

APPLICATION

PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences

Gavin H. Thomas^{1*†}, Klaas Hartmann^{2†}, Walter Jetz³, Jeffrey B. Joy⁴, Aki Mimoto⁴ and Arne O. Mooers⁴

¹Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK; ²Institute for Marine and Arctic Studies, University of Tasmania, Private Bag 49, Hobart TAS 7001, Australia; ³Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 06520-8106, USA; and ⁴Department of Biosciences, Simon Fraser University, 8888 University Dr., Burnaby, BC V5A 1S6, Canada

Summary

1. Phylogenetic trees that include all member lineages are necessary for many questions in macroevolution, biogeography and conservation. Currently, producing such trees when genetic data or phenotypic characters for some tips are missing generally involves assigning missing species to the root of their most exclusive clade, essentially grafting them onto existing and static topologies as polytomies.

2. We describe an R package, 'PASTIS', that enables a two-stage Bayesian method using MrBayes version 3.2 (or higher) to incorporate lineages lacking genetic data at the tree inference stage. The inputs include a consensus topology, a set of taxonomic statements (e.g. placing species in genera and aligning some genera with each other or placing subspecies within species) and user-defined priors on edge lengths and topologies. PASTIS produces input files for execution in MrBayes that will produce a posterior distribution of complete ultrametric trees that captures uncertainty under a homogeneous birth-death prior model of diversification and placement constraints. If the age distribution of a focal node is known (e.g. from fossils), the ultrametric tree distribution can be converted to a set of dated trees. We also provide functions to visualize the placement of missing taxa in the posterior distribution.

3. The PASTIS approach is not limited to the level of species and could equally be applied to higher or lower levels of organization (e.g. accounting for all recognized subspecies or populations within a species) given an appropriate choice of priors on branching times.

Key-words: bioinformatics, phylogenetics, reconstruction, software

Introduction

Analyses of diversification through time (e.g. Morlon, Parsons & Plotkin 2011; Stadler 2011; Jetz *et al.* 2012), evolutionary isolation (e.g. Isaac *et al.* 2007) and community phylogenetics (e.g. Cooper, Rodriguez & Purvis 2008) perform best if all lineages in a focal clade or region are accounted for. Existing methods that account for incomplete species sampling are restricted to quite specific methods – either to investigations of correlates of diversification (e.g. FitzJohn, Maddison & Otto 2009) or to investigations of the temporal dynamics of diversification (e.g. Cusimano, Stadler & Renner 2012). Given known sampling biases towards particular clades or geographic regions, more general solutions are required that (i) account for non-random distributions of missing species or nodes among lineages and (ii) yield distributions of fully resolved trees that reflect uncertainty in tree topology and branch lengths.

Phylogenetic trees inferred from genetic data on all, or nearly all, extant species are still rare and are generally restricted to smaller clades (Steeman *et al.* 2009; Aze *et al.* 2011; Near *et al.* 2011). At larger taxonomic scales, supertree (e.g. Bininda-Emonds *et al.* 2007) or agglomerative methods (Kraft & Ackerly 2010), which incorporate taxonomic information, are used, producing trees with many polytomies. Trees that are taxonomically complete but unresolved (as is common with supertree methods) can be resolved by inferring the timing of missing splits under a constant-rate birth-death model (Kuhn, Mooers & Thomas 2011). If the trees are incomplete, several approaches allow missing species to be accounted for but do not generate complete, resolved trees directly or are limited to specific types of analysis. First, if the missing splits are temporally non-random (e.g. if splits are expected to be biased towards the root of the tree), their placement can be simulated under a constant-rate or temporally varying rate birth-death model (Cusimano, Stadler & Renner 2012). Secondly, if missing species are taxonomically random (which is rarely the case), they can be treated as pseudo-extinct at the analysis stage

*Correspondence author. E-mail: gavin.thomas@sheffield.ac.uk

Tweetable abstract: PASTIS for R integrates taxa lacking genetic data into MrBayes version 3.2 input using topology constraints.

†These authors contributed equally.

(Pybus & Harvey 2000; FitzJohn, Maddison & Otto 2009). Finally, if the missing species can be assigned to subclades, these subclades can be augmented numerically (Alfaro *et al.* 2009; FitzJohn, Maddison & Otto 2009). While reasonable, none of these approaches make full use of available information.

We describe and implement a general approach that incorporates prior information on the placement of missing taxa and that is suited to any taxonomic group with missing genetic or phenotypic information. We provide new software (PASTIS; available from <http://cran.r-project.org/web/packages/pastis/>) written in the R environment to (i) generate MrBayes input files for tree inference and (ii) conduct post-MrBayes assessment of the placement of missing taxa. Our approach has recently been used to infer complete species phylogenies for birds (Jetz *et al.* 2012). We use a birth-death prior on the edge lengths, which means that (i) edge lengths for all species (and clades, including those with no genetic data) are sampled under a common framework and (ii) the inclusion of missing species is unbiased with respect to fitting rate-heterogeneous diversification models. We impose no model prior on topologies, which means that all labelled trees are equally likely before the data and topological constraints are imposed. Such a prior is consistent with our current understanding of phylogenetic tree shape (see, e.g. Blum & Francois 2006). The resulting tree distribution is expected to be broad, and the consensus tree created from it should reconstruct the constraint tree plus all the added species attached as polytomies to their most exclusive clade. However, the full distribution also represents all available taxonomic and genetic information and meaningful bifurcating topologies and edge lengths.

The PASTIS method

RUNNING PASTIS

Our objective is to integrate missing taxa into a posterior distribution of trees that includes all taxa in the clade and is consistent with raw sequence data and taxonomic information. To achieve this, the PASTIS method takes advantage of flexible topology constraints recently implemented in MrBayes version 3.2 (Ronqvist *et al.* 2012) and uses functions in the R packages *ape* (Paradis, Claude & Strimmer 2004) and *caper* (Orme *et al.* 2012). Topology constraints are statements dictating where species can and cannot be placed while inferring a phylogenetic tree (e.g. Swofford & Beagle 1993; Day, Cotton & Barraclough 2008; Thomas 2008; Lanfear & Bromham 2011). MrBayes allows constraints to be *hard* (enforcing monophyly on a taxon set), *negative* (preventing monophyly on a taxon set) or *partial*. Partial constraints define taxon sets that must be monophyletic with respect to a second taxon set while allowing taxa not defined in either set to move freely about the tree. Use of multiple complementary and hierarchical constraints provides a powerful approach with which to incorporate taxa with no genetic data into posterior distributions of phylogenetic trees by the application of a series of rules and assumptions on the placement of missing taxa. Our general approach is outlined in Fig. 1. The key components are (i) a constraint tree that defines relationships among taxa with sequence data and (ii) a taxonomy or other external data providing prior information on the affinities of taxa with no sequence data. Generating sets of tens or hundreds of constraints manually is non-trivial, tedious and prone to error. PASTIS automates the process by combining the two primary sources of information into a

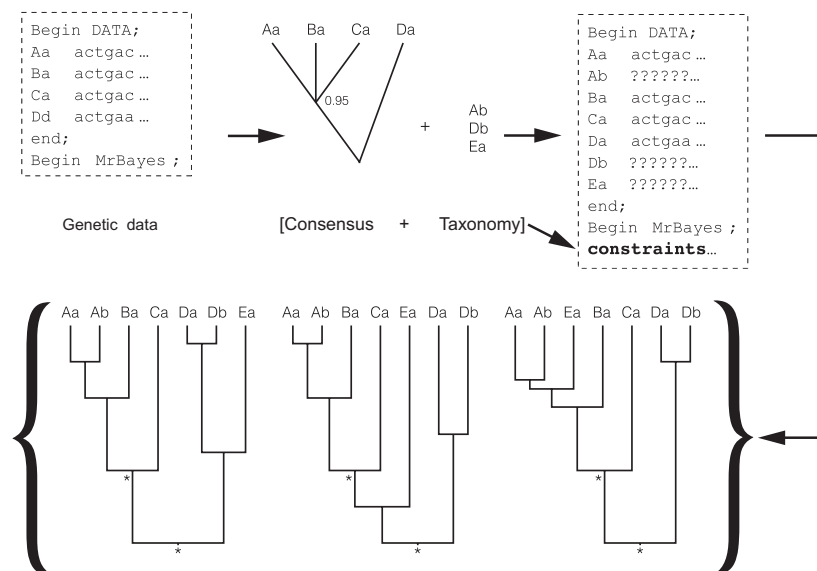


Fig. 1. A schematic of the tree inference process. A constraint tree (typically a consensus topology) is generated from the genetic data set or previously published phylogenetic hypotheses. This constraint tree and the taxonomic data associated with missing species dictate placement constraints generated by PASTIS. A posterior distribution of ultrametric trees containing all taxa is then generated in MrBayes. Asterisks denote the constraints from the consensus topology. In the example below, genus E has no sequence data and is free to move about the tree, but not to enter any other genus. For simplicity, these are depicted with no variation in depth; in a full posterior, depths will also vary.

Table 1. Accipitridae input used in PASTIS

PASTIS argument =suffix	Required	Description
constraint_tree=Accipitridae.tree	Yes	A constraint tree containing 177 species (175 Accipitridae and two outgroups). The tree is a 95% majority-rule consensus tree derived from an unconstrained MrBayes analysis of the alignment in Accipitridae.sequences. This structure will be included in all output trees.
taxa_list=Accipitridae.taxa	Yes	This is a list (.csv format) of all 245 taxa (including outgroups) to be included in the complete Accipitridae tree. Each species is assigned membership to its genus.
missing_clades=Accipitridae.missingclades	Optional	This is a list of six genera that are not represented in the constraint tree and where they may be placed in the tree.
sequences=Accipitridae.sequences	Optional	This is the aligned sequence data for the 177 species in the constraint tree. PASTIS expects the alignment in FASTA format. This is optional, but will be present in most typical analyses.
output_template=Accipitridae.template	Optional	This is a template file for the MrBayes output file. It outlines options such as the data partitions, number of iterations, burn in period, etc.

MrBayes input file with a potentially extensive set of topology constraints.

We define three categories of species: type 1 species have genetic information; type 2 species have no genetic information but are congeners of a species with genetic information; and type 3 species have no genetic data and are members of a genus that does not have genetic data. To integrate the three types of missing species, we make two important assumptions: (i) taxonomic groups (e.g. genera, subfamilies) are monophyletic unless there is positive evidence (i.e. genetic data) that suggests otherwise and (ii) reasonable edge-length and topology priors (i.e. birth-death models) exist.

We use the avian family Accipitridae to demonstrate how we integrate missing species into the phylogenetic pipeline with PASTIS. We provide all data, details of data sources and R scripts on figshare (<http://dx.doi.org/10.6084/m9.figshare.692180>). The Accipitridae consists of 243 species for which we obtained sequence data on 175 species from GenBank (Benson *et al.* 2009). We refer to these as type 1 species. The remaining 68 species have no sequence data, of which 60 have congeners with sequence data (type 2 species) and eight species from six genera do not have congeners with sequence data (type 3 species). We constructed a constraint tree (Accipitridae.tree, Table 1) based on an alignment including mitochondrial, nuclear coding and nuclear non-coding genes for the 175 type 1 species, plus two outgroups using MrBayes (Ronqvist *et al.* 2012). We collapsed nodes with <95% posterior probabilities to polytomies. Model parameters for the constraint tree are as defined in the supplementary file Accipitridae.template on figshare, and further details of construction can be obtained from Jetz *et al.* (2012). The constraint tree could alternatively have been derived from published sources or inferred from different data or methods. However, in practice, we have found that, in order to avoid conflict between data and imposed constraints, the constraint tree should preferentially be generated with the same underlying genetic data (e.g. new data or previously published sequences or alignments) that will be used in the generation of the full tree.

To integrate the 68 type 2 and type 3 species, we construct a simple taxon definition file, Accipitridae.taxa (Table 1), that lists all species (types 1, 2 and 3) along with a clade name.

The clade name in this example is simply the genus name but could in principle be any higher taxon or more inclusive clade. The file is in csv format, and the first few lines are as follows:

```
taxon,clade
Tyto_alba,Outgroup
Cathartes_aura,Outgroup
Accipiter_albogularis,Accipiter
Accipiter_badius,Accipiter
Accipiter_bicolor,Accipiter
Accipiter_brachyurus,Accipiter
```

PASTIS integrates the information in the taxa file with the constraint tree to formulate the simplest possible topology constraints that combine taxonomic data (type 2, 3) and genetic data (type 1). To generate a MrBayes input file using PASTIS, we simply run:

```
library(pastis)
pastis_main(constraint_tree='Accipitridae.tree',
taxa_list='Accipitridae.taxa',
sequences='Accipitridae.sequences',
output_template='Accipitridae.template',
output_file='Accipitridae.nexus')
```

The file Accipitridae.template provides an editable template for the MrBayes input file and is optional: PASTIS uses a simple default template if this file is omitted. In the above call to `pastis_main`, PASTIS will assume that all clades defined in the taxa file are monophyletic unless there is conflicting evidence to the contrary. Figure 2 describes the placements of exemplar type 2 species from the Accipitridae tree. The type 2 species *Butastur liventer* belongs to a clade (*Butastur*) that is monophyletic on the constraint tree. With no evidence to the contrary, PASTIS defines constraints that restricts *Butastur liventer* to the *Butastur* clade but allow it to move freely among branches within that clade. In contrast, the type 2 species *Buteo archeri* is member of a genus for which the constraint tree provides positive evidence of non-monophyly and implies that *Buteo* is part of a more inclusive clade including the genera *Leucopternis*, *Parabuteo* and *Geranoaetus*. Here, *Buteo archeri* is constrained to the clade that includes the most recent common ancestor of all type 1 species belonging to *Buteo* and the additional three genera listed above. We refer to this as a

supragenus. *Buteo archeri* can move freely within this broad clade but cannot break the monophyly of *Harpypaliaetus*, *Butastur* or *Ictinia* all of which are monophyletic genera nested within the supragenus (Fig. 2). Without further information, type 3 species are constrained to be monophyletic (e.g. the two members of *Harpagus* are forced to be sister taxa), they cannot break the monophyly of any genus or supragenus but can otherwise move freely throughout the tree. The output file (Accipitridae.nexus) contains full sets of constraints that meet these criteria and can be executed in MrBayes.

In the above example, we allowed type 3 taxa to move throughout the whole tree. However, we may have prior information that allows type 3 taxa to be constrained further. For the eight type 3 taxa, we define additional constraints in a second constraint file, Accipitridae.missingclades (Table 1). For brevity, the example below provides a hypothetical set of constraints for a genus, A, with no sequence data for any member species (but see the file Accipitridae.missingclades for a more complete example).

A,include,B,C,D,E

A,exclude,B,C

Here, the genus A is constrained to be included in a clade containing the genera B, C, D and E. However, taxonomy or other sources suggest that A is more closely affiliated to genera D and E than to B and C, but the constraint phylogeny

suggests that D and E are not monophyletic. In parenthetical format, the relationship between B, C, D and E is defined as: (D, (E, (C, B))). The exclude constraint prevents genus A from entering the clade (C, B). In practice, type 3 species can have multiple exclude constraints but require only a single include constraint (see examples in Accipitridae.missingclades). Figure 3a highlights one example, *Megatriorchis doriae*, showing the branches to which the genus can attach. To obtain MrBayes input run:

```
pastis_main(constraint_tree='Accipitridae.tree',
taxa_list='Accipitridae.taxa',missing_clades='
Accipitridae.missingclades',
sequences='Accipitridae.sequences',
output_template='Accipitridae.template',
output_file='Accipitridae.nexus')
```

Alternatively, we provide a wrapper function to `pastis_main` called `pastis_simple` which searches for and automatically loads relevant files. If all files are in the same location (in the example below PASTIS will search the current working directory but a full file path can be specified) with the appropriate suffix (tree, taxa, missingclades, sequences, template), the function `pastis_simple` is run as follows:

```
pastis_simple('Accipitridae')
```

In either example, PASTIS will generate an input file with the nexus suffix ready for MrBayes execution.

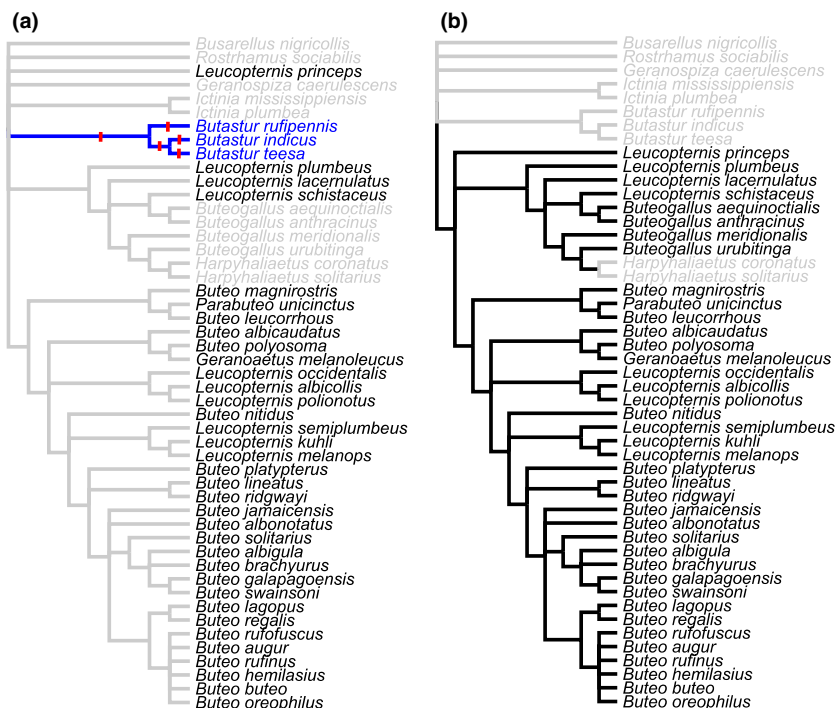


Fig. 2. Placement of missing (type 2 and type 3) species on a subclade of the Accipitridae phylogeny. (a) The type 2 species *Butastur liventer* is confined to the area of the tree containing the genus *Butastur* (highlighted in blue) and is allowed to attached to any branch marked with a red line. The construction of other constraints is more complex. The genus *Buteo* has several missing (type 2) species. *Buteo* is not a monophyletic genus and is inferred to form a clade with members of *Parabuteo*, *Leucopternis* and *Geranoaetus* (taxon names in black in panel a). We refer to this as a supragenus. The *Buteo* supragenus is further complicated by the broad spread of the included genus *Leucopternis* which forms part of a polytomy at the root of the subclade. To resolve the polytomy, we include the minimum possible set of genera defined by the most recent common ancestor of type 1 species belonging to *Buteo*, *Parabuteo*, *Leucopternis* or *Geranoaetus*. This additionally includes the genera *Buteogallus* and *Harpypaliaetus*. Type 2 *Buteo* species can attach to branches within this supragenus shown as black branches panel (b). However, within the resolved supragenus, type 2 species cannot break the monophyly of any nested genus or supragenus and so cannot attach to the grey branches belonging to the genus *Harpypaliaetus*.

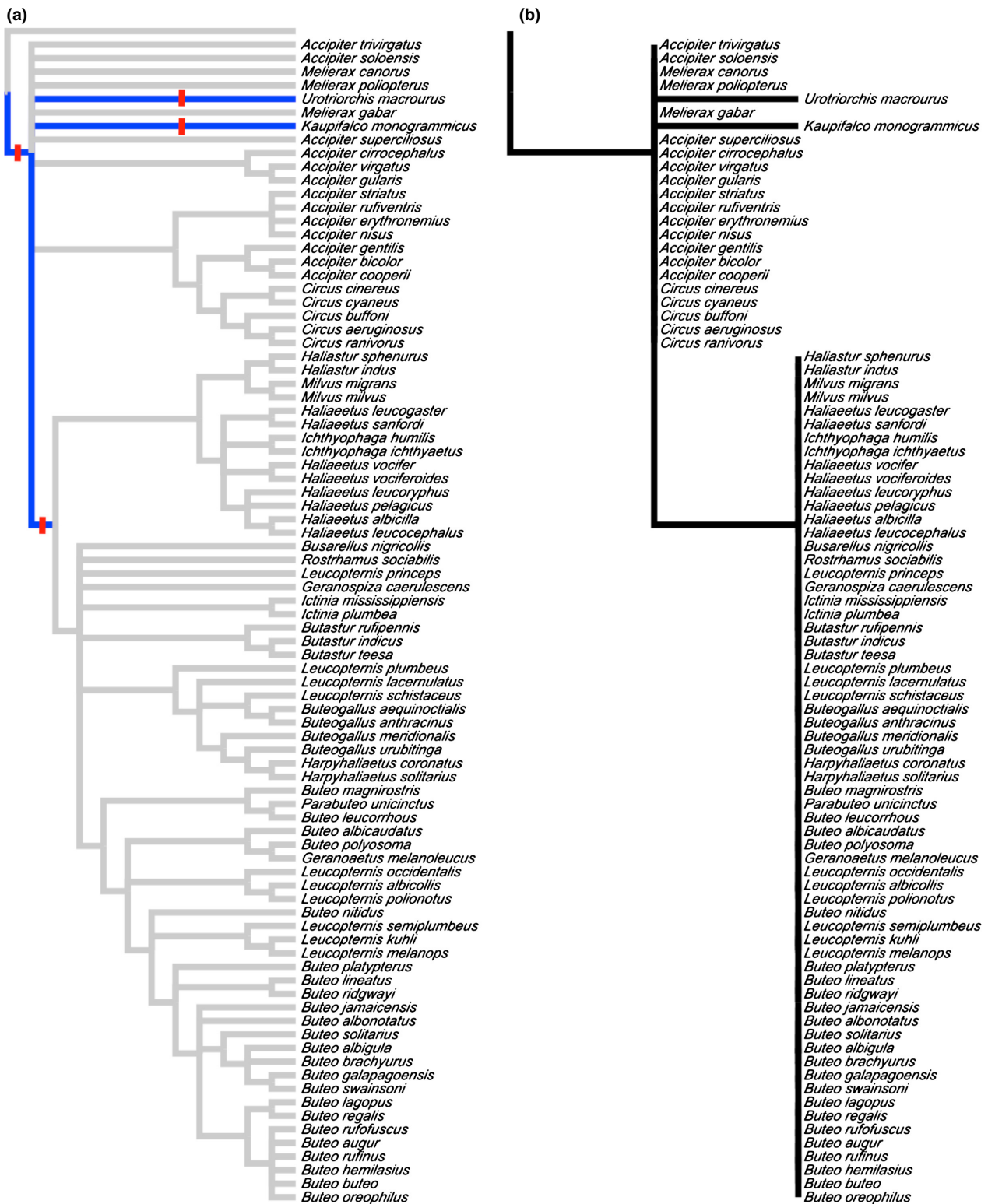


Fig. 3. Reviewing the placement of missing taxa using conch. The topology of the constraint tree showing branches (coloured blue with red line) to which the type 3 species *Megatriorchis doriae* is constrained to attach (a). Output from the function conch showing branches to which *Megatriorchis doriae* attached after a prior-only analysis in MrBayes 3.2 (b). Conch adjusts branch lengths of the input constraint tree to highlight branches where the focal missing species attaches in the posterior distribution of trees. If the missing species does not attach to a branch in any tree in the posterior distribution, then that branch is assigned a length of zero. All other branches are assigned a branch length >0 . In (b), branch lengths are set to unity for a given branch if the missing species attaches at least once to that branch in the posterior distribution of trees. Alternatively, conch allows branch lengths to be set proportional to the number of times the missing species attaches to the focal branch in the posterior distribution of trees.

DEALING WITH POLYTOMIES

In the above descriptions, we implicitly assumed that the constraint tree is fully bifurcating. This may not always be the case. Where possible, PASTIS resolves polytomies in the constraint topologies by ensuring that (i) genera are monophyletic if possible, (ii) supragenera are as small as possible and (iii) type 3 species are placed among the smallest group of genera consistent with the taxonomic information. This step uses taxonomic information and monophyly of named genera to further resolve the consensus topology (Fig. 2).

DIAGNOSING PLACEMENT OF TYPE 2 AND 3 TAXA

PASTIS provides functionality to visualize the placement of missing (type 2 and type 3) species in the posterior distribution of trees. This is primarily intended as a means to check that constraints have been correctly implemented. This can be applied to the posteriors generated by running the examples above. However, inclusion of sequence data means that some allowed placements may rarely occur, and we recommend that constraints be checked based on prior-only analyses. This can be done by omitting sequence data using either `pastis_main` or `pastis_simple`, for example:

```
pastis_simple('Accipitridae', omit_sequences=TRUE)
```

The corresponding MrBayes analysis will run substantially faster than a full analysis with sequence data. Topology constraints can be checked using the function `conch` (*constraint checker*):

```
conch(constraint_tree='Accipitridae.tree',
mrbayes_output='Accipitridae.nexus.t',
simple_edge_scaling=TRUE)
```

(where `Accipitridae.nexus.t` was created by a MrBayes execution of the `pastis_simple` output). For each type 2 or 3 taxon, *i*, not in the constraint tree, this will create a file called 'taxonposition_*i*.tree'. The above call to `conch` assesses all type 2 and type 3 taxa. Alternatively, subsets of species can be checked using the `species_set` argument, for example:

```
conch(constraint_tree='Accipitridae.tree',
mrbayes_output='Accipitridae.nexus.t',
simple_edge_scaling=TRUE,
species_set=c('Megatriorchis_doriae'))
```

The output `taxonposition` tree contains the original constraint tree with edge lengths that are either proportional to the number of sampled trees in which *i* was descendant from that edge (`simple_edge_scaling=FALSE`) or else branch lengths are set to one for all edges in which *i* was descendant from that edge at least once and zero for all other edges (`simple_edge_scaling=TRUE`). Figure 3b provides an example of the output, highlighting edges to which the type 3 species *Megatriorchis doriae* attaches based on a posterior distribution generated over 50 million generations of a prior-only analysis.

Conclusions

Complete trees can help to address questions in macroevolution, biogeography and conservation by removing or reducing

the impacts of non-random species sampling. The effects of including missing species in phylogenies remain to be fully tested (but see Kuhn, Mooers & Thomas 2011 for detailed simulations of one approach). For the analyses of diversification, inclusion of missing taxa under a homogeneous birth-death prior is expected to bias towards detection of a constant-rate model. Since a constant-rate model is also generally treated as the null model in analyses of diversification, trees generated with PASTIS and related approaches (Kuhn, Mooers & Thomas 2011; Cusimano, Stadler & Renner 2012) should retain acceptable type I error rates. We note that even with >30% missing species, trees inferred using PASTIS for the class Aves retained comparable temporal patterns in diversification with trees that omitted missing species (Supplementary Discussion Fig. 2 in Jetz *et al.* 2012). However, the effects of missing species placement or polytomy resolution are less clear for other phylogeny-based analyses (e.g. correlates of diversification, modelling trait evolution, community phylogenetics), and future work should test how the treatment of missing species influences both parameter estimation and type I and II errors. At the very least, we strongly suggest that analyses using trees generated with the PASTIS framework use a large sample of trees from the posterior distribution in order to capture the uncertainty in placement of missing taxa.

Our aim with PASTIS is to provide a straightforward tool to integrate missing taxa into a posterior distribution of trees that includes all taxa in the clade and is consistent with raw sequence data and taxonomic information. The PASTIS approach is not limited to the level of species and could equally apply to higher or lower levels of organization (e.g. accounting for all recognized subspecies or populations within a species) given an appropriate choice of priors on branching times. We envisage that PASTIS may be incorporated as an additional step in existing tree-building pipelines (e.g. Pearse & Purvis 2013; Roquet, Thuiller & Lavergne 2013).

Acknowledgements

We are indebted to Frederic Ronqvist and Maxim Teslenko for modifying MrBayes to allow flexible topology constraints. We thank Beth Forrestel for testing and providing feedback on PASTIS, an anonymous referee for referring to a tree built using this approach as a 'pastiche' and to two anonymous referees who provided helpful suggestions that improved both the manuscript and the code. This work was partly supported by NSF Grants DBI 0960550 and DEB 1026764 (W.J.); the Natural Environment Research Council (Post-doctoral Fellowship Grant number NE/G012938/1 and the NERC Centre for Population Biology) (G.H.T.); and NSERC Canada and the Yale Institute for Biospheric Sciences (A.O.M.).

Data accessibility

All example data sets referred to in the text are available from Figshare at <http://dx.doi.org/10.6084/m9.figshare.692180>. The package PASTIS is available from <http://cran.r-project.org/web/packages/pastis/>.

Author contributions

K.H. and A.O.M. conceived of the study; K.H., W.J., J.B.J., A.O.M., G.H.T. developed the methods; K.H., A.M. and G.H.T. developed software; A.O.M. and G.H.T. wrote the manuscript; K.H., W.J., G.H.T., A.O.M. contributed to the final version of the manuscript.

References

- Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G. & Harmon, L.J. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 13410–13414.
- Aze, T., Ezard, T.H.G., Purvis, A., Coxall, H., Stewart, D., Wade, B. & Pearson, P.N. (2011) A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews*, **86**, 900–927.
- Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J. & Sayers, E. (2009) GenBank. *Nucleic Acids Research*, **37**, D26–D31.
- Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R. *et al.* (2007) The delayed rise of present-day mammals. *Nature*, **446**, 507–512.
- Blum, M.G.B. & Francois, O. (2006) Which random processes describe the Tree of Life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology*, **55**, 685–691.
- Cooper, N., Rodriguez, J. & Purvis, A. (2008) A common tendency for phylogenetic overdispersion in mammalian assemblages. *Proceedings of the Royal Society of London Series B*, **275**, 2031–2037.
- Cusimano, N., Stadler, T. & Renner, S. (2012) A new method for handling missing species in diversification analysis applicable to randomly or nonrandomly sampled phylogenies. *Systematic Biology*, **61**, 785–792.
- Day, J.J., Cotton, J.A. & Barraclough, T.G. (2008) Tempo and mode of diversification of Lake Tanganyika cichlid fishes. *PLoS ONE*, **3**, e1730.
- FitzJohn, R.G., Maddison, W.P. & Otto, S.P. (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, **58**, 595–611.
- Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C. & Baillie, J.E.M. (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE*, **2**, e296.
- Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.
- Kraft, N. & Ackerly, D.D. (2010) Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. *Ecological Monographs*, **80**, 401–422.
- Kuhn, T.S., Mooers, A.O. & Thomas, G.H. (2011) A Simple Polytope Resolver For Dated Phylogenies. *Methods in Ecology and Evolution*, **2**, 427–436.
- Lanfear, R. & Bromham, L. (2011) Estimating phylogenies for species assemblages: a complete phylogeny for the past and present native birds of New Zealand. *Molecular Phylogenetics and Evolution*, **61**, 958–963.
- Morlon, H., Parsons, T.L. & Plotkin, J.B. (2011) Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences of USA*, **108**, 16327–16332.
- Near, T.J., Bossu, C.M., Bradburd, G.S., Carlson, R.L., Harrington, R.C., Hollingsworth, P.R., Keck, B.P. & Etnier, D.A. (2011) Phylogeny and temporal diversification of darters (Percidae: Etheostominae). *Systematic Biology*, **60**, 565–595.
- Orme, C.D.L., Freckleton, R.P., Thomas, G.H., Petzoldt, T., Fritz, S.A., Isaac, N.J.B. & Pearse, W. (2012) Caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5. <http://CRAN.R-project.org/package=caper>. [accessed 03/04/2013].
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pearse, W.D. & Purvis, A. (2013) phyloGenerator: an automated phylogeny generation tool for ecologists. *Methods in Ecology and Evolution*, **4**, 692–698.
- Pybus, O.G. & Harvey, P.H. (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society of London Series B*, **267**, 2267–2272.
- Ronqvist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S. *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
- Roquet, C., Thuiller, W. & Lavergne, S. (2013) Building megaphylogenies for macroecology: taking up the challenge. *Ecography*, **36**, 13–26.
- Stadler, T. (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences USA*, **108**, 6187–6192.
- Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R. *et al.* (2009) Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology*, **58**, 573–585.
- Swofford, D.L. & Beagle, D.P. (1993) *PAUP: Phylogenetic Analysis using Parsimony Version 3.1 Users Manual*. Illinois Natural History Survey, Champaign, IL.
- Thomas, G.H. (2008) Phylogenetic distributions of British birds of conservation concern. *Proceedings of the Royal Society of London Series B*, **275**, 2077–2083.

Received 26 July 2013; accepted 17 September 2013

Handling Editor: Emmanuel Paradis