators
- But fossil crocs were much more diverse
  - Active terrestrial forms (cursorial)
  - Small tree-climbing forms (arboreal)
  - Ocean-going forms

Crocodylians were probably much more active in the past, and living forms are secondarily sluggish.

- Living crocs still retain many active, bird-like adaptations (fully-divided hearts and unidirectional lungs).




## Crocodylians: Diversity and Anatomy

Diversity:

- All living crocs are sluggish, amphibious predators
- But fossil crocs were much more diverse
  - Active terrestrial forms (cursorial)
  - Small tree-climbing forms (arboreal)
  - Ocean-going forms

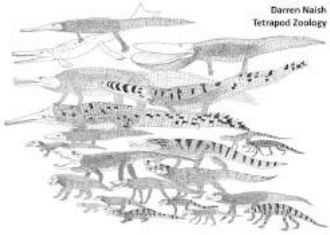
Crocodylians were probably much more active in the past, and living forms are secondarily sluggish.

- Living crocs still retain many active, bird-like adaptations (fully-divided hearts and unidirectional lungs).
- They can also gallop!
- Evidence of agile ancestors?

## Crocodylians: Diversity and Anatomy

Diversity:

- All living crocs are sluggish, amphibious predators
- But fossil crocs were much more diverse
  - Active terrestrial forms (cursorial)
  - Small tree-climbing forms (arboreal)
  - Ocean-going forms



Crocodylians: Long-snouted forms

Indian Gavial (*Gavialis*)

Malayan or "false" Gavial (*Tomistoma*)

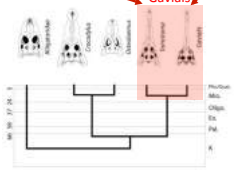
Crocodylians: Long-snouted forms

Malayan or "false" Gavial (*Tomistoma*)

Indian Gavial (*Gavialis*)

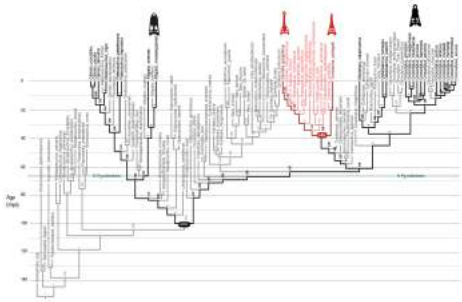
Gavials

Genomic tree:
Long-snouted forms
are a recent clade



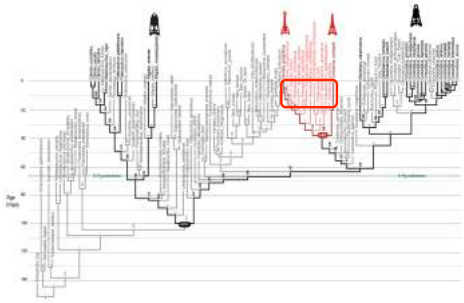
Crocodylians: Long-snouted forms

Most fossils fit the genomic scheme.... light = fossils



Crocodylians: Long-snouted forms

Most fossils fit the genomic scheme.... light = fossils



Crocodylians: Long-snouted forms

Most fossils fit the genomic scheme.... Except for thoracosaurs

Malayan or "false" Gavial (Tomistoma)

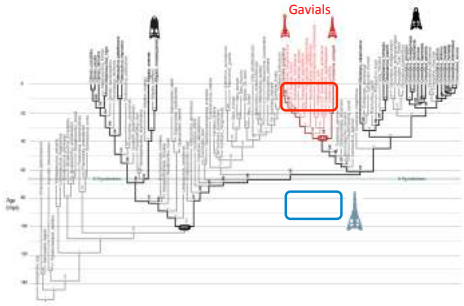
Indian Gavial (Gavialis)

Thoracosaur

Crocodylians: Long-snouted forms

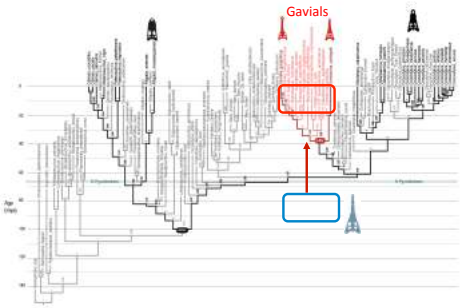
Most fossils fit the genomic scheme.... Except for thoracosaurs
Among the oldest crocs - three times as old as all other gavial fossils!

Gavials



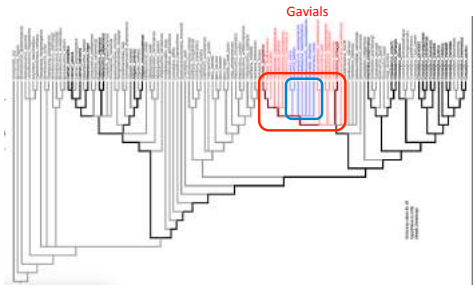
## Crocodylians: Long-snouted forms

Every undated (parsimony, MrBayes) analysis puts thoracosaurs within living gavials




## Crocodylians: Long-snouted forms

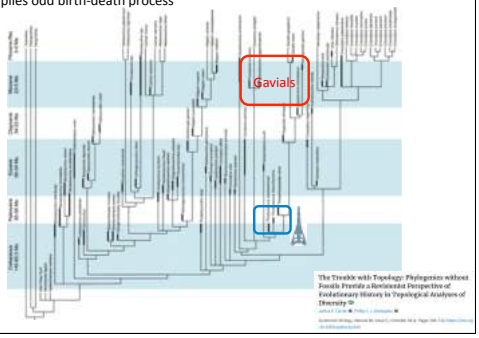
Every undated (parsimony, MrBayes) analysis puts thoracosaurs within living gavials




## Crocodylians: Long-snouted forms

Putting thoracosaurs within crown-clade gavials ruins the time-tree...

- 3x more ghost lineage (implying the croc fossil record is terrible)
- Poor stratigraphic concordance (oldest crocs are deeply nested!)
- Implies odd birth-death process

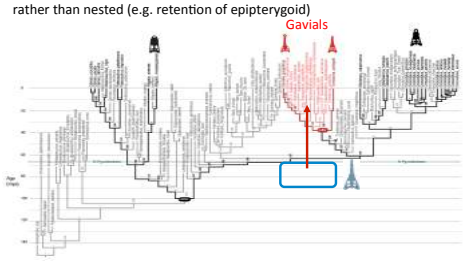



## Crocodylians: Long-snouted forms

Putting thoracosaurs within crown-clade gavials ruins the tree...

- 3x more ghost lineage (implying the croc fossil record is terrible)
- Poor stratigraphic concordance (oldest crocs are deeply nested!)
- Implies odd birth-death process

Notably, though, there are a few characters that do suggest thoracosaurs are basal rather than nested (e.g. retention of epipterygoid)



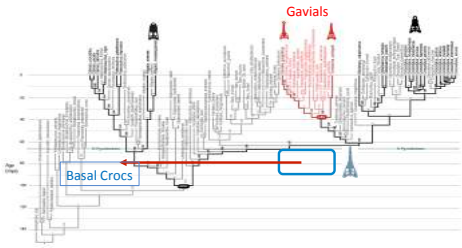

## Crocodylians: Long-snouted forms

Maybe thoracosaurs are basal crocs, which convergently evolved long snouts.

- Slightly less parsimonious (e.g. more convergence in snout characters)

But also

- Much more congruent with DNA
- Much more congruent with rest of stratigraphic record (all gavial fossils young)
- Consistent with some other morphological traits (epipterygoid)

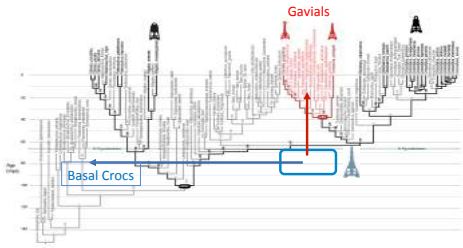



## Crocodylians: Long-snouted forms

2 hypotheses

Within gavials, favoured by parsimony

Basal, Favoured by everything else



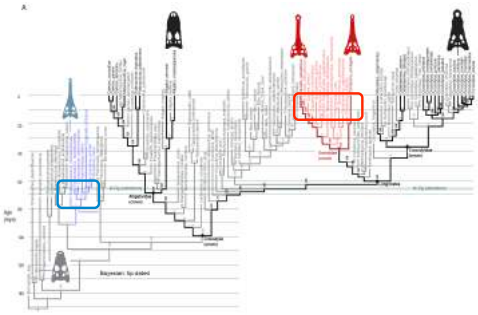
## BEAST to the rescue ?

Data
- 117 taxa including thoracosaurs
- 278 morphological traits
- 9200 bp

Total evidence tip-dating in BEAST / BEASTMC3
- Separate UCLN clocks
- Morph: Lewis, ascertainment, gamma, unpartitioned by state number

## BEAST to the rescue ?

Thoracosaurs are an early, basal lineage of crocs which convergently evolved gavial-like snouts (pp 0.99)



## BEAST to the rescue ?

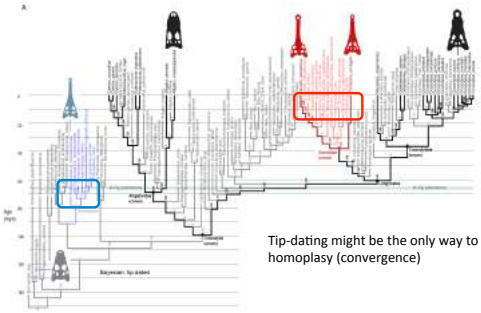
Thoracosaurs are an early, basal lineage of crocs which convergently evolved gavial-like snouts (pp 0.99)

Consistent with old age
Makes tree very congruent with stratigraphy
Reduces inferred ghost lineages
Explains why they retain some primitive bones

**But** slightly more homoplasy (i.e. more convergence)

## BEAST to the rescue ?

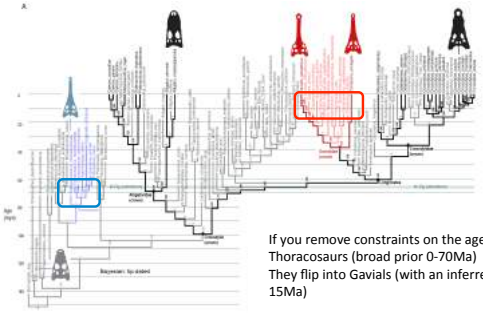
When similar organisms evolve multiple times at different ages ...



Tip-dating might be the only way to detect this homoplasy (convergence)

## BEAST to the rescue ?

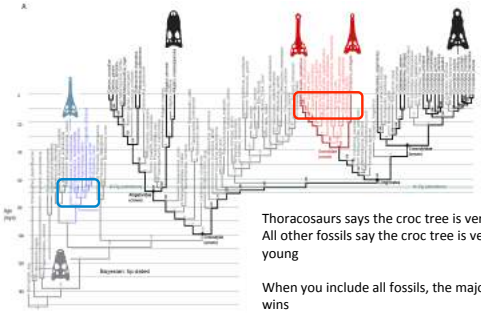
When similar organisms evolve multiple times at different ages ...



If you remove constraints on the age of Thoracosaurs (broad prior 0-70Ma)
They flip into Gavials (with an inferred age of 15Ma)

## BEAST to the rescue ?

Calibrations 'sort it out' among themselves in tip-dating ....



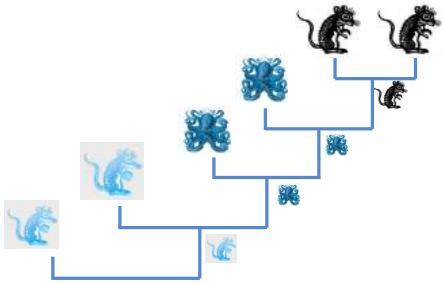
Thoracosaurs says the croc tree is very old
All other fossils say the croc tree is very young

When you include all fossils, the majority wins

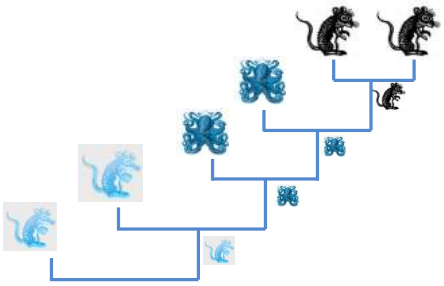
## Bayesian Inference and Phenotypic Data

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

## Ancestral State Reconstruction

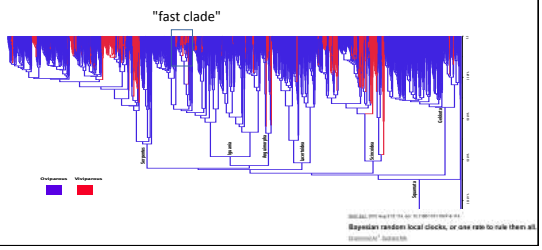


## Ancestral State Reconstruction



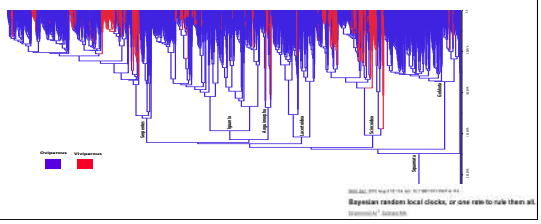
## Ancestral States: Why relaxed clocks matter!

- Biologists are tracing the evolution of traits on increasingly large trees (e.g. 10 000 birds, 6000 mammals, 4000 lizards and snakes).



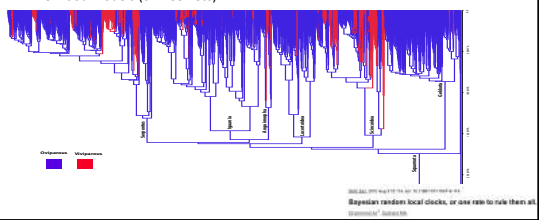

## Ancestral States: Why relaxed clocks matter!

- Biologists are tracing the evolution of traits on increasingly large trees (e.g. 10 000 birds, 6000 mammals, 4000 lizards and snakes).
- Traits are never going to evolve at the same rate across such large swathes of the tree of life …



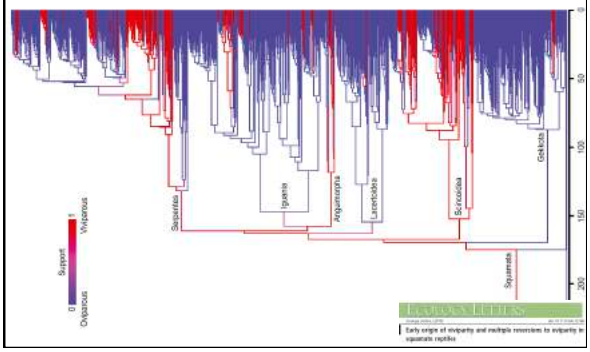
## Ancestral States: Why relaxed clocks matter!

- Biologists are tracing the evolution of traits on increasingly large trees (e.g. 10 000 birds, 6000 mammals, 4000 lizards and snakes).
- Traits are never going to evolve at the same rate across such large swathes of the tree of life …
- If you don't account for this (e.g. random local clocks) , you can get ***totally erroneous*** ancestral state reconstructions under Bayesian or Likelihood models (or BiSSE etc).



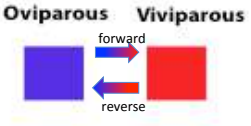
Ancestral state reconstruction (using strict clock) gave ludicrous results

- Ancestral lizard was live bearing (but all basal lineages are egg laying)
- Live-bearers reverted back to egg-laying over 100 times




The likelihood model used by most people..

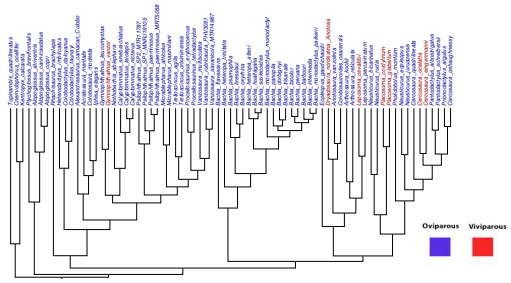
- Rate of evolution is homogenous throughout the tree (strict clock)
- But the asymmetry (forward-reverse bias) is estimated, ie a parameter that can be optimised and used to explain the observed data




- Likelihood's goofy solution to explain "fast evolving" clades:
  - Give one particular transition (e.g. reversals) a very high rate, so it should occur often
  - contrive a reconstruction that has lots of these transitions in fast clades

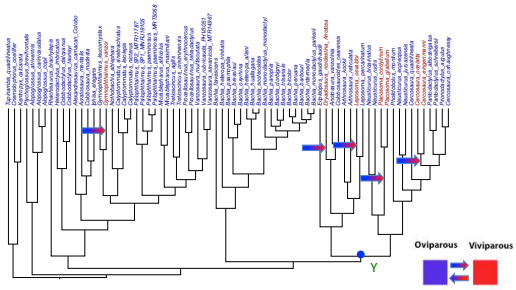
The Epic Fail explained …

- How would you map reproductive mode on this (hypothetical) tree?



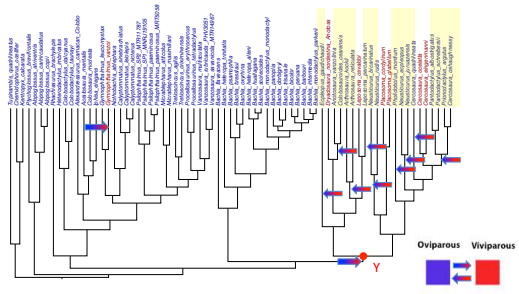

The Epic Fail explained …

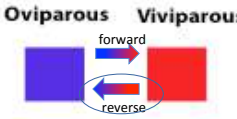
- Common-sense and parsimony says
  - lots of forward transformations (O-V)
  - Clade Y has a faster rate of change, ie 4 forward changes



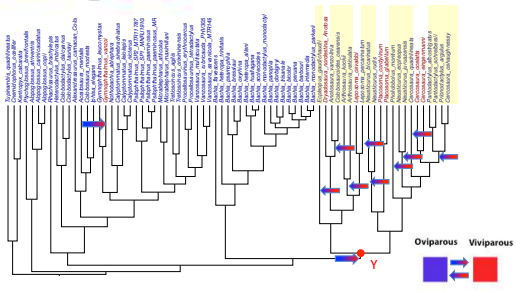

The Epic Fail explained …

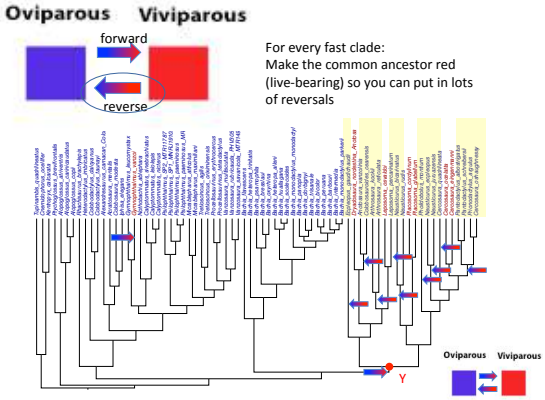
- Likelihood (rate-constant) reconstructs live-bearing at node Y on this tree, and then ~9 reversals back to egg-laying in lots of taxa (shaded)
- How does it converge on this ludicrous optimisation?




Make one transition
(e.g. reversals away from live bearing)
occur at a very fast rate

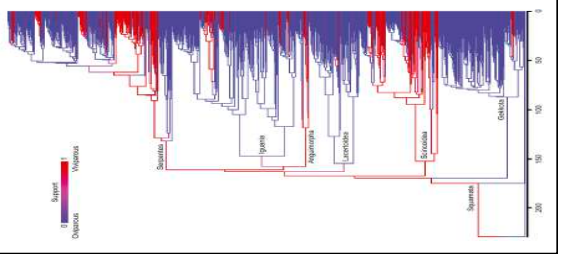





## Mis-paramaterised model

- The tree has lots of rate variability
- But model has no relaxed clock to account for it
- But model has asymmetric evolution – so that is hijacked to account for it

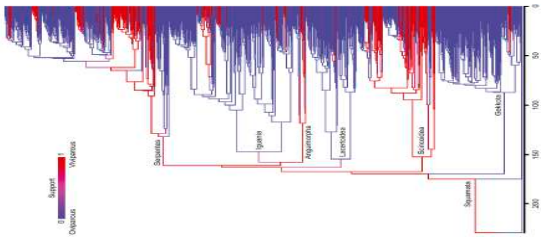
- It uses the substitution model to account for among-lineage rate heterogeneity.



## Mis-paramaterised model

Hijacking the substitution model to account for among-lineage rate heterogeneity, results:

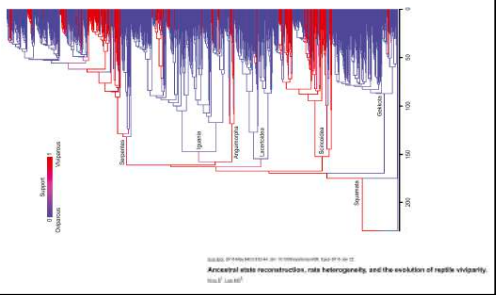
- Weird ancestral state reconstructions
- Weird substitution model
- Paper in high-impact journal



## BEAST to the rescue (again)

Reanalysed the data with BEAST using

- Random local clock
- Asymmetric (time-irreversible) matrix
- Traced character over a fixed tree (starting tree, turned off all branch swapping)



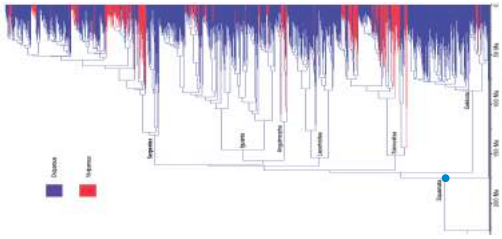
## BEAST to the rescue

Results – normality restored

## BEAST to the rescue

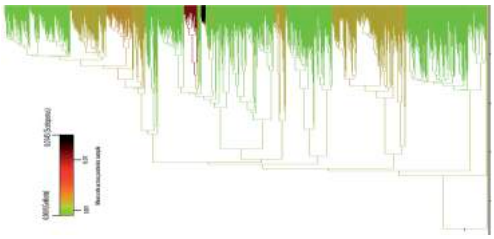
Results – normality restored

- Egg-laying is ancestral for lizards and snakes
- About 100 convergent evolutions of live-bearing
- Only 3-4 reversals back to egg laying (all biologically sensible)
- Rate shifts all very sensible



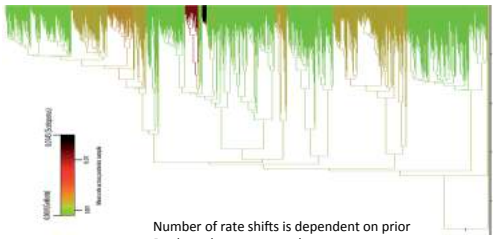
## BEAST to the rescue

Results – normality restored
- Egg-laying is ancestral for lizards and snakes
- About 100 convergent evolutions of live-bearing
- Only 3-4 reversals back to egg laying (all biologically sensible)
- Rate shifts all very sensible



## BEAST to the rescue

Results – normality restored
- Egg-laying is ancestral for lizards and snakes
- About 100 convergent evolutions of live-bearing
- Only 3-4 reversals back to egg laying (all biologically sensible)
- Rate shifts all very sensible



Number of rate shifts is dependent on prior
But branch rates are not!

## Bayesian Inference and Phenotypic Data

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

## Bayesian Inference and Phenotypic Data

- Bayesian phylogenetics (especially BEAST) has a lot to offer for morphologists and palaeontologists
- Peculiarities of morphological / palaeontological data
- Some case studies
  - Cambrian trilobites (evolutionary rates)
  - Crocodile evolution (detecting convergent evolution)
  - Lizard viviparity (ancestral state reconstruction)
- You don't need to be a programming whiz to good stuff in BEAST(2)