http://dx.doi.org/10.1590/1519-6984.20814 Original Article

An enhanced calibration of a recently released megatree for the analysis of phylogenetic diversity

M. Gastauer[a,b]* *and J. A. A. Meira-Neto*[a,b]

[a]Laboratório de Ecologia e Evolução de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa – UFV, Campus UFV, s/n, CEP 36570-000, Viçosa, MG, Brazil

[b]Centro de Ciências Ambientais Floresta-Escola – FLORESC, Av. Prof. Mário Palmeiro, 1000, CEP 38200-000, Frutal, MG, Brazil

*e-mail: markus.gastauer@ufv.br

Received: October 2, 2014 – Accepted: March 31, 2015 – Distributed: August 31, 2016
(With 2 figures)

Abstract

Dated or calibrated phylogenetic trees, in which branch lengths correspond to evolutionary divergence times between nodes, are important requirements for computing measures of phylogenetic diversity or phylogenetic community structure. The increasing knowledge about the diversification and evolutionary divergence times of vascular plants requires a revision of the age estimates used for the calibration of phylogenetic trees by the bladj algorithm of the Phylocom 4.2 package. Comparing the recently released megatree R20120829.new with two calibrated vascular plant phylogenies provided in the literature, we found 242 corresponding nodes. We modified the megatree (R20120829mod.new), inserting names for all corresponding nodes. Furthermore, we provide files containing age estimates from both sources for the updated calibration of R20120829mod.new. Applying these files consistently in analyses of phylogenetic community structure or diversity serves to avoid erroneous measures and ecological misinterpretation.

Keywords: R20120829.new, phylogenetic community analysis, ecological misinterpretation, phylogenetic diversity, evolutionary divergence times, phylogenetic dispersion, phylogeny of vascular plants.

Calibração aperfeiçoada de uma megatree lançada recentemente para a análise da diversidade filogenética

Resumo

Árvores filogenéticas datadas, ou calibradas, em que os comprimentos dos ramos correspondem ao tempo evolutivo de divergência entre os nós, são importantes requisitos para calcular medidas de diversidade filogenética ou de estrutura filogenética de comunidades. O conhecimento crescente sobre a diversificação e sobre o tempo de divergência evolutiva das plantas vasculares fez necessária uma revisão das estimativas de idades dos nós que são utilizadas para a calibração de árvores filogenéticas por meio do algoritmo bladj do pacote Phylocom 4.2. Comparando a mega-árvore R20120829.new, recentemente publicada, e outras duas filogenias calibradas de plantas vasculares, encontramos 242 nós correspondentes. Modificamos esta mega-árvore (R20120829mod.new), inserindo todos os nomes dos nós correspondentes. Além disso, providenciamos dois arquivos com todas as estimativas das idades para uma calibração mais atualizada. Utilizando esses arquivos de maneira consistente nas análises de diversidade ou de estrutura filogenética de comunidades, evita-se incorreções nas datações e imprecisões na interpretação de informações ecológicas.

Palavras-chave: R20120829.new, análise filogenética de comunidades, má interpretação ecológica, diversidade filogenética, tempo de divergência evolutiva, dispersão filogenética, filogenia de plantas vasculares.

1. Introduction

Dated or calibrated phylogenetic trees, in which branch lengths correspond to evolutionary divergence times, are important requirements for computing measures of phylogenetic diversity (Faith, 1992) or phylogenetic community structure (Webb et al., 2002; Gastauer and Meira-Neto, 2014). The importance of correct calibration has been highlighted to avoid ecological misinterpretation, i.e., the outcome of a correctly calibrated tree indicates a phylogenetic community structure different from the outcome of a wrongly calibrated tree (Gastauer and Meira-Neto, 2013), but achieving an updated calibration is still not an easy task. Furthermore, constantly increasing knowledge about the phylogenetic relationships among vascular plants, especially angiosperms (Smith et al., 2011), and about the diversification times of different clades (i.e., Magallón and Castillo, 2009; Lemaire et al., 2011; Magallón et al., 2013) require regular revisions of the database used to achieve updated tree calibrations.

The rapid increase in information results in the periodic release of new, updated megatrees, i.e., phylogenetic hypotheses containing all euphyllophyte families (GitHub, 2014). For phylogenetic community analyses or the computation of phylogenetic diversity measures, these megatrees may be pruned down to a user-supplied list of species by the phylomatic function of the Phylocom 4.2 package (Webb and Donoghue, 2005). By this procedure, all species are inserted as terminals, and branches of the megatree without terminals are cut. Internal node names are maintained. For the calibration of such community trees, the bladj algorithm from the Phylocom package is run (Webb et al., 2008). This algorithm recognizes the names of internal nodes and dates them according to clade dating information provided in a separate ages file. Not matched internal nodes are smoothed by mean dates between fixed divergence times.

The Phylocom package includes the wikstrom.ages file, which provides the minimum age estimates for 176 internal nodes within the angiosperms proposed by Wikström et al. (2001). Revised age estimates are available (e.g. Bell et al., 2010), but still not incorporated within the Phylocom package.

Therefore, the aim of this note is to outline how community trees derived from the megatree R20120829.new, recently released for phylogenetic community analysis and the computation of phylogenetic diversity among vascular plants, are calibrated in the most updated way. For that, we identified and provided recent age estimates from literature of a maximum number of internal nodes that are distributed over the entire megatree.

2. Material and Methods

We compared the topology of R20120829.new with two comprehensive revisions about vascular plant diversification (Hedges and Kumar, 2009; Bell et al., 2010). Corresponding nodes, i.e., nodes with the same descending clades or taxa, were identified. If not already named by the authors of R20120829.new, these corresponding nodes were named within the plain text archive of the megatree.

Age estimates from both revisions were pooled to provide updated ages files for tree calibration. In several cases, two or three subsequent nodes within the same lineage were estimated to identical ages by literature. In these cases, the bladj algorithm calibrates the node that appears first in the ages file and ignores age estimates for the other node(s), smoothing by mean dates between fixed divergence times. This causes distortions of the tree altering patterns of phylogenetic diversity, because the subsequent node(s) is/are always fixed to more recent divergence times than hypothetically assumed by literature. To avoid that, we corrected the age estimates of the most recent node of the sequence by –0.1 Myr, while in case of three subsequent nodes with identical age estimations, the oldest one is corrected by +0.1 Myr. Due to this procedure, the topology of R20120829.new is conserved.

To illustrate advances in tree calibration, age estimates of these corresponding nodes were used to calibrate a hypothetic community phylogeny that contains two species from each monophyletic family from R20120829.new. Resulting trees are visualized with FigTree v1.4.2.

3. Results and Discussion

Bell et al. (2010) revised Wikstrom's age estimates by dating an angiosperm phylogeny using a relaxed molecular clock calibrated by 36 fossils considered as minimum ages of the most recent common ancestor. They used two slightly different approaches to estimate node ages, one in which they included these minimum age constraints from fossil data into an exponential distribution (BEAST^a analysis in the paper's terminology) and one in which they used a lognormal distribution (BEAST^b), two different ages files are provided. This results in two sets of calibration points labeled as ages_exp and ages_logn available from the authors in the Supplementary Material*; either of them might be used to arrive at a revised calibration of the angiosperm clade in a user-specific community tree, given the user is aware of the possible shortcomings and pitfalls of both types of approaches. Furthermore, age estimates for gymnosperms and ferns, among other clades, have been compiled by Hedges and Kumar (2009).

We identified 242 corresponding nodes (Table 1). Of these nodes, 26 were identified by comparing gymnosperm and fern clades from Hedges and Kumar (2009) with R20120829.new. The other nodes represented correspondences between Bell et al. (2010) and R20120829.new. Node names lacking in the plain text archive of the megatree were added. This modified R20120829mod.new is available in the Supplementary Material*.

Age estimates of all 242 nodes are available as ages files from the authors (Table 1). Four groups of subsequent nodes within the same lineages were estimated to have identical ages by Bell et al. (2010). As shown in Table 2, we corrected the age estimates by + or –0.1 Myr to avoid distortion of the community tree.

Calibrating the modified version R20120829mod.new with provided ages files produces a phylogeny that differs substantially from the original file (R20120829.new) calibrated with the age estimates from Wikström et al. (2001), especially regarding gymnosperms and ferns (Figures 1 and 2), because these clades were previously not covered by the wikstrom.ages file. Important to note is that difference in trees results from varying divergence times, while tree topology is maintained. Such differences indicate that phylogenetic community analyses can generate relatively different results depending on the chosen databases, which can lead to ecological misinterpretation if based on incorrect data (Gastauer and Meira-Neto, 2013).

---

* Supplementary Material: http://www.leep.ufv.br/pt-BR/noticia/pesquisadores-da-floresta-escola-lancam-artigo-cientifico-sobre-computacao-da-diversidade-filogenetica

Table 1. Age estimates (ages_exp[1] and ages_logn[2]) of corresponding nodes between R20120829.new and vascular plant phylogenies provided by Hedges and Kumar (2009) as well as Bell et al. (2010) using syntax of R20120829mod.new and their comparison to former age estimates given by Wikström et al. (2001, wikstrom.ages). Nodes with age estimates according to Hedges and Kumar (2009) are marked with asterisks*.

Node name	wikstrom. ages	ages_exp[1]	ages_logn[2]
euphyllophyte*	400	466	466
monilophyte*		364	364
equisetales_to_polypodiales*		360	360
seedplants*		355	355
gymnosperms*		354	354
ginkgoaceae_to_gnetales*		346	346
osmundales_to_polypodiales*		323	323
ophioglossaceae_to_psilotaceae*		306	306
gnetales_to_pinales*		298	298
pinales*		288	288
hymenophyllaceae_to_polypodiales*		286	286
cycadales*		283	283
lygopodiaceae_to_polypodiales*		266	266
gleicheniales*		263	263
araucariaceae_to_podocarpaceae*		257	257
taxaceae_to_cupressaceae*		227	227
marsileaceae_to_polypodiales*		220	220
schizaeales*		212	212
cyatheales_to_polypodiales*		211	211
cyatheales*		183	183
angiosperms	179	147	183
polypodiales*		176	176
salviniales		173	173
austrobaileyales_to_asterales	165	144	173
gnetales*		159	159
ceratophyllales_and_eudicots		129	156
dennstaedtiaceae_to_polypodiaceae*		151	151
magnoliales_to_asterales	162	130.1	146
monocots		129.9	146
poales_to_asterales	161	130	146
narthecidae		118	136
eudicots	147	128.9	129
sabiales_to_asterales	137	126	129
subrosid	121	116.8	128
petrosaviidae		109	124
alismatales	124	107	122
magnoliids		125	122
core_eudicots	127	117	121
pentapetalids		116.9	120
subasterids		116.8	120
laurales	111	112	119
piperales	132	104	119
proteales	135	110	116
rosids		108	116
malvids	95	107	115
liliales	96	86	114
asterids		104	110
ericales_to_asterales	117	104	110
lamiids_to_campanulids		99	108
ranunculales	140	100	108
fabids	98	103	107

[1]Exponential distribution of minimum age constraints from fossil data. [2]Lognormal distribution.

Table 1. Continued...

Node name	wikstrom. ages	ages_exp[1]	ages_logn[2]
caryophyllales	84	99	106
aristolochiaceae	108	91	104
celastrales_to_malpighiales		101	104
lamiids		96	104
asparagales	107	92	103
commelinids		96	103
saxifragales	111	95	103
campanulids	107	93	100
buxaceae	97	98	99
nitrogenfixing		99	99
platanaceae_to_proteaceae		98	99
santalales	69	91	99
calycanthaceae	52	98	98
chloranthaceae	132	121	98
fagales_to_cucurbitales		96	96
aspleniaceae_to_onocleaceae		95	95
drypteridaceae_to_polypodiaceae		94	94
geraniales	92	87	93
oxalidales	77	89	93
poales	72	85	93
crossosomatales	62	84	92
dioscoreales	95	83	92
ericales	100	92	92
malpighiales	81	89	92
araceae	98	79	89
austrobaileyales		89	89
myrtales	88	89	89
sabiaceae	129	87	89
papaveraceae	112	82	88
aquifoliales	97	88	87
cornales	101	87	87
lardizabalaceae_to_circaeasteraceae		81	87
rosales	76	82	85
asterales	90	77	84
zingiberales	62	84	84
fabales	79	79	83
malvales	68	78	83
brassicales	79	83	82
pandanales	88	72	82
achariaceae		79	81
canellales	99	77	80
garryales	93	70	77
celastrales	42	71	76
commelinales	68	70	76
magnoliales	113	69	76
solanales	78	71	76
lamiales	63	69	74
linaceae	33	71	73
escalloniaceae		65	72
ranunculaceae_to_berberidaceae		67	72
passifloraceae	26	68	71
sapindales	61	71	71
stylidiaceae		65	71

[1]Exponential distribution of minimum age constraints from fossil data. [2]Lognormal distribution.

Table 1. Continued...

Node name	wikstrom. ages	ages_exp[1]	ages_logn[2]
gentianales	71	65	69
cornaceae	64	67	67
cucurbitales	65	61	67
melanthiaceae		59	67
moringaceae_to_caricaceae		64	67
poaceae_to_flagellariaceae		58	65
bursa_to_anaca	51	50	64
tofieldiaceae		61	64
fabaceae	56	61	63
rhizophoraceae		60	63
salicaceae	23	61	63
rhamnaceae	55	59	62
solanaceae_to_convolvulaceae		59	62
cyperaceae_to_juncaceae		55	61
primulaceae		57	61
dipsacales	81	57	60
malpighiaceae	32	61	60
plumbaginaceae_to_polygonaceae		58	60
amaranthaceae_to_caryophyllaceae		55	59
ranunculaceae	65	55	59
dilleniaceae	52	55	58
liliaceae	48	52	58
boraginaceae	59	54	57
celastraceae	34	53	57
rubiaceae	56	57	57
campanulaceae	33	53	56
cannabaceae_to_moraceae		54	56
gesneriaceae	38	52	56
fagales	61	52	55
geraniaceae	38	48	54
meliaceae_to_rutaceae		53	54
saururaceae	75	47	54
styracaceae_to_diapensiaceae		51	54
apiales	69	49	53
argophyllaceae_to_alseuosmiaceae		49	53
scrophulariaceae		51	53
tecophilaeaceae		45	52
penthoraceae_to_haloragaceae		48	51
crassulaceae	41	47	50
lythraceae	37	46	50
santalaceae	37	43	50
ochnaceae		49	49
circaeasteraceae	54	45	48
lecythidaceae	65	46	48
orchidaceae	26	42	48
vitaceae	92	65	48
menyanthaceae	51	44	47
plantaginaceae		42	46
polygalaceae	28	44	46
helwingiaceae_to_aquifoliaceae		42	45
oleaceae	47	41	45
asparagaceae	27	42	44

[1] Exponential distribution of minimum age constraints from fossil data. [2] Lognormal distribution.

Table 1. Continued...

Node name	wikstrom. ages	ages_exp[1]	ages_logn[2]
rosaceae	47	40	44
asteraceae	44	40	43
brassicaceae_to_cleomaceae		41	43
hydrangeaceae	78	44	43
plumbaginaceae	27	43	43
dasypogonaceae		38	42
elaeocarpaceae	59	38	42
lentibulariaceae		37	42
nymphaeales		38	42
polygonaceae	28	41	42
sapindaceae	36	41	42
strelitziaceae_to_lowiaceae		40	42
acanthaceae	31	38	41
dioncophyllaceae_to_ancistrocladaceae		37	41
monimiaceae	71	35	41
lamiaceae	23	38	40
piperaceae	43	32	40
rutaceae	39	40	40
amaranthaceae	19	37	39
malvaceae	34	66	39
lardizabalaceae	77	35	38
meliaceae	30	39	38
sarcolaenaceae_to_dipterocarpaceae		35	38
saxifragaceae	51	38	38
solanaceae	41	37	38
styracaceae	45	36	38
cannabaceae		36	37
caprifoliaceae	10	36	37
cyperaceae	16	32	37
oxalidaceae	43	34	37
berberidaceae	44	33	36
magnoliaceae	70	33	36
musaceae	50	34	36
polemoniaceae	35	32	36
proteaceae	47	35	36
thymelaeaceae	26	36	36
melianthaceae		34	35
menispermaceae	53	33	35
moraceae	23	33	35
urticaceae	22	34	35
iridaceae	32	31	34
adoxaceae	57	31	33
arecaceae	73	31	33
brassicaceae	24	31	33
cyclanthaceae	45	30	33
loasaceae	67	31	33
amaryllidaceae	28	30	32
apiaceae	33	29	32
juglandaceae_to_myricaceae		29	32
philydraceae		29	32
fagaceae	34	28	31
poaceae	12	28	31
verbenaceae	20	29	30

[1] Exponential distribution of minimum age constraints from fossil data. [2] Lognormal distribution.

Table 1. Continued...

Node name	wikstrom. ages	ages_exp[1]	ages_logn[2]
cunoniaceae	42	27	29
hamamelidaceae	104	25	27
crossosomataceae_to_stachyuraceae		24	26
annonaceae	26	23	25
bignoniaceae	38	25	25
convolvulaceae	20	24	25
onagraceae	20	23	25
strelitziaceae	29	23	25
combretaceae	27	21	24
haloragaceae	17	23	24
apocynaceae	18	21	23
cucurbitaceae	20	21	23
nyctaginaceae	13	22	23
cactaceae_to_portulaceae		21	22
bromeliaceae	33	20	21
costaceae	27	19	21
betulaceae	19	18	20
elaeagnaceae	13	20	20
trochodendraceae	106	19	20
araliaceae	26	18	19
winteraceae	42	18	19
zingiberaceae	10	18	19
marantaceae	23	17	18
chrysobalanaceae	8	16	17
ericaceae	12	14	15
lauraceae	34	12	13
limnanthaceae	9	12	13
pittosporaceae	14	11	12
myristicaceae	23	12	11
canellaceae	17	10	10
altingiaceae	56	7	7
akaniaceae	31	6	6
juglandaceae	6	4	4

[1]Exponential distribution of minimum age constraints from fossil data. [2]Lognormal distribution.

Table 2. Subsequent nodes in R20120829mod.new with identical age estimates according to Bell et al. (2010) exponential distribution (BEAST[a]) or lognormal distribution (BEAST[b]) of minimum age constraints plus corrected diversification time in ages_exp and ages_logn (available from Supplementary Material*).

Node name	Bell's BEAST[a] age estimate	Corrected version (ages_exp)	Bell's BEAST[b] age estimate	Corrected version (ages_logn)
magnoliales_to_asterales	130.0	130.1	146	146.1
poales_to_asterales	130.0		146.0	
monocots	130.0	129.9	146.0	145.9
caryophylales_and_eudicots			156.0	145.9
caryophylales_and_eudicots	129.0		156.0	145.9
eudicots	129.0	128.9	129.0	
core_eudicots	117.0	117.1	121.0	
pentapetalids	117.0		120.0	
subasterids, subrosids	117.0	116.9	120.0	119.9
asterids			110.0	
ericales_to_asterales			110.0	109.9

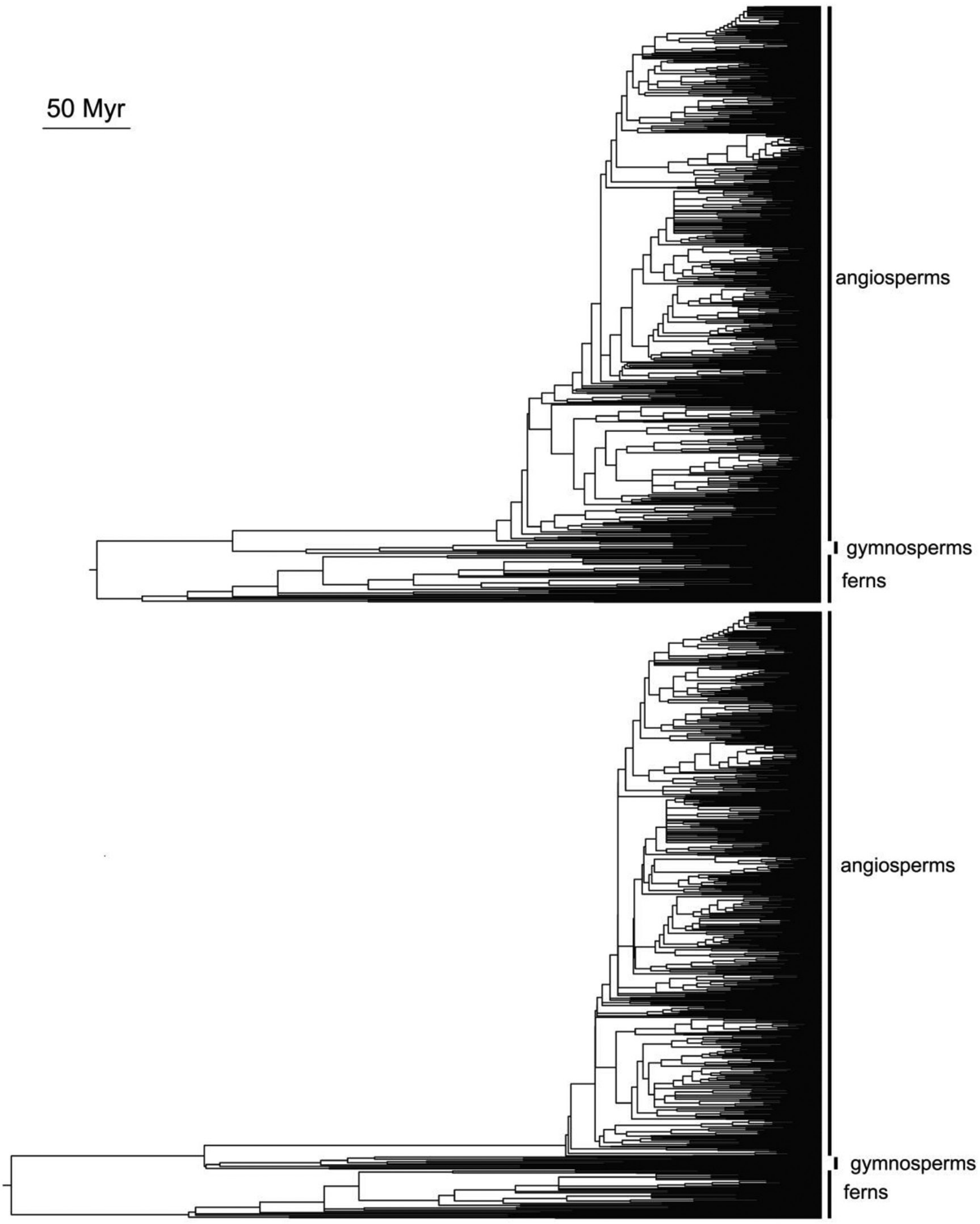

Figure 1. Hypothetical trees of a community composed of two species from all monophyletic families from APG III (APG III, 2009) that were inserted in the original megatree R20120829.new that was calibrated by agesclIII (see Gastauer and Meira-Neto, 2013 for details, upper tree) and R20120829mod.new calibrated by ages_exp (lower tree).

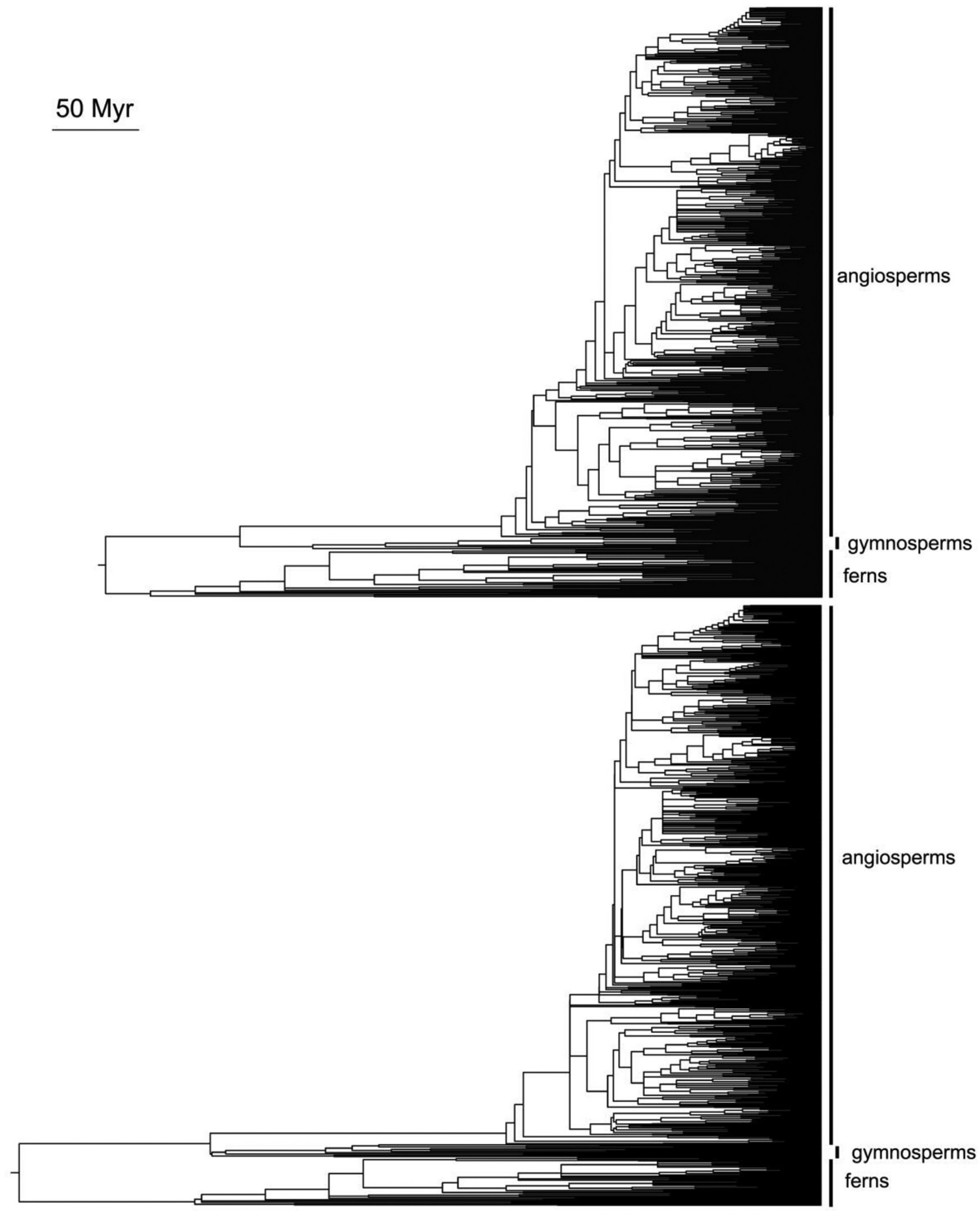

Figure 2. Hypothetical community trees containing two species of all monophyletic families from APG III (APG III, 2009) and resulting from insertion in the original megatree R20120829.new that was calibrated by agesclIII (see Gastauer and Meira-Neto, 2013 for details, upper tree) and R20120829mod.new calibrated by ages_logn (lower tree).

4. Conclusions

Differences in tree branch calibrations may produce different results regarding phylogenetic diversity and/or change the ecological interpretation of measures of phylogenetic community structure. Our goal was to gather and make available the most updated vascular plant phylogeny as well as updated minimum divergence times, as available from the recent literature. For that, we recommend the consistent application of the R20120829mod.new file in combination with ages_exp or ages_logn (available in the Supplementary Material*). This ensures a more inclusive and updated calibration of phylogenetic trees including not only angiosperms, but also gymnosperms and ferns, due to a higher number of internal nodes being dated, distributed among the entire tree of euphyllophytes. Furthermore, users have the possibility to choose between underlying statistical extrapolation of age estimates in their studies of evolutionary ecology, phylogenetic comparisons and analysis of most recent common ancestor analysis in plants.

Acknowledgements

We are grateful to the Floresta-Escola Environmental Science Center, SECTES-MG, CNPq, FAPEMIG and CAPES for grants and scholarships.

References

ANGIOSPERM PHYLOGENY GROUP – APG III, 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, vol. 161, no. 2, pp. 105-121. http://dx.doi.org/10.1111/j.1095-8339.2009.00996.x.

BELL, C.D., SOLTIS, D.E. and SOLTIS, P.S., 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany*, vol. 97, no. 8, pp. 1296-1303. http://dx.doi.org/10.3732/ajb.0900346. PMid:21616882.

FAITH, D.P., 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation*, vol. 61, no. 1, pp. 1-10. http://dx.doi.org/10.1016/0006-3207(92)91201-3.

GASTAUER, M. and MEIRA-NETO, J.A.A., 2013. Avoiding inaccuracies in tree calibration and phylogenetic community analysis using Phylocom 4.2. *Ecological Informatics*, vol. 15, no. 1, pp. 85-90. http://dx.doi.org/10.1016/j.ecoinf.2013.03.005.

GASTAUER, M. and MEIRA-NETO, J.A.A., 2014. Interactions, environmental sorting and chance: phylostructure of a tropical forest assembly. *Folia Geobotanica*, vol. 49, no. 3, pp. 443-459. http://dx.doi.org/10.1007/s12224-013-9181-1.

GITHUB, 2014 [viewed 3 January 2014]. [online]. Available from: https://github.com/camwebb/tree-of-trees/tree/master/megatrees

HEDGES, S.B. and KUMAR, S., 2009. *The time tree of life*. New York: Oxford University Press.

LEMAIRE, B., HUYSMANS, S., SMETS, E. and MERCKX, V., 2011. Rate accelerations in nuclear 18S rDNA of mycoheterotrophic and parasitic angiosperms. *Journal of Plant Research*, vol. 124, no. 5, pp. 561-576. http://dx.doi.org/10.1007/s10265-010-0395-5. PMid:21188459.

MAGALLÓN, S. and CASTILLO, A., 2009. Angiosperm diversification through time. *American Journal of Botany*, vol. 96, no. 1, pp. 349-365. http://dx.doi.org/10.3732/ajb.0800060. PMid:21628193.

MAGALLÓN, S., HILU, K.W. and QUANDT, D., 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, vol. 100, no. 3, pp. 556-573. http://dx.doi.org/10.3732/ajb.1200416. PMid:23445823.

SMITH, S.A., BEAULIEU, J.M., STAMATAKIS, A. and DONOGHUE, M.J., 2011. Understanding Angiosperm diversification using small and large phylogenetic trees. *American Journal of Botany*, vol. 98, no. 3, pp. 404-414. http://dx.doi.org/10.3732/ajb.1000481. PMid:21613134.

WEBB, C.O. and DONOGHUE, M.J., 2005. Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes*, vol. 5, no. 1, pp. 181-183. http://dx.doi.org/10.1111/j.1471-8286.2004.00829.x.

WEBB, C.O., ACKERLY, D.D., MCPEEK, M.A. and DONOGHUE, M.J., 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, vol. 33, no. 1, pp. 475-505. http://dx.doi.org/10.1146/annurev.ecolsys.33.010802.150448.

WEBB, C.O., ACKERLY, D.D. and KEMBEL, S.W., 2008. Phylocom: software for the analysis of phylogenetic community structure and character evolution. *Bioinformatics*, vol. 24, no. 18, pp. 2098-2100. http://dx.doi.org/10.1093/bioinformatics/btn358. PMid:18678590.

WIKSTRÖM, N., SAVOLAINEN, V. and CHASE, M.W., 2001. Evolution of angiosperms: calibrating the family tree. *Proceedings. Biological Sciences*, vol. 268, no. 1482, pp. 2211-2220. http://dx.doi.org/10.1098/rspb.2001.1782. PMid:11674868.