Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 68(2) 2014 285-292

Bayesian Multiscale Phylogenetics

Marco A.R. Ferreira[1] and M. Alejandra Jaramillo[2]
[1]*Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100*
[2]*Centro de Investigación para el Manejo Ambiental y el Desarrollo - CIMAD, Bogotá Colombia*

Received 31 January 2014; Revised 16 June 2014; Accepted 18 June 2014

SUMMARY

We propose a computational approach for the construction of Bayesian multiscale phylogenetic trees. Specifically, first we classify the DNA sites or nucleotides in different scales of evolutionary resolution using entropy. After that, for each evolutionary resolution level we run a Markov chain Monte Carlo (MCMC) analysis that uses the molecular data up to that resolution level, as well as the last phylogenetic trees simulated from the immediate coarser level. We illustrate the use of our multiscale phylogenetics framework with an application to a large molecular dataset for primates.

Keywords: Bayesian inference, Entropy, Markov chain Monte Carlo, Multiscale analysis.

1. INTRODUCTION

The construction of phylogenies with many species continues to be a major challenge for at least two reasons: large analyses become computationally infeasible, and different levels of variation are encountered at different hierarchical levels. Many authors have addressed this problem in different manners. One proposal is to perform independent analyses, followed by matrix representation of the tree to reconstruct a consensus super-tree (Swenson *et al*. 2012, Bininda-Emonds 2004). Another proposal is to perform compartmentalized analysis, where a local analysis among closely related taxa is carried out, and then those taxa are substituted by a representative "archetype" (a consensus of the taxa analyzed) in a more inclusive analysis (Mishler 1994). In addition, to accelerate computations for phylogenetic studies there have been many recent advances on the use of parallelization and multithreading using graphics processing units (GPUs) (Suchard and Rambaut 2009, Ayres *et al*. 2012, Bao *et al*. 2013). Lastly, in view of the rapid improvement on computational power, many have just disregarded the matter and are running very large analyses, while the problem of different evolutionary/resolution levels still exists.

To address this problem we use a divide and conquer approach to build multiscale phylogenetic trees. Our multiscale phylogenetic methodology builds upon similar multiscale concepts previously used for the analysis of time series, spatial, and spatiotemporal data (Ferreira *et al*. 2006, Ferreira and Lee 2007, Ferreira *et al*. 2011, Sanyal and Ferreira 2012). First, we classify the DNA sites or nucleotides in different scales of evolutionary resolution using entropy. For example, we may create three levels of evolutionary resolution. A first level, that provides coarse evolutionary information, contains sites or nucleotides with low entropy that evolve slower and thus contain information about the parts of the phylogenetic tree that are closer to the oldest common ancestor. A second level, that

Corresponding author: Marco A.R. Ferreira
E-mail address: ferreiram@missouri.edu

provides intermediate evolutionary information, contains sites or nucleotides with an intermediate level of entropy. Finally, a third level that provides fine evolutionary information contains sites with high entropy that evolve faster and thus contain information about the most recent splits in the tree of life.

After we classify the sites or nucleotides in different scales of resolution, we implement a Markov chain Monte Carlo (MCMC) analysis for the coarse level nucleotides. We then use the output of this MCMC analysis as the starting point for an MCMC analysis that adds the intermediate level nucleotides to the coarse level nucleotides. In this analysis, there are distinct evolutionary rate parameters for the coarse level and for the intermediate level partitions. We then use the output of this latter MCMC analysis as the starting point for an MCMC analysis that uses all the nucleotides. For this final analysis, there are distinct evolutionary rate parameters for each of the evolutionary resolution level partitions. This multiscale approach has two benefits: first, the use of distinct evolutionary rate parameters for each of the evolutionary resolution level partitions provides a more adequate analysis than assuming the same evolutionary rate parameters for all sites or nucleotides. Second, the coarse level analysis runs faster and converges faster, thus guiding the intermediate level analysis, that then guides the fine level analysis. So, in this way there is an acceleration of the computations.

Our use of entropy to partition molecular data is a statistical data-driven alternative to current methods that partition the data in a more subjective manner. For example, it is current common practice to partition molecular data in a biologically informed manner according to the three codon positions or to separate introns from exons (Yang 1996). This current practice has two main objectives: to allow for distinct patterns of evolution at different nucleotide sites, and to ameliorate the impact of saturation at introns and third codon positions. However, the current practice is not enough when saturation occurs at sites not located at introns and third codon positions (Breinholt and Kawahara 2013). In contrast, our entropy-based methodology partitions the data according to scales of evolutionary resolution as estimated from the molecular data. Therefore, our entropy-based methodology is better able to account for saturation when needed and also allows for distinct patterns of evolution at different resolution levels.

There are many methods to account for evolutionary rate heterogeneity among sites. The most widely used way to deal with rate heterogeneity is to assume that the rate for each site follows a gamma distribution (Yang 1993, 1996). For computational reasons, this is usually implemented using a discrete gamma approximation Yang (1994). Another way to account for rate heterogeneity is to use discrete-class models that assume a small number of distinct rates. In particular, discrete-class models usually assume a class of invariant sites. Another way to account for rate heterogeneity is to partition the data according to biological considerations such as for example codon position or intron/exon position. In contrast, our proposal partitions the data according to estimated evolutionary resolution. Finally, within each evolutionary resolution partition we assume a discrete gamma model for the rates.

We illustrate the use of our multiscale phylogenetics framework with an application to a molecular dataset for primates previously analyzed by Perelman *et al*. (2011). There is a tremendous interest in the phylogenetic tree of primates, and several researchers have contributed to that literature. In particular, Purvis (1995) used existing methods of matrix representation and data on primates to reconstruct the phylogenetic relationships among primates. Others have reconstructed the primate phylogeny with supermatrix approaches, based on datasets that differ in their level of completeness of (a) number of species sampled from 16-76% and (b) percentage of matrix completeness from 18-ca. 100% (Perelman *et al*. 2011, Springer *et al*. 2012, Chatterjee *et al*. 2009, Finstermeier *et al*. 2013, Fabre *et al*. 2009). The phylogenetic tree built by Purvis (1995) was used to evaluate several distinct evolutionary issues in a wide range of areas such as molecular evolution (Smith *et al*. 2003), behavior (Nunn 2003), species diversity (Conroy 2003), imunology (Boniotto *et al*. 2003), and biodiversity conservation (Sechrest *et al*. 2002). Similarly, the tree built by Perelman *et al*. (2011) has been used to study the evolution of immunodeficiency viruses (Compton and Emerman 2013), body size (Dunham *et al*. 2013), muscle morphology (Diogo and Wood 2013), and chromosomes (de Oliveira *et al*. 2012). Therefore, the phylogenetic tree of primates is of wide interest and deserves to be studied using multiscale phylogenetics.

The remainder of the article is organized as follows. Section 2 introduces the concept of sitespecific entropy. Section 3 describes our multiscale phylogenetic framework. Section 4 illustrates the use of our multiscale phylogenetic framework with an application to the phylogeny of primates. Section 5 concludes with a discussion and possible future research directions.

2. SITE-SPECIFIC ENTROPY

This section introduces the concept of sitespecific entropy. In a DNA sequence each site or nucleotide may contain one of four bases: adenine (A), cytosine (C), guanine (G), and thymine (T). When we consider the molecular sequences of multiple species for the purpose of phylogenetic reconstruction, an initial preprocessing step is the alignment of the molecular sequences. After the sequences are aligned, it is assumed that each nucleotide in a given site across species has evolved from the same ancestral nucleotide. Nucleotides usually evolve at different rates depending on how crucial to survival are the proteins that they encode. For example, for the primates data that we analyze in Section 4, some nucleotides are tremendously well conserved and have not changed at all. Those well conserved nucleotides have the same bases across all considered species in the dataset and therefore do not provide any information about the phylogenetic tree. Next are nucleotides that evolve slowly and are reasonably well preserved but still present some variability across species. These nucleotides are important to resolve the parts of the phylogenetic tree closest to the ancestral root. Next, there are the nucleotides that have an intermediate rate of evolution. And finally, there are the nucleotides that have a fast rate of evolution and thus contain information about the most recent parts of the phylogenetic tree. Because the rate of evolution of a nucleotide is related to its variability, we may use a preprocessing step where we partition the nucleotides according to their variability and then apply Bayesian phylogenetic tree reconstruction with partition-specific rates of evolution. Here we propose the use of entropy as the measure of variability to be used to partition the nucleotides.

To be specific, consider site or nucleotide s and let p_{si}, $s = 1, ..., S$, $i = 1, ..., 4$, be the probability that a randomly chosen species will have at site s a base i, where $i = 1$ corresponds to adenine, $i = 2$ to cytosine, $i = 3$ to guanine, and i = 4 to thymine. Then, the entropy of site s is defined as

$$e_s = -\sum_{i=1}^{4} p_{si} \log(p_{si}), \tag{1}$$

where we use the convention that if for some pair (s, i) we have $p_{si} = 0$, then $p_{si} \log(p_{si}) = 0$.

In practice, we estimate the probability p_{si} by the relative frequency $\hat{p}_{si}$ of base i observed at site s. We then estimate the entropy of site s as

$$\hat{e}_s = -\sum_{i=1}^{4} \hat{p}_{si} \log(\hat{p}_{si}). \tag{2}$$

The smallest possible entropy for a site s is obtained when all the nucleotides at site s have the same base. In that case, the entropy will be equal to 0. The largest possible entropy is obtained when the distribution of the four possible bases is uniform, *i.e.*, when each of the four bases has probability of occurrence equal to 0.25. Thus, the maximum possible entropy is equal to 1.3863. In the application we present in Section 4 we use the maximum possible entropy as reference to define the multiscale evolutionary partitions.

3. MULTISCALE PHYLOGENETIC RECONSTRUCTION

In this section we describe the multiscale phylogenetic reconstruction that we propose. We start by defining multiscale evolutionary partitions based on the estimated site-specific entropies described in Section 2. Specifically, we consider L levels of evolutionary resolution where the lowest resolution level $l = 1$ corresponds to the slowest evolving nucleotides, and higher resolution levels correspond to faster evolving nucleotides. In particular, we define the several resolution levels based on thresholds $t_0 < t_1 < ... < t_L$ such that site s belongs to the lth level of evolutionary resolution if $t_{l-1} < \hat{e}_s \leq t_l$, $l = 1, ..., L$. Here $t_0 = 0$ is the lowest possible entropy value and t_L is the highest possible entropy value for a multinomial distribution with four cells. After each site is classified in different scales of evolutionary resolution, we have the dataset partitioned in L evolutionary partitions. One important issue is the choice of the number of levels of resolution or evolutionary partitions L. In this work we are going to consider three levels of evolutionary resolution: coarse, intermediate, and fine resolution levels.

The lowest possible entropy value $t_0 = 0$ is obtained by invariant sites that have no base variation. In the context considered here of phylogenetic reconstruction of a large set of species, this lack of base variability and resistance to change indicates that mutations at these sites may be strongly harmful or fatal. Thus, we assume that sites that have entropy equal to $t_0 = 0$ form a partition of their own. Moreover, for this zero-entropy partition the maximum likelihood estimate of the mutation rate is equal to zero. Note that the assumption that there is a rate class with a mutation rate equal to zero has been used by a number of authors in the phylogenetics literature (*e.g*., see Hasegawa *et al*. 1985, Palumbi 1989, Gu *et al*. 1995, Waddell and Steel 1997, Jia *et al*. 2014). Finally, because sites with zero mutation rate do not provide information about the phylogenetic tree, the zero-entropy sites are eliminated from the multiscale phylogenetic analysis.

Let $\boldsymbol{y}_l$ denote the molecular data for the sites that belong to the evolutionary partition l. In addition, let $\mathcal{T}$ be the underlying true phylogenetic tree and θ_l be the parameter vector for the evolutionary rate parameters for the evolutionary partition l, $l = 1, ..., L$. Here $\mathcal{T}$ includes the tree topology and the branch lengths. Assuming conditional independence of the partitions, the likelihood function admits the multiscale factorization

$$f(y_1, ..., y_L|\theta_1, ..., \theta_L, \mathcal{T}) = \prod_{l=1}^{L} f(\boldsymbol{y}_l|\theta_l, \mathcal{T}). \quad (3)$$

For a Bayesian analysis, one needs to assume priors on the multiscale evolutionary parameters $\theta_1, ..., \theta_L$ and on the phylogenetic tree $\mathcal{T}$ (*e.g*., see Ferreira and Suchard 2008). Let us denote those priors by $\pi(\theta_1), ..., \pi(\theta_L)$ and $\pi(\mathcal{T})$. Then, by Bayes Theorem the posterior density of $(\theta_1, ..., \theta_L, \mathcal{T})$ is

$$\pi(\theta_1, ..., \theta_L, \mathcal{T}|y_1, ..., y_L) \propto f(y_1, ..., y_L|\theta_1, ..., \theta_L, \mathcal{T}) \; \pi(\theta_1)\times \cdots \times \pi(\theta_L)\pi(\mathcal{T})$$

$$= \pi(\mathcal{T})\prod_{l=1}^{L} f(\boldsymbol{y}_l|\theta_l, \mathcal{T})\pi(|\theta_l), \quad (4)$$

where $\propto$ means "proportional to." Usually Bayesian analysis for phylogenetic models proceeds by simulating $(\theta_1, ..., \theta_L, \mathcal{T})$from the posterior distribution through the use of an MCMC algorithm such as that implemented in the phylogenetic software Mr. Bayes (Ronquist and Huelsenbeck 2003). Note that the posterior density given in Equation (4) admits a multiscale factorization similar to that of the likelihood function in Equation (3). We propose to explore this multiscale factorization by an MCMC algorithm with multiple stages.

Specifically, we start at the coarsest evolutionary resolution $l = 1$ by simulating $(\theta_1, \mathcal{T})$ from the coarsest level posterior density

$$\pi(\theta_1, \mathcal{T}|y_1) \propto f(y_1|\theta_1, \mathcal{T}) \; \pi(\theta_1) \; \pi(\mathcal{T}). \quad (5)$$

The coarsest level analysis runs and converges faster than the full dataset analysis. We use the output of the MCMC from the coarsest level analysis as starting point for the MCMC algorithm that simulates $(\theta_1, \theta_2, \mathcal{T})$ from the 1-to-2-level posterior density

$$\pi(\theta_1, \theta_2, \mathcal{T}|y_1, y_2) \propto f(y_1|\theta_1, \mathcal{T}) f(y_2|\theta_2, \mathcal{T}) \times\pi(\theta_1)\pi(\theta_2)\pi(\mathcal{T}). \quad (6)$$

We proceed in this manner using the output from the 1-to-$(l - 1)$-level MCMC as starting point for the MCMC algorithm that simulates $(\theta_1, ..., \theta_l, \mathcal{T})$ from the 1-to-l-level posterior density

$$\pi(\theta_1, ..., \theta_l, \mathcal{T}|y_1, ..., y_l) \propto \pi(\mathcal{T}) \prod_{j=1}^{L} f(\boldsymbol{y}_j|\theta_j, \mathcal{T})\pi(\theta_j), \quad (7)$$

until we have an MCMC output from the full dataset posterior density given in Equation (4). Note that for the full dataset analysis, there are distinct evolutionary rate parameters for each of the evolutionary resolution level partitions. This multiscale approach has two benefits: first, the use of distinct evolutionary rate parameters for each of the evolutionary resolution level partitions provides a more adequate analysis than assuming the same evolutionary rate parameters for all sites or nucleotides. Second, the coarser level analyses run faster and converge faster, thus guiding the finer level analyses. Therefore, proceeding in this multi-stage manner accelerates computations.

4. APPLICATION

We illustrate our multiscale phylogenetics framework with an application to a molecular dataset for primates previously analyzed by Perelman *et al.* (2011). To facilitate the computation of site-speciûc entropies, in our analysis all multistate characters were considered uncertainties. We could incorporate the multistate characters in the analysis by assuming that

for each species its possible characters were equally likely and by incorporating that information in the computation of the probability estimates $\hat{p}_{si}$. However, there are only 2,155 multistate characters in this dataset, comprising only 0.03% of the total number of characters. Because the number of multistate character sites is relatively small for the molecular dataset considered here, we do not incorporate the multi-state characters in the analysis. In addition, we have used the *R* package ape (Paradis *et al*. 2004) to read the molecular data and perform basic statistics computations. Further, we have used the software Mr. Bayes (Ronquist and Huelsenbeck 2003) to implement our multiscale phylogenetic framework.

The total number of sites in the dataset is 34,941 and the number of sites with entropy equal to zero is 18,337. As discussed in Section 3, we have removed sites with zero entropy from the analysis. Hence, the total number of sites used in our analysis is 16,604. To define the multiscale partitions, we have specified the threshold values in terms of the maximum possible entropy. Specifically, we have defined the three multiscale partitions as follows: the coarse level partition contains the sites that have entropy between 0 and 25% of the maximum possible entropy. The intermediate level partition contains the sites that have entropy between 25% and 50% of the maximum possible entropy. Finally, the fine level partition contains the sites that have entropy larger than 50% of the maximum possible entropy.

At all resolution levels, we have used the generalized time reversible (GTR) substitution model with gamma distributed rate variation across sites. To speed up computations, Mr. Bayes uses discrete categories to approximate the gamma distribution; in that regards, we have used the default number of four categories. Moreover, at each resolution level we have run two independent chains with 100,000 iterations per chain in a total of 200,000 iterations. A total of 60,000 iterations per chain were discarded as burn in. Further, to reduce the dependence between iterations and reduce the demand for computer storage, we have performed thinning with a sampling frequency of 1 simulation for each 100 iterations. Therefore, for each resolution level we have kept a total of 1,000 simulations for analysis. To further accelerate convergence, we use the Mr. Bayes implementation of Metropolis coupling where for each run we use four chains with a temperature parameter equal to 0.2. For each of the coarse level runs, we have used as initial tree a neighbor joining tree. After running the algorithm at the coarse level, we use the last sampled tree in each run at the coarse level as initial tree for each run at the intermediate level. Finally, after running the algorithm at the intermediate level, we use the last sampled tree in each run at the intermediate level as initial tree for each run at the fine level.

After running the algorithm at the fine level, we have a sample of the posterior distribution of tree topologies and evolutionary parameters for the GTR model. In particular, while the tree topology is common to the three resolution level partitions, the evolutionary parameters are specific to each partition. The estimated infinitesimal rate matrices for the three partition are similar to each other. However, Table 1 provides some evidence that the implied stationary distributions of nucleotides are somehow different for the different partitions. In particular, the stationary probability of thymine seems to be higher at finer resolution levels. Even more remarkably, Table 2 shows that the rate multiplier parameter differs substantially among the different resolution level partitions. In particular, the estimated rate multiplier at the fine level partition is four times the rate multiplier at the coarse level partition. Therefore, as intuitively expected, nucleotides at the fine resolution level partition undergo mutations at a faster rate.

Table 1. Stationary distribution of nucleotides per resolution level partition. Posterior mean (Mean) and standard deviation (SD).

Nucleotide	Coarse		Intermediate		Fine	
	Mean	SD	Mean	SD	Mean	SD
Adenosine	0.269	0.0023	0.262	0.0042	0.265	0.0088
Cytosine	0.240	0.0018	0.238	0.0034	0.227	0.0078
Guanine	0.236	0.0023	0.224	0.0042	0.228	0.0070
Thymine	0.255	0.0025	0.276	0.0047	0.280	0.0100

Table 2. Rate multiplier parameter. Posterior mean (Mean) and standard deviation (SD).

Partition	Mean	SD
Coarse	0.744	0.0044
Intermediate	1.602	0.0157
Fine	2.505	0.0525

Fig. 1 presents the phylogeny of primates based on our proposed multiscale phylogenetic reconstruction. The numbers beside each branch indicate the branch estimated posterior probability (X100). Specifically, green corresponds to the coarse resolution level, red to the intermediate level, and blue to the fine level. The absence of a number indicates estimated posterior probability equal to one at the corresponding resolution level. For example, no number beside a branch means that branch has an estimated posterior probability equal to one at all three resolution levels. As another example, if there is only a green number then the support at the coarse level is less than one, but the support at the intermediate and fine levels is equal to one. We find that the phylogeny obtained in our multiscale analysis is topologically identical to that obtained by Perelman *et al*. (2011). Differences are evident in the support for backbone (more inclusive/older) nodes. In particular, node support was lower in the coarse level resolution analysis. This result is not surprising because coarse level analysis includes the least variable sites. Finally, the intermediate and fine resolution level analyses provide similar support to tree nodes.

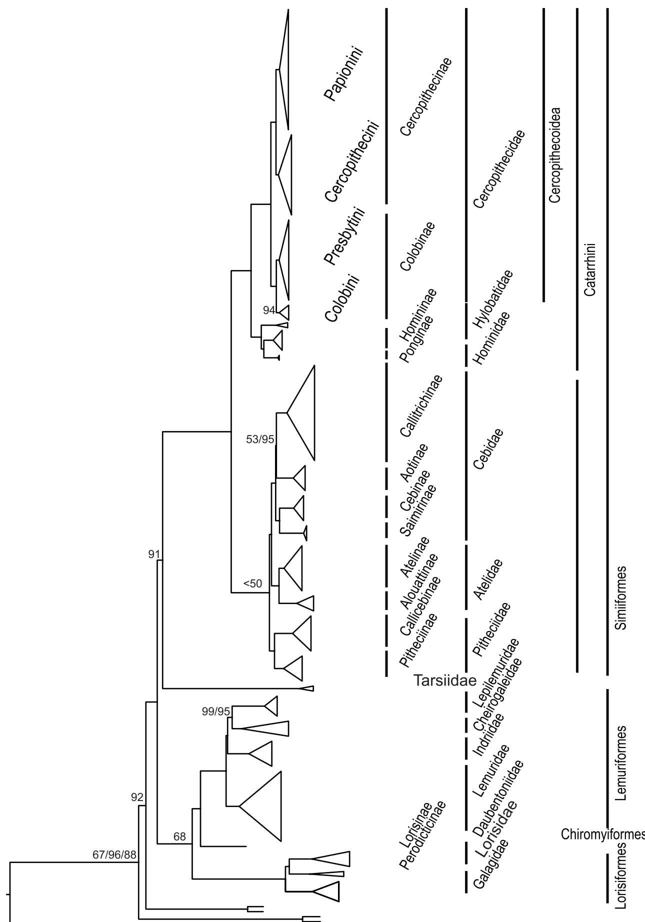

Fig. 1. Phylogeny of primates based on multiscale phylogenetic reconstruction. The numbers beside each branch indicate the branch posterior probability. Specifically, green corresponds to the coarse resolution level, red to the intermediate level, and blue to the fine level. The absence of a number indicates posterior probability equal to one at the corresponding resolution level.

5. CONCLUSIONS

We have made two main contributions in this manuscript. First, we have proposed the use of entropy as a statistical data-driven way to partition molecular data into distinct evolutionary scales of resolution. Second, we have used this multiscale partition to implement an MCMC scheme where results from coarser resolutions guide computations for finer resolutions to achieve faster convergence and accelerate computations.

We think multiscale phylogenetics may be useful for the construction of the tree of life. A particularly important challenge in the construction of the tree of life is the amount of missing molecular data. Specifically, molecular sequences datasets usually are not complete and contain large amounts of missing data for many different reasons. For example, some genes do not exist across all species. As another example, some genes do not vary enough to give information about phylogenies for species that are closely related and, as a result, biologists do not sequence such genes when studying closely related species. However, these genes are sequenced when biologists are interested in the relationship between species that are far apart in the tree of life. Further, genes that vary enough to give phylogenetic information for species that are closely related cannot be used to relate far apart species: some of these genes may not even exist for different species, or they may be so different among species that it is impossible to align them. Hence, the patterns of missing molecular data are directly related to the distinct evolutionary resolutions that govern the evolution of each nucleotide.

We think that a further developed multiscale phylogenetics methodology may help address these missing data issues. One way to deal with missing molecular data would be to incorporate morphological data in the analysis. Specifically, usually biologists decide which genes to sequence or not depending on their beliefs about the proximity of the species of interest in each particular study, and their beliefs depend

quite strongly on the morphological characteristics of the species of interest in each study. Thus, a possible solution is to impute the missing molecular data using a model that predicts the molecular data using the available data about the morphological characteristics of the several species. In addition, entropy can once again inform about the distinct evolutionary resolutions that govern the evolution of each nucleotide. Finally, this partitioning of the molecular dataset based on entropy together with the use of evolutionary parameters specific to each partition may lead to better estimates of the tree of life.

Finally, there have been many recent advances on the use of parallelization and multithreading using GPUs applied to phylogenetic studies to accelerate computations (Suchard and Rambaut 2009, Ayres *et al*. 2012, Bao *et al*. 2013). The development of the application of these advanced computational technologies to multiscale phylogenetic reconstruction is a promising research direction.

ACKNOWLEDGEMENTS

This work was supported in part by CNPq grant 402010/2003-5.

REFERENCES

Ayres, D.L. *et al*. (2012). BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Syst. Bio.*, **61,** 170-173.

Bao, J., Xia, H., Zhou, J., Liu, X. and Wang, G. (2013). Efficient implementation of Mr Bayes on multi-GPU. *Molecular Biol. Evol.*, **30,** 1471-1479.

Bininda-Emonds, O.R. (2004). *Phylogeneticsupertrees: Combining Information to Reveal the Tree of Life.* **4,** Springer.

Boniotto, M., Tossi, A., DelPero, M., Sgubin, S., Antcheva, N., Santon, D., Masters, J. and Crov-ella, S. (2003). Evolution of the beta defensin 2 gene in primates. *Genes Immunity*, **4(4)**, 251-257.

Breinholt, J.W. and Kawahara, A.Y. (2013). Phylotranscriptomics: Saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biol. Evol.*

Chatterjee, H., Ho, S., Barnes, I. and Groves, C. (2009). Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol. Biol.*, **9(1)**, 259.

Compton, A.A. and Emerman, M. (2013). Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. *PLoS Pathogens*, **9(1)**, e1003135.

Conroy, G.C. (2003). The inverse relationship between species diversity and body mass: do primates play by the rules? *J. Human Evol.*, **45(1)**, 43-55.

de Oliveira, E., Neusser, M. and Müller, S. (2012). Chromosome evolution in new world monkeys (Platyrrhini). *Cytogenetic Genome Res.*, **137(2-4),** 259-272.

Diogo, R. and Wood, B. (2013). The broader evolutionary lessons to be learned from a comparative and phylogenetic analysis of primate muscle morphology. *Biol. Rev.*, **88(4)**, 988-1001.

Dunham, A., Maitner, B., Razafindratsima, O., Simmons, M. and Roy, C. (2013). Body size and sexual size dimorphism in primates: influence of climate and net primary productivity. *J. Evol. Biol.*, **26**, 2312-2320.

Fabre, P.-H., Rodrigues, A. and Douzery, E. (2009). Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear {DNA}. *Molecular Phylogenetics Evol.*, **53(3),** 808-825.

Ferreira, M.A.R., Holan, S.H. and Bertolde, A.I. (2011). Dynamic multiscale spatio-temporal models for Gaussian areal data. *J. Roy. Statist. Soc.*, *Series B*, **73**, 663-688.

Ferreira, M.A.R. and Lee, H.K.H. (2007). *Multiscale Modeling: A Bayesian Perspective.* Springer Series in Statistics. Springer, New York.

Ferreira, M.A.R. and Suchard, M.A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *The Canad. J. Statist.,* **36,** 355-368.

Ferreira, M.A.R., West, M., Lee, H.K.H. and Higdon, D. (2006). Multi-scale and hidden resolution time series models. *Bayesian Anal.*, **1,** 947-968.

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M. and Roos, C. (2013). A mitogenomic phylogeny of living primates. *PloSONE*, **8(7)**, e69504.

Gu, X., Fu, Y.-X. and Li, W.-H. (1995). Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Molecular Biol. Evol.*, **12,** 546-557.

Hasegawa, M., Kishino, H. and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Molecular Evol.*, **22,** 160-174.

Jia, F., Lo, N. and Ho, S.Y.W. (2014). The impact of modelling rate heterogeneity among sites on phylogenetic

estimates of intraspecific evolutionary rates and timescales. *PLoSONE*, **9(5),** e95722.

Mishler, B.D. (1994). Cladistic analysis of molecular and morphological data. *Amer. J. Physical Anthropology*, **94(1),** 143-156.

Nunn, C.L. (2003). Behavioural defences against sexually transmitted diseases in primates. *Animal Behaviour*, **66(1),** 37-48.

Palumbi, S.R. (1989). Rates of molecular evolution and the fraction of nucleotide positions free to vary. *J. Molecular Evol.*, **29,** 180-187.

Paradis, E., Claude, J. and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.

Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A., Kessing, B., Pontius, J., Roelke, M. and Rumpler, Y. *et al*. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, **7(3)**, e1001342.

Purvis, A. (1995). A composite estimate of primate phylogeny. *Phil. Trans. Roy. Soc. London. Series B: Biological Sciences,* **348(1326),** 405-421.

Ronquist, F. and Huelsenbeck, J.P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19,** 1572-1574.

Sanyal, N. and Ferreira, M.A.R. (2012). Bayesian hierarchical multi-subject multiscale analysis of functional MRI data. *Neuro Image*, **63,** 1519-1531.

Sechrest, W., Brooks, T.M., da Fonseca, G.A., Konstant, W. R., Mittermeier, R.A., Purvis, A., Rylands, A.B. and Gittleman, J.L. (2002). Hotspots and the conservation of evolutionary history. *Proceedings of the National Academy of Sciences,* **99(4),** 2067-2071.

Smith, N.G., Webster, M.T. and Ellegren, H. (2003). A low rate of simultaneous double-nucleotide mutations in primates. *Molecularbiology Evol.*, **20(1),** 47-53.

Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, D.L., Stadler, T., Steiner, C., Ryder, O.A., Janecka, J. E., *et al*. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PloS ONE*, **7(1)**, e49521.

Suchard, M.A. and Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics*, **25**, 1370-1376.

Swenson, M.S., Suri, R., Linder, C.R. and Warnow, T. (2012). SuperFine: fast and accurate supertree estimation. *Systematic Biology*, **61(2),** 214-227.

Waddell, P.J. and Steel, M.A. (1997). General time-reversible distances with unequal rates across sites: mixing and inverse Gaussian distributions with invariant sites. *Molecular phylogenetics Evol.,* **8**, 398-414.

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when sub-stitution rates differ over sites. *Molecular Biol. Evol.,* **10**, 1396-1401.

— (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Molecular Evol.*, **39**, 306-314.

— (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.*, **11(9)**, 367-372.