Syst. Biol. 52(5):649–664, 2003
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150390238879

Hierarchical Phylogenetic Models for Analyzing Multipartite Sequence Data

Marc A. Suchard,[1] Christina M. R. Kitchen,[2] Janet S. Sinsheimer,[1,2,3] and Robert E. Weiss[2]

[1] *Department of Biomathematics, David Geffen School of Medicine, University of California–Los Angeles, Los Angeles, California 90095-1766, USA; E-mail: msuchard@ucla.edu*
[2] *Department of Biostatistics, School of Public Health, University of California–Los Angeles, Los Angeles, California 90095-1772, USA*
[3] *Department of Human Genetics, David Geffen School of Medicine, University of California–Los Angeles, Los Angeles, California 90095-1766, USA*

Abstract.—Debate exists over how to incorporate information from multipartite sequence data in phylogenetic analyses. Strict combined-data approaches argue for concatenation of all partitions and estimation of one evolutionary history, maximizing the explanatory power of the data. Consensus/independence approaches endorse a two-step procedure where partitions are analyzed independently and then a consensus is determined from the multiple results. Mixtures across the model space of a strict combined-data approach and a priori independent parameters are popular methods to integrate these methods. We propose an alternative middle ground by constructing a Bayesian hierarchical phylogenetic model. Our hierarchical framework enables researchers to pool information across data partitions to improve estimate precision in individual partitions while permitting estimation and testing of tendencies in across-partition quantities. Such across-partition quantities include the distribution from which individual topologies relating the sequences within a partition are drawn. We propose standard hierarchical priors on continuous evolutionary parameters across partitions, while the structure on topologies varies depending on the research problem. We illustrate our model with three examples. We first explore the evolutionary history of the guinea pig (*Cavia porcellus*) using alignments of 13 mitochondrial genes. The hierarchical model returns substantially more precise continuous parameter estimates than an independent parameter approach without losing the salient features of the data. Second, we analyze the frequency of horizontal gene transfer using 50 prokaryotic genes. We assume an unknown species-level topology and allow individual gene topologies to differ from this with a small estimable probability. Simultaneously inferring the species and individual gene topologies returns a transfer frequency of 17%. We also examine HIV sequences longitudinally sampled from HIV+ patients. We ask whether posttreatment development of CCR5 coreceptor virus represents concerted evolution from middisease CXCR4 virus or reemergence of initial infecting CCR5 virus. The hierarchical model pools partitions from multiple unrelated patients by assuming that the topology for each patient is drawn from a multinomial distribution with unknown probabilities. Preliminary results suggest evolution and not reemergence. [Bayes factor; *Cavia*; CXCR4/CCR5 coreceptor; HIV evolution; horizontal gene transfer; MCMC; phylogeny.]

Efforts to sequence an ever-growing number of genomes are speeding ahead. These advances are producing larger data sets increasingly labeled with definable substructures. A standard example of these substructures includes the multiple genes found within sequences from a set of taxa, defining natural partitions of the data. As a result, phylogeneticists are confronted with a dilemma over how to incorporate information about these multiple data partitions in their analyses.

A long-standing debate rages over this issue (Bull et al., 1993) with at least two opposing solutions. One solution, a strict combined-data approach, also called the total evidence approach by Kluge (1989), combines multiple partitions into a single undifferentiated partition (Miyamoto, 1985) before phylogenetic analysis. A strict combined-data approach ignores the partition information, pools the data into a single sample, and infers a single evolutionary history. This approach implies that given the evolutionary history for the concatenated data, results from the individual partitions are irrelevant.

The competing consensus/independence approach takes a two stage approach (Penny and Hendy, 1986; Miyamoto and Fitch, 1995). In the first stage of analysis, a phylogenetic model is independently fit to each partition and used to estimate separate evolutionary histories for each partition. The second stage forms a consensus from the resulting topologies.

An apparent advantage of a strict combined-data approach is that parameter estimates and inference regarding the single evolutionary history are more robust than those from the individual partitions, as the individual partitions are more subject to the effects of sampling variation in their potentially sparse data. However, separate analyses yield separate insights into the histories of individual partitions. Individual partitions may reconstruct different topologies, suggesting nonorthology problems, gene conversion, or recombination. Also, individual partitions may evolve at substantially different rates or under different pressures, causing the model of evolution for one or more partitions to vary from that in the remaining partitions. Allowing for this variability is important, for assuming inappropriate models may lead to inconsistent or biased inference (Yang, 1995a; Buckley et al., 2001; Dorman et al., 2002).

Still, the consensus approach also possesses difficulties. When attempting to draw inference from across a number of different partitions, simple consensus measures can fail, particularly when the evolutionary reconstruction model provides only the most likely topological relationship with no further estimates of uncertainty. As a contrived example, consider the case of three partitions with phylogenetic data for the same four taxa. Suppose that support for topology A over topologies B or C is only marginally higher in two partitions while the last partition shows close to 100% support for B. The consensus estimate is A, while a method that can also accommodate the uncertainty across partitions should yield greater support for B.

To overcome some of the difficulties inherent in choosing a strict combined-data approach versus consensus framework for multipartite data, many phylogenetic analyses employ a mixture of these approaches by dividing the parameter space within partitions into two sets (Yang, 1996). In one set, parameters are fixed equal (or proportional) across partitions, following the strict combined-data paradigm; in the remaining set, parameters remain a priori independent from partition to partition, following the first stage in a consensus approach. Popular phylogenetic software, such as Bambe, MrBayes, PAML, PAUP*, and Phylip, allow for these analyses (Felsenstein, 1993; Yang, 1996; Larget and Simon, 1999; Huelsenbeck and Ronquist, 2001; Swofford, 2003). In general, the models implemented in these programs fix the topologies across partitions to be equal, side-stepping the need to take a consensus over varying topologies, while evolutionary pressure parameters, e.g., transition:transversion ratios and evolutionary rates, vary independently across partitions. However, alternative models that constrain equal (or proportional) rates and pressures across partitions are also commonly implemented (Newton et al., 1999).

Bayesian hierarchical models can be adapted to provide another alternative in analyzing multiple partitions. Bayesian methods naturally handle averaging across uncertain discrete quantities such as topologies across partitions and are gaining popularity in phylogenetics (Sinsheimer et al., 1996; Yang and Rannala, 1997; Larget and Simon, 1999; Mau et al., 1999; Li et al., 2000; Huelsenbeck et al., 2001; Suchard et al., 2001). Bayesian approaches focus on the estimation of the posterior distribution of unknown model parameters given the observed data via Bayes theorem. Let the parameters be θ and data be Y. Bayes theorem states that the posterior distribution $p(\theta \mid Y)$ of θ given Y is proportional to the product of the sampling density $f(Y \mid \theta)$ of Y given θ, referred to as the model likelihood, and the prior distribution $q(\theta)$ of θ. Specifically,

$$p(\theta \mid Y) = \frac{f(Y \mid \theta) q(\theta)}{m(Y)}, \tag{1}$$

where the constant of proportionality $m(Y) = \int_\theta f(Y|\theta) q(\theta)\, d\theta$ is the marginal likelihood of the data Y.

Buckley et al. (2002) extended the first level of a consensus approach into a Bayesian framework to examine the congruence of topologies across multiple partitions. Like previous consensus-based analyses, inference regarding tendencies across the partitions relies on an ad hoc two-step approach in which all data are not used simultaneously. The first step estimates individual partition parameters independently, and then the second step either averages these values or compares them with a fixed point. Further, to the best of our knowledge, no previous methods share information about evolutionary pressure parameters without assuming absolute equality or proportionality across partitions.

In this article, using a Bayesian hierarchical framework, we define a middle ground between a strict combined-data approach and a consensus approach that combines the strengths of both. We use all of the data in a single analysis, akin to a strict combined-data approach. At the same time, we allow for different phylogenetic parameters in the individual partitions, as in the consensus approach. Our framework includes a formal statistical model that combines the results from the individual partitions to provide overall or across-partition-level summaries of all evolutionary parameters, including topologies for the entire data set. This across-partition-level model and the individual partition models are fit simultaneously, enabling the across-partition-level model to feed back information, in the form of a prior, into the estimation for the individual partitions. The feedback results in a borrowing of strength of information from one partition by another, producing more precise partition-level estimates. These hierarchical relations are found in many statistical frameworks (Laird and Ware, 1982; Gelman et al., 1995).

The Bayesian model described in Equation 1 is not inherently hierarchical. To define a hierarchical structure in the model, we start with the natural divide of the multipartite sequence data $Y = (Y_1, \ldots, Y_K)$ into K separate partitions with K copies of the model parameters, $\theta = (\theta_1, \ldots, \theta_K)$. This yields

$$p(\theta \mid Y) \propto f(Y \mid \theta) q(\theta) = \left(\prod_{k=1}^{K} f(Y_k \mid \theta_k) \right) q(\theta_1, \ldots, \theta_K). \tag{2}$$

Key to the hierarchical construction is modeling the prior $q(\theta_1, \ldots, \theta_K)$ such that it depends on unknown, but estimable, parameters ϕ in which θ_k are only conditionally independent given ϕ, suggesting

$$q(\theta_1, \ldots, \theta_K) = \int_\phi \prod_{k=1}^{K} q(\theta_k \mid \phi) q(\phi)\, d\phi. \tag{3}$$

Employing unknown hyperparameters ϕ that, in turn, have their own prior $q(\phi)$ enables the borrowing of strength of information from Y_k through θ_k and ϕ to the remaining $K - 1$ partitions and their respective parameters. The hierarchical model becomes

$$p(\theta, \phi \mid Y) \propto \left(\prod_{k=1}^{K} f(Y_k \mid \theta_k) q(\theta_k \mid \phi) \right) q(\phi). \tag{4}$$

In contrast, independent analyses or mixed models combining both constrained and a priori independent parameters assume that ϕ is a fixed constant across the independent portions of the parameter space. As a result, no sharing of information is possible and so these models are not hierarchical. Further, constraining $\theta_1 = \cdots = \theta_K$

is also not a hierarchical model as $q(\theta_k \mid \phi)$ becomes a point-mass on $\theta_k = \phi$ and all variability is lost.

In a phylogenetic setting, hierarchical models serve as a generalization of discrete site class models (Hasegawa et al., 1993; Yang, 1995b), where individual partitions possess independent evolutionary parameters, and the mixture models that combine constrained equal or independent parameters across partitions. Exploiting the hierarchical dependency between parameters allows the generalized model to pool information across partitions to improve the precision of estimates in individual partitions and conveniently enables estimating and testing tendencies in topologies across partitions while allowing the evolutionary parameters in individual partitions to vary.

We outline the remainder of this paper as follows. In the next two sections, first we review models for the reconstruction of evolutionary histories from a single partition, then formally introduce a hierarchical phylogenetic model to incorporate data from multiple partitions and finally describe methods to sample from the Bayesian model. In the *Examples* section, we illustrate the utility of our hierarchical phylogenetic model in three disparate problems. The first example uses data from the recent debate regarding the classification of guinea pigs as rodents, where the 13 mitochondrial genes from four taxa define 13 data partitions. This example employs the hierarchical model in a data set typical of the strict combined-data versus consensus/independence approach debate. The second example explores the frequency of horizontal gene transfer (HGT) among four prokaryotic species. Individual genes define the separate partitions, and we simultaneously estimate the most likely species topology along with all gene topologies. The last example examines the intrahost evolution of HIV and utilizes the hierarchical model to draw conclusions across multiple hosts simultaneously. Here, individual data partitions comprise independent hosts, and the taxa within partitions represent HIV sequences collected at corresponding time points. Strict combined-data or independent analyses approaches provide inadequate frameworks to broach many phylogenetic problems like these latter two. We conclude the article with a brief discussion in the REMARKS section.

EVOLUTIONARY RECONSTRUCTION WITHIN A PARTITION

We begin with aligned molecular sequence data Y, in particular DNA or RNA sequences. Data $Y = (Y_1, \ldots, Y_K)$ consist of K disjoint sets, called partitions. Data $Y_k = (Y_{k1}, \ldots, Y_{kC})$ within partition k for $k = 1, \ldots, K$ consist of aligned sequences from N equivalent taxa and can be further subdivided into C evolutionary site classes. Within class c for $c = 1, \ldots, C$, Y_{kc} represent L_{kc} aligned sites, such that $Y_{kc} = (Y_{kc1}, \ldots, Y_{kcL_{kc}})$. Site data $Y_{kcl} = (Y_{kcl1}, \ldots, Y_{kclN})^t$ contain one nucleotide from each taxon, such that $Y_{kcli} \in$ (A, G, C, T) for $i = 1, \ldots, N$, where A stands for adenosine, G for guanine, C for cytosine, and T for thymidine (or U for uracil in RNA sequences). Rather than deleting sites that contain alignment indels or ambiguous nucleotides, we integrate over all their possible values, where an indel can be A, G, C, or T (Felsenstein, 1981).

Many Bayesian evolutionary reconstruction methods (e.g., Mau and Newton, 1997; Yang and Rannala, 1997; Larget and Simon, 1999; Suchard et al., 2001) assume that sites are independent and identically distributed (iid) within partitions and site classes, and the likelihood of observing Y_{kcl} is given by a multinomial distribution over the 4^N possible outcomes. The multinomial probabilities are functions of an unknown topology τ_k relating the N taxa within partition k and branch lengths $t_{kc} = (t_{kc1}, \ldots, t_{kcS})$, where $s = 1, \ldots, S$ and $S = 2N - 3$, and a model to describe the mutation of nucleotides along these branch lengths within partition k and site class c. Popular models include the continuous-time Markov chain (CTMC) model for nucleotide substitution (Felsenstein, 1981). CTMC models assume that the substitution mechanism is independent across branches and follows a memoryless process, with the probability of nucleotide X mutating to Z along a branch with length t_{kcs} is equal to $\exp(t_{kcs}\Lambda_{kc})$, where Λ_{kc} is a 4×4 infinitesimal rate matrix. We use a parameterization of Λ_{kc} similar to that of Tamura and Nei (1993, TN93).

$$\Lambda_{kc}^{\mathrm{TN93}} = f_{kc} \begin{pmatrix} - & \alpha_{kc}\pi_{kcG} & \pi_{kcC} & \pi_{kcT} \\ \alpha_{kc}\pi_{kcA} & - & \pi_{kcC} & \pi_{kcT} \\ \pi_{kcA} & \pi_{kcG} & - & \gamma_{kc}\pi_{kcT} \\ \pi_{kcA} & \pi_{kcG} & \gamma_{kc}\pi_{kcC} & - \end{pmatrix}, \qquad (5)$$

where the minus sign in each row represents minus the sum of the remaining elements in that row. Parameter f_{kc} is the transversion rate, α_{kc} is the transition:transversion rate ratio for transitions between the purines A and G, and γ_{kc} is the transition:transversion rate ratio for transitions between the pyrimidines C and T in partition k and site class c. The vector $\pi_{kc} = (\pi_{kcA}, \pi_{kcG}, \pi_{kcC}, \pi_{kcT})$ is the stationary distribution of the Markov chain generated by Λ_{kc}. Only the product $t_{kcs} \times \Lambda_{kc}$ enters into the model likelihood, so without loss of generality, we fix

$$f_{kc} = \frac{1}{2[\alpha_{kc}\pi_{kcA}\pi_{kcG} + \gamma_{kc}\pi_{kcC}\pi_{kcT} + (\pi_{kcA} + \pi_{kcG})(\pi_{kcC} + \pi_{kcT})]}. \qquad (6)$$

This constraint enforces

$$\sum_{m \in (A,G,C,T)} \pi_{kcm}(\Lambda_{kc})_{m,m} = -1, \qquad (7)$$

such that each branch length is the expected number of nucleotide substitutions per site between the two nodes that the branch connects (Yang et al., 1994).

Not all branch lengths retain definition between topologies, and topologies may change across partitions.

To overcome this difficulty when attempting to pool branch length information across partitions, we first model branch lengths t_{kcs} within partition k and site class c as exponentially distributed

$$t_{kcs} \sim \text{Exponential}(\mu_{kc}). \tag{8}$$

The parameter μ_{kc} is the prior expected divergence between nodes within partition k and site class c, retains definition across topologies, and enables us to conveniently share branch length information across partitions. A conditionally Exponential(μ_{kc}) random variable has density $p(t_{kcs} \mid \mu_{kc}) = 1/\mu_{kc} \times \exp(-t_{kcs}/\mu_{kc})$, has mode 0, has nonzero density on the entire positive real line, and has been used previously in Bayesian phylogenetics to model branch lengths (Suchard et al., 2001, 2003a). Given its mode at 0, we suspect the conditional prior to bias toward a starlike topology and hence would be conservative. In our hands, a conditionally Exponential prior has not had a noticeable impact on branch length estimation for up to 15 taxa (Suchard et al., 2001). Further, the Exponential prior is motivated by a Yule process of neutral evolution and is vague but remains proper. The usual Jeffreys' prior on $t_{kcs} \in [0, \infty)$ is $1/t_{kcs}$ (Jeffreys, 1998). This prior choice is not integrable and may preclude calculating Bayes factors for hypotheses of interest.

Hierarchical Priors for Combining Partition Models

We build a Bayesian hierarchical framework that allows us to unite the separate partition models in a single comprehensive model. The second level (across-partition level) of our model functions as a prior for the partition-level parameters. Hyperparameters of this second level are themselves unknown parameters, and this allows information about the values of the parameters in one set of partitions to be transferred to help in estimating the parameters in other partitions. This transfer is particularly useful if the partition-level parameter is poorly estimated, perhaps due to small L_{kc} or otherwise uninformative data. Further, estimation of the higher level parameters reveals tendencies across partitions. In the language of classical statistics, across-partition-level parameters represent fixed effects, while partition-level parameters represent random effects. The random effects describe how the partition-level parameters vary from the higher across-partition-level parameter means. Different hierarchical priors are required for continuous and discrete parameters. The next two subsections illustrate how we model these priors.

Continuous Evolutionary Parameters

Transition:transversion rate ratios α_{kc} and γ_{kc} and expected divergence μ_{kc} exist on the positive half of the real line. The natural transform for ratios onto the entire real line is the logarithmic transformation, placing equal rates at zero. To maintain consistency and allow for future multivariate analysis, we also transform the expected divergence. Given transformed parameters spanning the entire real line, we follow common practice (Gelman et al., 1995) and model the corresponding parameters across partitions using a multivariate normal prior,

$$\begin{pmatrix} \log \alpha_{kc} \\ \log \gamma_{kc} \\ \log \mu_{kc} \end{pmatrix} \sim \text{Normal}(V_c, \Sigma), \tag{9}$$

with log-scale, across-partition-level unknown mean vector $V_c = (A_c, G_c, M_c)^t$ for each site class c and unknown variance–covariance matrix Σ. In this article, we assume a shared Σ across site classes, due to a lack of prior knowledge about its variability, with a simple diagonal form, such that $\Sigma = \text{diag}\,(\sigma^2_\alpha, \sigma^2_\gamma, \sigma^2_\mu)$. To ease computation by allowing direct Gibbs sampling of V_c and Σ, we specify conjugate hyperpriors

$$V_c \sim \text{Normal}(\Psi_{V,1}, \Psi_{V,2})$$

and

$$\frac{1}{\sigma^2_x} \sim \text{Gamma}(\psi_{\sigma^2,1}, \psi_{\sigma^2,2}) \tag{10}$$

for $x \in (\alpha, \gamma, \mu)$. We choose relatively uninformative priors by setting $\Psi_{V,1} = \psi_{V,1} \times (1, 1, 1)^t$ for $\psi_{V,1} = 0$, $\Psi_{V,2} = \psi_{V,2} \times I$ for $\psi_{V,2} = 10$, where I is the identity matrix, $\psi_{\sigma^2,1} = 2.1$ and $\psi_{\sigma^2,2} = 1.1$. Under these choices, σ^2_α, σ^2_γ, and σ^2_μ have prior expectation 1 and variance 10. In situations where K or C are large and one is interested in estimating the posterior covariation between α_{kc}, γ_{kc}, and μ_{kc}, less structured forms of Σ_c specified for each site class c are possible, and the conjugate hyperprior becomes the Wishart distribution over Σ_c.

Stationary distributions π_{kc} exist on the simplex in $\Re^4$ and are naturally modeled by a Dirichlet distribution,

$$\pi_{kc} \sim \text{Dirichlet}(N_\Pi \times \Pi_c), \tag{11}$$

where $\Pi_c = (\Pi_{cA}, \Pi_{cG}, \Pi_{cC}, \Pi_{cT})$ are the across-partition-level proportions for each nucleotide type for site class c and N_Π is a pseudocount measure of precision across π_{kc}. For hyperpriors on Π_c and N_Π, we assume

$$\Pi_c \sim \text{Dirichlet}(\phi_\Pi)$$

and

$$N_\Pi \sim \text{Gamma}(\psi_{N,1}, \psi_{N,2}) \tag{12}$$

and set $\phi_\Pi = (1, 1, 1, 1)$, providing a flat prior on Π_c, $\psi_{N,1} = 0.1$ and $\psi_{N,2} = 0.1$, providing a proper yet vague prior on N_Π.

In most cases, the number of site classes $C \leq 3$ and as such we have chosen to keep across-partition-level parameters V_c and Π_c a priori independent and to use shared measures of variability Σ and N_Π across site classes. When C is large, further borrowing strength can be obtained by specifying a third hierarchical level across classes. One accomplishes this by changing constants $\Psi_{V,1}$, $\Psi_{V,2}$, and ϕ_Π into unknowns, specifying prior distributions over them, and relaxing the shared variability assumption.

In interpreting the hierarchical distributions, $x \sim \text{Normal}(y, z)$ has density $p(x) = (2\pi|z|)^{-1/2} \exp[-(1/2)(x-y)^t z^{-1}(x-y)]$, $x \sim \text{Gamma}(y, z)$ has density $p(x) \propto x^{y-1}\exp(-zx)$, and $(x_1, \ldots, x_I) \sim \text{Dirichlet}(y_1, \ldots, y_I)$ has density $p(x_1, \ldots, x_I) \propto x_1^{y_1-1} \cdots x_I^{y_I-1}$, where I counts the number of components in $(x_1, \ldots, x_I)$.

Discrete Evolutionary Parameters

Whereas standard hierarchical distributions exist for continuous parameters, hierarchical models for discrete parameters, like the topologies τ_k across partitions in our case, present more of a challenge. We develop three options here.

Completely linked partitions model.—In many phylogenetic problems, the separate partitions are believed to have experienced equivalent evolutionary histories and should yield the same common topology Υ. This situation presents itself, for example, when the partitions represent various genes sampled across the same sets of organisms and recombination, HGT, or lineage sorting (Robertson et al., 1995; Jain et al., 2002; Suchard et al., 2003b) are highly unlikely a priori. We refer to these partitions as completely linked, following the idea that the genes from the different partitions may be linked together on shared chromosomes or at least across evolutionary time, although we need not assume such a rigorous relationship between partitions. Under the completely linked assumption, we constrain $\Upsilon = \tau_1 = \cdots = \tau_K$ and assume that

$$\Upsilon \sim \text{Multinomial}(Q), \tag{13}$$

where $Q = (Q_1, \ldots, Q_E)$ are constants, the prior probabilities of the $E = (2N-5)!/2^{N-3}(N-3)!$ possible N-taxon topologies. When little or no information is known about Υ, a reasonable choice is that a priori all possible common topologies are equally likely, such that $p(\Upsilon) = 1/E$. Alternatively, in a hypothesis testing setting, one may choose Q such that the prior odds of the dueling hypotheses regarding Υ are 1.

The completely linked partitions model is a direct extension of the mixed models under which topologies are taken to be equal and evolutionary pressure parameters vary independently across partitions. Although our prior over τ_k is not hierarchical here, our prior over the remaining parameters is hierarchical, enabling more precise parameter estimation through the borrowing of strength than mixed models afford.

Partially linked partitions model.—Situations where infrequent recombination, HGT, or lineage sorting may occur warrant a full hierarchical prior. Our partially linked partitions model speculates, as above, that there exists an estimable common topology Υ for all partitions. In contrast to the completely linked model, each partition topology τ_k may differ from Υ with an estimable probability p. If τ_k is incongruent with Υ, then τ_k is equally likely to be any of the remaining $E - 1$ possible topologies. The multinomial hierarchical prior on τ_k becomes

$$q(\tau_k \mid \Upsilon) = \begin{cases} 1 - p & \text{if } \tau_k = \Upsilon \\ p \times \dfrac{1}{E-1} & \text{otherwise.} \end{cases} \tag{14}$$

Following the completely linked partitions model, we assume a multinomial prior for Υ with prior probabilities Q. To maintain identifiability between Υ and p, we restrict $0 \leq p \leq (E-1)/E$ by assuming

$$p \sim \text{Beta}(\psi_{p,1}, \psi_{p,2}) \times 1\left\{p \leq \frac{E-1}{E}\right\}, \tag{15}$$

where $\psi_{p,1}$ and $\psi_{p,2}$ specify our prior information regarding p and $1\{\cdot\}$ is the indicator function. An advantage of this hierarchical structure is that we allow τ_k to vary and, hence, can simultaneously estimate the posterior probabilities of incongruence between partitions while inferring Υ.

One caveat in interpreting incongruent topologies is that such incongruence can also result from the estimation process itself. These mechanisms include stochastic error due to sparse data, evolutionary model misspecification, and parallel/convergent evolution (Cao et al., 1998). As a consequence, probability p becomes an upper bound on the frequency of the underlying biological process producing incongruent topologies. On the other hand, the ability of the partially linked partitions model to identify model misspecification when no recombination, HGT, or lineage sorting is suspected may be capitalized on as a diagnostic tool.

Unlinked partitions model.—In contrast to both linked partition problems, there also exist phylogenetic questions focused on the average behavior of unlinked partitions. Unlinked partitions arise when the data within one partition result from evolutionary events that are independent from those producing the data in another partition. For example, consider the evolutionary histories of independent populations of the same organism and let each partition contain sampled sequences from corresponding phenotypes or time points in the history of one population. In this situation, the researcher may be interested in identifying the occurrence of similar evolutionary events across population partitions or estimating the most probable sequence of events using all partitions simultaneously.

Given the independence of unlinked partitions, our model for them is simpler than those for their linked counterparts. We posit that

$$\tau_k \sim \text{Multinomial}(T), \tag{16}$$

where $T = (T_1, T_2, \ldots, T_E)$ are the unknown across-partition-level probabilities for each of the E possible topologies. To complete the hierarchical prior, we assume

$$T \sim \text{Dirichlet}(N_Q \times Q), \tag{17}$$

where constant prior probabilities Q retain their earlier definition and specification suggestions and N_Q is the number of pseudocounts or weight given to the prior probabilities Q. This prior imparts as much information to the posterior of T as N_Q additional partitions with topologies drawn directly from Multinomial (Q). For large N_Q, the posterior of T approaches a point mass at Q. On the other end of the spectrum, $N_Q \rightarrow 0$ results in an uninformative prior but, in the limit, does not guarantee that the posterior of T remains proper. We recommend moderate choices of N_Q to avoid these extremes.

Posterior Model Sampler

For each partition, let $\theta_k = (\tau_k, t_{k1}, \ldots, t_{kC}, \alpha_{k1}, \ldots, \alpha_{kC}, \gamma_{k1}, \ldots, \gamma_{kC}, \mu_{k1}, \ldots, \mu_{kC}, \pi_{k1}, \ldots, \pi_{kC})$. Also, let $\theta = (\theta_1, \ldots, \theta_K)$ and $\phi = (V_1, \ldots, V_C, \Sigma, \Pi_1, \ldots, \Pi_C, N_\Pi, \Upsilon, p)$ for linked partitions or $\phi = (V_1, \ldots, V_C, \Sigma, \Pi_1, \ldots, \Pi_C, N_\Pi, T)$ for unlinked partitions, then the parameter pair (θ, ϕ) specifies the complete model. Having formalized our hierarchical phylogenetic model with parameters (θ, ϕ) in the previous sections, we now discuss a Markov chain Monte Carlo (MCMC) approach to sample from the model's joint posterior distribution $p(\theta, \phi \,|\, Y)$. MCMC has been used extensively to sample from nonhierarchical partition models (Larget and Simon, 1999; Mau et al., 1999; Li et al., 2000; Huelsenbeck and Ronquist, 2001; Suchard et al., 2001) parameterized in terms of just θ. Most of these samplers utilize a Metropolis-within-Gibbs (Tierney, 1994) algorithm that cycles through blocks of parameters within θ_k, updating them via a Metropolis–Hastings proposal (Metropolis et al., 1953; Hastings, 1970) conditional on the current values of the remaining parameters.

We construct our hierarchical sampler out of two nested Metropolis-within-Gibbs cycles. The outer cycle first iterates over partitions $k = 1, \ldots, K$ and then over the across-partition-level parameters ϕ. Within each partition k, the inner cycle proceeds over parameters θ_k, following the proposals of Suchard et al. (2001) with minor exception. The within-partition update blocks are

$$\begin{gathered}
\tau_k, t_{kc} \mid \alpha_{kc}, \gamma_{kc}, \mu_{kc}, \pi_{kc}, \phi, Y \\
t_{kc} \mid \tau_k, \alpha_{kc}, \gamma_{kc}, \mu_{kc}, \pi_{kc}, \phi, Y \\
\alpha_{kc}, \gamma_{kc} \mid \tau_k, t_{kc}, \mu_{kc}, \pi_{kc}, \phi, Y \\
\mu_{kc} \mid \tau_k, t_{kc}, \alpha_{kc}, \gamma_{kc}, \pi_{kc}, \phi, Y \\
\pi_{kc} \mid \tau_k, t_{kc}, \alpha_{kc}, \gamma_{kc}, \mu_{kc}, \phi, Y.
\end{gathered} \tag{18}$$

Suchard et al. (2001) restrict Trace$(\Lambda_{kc}) = -1$, constraining $(\alpha_{kc}, \gamma_{kc}) \in \text{U}[0, 1) \times \text{U}[0, 1)$, and propose new values for α_{kc} and γ_{kc} by generating Normal random variates centered at their current values and reflected about both zero and 1. In our current formulation $(\alpha_{kc}, \gamma_{kc}) \in [0, \infty) \times [0, \infty)$. As a consequence, we modify the proposal step such that it reflects only about zero. For the completely linked partitions model, the first update block above is skipped, as $\tau_k = \Upsilon$ for all k.

In order to sample the across-partition-level parameters ϕ in the outer Metropolis-within-Gibbs cycle, we must either derive their full conditional distributions to use Gibbs sampling or develop Metropolis–Hastings proposals for each parameter. We begin with V_c and Σ. By assuming a prior diagonal form for Σ and $\Psi_{V,2}$, the full conditional distributions of (A_c, σ_α^2), (G_c, σ_γ^2), and (M_c, σ_μ^2) are independent. Let $\alpha_c = (\alpha_{1c}, \ldots, \alpha_{Kc})$, then the full condition distributions of A_c and σ_α^2 are

$$A_c \,|\, \alpha_c, \sigma_\alpha^2 \sim \text{Normal}(\nu_A, \sigma_A^2) \;\text{ for } c = 1, \ldots, C$$

and

$$\frac{1}{\sigma_\alpha^2} \,|\, \alpha_c, A_c \sim \text{Gamma}\left(\psi_{\sigma^2,1} + \frac{KC}{2}, \psi_{\sigma^2,2} + \frac{1}{2}\sum_{c=1}^{C} SS_{Ac}\right),$$

where

$$\begin{aligned}
\nu_A &= \frac{\psi_{V,1}\sigma_\alpha^2 + \psi_{V,2}\sum_{k=1}^{K}\log\alpha_{kc}}{\sigma_\alpha^2 + K\psi_{V,2}}, \\
\sigma_A^2 &= \left(\frac{K}{\sigma_\alpha^2} + \frac{1}{\psi_{V,2}}\right)^{-1} \text{ and} \\
SS_{Ac} &= \sum_{k=1}^{K}(A_c - \log\alpha_{kc})^2.
\end{aligned} \tag{19}$$

Similar conditional distributions exist for (G_c, σ_γ^2) and (M_c, σ_μ^2) by modifying the indices in Equation 19.

The full conditional distributions of N_Π and Π_c are not of standard form. We sample these parameters using Metropolis–Hastings proposals. For N_Π, we propose new values by generating a Normal random variate centered at the current value of N_Π and reflected about zero to maintain positivity. For Π_c, we draw new values Π_{cm}^* from a Normal distribution centered at the current values Π_{cm} for $m \in (A, G, C)$, set $\Pi_{cT}^* = 1 - \Pi_{cA}^* - \Pi_{cG}^* - \Pi_{cC}^*$, and automatically reject any proposal where Π_c^* lies outside the simplex in $\Re^4$. These two Normal distributions have tunable variances. We adjust these variances such that proposals have acceptance rates of 30%–40% (Gelman et al., 1996).

For the completely linked partitions model, we sample Υ using the topology proposal of Suchard et al. (2001). For the partially linked partitions model, it is possible to directly sample Υ from its full conditional distribution when E is moderately small. Let $\tau = (\tau_1, \dots, \tau_K)$, $\Omega_{-\Upsilon}$ be the vector of all model parameters from (θ, ϕ) excluding Υ, and $\Omega_{-(\Upsilon,\tau)}$ be all parameters excluding both Υ and τ. Then, given τ, Υ is independent of the pair $(Y, \Omega_{-(\Upsilon,\tau)})$. As a consequence,

$$p(\Upsilon \mid Y, \Omega_{-\Upsilon}) = \frac{q(\tau \mid \Upsilon)q(\Upsilon)}{\sum_{\Upsilon=1}^{E} q(\tau \mid \Upsilon)q(\Upsilon)}. \tag{20}$$

Finally, given Υ, the K topologies τ_k are a priori iid, yielding

$$q(\tau \mid \Upsilon) = \prod_{k=1}^{K} q(\tau_k \mid \Upsilon). \tag{21}$$

We draw Υ from a Multinomial distribution with E state probabilities calculated using Equations 20 and 21. Probability p is updated via a Metropolis–Hastings proposal similar to N_Π with reflections about both 0 and $(E-1)/E$.

Under the unlinked partitions model, the Dirichlet prior on T is conjugate to the Multinomial distribution on τ_k. This conjugacy makes Gibbs sampling of T convenient, as its full conditional distribution remains Dirichlet,

$$T \mid \tau, N_Q, Q \sim \text{Dirichlet}(N_Q \times Q + C), \tag{22}$$

where $C = (C_1, \dots, C_E)$ with elements C_e counting the number of partitions in which the topology corresponding to e is observed,

$$C_e = \sum_{k=1}^{K} 1\{\tau_k = e\}. \tag{23}$$

We run our MCMC chains for 5.1×10^6 outer Metropolis-within-Gibbs cycles, discard the first 10^5 cycles as burn-in, and subsample every 5×10^2 cycles. This process retains $P = 10^4$ posterior samples with decreased autocorrelation. To help ensure convergence, these burn-in times and total chain lengths are significantly longer than what appears to be required by examining time-series plots of the log likelihood of each sample. Further, we compare the estimates obtained from the simulation of at least five independent chains with starting values drawn directly from the model priors to assess convergence.

Bayes Factors

Bayes factors measure the change in the support of the data in favor of one statistical model relative to another model. Formally, a Bayes factor (B_{10}) in favor of model 1 (M_1) over model 0 (M_0) is the ratio of the marginal likelihood $m(Y \mid M_1)$ of M_1 over the marginal likelihood $m(Y \mid M_0)$ of M_0 (Kass and Raftery, 1995):

$$B_{10} = \frac{m(Y \mid M_1)}{m(Y \mid M_0)}. \tag{24}$$

To calculate Bayes factors for nonnested models, it is frequently more convenient to estimate the posterior probabilities $p(M_0 \mid Y)$ and $p(M_1 \mid Y)$ by MCMC sampling over the joint space of the competing models rather than estimating the multidimensional integrals hidden in Equation 24 directly. Applying Bayes theorem to Equation 24 yields

$$B_{10} = \frac{p(M_1 \mid Y)}{p(M_0 \mid Y)} \Big/ \frac{q(M_1)}{q(M_0)} = \frac{\text{Posterior Odds}}{\text{Prior Odds}}, \tag{25}$$

where $q(M_0)$ and $q(M_1)$ are the prior probabilities of models M_0 and M_1.

Bayes factors are the Bayesian analogue of the likelihood ratio test (LRT). LRTs have been used effectively in phylogenetics (Huelsenbeck and Rannala, 1997) but can be remiss in that the data are sparse and the space of possible evolutionary topologies is discrete so that standard likelihood asymptotics may not apply (Goldman, 1993; Sinsheimer et al., 1996; Whelan and Goldman, 1999). Bayes factors have fewer difficulties in discrete spaces; probability mass functions naturally substitute for continuous distributions, and Bayes factors do not rely on large sample asymptotics. Huelsenbeck et al. (2001) and Suchard et al. (2001, 2003a) discussed further advantages of Bayes factors in phylogenetics, namely that tests of competing nucleotide substitution models and branch length restrictions do not require conditioning on known topologies.

More Efficient Estimators

To calculate Bayes factors comparing across-partition-level topological hypotheses regarding Υ and T, we estimate the posterior odds of the competing hypotheses and divide this value by the prior odds (Eq. 25). Three specific tests for nonnested hypotheses are illustrated in the examples. For the completely and partially linked partitions models, a standard estimator of the posterior probability $\lambda_e = p(\Upsilon = e \mid Y)$ for some topology $e \in (1, \dots, E)$ is the average of the indicator that $\Upsilon = e$ is true over all posterior draws. Using this naive estimator, the smallest nonzero estimate possible is $<1/P$, where P is the posterior sample length and no samples support $\Upsilon = e$. Further, if we approximate the posterior draws as independent, the relative SE of estimation of λ_e is approximately 100% when λ_e is small. This error grows substantially when one considers the dependence among the Markovian posterior samples.

For the completely linked partitions model, a more efficient estimator is possible by first conditioning on $\Upsilon = e$ and then calculating $p(Y \mid \Upsilon = e)$ using importance sampling integration (Newton and Raftery, 1994),

$$p(Y \mid \Upsilon = e) = \int_{\Omega_{-\Upsilon}} \frac{f(Y \mid \Omega_{-\Upsilon}, \Upsilon = e) q(\Omega_{-\Upsilon})}{g(\Omega_{-\Upsilon})} g(\Omega_{-\Upsilon})\, \mathrm{d}\Omega_{-\Upsilon}, \quad (26)$$

for each $e \in (1, \ldots, E)$. Importance sampling reduces the variance of Monte Carlo integration by utilizing an importance sampling function $g(\cdot)$ that places increased weight on random samples where $f(Y \mid \Omega_{-\Upsilon}, \Upsilon = e)$ is large. Following Newton and Raftery (1994), we employ an importance sampling function that is a mixture of samples from the model prior and posterior samples from our MCMC algorithm conditional on $\Upsilon = e$. Application of Bayes theorem allows us then to recover estimates $\hat{\lambda}_e$.

For the partially linked partitions model, we take advantage of the Gibbs sampling step for Υ. Based on intermediate calculations saved during Gibbs sampling (Weiss et al., 1999), we use Rao-Blackwellization to provide an efficient estimator (Casella and Robert, 1996). Rao-Blackwellization reduces the variance of an unbiased estimator by replacing the estimator by its conditional expectation given a sufficient statistic.

By conditioning,

$$p(\Upsilon \mid Y) = \int_{\Omega_{-\Upsilon}} p(\Upsilon \mid Y, \Omega_{-\Upsilon}) p(\Omega_{-\Upsilon} \mid Y)\, \mathrm{d}\Omega_{-\Upsilon}. \quad (27)$$

Let $\Omega_{-\Upsilon}^{(p)}$ for $p = 1, \ldots, P$ be a marginal posterior sample from our model, then the Rao-Blackwell estimator is

$$\hat{\lambda}_e = \frac{1}{P} \sum_{p=1}^{P} p\big(\Upsilon = e \,\big|\, Y, \Omega_{-\Upsilon}^{(p)}\big) \quad \text{for } e = 1, \ldots, E. \quad (28)$$

Conveniently, this estimator is the posterior mean of the full conditional probabilities calculated during each Gibbs cycle (Eq. 20).

Examples

Guinea Pigs as Rodents

Common lore holds that guinea pigs (*Cavia porcellus*), originally domesticated in South America as a food source and later brought to Europe as pets, are rodents. As far back as Linnaeus in 1758, guinea pigs have been classified within the order Rodentia. Graur et al. (1991) questioned this classification using a maximum-parsimony analysis of 15 protein sequences and found that the order Rodentia is polyphyletic, suggesting that guinea pigs should be separated out of the order. Since that time, several research groups have broached the guinea-pig-as-rodent hypothesis using various nucleotide and protein sequences with parsimony and maximum-likelihood–based approaches (Hasegawa et al., 1992; Cao et al., 1994, 1997; Frye and Hedges, 1995; D'Erchia et al., 1996; Sullivan and Swofford, 1997). Results from these analyses have been contradictory and sometimes inconclusive regarding the placement of the guinea pig. One reason for the disparate findings has been the lack of methodology to reconstruct the evolutionary histories from multiple data partitions simultaneously and infer tendencies across the partitions. Our intention in this small example here is not to resolve the guinea pig issue once and for all; for example, Sullivan and Swofford (1997) show that nucleotide substitution model choice, in particular within-site-class rate variation, can greatly affect inference in this problem. Rather, our intention is to illustrate how a hierarchical phylogenetic model can provide more efficient parameter estimation and could possibly shed additional light on the question.

Following D'Erchia et al. (1996), we attack the guinea pig problem using nucleotide sequences from all $K = 13$ protein-coding mitochondrial genes. For each of the 13 partitions, we construct a four-taxon alignment consisting of a single gene sequence from a (1) guinea pig, (2) rat, (3) human, and (4) opossum. We divide partitions into $C = 3$ site classes based on first, second, or third codon position. Partition lengths range from 207 nucleotides for *atp8* to 1,842 nucleotides for *nd5*, with an average length of 890 nucleotides. All sequence alignments originate from AMmtDB, a database of multiply aligned metazoan mitochondrial DNA (Lanave et al., 2000). Serving as an outgroup, the opossum is a marsupial and is assumed to have evolutionarily diverged before the remaining three eutherial (placental) taxa in our analysis.

Among the $N = 4$ taxa, there exist $E = 3$ possible topologies for each partition. Figure 1 depicts these topologies. Two topologies $\tau_{\overline{R}_1}$ and $\tau_{\overline{R}_2}$ are inconsistent with the guinea-pig-as-rodent hypothesis. In these topologies, the guinea pig is nearest neighbor to either the human or opossum, suggesting that the order Rodentia is polyphyletic. The remaining topology τ_R is consistent with a monophyletic order Rodentia but does not offer definitive support of the guinea pig as a rodent; this support may change depending on taxon choice. Hence, estimating the posterior probability of the inconsistent topologies versus the consistent topology provides a conservative test of polyphyly. We utilize the completely linked partitions model. To assign a prior probability distribution over all possible common topologies Υ, we first assume equal prior probability over monophyly versus polyphyly and further assume that the two polyphyletic topologies are also equally probable. As a result, $Q_R = 1/2$ and $Q_{\overline{R}_1} = Q_{\overline{R}_2} = 1/4$.

Table 1 presents the partition-level evolutionary estimates for the 13 mitochondrial genes divided into three codon positions or site classes. Listed in the table are estimates of the transition:transversion rate ratios α_{kc} and γ_{kc}, expected divergences μ_{kc}, and stationary

TABLE 1. Hierarchical partition-level estimates for the guinea pig example. For each gene k and codon position (site) class c, estimates of α_{kc}, γ_{kc}, and μ_{kc} are posterior means and 95% confidence intervals and estimates of π_{kc} are posterior means.

					π_{kc}			
k	c	α_{kc}	γ_{kc}	$\mu_{kc} \times 10$	A	G	C	T
atp6	1	1.4 (0.8–2.1)	3.2 (2.0–4.7)	1.5 (0.7–2.8)	0.38	0.16	0.28	0.19
	2	2.2 (0.9–4.4)	2.5 (1.5–4.1)	0.6 (0.3–1.2)	0.16	0.10	0.30	0.44
	3	4.4 (2.5–6.9)	12.2 (4.9–27.1)	10.0 (4.2–20.9)	0.41	0.05	0.28	0.26
atp8	1	2.7 (1.2–5.3)	1.5 (0.8–2.6)	1.9 (1.0–3.6)	0.40	0.09	0.25	0.27
	2	1.4 (0.5–2.8)	2.6 (1.4–4.4)	1.3 (0.7–2.3)	0.27	0.06	0.34	0.34
	3	3.3 (1.5–6.3)	12.6 (4.8–26.8)	8.3 (3.3–17.8)	0.45	0.06	0.26	0.23
co1	1	2.3 (1.5–3.5)	7.3 (4.9–10.6)	1.0 (0.5–2.1)	0.27	0.28	0.19	0.26
	2	1.5 (0.6–3.2)	3.8 (2.0–7.0)	0.4 (0.1–0.8)	0.19	0.14	0.26	0.41
	3	8.4 (5.3–13.9)	13.1 (6.4–34.4)	10.3 (4.6–21.7)	0.38	0.06	0.29	0.27
co2	1	1.8 (1.1–2.8)	2.7 (1.7–4.2)	1.3 (0.6–2.6)	0.30	0.23	0.23	0.23
	2	2.0 (0.8–3.8)	3.1 (1.7–5.2)	0.6 (0.3–1.1)	0.25	0.11	0.24	0.40
	3	5.3 (3.0–9.0)	13.0 (5.6–29.9)	9.9 (4.2–20.6)	0.40	0.06	0.28	0.27
co3	1	2.5 (1.4–4.0)	3.4 (2.1–5.3)	1.1 (0.5–2.2)	0.27	0.21	0.24	0.29
	2	1.5 (0.6–3.0)	2.9 (1.5–4.9)	0.5 (0.2–1.0)	0.22	0.15	0.25	0.37
	3	7.1 (3.9–14.5)	12.8 (5.7–28.3)	9.6 (4.2–19.8)	0.41	0.05	0.30	0.25
cytb	1	1.9 (1.3–2.8)	2.3 (1.5–3.2)	1.3 (0.6–2.5)	0.31	0.20	0.25	0.25
	2	1.6 (0.8–3.0)	3.0 (1.8–4.5)	0.6 (0.3–1.1)	0.20	0.13	0.26	0.41
	3	6.7 (3.9–11.3)	13.3 (6.0–28.5)	10.3 (4.5–21.6)	0.40	0.04	0.38	0.18
nd1	1	2.0 (1.3–2.9)	2.6 (1.8–3.7)	1.6 (0.8–2.9)	0.32	0.19	0.26	0.24
	2	2.2 (1.0–4.2)	2.9 (1.8–4.5)	0.6 (0.3–1.2)	0.19	0.10	0.29	0.42
	3	8.8 (4.4–17.3)	10.5 (4.3–23.0)	10.4 (4.6–21.3)	0.43	0.04	0.34	0.18
nd2	1	1.1 (0.7–1.5)	1.6 (1.1–2.2)	1.9 (1.0–3.6)	0.40	0.12	0.25	0.22
	2	1.4 (0.7–2.5)	2.4 (1.7–3.3)	1.0 (0.6–1.9)	0.20	0.08	0.31	0.41
	3	7.4 (3.6–14.6)	11.9 (4.7–24.6)	11.4 (5.0–23.1)	0.45	0.04	0.32	0.20
nd3	1	1.2 (0.6–2.0)	2.2 (1.3–3.5)	1.8 (0.9–3.3)	0.32	0.17	0.23	0.28
	2	1.8 (0.7–3.8)	4.8 (2.6–8.3)	0.8 (0.4–1.5)	0.18	0.11	0.27	0.44
	3	8.9 (4.0–20.3)	10.3 (3.9–22.4)	10.3 (4.5–21.3)	0.45	0.05	0.27	0.24
nd4	1	1.4 (1.0–1.9)	2.6 (1.9–3.4)	1.7 (0.9–3.1)	0.38	0.13	0.25	0.24
	2	1.9 (1.0–3.2)	2.0 (1.3–2.7)	0.7 (0.4–1.3)	0.20	0.11	0.27	0.41
	3	6.3 (3.6–12.2)	11.9 (5.3–25.1)	10.8 (4.8–22.5)	0.44	0.04	0.31	0.21
nd4l	1	1.4 (0.7–2.4)	2.3 (1.2–3.8)	1.7 (0.8–3.1)	0.33	0.19	0.22	0.27
	2	1.4 (0.5–3.1)	4.8 (2.5–8.5)	0.8 (0.4–1.5)	0.18	0.09	0.26	0.47
	3	6.3 (2.6–14.0)	10.0 (3.5–21.2)	10.8 (4.6–22.3)	0.43	0.04	0.30	0.23
nd5	1	1.1 (0.8–1.4)	2.1 (1.6–2.7)	1.7 (0.8–3.1)	0.38	0.15	0.23	0.24
	2	1.5 (0.9–2.3)	2.2 (1.6–2.9)	0.8 (0.4–1.6)	0.22	0.10	0.28	0.40
	3	8.3 (4.8–14.2)	11.1 (5.1–22.3)	10.9 (4.8–22.1)	0.41	0.03	0.32	0.23
nd6	1	2.0 (1.3–2.9)	2.5 (1.3–4.2)	2.1 (1.1–3.8)	0.26	0.35	0.09	0.31
	2	2.5 (1.4–4.1)	1.8 (1.1–2.8)	1.1 (0.6–2.0)	0.18	0.21	0.16	0.45
	3	4.4 (2.1–10.1)	10.8 (3.2–26.8)	11.9 (5.3–24.4)	0.27	0.21	0.08	0.43

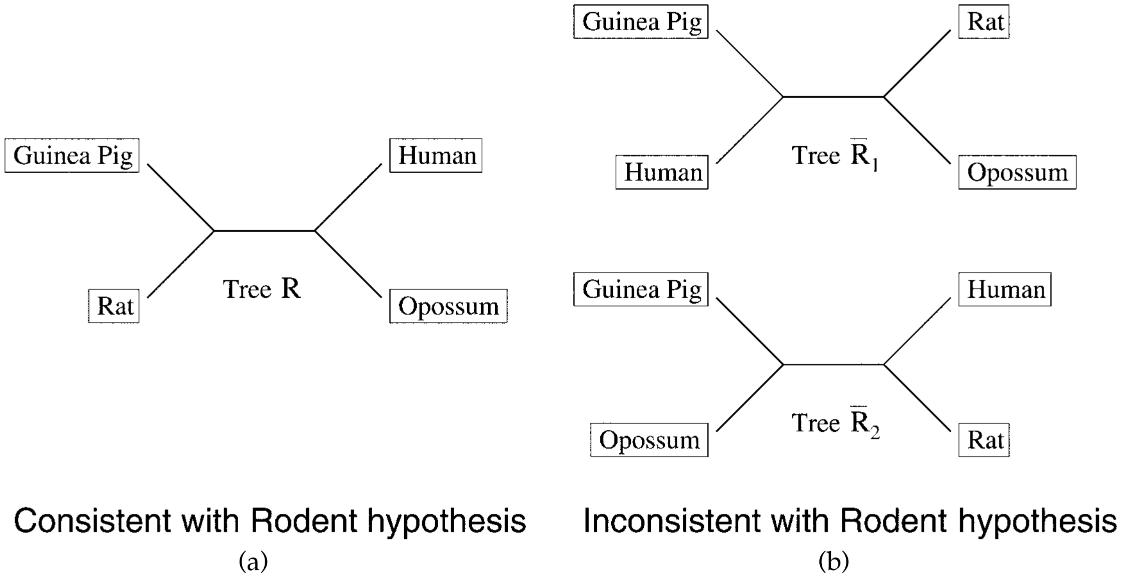

FIGURE 1. Three possible topologies relating guinea pigs, rats, humans, and opossum. (a) Topology R is consistent with the guinea-pig-as-rodent hypothesis. (b) Topologies $\overline{R}_1$ and $\overline{R}_2$ are inconsistent with this hypothesis.

distributions π_{kc}. As expected, the most noticeable feature of the table is the universally larger estimates of the transition:transversion rate ratios and divergence of the third codon position as compared with the first and second positions. However, even within the hierarchical structure of a single site class, variability between corresponding parameters exists. Among ratios α_{kc} and γ_{kc}, γ_{k1} has the largest range of point estimates, 1.5–7.3. Ratio α_{k2} has the smallest range, 1.4–2.5. Within classes, ratios $\alpha_{co1,1}$ and $\alpha_{nd5,1}$ significantly differ; their 95% Bayesian credible intervals do not overlap. Further, $\gamma_{co1,1}$ is significantly larger than γ_{k1} for 11 other genes, overlapping only with $\gamma_{co3,1}$. These differences highlight the need for a hierarchical model when analyzing multiple partitions, as these model accommodate such variability in a parsimonious manner. Expected divergences μ_{kc} also express some variability across partitions, with the *co1* gene demonstrating the slowest rate of evolution across first and second codon positions and *atp8* and *nd6* demostrating the fastest rates.

Table 2 lists the across-partition-level estimates across genes. We calculate posterior probability estimates $\hat{\lambda}_e$ that the common tree $\Upsilon = e$ using importance sampling. The conditional marginal log likelihoods are (−39628, −39592, −39597) for $e \in (R, \overline{R}_1, \overline{R}_2)$, respectively. Using the posterior probability estimates, the Bayes factor in favor of polyphyly,

$$B_{\overline{R},R} = \left(\frac{\hat{\lambda}_{\overline{R}_1} + \hat{\lambda}_{\overline{R}_2}}{\hat{\lambda}_R}\right) \Big/ \left(\frac{Q_{\overline{R}_1} + Q_{\overline{R}_2}}{Q_R}\right), \qquad (29)$$

equals 1.4×10^{15}, decisively rejecting the guinea pig as a rodent given this restricted nucleotide substitution model.

Returning to Table 2, we appreciate little difference between our point estimates for the across-partition-level SDs σ_α, σ_γ, and σ_μ. This coincidence is not surprising; the difference between $\sum_c SS_{Ac}$, $\sum_c SS_{Gc}$, and $\sum_c SS_{Mc}$ is small relative to the uninformative hyperprior parameters $\psi_{\sigma^2,1} + (K \times C)/2$ and $\psi_{\sigma^2,2}$, found in the full conditions distributions of σ_α^2, σ_γ^2, and σ_μ^2. This observation suggests that informative prior information, when available, may aid inference. Finally, site class-specific across-partition-level parameters A_c, G_c, and M_c follow the pattern of variability observed at the partition-level across classes; notably, rates of evolution, as measured by M_c, are significantly larger for the third codon position than for the first or second positions. We also observe significantly heterogeneous nucleotide composition across codon positions, as seen through Π_c.

For comparison to the hierarchical approach, Table 3 reports the posterior estimates of α_{kc}, γ_{kc}, μ_{kc}, and π_{kc} when the data are analyzed using a mixture model consisting of a single topology and a priori independent continuous parameters across partitions and site classes. Although the mixture model returns the same most probable topology, $\tau_{\overline{R}_1}$, as the hierarchical model, we observe from Table 3 that estimates of the continuous parameters are in general less precise. To quantify the gain in efficiency bought by the hierarchical model, we calculate across partitions the average percentage reduction in the length of the 95% confidence intervals when comparing the hierarchical and the mixture results. These gains in efficiency range from 12% for γ_{k1} to 61% for γ_{k3}.

HGT Among Prokaryotes

Biologists increasingly have recognized HGT between different species as an important mechanism of evolution (Syvanen, 1994; Lawrence, 1999; Jain et al., 2002) and in particular among prokaryotes (Jain et al., 1999; Koonin et al., 2001). The ability of prokaryotes to quickly adapt to new environments often results from the acquisition of new genes through HGT rather than by the alteration of current gene function by random mutation (Lawrence, 1999). Within prokaryotes, genes or complete operans can be horizontally transferred by means of transformation, conjugation, and transduction (Jain et al., 2002). The rate of HGT has traditionally been difficult to estimate (Lawrence, 1999).

Both phylogenetic reconstruction using orthologous genes (Jain et al., 1999) and similarity approaches based on gene content (Lawrence and Ochman, 1997) are popular methods to examine HGT across species. Phylogenetic methods offer an advantage over similarity based approaches; the reconstructed topologies from a phylogenetic method have direct biological interpretability as descriptions of the underlying evolutionary histories of the different genes (Doolittle, 1999). If the reconstructed topology for a gene differs from the known phylogeny of the species, then HGT is suggested as a possible explanation (Syvanen, 1994). In many situations, the species

TABLE 2. Hierarchical across-partition-level estimates for the guinea pig example. Parameters $\hat{\lambda}_e$ estimate the posterior probability that the common topology $\Upsilon = e$. Remaining continuous parameter estimates are posterior means and 95% confidence intervals.

Across site class estimates		Site class specific estimates (for class c)			
Parameter		Parameter	1	2	3
$\hat{\lambda}_R$	7.2×10^{-16}	A_c	0.48 (0.18–0.78)	0.49 (0.08–0.87)	1.78 (1.44–2.13)
$\hat{\lambda}_{\overline{R}_1}$	9.9×10^{-1}	G_c	0.92 (0.63–1.20)	1.01 (0.71–1.32)	2.36 (1.83–2.96)
$\hat{\lambda}_{\overline{R}_2}$	8.6×10^{-3}	M_c	−1.92 (−2.28––1.55)	−2.70 (−3.06––2.32)	−0.05 (−0.52–0.46)
σ_A	0.47 (0.34–0.64)	Π_{cA}	.33 (.29–.37)	.21 (.17–.24)	.41 (.37–.45)
σ_G	0.46 (0.35–0.62)	Π_{cG}	.19 (.15–.22)	.12 (.09–.14)	.06 (.04–.08)
σ_M	0.47 (0.34–0.64)	Π_{cC}	.23 (.19–.26)	.27 (.23–.30)	.28 (.25–.32)
N_Π	50.5 (36.9–67.0)	Π_{cT}	.26 (.22–.29)	.41 (.37–.45)	.25 (.21–.28)

TABLE 3. Independent estimates for the guinea pig example. For each gene k and site class c, estimates of α_k, γ_k, and μ_k are posterior means and 95% confidence intervals, and estimates of π_k are posterior means.

					π_k			
k	c	α_k	γ_k	$\mu_k \times 10$	A	G	C	T
atp6	1	1.3 (0.7–2.2)	3.3 (2.0–5.2)	1.6 (0.6–3.9)	0.39	0.15	0.28	0.18
	2	2.8 (0.6–6.9)	2.4 (1.3–4.2)	0.7 (0.2–1.7)	0.15	0.10	0.30	0.45
	3	3.6 (1.9–6.1)	12.0 (3.7–55.0)	10.1 (2.7–37.1)	0.41	0.05	0.28	0.26
atp8	1	6.8 (1.7–20.3)	0.9 (0.3–2.0)	3.1 (1.1–8.4)	0.39	0.05	0.27	0.29
	2	0.7 (0.0–2.5)	2.3 (1.1–4.3)	2.4 (0.9–6.1)	0.29	0.05	0.36	0.31
	3	1.3 (0.3–3.3)	11.8 (2.8–51.8)	6.4 (1.7–21.9)	0.47	0.08	0.24	0.21
co1	1	2.9 (1.7–4.6)	9.7 (6.2–14.5)	0.7 (0.3–1.9)	0.26	0.29	0.19	0.26
	2	1.0 (0.1–3.7)	4.3 (1.8–9.0)	0.2 (0.1–0.5)	0.19	0.14	0.26	0.41
	3	14.7 (5.3–65.7)	19.5 (5.9–78.7)	15.2 (3.8–52.6)	0.38	0.06	0.29	0.27
co2	1	1.8 (1.0–3.0)	2.8 (1.6–4.5)	1.3 (0.5–3.3)	0.30	0.24	0.23	0.23
	2	2.4 (0.6–5.9)	3.4 (1.6–6.5)	0.5 (0.2–1.3)	0.26	0.11	0.24	0.39
	3	5.0 (2.5–9.5)	15.2 (4.6–63.6)	11.4 (3.0–39.9)	0.39	0.06	0.28	0.27
co3	1	3.2 (1.7–5.4)	3.9 (2.2–6.4)	0.9 (0.3–2.3)	0.25	0.21	0.24	0.30
	2	1.1 (0.2–3.0)	2.6 (1.2–5.0)	0.4 (0.1–1.1)	0.23	0.16	0.25	0.36
	3	10.7 (3.5–38.0)	18.0 (5.3–72.0)	12.2 (3.1–44.1)	0.41	0.05	0.30	0.25
cytb	1	2.0 (1.2–2.9)	2.2 (1.4–3.2)	1.3 (0.5–3.2)	0.31	0.20	0.25	0.25
	2	1.5 (0.5–3.4)	3.0 (1.7–4.8)	0.5 (0.2–1.4)	0.20	0.14	0.26	0.40
	3	8.6 (3.6–30.5)	19.3 (5.4–76.3)	14.3 (3.5–51.1)	0.39	0.04	0.38	0.18
nd1	1	2.1 (1.3–3.0)	2.6 (1.7–3.8)	1.8 (0.7–4.4)	0.31	0.19	0.26	0.24
	2	2.9 (0.9–6.3)	2.9 (1.7–4.9)	0.6 (0.2–1.6)	0.19	0.10	0.29	0.42
	3	21.9 (4.3–102)	13.5 (3.5–59.8)	14.9 (3.6–55.2)	0.43	0.04	0.34	0.18
nd2	1	0.9 (0.6–1.4)	1.4 (0.9–2.0)	2.7 (1.1–6.6)	0.41	0.11	0.25	0.22
	2	1.3 (0.5–2.6)	2.3 (1.6–3.2)	1.7 (0.6–4.1)	0.19	0.08	0.31	0.41
	3	16.8 (3.2–82.1)	15.8 (3.9–67.0)	17.5 (4.2–62.8)	0.45	0.04	0.32	0.19
nd3	1	0.9 (0.3–1.8)	2.0 (1.0–3.4)	2.3 (0.9–5.6)	0.31	0.17	0.23	0.29
	2	2.1 (0.2–7.0)	6.8 (3.0–13.9)	1.1 (0.4–2.9)	0.18	0.11	0.27	0.44
	3	20.5 (3.6–93.0)	9.0 (2.5–41.6)	12.3 (3.3–42.3)	0.45	0.05	0.27	0.24
nd4	1	1.4 (0.9–1.9)	2.6 (1.8–3.4)	2.1 (0.8–5.1)	0.38	0.13	0.25	0.24
	2	2.0 (0.9–3.5)	1.8 (1.2–2.6)	0.8 (0.3–2.1)	0.20	0.11	0.28	0.41
	3	8.4 (3.2–36.5)	16.0 (4.5–67.9)	14.7 (3.8–52.0)	0.44	0.04	0.31	0.21
nd4l	1	1.2 (0.5–2.5)	2.0 (0.9–3.8)	2.1 (0.8–5.1)	0.32	0.20	0.21	0.27
	2	0.6 (0.0–3.0)	6.9 (2.8–14.9)	1.1 (0.4–2.9)	0.18	0.09	0.25	0.48
	3	9.2 (1.4–50.6)	9.0 (1.9–42.7)	11.6 (2.9–43.0)	0.43	0.04	0.31	0.22
nd5	1	1.0 (0.7–1.3)	2.0 (1.5–2.6)	2.1 (0.8–5.0)	0.38	0.15	0.23	0.24
	2	1.4 (0.7–2.4)	2.1 (1.6–2.8)	1.2 (0.5–3.0)	0.22	0.10	0.28	0.40
	3	21.2 (5.0–101)	16.1 (4.5–70.0)	17.1 (4.1–62.7)	0.41	0.03	0.32	0.23
nd6	1	2.0 (1.2–3.1)	2.9 (1.2–5.5)	3.0 (1.2–7.4)	0.25	0.37	0.07	0.31
	2	2.7 (1.4–4.7)	1.7 (0.9–2.8)	1.8 (0.7–4.4)	0.18	0.22	0.15	0.45
	3	4.7 (1.5–18.9)	8.1 (1.2–42.8)	12.9 (3.7–43.8)	0.26	0.22	0.07	0.44

phylogeny is not known with absolute certainty and should be jointly estimated along with the individual gene topologies.

Here, we briefly illustrate the utility of a hierarchical phylogenetic model to examine HGT using multiple orthologous gene alignments simultaneously. Jain et al. (1999) construct a data set of 144 separate gene alignments. Each alignment contains orthologous copies of a gene from the same six prokaryotes. To limit computational demand, we randomly subsample $K = 50$ alignments of genes from $N = 4$ taxa: *Escherichia coli* (Ec) and *Synechocystis* 6803 (S6), both Eubacteria, and *Methanococcus jannaschii* (Mj) and *Archeoglobus fulgidus* (Af), both Archaea. Maintaining consistency with Jain et al. (1999), we exclude the third codon position from analysis and assumed a single site class ($C = 1$) for the first and second positions within a gene. Partition lengths range from 156 to 1,580 nucleotides, with an average length of 585 nucleotides. Among the $E = 3$ possible topologies, few would doubt that $\tau_{AE} =$ (Ec, S6, (Mj, Af)) describes the true species phylogeny, splitting the Archaea from the Eubacteria. However, for illustrative purposes, we assume that the species topology is completely unknown and simultaneously infer the common topology Υ among the four species and an upper estimate of the frequency of HGT p. In doing so, we take $Q = (1/3, 1/3, 1/3)$ and $\psi_{p,1} = \psi_{p,2} = 1$, such that the prior on p is uniform over 0 to 2/3.

Figure 2 plots the partition-level, cumulative posterior probabilities of the three possible topologies for all 50 genes. The vast majority of the genes support τ_{AE} as most probable (open boxes), while eight genes support alternative topologies. Four genes recover $\tau_{\overline{AE}_1} =$ (Mj, S6, (Ec, Af)) as most probable (shaded boxes), and four genes recover $\tau_{\overline{AE}_2} =$ (Ec, Mj, (S6, Af)) as most probable (solid boxes). For brevity, partition-level estimates of the evolutionary pressure parameters are not shown. However, when compared with estimates obtained while analyzing each gene independently, the hierarchical estimates show increased precision, similar to the guinea pig example. Table 4 presents the across-partition-level estimates

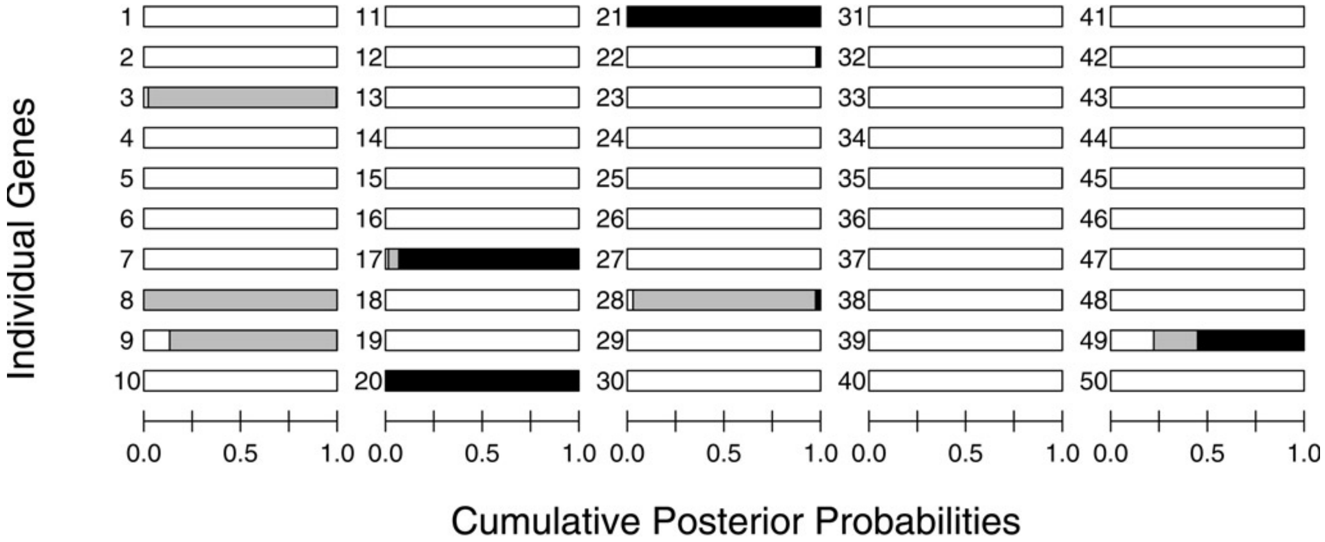

FIGURE 2. HGT among prokaryotes. For each gene, boxes report the cumulative posterior probabilities of the three possible topologies (open $= \tau_{AE}$; shaded $= \tau_{\overline{AE}_1}$; solid $= \tau_{\overline{AE}_2}$). A majority of genes support τ_{AE} as the common topology. Support for alternative topologies suggests HGT.

of the common topology Υ and HGT frequency p, in addition to the remaining across-partition-level parameter estimates. Estimates of p are consistent with previous studies of HGT rates among prokaryotes using similarity based methods; the estimates include 17% in *Escherichia* (Lawrence and Ochman, 1997) and range from 2% to 14% across other prokaryotes (Garcia-Vallve et al., 2000). A strength of the hierarchical approach is that we are able to simultaneously determine the most probable species phylogeny. The Bayes factor in favor of τ_{AE} versus its two alternatives,

$$B_{AE,\overline{AE}} = \left(\frac{\hat{\lambda}_{AE}}{\hat{\lambda}_{\overline{AE}_1} + \hat{\lambda}_{\overline{AE}_2}}\right) \Big/ \left(\frac{Q_{AE}}{Q_{\overline{AE}_1} + Q_{\overline{AE}_2}}\right), \quad (30)$$

equals 4.4×10^{23}, decisively accepting the known tree without any a priori information.

Intrahost Evolution of HIV

Within infected patients, HIV exists as a population of rapidly mutating viruses. Evolution of the viruses is due in part to sloppy replication as they attempt to infect new cells in the body. Evolution of the envelope gene, *env*, has important consequences on disease progression because the gene is implicated in differential coreceptor usage (Connor et al., 1997; Shankarappa et al., 1999; Philpott et al., 2001). Most strains transmitted in vivo utilize the CCR5 coreceptor (R5) (Scarlatti et al., 1997; Berger et al., 1998; Shankarappa et al., 1999). As infection progresses, viral isolates that utilize the CXCR4 coreceptor (X4) emerge in about 50% of hosts (Connor et al., 1997; Scarlatti et al., 1997; Berger et al., 1998). The emergence of X4 strains is correlated with disease progression and death (Connor et al., 1997). Recently, it has been shown that potent antiretroviral therapy (PART) can alter this course and shift the viral populations back to R5-utilizing strains (Este et al., 1999; Philpott et al., 2001). It is not known whether these reemergent R5 strains are latent archived virus or represent further mutation from the X4 strains in light of PART suppression.

TABLE 4. Hierarchical across-partition-level estimates for the horizontal gene transfer example. Parameters $\hat{\lambda}_e$ estimate the posterior probability that the common topology $\Upsilon = e$. Remaining continuous parameter estimates are posterior means and 95% confidence intervals.

Parameter	Estimate	Parameter	Estimate
$\hat{\lambda}_{AE}$	≈1	A	0.44 (0.35–0.53)
$\hat{\lambda}_{\overline{AE}_1}$	3.8×10^{-24}	G	0.21 (0.10–0.32)
$\hat{\lambda}_{\overline{AE}_2}$	7.7×10^{-25}	M	−1.35 (−1.51––1.19)
p	0.17 (0.08–0.28)	Π_A	.34 (.34–.35)
σ_A	0.28 (0.23–0.34)	Π_G	.30 (.29–.31)
σ_G	0.31 (0.25–0.39)	Π_C	.17 (.16–.17)
σ_M	0.35 (0.27–0.46)	Π_T	.19 (.18–.20)
N_Π	334 (252–430)		

To infer the evolutionary ancestry of the reemergent R5 strains from multiple patients simultaneously, we begin with serially sampled HIV-1 sequences, comprising the 105-nucleotide V3 region of *env*, from $K = 4$ infected women who demonstrate the R5 → X4 → R5 progression. For each patient, we construct one data partition. Each partition contains three patient sequences, that of the infecting R5 (R5-i) virus, the disease progressing X4 virus, and the reemergent R5 (R5-r) virus after PART treatment, and an outgroup (O) sequence. Our outgroup sequence comes from HIV-1 clone JRCSF and utilizes the R5 coreceptor (O'Brien et al., 1990). Due to the short length of available sequences, we used only $C = 1$ site class.

Among the $N = 4$ taxa in each partition, there exist $E = 3$ possible topologies. Figure 3 superimposes two of these topologies, E (evolution) and L (latent). When R5-r

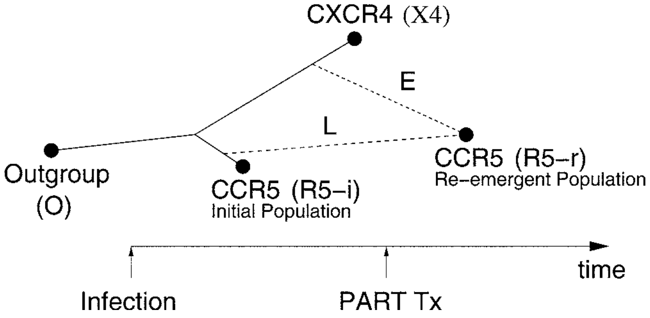

FIGURE 3. Evolution of coreceptor utilization in HIV-1. Two possible hypotheses, continued evolution (E) or latency (L), about the origin of reemergent CCR5 tropic virus are shown.

and X4 are nearest neighbors, the final R5 population most likely evolved from the X4 population. When R5-r and R5-i are nearest neighbors, the final R5 population most likely stems from a latent reservoir of the initial R5 population. The final topology O (outgroup) has R5-r and the outgroup as nearest neighbors and is much less likely a priori given the time sequence of events. Unlike the guinea pig and HGT examples, evolution across partitions is independent, so we employ the unlinked partitions model for this example. We assume a prior pseudocount $N_Q = 1$ and that topologies L and E are equally likely and together are ≈10 times more likely than topology O, such that $Q = (Q_O, Q_E, Q_L) = (0.10, 0.45, 0.45)$.

Table 5 reports the partition-level evolutionary estimates for the four HIV+ patients. Three of the four patients show >95% support for the evolution topology τ_E; the remaining patient demonstrates equivocal support for τ_E and the latent topology τ_L. Table 6 presents the across-partition-level estimates. We determine whether τ_E is more likely than τ_L across all patients using a Bayes factor,

$$B_{E,L} = \left(\frac{T_E}{T_L}\right) \Big/ \left(\frac{Q_E}{Q_L}\right) = 4.1. \tag{31}$$

This result offers evidence in favor of continued evolution of HIV in light of PART suppression. However, overall support is weak, not surprisingly given the small number of patients and the short length of alignments available.

TABLE 6. Hierarchical across-partition-level estimates for the HIV example. Parameter estimates are posterior means and 95% confidence intervals.

Parameter	Estimate	Parameter	Estimate
T_O	0.02 (<0.01–0.23)	A	1.31 (0.38–2.22)
T_L	0.19 (<0.01–0.66)	G	0.49 (−0.90–1.69)
T_E	0.78 (0.31–>0.99)	M	−3.34 (−4.20–−2.44)
σ_A	0.72 (0.41–1.33)	Π_A	.43 (.33–.52)
σ_G	0.81 (0.43–1.62)	Π_G	.21 (.14–.29)
σ_M	0.69 (0.40–1.23)	Π_C	.18 (.11–.26)
N_Π	37.4 (12.7–77.1)	Π_T	.18 (.11–.25)

Finally, Table 7 reports the posterior model estimates when the patients are analyzed completely independently and when concatenated into a single partition. In general, the independent analysis provides less precise continuous parameter estimates. Considering the topology estimates, a consensus conclusion of the odds in favor of the evolution hypothesis versus the latent hypothesis is 3:1. On the other hand, the concatenated analysis overestimates the posterior support. The analysis returns >99% support for τ_E and <1% support for τ_L, even when one of the four patients offers modest support in favor of the latent topology. The hierarchical estimates of $T_E = 0.78$ and $T_L = 0.19$ provide a more balanced summary of the data.

REMARKS

In this article, we introduce a hierarchical phylogenetic model that enables researchers to estimate both the variability between multiple data partitions and tendencies across partitions simultaneously. Like mixture models, the hierarchical model offers a middle ground to the divide between strict combined-data and consensus/independence approaches, drawing on both of their strengths. We illustrate some of the advantages of hierarchical models using three examples. In the first example, we employ a hierarchical approach that is a direct extension of the mixture models readily available in current phylogenetic software. The example demonstrates improved parameter estimation by borrowing strength through hierarchical priors without losing the salient features of the data. In the HGT example, a single topology assumption would miss evidence of incongruent partitions completely, while allowing all partitions to be a priori independent would preclude estimation of a common species topology. A hierarchical approach overcomes both issues simultaneously. Finally, the HIV example shows that a hierarchical solution provides a proper compromise between strict combined-data and

TABLE 5. Hierarchical partition-level estimates for the HIV example. For each subject k, estimates of τ_k are posterior probabilities, estimates of α_k, γ_k, and μ_k are posterior means and 95% confidence intervals, and estimates of π_k are posterior means.

	τ_k						π_k			
k	O	L	E	α_k	γ_k	$\mu_k \times 100$	A	G	C	T
1	<0.01	0.02	0.97	4.7 (1.8–10.4)	2.3 (0.4–6.5)	4.3 (1.7–9.6)	0.43	0.23	0.18	0.17
2	0.02	0.48	0.50	4.4 (1.6–9.9)	1.3 (0.1–4.0)	3.8 (1.5–8.5)	0.46	0.20	0.17	0.17
3	<0.01	<0.01	>0.99	3.0 (0.9–7.2)	2.8 (0.5–8.4)	3.1 (1.1–7.1)	0.44	0.21	0.17	0.18
4	<0.01	<0.01	>0.99	5.3 (2.0–11.8)	2.5 (0.5–7.1)	4.1 (1.6–9.0)	0.45	0.20	0.18	0.17

TABLE 7. Independent and concatenated estimates for the HIV example. For each subject k, estimates of τ_k are posterior probabilities, estimates of α_k, γ_k, and μ_k are posterior means and 95% confidence intervals, and estimates of π_k are posterior means.

	τ_k						π_k			
k	O	L	E	α_k	γ_k	$\mu_k \times 100$	A	G	C	T
Independent										
1	0.06	0.07	0.86	5.1 (1.6–12.9)	2.8 (0.3–9.4)	5.4 (1.8–14.2)	0.43	0.23	0.17	0.17
2	0.11	0.75	0.13	3.9 (1.1–9.6)	0.3 (0.0–1.9)	4.4 (1.3–11.8)	0.47	0.19	0.17	0.17
3	0.04	<0.01	0.96	2.3 (0.4–7.1)	3.5 (0.3–12.1)	3.0 (0.9–8.1)	0.45	0.21	0.17	0.18
4	0.05	0.02	0.94	6.5 (1.9–17.2)	3.5 (0.3–11.8)	5.0 (1.6–13.5)	0.45	0.20	0.17	0.17
Concatenated										
	<0.01	<0.01	>0.99	4.0 (2.3–6.8)	2.1 (0.7–4.6)	4.3 (1.6–11.1)	0.45	0.21	0.17	0.17

consensus estimates of the most likely topology across partitions.

The general hierarchical phylogenetic model proposed here can be easily extended. For example, one can incorporate more complicated models of nucleotide substitution, including those that consider within-site-class rate variation (Yang, 1994) and codon models (Goldman and Yang, 1994; Schadt et al., 2002). The hierarchical model can also be applied to other forms of sequence information as well, i.e., amino acid sequences, given appropriate changes to the sequence mutation models used (Whelan and Goldman, 2001). Our current framework and presentation assumes that the same equivalent taxa are present in each partition. With simple adaptation of the hierarchical structure placed on topologies, overlapping sets of taxa can also be accommodated and supertrees can be constructed (Sanderson et al., 1998; Salamin et al., 2002).

One feature central to the hierarchical framework is the exchangeability assumption between partitions. Exchangeability implies that inference is invariant to permutation of the partition labels. Although exchangeability is reasonable between independent patients in our unlinked model example, the order of mitochondrial or prokaryotic genes in our first two examples may contain additional information and impart further correlation between partitions. For example, a recombination event along a chromosome may cause the topologies to vary between genes on either side of the event. One potential extension to our model accommodates partial exchangeability. Here, continuous model parameters remain exchangeable, while the topologies follow a Markovian prior (Suchard et al., 2002).

The hierarchical framework we propose over topologies requires estimation of $E - 1$ free across-partition-level probabilities in the unlinked partitions model and enumeration over all E possible common topologies in the partially linked partitions model. For the examples considered in this paper, $E = 3$; thus, the estimation or enumeration remains quite modest. However, E increases superexponentially with the number of taxa N in a partition. Although many phylogenetic questions can be reduced to four-taxon problems, at least two solutions exist for handling large E in a hierarchical framework. The first solution proposes grouping disjoint subsets of all possible topologies into a modest number of cliques and estimating their probabilities. Within a clique, all topologies are then considered equally likely. The definition of the cliques depends on the research question at hand. A second solution advocates sampling the common topology Υ using a Metropolis–Hastings algorithm, as is done in the completely linked partitions model, rather than Gibbs sampling.

ACKNOWLEDGMENTS

We thank the Berger-Weiser lab, in particular Sean Philpott, for providing the HIV viral sequences and the Lake lab for providing the prokaryote sequences. We make our complete datasets available to interested readers at http://www.biomath.medsch.ucla.edu/msuchard/datasets.html. M.A.S. and C.M.R.K. are supported by the UCLA AIDS Institute Center for AIDS Research (AI28697). M.A.S. and J.S.S. are supported by USPHS grant CA16042.

REFERENCES

BERGER, E., R. DOMS, E. FENYO, B. KORBER, D. LITTMAN, J. MOORE, Q. SATTENTAU, H. SCHUITEMAKER, J. SODROSKI, AND R. WEISS. 1998. A new classification for HIV-1. Nature 391:240.

BUCKLEY, T., P. ARENSBURGER, C. SIMON, AND G. CHAMBERS. 2002. Combined data, Bayesian phylogenetics, and the origins of the New Zealand cicada genera. Syst. Biol. 51:4–18.

BUCKLEY, T., C. SIMON, AND G. CHAMBERS. 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: The effects of model assumptions on estimates of topology, branch lengths and bootstrap support. Syst. Biol. 50:67–86.

BULL, J., J. HUELSENBECK, C. CUNNINGHAM, D. SWOFFORD, AND P. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

CAO, Y., J. ADACHI, T. YANO, AND M. HASEGAWA. 1994. Phylogenetic place of guinea pigs: No support of the rodent-polyphyly hypothesis from maximum-likelihood analyses of multiple protein sequences. Mol. Biol. Evol. 11:593–604.

CAO, Y., A. JANKE, P. WADDELL, M. WESTERMAN, O. TAKENAKA, S. MURATA, N. OKADA, S. PÄÄBO, AND M. HASEGAWA. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. 47:307–322.

CAO, Y., N. OKADA, AND M. HASEGAWA. 1997. Phylogenetic position of guinea-pigs revisited. Mol. Biol. Evol. 14:461–464.

CASELLA, G., AND C. ROBERT. 1996. Rao-Blackwellisation of sampling schemes. Biometrika 83:81–94.

CONNOR, R., K. SHERIDAN, D. CERADINI, S. CHOE, AND N. LANDAU. 1997. Change in coreceptor use correlates with disease progression in HIV-1 infected individuals. J. Exp. Med. 185:621–628.

D'ERCHIA, A., C. GISSI, G. PESOLE, C. SACCONE, AND U. ARNASON. 1996. The guinea-pig is not a rodent. Nature 381:597–600.

DOOLITTLE, W. 1999. Lateral gene transfer, genome surveys and the phylogeny of prokaryotes [technical comments]. Science 286: 1443a.

DORMAN, K., A. KAPLAN, AND J. SINSHEIMER. 2002. Bootstrap confidence levels for HIV-1 recombination. J. Mol. Evol. 54:200–209.

ESTE, J., C. CABRERA, J. BLANCO, A. GUTIERREZ, G. BRIDGER, G. HENSON, D. SCHOLS, AND E. DECLERCQ. 1999. Shift of clinical human immunodeficiency virus type 1 isolates from X4 to R5 and prevention of emergence of the syncytium-inducing phenotype by blockade of CXCR4. J. Virol. 73:5577–5585.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

Felsenstein, J. 1993. Phylip (phylogenetic inference package), version 3.5. Distributed by the author, Department of Genetics, Univ. Washington, Seattle.

Frye, M., and S. Hedges. 1995. Monophyly of the order rodentia inferred from mitochondrial DNA sequences of the genes for 12S ribosomal RNA, 16S ribosomal RNA and transfer RNA valine. Mol. Biol. Evol. 12:168–176.

Garcia-Vallve, S., A. Romeu, and J. Palau. 2000. Horizontal gene transfer in bacterial and archeal complete genomes. Genome Res. 10:1719–1725.

Gelman, A., J. Carlin, H. Stern, and D. Rubin. 1995. Bayesian data analysis. Chapman and Hall/CRC, New York.

Gelman, A., G. Roberts, and W. Gilks. 1996. Efficient Metropolis jumping rules. Pages 599–608 *in* Bayesian Statistics, Volume 5 (J. Bernardo, J. Berger, A. Dawid, and A. Smith, eds.) Oxford Univ. Press, Oxford, U.K.

Goldman, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182–198.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11:725–736.

Graur, D., W. Hide, and W. Li. 1991. Is the guinea-pig a rodent? Nature 351:649–652.

Hasegawa, M., Y. Cao, J. Adachi, and T. Yano. 1992. Rodent polyphyly. Nature 355:595.

Hasegawa, M., A. Rienzo, T. Kocher, and A. Wilson. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. J. Mol. Evol. 37:347–354.

Hastings, W. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.

Huelsenbeck, J., and B. Rannala. 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. Science 276:227–232.

Huelsenbeck, J., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Huelsenbeck, J., F. Ronquist, R. Nielsen, and J. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

Jain, R., M. Rivera, and J. Lake. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. Proc. Natl. Acad. Sci. USA 96:3801–3806.

Jain, R., M. Rivera, J. Moore, and J. Lake. 2002. Horizontal gene transfer in microbial genome evolution. Theor. Popul. Biol. 61:489–495.

Jeffreys, H. 1998. Theory of probability. Oxford classic texts in the physical sciences, 3rd edition. Oxford Univ. Press, New York.

Kass, R., and A. Raftery. 1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.

Kluge, A. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38:7–25.

Koonin, E., K. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokaryotes: Quantification and classification. Annu. Rev. Microbiol. 55:709–742.

Laird, N., and J. Ware. 1982. Random-effects models for longitudinal data. Biometrics 38:963–974.

Lanave, C., S. Liuni, F. Licciulli, and M. Attimonelli. 2000. Update of AMmtDB: A databsae of multi-aligned Metazoa mitochondrial DNA sequences. Nucleic Acids Res. 28:153–154.

Larget, B., and D. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750–759.

Lawrence, J. 1999. Gene transfer, speciation and the evolution of bacterial genomes. Curr. Opin. Microbiol. 2:519–523.

Lawrence, J., and H. Ochman. 1997. Amelioration of bacterial genomes: Rates of change and exchange. J. Mol. Evol. 44:383–397.

Li, S., D. Pearl, and H. Doss. 2000. Phylogenetic tree construction using Markov chain Monte Carlo. J. Am. Stat. Assoc. 95:493–508.

Mau, B., and M. Newton. 1997. Phylogenetic inference for binary data on dendograms using Markov chain Monte Carlo. J. Comput. Graph. Stat. 6:122–131.

Mau, B., M. Newton, and B. Larget. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55:1–12.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092.

Miyamoto, M. 1985. Consensus cladograms and general classification. Cladistics 1:186–189.

Miyamoto, M., and W. Fitch. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

Newton, M., B. Mau, and B. Larget. 1999. Markov chain Monte Carlo for the Bayesian analysis of evolutionary trees from aligned molecular sequences. Pages 143–162 *in* Statistics in molecular biology and genetics: Selected proceedings of a 1997 joint AMS-IMS-SIAM summer conference on statistics in molecular biology, Volume 33 (F. Seillier-Moiseiwitsch, ed.). IMS, Hayward, California.

Newton, M., and A. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. J. R. Stat. Soc. B 56:3–48.

O'Brien, W., Y. Koyanagi, A. Namazie, J. Zhao, A. Diagne, K. Idler, J. Zack, and I. Chen. 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside of the CD4-binding domain. Nature 348:69–73.

Penny, D., and M. Hendy. 1986. Estimating the reliability of evolutionary trees. Mol. Biol. Evol. 3:403–417.

Philpott, S., B. Weiser, K. Anastos, C. Kitchen, E. Robison, W. Meyer III, H. Sacks, U. Mathut-Wagh, C. Brunner, and H. Burger. 2001. Preferential suppression of CXCR4-specific strains of HIV-1 by antiviral therapy. J. Clin. Invest. 107:431–438.

Robertson, D., B. Hahn, and P. Sharp. 1995. Recombination in AIDS viruses. J. Mol. Evol. 40:249–259.

Salamin, N., T. Hodkinson, and V. Savolainen. 2002. Building supertrees: An empirical assessment using the grass family (Poaceae). Syst. Biol. 51:136–150.

Sanderson, M., A. Purvis, and C. Henze. 1998. Phylogenetic supertrees: Assembling the trees of life. Trends Ecol. Evol. 13:105–110.

Scarlatti, G., E. Tresoldi, A. Bjorndal, R. Fredrikkson, C. Colognesi, H. Deng, M. Lalnati, A. Plebani, E. Fenyo, and P. Lusso. 1997. In vivo evolution of HIV-1 coreceptor usage and sensitivity to chemokine mediated suppression. Nat. Med. 3:1259–1265.

Schadt, E., J. Sinsheimer, and K. Lange. 2002. Applications of codon and rate variation models in molecular phylogeny. Mol. Biol. Evol. 19:1550–1562.

Shankarappa, R., J. Margolick, S. Gange, A. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. Rinaldo, G. Learn, X. He, X. Huang, and J. Mullins. 1999. Consistent viral evolutionary changes associated with the disease progression of human immunodeficiency virus type 1 infection. J. Virol. 73:10489–10502.

Sinsheimer, J., J. Lake, and R. Little. 1996. Bayesian hypothesis testing of four-taxon topologies using molecular sequence data. Biometrics 52:193–210.

Suchard, M., R. Weiss, K. Dorman, and J. Sinsheimer. 2002. Oh brother, where art thou? A Bayes factor test for recombination with uncertain heritage. Syst. Biol. 51:715–728.

Suchard, M., R. Weiss, and J. Sinsheimer. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. 18:1001–1013.

Suchard, M., R. Weiss, and J. Sinsheimer. 2003a. Testing a molecular clock without an outgroup: Derivations of induced priors on branch length restrictions in a Bayesian framework. Syst. Biol. 52:48–54.

Suchard, M., R. Weiss, J. Sinsheimer, K. Dorman, M. Patel, and E. McCabe. 2003b. Evolutionary similarity among genes. J. Am. Stat. Assoc. 98:653–662.

Sullivan, J., and D. Swofford. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4:77–86.

Swofford, D. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, Massachusetts.

Syvanen, M. 1994. Horizontal gene transfer: Evidence and possible consequences. Annu. Rev. Genet. 28:237–261.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512–526.

Tierney, L. 1994. Markov chains for exploring posterior distributions (with discussion). Ann. Stat. 22:1701–1762.

Weiss, R., M. Cho, and M. Yanuzzi. 1999. On Bayesian calculations for mixture likelihoods and priors. Stat. Med. 18:1555–1570.

Whelan, S., and N. Goldman. 1999. Distribution of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol. Biol. Evol. 16:1292–1299.

Whelan, S., and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18:691–699.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39:306–314.

Yang, Z. 1995a. Evaluation of several methods for estimating phylogenetic trees when substitution rates differ over nucleotide sites. J. Mol. Evol. 40:689–697.

Yang, Z. 1995b. A space–time process model for the evolution of DNA sequences. Genetics 139:993–1005.

Yang, Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42:587–596.

Yang, Z., N. Goldman, and A. Friday. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimatation. Mol. Biol. Evol. 11:316–324.

Yang, Z., and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. Mol. Biol. Evol. 14:717–724.

First submitted 9 December 2002; reviews returned 14 April 2003; final acceptance 29 June 2003
Associate Editor: Ziheng Yang