An Introduction to Bayesian Inference of Phylogeny

John P. Huelsenbeck[1], Bruce Rannala[2], and John P. Masly[1]

[1] *Department of Biology, University of Rochester, Rochester, NY 14627, U.S.A.*

[2] *Department of Medical Genetics, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada*

A simple example of Bayesian inference

We will illustrate Bayesian inference using a simple example involving dice. Consider a box with 100 dice, 90 of which are fair and 10 of which are biased. The probability of observing some number of pips after rolling a fair or biased die is given in the following table:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

The probability of a high roll is larger for the biased dice than for the fair dice. Suppose that you draw a die at random from the box and roll it twice, observing a four on the first roll and a six on the second roll. What is the probability that the die is biased?

A Bayesian analysis combines ones prior beliefs about the probability of a hypothesis with the likelihood. The likelihood is the vehicle that carries the information about the hypothesis contained in the observations. In this case, the likelihood is simply the probability of observing a four and a six given that the die is biased or fair. Assuming independence of the tosses, the probability of observing a four and a six is

$$\Pr[⚃, ⚅ \mid \text{Fair}] = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

for a fair die and

$$\Pr[⚃, ⚅ \mid \text{Biased}] = \frac{4}{21} \times \frac{6}{21} = \frac{24}{441}$$

for a biased die. The probability of observing the data is 1.96 times greater under the hypothesis that the die is biased. In other words, the ratio of the likelihoods under the two hypotheses suggests that the die is biased.

Bayesian inferences are based upon the posterior probability of a hypothesis. The posterior probability that the die is biased can be obtained using Bayes' *(1)* formula:

$$\Pr[\text{Biased} \mid ⚃, ⚅] = \frac{\Pr[⚃, ⚅ \mid \text{Biased}] \times \Pr[\text{Biased}]}{\Pr[⚃, ⚅ \mid \text{Biased}] \times \Pr[\text{Biased}] + \Pr[⚃, ⚅ \mid \text{Fair}] \times \Pr[\text{Fair}]}$$

where Pr[Biased] and Pr[Fair] are the prior probabilities that the die is biased or fair, respectively. As we set up the problem, a reasonable prior probability that the die is biased would be the proportion of the dice in the box that were biased. The posterior probability is then

$$\Pr[\text{Biased} \mid ⚃, ⚅] = \frac{\frac{24}{441} \times \frac{1}{10}}{\frac{24}{441} \times \frac{1}{10} + \frac{1}{36} \times \frac{9}{10}} = 0.179$$

This means that our opinion that the die is biased changed from 0.1 to 0.179 after observing the four and six.

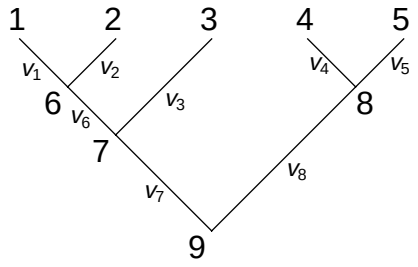

Figure 1.—An example of a phylogenetic tree for $s = 5$ species. The branch lengths are denoted v_i.

Depending upon one's viewpoint, the incorporation of prior beliefs about a parameter is either a strength or a weakness of Bayesian inference. It is a strength in as much as the method explicitly incorporates prior information in inferences about a hypothesis. However, it can often be difficult to specify a prior. For the dice example, it is easy to specify the prior as we provided information on the number of fair and biased dice in the box and also specify that a die was randomly selected. However, if we were to simply state that the die is either fair or biased, but did not specify a physical description of how the die was chosen, it would have been much more difficult to specify a prior specifying the probability that the die is biased. For example, one could have taken the two hypotheses to have been *a priori* equally probable or given much more weight to the hypothesis that had the die fair as severely biased dice are rarely encountered (or manufactured) in the real world.

Bayesian inference of phylogeny

Bayesian inference of phylogeny is based upon the posterior probability of a phylogenetic tree, τ. The posterior probability of the ith phylogenetic tree, τ_i, conditioned on the observed matrix of aligned DNA sequences ($\boldsymbol{X}$) is obtained using Bayes formula:

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{B(s)} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

[throughout, we denote conditional probabilities as $f(\cdot|\cdot)$]. Here, $f(\tau_i|\boldsymbol{X})$ is the posterior probability of the ith phylogeny and can be interpreted as the probability that τ_i is the correct tree given the DNA sequence data. The likelihood of the ith tree is $f(\boldsymbol{X}|\tau_i)$ and the prior probability of the ith tree is $f(\tau_i)$. The summation in the denominator is over all $B(s)$ trees that are possible for s species. This number is $B(s) = \frac{(2s-3)!}{2^{s-2}(s-2)!}$ for rooted trees, $B(s) = \frac{(2s-5)!}{2^{s-3}(s-3)!}$ for unrooted trees, and $B(s) = \frac{s!(s-1)!}{2^{s-1}}$ for labelled histories. Typically, an uninformative prior is used for trees, such that $f(\tau_i) = \frac{1}{B(s)}$

DNA sequence data.—We consider an aligned matrix of s DNA sequences:

$$\boldsymbol{X} = \{x_{ij}\} = \begin{array}{l} \text{Species 1} \\ \text{Species 2} \\ \text{Species 3} \\ \vdots \\ \text{Species } s \end{array} \left\{ \begin{array}{ccccc} A & A & C & C & T \\ A & A & C & G & G \\ A & C & C & C & T \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A & C & C & C & T \end{array} \right\}$$

The data matrix consists of the sequences for s species for $c = 5$ sites from a gene (c is the length of the aligned DNA sequences). The observations at the first site are $\boldsymbol{x}_1 = \{A, A, A, \ldots, A\}'$. In general, the information at the ith site in the matrix is denoted $\boldsymbol{x}_i$.

Phylogenetic models.—What is the probability of observing the data at the ith site? To calculate this probability, we assume a phylogenetic model. A phylogenetic model consists of a tree (τ_i) with branch

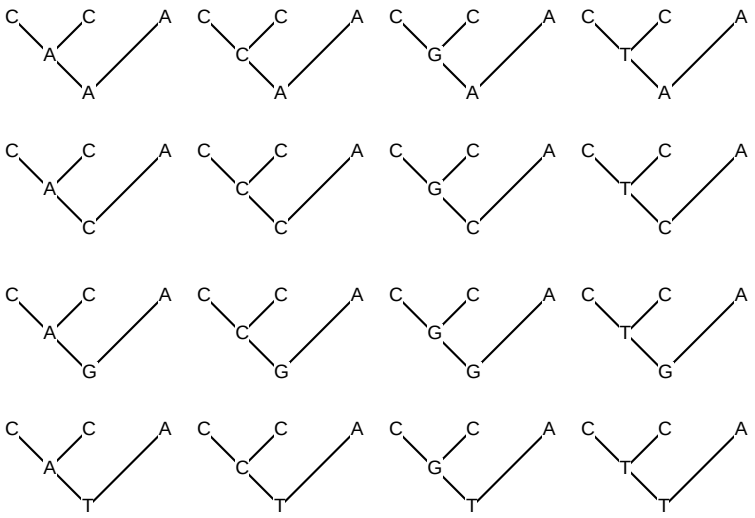

Figure 2.—The 16 possible assignments of nucleotides to the internal nodes of a tree of $s = 3$ species. The observations at the site are $\boldsymbol{x}_i = \{C, C, A\}$ and the unobserved nucleotides at the internal nodes of the tree are denoted $\boldsymbol{y}$.

lengths specified on the tree ($\boldsymbol{v}_i$) and a stochastic model of DNA substitution. Figure 1 shows an example of a phylogenetic tree of $s = 5$ species. The tips of the tree are labeled $1, 2, \ldots, s$ and the internal nodes of the tree are labeled $s+1, s+2, \ldots, 2s-1$; the root of the tree is always labeled $2s-1$. The lengths of the branches are denoted v_i and are in terms of the number of substitutions expected to occur along the ith branch. In general, the ancestor of node k will be denoted $\sigma(k)$; the ancestor of node 4 is $\sigma(4) = 8$. The ancestor of the root is $\sigma(2s-1) = \emptyset$.

The second part of the phylogenetic model consists of a stochastic model of DNA substitution. Here, the typical assumption is that DNA substitution follows a time-homogeneous Poisson process. The heart of the model is a matrix specifying the instantaneous rate of substitution from one nucleotide state to another:

$$\boldsymbol{Q} = \{q_{ij}\} = \begin{pmatrix} \cdot & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\ \pi_A r_{AC} & \cdot & \pi_G r_{CG} & \pi_T r_{CT} \\ \pi_A r_{AG} & \pi_C r_{CG} & \cdot & \pi_T r_{GT} \\ \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{CT} & \cdot \end{pmatrix}$$

where the matrix specifies the rate of change from nucleotide i (row) to nucleotide j (column). The nucleotides are in the order A, C, G, T. The diagonals of the matrix are specified such that the rows each sum to 0. The equilibrium (or stationary) frequencies of the four nucleotides are denoted π_i ($\boldsymbol{\pi} = \{\pi_A, \pi_C, \pi_G, \pi_T\}$). This matrix specifies the most general time-reversible model of DNA substitution and is referred to as the GTR model *(2)*. Because the rate of substitution and time are confounded, the $\boldsymbol{Q}$ matrix is rescaled such that $-\sum \pi_i q_{ii} = 1$ for all i (making the average rate of substitution 1). Over a branch of length v the transition probabilities are calculated as $\boldsymbol{P}(v, \boldsymbol{\theta}) = \{p_{ij}(v, \boldsymbol{\theta})\} = e^{\boldsymbol{Q}v}$. The parameters of the substitution model are contained in a vector $\boldsymbol{\theta}$.

The likelihood of a phylogeny.—The phylogenetic model consists of a tree (τ_i) with branch lengths ($\boldsymbol{v}_i$) and a stochastic model of DNA substitution that is specified by a matrix of instantaneous rates. The probability of observing the data at the ith site in the aligned matrix is a sum over all possible assignments of nucleotides to the internal nodes of the tree:

$$f(\boldsymbol{x}_i | \tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}) = \sum_{\boldsymbol{y}} \left[\pi_{y_{2s-1}} \left(\prod_{k=1}^{s} p_{y_{i\sigma(k)}, x_{ik}}(v_k, \boldsymbol{\theta}) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{i\sigma(k)}, y_{ik}}(v_k, \boldsymbol{\theta}) \right) \right]$$

Here, y_{ij} is the (unobserved) nucleotide at the jth node for the ith site. The summation is over all 4^{s-1} ways that nucleotides can be assigned to the internal nodes of the tree. Figure 2 illustrates the possible nucleotide assignments for a simple tree of $s = 3$ species. Felsenstein *(3)* introduced a pruning algorithm that efficiently calculates the summation. Often, the rate at the site is assumed to be drawn from a gamma distribution. This allows one to relax the assumption that the rate of substitution is equal across all sites. If gamma-distributed rate variation is assumed, then the probability of observing the data at the ith site becomes:

$$f(\boldsymbol{x}_i|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha) = \int_0^\infty \left\{ \sum_{\boldsymbol{y}} \left[\pi_{y_{2s-1}} \left(\prod_{k=1}^{s} p_{y_{i\sigma(k)}, x_{ik}}(v_k r, \boldsymbol{\theta}) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{i\sigma(k)}, y_{ik}}(v_k r, \boldsymbol{\theta}) \right) \right] \right\} f(r|\alpha)\, dr$$

where $f(r|\alpha)$ is the density of the rate r under the gamma model *(4)*. The parameter α is the shape parameter of the gamma distribution (here, the shape and the scale parameters of the gamma distribution are both set to α). Typically, this integral is impossible to evaluate. Hence, an approximation first suggested by Yang *(5)* is used in which the continuous gamma distribution is broken into K categories, each with equal weight. The mean rate from each category represents the rate for the entire category. The probability of observing the data at the ith site then becomes:

$$f(\boldsymbol{x}_i|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha) = \sum_{n=1}^{K} \left\{ \sum_{\boldsymbol{y}} \left[\pi_{y_{2s-1}} \left(\prod_{k=1}^{s} p_{y_{i\sigma(k)}, x_{ik}}(v_k r_n, \boldsymbol{\theta}) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{i\sigma(k)}, y_{ik}}(v_k r_n, \boldsymbol{\theta}) \right) \right] \right\} \frac{1}{K}$$

Assuming independence of the substitutions across sites, the probability of observing the aligned matrix of DNA sequences is

$$f(\boldsymbol{X}|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha) = \prod_{i=1}^{c} f(\boldsymbol{x}_i|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha)$$

Importantly, the likelihood can be calculated under a number of different models of character change. For example, the codon model describes the substitution process over triplets of sites (a codon) and allows the estimation of the nonsynonymous/synonymous rate ratio *(6, 7)*. Similarly, models of DNA substitution have been described that allow nonindenpendent substitutions to occur in stem regions of rRNA genes *(8)*. Finally, one can calculate likelihoods for amino acid *(9)*, restriction site *(10)*, and, more recently, morphological data *(11)*.

Bayesian inference of phylogeny.—As described so far, the likelihood depends upon several unknown parameters; generally, the phylogeny, branch lengths, and substitution parameters are unknown. The method of maximum likelihood estimates these parameters by finding the values of the parameters which maximize the likelihood function. Currently, programs such as PAUP* *(12)*, PAML *(13)*, and PHYLIP *(14)* estimate phylogeny using the method of maximum likelihood.

Bayesian inference is based instead upon the posterior probability of the parameter. As described above, the posterior probability of the ith tree is

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i) f(\tau_i)}{\sum_{j=1}^{B(s)} f(\boldsymbol{X}|\tau_j) f(\tau_j)}$$

where the likelihood function is integrated over all possible values for the branch lengths and substitution parameters:

$$f(\boldsymbol{X}|\tau_i) = \int_{v_i} \int_{\theta} \int_{\alpha} f(\boldsymbol{X}|\tau_i, \boldsymbol{v}_i, \boldsymbol{\theta}, \alpha) f(\boldsymbol{v}_i) f(\boldsymbol{\theta}) f(\alpha)\, d\boldsymbol{v}_i\, d\boldsymbol{\theta}\, d\alpha$$

Markov chain Monte Carlo.—Typically, the posterior probability cannot be calculated analytically. However, the posterior probability of phylogenies can be approximated by sampling trees from the posterior probability distribution. Markov chain Monte Carlo (MCMC) can be used to sample phylogenies according to their posterior probabilities. The Metropolis-Hastings-Green (MHG) algorithm *(15, 16, 17)* is an MCMC algorithm that has been used successfully to approximate the posterior probabilities of trees *(18, 19)*.

The MHG algorithm works as follows. Let $\boldsymbol{\Psi} = \{\tau, \boldsymbol{v}, \boldsymbol{\theta}, \alpha\}$ be a specific tree, combination of branch lengths, substitution parameters, and gamma shape parameter. The MHG algorithm constructs a Markov chain that has as its stationary frequency the posterior probability of interest (in this case, the joint posterior probability of τ, $\boldsymbol{v}$, $\boldsymbol{\theta}$, and α). The current state of the chain is denoted $\boldsymbol{\Psi}$. If this is the first generation of the chain, then the chain is initialized (perhaps by randomly picking a state from the prior). A new state is then proposed, $\boldsymbol{\Psi}'$. The probability of proposing the new state given the old state is $f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})$ and the probability of making the reverse move (which is never actually made) is $f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')$. The new state is accepted with probability

$$\begin{aligned} R &= \min\left(1, \frac{f(\boldsymbol{\Psi}'|\boldsymbol{X})}{f(\boldsymbol{\Psi}|\boldsymbol{X})} \times \frac{f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})}\right) \\ &= \min\left(1, \frac{f(\boldsymbol{X}|\boldsymbol{\Psi}')f(\boldsymbol{\Psi}')/f(\boldsymbol{X})}{f(\boldsymbol{X}|\boldsymbol{\Psi})f(\boldsymbol{\Psi})/f(\boldsymbol{X})} \times \frac{f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})}\right) \\ &= \min\left(1, \underbrace{\frac{f(\boldsymbol{X}|\boldsymbol{\Psi}')}{f(\boldsymbol{X}|\boldsymbol{\Psi})}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{f(\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi})}}_{\text{Prior Ratio}} \times \underbrace{\frac{f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})}}_{\text{Proposal Ratio}}\right) \end{aligned}$$

A uniform random variable between 0 and 1 is drawn. If this number is less than R, then the proposed state is accepted and $\boldsymbol{\Psi} = \boldsymbol{\Psi}'$. Otherwise, the chain remains in the original state. This process of proposing a new state, calculating the acceptance probability, and either accepting or rejecting the proposed move is repeated many thousands of times. The sequence of states visited forms a Markov chain. This chain is sampled (either every step, or the chain is "thinned" and samples are taken every so often). The samples from the Markov chain form a valid, albeit dependent, sample from the posterior probability distribution *(20)*. As described here, the Markov chain samples from the joint probability density of trees, branch lengths, and substitution parameters. The marginal probability of trees can be calculated by simply printing to a file the trees that are visited during the course of the MCMC analysis. The proportion of the time any single tree is found in this sample is an approximation of the posterior probability of the tree.

An example of Bayesian inference of phylogeny.—Here we will demonstrate Bayesian inference of phylogeny for a simple example of five species. The DNA sequences are *albumin* and *c-myc* sequences sampled from a fish, frog, bird, rodent, and human (*albumin:* Actinoptergyii, *Salmo salar*, X52397; Amphibia, *Xenopus laevis*, M18350; Aves, *Gallus gallus*, X60688; Rodentia, *Rattus norvegicus*, J00698; Primates, *Homo sapiens*, L00132; *c-myc:* Actinoptergyii, *Salmo gairdneri*, M13048; Amphibia, *Xenopus laevis*, M14455; Aves, *Gallus gallus*, M20006; Rodentia, *Rattus norvegicus*, Y00396; Primates, *Homo sapiens*, V00568). There are a total of $B(5) = 15$ unrooted trees possible for the five sequences. The prior probability of any single tree, then, is $\frac{1}{15} = 0.067$.

We first analyzed the *c-myc* DNA sequences using a program written by J.P.H. The HKY85+Γ model of DNA substitution was assumed *(5, 21)*. This model allows there to be a different rate of transitions and transversions, different stationary nucleotide frequencies, and among-site rate variation (as described by a discrete gamma distribution). The Markov chain was run for 100,000 generations and sampled every 100 generations. The first 10,000 generations of the chain were discarded; the chain was started from a random tree and branch lengths and it took some time for the chain to reach apparent stationarity. Hence, inferences were based upon a sample of 900 trees. Figure 3 summarizes the results of the analysis. The tree with the largest posterior probability was (Fish,Frog,(Bird,(Rodent,Human))) and the posterior probability of this tree was 0.964. Figure 3 shows the posterior probability of the clades on the tree with the maximum posterior probability.

One of the advantages of Bayesian inference of phylogeny is that the results are easy to interpret. For example, the sum of the posterior probabilities of all trees will sum to 1. Moreover, the posterior probability of any single clade is simply the sum of the posterior probabilities of all trees that contain that clade. Finally,

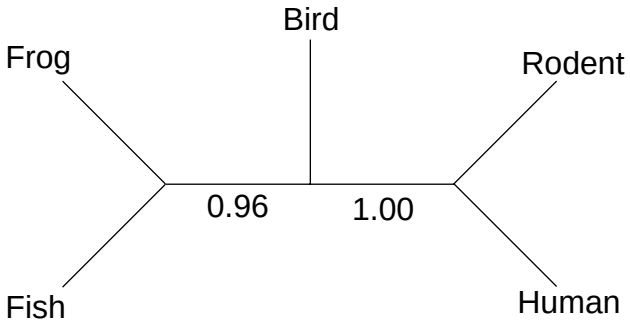

Figure 3.—The tree with the maximum posterior probability for the analysis of the c-myc sequences. The numbers at the internal branches represent the posterior probability that the clade is correct.

a credible set of trees can be formed by ordering all of the trees from largest to smallest posterior probability and then adding those trees with the highest posterior probability to a set until the cumulative posterior probability is 0.95. A 95% credible set of trees for the *c-myc* gene would contain only one tree.

The posterior probability of trees can form the prior for any subsequent analysis of the species. For example, let us imagine that the *albumin* sequences were analyzed after the *c-myc* sequences. The posterior probabilities of phylogenies from the analysis of the *c-myc* sequences is the prior for the analysis of the *albumin* sequences. The posterior probability of the trees after analysis of the *albumin* sequences is shown in the table. The posterior probability of τ_1 is now 0.996. Our beliefs about the phylogeny of the five species have changed throughout the analysis. For example, our initial belief about the the phylogeny τ_1 was 0.067. After observing the *c-myc* sequences, our belief that this is the true phylogeny increased from 0.067 to 0.964. The *albumin* sequences strengthened our beliefs about this phylogeny. The final posterior probability of this phylogeny was 0.996. This probability could form the prior probability for tree 1 for any subsequent analysis.

i	τ_i	$f(\tau_i)$	$f(\tau_i\|\text{c-myc})$	$f(\tau_i\|\text{albumin})$
1	(Fish,Frog,(Bird,(Rodent,Human)))	0.067	0.964	0.996
2	(Fish,Frog,(Rodent,(Bird,Human)))	0.067	0.000	0.000
3	(Fish,Frog,(Human,(Rodent,Bird)))	0.067	0.000	0.000
4	(Fish,Bird,(Frog,(Rodent,Human)))	0.067	0.012	0.003
5	(Fish,Bird,(Rodent,(Frog,Human)))	0.067	0.000	0.000
6	(Fish,Bird,(Human,(Rodent,Frog)))	0.067	0.000	0.000
7	(Fish,Rodent,(Bird,(Frog,Human)))	0.067	0.000	0.000
8	(Fish,Rodent,(Frog,(Bird,Human)))	0.067	0.000	0.000
9	(Fish,Rodent,(Human,(Bird,Frog)))	0.067	0.000	0.000
10	(Fish,Human,(Bird,(Rodent,Frog)))	0.067	0.000	0.000
11	(Fish,Human,(Frog,(Rodent,Bird)))	0.067	0.000	0.000
12	(Fish,Human,(Rodent,(Frog,Bird)))	0.067	0.000	0.000
13	(Frog,Bird,(Fish,(Rodent,Human)))	0.067	0.023	0.001
14	(Frog,Human,(Fish,(Rodent,Bird)))	0.067	0.000	0.000
15	(Frog,Rodent,(Fish,(Bird,Human)))	0.067	0.000	0.000

One modification of the analysis of the vertebrate sequences would be to modify the prior probabilities of trees. There is overwhelming morphological and paleontological evidence that the correct phylogeny for fish, frogs, birds, rodents, and humans is tree τ_1. Hence, a systematist might reflect this prior information as a different prior probability on the trees. For example, he or she may decide to put almost all of the prior probability on τ_1 and very little prior probability on the other trees.

Programs for Bayesian inference of phylogeny.—There are a few programs for the Bayesian analysis

of phylogenetic trees. BAMBE *(22)* approximates the posterior probability of phylogenies using MCMC (specifically, BAMBE uses the MHG algorithm). BAMBE assumes uniform priors on phylogenies and branch lengths. The program uses an improved method for calculating likelihoods that is very fast. Another program, MCMCTREE in the PAML package of programs *(13)* calculates posterior probabilities of trees using a combination of Monte Carlo and MCMC integration. The program works for up to $s = 11$ species. Besides the algorithm for approximating posterior probabilities, the program differs from BAMBE in assuming a birth-death process prior on phylogenies. This prior places equal weight on labelled histories (where a labelled history differs from a rooted tree in considering the relative speciation times).

References and Notes

1. T. Bayes, *Phil. Trans. Roy. Soc.*, 330 (1763).
2. S. Tavaré, *Lectures in Mathematics in the Life Sciences*, **17**, 57 (1986).
3. J. Felsenstein, J, *J. Mol. Evol.*, **17**, 368 (1981).
4. Z. Yang, *Mol. Biol. Evol.* **10**, 1396 (1993).
5. Z. Yang, *J. Mol. Evol.* **39**, 306 (1994).
6. N. Goldman and Z. Yang, *Mol. Biol. Evol.* **11**, 725 (1994).
7. S. Muse and B. Gaut, *Mol. Biol. Evol.* **11**, 715 (1994).
8. M. Schöniger and A. von Haeseler, *Mol. Phyl. Evol.* **3**, 240 (1994).
9. J. Adachi and M. Hasegawa, MOLPHY: Programs for molecular Phylogenetics I—PROTML: Maximum likelihood inference of protein phylogeny. *Comp. Sci. Mono.* **27** (1992).
10. P. E. Smouse and W.-H. Li, *Evol.* **41**, 1162 (1987).
11. P. O. Lewis, Talk at the Annual meeting of the SSE, SSB, and ASN societies held in Madison, Wisconsin (1999).
12. D. L. Swofford, *PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods.* Sinauer Associates, Sunderland, MA (1998).
13. Z. Yang, *CABIOS* **15**, 555 (1997).
14. J. Felsenstein, PHYLIP (Phylogeny Inference Package, version 3.5c, Dept. Genet., Univ. Wash.
15. N. Metropolis, A. W. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
16. W. Hastings, *Biometrika* **57**, 97 (1970).
17. P. J. Green, *Biometrika* **82**, *711* (1995).
18. Z. Yang and B. Rannala, *Mol. Biol. Evol.* **14**, 717 (1997).
19. B. Larget and D. Simon, *Mol. Biol. Evol.* **16**, 750 (1999).
20. L. Tierney, *Annuals of Statistics* **22**, 1701 (1994).
21. M. H. Hasegawa, H. Kishino, T. Yano *J. Mol. Evol.* **22**, 160 (1985).
22. D. Simon and B. Larget, *Bayesian Analysis in Molecular Biology, BAMBE* (Dept. Math. Comp. Sci., Duquesne, Univ, 1999).