

# Using the BirdTree.org website to obtain robust phylogenies for avian comparative studies: A primer

Diego RUBOLINI<sup>1\*</sup>, András LIKER<sup>2</sup>, László Z. GARAMSZEGI<sup>3</sup>,  
Anders P. MØLLER<sup>4</sup>, Nicola SAINO<sup>1</sup>

<sup>1</sup> Dipartimento di Bioscienze, Università degli Studi di Milano, via Celoria 26, I-20133 Milano, Italy

<sup>2</sup> Department of Limnology, University of Pannonia, Pf. 158, H-8201 Veszprém, Hungary

<sup>3</sup> Department of Evolutionary Ecology, Estación Biológica de Doñana-CSIC, c/ Americo Vespucio s/n, E-41092 Seville, Spain

<sup>4</sup> Laboratoire Ecologie, Systematique et Evolution, UMR 8079 CNRS, Université Paris-Sud XI-AgroParisTech, Batiment 362 Université Paris-Sud XI, F-91405 Orsay Cedex, France

**Abstract** Comparative studies of trait evolution require accounting for the shared evolutionary history. This is done by including phylogenetic hypotheses into statistical analyses of species' traits, for which birds often serve as excellent models. The online publication of the most complete molecular phylogeny of extant bird species ([www.birdtree.org](http://www.birdtree.org), BirdTree hereafter) now allows evolutionary biologists to rapidly obtain sets of equally plausible phylogenetic trees for any set of species to be incorporated as a phylogenetic hypothesis in comparative analyses. We discuss methods to use BirdTree tree sets for comparative studies, either by building a consensus tree that can be incorporated into standard comparative analyses, or by using tree sets to account for the effect of phylogenetic uncertainty. Methods accounting for phylogenetic uncertainty should be preferred whenever possible because they should provide more reliable parameter estimates and realistic confidence intervals around them. Based on a real comparative dataset, we ran simulations to investigate the effect of variation in the size of the random tree sets downloaded from BirdTree on the variability of parameter estimates from a bivariate relationship between mass-specific productivity and body mass. Irrespective of the method of analysis, using at least 1,000 trees allows obtaining parameter estimates with very small (< 0.15%) coefficients of variation. We argue that BirdTree, due to the ease of use and the major advantages over previous 'traditional' methods to obtain phylogenetic hypotheses of bird species (e.g. supertrees or manual coding of published phylogenies), will become the standard reference in avian comparative studies for years to come [*Current Zoology* 61 (6): 959–965, 2015].

**Keywords** Birds, BirdTree, Comparative method, PGLM, PGLS, Phylogenetic uncertainty

## 1 Introduction

The study of trait evolution, a landmark of evolutionary biology, requires the reconstruction of ancestral character states and the mapping of character evolution on phylogenetic trees, and it requires the use of comparative methods (Harvey and Pagel, 1991). These methods allow the comparison of traits from different populations or species varying in their evolutionary history, and they are powerful tools to distinguish patterns of phenotypic convergent evolution and adaptation from patterns caused by common phylogenetic descent (Harvey and Pagel, 1991).

Various statistical methods have been devised to account for shared evolutionary history in comparative studies testing evolutionary hypotheses by analyzing current traits of species (e.g. Felsenstein, 1985; Grafen, 1989; Møller and Birkhead, 1992; Pagel, 1999; Maddi-

son, 2000; Freckleton et al., 2002; Nunn, 2011; Paradis, 2012; Garamszegi, 2014). Although the numerical computations may differ depending on whether the traits of interest are discrete or continuous (Maddison, 1990; Harvey and Pagel, 1991), these statistical methods require a phylogenetic hypothesis as a basis for making inferences. The most commonly used techniques to overcome the problem of non-independence of closely related taxa has been the independent contrast method (Felsenstein, 1985), and more recently, the approach based on generalized least-squares (Grafen, 1989; Martins and Hansen, 1997). Irrespective of the specific method of analysis, a phylogenetic hypothesis in the form of tree topology and branch lengths is essential for making robust statistical inferences about character (co) evolution.

Phylogenetic hypotheses used in comparative studies are usually a combination of phylogenetic trees ob-

---

Received May 28, 2014; accepted Sep. 22, 2014.

\* Corresponding author. E-mail: [diego.rubolini@unimi.it](mailto:diego.rubolini@unimi.it)

© 2015 *Current Zoology*

tained from various literature sources, because complete phylogenetic trees including the entire set of species of interest for a given study are seldom available. Combinations of phylogenetic trees are often made by hand using a qualitative evaluation of published phylogenies (e.g. Garamszegi and Gonzalez-Voyer, 2014) or via quantitative methods such as the ‘supertree’ approach (e.g. Thomas et al., 2004; Jönsson et al. 2006; Bininda-Emonds, 2014), which may be useful to obtain very large phylogenies for the set of species under scrutiny starting from a set of phylogenies from different subsets of species, built using a variety of primary sources (e.g. DNA-DNA hybridization, molecular data, or morphology).

Compared to other taxa, birds are a favourite subject of comparative studies, mostly due to the huge variability and thorough knowledge of their life-history traits and behaviour (see e.g. Bennett and Owens, 2002). In addition, their phylogenetic relationships are rather well known: in the early 1990’s, Sibley and Ahlquist (1991) published a so-called ‘tapestry’ (or ‘backbone’) phylogeny, based on DNA-DNA hybridization that has constituted the foundation of most large-scale comparative analyses of this class carried out so far. Later, the Sibley and Ahlquist phylogeny has been supplemented (and often challenged) with extensive molecular studies by Barker et al. (2002), Ericson et al. (2006) and Hackett et al. (2008), the latter being perhaps the most complete study of the deep phylogenetic relationships among extant bird clades to date. While the deep phylogenetic relationships of bird species were touched upon by only a few major reference works, thanks to the advances and the ever-increasing availability of molecular biology techniques, the past 20 years have seen a flourishing of small-scale studies addressing phylogenetic relationships within avian genera, families, or orders, generally based on molecular analyses of a set of conserved genes. Recently, Jetz et al. (2012) assembled this huge amount of molecular information to investigate patterns of speciation in avian lineages through their evolutionary history. To do so, they assembled what is definitely the most complete molecular phylogeny of all 9,993 extant bird species. The paper was accompanied by a website (<http://www.birdtree.org>, BirdTree hereafter), which allows users to freely download the original data, either in the form of molecular sequences or phylogenetic trees. Remarkably, the authors also developed a tool to easily select a subset of any of the extant bird species and to download phylogenetic trees for that set of species. Within a matter of minutes, researchers aiming at conducting comparative studies of birds can thus

obtain complete and validated phylogenies from any set of extant species. This is a tremendous advance over previously available methods to obtain phylogenies for comparative analyses, such as supertree building, which necessitates considerable phylogenetic software as well as molecular biology knowledge, or manual coding of existing phylogenies from published studies, which is time-consuming and error-prone. As is already occurring with a similar online tool providing primate as well as other mammal phylogenies, the 10kTrees project (Arnold et al., 2010; see <http://10ktrees.fas.harvard.edu>), we assume that phylogenetic trees downloaded from BirdTree will be used as a standard reference in avian comparative studies for years to come, and here we discuss methods that might be used to incorporate such trees in phylogenetic analyses of trait (co)evolution, such as phylogenetically-corrected multiple regression models (e.g. phylogenetic generalized least-squares (PGLS) models; Pagel, 1997; Freckleton et al., 2002). We then ran simulations, based on a real comparative dataset, to investigate the variability in parameter estimates according to the number of trees downloaded from BirdTree, in order to provide guidance on the number of trees required to obtain robust statistical inferences.

## 2 How to Use Phylogenetic Trees Downloaded from BirdTree

Jetz et al. (2012) built a complete tree of all extant species by combining time-calibrated trees of deep avian relationships with separately inferred trees of well-supported constituent clades, using genetic data from a total of 6,663 taxa. As a backbone for their phylogenetic reconstruction, Jetz et al. (2012) used the two major reference works by Hackett et al. (2008) and Ericson et al. (2006), which differ in their phylogenetic reconstruction of the Neoaves according to the exclusion or inclusion of the  $\beta$ -fibrinogen gene, whose usefulness in avian phylogenetic studies has been debated (Morgan-Richards et al., 2008). These two ‘backbone’ phylogenies were supplemented with additional topological constraints from well-resolved phylogenies of major avian clades (e.g. Barker et al., 2004). All existing species were assigned to one of the major clades (‘crown clades’) of the backbone phylogeny. Genetic data from multiple loci, retrieved from GenBank, were used to reconstruct the phylogeny of crown clades and to combine these with the backbone phylogenies in a Bayesian framework, using a number of additional constraints and informed priors [thoroughly described in the supplement-

tary material of Jetz et al. (2012)]. Using BEAST (Drummond and Rambaut, 2007), Jetz et al. (2012) obtained two sets of ultrametric, fully resolved (bifurcating) Markov chain Monte Carlo (MCMC) trees of all bird species, either based on the Hackett et al. (2008) or the Ericson et al. (2006) backbone phylogeny, and made 10,000 post-burn-in trees freely available to download from the BirdTree online platform. These tree sets are named ‘Hackett All Species’ and ‘Ericson All Species’ on the BirdTree website. Users are allowed to download random sets of trees (up to 9999 trees, in NEXUS multiple tree format) for any combination of all the extant bird species either based on the ‘Hackett’ or the ‘Ericson’ backbone phylogeny. BirdTree does not provide any consensus tree, as phylogenetic hypotheses are formed as sets of equally plausible (but variable) trees.

We suggest ways by which tree sets downloaded from BirdTree can be implemented in comparative phylogenetic analyses. A basic approach would be to summarize a tree set into a single consensus tree that can be incorporated as a phylogenetic hypothesis in statistical models, such as PGLS models, or independent contrast analyses. In Appendix 1 we provide a suggestion to obtain an optimal consensus tree (the 50% majority-rule consensus tree, MRC tree; Holder et al., 2008; Sukumaran and Holder, 2010) out of a set of trees downloaded from BirdTree. Using single trees and disregarding phylogenetic uncertainty has been the most popular option in the vast majority of comparative studies carried out to date. For example, a scan of the comparative studies on birds published over the past 10 years (2004–2013) in the *Journal of Evolutionary Biology* (search terms in ISI Web of Science™: (*comparat\** AND (*avian* OR *bird\**))) yielded 28 relevant papers, all of which used a single tree as a phylogenetic hypothesis, obtained either by manually coding phylogenies ( $n = 21$ ), by adopting a supertree approach ( $n = 4$ ) or by using a consensus tree out of a set of multiple trees ( $n = 3$ ).

However, the assumption of no phylogenetic uncertainty is unwarranted: for instance, phylogenetic tree reconstruction using MCMC return trees (in proportion to their posterior probability) that vary somewhat, either in topology or branch length, as it is the case for BirdTree phylogenetic trees. Therefore, methods that incorporate phylogenetic uncertainty in regression models (or in any statistical method relying on a phylogenetic tree) are expected to provide more reliable parameter estimates and standard errors compared to methods ignoring phylogenetic uncertainty, and should be preferred whenever possible (e.g. Grafen, 1989; Martins, 1996;

Martins and Hansen, 1997; Huelsenbeck and Rannala, 2003; de Villemereuil et al., 2012, Garamszegi and Mundry, 2014; see also the popular software BayesTraits, <http://www.evolution.rdg.ac.uk/BayesTraitsV2Beta.html>). Particularly, incorporating uncertainties into the phylogenetic analyses puts an emphasis on the confidence interval around parameter estimates, which delineates how much a comparative result can be trusted.

To account for phylogenetic uncertainty and to fully exploit the possibilities offered by downloading tree sets from BirdTree to obtain robust comparative inferences, we envisage several possibilities that rely on an entire set of phylogenetic trees instead of a single consensus tree.

First, phylogenetic analyses of trait (co)evolution can be computed using an entire set of  $n$  trees, by re-running the analysis  $n$  times, each time using a different tree, and obtaining a distribution of  $n$  coefficients and/or associated statistics, allowing the estimation of the effect of phylogenetic uncertainty on the phylogenetic correlation between the traits of interest (e.g. Martins, 1996; Liker et al., 2013). This approach does not consider potential differences between different models relying on different phylogenetic trees, implying that summary statistics are made by assigning equal weights to each observation. Another flexible and easy to implement method to incorporate phylogenetic uncertainty into statistical modelling applies Information Theoretic (IT) approaches to differentially weight particular models built using different trees (Garamszegi and Mundry, 2014). This method assumes that accounting for phylogenetic uncertainty in comparative analyses is a classical model selection problem that can be solved by multi-model inference. Therefore, one can run a statistical model (e.g. a PGLS regression model) testing one (or more) comparative hypotheses using an entire set of  $n$  equally likely trees, and then perform model-averaging of parameter estimates from the set of  $n$  models, weighing each model based on its fit to the data using, for example, its Akaike weight (Burnham et al., 2011). The benefit of the weighted approach might be that, for parameter estimation, it gives less emphasis to particular models that offer a bad fit to the data. Both the unweighted and weighted forms of multimodel inference over an entire set of models conditioned to different trees provide a single average effect (such as is the case when using a consensus tree), but the output also includes a confidence or error range around this mean estimate as obtained from the variance of model parameters across different models. This interval can be

interpreted as a proxy of the uncertainty associated with the phylogenetic hypothesis, as the confidence range will essentially capture variations in parameter estimates that the consideration of alternative phylogenetic hypotheses can cause (Garamszegi and Mundry, 2014). We emphasize that the performance of the above multimodel approaches in comparative analyses would deserve further evaluation by simulation studies as well as challenges by empirical data, which might be helpful to decide whether the weighted or unweighted modes of multimodel inference are more reliable. For illustrative purposes, in the subsequent simulation (see next section), we relied on the approach that uses model averaging.

Second, Bayesian methods may offer additional solutions: for instance, de Villemereuil et al. (2012) devised computationally efficient and rather flexible MCMC methods to incorporate multiple trees in phylogenetic regression models in order to account for phylogenetic uncertainty. Although we will not deal in detail with Bayesian methods here, we advise researchers that are familiar with Bayesian statistics to explore this possibility. From their simulations, de Villemereuil et al. (2012) suggest that Bayesian models incorporating multiple trees may outperform (in terms of lower uncertainty of parameter estimates and anti-conservative error estimates) 'traditional' consensus-tree-based PGLS models. However, they compared the performance of Bayesian models with that of PGLS models using a 'strict-consensus tree', which might not be the optimal way to summarize a set of MCMC trees (see Appendix 1). Moreover, Bayesian methods, such as the one proposed by de Villemereuil et al. (2012), have yet to be tested extensively in a variety of real dataset conditions, including multiple regression settings, and may become computationally problematic when the number of species is very large ( $> 200$ ) (de Villemereuil et al., 2012). Finally, Bayesian approaches may introduce unnecessary variance in posterior estimates whenever no informative priors are provided (i.e. in most cases). Therefore, albeit Bayesian approaches also provide mean parameter estimates and confidence intervals (in this case they are called credible intervals), the latter variance will reflect the effect of both the variance across the considered phylogenetic trees and the use of uninformative priors (Garamszegi and Mundry, 2014).

### 3 How Many Trees?

The reliability of phylogenetically-corrected analyses may vary according to the number and specific sets of

species (taxonomic bias, Arnold and Nunn, 2010) and intra-specific variance in traits (Garamszegi and Møller, 2010). Moreover, it may vary according to the number of phylogenetic trees used to draw inferences, since fitting models using a variable number of trees that differ to some extent between each other in their phylogenetic relationships among species (either in terms of topology or branch lengths) will result in variation of estimated model parameters.

Here we addressed two of these issues by evaluating the effect of variation in the number of species and in the number of phylogenetic trees on the precision of parameter estimates from PGLS models based on a real comparative dataset. An effect of the number of trees on the precision of parameter estimates may emerge both if a consensus or a multiple tree approach is adopted. Although we aimed at devising a 'rule-of-thumb' to assist researchers when deciding the number of trees that should be downloaded from BirdTree and incorporated in subsequent comparative analyses, we emphasize that the issue regarding the effect of the number of trees on the precision of parameter estimates is a general one, and applies to any case when multiple equally likely phylogenetic trees are available for comparative analyses. In the following simulation, to contrast two radically different methods, we focused on the precision of parameter estimates from PGLS models obtained with the MRC tree-based method and with the model averaging approach.

As phylogenetic information, we used random draws of either 250, 500, 1,000 or 2,500 trees (tree-size classes hereafter) [we choose 2,500 as the higher limit because downloading tree sets  $> 2,000$  trees from BirdTree can be problematic, as download often fails (D. Rubolini pers. obs.)]. Our simple simulation was based on the large-scale comparative analysis of the relationship between mass-specific productivity and life-history traits of 980 bird species carried out by Sibly et al. (2012). We assessed variation in parameter estimates of the (negative) relationship between mass-specific productivity and body mass (both  $\log_{10}$ -transformed; see Sibly et al., 2012) in two random samples of 50 and 100 species (a typical sample size in most comparative analyses of bird species) from the original 980 species set (50 and 100 species test datasets hereafter). For each test dataset, we downloaded from BirdTree ('Hackett All Species') 250, 500, 1,000 and 2,500 randomly sampled trees. We repeated downloading 10 times for each tree size class, thus obtaining 10 random replicate tree sets for each tree size class. For each replicate set of trees,

we built a MRC tree. We then ran a PGLS model for each test dataset and each replicate tree-size using as phylogenetic information either the MRC tree (MRC models) or all the trees in each replicate tree set using the model-averaging method (MA models). PGLS models were ran with the caper R package (Orme et al., 2012), with  $\lambda$  estimated from the data by maximum likelihood, using R (2.15.2) (R Core Team, 2013). For each run, we noted the estimated slope and standard error of MRC models, and the model-averaged slope and standard error from the MA models. We obtained a total of 80 slopes and standard errors for each method (10 replicates  $\times$  2 test datasets  $\times$  4 tree-size classes). From each pair of slope and standard error, we calculated a z-value as the ratio between the two (absolute value; this was done also for model-averaged slopes and standard errors).

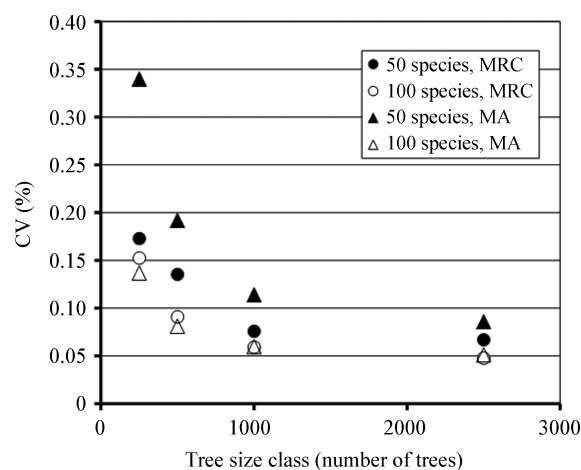
To evaluate precision of model parameters, we calculated the coefficient of variation (CV out of the 10 replicate analyses) of each model parameter (slope, standard error, z) for each of the four tree-size classes and method of analysis. The variability of CVs of z-values according to the test dataset, tree size class and method of analysis is illustrated in Fig. 1. CVs were below 0.40% even for analyses based on 250 trees, and were consistently lower for the 100 than for the 50 species test dataset, especially for estimates obtained with the MA method. CVs of z-values dropped to values varying between 0.05% and 0.12% for models based on at least 1,000 trees (Fig. 1). A pairwise comparison of z-values for each tree set revealed that the MA method returned consistently smaller z-values than the MRC method for the 50 species test dataset ( $4.61 \pm 0.01 SE$  vs.  $4.65 \pm 0.01 SE$ ,  $t_{39} = 23.1$ ,  $P < 0.001$ ), while the opposite was the case for the 100 species test dataset ( $11.56 \pm 0.01 SE$  vs.  $11.51 \pm 0.01 SE$ ,  $t_{39} = 27.2$ ,  $P < 0.001$ ). Trends of CVs of other model parameters according to tree size-class were similar as those shown in Fig. 1 and were of the same magnitude as those of z-values (min-max, slope: 0.01%–0.32%; standard error: 0.04%–0.17%) (details not shown). Although there is no generally recommended threshold for precision estimates, we are confident that CVs of parameter estimates  $< 0.5\%$  will not strongly affect the conclusions of phylogenetically corrected analyses. However, downloading from BirdTree and including in the analyses at least 1,000 trees should result in very precise model parameters, regardless of the method of analysis. Clearly, our simulations on a real dataset suggest that inclusion in PGLS models of  $> 1,000$  trees does not improve the precision

of parameter estimates. Moreover, we suggest to run the analyses either using trees based on the ‘Hackett’ and ‘Ericson’ backbones, to check for consistency of results, although it is likely that for most comparative datasets the ‘Hackett’ and ‘Ericson’ backbones will give consistent results. Indeed, as it is also clear from our simulations, comparative regression models appear relatively robust to some degree of misspecifications in the underlying trees, either in terms of branch lengths or topology (Purvis et al., 2004; Stone, 2011), although the extent to which such misspecifications add bias to the analysis deserves further investigation.

Although simulations with Bayesian methods indicate that even a small sample of trees (100 trees) can be sufficient to obtain reliable parameter and error estimates, in real studies de Villemereuil et al. (2012) advise that a large number of trees is used because using more trees is expected to better represent their true probability distribution, also considering that computation time and memory usage in MCMC methods seem little affected by the number of trees (see de Villemereuil et al., 2012).

## 4 Conclusions

To date, most comparative analyses of birds have been carried out by combining available trees using a



**Fig. 1** Coefficient of variation (CV, %) of z-values (ratio of slope to standard error) of the bivariate relationship between productivity and body mass (both  $\log_{10}$ -transformed, see Sibly et al. 2012) in the two test datasets (50 and 100 species; see main text) in relation to the number of random trees (downloaded from BirdTree)

CVs were calculated based on 10 replicate z-values from 10 PGLS models. MRC = z-values obtained from PGLS models ran using the 50% majority-rule consensus tree. MA = z-values calculated as the ratio of the model-averaged slope and standard errors from PGLS models averaged across an entire tree set.

variety of approaches, often obtained using incomparable genetic methods (e.g. DNA-DNA hybridization and gene-based phylogenies). To our knowledge, the BirdTree project constitutes the first attempt to produce a complete phylogenetic tree of avian taxa based on a robust Bayesian phylogenetic framework. We believe that, if *ad hoc* phylogenetic analyses based on novel, original molecular data are not available (as is the case with the vast majority of evolutionary comparative studies of birds carried out to date), phylogenetic trees obtained from BirdTree will provide evolutionary ecologists with a unique opportunity to standardize the presentation of their phylogenetically-corrected statistical analyses using a robust and validated phylogenetic background. Since it is unlikely that more comprehensive (and user-friendly) data will become available in the coming years, incorporation of BirdTree phylogenies into comparative studies, either using a multiple tree modelling approach (either using an IT or Bayesian framework) or a more traditional phylogenetic regression modelling approach using consensus trees, could become a standard in comparative studies of avian species for many years to come.

**Acknowledgements** We are grateful to D. Horner and C. Gissi for assistance with Python language, and to G. Thomas and two anonymous referees for useful comments on previous drafts of the paper. LZG was supported by funds from the Spanish government within the frame of the 'Plan Nacional' program (ref. no. CGL2012-38262, CGL2012-40026-C02-01).

## References

- Arnold C, Nunn CL, 2010. Phylogenetic targeting of research effort in evolutionary biology. *Am. Nat.* 176: 601–612
- Arnold C, Matthews LJ, Nunn CL, 2010. The 10kTrees website: A new online resource for primate phylogeny. *Evol. Anthropol.* 19: 114–118.
- Barker FK, Cibois A, Schilker P, Feinstein J, Cracraft J, 2004. Phylogeny and diversification of the largest avian radiation. *Proc. Natl. Acad. Sci. USA* 101: 11040–11045.
- Bennett PM, Owens IPF, 2002. *Evolutionary Ecology of Birds: Life Histories, Mating Systems, and Extinction*. Oxford: Oxford University Press.
- Bininda-Emonds ORP, 2014. An introduction to supertree construction (and partitioned phylogenetic analyses) with a view toward the distinction between gene trees and species trees. In: Garamszegi LZ ed. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Theory and Practice*. Berlin: Springer-Verlag, 49–76.
- Burnham KP, Anderson DR, Huyvaert KP, 2011. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* 65: 23–35.
- de Villemeireuil P, Wells JA, Edwards RD, Blomberg SP, 2012. Bayesian models for comparative analysis integrating phylogenetic uncertainty. *BMC Evol. Biol.* 12: 102.
- Drummond A, Rambaut A, 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.
- Ericson PG, Anderson CL, Britton T, Elzanowski A, Johansson US et al., 2006. Diversification of Neoaves: Integration of molecular sequence data and fossils. *Biol. Lett.* 2: 543–547.
- Felsenstein J, 1985. Phylogenies and the comparative method. *Am. Nat.* 125: 1–15.
- Freckleton RP, Harvey PH, Pagel M, 2002. Phylogenetic analysis and comparative data: A test and review of evidence. *Am. Nat.* 160: 712–726.
- Garamszegi LZ, 2014. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Theory and Practice*. Berlin: Springer-Verlag.
- Garamszegi LZ, Møller AP, 2010. Effects of sample size and intraspecific variation in phylogenetic comparative studies: A meta-analytic review. *Biol. Rev.* 85: 797–805.
- Garamszegi LZ, Mundry R, 2014. Multimodel-inference in comparative analyses. In: Garamszegi LZ ed. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Theory and Practice*. Berlin: Springer-Verlag, 305–331.
- Garamszegi LZ, Gonzalez-Voyer A, 2014. Working with the tree of life in comparative studies: How to build and tailor phylogenies to interspecific datasets. In: Garamszegi LZ ed. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Theory and Practice*. Berlin: Springer-Verlag, 19–48.
- Grafen, A, 1989. The phylogenetic regression. *Phil. Trans. R. Soc. Lond. B* 326: 119–157.
- Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL et al., 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320: 1763–1768.
- Harvey PH, Pagel MD, 1991. *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Holder MT, Sukumaran J, Lewis PO, 2008. A justification for reporting the majority-rule consensus tree in Bayesian phylogenetics. *Syst. Biol.* 57: 814–821.
- Huelsenbeck JP, Rannala B, 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* 57: 1237–1247.
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO, 2012. The global diversity of birds in space and time. *Nature* 491: 444–448.
- Jönsson KA, Fjeldså J, 2006. A phylogenetic supertree of oscine passerine birds (Aves: Passeri). *Zool. Scripta* 35: 149–186.
- Liker A, Székely T, 2005. Mortality costs of sexual selection and parental care in natural populations of birds. *Evolution* 59: 890–897.
- Liker A, Freckleton RP, Székely T, 2013. The evolution of sex roles in birds is related to adult sex ratio. *Nature Comm.* 4: 1587.
- Maddison WP, 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution* 44: 539–557.
- Maddison WP, 2000. Testing character correlation using pairwise comparisons on a phylogeny. *J. Theor. Biol.* 202: 195–204.
- Martins EP, 1996. Conducting phylogenetic comparative analyses

- when the phylogeny is not known. *Evolution* 50: 12–22.
- Martins EP, Hansen TF, 1997 Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149: 646–667
- Møller AP, Birkhead TR, 1992. A pairwise comparative method as illustrated by copulation frequency in birds. *Am. Nat.* 139: 644–656.
- Morgan-Richards M, Trewick SA, Bartosch-Harlid A, Kardailsky O, Phillips MJ et al., 2008. Bird evolution: Testing the metaves clade with six new mitochondrial genomes. *BMC Evol. Biol.* 8: 20.
- Nunn CL, 2011. *The Comparative Approach in Evolutionary Anthropology and Biology*. Chicago: The University of Chicago Press.
- Orme D, Freckleton RP, Thomas G, Petzoldt T, Fritz S et al., 2012. Caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5. <https://cran.r-project.org/web/packages/caper/>
- Pagel MD, 1997 Inferring evolutionary processes from phylogenies. *Zool. Scripta* 26: 331–348.
- Pagel MD, 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.
- Paradis E, 2012. *Analysis of Phylogenetics and Evolution with R*. 2<sup>nd</sup> edn. Berlin: Springer-Verlag.
- Purvis A, Gittleman JL, Luh H-K, 1994. Truth or consequences: Effects of phylogenetic accuracy on two comparative methods. *J. Theor. Biol.* 167: 293–300.
- R Core Team, 2013. *R: A Language and Environment For Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Saino N, Rubolini D, Serra L, Caprioli M, Morganti M et al., 2010. Sex-related variation in migration phenology in relation to sexual dimorphism: A test of competing hypotheses for the evolution of protandry. *J. Evol. Biol.* 23: 2054–2065.
- Sibley CG, Ahlquist JE, 1990. *Phylogeny and Classification of Birds: A study in molecular evolution*. New Haven: Yale University Press.
- Sibly RM, Witt CC, Wright NA, Venditti C, Jetz W et al., 2012. Energetics, lifestyle, and reproduction in birds. *Proc. Natl. Acad. Sci. USA* 109: 10937–10941.
- Stone EA, 2011. Why the phylogenetic regression appears robust to tree misspecification. *Syst. Biol.* 60: 245–260.
- Sukumaran J, Holder MT, 2010. DendroPy: A Python library for phylogenetic computing. *Bioinformatics* 26: 1569–1571.
- Thomas GH, Wills MA, Székely T, 2004. A supertree approach to shorebird phylogeny. *BMC Evol. Biol.* 4: 28.

## Appendix 1 Suggested procedure for summarizing BirdTree tree sets into an optimal consensus tree

We suggest that researchers aiming at summarizing BirdTree tree sets into a single consensus tree adopt a 50% majority-rule consensus tree (MRC tree hereafter). According to Holder et al. (2008), the MRC trees can be viewed as an optimal summary of the posterior distribution of MCMC trees. However, Holder et al. (2008) clearly point out that when phylogenetic trees are to be used as ‘nuisance parameters’ (as it is the case in comparative analyses), the entire sample of trees from the posterior distribution can be used to characterize phylogenetic uncertainty. In a MRC tree, a branching event is considered as supported if it occurs in >50% of the MCMC trees (Holder et al. 2008). In cases when support for a given branch is below 50%, a polytomy is formed. From a set of ultrametric MCMC trees, such as those downloaded from BirdTree, an MRC tree can be obtained by using the SumTrees program (<http://pythonhosted.org/DendroPy/scripts/sumtrees.html>). SumTrees is part of DendroPy, a Python library for phylogenetic computing (Sukumaran and Holder 2010). In SumTrees, the mean branch lengths of the MRC tree derived from a set of ultrametric trees are adjusted such that the ages of the subtended nodes correspond to the median age of the corresponding nodes of the input trees (see details on the SumTrees website).

We have tested the current SumTrees official release version (in DendroPy 3.12.0) with several BirdTree tree sets, and noted that the downloaded trees, although they are ultrametric by definition, were not recognized as such by SumTrees, probably because of very minor inconsistencies in edge lengths of the tree subsets. The authors of SumTrees have already added a useful option to cope with this, that allows skipping ultrametricity checks, which unfortunately is not yet included in the official release, but can be found in the active development source code repository at <https://github.com/jetsukumaran/DendroPy>. This minor adjustment allows SumTrees to readily accept and correctly process BirdTree tree sets as ultrametric trees.

A basic SumTree script to process a BirdTree NEXUS tree set to obtain an MRC tree, run on the development source code with the ‘ultrametricity-precision’ option set to an arbitrarily very high value (i.e. skipping ultrametricity checks), is thus as follows:

```
sumtrees.py --rooted --ultrametric -- ultrametricity-precision 10000000 YOUR_
INPUT_TREES.tre > YOUR_OUTPUT_TREE.tre
```

The MRC tree is a standard NEXUS tree that can be included in any statistical analysis, including PGLS models or independent-contrast analyses.