Ground-truthing tip-dating methods
(And: Bayesian tip-dating with continuous characters using BEASTmaster)

Ground truth (Tedford et al.)
Uniform node age prior
BEASTmaster, Beast2 SABD

Technical Session V:
Histology and Methods
Wednesday, October 14, 2015; 1:45-4:15 PM
3:45 pm, Landmark D
Hyatt Regency, SVP 2015, Dallas, TX

Nicholas J. Matzke, DECRA Fellow
Division of Ecology, Evolution, and Genetics
Research School of Biology
The Australian National University
phylo.wikidot.com/nicholas-j-matzke

Ground truth from: Summary tree of Tedford et al. (2008)
Character data from: Slater (2015)

Thanks for support from:
PLEASE TWEET ME

Also:
Slides are online at:
phylo.wikidot.com/nicholas-j-matzke

#BEASTmasteR
Background

Tip-dating is taking off!
Tip-dating is taking off!

Recent papers on tip-dating


Tip-dating is taking off!

“Putting Fossils in Trees” symposium last year


Tip-dating: Original motivation

Best practices include:

• Total-evidence analysis to place fossils
  
  Requires:
  - morphology matrix for living and fossil taxa
  - good resolution of tree

• Statistical determination of node-date prior distributions
  (e.g. methods by Charles Marshall, Tracy Heath)

  Requires:
  - Fossils with secure dates
  - Information about sampling through time
  - Sampling model
  - Lots of fossils

Why not just go all the way and put the dated fossils into the tree?

If you have all of this anyway...
Tip-dating: more reasons to try it

Node-dating procedure
(ideal; Parham et al. 2012 best practices)

- Sequence/download DNA
  - Make undated molecular phylogeny

  - Collaborate with a paleontologist
    - Undated total-evidence analysis with morphology + DNA

    - Fossils that are placed with high confidence provide minimum ages;
      make hard min, soft max distribution using expert knowledge or method

    - Dated tree ➔ downstream analysis ➔ Publication

Tip-dating procedure

- Sequence/download DNA

  - Add morphology matrix from living & fossil specimens

    - Include fossil specimen dates, with whatever uncertainty exists

    - Dated trees with fossils included ➔ downstream analysis ➔ Publication
Tip-dating: more reasons to try it

Node-dating

Results of node-dating
- Dated molecules-only tree (no fossils, no extant morphology-only taxa)
- Comparative methods can be run on the dated tree, but any fossil trait data is lost, or has to be added ad hoc (e.g., Slater et al.)
- Uncertainty in fossil placement not included; using only confidently-placed fossils might cause biases
- Most hard-won paleontological data is “thrown away” by node-dating

Paleo a mere “source of dates”

Tip-dating

Results of tip-dating
- Dated tree including fossils
- Comparative methods can be run on the dated tree, including any fossil data naturally
- Uncertainty in fossil placement taken into account in the dates and the posterior tree sample (and the comparative methods)
- Parameter estimates such as:
  - sampling rate through time
  - rate of morphological change
  - variability in morph. rate

Paleontologists & their data become key to both dating & downstream analyses!
Node-based dating

Tip-based dating
Node-based dating
Prior distribution on node date:

Tip-based dating
Node-based dating

Prior distribution on node date:

Tip-based dating
Node-based dating

Prior distribution on node date:

Tip-based dating
Node-based dating

Prior distribution on node date:

Tip-based dating
Node-based dating

Prior distribution on node date:

MCMC sample #1

MCMC sample #2

Tip-based dating

Prior distribution on specimen dates:

MCMC sample #1

MCMC sample #2

MCMC sample #3
Node-based dating

Prior distribution on node date:

MCMC sample #1
MCMC sample #2

Tip-based dating
Node-based dating

Prior distribution on node date:

MCMC sample #1

MCMC sample #2

MCMC sample #3

Tip-based dating
Node-based dating

Prior distribution on node date:

MCMC generation:
102,000
112,000

Tip-based dating
Node-based dating

Prior distribution on node date:

MCMC sample #1

MCMC sample #2

MCMC sample #3

Tip-based dating
Node-based dating

Prior distribution on node date:

MCMC sample #1
MCMC sample #2
MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

Fossils #1, 2, 3
**Node-based dating**

Prior distribution on node date:

- MCMC sample #1
- MCMC sample #2
- MCMC sample #3

**Time (Ma)**

- 200
- 100
- 0

**Tip-based dating**

Prior distribution on specimen dates:

- MCMC sample #1

**Fossils #1, 2, 3**
Node-based dating

Prior distribution on node date:

MCMC sample #1
MCMC sample #2
MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

MCMC sample #1

Fossils #1, 2, 3
Node-based dating

Prior distribution on node date:

MCMC sample #1

MCMC sample #2

MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

Fossils #1, 2, 3

MCMC sample #1
Node-based dating

Prior distribution on node date:

MCMC sample #1

MCMC sample #2

MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

MCMC sample #1

Fossils #1, 2, 3
Node-based dating

Prior distribution on node date:

- MCMC sample #1
- MCMC sample #2
- MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

Fossils #1, 2, 3

MCMC sample #1

MCMC sample #2

MCMC sample #3

Tip-based dating
**Node-based dating**

Prior distribution on node date:

- MCMC sample #1
- MCMC sample #2
- MCMC sample #3

Time (Ma)

**Tip-based dating**

Prior distribution on specimen dates:

- MCMC sample #1
- MCMC sample #2

Fossils #1, 2, 3

Time (Ma)
Node-based dating

Prior distribution on node date:

MCMC sample #1
MCMC sample #2
MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

MCMC sample #1
MCMC sample #2

Fossils #1, 2, 3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3
Fossil #1
Fossil #2
Fossil #3

Time (Ma)
Node-based dating

Prior distribution on node date:

MCMC sample #1
MCMC sample #2
MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

MCMC sample #1
MCMC sample #2
MCMC sample #3

Fossils #1, 2, 3
Node-based dating

Prior distribution on node date:

- MCMC sample #1
- MCMC sample #2
- MCMC sample #3

Tip-based dating

Prior distribution on specimen dates:

- Fossils #1, 2, 3
Tip-dating is taking off!

“Putting Fossils in Trees” symposium last year
Background

Tip-dating is taking off!

“Putting Fossils in Trees” symposium last year

Now 15+ studies in the literature
Background

Tip-dating is taking off!

“Putting Fossils in Trees” symposium last year

Now 15+ studies in the literature

Two major reviews:
Background

Tip-dating is taking off!

“Putting Fossils in Trees” symposium last year

Now 15+ studies in the literature

Two major reviews:
Background

Tip-dating is taking off!

“Putting Fossils in Trees” symposium last year

Now 15+ studies in the literature

Two major reviews:

Review
Dating Tips for Divergence-Time Estimation
Joseph E. O’Reilly,1 Mario dos Reis,2,† and Philip C.J. Donoghue1,∗

Morphological Phylogenetics in the Genomic Age
Michael S.Y. Lee1,2,∗ and Alessandro Pauci1,∗
1Earth Sciences Section, South Australian Museum, North Terrace, Adelaide SA 5000, Australia
2School of Biological Sciences, University of Adelaide, SA 5005, Australia
∗Correspondence: mike.lee@samuseum.sa.gov.au
http://dx.doi.org/10.1016/j.cub.2015.07.009
Background

Tip-dating is taking off!
“Putting Fossils in Trees” symposium last year
Now 15+ studies in the literature
Two major reviews:

But...does it work?
Tip-dating studies: results...mixed

Examples:

Authors judged date results implausible


Very high uncertainty


Should we expect current models to work on *anything*?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
</table>


Should we expect current models to work on *anything*?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
</tbody>
</table>
Should we expect current models to work on *anything*?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
</tbody>
</table>
Should we expect current models to work on *anything*?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
</tr>
</tbody>
</table>
Should we expect current models to work on anything?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
</tr>
</tbody>
</table>
Should we expect current models to work on *anything*?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
</tr>
</tbody>
</table>
Should we expect current models to work on anything?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
</tr>
<tr>
<td>Clade features that make tip-dating challenging</td>
<td>Examples</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
</tr>
<tr>
<td>Likely changes in molecular/ morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
</tr>
</tbody>
</table>
Should we expect current models to work on *anything*?

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
</tr>
<tr>
<td>Likely changes in molecular/morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
</tr>
</tbody>
</table>
The crucial question is:

When are the problems due to the method/model, and when are the problems due to the data?

(i.e., when are the data hopeless, given available models)
We need to “ground-truth” our tip-dating methods

Simulations are fine, but are always too simple

If a method/model works to produce basically “known” truth, then it has potential for making inferences on less ideal datasets

But, if a model/method returns nonsense even on an ideal dataset, then probably it shouldn’t be used.
Ground-truth using:
Ground-truth using: Fossil Canidae
Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods. Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
</tr>
<tr>
<td>Likely changes in molecular/ morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
</tr>
</tbody>
</table>
# Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods. Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td></td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td></td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td></td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td></td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td></td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td></td>
</tr>
<tr>
<td>Likely changes in molecular/morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td></td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods). Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td></td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td></td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td></td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td></td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td></td>
</tr>
<tr>
<td>Likely changes in molecular/ morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td></td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
# Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods. Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td>Approximately constant macroevolutionary regime</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td></td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td></td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td></td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td></td>
</tr>
<tr>
<td>Likely changes in molecular/ morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td></td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
## Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods. Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td>Approximately constant macroevolutionary regime</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td>Mostly endemic to a single region (North America) for most of Canidae history</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td></td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td></td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td></td>
</tr>
<tr>
<td>Likely changes in molecular/morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td></td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
## Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods. Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td>Approximately constant macroevolutionary regime</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td>Mostly endemic to a single region (North America) for most of Canidae history</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td>Fossils have many characters (100+)</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td></td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td></td>
</tr>
<tr>
<td>Likely changes in molecular/ morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td></td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O’Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods). Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td>Approximately constant macroevolutionary regime</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td>Mostly endemic to a single region (North America) for most of Canidae history</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td>Fossils have many characters (100+)</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td>Fossils preserved continuously throughout clade history (40-0 Ma)</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td></td>
</tr>
<tr>
<td>Likely changes in molecular/morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td></td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O’Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods. Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td>Approximately constant macroevolutionary regime</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td>Mostly endemic to a single region (North America) for most of Canidae history</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td>Fossils have many characters (100+)</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td>Fossils preserved continuously throughout clade history (40-0 Ma)</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td>Approximately constant preservability</td>
</tr>
<tr>
<td>Likely changes in molecular/ morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td></td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
# Ground-truth using: Fossil Canidae

Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods. Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td>Approximately constant macroevolutionary regime</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td>Mostly endemic to a single region (North America) for most of Canidae history</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td>Fossils have many characters (100+)</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td>Fossils preserved continuously throughout clade history (40-0 Ma)</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td>Approximately constant preservability</td>
</tr>
<tr>
<td>Likely changes in molecular/ morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td>Moderate change</td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Clade features that present challenges to tip-dating methods (or any dating estimation methods). Canidae exhibit few of the issues that may confound dating in other clades (e.g. angiosperms, mammals, birds).

<table>
<thead>
<tr>
<th>Clade features that make tip-dating challenging</th>
<th>Examples</th>
<th>Canidae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade occupies widely disparate niches (habitats, body form, etc.)</td>
<td>angiosperms, mammals</td>
<td>Clade in about the same ecological niche (carnivore)</td>
</tr>
<tr>
<td>Clade spans a mass extinction and post-extinction diversification</td>
<td>mammals, birds</td>
<td>Approximately constant macroevolutionary regime</td>
</tr>
<tr>
<td>Clade has a massive worldwide radiation</td>
<td>angiosperms, mammals, birds</td>
<td>Mostly endemic to a single region (North America) for most of Canidae history</td>
</tr>
<tr>
<td>Fossils have few characters</td>
<td>angiosperms (pollen), bivalves</td>
<td>Fossils have many characters (100+)</td>
</tr>
<tr>
<td>Fossils episodic or scarce near possible clade origin</td>
<td>placentals, angiosperms, Cambrian arthropods</td>
<td>Fossils preserved continuously throughout clade history (40-0 Ma)</td>
</tr>
<tr>
<td>Possibly major changes in preservability</td>
<td>angiosperms (woody vs. herbaceous); Cambrian phyla (soft vs. hard parts); hominids (forest vs. savanna habitats)</td>
<td>Approximately constant preservability</td>
</tr>
<tr>
<td>Likely changes in molecular/morphological rate (due to major changes in body size, population size, growth rate, etc.)</td>
<td>angiosperms (woody vs. herbaceous, annuals vs. perennials)</td>
<td>Moderate change</td>
</tr>
<tr>
<td>Available coded fossils represent only a small proportion of total known diversity</td>
<td>E.g. O'Leary et al. (2012) placental dataset</td>
<td>Coded fossil diversity greatly exceeds extant diversity</td>
</tr>
</tbody>
</table>
**Ground-truth using: Fossil Canidae**

<table>
<thead>
<tr>
<th>Eocene</th>
<th>Oligocene</th>
<th>Miocene</th>
<th>Pliocene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chadronian</td>
<td>Orellan</td>
<td>Whit-ryan</td>
<td>Arkareean</td>
</tr>
<tr>
<td>38</td>
<td>32</td>
<td>28</td>
<td>24</td>
</tr>
</tbody>
</table>

**Fig 6c.** Stratigraphic ranges and postulated phylogenetic relationships for taxa of the subfamily Caninae. The relationships is largely based on the preferred cladogram in figure 6b, modified by speculations about cladogenetic or anagenetic events. These speculations are based on morphological features too subtle to be coded in the data matrix but which, nonetheless, offer clues about possible relationships.

Parsimony + expert time-scaling of Caninae subtree: Tedford et al. 2008
Ground-truth using: Fossil Canidae

North America: Fossil Canidae are shown that schematically reflect dispersal to South America and extinction in the Pliocene. This schematic framework is based on Tedford et al. (2004) Geochronologic time frames and changes in geographic range chronologies are also plotted on this chart. Geographic range framework is based on Tedford et al. (2004) Geochronologic time frames and changes in geographic range chronologies are also plotted on this chart.

Parsimony + expert time-scaling of Caninae subtree: Tedford et al. 2008

Tedford et al. 2008
Methods: Digitizing the ground-truth tree

Used TreeRogue to digitize and combine trees from Hesperocyoninae, Borophaginae, Caninae (monographs of Wang & Tedford)
Methods: Canidae ground-truth

(Figure 6.6)

Diversity of species through time in the three subfamilies of Canidae

The width of each subfamily at a particular time slice in the figure corresponds to the number of species within the subfamily (shown in the scale above the Hesperocyoninae). Although both the Hesperocyoninae and the Borophaginae had a substantial diversity during the Oligocene, members of these subfamilies occupied different niches. The early borophagines were predominantly hypocarnivores, which complemented the hesperocyonines' hypercarnivory. By the time the Hesperocyoninae underwent its drastic decline in the early Miocene, the borophagines began to occupy the hypercarnivorous niches left open by the disappearing hesperocyonines. A similar complementary relationship also existed later between the borophagines and canines.

(Figure 7.1)

Phyletic relationships and intercontinental migrations of canid genera

Only selected genera are listed, and the actual diversity in each subfamily (Hesperocyoninae, Borophaginae, and Caninae) is somewhat higher than shown. The width of each lineage in the figure roughly corresponds to the diversity of the lineage through time. For a more realistic sense of diversity through time, see figure 6.6. Arrows indicate directions of migrations.
Methods: Digitizing the ground-truth tree

Used *TreeRogue* to digitize and combine trees from Hesperocyoninae, Borophaginae, Caninae (monographs of Wang & Tedford)
**Methods:** Digitizing the ground-truth tree

Used *TreeRogue* to digitize and combine trees from Hesperocyoninae, Borophaginae, Caninae (monographs of Wang & Tedford)
**Methods: Digitizing the ground-truth tree**

Used *TreeRogue* to digitize and combine trees from Hesperocyoninae, Borophaginae, Caninae (monographs of Wang & Tedford)

**Example dates:**

1. LCA crown *Canis*: ~4 Ma
2. LCA crown Caninae: ~12 Ma
3. LCA total-group Canidae: ~41 Ma
Methods: Digitizing the ground-truth tree

Used TreeRogue to digitize and combine trees from Hesperocyoninae, Borophaginae, Caninae (monographs of Wang & Tedford)

Example dates:

1. LCA crown *Canis*: ~4 Ma
2. LCA crown Caninae: ~12 Ma
3. LCA total-group Canidae: ~41 Ma
Methods: Character data

Character data and dates:
from Slater (2015), *PNAS*
(revision of Wang/Tedford datasets)
Methods: Models / programs

Compared methods:
MrBayes, Beast2

Compared models:

MrBayes
- Slater (2015) setup (tip and node dates)
- Tip-dates only
- New fossilized BD process

Beast2
- BEASTmaster
- Tip-dates only, BDSS
- Tip-dates only, sampled ancestors
**Methods:** Models /
programs

MrBayes tip-dating: Ronquist et al. (2012)

First easy-to-use tip-dating method

Until 2015, assumed the “uniform node age” prior

Argument: “uniform node age” was more agnostic prior than the Birth-Death Serial Sampling (BDSS) prior of Stadler
<table>
<thead>
<tr>
<th></th>
<th>MrBayes</th>
<th>Beast1</th>
<th>Beast2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setup</strong></td>
<td>easy</td>
<td>hard</td>
<td>hard*</td>
</tr>
<tr>
<td><strong>Tree models</strong></td>
<td>uniform, BDSS, SABD</td>
<td>BDSS only</td>
<td>BDSS, SABD</td>
</tr>
<tr>
<td><strong>New features</strong></td>
<td>some</td>
<td>no</td>
<td>many</td>
</tr>
<tr>
<td><strong>flexible</strong></td>
<td>no</td>
<td>sorta</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Works?</strong></td>
<td>sorta</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
BEASTmaster: setup

Fossils

Characters
1011101
1010111
0110122

Dated phylogenies

Settings
models, priors, settings

XML
BEAST requires XML input
BEAST requires XML input

```xml
<!-- Prior on samplingRate(s) -->

<!-- Prior probability density on the value of the parameter `samplingRate_bin0`, according to a uniform distribution with lower=0, upper=10. -->
    <prior id="prP_UniformDistrib_on_param_samplingRate_bin0" name="distribution" x="@samplingRate_bin0">
        <Uniform id="UniformDistrib_on_param_samplingRate_bin0" name="distr" offset="0" lower="0" upper="10"/>
    </prior>

<!-- Also: we are assuming for now that rho at time 0 is fixed -->

<!-- Probability of MCMC sampled tree given treeModel parameters -->
    <distribution id="BDSKY_prior_shared_tree" spec="beast.evolution.speciation.BirthDeathSkylineModel" tree="@shared_tree" origin="@originTime" birthRate="@birthRate" birthRateChangeTimes="@birthRateChangeTimes_tops" deathRate="@deathRate" deathRateChangeTimes="@deathRateChangeTimes_tops" samplingRate="@samplingRate" samplingRateChangeTimes="@samplingRateChangeTimes_tops" rho="@rho" rhoSamplingTimes="@rhoSamplingTimes" reverseTimeArrays="@reverseTimeArrays"/>
    <distribution id="likelihood" spec="util.CompoundDistribution">
```
BEAST requires XML input

The Matrix
BEAST2: BEAUTi

Fossils

NEXUS
1011101
1010111
0110122

Beauti models, priors, settings

XML

Dated phylogenies
**BEAST2: BEAUTi**

Things that are hard in BEAUTi

(* although it’s always being updated*)

1. Importing NEXUS morphology
2. Tip dates, tip-date priors
3. Clade constraints and node age priors
5. Remembering everything.
BEAST2: BEAUTi

Fossils

NEXUS
1011101
1010111
0110122

Dated phylogenies

Beauti models, priors, settings

XML
BEASTmasteR: setup

Fossils

NEXUS
1011101
1010111
0110122

Excel models, priors, settings

XML

Dated phylogenies
BEAST2: BEAUTi

Things that are hard in BEAUTi

(* although it’s always being updated)

1. Importing NEXUS morphology
2. Tip dates, tip-date priors
3. Clade constraints and node age priors
5. Remembering everything.
BEASTmaster: things that are easy

1. Importing NEXUS morphology (discrete or continuous)
2. Tip dates, tip-date priors
3. Clade constraints and node age priors
   3a. Changing priors, e.g. to Normal
4. BDSS, Sampled Ancestors, Skyline setup
5. Relative priors
6. Saving/loading/changing
# Tip-dating: data options

<table>
<thead>
<tr>
<th></th>
<th>MrBayes</th>
<th>Beast2/BEAUTi</th>
<th>Beast2/BEASTmaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>easy</td>
<td>easy</td>
<td>easy</td>
</tr>
<tr>
<td>Amino acids</td>
<td>easy</td>
<td>easy</td>
<td>easy</td>
</tr>
<tr>
<td>Discrete morphology</td>
<td>easy</td>
<td>morph-models</td>
<td>easy</td>
</tr>
<tr>
<td>Continuous</td>
<td>crashes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Combine all</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
Results

Ground-truth tree:

Key dates:
1. LCA crown *Canis*: ~4 Ma
2. LCA crown Caninae: ~12 Ma
3. LCA total-group Canidae: ~41 Ma
Results

Slater (2015) analysis: tip+node dates
MrBayes uniform prior

Key dates:

1. LCA crown *Canis*: ~7 Ma
2. LCA crown Caninae: ~22 Ma
3. LCA total-group Canidae: ~41 Ma
Results

Slater (2015) data:
tip dates only
MrBayes uniform prior

Key dates:
1. LCA crown Canis: ~27 Ma
2. LCA crown Caninae: ~36 Ma
3. LCA total-group Canidae: ~50 Ma
Results

Slater (2015) data:
tip dates only

BEASTmasterR, BDSS prior

Key dates:

1. LCA crown *Canis*: ~4 Ma
2. LCA crown Caninae: ~12 Ma
3. LCA total-group Canidae: ~37 Ma
Results

Slater (2015) data: tip dates only

**BEASTmasterR, SABD**

Key dates:

1. LCA crown *Canis*: ~3 Ma
2. LCA crown Caninae: ~11 Ma
3. LCA total-group Canidae: ~37 Ma
Results

Ground-truth tree:

Key dates:

1. LCA crown *Canis*: ~4 Ma
2. LCA crown Caninae: ~12 Ma
3. LCA total-group Canidae: ~41 Ma
Results

Slater (2015) data: tip dates only
MrBayes fossilized BD

Key dates:
1. LCA crown *Canis*: NOT FOUND
2. LCA crown Caninae: NOT FOUND
3. LCA total-group Canidae: ~37 Ma
Results

Slater (2015) data: tip dates only
MrBayes fossilized BD

Key dates:
1. LCA crown *Canis*: NOT FOUND
2. LCA crown Caninae: NOT FOUND
3. LCA total-group Canidae: ~37 Ma

Ruh-roh!!
African Wild Dogs in the Borophagines
Results

Slater (2015) data:
tip dates only
MrBayes fossilized BD, more thorough search

1. LCA crown *Canis*: 4 Ma
2. LCA crown Caninae: 12 Ma
3. LCA total-group Canidae: ~38 Ma
Results

Slater (2015) data: tip dates only

**BEASTmasterR, BDSS prior**

Key dates:

1. LCA crown *Canis*: ~4 Ma
2. LCA crown Caninae: ~12 Ma
3. LCA total-group Canidae: ~47 Ma
Results

Ground-truth tree:

Key dates:

1. LCA crown *Canis*: ~4 Ma
2. LCA crown Caninae: ~12 Ma
3. LCA total-group Canidae: ~41 Ma
Adding continuous characters:

Beast2/BEASTmasteR

Brownian motion/
- random walk, or
- relaxed random walk
(RRW)
(relaxed clock)
Adding continuous characters:

Methods

Simulate characters with different rates on an inferred tree

Add to standard characters

Infer tree again
Adding continuous characters:

Results

More traits = better posterior probability (PP)

But: mixing problems increase with more traits

(too many independent rate parameters)
Adding continuous characters:

Results

More traits = better posterior probability (PP)

But: mixing problems increase with more traits

(too many independent rate parameters)
Adding continuous characters:

Results

More traits = better posterior probability (PP)

But: mixing problems increase with more traits

(too many independent rate parameters)
Adding continuous characters:

Results

More traits = better posterior probability (PP)

But: mixing problems increase with more traits

(too many independent rate parameters)
Conclusions

1. BEASTmasteR eases tip-dating
2. Priors matter: Uniform=bad BDsamp=good
3. Combine data, incl. continuous

Try it at:
phylo.wikidot.com /beastmaster
Ground-truthing tip-dating methods
(And: Bayesian tip-dating with continuous characters using BEASTmaster)

Acknowledgements:
April Wright
Mike Lee
Graeme Lloyd
David Bapst

Technical Session V:
Histology and Methods
Wednesday, October 14, 2015; 1:45-4:15 PM
Hyatt Regency, SVP 2015, Dallas, TX

Nicholas J. Matzke, DECRA Fellow
Division of Ecology, Evolution, and Genetics
Research School of Biology
The Australian National University
phylo.wikidot.com/nicholas-j-matzke