Ecological Informatics 15 (2013) 85–90

Contents lists available at SciVerse ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Avoiding inaccuracies in tree calibration and phylogenetic community analysis using Phylocom 4.2

Markus Gastauer, João Augusto Alves Meira-Neto *

Laboratory of Ecology and Evolution of Plants, Plant Biology Department, Federal University of Viçosa, Campus UFV s/n, 36570-000, Viçosa, Minas Gerais, Brazil
Floresta-Escola Environmental Science Center, Av. Prof. Mário Palmeiro 1000, 38200-000, Frutal, Minas Gerais, Brazil

ARTICLE INFO

Article history:
Received 9 July 2012
Received in revised form 21 March 2013
Accepted 21 March 2013
Available online 5 April 2013

Keywords:
Mean pairwise distance
Nearest taxon index
Net relatedness index
Phylogenetic diversity
Tree dating
Wikstrom ages

ABSTRACT

There is an increased interest in phylogenetic approaches for conservation biology and community analysis. Many of these analyses are carried out using the Phylocom 4.2 package. With this computational tool, already existing trees are pruned to species from community to be studied. For plant communities, a variety of megatrees including all angiosperm families are available for phylogenetic community analysis. Using the bladj algorithm, internal nodes of community trees derived from these megatrees are calibrated on time scales from fossil or molecular data provided in an ages file. The higher precision of tree calibration, the better is the ecological interpretation if we assume that the closest related species have the most superposed set of traits, the highest probability of co-occurrence in case of environmental filter effects and the highest ratio of competitive exclusion. Together with the Phylocom 4.2 package comes an ages file based on Wikstrom's dating of angiosperm families (wikstrom.ages). But there are inconsistencies in syntax and/or nomenclature between internal node names of trees and the ages file from phylocom that influence the tree calibration and the subsequent analysis. To avoid that, we classified all online available megatrees according to their syntax and nomenclatureof internal nodes. For each of the four classes we provide a new, fully compatible ages file in the supplement material. Each online available megatree, pruned to the species from an example community from the Atlantic Rainforest, was calibrated twice running the bladj algorithm using once the original wikstrom.ages file and, additionally, the new ages file prepared for that tree class. Outcomes from trees calibrated by different methods have been compared. To avoid inconsistencies that push results beyond the realistic, we recommend a strict application of the four ages files provided as supplementary files.

1. Introduction

To gauge the phylogenetic diversity (PD) of a community (i.e. Davies and Buckley, 2011; Donoghue, 2008; Faith, 1992, 2008; Faith et al., 2010; Forest et al., 2007) and to detect non-random processes in community ecology (i.e. Cadotte et al., 2010; Cavender-Bares et al., 2011; Emerson et al., 2011; Paine et al., 2012; Silver et al., 2012), the analysis of phylogenetic community structure produces interesting insights. Thanks to high computing power, available software packages like Phylocom 4.2 (Webb et al. 2002) and numerous different megatrees (Davies et al., 2004; Soltis et al., 1998), this field of research is of growing interests, forming a bridge between community ecology and evolutionary, biogeography and other research areas (e.g. Cavender-Bares et al., 2009; Johnson and Stinchcombe, 2007). Different indices of phylogenetic community structure and diversity have been proposed (Hardy, 2008; Vellend et al., 2011; Webb et al., 2002), and all of them are computed from the matrix of phylogenetic distances, i.e. the evolutionary divergence time, between focused species. Because of their key role, the computation of these phylogenetic distances should be performed with maximum accuracy to avoid misinterpretations of the results. Thus, improving the phylogenetic tree's precision is fundamental for ecological interpretation if we assume that the closest related species have the most superposed set of traits, the highest probability of co-occurrence in case of environmental filtering and the highest ratio of competitive exclusion.

Commonly, phylogenetic distances between single species are computed from age estimates of speciation or radiation events retrieved from molecular and/or fossil data. For a larger number of species, the matrix of phylogenetic distances between all species is commonly computed from a calibrated phylogenetic hypothesis, i.e. a phylogenetic tree, comprising all species in focus in which root and internal nodes are dated referring to such age estimates (Webb et al., 2008).

The Phylocom 4.2 package, built for the phylogenetic analysis of plant communities (Webb et al., 2011), offers the bladj algorithm (from branch length adjustment) to calibrate phylogenetic trees. This bladj algorithm dates the internal nodes of the tree according

---

* Corresponding author at: Laboratory of Ecology and Evolution of Plants, Plant Biology Department, Federal University of Viçosa, Campus UFV s/n, 36570-000, Viçosa, Minas Gerais, Brazil.
E-mail addresses: markus.gastauer@ufv.br (M. Gastauer), j.meira@ufv.br (J.A.A. Meira-Neto).

1574-9541/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.03.005

Table 1
Tree classes, affiliated megatrees, syntax differences between trees and wikstrom.ages, syntax in corrected version and file name of compatible ages versions. Trees marked in bolt were selected to outline the influence of inappropriate tree calibration on outcomes of community analysis.

Class	Trees	Not matching internal node names	Corrected version	File name of correct version
Class Ia	C20050610.new	seedplants[a]	seedplant	agescl1a
	C20061022.new			
	C20070305.new			
	R20050610.new			
	R20061022.new			
	R20070305.new			
Class Ib	R20070606.new	seedplants[a]	seedplant	agescl1b
	R20080401.new	Upper-case for taxa below order level	Lower-case for taxa below order level	
	R20080410.new			
	R20080417.new			
	R20081027.new			
	R20090303.new			
Class II	C20031202.new	seedplants[a]	seedplant	agescl2
	L20030804.new	monocotneudicot	monocot + eudicot	
	L20031115.new	nym2ast	nym-ast	
	R20031202.new	aus2ast	aus-ast	
		bux2ast	bux-ast	
		bursa2anaca	bursa-anaca	
Class III	C20091110.new	seedplant[b]	seedplants	agescl3
	R20091107.new	monocotneudicot	magnoliales_to_asterales	
	R20091120.new	nym2ast	nymphaeales_to_asterales	
	R20100318.new	aus2ast	austrobaileyales_to_asterales	
	R20100701.new	bux2ast	sabiales_to_asterales	
	Online phylomatic	asterid	ericales_to_asterales	
		euasterid2	campanulids	
		euasterid1	lamids	
		eurosid1	fabids	
		eurosid2	malvids	
		Upper-case for taxa below order level	Lower-case for taxa below order level	
Outgroup	R20070305.bl.new	Already dated trees		
	davies04.bl.new			

[a] online wikstrom.ages only; [b] not in online wikstrom.ages

to its ages listed in an ages file so that tree's branch lengths correspond to evolutionary divergence time between ramifications (Webb et al., 2002, 2011). Internal nodes not listed in the ages file are distributed equally between two dated nodes. Because the bladj algorithm is based on the recognition of strings, it matches neither syntax differences nor different nomenclature for internal nodes between phylogenetic trees and the ages file. Furthermore, the algorithm is case sensitive.

For the examination of plant communities, a lot of different phylogenetic hypotheses comprising all accepted plant families, so-called megatrees, are available at http://svn.phylodiversity.net/tot/megatrees. Like the Phylocom 4.2 package, these megatrees are provided by the Phylodiversity Network in cooperation with the Angiosperm Phylogeny Group (http://phylodiversity.net).

These megatrees are pruned to a user supplied list of species (a .txt-file containing family and genus name of each species called t-file in the phylocom terminology) by the phylomatic function of the Phylocom 4.2 package. This function inserts all species as terminals and cuts all branches of the pre-existing megatree without terminals. Like that, phylogenetic trees describing the phylogenetic relationship among a set of species are built immediately.

The Phylocom 4.2 package includes the wikstrom.ages file that provides the age estimates for 176 internal nodes from the angiosperm clade as proposed by Wikstrom et al. (2001). Therefore, the calibration of megatrees pruned to a set of species, so-called community trees, seems to be easy: Just run bladj and your tree is dated!

But tree calibration is aggravated, because syntax and nomenclature of internal nodes of megatrees available at http://svn.phylodiversity.net/tot/megatrees are in many cases incompatible with the wikstrom.ages file from the Phylocom 4.2 package or its online version (http://svn.phylodiversity.net/tot/ages/wikstrom.ages). Therefore, objective of this work is to outline the consequences of inaccurate tree calibration by using incompatible ages files for the following analysis of phylogenetic community structure and to provide consistent alternatives to avoid these problems.

2. Methods

We checked the syntax and nomenclature of internal nodes from all 24 megatrees available on http://svn.phylodiversity.net/tot/megatrees as well as the tree behind the online phylomatic (http://phylodiversity.net/phylomatic/html/pm2_form.html). According to syntax and nomenclature, we grouped these megatrees in different classes. For each class, we compared syntax and nomenclature of internal nodes with that from the orginal wikstrom.ages file from the Phylocom 4.2 package and its online version from http://svn.phylodiversity.net/tot/ages/wikstrom.ages. As we detected at least small differences in syntax between one of the original wikstrom.ages files and all megatrees from each class, we created new ages files for each megatree class that matches all internal nodes for which dating was proposed by Wikstrom et al. (2001).

To test the influence of imprecise tree calibration on the outcomes of phylogenetic community analysis, one megatree form each class was selected. These selected megatrees were pruned to the species list from a plant community from the Forest of Seu Nico, Viçosa, Minas Gerais, Brazil (Gastauer and Meira-Neto, in press) using the phylomatic command. The Forest of Seu Nico is a primary Seasonal Semidecidual Atlantic Rainforest patch of about 35 hectare situated at 20°48' S and 42°51' W. Within this patch, a census of all treelets and trees with a diameter at breast height equal or greater than 3.2 cm within a one-hectare permanent plot was carried out between December 2009 and February 2010. The one-hectare plot was subdivided in 100 subplots of 10 x 10 m. The census comprises 2529 individuals belonging to 218 species from all major angiosperm groups (Gastauer and Meira-Neto, in press).