

ECOGRAPHY

Software note

V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants

Yi Jin and Hong Qian

EDITOR'S
CHOICE

Y. Jin ✉ (codon@126.com), School of Life Science, Guizhou Normal Univ., Guiyang, China. – H. Qian (<https://orcid.org/0000-0002-1381-7496>), Research and Collections Center, Illinois State Museum, IL, USA, and CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Inst. of Botany, Chinese Academy of Sciences, Kunming, China.

Ecography

42: 1353–1359, 2019

doi: 10.1111/ecog.04434

Subject Editor: Thiago Rangel

Editor-in-Chief: Miguel Araújo

Accepted 11 March 2019

We present V.PhyloMaker, a freely available package for R designed to generate phylogenies for vascular plants. The mega-tree implemented in V.PhyloMaker (i.e. GBOTB.extended.tre), which was derived from two recently published mega-trees and includes 74 533 species and all families of extant vascular plants, is the largest dated phylogeny for vascular plants. V.PhyloMaker can generate phylogenies for very large species lists (the largest species list that we tested included 314 686 species). V.PhyloMaker generates phylogenies at a fast speed, much faster than other phylogeny-generating packages. Our tests of V.PhyloMaker show that generating a phylogeny for 60 000 species requires less than six hours. V.PhyloMaker includes an approach to attach genera or species to their close relatives in a phylogeny. We provide a simple example in this paper to show how to use V.PhyloMaker to generate phylogenies.

Keywords: phylogeny, software, vascular plants

Background

Species composition of a local or regional assemblage results from the interplay between evolutionary and ecological processes. Because the evolutionary history of each clade (e.g. angiosperms) and the relationships among all members of the clade at various taxonomic levels (e.g. order, family, genus, species) are imprinted in an evolutionary tree (i.e. phylogeny), phylogenies have been increasingly used in ecological and biogeographic studies. Due to the recent availability of tools for building large phylogenies and to conducting sophisticated phylogenetic analyses, the number of phylogeny-based studies in community ecology has exponentially increased in the past two decades (Supplementary material Appendix 1). Recent years have seen an explosion of interest in using phylogenetic relationships to achieve ecological and evolutionary insights.

There are ~391 000 vascular plant species in the world (Royal Botanic Gardens 2016) but only ~20% of these species have been sequenced, according to gene sequence data with GenBank. Because well resolved phylogenies that include all plant species of a study area are rare, botanists and ecologists often use mega-tree approaches to generate plant phylogenies. A broadly used mega-tree approach is to use a mega-tree



available at <<https://github.com/camwebb/tree-of-trees/tree/master/megatrees>> (e.g. R20120829) as a backbone and use Phylomatic and BLADJ implemented in Phylocom (Webb et al. 2008) to assign plant clade age estimates available in the 'wikstrom.ages' file (available at <https://github.com/phylocom/phylocom/blob/master/example_data/bladj/wikstrom.ages>; Wikström et al. 2001) to families or some more basal clades. Because such mega-trees are primarily family-level phylogenetic trees and because the 'wikstrom.ages' file provides age estimates only for 176 internal nodes (Gastauer and Meira-Neto 2013) and includes less than 30% of the families of vascular plants, node ages in plant phylogenies generated with this mega-tree approach are poorly calibrated. Inappropriate calibration of node ages based on the 'wikstrom.ages' file would influence the results of phylogenetic community analysis (Gastauer and Meira-Neto 2013).

Zanne et al. (2014) published a phylogeny that includes 31 749 plant species. This phylogeny was the largest plant phylogeny at that time. Qian and Jin (2016) updated Zanne et al.'s (2014) phylogeny for vascular plants by correcting errors in name spelling, standardizing botanical nomenclature according to The Plant List (<www.theplantlist.org>), and including six families of seed plants that are absent from Zanne et al.'s (2014) phylogeny. The updated phylogeny (i.e. PhytoPhylo.tre; Qian and Jin 2016) includes all families of extant seed plants. In addition, Qian and Jin (2016) developed a software package, S.PhyloMaker, to generate phylogenies for seed plants. This package has been used in over 80 published works (according to Google Scholar, as in 7 January 2019) since it was made available to the public. However, the package has some drawbacks. First, because species of pteridophytes (ferns and fern allies) were not assigned to families in Zanne et al.'s mega-tree, S.PhyloMaker may not be used to generate phylogenies for pteridophytes. Second, while S.PhyloMaker can conveniently generate phylogenies for small- to medium-sized species lists (e.g. a few hundreds of species), it may not work well when generating a large phylogeny with tens of thousands of species and may take a long time to generate a phylogeny for a large species list (e.g. a test of using S.PhyloMaker to generate a phylogeny for a list of 65 536 species took 128 h). Third, as in all available packages that generate phylogenies for plants, S.PhyloMaker does not take into account the relationships of the species and genera in a focal species list that are absent from the mega-tree with those in the mega-tree when the relationships are known in the literature.

Recently, Smith and Brown (2018) published a phylogeny (i.e. GBOTB) that was generated based on molecular data from GenBank and incorporated data from the Open Tree of Life project. GBOTB includes 79 881 taxa of seed plants (i.e. spermatophytes) at and below the species rank, and is the largest dated phylogeny for seed plants. Here, we introduce a new software package, V.PhyloMaker (the prefix 'V' stands for vascular plants), which was implemented with a mega-tree derived primarily from Smith and Brown's (2018) phylogeny for seed plants (i.e. GBOTB) and Zanne et al.'s (2014) phylogeny for pteridophytes. V.PhyloMaker can

generate very large phylogenies for vascular plants at a relatively fast speed.

Package description

This package includes two major components: the R package 'V.PhyloMaker' and the mega-tree 'GBOTB.extended.tre'.

The mega-tree

The mega-tree GBOTB.extended.tre is a combination of GBOTB for seed plants (Smith and Brown, 2018) and the clade in Zanne et al.'s (2014) phylogeny for pteridophytes, with updates, corrections and expansion. The mega-tree is in the Newick format. We took the following steps to generate this mega-tree.

- 1) We extracted the clade of pteridophytes with 518 terminal taxa from Zanne et al.'s (2014) phylogeny and attached it to GBOTB. The divergence time between pteridophytes and seed plants was set according to Zanne et al.'s (2014) phylogeny.

- 2) We standardized spelling and nomenclature of plant names according to The Plant List (TPL, ver. 1.1; <www.theplantlist.org/>), combined infraspecific taxa (e.g. subspecies and variety) with their parent species, and removed duplicate species.

- 3) We determined family-level clades according to TPL for pteridophytes and gymnosperms, and APG4 (Angiosperm Phylogeny Group 2016) for angiosperms. TPL recognized 48 families for pteridophytes (note that Aplenaceae and Dryopteridaceae are misspelled family names of Aspleniaceae and Dryopteridaceae, respectively) and 12 families for gymnosperms while APG4 recognizes 416 angiosperm families. For pteridophytes, we found that three species (i.e. *Bolbitis heteroclita*, *Cyclosorus parasiticus*, *Dicksonia sellowiana*) were nested within members of other families, accordingly, these three species were removed. For seed plants, 122 species were nested within members of other families; they were also removed.

- 4) For pteridophytes, there are thirteen families that are recognized by TPL but are absent from Zanne et al.'s (2014) phylogeny, which are Culcitaceae, Cystodiaceae, Diplaziopsidaceae, Dipteridaceae, Lonchitidaceae, Loxsomataceae, Matoniaceae, Metaxyaceae, Oleandraceae, Plagiogyriaceae, Rhachidosoraceae, Thyrsopteridaceae, and Woodsiaceae. In addition, two fern families are recently recognized (i.e. Desmophlebiaceae and Hemidictyaceae; Mynssen et al. 2016) and are absent from TPL. We added these fifteen families to the mega-tree according to the relationships between them and other families reported in Fitz-Palacios et al. (2011), Lehtonen (2011) and Mynssen et al. (2016), with branch lengths being adjusted proportionally. For angiosperms, the following six families are recognized in APG4 but are absent from GBOTB: Apodanthaceae, Cytinaceae, Mitrastemonaceae, Physenaceae, Rafflesiaceae, and Setchellanthaceae, we added them to the mega-tree

according to the relationships between them and their closely related families reported in Zanne et al. (2014) and Gastauer and Meira-Neto (2017), with branch lengths being adjusted proportionally.

As a result, the final mega-tree (i.e. GBOTB.extended.tre) included 10 587 genera and 74 533 species of vascular plants. GBOTB.extended.tre includes age estimates for all 479 families of extant vascular plants in the world (Supplementary material Appendix 2; note that Isoëtaceae is spelled as Isoetaceae), unlike the ‘wikstrom.ages’ file (Wikström et al. 2001) which includes less than 30% of the families of extant vascular plants. GBOTB.extended.tre is the largest dated mega-tree for vascular plants. All families in the mega-tree are monophyletic. V.PhyloMaker takes this mega-tree as a backbone to generate phylogenies for vascular plants.

The V.PhyloMaker package

Package overview

V.PhyloMaker contains seven functions and four data files. The main function is *phylo.maker*, the other six functions are *at.node*, *bind.relative*, *build.nodes.1*, *build.nodes.2*, *ext.node* and *int.node* (Table 1). (Note that when the user loads V.PhyloMaker in R by the command `library('V.PhyloMaker')`, all these seven functions are automatically loaded.) The four data files are the mega-tree GBOTB.extended, the data frame with tip information (i.e. *tips.info*) that contains the family and genus assignments of every tip species in GBOTB.extended, and the two data frames (i.e. *nodes.info.1* and *nodes.info.2*, respectively generated by *build.nodes.1* and *build.nodes.2*) that contain the genus- and family-level node and age information of GBOTB.extended. The V.PhyloMaker package, which was written in RStudio (RStudio Team 2015), is for use in the R language (R Core Team), requires a standard installation of R (ver. 3.4.0 or newer) and the ‘ape’ package (Paradis et al. 2004). When one loads V.PhyloMaker, the ‘ape’ package will automatically be loaded.

V.PhyloMaker is an open-source software package (published under GPL-2), and is available as a source package V.PhyloMaker_0.1.0.tar.gz. The R package of V.PhyloMaker, together with documentation, is available from GitHub (<<https://github.com/jinyizju/V.PhyloMaker>>).

The user determines whether the function *build.nodes.1* or *build.nodes.2* is used to generate a phylogeny

A genus may not be monophyletic in the mega-tree GBOTB.extended. The user needs to determine whether to use the function *build.nodes.1* or *build.nodes.2* to extract the information of root and basal nodes of the genera in the mega-tree GBOTB.extended (e.g. *tips.info* in the package), which will be used to generate the phylogenetic hypothesis of the user-specified species list, using the functions *phylo.maker* and *bind.relative*. The *build.nodes.1* and *build.nodes.2* functions return the same node information for a genus whose species are distributed in a single cluster (i.e. monophyletic) in the mega-tree. However, if the species of a genus are distributed in more than one cluster, *build.nodes.1* finds and returns the most recent common ancestor (MRCA) of all the tips in the largest cluster of the genus, and define it as the basal node of the genus, whereas *build.nodes.2* finds and returns the MRCA of all the tips of the genus, and define it as the basal node of the genus. In both functions, when the basal node of the genus is determined, the node directly above it is defined as the root node for the genus. In the extreme case that a genus is represented by only one tip in the mega-tree, the basal node of that genus is the same as the root node, which is the MRCA of the genus and its closest genus relative in the mega-tree.

The user determines which scenario is used to generate a phylogeny

The *phylo.maker* function makes phylogenetic hypotheses under three scenarios (i.e. scenarios 1–3), which are the same three scenarios (approaches) as in S.PhyloMaker (Qian and Jin 2016). Specifically, in scenario 1, a new tip is binded to

Table 1 Description of the functions and data.

Name	Description
Function	
<i>at.node</i>	Bind tip to a node in a mega-tree
<i>bind.relative</i>	Bind tip to its designated relative as polytomy
<i>build.nodes.1</i>	Extract the genus- or family-level largest cluster's root and basal node information from the mega-tree
<i>build.nodes.2</i>	Extract the genus- or family-level root and basal nodes information from a phylogeny, regardless of whether the genus or family is monophyletic in the mega-tree
<i>ext.node</i>	Bind tip to a terminal branch in a mega-tree
<i>int.node</i>	Bind tip to an interior branch in a mega-tree
<i>phylo.maker</i>	Generate phylogenies under three scenarios using user-specified species list based on a backbone mega-tree (e.g. GBOTB.extended.tre)
Data	
GBOTB.extended	A phylo object; a combination of GBOTB for seed plants (Smith and Brown 2018) and the clade in Zanne et al.'s (2014) phylogeny for pteridophytes
nodes.info.1	A data.frame; the genus- and family-level nodes information of GBOTB.extended, which was generated by function <i>build.nodes.1</i>
nodes.info.2	A data.frame; the genus- and family-level nodes information of GBOTB.extended, which was generated by function <i>build.nodes.2</i>
tips.info	A data.frame; all the tip species, as well as their genus and family assignments, in GBOTB.extended

genus- or family-level basal node; in scenario 2, the new tip is binded to a randomly selected node at and below the genus- or family-level basal node. In scenario 3, the tip for a new genus is binded to the 1/2 point of the family branch (the branch between the family root node and basal node), unless the family branch length is longer than 2/3 of the whole family branch (from the family root node to the tip) length, the tip of a new genus will be binded to the upper 1/3 point of the whole family branch length. Otherwise, the new tip of an existing genus is binded to the basal node of that genus. In scenarios 2 and 3, if a family has only one tip in the mega-tree, the branch of the tip is evenly divided into the family-, genus- and species-level sections at the 1/3 and 2/3 points of the branch. The tip of a new genus is binded to the upper 1/3 point of the branch. If a genus has only one tip in the mega-tree, the branch of the tip is evenly divided into the genus- and species-level sections at the 1/2 point of the branch. In this case, the tip of a new species is binded to the 1/2 point of the branch. All the three scenarios have been used in the literature; Qian and Jin (2016) conducted an analysis comparing the three scenarios (see Qian and Jin 2016 for details).

Species list preparation

A species list to be used by V.PhyloMaker should be in the format of comma-separated values (csv) with five columns as shown in Supplementary material Appendix 3. The first three columns have to be filled and the last two columns are optional.

Because the spelling and nomenclature of species names in the mega-tree GBOTB.extended were standardized according to TPL, to maximize match in species names between the user's species list and GBOTB.extended, we suggest the user standardize spelling and nomenclature of species according to TPL. All names in the user's species list that are considered as synonyms in TPL should be replaced with their accepted names in TPL. Because taxa with terminal branches in the mega-tree GBOTB.extended are species-level taxa (i.e. binomials), infraspecific taxa (e.g. subspecies, variety and forma) in the user's species list should be combined with their parental species.

Supplementary material Appendix 2 included a complete list of families of the world's flora of vascular plants. All the families in Supplementary material Appendix 2 were included in the mega-tree GBOTB.extended. Each species in the user's species list must be assigned with one of the families shown in Supplementary material Appendix 2. The relationships between genera and families for pteridophytes and gymnosperms should follow TPL except for *Desmophlebium* and *Hemidictyum*, which belong to the families Desmophlebiaceae and Hemidictyaceae, respectively (Mynssen et al. 2016). For angiosperms, the relationships between genera and families are available in Angiosperm Phylogeny Website (<www.mobot.org/MOBOT/research/APweb/>). Species in the user's species list that do not have a family name in column 'family' will not be included in a phylogeny generated by V.PhyloMaker.

Many of the species or genera that are absent from the mega-tree GBOTB.extended are sister to or closely related with some species or genera in the mega-tree GBOTB.extended. When the information of their closely related species or genera is given in the user's species list (columns 'species.relative' and 'genus.relative'), these species or genera will be attached to their closely related species or genera as sister species or genera in a phylogeny generated by V.PhyloMaker. Without the information about relatives, the locations in the mega-tree where they are added may be distant from their close relatives (see details in the 'Example application of V.PhyloMaker' section below).

Package installation

The package V.PhyloMaker is an open source package hosted on GitHub, and is freely available at <<https://github.com/jinyizju/V.PhyloMaker>>. The package can be installed in R either using the *install_github* function in the 'devtools' package (Wickham et al. 2018) or using the *githubinstall* function in the 'githubinstall' package (Makiyama 2018). For example, the R code for installation of V.PhyloMaker using the 'devtools' package is as follows:

```
devtools::install_github("jinyizju/V.
PhyloMaker")
```

Speed of V.PhyloMaker

To generate species lists for evaluating the running speed of V.PhyloMaker, we extracted all 'Accepted' species names (binomials) from TPL and excluded those species which do not belong to the families included in GBOTB.extended, which left with 314 686 species names. This checklist includes about 80% of the estimated vascular plant species in the world (Royal Botanic Gardens 2016). We tested V.PhyloMaker with this very large species list. After running 71.5 h, V.PhyloMaker successfully generated a phylogeny for these species. (The computer that we used to test V.PhyloMaker was an HP EliteBook Laptop with the following features: Windows 7 Enterprise; Intel® Core (TM) i5-2540 CPU @ 2.60 GHz; 4 GB RAM; 64-bit Operating System.) This test indicates that V.PhyloMaker is able to generate very large phylogenies for vascular plants.

We also ran two sets of tests with smaller species lists. In the first set of tests, we randomly extracted five subsets of species with varying numbers of species (i.e. 256, 1024, 4096, 16 384, 65 536) from the 314 686 species in the world checklist described above and used V.PhyloMaker to generate a phylogeny for each subset of species. We repeated this five times (i.e. 25 tests in total). The time that each run took was 0.030 ± 0.001 (mean \pm SD), 0.095 ± 0.001 , 0.355 ± 0.002 , 1.390 ± 0.080 and 5.896 ± 0.106 h, respectively, for the subsets of 256, 1024, 4096, 16 384 and 65 536 species.

In the second set of tests, we fixed the number of species (i.e. 5000) in each species list but varied the number of

species that are absent from the mega-tree GBOTB.extended. Specifically, we first randomly assembled a species list with 5000 species from the world checklist, in which 1000 species are absent from the mega-tree, and used V.PhyloMaker to generate a phylogeny for this species list. We repeated this four times. These resulted in five phylogenies. We then repeated these tests for different numbers of species that are absent from the mega-tree, which are 2000, 3000, 4000 and 5000. In total, we ran 25 tests. The time that each run took increased linearly with increasing the number of species that are absent from the mega-tree. For example, for a species list with 5000 species, when the number of species that are absent from the mega-tree increased from 1000 through 2000, 3000, 4000–5000, the running time was, respectively, 0.030 ± 0.001 (mean \pm SD), 0.132 ± 0.003 , 0.221 ± 0.002 , 0.314 ± 0.001 , 0.411 ± 0.001 and 0.507 ± 0.002 h.

Example application of V.PhyloMaker

The example species list included 92 species in 47 genera and 13 families (Supplementary material Appendix 3). Sixty seven species and 41 genera in this species list are present in the mega-tree GBOTB.extended. The genus *Platycarya*, which is not present in the mega-tree, is, for illustration purpose, considered as a close relative of the genus *Cyclocarya*, which is present in the mega-tree; thus *Cyclocarya* was placed in the column ‘genus.relative’ for *Platycarya strobilacea* in the species list. Similarly, *Carya pallida* is not present in the mega-tree but its closely related species *Carya floridana* is in the mega-tree;

thus *Carya floridana* was placed in the column ‘species.relative’ for *Carya pallida* in the species list. As an example, we used the options of build.nodes.1 and scenario 3 with V.PhyloMaker to generate two phylogenies for this example species list: one used the information in the columns species.relative and genus.relative, and the other did not use the information in the two columns. The two phylogenies were shown in Fig. 1 (also see Supplementary material Appendix 4). The R codes used to generate the phylogenies presented in Supplementary material Appendix 4 are as follows.

```
# load the package
library("V.PhyloMaker")

# input example species list
example <- read.csv("example species
list.csv")

# generate the phylogeny presented in
Figure 1a
tree.a <- phylo.maker(sp.list=exam-
ple, tree=GBOTB.extended, nodes=nodes.
info.1, scenarios="S3")
write.tree(tree.a$scenario.3,
"Figure.1a.tre")

# generate the phylogeny presented in
Figure 1b
rel <- bind.relative(sp.list=exam-
ple, tree=GBOTB.extended, nodes=nodes.
info.1)
```

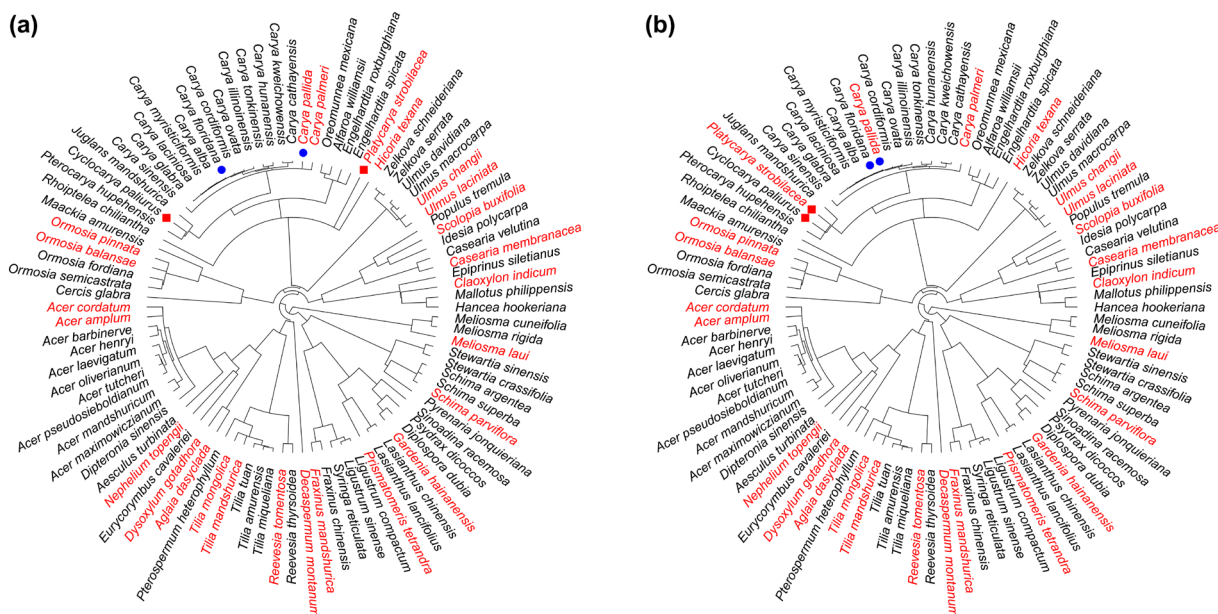


Figure 1. Cladograms of the two phylogenies generated by V.PhyloMaker for the species list in Supplementary material Appendix 3: (a) removing the information in the columns species.relative and genus.relative, and (b) retaining the information in the two columns. Species names in black color are present in the mega-tree GBOTB.extended whereas species names in red color are absent from the mega-tree. The two red square symbols indicate two closely related genera and the two blue circle symbols indicate two closely species within the genus *Carya*. See the main text and Supplementary material Appendix 3 for details. The two phylogenies used to generate cladograms (a) and (b) were shown in Supplementary material Appendix 4.

```
tree.b      <-      phylo.maker(sp.
list=rel$species.list, tree=rel$phylo,
nodes=rel$nodes.info, scenarios="S3")
write.tree(tree.b$scenario.3,
"Figure.1b.tre")
```

As shown in Fig. 1a, without the information in the columns *species.relative* and *genus.relative* of the example species list, V.PhyloMaker placed *Platycarya* in the phylogeny as a basal branch of the family Juglandaceae, which was distant from its close relative genus (i.e. *Cyclocarya*) in the phylogeny, and placed *Carya pallida* in the phylogeny as a basal branch of the genus *Carya*, which was distant from its close relative species (i.e. *Carya floridana*) in the phylogeny. However, with the information in the columns *species.relative* and *genus.relative* of the example species list, V.PhyloMaker placed *Platycarya* close to *Cyclocarya* and placed *Carya pallida* close to *Carya floridana* (Fig. 1b). Thus, we consider the function of V.PhyloMaker to attach genera and species close to their relatives very useful.

Because scenario 2 attaches a new tip to a randomly selected node with its genus or family, there is uncertainty in the insertion of a tip into the backbone tree. To account for such phylogenetic uncertainty, users of V.PhyloMaker can generate a large number of trees by scenario 2, and replicate their analyses using multiple phylogenetic trees (outside V.PhyloMaker) to estimate phylogenetic uncertainty in their study. V.PhyloMaker can simultaneously generate a large number of phylogenetic trees when scenario 2 is used. The R codes for simultaneously generating phylogenetic trees with scenario 2 for the example species list in Supplementary material Appendix 3 are shown in Supplementary material Appendix 5.

Discussion

Recently, Gastauer et al. (2018) published the ComTreeOpt package, a tool for generating phylogenetic trees for plants. V.PhyloMaker differs substantially from ComTreeOpt in many aspects, two of which are highlighted below. First, V.PhyloMaker is based on a dated mega-tree, which was in turn built based on fossil records, molecular data from GenBank and phylogenetic data from the Open Tree of Life (Smith and Brown 2018). It includes all families, the majority (> 10 000) of genera, and over 70 000 species of vascular plants in the world. Thus, when using V.PhyloMaker to generate a phylogenetic tree for vascular plants, all families and the majority of genera in the resultant phylogenetic tree will be resolved. In contrast, when using ComTreeOpt to generate a phylogenetic tree for vascular plants, the resultant phylogenetic tree, which is generated based on phylogenetic data from the Open Tree of Life, is generally not resolved; and a branch for a higher-level clade is evenly divided among its lower-level clades, as shown in Fig. 2 of Gastauer et al. (2018). Second, V.PhyloMaker offers three different ways to attach plant names to nodes of a mega-tree and allows

the user to attach unresolved genera or species to their close relatives; such functions are not available in ComTreeOpt.

The functions in V.PhyloMaker work not only with the mega-tree GBOTB.extended, but also can work with other mega-trees for plants and animals as long as they are properly formatted. Specifically, with a mega-tree and a species list that includes the genus and family assignments of each tip species in the mega-tree, the *build.nodes.1* or *build.nodes.2* function can extract the node information in the mega-tree. Then the mega-tree, together with the node information, could be used in the *phylo.maker* and *bind.relative* functions to make phylogenetic hypotheses for the user-specified species lists.

To cite V.PhyloMaker or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0':

Jin, Y. and Qian, H. 2019. V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. – *Ecography* 42: 1353–1359 (ver. 0).

Acknowledgements – We thank the subject editor Thiago Rangel and two anonymous reviewers for their helpful comments.

Funding – This project was supported by the National Natural Science Foundation of China (grant no. 31600340), and Key Projects of the Joint Fund of the National Natural Science Foundation of China (grant no. U1802232).

References

- Angiosperm Phylogeny Group 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. – *Bot. J. Linn. Soc.* 181: 1–20.
- Fiz-Palacios, O. et al. 2011. Diversification of land plants: insights from a family level phylogenetic analysis. – *BMC Evol. Biol.* 11: 341.
- Gastauer, M. and Meira-Neto, J. A. A. 2013. Avoiding inaccuracies in tree calibration and phylogenetic community analysis using Phylocom 4.2. – *Ecol. Inform.* 15: 85–90.
- Gastauer, M. and Meira-Neto, J. A. A. 2017. Updated angiosperm family tree for analyzing phylogenetic diversity and community structure. – *Acta Bot. Bras.* 31: 191–198.
- Gastauer, M. et al. 2018. Optimizing community trees using the open tree of life increases the reliability of phylogenetic diversity and dispersion indices. – *Ecol. Inform.* 46: 192–198.
- Lehtonen, S. 2011. Towards resolving the complete fern tree of life. – *PLoS One* 6: e24851.
- Makiyama, K. 2018. githubinstall: a helpful way to install R packages hosted on GitHub. – R package ver. 0.2.2, <<https://CRAN.R-project.org/package=githubinstall>>.
- Mynssen, C. M. et al. 2016. *Desmophlebiaceae* and *Desmophlebiium*: a new family and genus of Eupolypod II ferns. – *Taxon* 65: 19–34.
- Paradis, E. et al. 2004. APE: analyses of phylogenetics and evolution in R language. – *Bioinformatics* 20: 289–290.
- Qian, H. and Jin, Y. 2016. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. – *J. Plant Ecol.* 9: 233–239.
- Royal Botanic Gardens, Kew 2016. The state of the World's Plants, 2016. – Royal Botanic Gardens, Kew.

- RStudio Team 2015. RStudio: integrated development for R. – RStudio, Boston, MA, USA, <www.rstudio.com/>.
- Smith, S. A. and Brown, J. W. 2018. Constructing a broadly inclusive seed plant phylogeny. – *Am. J. Bot.* 105: 302–314.
- Webb, C. O. et al. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. – *Bioinformatics* 24: 2098–2100.
- Wickham, H. et al. 2018. devtools: tools to make developing R packages easier. – R package ver. 1.13.6, <<https://CRAN.R-project.org/package=devtools>>.
- Wikström, N. et al. 2001. Evolution of the angiosperms: calibrating the family tree. – *Proc. R. Soc. B* 268: 2211–2220.
- Zanne, A. E. et al. 2014. Three keys to the radiation of angiosperms into freezing environments. – *Nature* 506: 89–92.

Supplementary material (available online as Appendix ecog-04434 at <www.ecography.org/appendix/ecog-04434>). Appendix 1–5.