The 10kTrees Website: A New Online Resource for Primate Phylogeny

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation	Arnold, Christian, Luke J. Matthews, and Charles Lindsay Nunn. 2010. The 10kTrees website: A new online resource for primate phylogeny. Evolutionary Anthropology 19(3): 114–118.
Published Version	doi:10.1002/evan.20251
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:4318029
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

The 10kTrees Website: A New Online Resource for Primate Phylogeny

Christian Arnold[1,2], Luke J. Matthews[1], and Charles L. Nunn[1] [*]

[1] *Harvard University, Department of Human Evolutionary Biology, Peabody Museum, 11 Divinity Avenue, Cambridge MA 02138*

[2] *Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*

 * to whom correspondence should be addressed (cnunn@oeb.harvard.edu, tel. 617-495-4710, fax 617-496-8041)

 17 pages, 1 figure, 0 tables

Key words: primate phylogeny, comparative method, Bayesian phylogenetics, phylogenetic uncertainty

About the authors

Christian Arnold received his Master's in Bioinformatics in June 2008. He then worked for one year at Harvard University in the Department of Human Evolutionary Biology. Since October 2009, he is a Ph.D. candidate at University of Leipzig, Germany. His bioinformatics research focuses on epigenetics, in particular non-coding RNAs and chromatin. Generally, he is interested in finding creative solutions for a wide variety of bioinformatic-related problems in phylogenetics, phylogenetic comparative methods, and epigenetics.

Luke Matthews is an anthropologist and primatologist whose interests include the evolution of primate behavior, primate traditions, and human cultural evolution. He has conducted primatological fieldwork in Ecuador and Argentina. He employs phylogenetics, network analysis, and cluster analysis to study inheritance systems that range from DNA sequences to primate foraging traditions and human material culture. Luke received his Ph.D. from New York University in Physical Anthropology in 2008. Since September 2008, he has been a Postdoctoral Fellow at Harvard University in Charles Nunn's laboratory.

Charles Nunn is an evolutionary anthropologist with interests in primate disease ecology, behavior, and cultural evolution. His research uses phylogenetic methods and mathematical modeling, and he has conducted fieldwork in Costa Rica and Madagascar. He is the author of *Infectious Diseases of Primates: Behavior, Ecology and Evolution*

(with Sonia Altizer) and a forthcoming book on *Phylogenetic Comparative Methods in Evolutionary Anthropology and Biology*. Charlie received his Ph.D. from Duke University in 1999 and has held positions at the University of California Berkeley and Davis, the Max Planck Institute, and University of Virginia. Currently, he is an Associate Professor at Harvard University.

The comparative method plays a central role in efforts to uncover the adaptive basis for primate behaviors, morphological traits and cognitive abilities.[1-4] The comparative method has been used, for example, to infer that living in a larger group selects for a larger neocortex,[5-6] that primate territoriality favors a longer day range relative to home range size,[7] and that sperm competition can account for the evolution of primate testes size.[8-9] Comparison is fundamental for reconstructing behavioral traits in the fossil record, for example in studies of locomotion and diet.[10-13] Recent advances in comparative methods require phylogenetic information,[2,14-16] but our knowledge of phylogenetic information is imperfect. In the face of uncertainty about evolutionary relationships, which phylogeny should one use? Here we provide a new resource for comparative studies of primates that enables users to run comparative analyses on multiple primate phylogenies Importantly, the 10,000 trees that we provide are not random, but instead use recent systematic methods to create a plausible set of topologies that reflect our certainty about some nodes on the tree and uncertainty about other nodes given the dataset. The trees also reflect uncertainty about branch lengths.

The comparative method has undergone a revolution in the past 20 years.[2,14-16] Specifically, new phylogenetic methods provide a way to incorporate evolutionary history directly into comparative research. Phylogeny is essential to comparative research because related species tend to resemble one another, resulting in non-independent data points.[2,17-18] Phylogenetic comparative methods can be used to investigate whether two traits change in tandem through time, while also providing the historical scaffolding to identify independent evolutionary origins of the traits of interest. More recently,

phylogenetic methods have provided a toolkit to investigate the tempo and mode of evolution,[19-20] to quantify phylogenetic signal in comparative data,[21-22] and to study the factors that influence diversification rates.[23-24] Computer simulations have revealed that it is usually preferable to conduct comparative tests with some form of phylogenetic method because this reduces false positives (Type I errors) and increases statistical power.[17-18,25-26] This latter point is often under-appreciated, but it is a logical outcome of phylogenetic comparative analyses that reduce error associated with the estimation of statistics and thus enhance the probability of detecting real effects.[26]

Researchers generally want to include as many species as possible in a comparative analysis. To incorporate phylogeny in comparative studies of primates, previous researchers have used either published primate-wide "supertrees" such as the Purvis phylogeny,[27] or they compiled smaller trees from the literature, often patching these together from among existing phylogenies based on morphology or genetics.[28-29] More recently, Bininda-Emonds et al.[30-31] produced a new supertree of mammals, and researchers have begun to use the primate portion of this tree in comparative studies of primates.[32-34]

The actual tree topology and timing of speciation events is, however, never known with certainty. In addition, phylogenetic relationships should be continually reassessed as new data become available, which recommends against the continued use of older phylogenies such as Purvis',[27] as better data are now available. Furthermore, when conducting a comparative test, it is desirable to incorporate the current level of uncertainty for specific nodes and branch lengths. Indeed, different trees can produce different results in a comparative analysis, and thus it is unwise to condition comparative

analyses on a single hypothesis of evolutionary relationships when that hypothesis is legitimately uncertain.[35]

Some evolutionary anthropologists have accounted for phylogenetic uncertainty by conducting multiple analyses using more than one tree.[28,36-38] But this raises an important question: how should we systematically decide on the trees to use? A number of authors have proposed that Bayesian phylogenetic approaches provide a way to systematically incorporate phylogenetic uncertainty into comparative research.[35,39-40] In particular, Bayesian methods allow the user to obtain a set of trees that are sampled in proportion to their posterior probability (see Box 1). The set of trees obtained reflects uncertainty in the phylogeny given the substitution model and data; more certain nodes are found across a greater proportion of the sample of trees, while less certain nodes are found less often. The user can create as many trees as he or she wishes – hundreds, even thousands of phylogenies, all fully bifurcating and with branch lengths, and not simply random permutations of the species in the study.[41-42] It is possible to then run comparative analyses on this sample of trees, and in this way the results of a comparative study are not conditioned on a particular phylogeny or set of branch lengths.[40]

Here we describe a new online resource for comparative studies of primates, which we call *10kTrees* and make available at http://10kTrees.fas.harvard.edu. The *10kTrees* website provides a way for users to download up to 10,000 primate phylogenies with branch lengths obtained from a Bayesian phylogenetic analysis. The trees in Version 1 include 189 primate species that are commonly used in comparative research on primates. The data are provided in a standard format[43] that can be read by a wide variety of comparative methods programs,[44-45] including phylogenetics packages for *R*.[46] We

request that people using this resource cite two papers: the current paper, which describes the *10kTrees* website and Version 1 of the trees, and another paper that applies Version 1 to study primate diversification in relation to body mass.[47]

A Bayesian inference of primate phylogeny. Bayesian phylogenetic methods provide a way to sample a set of trees in proportion to their posterior probabilities using Metropolis-coupled Markov chain Monte Carlo (MCMCMC, or MC^3) algorithms (see Box 1).[39,48-50] Nodes that the data strongly support are identical or nearly so across most of the "tree block" (i.e., the sample of trees obtained from the analysis). On the consensus tree of the *10kTrees* block, these nodes are thus depicted with high clade credibility values. Some nodes are not well supported, which indicates that alternative arrangements produce similar likelihoods, and these nodes vary across the tree block in proportion to their posterior probabilities. ***By running comparative analyses across this set of trees rather than using a single tree, the results are no longer conditioned on a single tree being correct.***

We expect this resource to be especially important for primate phylogenetic comparative studies because it provides a statistically rigorous and principled way to control for uncertainty at various nodes in primate phylogeny.[51-55] Importantly, our goal for this project is not to produce the definitive primate phylogeny; that goal will be best achieved with more focused studies of gene insertions, whole genomes, and standardized data collection, and will involve longer-term concerted effort by experts in primate phylogenetics (e.g., ref. 56). Instead, ***the goal is to produce a set of phylogenetic trees from available data that is appropriate for comparative research on primates***. We will,

however, regularly update the dataset to accommodate ever-increasing availability of sequence data and advances in tree inference methods.

For Version 1, we collected data on four mitochondrial genes and one autosomal gene from GenBank. To create the multiple sequence alignments (MSA), we used Muscle 3.7 with the default parameters.[57] Because alignment quality can have a substantial impact on the inferred tree,[58-62] we manually excluded poorly aligned sites or sites with a high percentage of missing data (especially at the beginning and end of the MSA). We constrained 29 major nodes if they were well characterized by at least three genomic Alu insertions.[63-69] These constraints eliminate uncertainty at constrained nodes, which we think is reasonable because Alu insertion events are generally regarded as more reliable cladistic indicators that are less prone to homoplasy than DNA sequence data.[63-65] However, we are likely to relax these constraints in future versions of *10kTrees* that use additional autosomal loci (and thus users should refer to the website for details appropriate for the version they use).

For tree inference in Version 1, we used the program MrBayes 3.1.2.[48] *Galeopterus variegatus* (Sunda flying lemur or colugo) was identified as the outgroup, as it has been shown that colugos are the closest living relatives to the order Primates.[70] We ran a Bayesian analysis with two runs and eight chains (one cold chain and seven heated chains) in each run. We used a GTR+I+G substitution model for each of the five genes in a partitioned dataset, which was identified as the best substitution model in the program FindModel.[71] The analysis for Version 1 was run for eight million generations, with trees sampled every 1,000 generations. We assessed the heating (changed to 0.02) and excluded the first three million generations as burn-in (see Box 1). We summarized these

topologies by constructing a 50% majority rule consensus tree, which we provide in various graphical formats on the *10kTrees* website. Branch lengths were calculated as the mean branch length from all trees in the posterior distribution in which the branch was present. In future versions of *10kTrees*, we will provide both molecular branch lengths and, by using fossil calibration points, branches that reflect the time since two species last shared a common ancestor.

Applications to primate comparative biology. *An important goal of our project is to make the trees readily available for comparative research.* To that end, users can download trees in NEXUS[43] format. On the *10kTrees* website, users can select the number of trees to download. These are sampled from the tree block such that they cover full range of variation in the analysis (i.e., they are sampled evenly along the stored chain of trees, rather than simply taking the first *n* trees in the sample, where *n* is the number of trees requested by the user). The consensus tree of the full sample is also available to download. In addition, the user has the option to select specific species of interest; the trees are then pruned to the selected species prior to download. In terms of the actual data used to generate the trees, users can obtain the original sequence data, the list of species that were studied, an availability matrix for the distribution of genetic data across species, and details on how the data were analyzed (i.e., the substitution model and parameters of the MC^3 analysis, such as sampling rate, number of chains, and number of sampled generations). The phylogenetic constraints based on Alu insertions are also downloadable from the website. By providing the raw data files, users can easily rerun the phylogenetic analysis in a different computer package or with different settings, including without the constraints.

The phylogenies available from *10kTrees* provide a significant improvement over the two primate-wide phylogenies that are currently most commonly used in comparative research.[27,30] The older of these phylogenies, published by Andy Purvis in 1995[27], was a ground-breaking contribution to comparative primatology when it was published. However, fewer genetic data were available in 1995, and many polytomies were present in the tree. In particular, the Asian colobines showed almost no resolution because phylogenetic information for this clade was generally unavailable when the tree was constructed. In addition to its lack of resolution, the Purvis phylogeny includes a number of topological misplacements even at the generic level as assessed by comparison to the preponderance of DNA sequence and Alu insertion data used to infer the *10kTrees*. For example, the basal bifurcation of the platyrrhines is incorrect on the basis of Alu insertions, as are the positions of *Aotus* and *Callicebus*.[65] Recent sequence-based studies and our research clearly contradict the topology given for other genera such as *Callimico* and *Lophocebus*.[72-73] Given its lack of resolution and topological misplacements at the generic level, the continued use of the Purvis phylogeny[27] cannot be recommended.

The primate portion of the more recent Bininda-Emonds supertree[30] is an improvement over Purvis' phylogeny,[27] but it still suffers from excessive polytomies within the guenons and the Asian colobines. The Bininda-Emonds tree contains an incorrect topology for the basal bifurcations among the platyrrhines given the Alu insertion data and whole mitochondrial genome evidence.[65,72] It also includes apparent anomalies or errors, such as the paraphyletic placement of *Callicebus personatus* as separate from its congeners and basal to Cebidae. While the Bininda-Emonds tree is

useful for mammal-wide comparative analyses, we recommend *10kTrees* for studies focused on primates.

We designed the website so that it can be easily updated as new genetic data become available to infer primate phylogeny. Thus, the website will provide multiple versions, including an "archive" where previous versions can be accessed. We are already working on Version 2 of the dataset, which will include over 230 species and more genes. We also expect that the **website itself will evolve to provide more tools for primate comparative biology**. In future versions, for example, we plan to provide a taxonomic translation tool. Thus, readers will be able to select species based on their names from GenBank, or from lists of names in which the original species designations are translated to commonly used taxonomies, such as the taxonomies by Corbet and Hill[74] and Groves in Wilson and Reeder.[75] In addition to giving trees with branch lengths proportional to genetic change, a future version of the website will provide dated trees based on fossil calibration points. We will also make available scripts to link the trees to comparative data to produce a single NEXUS[43] file for analysis, and we will provide advice for importing the trees into analysis programs, such as BayesTraits,[44] R[76] and Mesquite.[45] Ultimately, we aim for a flexible, open-access and user-friendly platform that will enhance the use of phylogenetic approaches in primate evolution, and that will grow as new sequence data and methods become available.

References

1 Clutton-Brock TH, Harvey PH. 1977. Primate ecology and social organization. Journal of Zoology 183:1-39.

2 Nunn CL, Barton RA. 2001. Comparative methods for studying primate adaptation and allometry. Evolutionary Anthroplogy 10:81-98.

3 Martin RD. 1990. Primate Origins and Evolution. London: Chapman and Hall.

4 Plavcan JM, van Schaik CP. 1992. Intrasexual Competition and Canine Dimorphism in Anthropoid Primates. American Journal of Physical Anthropology 87(4):461-477.

5 Dunbar RIM. 1998. The social brain hypothesis. Evolutionary Anthropology 6(5):178-190.

6 Barton RA. 1996. Neocortex size and behavioural ecology in primates. Proceedings Royal Society London, B 263:173-177.

7 Mitani JC, Rodman PS. 1979. Territoriality: the relation of ranging pattern and home range size to defendability, with an analysis of territoriality among primate species. Behavioral Ecology and Sociobiology 5:241-251.

8 Harcourt AH, Harvey PH, Larson SG, Short RV. 1981. Testis weight, body weight and breeding system in primates. Nature 293:55-57.

9 Harcourt AH, Purvis A, Liles L. 1995. Sperm competition: mating system, not breeding season, affects testes size of primates. Functional Ecology 9:468-476.

10 Spoor F, Garland T, Krovitz G, Ryan TM, Silcox MT, Walker A. 2007. The primate semicircular canal system and locomotion. Proceedings of the National Academy of Sciences of the United States of America 104(26):10808-12.

11 Jungers WL, Godfrey LR, Simons EL, Chatrath PS. 1997. Phalangeal curvature and positional behavior in extinct sloth lemurs (Primates, Palaeopropithecidae). Proceedings of the National Academy of Sciences of the United States of America 94(22):11998-12001.

12 Napier JR, Walker AC. 1967. Vertical Clinging and Leaping - a Newly Recognized Category of Locomotor Behaviour of Primates. Folia Primatologica 6(3-4):204-&.

13 Kay RF. 1984. On the use of anatomical features to infer foraging behavior in extinct primates. In: Rodman P, Cant J, editors. Adaptations for Foraging in Nonhuman Primates. New York: Columbia University Press.

14 Garland T, Harvey PH, Ives AR. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Systematic Biology 4:18-32.

15 Felsenstein J. 1985. Phylogenies and the comparative method. American Naturalist 125(1):1-15.

16 Harvey PH, Pagel MD. 1991. The Comparative Method in Evolutionary Biology. Oxford: Oxford University Press. 239 p.

17 Martins EP, Garland T. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534-557.

18 Purvis A, Gittleman JL, Luh H. 1994. Truth or consequences: effects of phylogenetic accuracy on two comparative methods. Journal of Theoretical Biology 167:293-300.

19 Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zoologica Scripta 26(4):331-348.

20 Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401(6756):877-884.

21 Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution 57(4):717-745.

22 Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. American Naturalist 160(6):712-726.

23 Nee S, Holmes EC, May RM, Harvey PH. 1994. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 344(1307):77-82.

24 Maddison WP. 2006. Confounding Asymmetries in Evolutionary Diversification and Character Change. Evolution 60(8):1743-1746.

25 Symonds MRE. 2002. The effects of topological inaccuracy in evolutionary trees on the phylogenetic comparative method of independent contrasts. Systematic Biology 51(4):541-553.

26 Rohlf FJ. 2006. A comment on phylogenetic correction. Evolution 60(7):1509-1515.

27 Purvis A. 1995. A composite estimate of primate phylogeny. Philosophical Transactions of the Royal Society, London, Series B 348:405-421.

28 Mitani JC, GrosLouis J, Manson JH. 1996. Number of males in primate groups: Comparative tests of competing hypotheses. American Journal of Primatology 38(4):315-332.

29 Smith RJ, Cheverud JM. 2002. Scaling of sexual dimorphism in body mass: A phylogenetic analysis of Rensch's rule in primates. International Journal of Primatology 23(5):1095-1135.

30 Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. Nature 446:507-512.

31 Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2008. Corrigendum: The delayed rise of present-day mammals. Nature in press.

32 Davies TJ, Pedersen AB. 2008. Phylogeny and geography predict pathogen community similarity in wild primates and humans. Proceedings of the Royal Society B-Biological Sciences 275(1643):1695-1701.

33 Walker RS, Gurven M, Burger O, Hamilton MJ. 2008. The Trade-off Between Number and Size of Offspring in Humans and Other Primates. Proceedings of the Royal Society B: Biological Sciences 275(1636):827-833.

34 Nunn CL, McNamara P, Capellini I, Preston P, Barton RA. 2009. Primate sleep in phylogenetic perspective. In: McNamara P, Barton RA, Nunn CL, editors. Evolution of sleep: Phylogenetic and Functional Perspectives. Cambridge: Cambridge University Press.

35 Lutzoni F, Pagel M, Reeb V. 2001. Major fungal lineages are derived from lichen symbiotic ancestors. Nature 411(6840):937-940.

36 Kamilar J, Martin S, Tosi A. 2009. Combining Biogeographic and Phylogenetic Data to Examine Primate Speciation: An Example Using Cercopithecin Monkeys. Biotropica 41(4):514-519.

37 Thierry B, Iwaniuk A, Pellis S. 2000. The influence of phylogeny on the social behaviour of macaques (Primates: Cercopithecidae, genus *Macaca*). Ethology 106(8):713-728.

38 Nunn CL. 1999. The number of males in primate social groups: a comparative test of the socioecological model. Behavioral Ecology and Sociobiology 46:1-13.

39 Huelsenbeck JP, Rannala B, Masly JP. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science 288(5475):2349-2350.

40 Pagel M, Lutzoni F. 2002. Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. In: Lässig M, Valleriani A, editors. Biological Evolution and Statistical Physics. Berlin: Springer-Verlag. p 148-161.

41 Martins EP. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. Evolution 50:12-22.

42 Abouheif E. 1998. Random trees and the comparative method: A cautionary tale. Evolution 52(4):1197-1204.

43 Maddison DR, Swofford DL, Maddison WP. 1997. Nexus: An extensible file format for systematic information. Systematic Biology 46(4):590-621.

44 Pagel M, Meade A. 2007. BayesTraits (www.evolution.rdg.ac.uk). Version 1.0. Reading, UK.

45 Maddison WP, Maddison DR. 2006. Mesquite: a modular system for evolutionary analysis. Version 2.5: http://mesquiteproject.org.

46 R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

47 Matthews LJ, Arnold C, Machanda Z, Nunn C. 2009. Body size effects on speciation and extinction rates in primates. In prep.

48 Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

49 Larget B, Simon DL. 1999. Markov Chasin Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees. Mol Biol Evol 16(6):750-759.

50 Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. Mol Biol Evol 14(7):717-724.

51 Martin RD. 2008. Evolution of Placentation in Primates: Implications of Mammalian Phylogeny. Evolutionary Biology 35(2):125-145.

52 Ting N, Tosi AJ, Li Y, Zhang YP, Disotell TR. 2008. Phylogenetic incongruence between nuclear and mitochondrial markers in the Asian colobines and the evolution of the langurs and leaf monkeys. Molecular Phylogenetics and Evolution 46(2):466-474.

53 Finarelli JA, Clyde WC. 2004. Reassessing hominoid phylogeny: evaluating congruence in the morphological and temporal data. Paleobiology 30(4):614-651.

54 Schneider H. 2000. The current status of the New World monkey phylogeny. Anais Da Academia Brasileira De Ciencias 72(2):165-172.

55 Schmitz J, Roos C, Zischler H. 2005. Primate phylogeny: molecular evidence from retroposons. Cytogenet Genome Res 108(1-3):26-37.

56 Disotell TR. 2008. Primate Phylogenetics. Encyclopedia of Life Sciences. Chinchester: John Wiley and Sons, Ltd.

57 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32(5):1792.

58 Ogden TH, Rosenberg MS. 2006. Multiple sequence alignment accuracy and phylogenetic inference. Systematic Biology 55(2):314-328.

59 Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56(4):564-577.

60 Smythe AB, Sanderson MJ, Nadler SA. 2006. Nematode small subunit phylogeny correlates with alignment parameters. Systematic Biology 55(6):972-992.

61 Kjer KM. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Molecular Phylogenetics and Evolution 4(3):314-330.

62 Morrison DA, Ellis JT. 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of Apicomplexa. Molecular Biology and Evolution 14(4):428-441.

63 Ray DA, Batzer MA. 2005. Tracking Alu evolution in New World primates. BMC Evolutionary Biology 5(1):51.

64 Ray DA, Xing JC, Hedges DJ, Hall MA, Laborde ME, Anders BA, White BR, Stoilova N, Fowlkes JD, Landry KE and others. 2005. Alu insertion loci and platyrrhine primate phylogeny. Molecular Phylogenetics and Evolution 35(1):117-126.

65 Xing JC, Witherspoon DJ, Ray DA, Batzer MA, Jorde LB. 2007. Mobile DNA elements in primate and human evolution. American Journal of Physical Anthropology:2-19.

66 Xing J, Wang H, Han KD, Ray DA, Huang CH, Chemnick LG, Stewart CB, Disotell TR, Ryder OA, Batzer MA. 2005. A mobile element based phylogeny of Old World monkeys. Molecular Phylogenetics and Evolution 37(3):872-880.

67 Schmitz J, Ohme M, Zischler H. 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157(2):777-784.

68 Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA. 2003. Alu elements and hominid phylogenetics. Proceedings of the National Academy of Sciences of the United States of America 100(22):12787-12791.

69 Roos C, Schmitz J, Zischler H. 2004. Primate jumping genes elucidate strepsirrhine phylogeny. Proceedings of the National Academy of Sciences 101(29):10650-10654.

70 Janecka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ. 2007. Molecular and genomic data identify the closest living relative of primates. Science 318(5851):792-794.

71 Tao N, Richardson R, Bruno W, Kuiken C. 2005. FindModel (http://hcv.lanl.gov/content/hcv–db/findmodel/findmodel.html).

72 Hodgson JA, Sterner KN, Matthews LJ, Burrell AS, Jani RA, Raaum RL, Stewart CB, Disotell TR. 2009. Successive radiations, not stasis, in the South American primate fauna. Proceedings of the National Academy of Sciences of the United States of America 106:5534-5539.

73 Burrell A, Jolly C, Tosi A, Disotell T. 2009. Mitochondrial evidence for the hybrid origin of the kipunji, *Rungwecebus kipunji* (Primates: Papionini). Molecular Phylogenetics and Evolution 51(2):340-348.

74 Corbet GB, Hill JE. 1991. A world list of mammalian species. Oxford: Oxford University Press.

75 Wilson DE, Reeder DM. 2005. Mammal Species of the World: Johns Hopkins University Press.

76 R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Text Box and Figure

Box 1. Schematic of Bayesian phylogenetics. Bayesian methods in phylogenetics typically use Metropolis-coupled Markov chain Monte Carlo algorithms (MCMCMC, or MC^3) to generate posterior probability distributions for a set of parameters; that is, the conditional distribution of the parameter given the data. Note that in Bayesian statistics, probability cannot be interpreted in its original meaning; rather, it is used to represent parameter uncertainty. In Bayesian phylogenetics, the parameters are comprised of a phylogenetic tree and a specific model of evolution, which is based on the individual priors for these parameters (*a priori* knowledge or beliefs about a parameter distribution) and the likelihood of the data.

Bayesian MC^3 methods start with a random tree and arbitrary initial values for branch lengths and model parameters. In each generation, either a new tree or a new model parameter is proposed. Typically, each proposed change is small; thus, the samples are not random because they are based on the parameter value of the previous generation. The proposal can then be accepted or rejected, depending on the ratio of the posterior densities of the new state to the old state (R). If R>1 (an "uphill" step, indicating a state with higher posterior probability) the proposed change is always accepted. If R<1 (a "downhill" step), the change is accepted with probability R. The smaller R, the smaller is the proposal acceptance probability. Thus, after a particular number of generations, the region of the parameter space with the highest posterior probability is reached and sampled most often. If the parameter space has multiple peaks that are separated by deep valleys, however, the algorithm may become stuck on a local peak. To rectify that

problem, additional chains (so-called hot or heated chains) are used in MC^3 that independently search the tree space, more readily accept proposals with a small R value, and regularly swap states with the original chain (cold chain). Heated chains flatten the posterior probability distribution and thus have shallower valleys, which allows them to more easily cross those valleys.

During MC^3, a "chain" of trees is produced that reflects the accepted modifications in phylogenetic information and model parameters. At the beginning of the chain, the likelihood typically climbs quickly (which is called burn-in, see Figure) until the chain eventually reaches its equilibrium distribution (indicated by the putative plateau in the Figure). Every k generations, the cold chain is sampled, which simply entails saving the tree topology, branch lengths and all model parameters, and further analyses generally make use of post-burnin samples. If the posterior probability distribution for each parameter has been approximated adequately, their post-burn-in sampling reflects the true parameter uncertainty, and the quality of the sample generally improves as a function of the number of steps (generations). The sample can then be summarized using various statistics, such as histograms, means, or credible intervals. Topology and branch lengths can also be summarized by constructing a majority rule consensus tree with support values in the form of clade credibility values and mean branch lengths. One can also use the whole sample of trees ("tree block") to incorporate topological and branch length uncertainty into comparative analyses that make use of the phylogenetic information. This is the approach that we advocate in this article and is also recommended by others.[39-40]