Infer relationships among three species:

Outgroup:
Three possible trees (topologies):
Prior distribution

Model

Posterior distribution

Data (observations)
Bayes' theorem

\[ f(\theta \mid X) = \frac{f(\theta) f(X \mid \theta)}{\int f(\theta) f(X \mid \theta) d\theta} \]

Posterior distribution

Prior distribution

“Likelihood”

Normalizing constant
Model: topology AND branch lengths

\[ \theta \quad \text{Parameters} \]

\[ \text{topology} (\tau) \]
\[ \text{branch lengths} \ (v_i) \]
\[ \text{(expected amount of change)} \]

\[ \theta = (\tau, v) \]
Posterior probability distribution

\[ f(\theta | X) \]

Parameter space

\( \theta \)

Posterior probability

tree 1  tree 2  tree 3
We can focus on any parameter of interest (there are no nuisance parameters) by marginalizing the posterior over the other parameters (integrating out the uncertainty in the other parameters).

(Percentages denote marginal probability distribution on trees)
Why is it called marginalizing?

<table>
<thead>
<tr>
<th>branch length vectors</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
<th>$\tau_3$</th>
<th>joint probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu^1$</td>
<td>0.10</td>
<td>0.07</td>
<td>0.12</td>
<td>0.29</td>
</tr>
<tr>
<td>$\nu^2$</td>
<td>0.05</td>
<td>0.22</td>
<td>0.06</td>
<td>0.33</td>
</tr>
<tr>
<td>$\nu^3$</td>
<td>0.05</td>
<td>0.19</td>
<td>0.14</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.48</td>
<td>0.32</td>
<td></td>
</tr>
</tbody>
</table>

marginal probabilities
Markov chain Monte Carlo

1. Start at an arbitrary point
2. Make a small random move
3. Calculate height ratio \( (r) \) of new state to old state:
   1. \( r > 1 \) \( \rightarrow \) new state accepted
   2. \( r < 1 \) \( \rightarrow \) new state accepted with probability \( r \). If new state not accepted, stay in the old state
4. Go to step 2

The proportion of time the MCMC procedure samples from a particular parameter region is an estimate of that region’s posterior probability density.
stationary phase sampled with thinning
(rapid mixing essential)

\[ \text{InL} \]

\[ 0 \quad 5000000 \quad 10000000 \quad 15000000 \quad 20000000 \]

**Generation**
Metropolis-coupled Markov chain Monte Carlo
a. k. a.
MCMC
a. k. a.
(MC)^3

cold chain

heated chain
cold chain

hot chain
cold chain

unsuccessful swap

hot chain
cold chain

hot chain
cold chain

hot chain
cold chain

successful swap

hot chain
cold chain

hot chain
cold chain

hot chain
cold chain

successful swap

hot chain
Incremental Heating

*T* is temperature, *λ* is heating coefficient

\[ T = \frac{1}{1 + \lambda i} \quad i = \{0,1,...,n-1\} \]

**Example for \( \lambda = 0.2 \):**

<table>
<thead>
<tr>
<th>( i )</th>
<th>( T )</th>
<th>Distr.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00</td>
<td>( f(\theta \mid X)^{1.00} )</td>
<td>cold chain</td>
</tr>
<tr>
<td>1</td>
<td>0.83</td>
<td>( f(\theta \mid X)^{0.83} )</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.71</td>
<td>( f(\theta \mid X)^{0.71} )</td>
<td>heated chains</td>
</tr>
<tr>
<td>3</td>
<td>0.62</td>
<td>( f(\theta \mid X)^{0.62} )</td>
<td></td>
</tr>
</tbody>
</table>
Assessing Convergence

- Plateau in the trace plot (unreliable!!)
- Compare windows within the same run (better)
- Compare independent runs starting from different randomly chosen topologies (best)
Astragalus (ITS)

$s = 140$
$c = 686$
GTR+$\Gamma$
Clade probability in analysis 1 vs. Clade probability in analysis 2 for Angiosperms (atpB).

- $s = 357$
- $c = 1497$
- GTR+SS
Improving Convergence

(Only if convergence diagnostics indicate problem!)

- Change tuning parameters of proposals to bring acceptance rate into the range 10% to 70%
- Propose changes to 'difficult' parameters more often
- Use different proposal mechanisms
- Change heating temperature to bring acceptance rate of swaps between adjacent chains into the range 10% to 70%.
- Run the chain longer
- Increase the number of heated chains
- Make the model more realistic
Target distribution

Too modest proposals
Acceptance rate too high
Poor mixing

Too bold proposals
Acceptance rate too low
Poor mixing

Moderately bold proposals
Acceptance rate intermediate
Good mixing
Running MrBayes

- Use `execute` to bring data in a Nexus file into MrBayes
- Set the model and priors using `lset` and `prset`
- Run the chain using `mcmc`
- Summarize the parameter samples using `sump`
- Summarize the tree samples using `sumt`
- Note that MrBayes 3.1 runs two independent analyses by default
Ver. 3.1 Convergence Diagnostics

- By default performs two independent analyses starting from different random trees (mcmc nruns=2)
- Average standard deviation of clade frequencies calculated and presented during the run (mcmc mcmcdiagn=yes diagnfreq=1000) and written to file (.mcmc)
- Standard deviation of each clade frequency and potential scale reduction for branch lengths calculated with sumt
- Potential scale reduction calculated for all substitution model parameters with sump
Bayes' theorem

\[ f(\theta \mid X) = \frac{f(\theta)f(X \mid \theta)}{\int f(\theta)f(X \mid \theta) \, d\theta} = \frac{f(\theta)f(X \mid \theta)}{f(X)} \]

Marginal likelihood (of the model)

We have implicitly conditioned on a model:

\[ f(\theta \mid X, M) = \frac{f(\theta \mid M)f(X \mid \theta, M)}{f(X \mid M)} \]
Bayesian Model Choice

Posterior model odds:

$$\frac{f(M_1)f(X|M_1)}{f(M_0)f(X|M_0)}$$

Bayes factor:

$$B_{10} = \frac{f(X|M_1)}{f(X|M_0)}$$
Bayesian Model Choice

- The normalizing constant in Bayes’ theorem, the marginal probability of the model, $f(X)$ or $f(X|M)$, can be used for model choice.
- $f(X|M)$ can be estimated by taking the harmonic mean of the likelihood values from the MCMC run (MrBayes will do this automatically with ‘sump’).
- Any models can be compared: nested, non-nested, data-derived.
- No correction for number of parameters.
- Can prefer a simpler model over a more complex model.
- Critical values in Kass and Raftery (1997).
## Bayes Factor Comparisons

### Interpretation of the Bayes factor

<table>
<thead>
<tr>
<th>$2\ln(B_{10})$</th>
<th>$B_{10}$</th>
<th>Evidence against $M_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 2</td>
<td>1 to 3</td>
<td>Not worth more than a bare mention</td>
</tr>
<tr>
<td>2 to 6</td>
<td>3 to 20</td>
<td>Positive</td>
</tr>
<tr>
<td>6 to 10</td>
<td>20 to 150</td>
<td>Strong</td>
</tr>
<tr>
<td>&gt; 10</td>
<td>&gt; 150</td>
<td>Very strong</td>
</tr>
</tbody>
</table>
Resources

- MrBayes web site (www.mrbayes.net)
- MrWiki manual (on the MrBayes web site)
- MrBayes 3.1 command reference
- MrBayes 3.1 manual (pdf) with two tutorials: a simple analysis (primates.nex) and an analysis of partitioned data (cynmix.nex)
- Last pages of the manual contain graphical summaries of the MrBayes 3 models and most common types of proposal mechanisms
- MrBayes 3 chapter: complex partitioned analysis (kim.nex)
Models supported by MrBayes 3 (simplified)

**Data Type**
- Restriction 0 - 1
  - State Frequencies (Substitution Rates): fixed/estimated (Dirichlet)
    - prset statefreqapr
  - Across-Site Rate Variation: equal/gamma
    - lset rates
  - Coding Bias: all/variable/nopresencesites/noabsencesites
    - lset coding
  - Misc.: unordered/ordered
c
type

- Standard 0 - 9
  - State Frequencies (Substitution Rates): equal/estimated (SymmDir)
    - prset symdirihyperpr
  - Across-Site Rate Variation: equal/gamma
    - lset rates
  - Coding Bias: all/variable/informative
    - lset coding
  - Misc.: unordered/ordered
c
type

**Data Type**
- DNA
  - Model Type: 4by4, doublet, codon
  - State Frequencies: fixed/est. (Dirichlet)
    - prset statefreqapr
  - Substitution Rates: F81/HKY/GTR
    - lset nst = 1/2/6
  - Across-Site Rate Variation: equal/gamma
    - propinv/invgamma/adgamma
    - lset rates
  - Across-Tree Rate Variation: yes/no
    - lset covarion
  - Across-Site Omega Variation
  - Omega Variation: equal/Ny98/M3
    - lset omegavar
### Models supported by MrBayes 3 (simplified)

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Model Type</th>
<th>State Frequencies</th>
<th>Substitution Rates</th>
<th>Across-Site Rate Variation</th>
<th>Across-Tree Rate Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein A-Y</td>
<td>GTR prset aaamodelpr</td>
<td>fixed/est. (Dirichlet) prset statefreqpr</td>
<td>fixed/est. (Dirichlet) prset statefreqpr</td>
<td>equal/gamma/propinv/invgamma/adgamma/iset rates</td>
<td>yes/no lset covarion</td>
</tr>
<tr>
<td></td>
<td>Equalin prset aaamodelpr</td>
<td>fixed/est. (Dirichlet) prset statefreqpr</td>
<td>fixed to equal</td>
<td>equal/gamma/propinv/invgamma/adgamma/iset rates</td>
<td>yes/no lset covarion</td>
</tr>
<tr>
<td></td>
<td>Poisson/Jones/Dayhoff/Mtrev/Mtmam/Wag/Rtrev/Cprev/Vt/Blossum/mixed prset aaamodelpr</td>
<td>fixed/mixed</td>
<td>fixed/mixed</td>
<td>equal/gamma/propinv/invgamma/adgamma/iset rates</td>
<td>yes/no lset covarion</td>
</tr>
</tbody>
</table>

### Tree Type

- **Non-clock**
  - Unconstrained prset brelenspr
  - Exponential/Uniform prset brelenspr

- **Clock**
  - Clock prset brelenspr
  - Uniform prset brelenspr
  - Coalescence prset brelenspr
  - Birth-Death prset brelenspr

### Brlens Prior

- Exponential/Uniform prset brelenspr
- Uniform prset brelenspr
- Coalescence prset brelenspr
- Birth-Death prset brelenspr

### Additional parameters

- `see prset (Isset for ploidy)`
- Treeheight
- Theta, Ploidy Growth
- Speciation Extinction Sampleprob

### Brlens Variation Across Partitions

- Equal/proportional prset ratepr
- Unlinked
- unlink brelens

### Topology Variation Across Partitions

- Same/unlinked
- link topology
- unlink topology

---

This diagram illustrates the various models and parameters supported by MrBayes 3 for protein data, including data types, model types, state frequencies, substitution rates, and across-site and tree rate variations.
Sliding Window Proposal

\[ \delta \]

\[ x \]

New values are picked uniformly from a sliding window of size \( \delta \) centered on \( x \).

Tuning parameter: \( \delta \)

Bolder proposals: increase \( \delta \)

More modest proposals: decrease \( \delta \)

*Works best when the effect on the probability of the data is similar throughout the parameter range*
Dirichlet proposal

New values are picked from a Dirichlet (or Beta) distribution centered on $x$.
Tuning parameter: $\alpha$
Bolder proposals: decrease $\alpha$
More modest proposals: increase $\alpha$

*Works well for proportions, such as revmat and statefreqs.*
Multiplier Proposal

New values are picked from the equivalent of a sliding window on the log-transformed $x$ axis.

Tuning parameter: $\lambda = 2 \ln a$

Bolder proposals: increase $\lambda$

More modest proposals: decrease $\lambda$

Works well when changes in small values of $x$ have a larger effect on the probability of data than changes in large values of $x$. Example: branch lengths.
Two adjacent branches $a$ and $b$ are chosen at random.
The length of $a + b$ is changed using a multiplier with tuning parameter $\lambda$.
The node $x$ is randomly inserted on $a + b$ according to a uniform distribution.

Bolder proposals: increase $\lambda$.
More modest proposals: decrease $\lambda$.

The boldness of the proposal depends heavily on the uniform reinsertion of $x$, so changing $\lambda$ may have limited effect.
LOCAL

Three internal branches - $a$, $b$, and $c$ - are chosen at random. Their total length is changed using a multiplier with tuning parameter $\lambda$.
One of the subtrees A or B is picked at random. It is randomly reinserted on $a + b + c$ according to a uniform distribution.

Bolder proposals: increase $\lambda$
More modest proposals: decrease $\lambda$
Extending TBR

An internal branch $a$ is chosen at random
The length of $a$ is changed using a multiplier with tuning parameter $\lambda$
The node $x$ is moved, with one of the adjacent branches, in subtree A, one node at a time, each time the probability of moving one more branch is $p$ (the extension probability).
The node $y$ is moved similarly in subtree B.

Bolder proposals: increase $p$
More modest proposals: decrease $p$
Changing $\lambda$ has little effect on the boldness of the proposal.
Why Bayesian?

- Priors: accumulate scientific knowledge
- Easy to deal with complex models, and current models need to be improved
- Computational efficiency: today hundreds of taxa, can probably be extended considerably
- Convergence diagnostics to detect problems with convergence
- Model testing with Bayes factors
- Model averaging using reversible-jump MCMC sampling