New Zealand Journal of Botany
Vol. 50, No. 3, September 2012, 257–265

Taylor & Francis
Taylor & Francis Group

Variability in phylogenetic diversity (PD) estimates illustrated with plant data for the high Andes of South America

RA Scherson[a]*, PI Naulin[a], AA Albornoz[b], T Hagemann[c] and MTK Arroyo[c]

[a]*Departamento de Silvicultura y Conservación de la Naturaleza, Universidad de Chile, Casilla, Santiago, Chile;* [b]*Center for Advanced Studies in Ecology and Biodiversity (CASEB), Pontificia Universidad Católica de Chile, Casilla, Santiago, Chile;* [c]*Instituto de Ecología y Biodiversidad, Universidad de Chile, Casilla, Santiago, Chile*

(*Received 18 October 2011; final version received 24 January 2012*)

Phylogenetic diversity (PD) is commonly calculated by adding together the branch lengths or ages of a subset of taxa present in an area, using a single phylogeny as a backbone. However, a phylogenetic hypothesis is the consensus of a range of equally likely trees, inherently variable in branch lengths and sometimes in topology. This study incorporates confidence intervals into PD calculations in order to account for such variability. Using the genera of Fabaceae and Solanaceae present in the high Andes, we calculated PD for three macro-zones (Puna, Paramo and Southern Andean Steppe) and studied its correlation with generic richness for these areas. We found a similar pattern between PD and richness in the Fabaceae, but not in the Solanaceae. Variability proved useful in interpreting the results, especially in the Solanaceae which showed alternative topologies. Further studies are needed to address the possible effects of this variability on the PD index.

Keywords: Fabaceae; high Andes; hotspot; phylogenetic diversity; Solanaceae; South America; species richness

Phylogenetic diversity (PD) measures assess the evolutionary pathways that connect taxa in a geographical area (Faith 1992) and address the question of how much evolutionary history would be lost from an area if its biodiversity were not preserved (Faith 1992; Purvis & Hector 2000). PD has been measured on a dated phylogeny or a phylogram as the sum of millions of years or branch lengths contained within the subtree of the taxa present in an area (Vane-Wright et al. 1991; Diniz-Filho et al. 2004; Faith & Baker 2006; Allen et al. 2009). These measures are commonly obtained from a single phylogenetic tree, chosen from a set of equally likely phylogenies, with their inherent differences in branch lengths and/or topology. Even if the chosen tree is very well supported, by using a single set of branches, variability among trees is overlooked (but see examples of the use of variability in Swenson 2009; Thuiller et al. 2011). Because PD calculations consider both branch lengths and the topology of the tree, we considered both of these factors, generating PD calculations that incorporate confidence intervals. For simplicity in the interpretation of the results, the PD of an area was calculated in terms of the relative or proportional contribution of an area to the PD of the full phylogeny (as in Potter 2008)

Another recurrent source of discussion is whether taxon richness is a good predictor of PD (Rodrigues & Gaston 2002; Torres &

*Corresponding author. Email: rscherson@uchile.cl

ISSN 0028-825X print/ISSN 1175-8643 online

http://dx.doi.org/10.1080/0028825X.2012.665808
http://www.tandfonline.com

Diniz-Filho 2004; Forest et al. 2007; McGoogan et al. 2007). It is now recognised that this is not always the case, due to the many processes that influence patterns of speciation and radiation in an area (Diniz-Filho et al. 2004; Forest et al. 2007).

Using two plant families well represented in the high Andes of South America, Fabaceae and Solanaceae, we present a contribution to the calculation and interpretation of PD and address the question of the correspondence between taxon richness and the evolutionary history contained by such taxa in a given area.

Materials and methods

Study area and taxon sampling

Lists of genera were compiled for three Andean sectors altitudinally roughly corresponding to: Páramo, hereafter PAR; Puna, hereafter PUN; and Southern Andean Steppe, hereafter SAS, using the literature (Brako & Zarruchi 1993; Luteyn 1999; Ulloa et al. 2004; Zuloaga et al. 2008), miscellaneous floristic lists, taxonomic treatments and herbarium records (SGO, CONC, TROPICOS).

To the extent of our knowledge, all genera of Fabaceae and Solanaceae used in this study are monophyletic. Thus, we chose one species for every genus present in the high Andes, depending upon the availability of material. For the Fabaceae, we relied mostly on previously sequenced species available for the published *matK* family-level phylogeny (Wojciechowski et al. 2004). For the Solanaceae, we obtained sequences available in GenBank for the commonly used chloroplast spacer trnL-trnF. Sequences not available in the literature or GenBank were obtained directly in our laboratory (Appendix).

Outgroups for Fabaceae were *Suriana maritima* L. (Surianaceae) and *Quillaja saponaria* Molina (Quillajaceae), and for Solanaceae, we used *Astripomoea malvacea* (Klotzsch) A. Meeuse (Convulvulaceae) and *Convolvulus arvensis* L. (Convulvulaceae).

DNA amplification and sequencing

Genomic DNA was isolated from herbarium dried specimens using the DNeasy Plant Minikit (Qiagen, Valencia, CA, USA) following the manufacturer's recommendations, slightly modified for herbarium material. The chloroplast *matK* and flanking 3′ *trnK* intron region were obtained by polymerase chain reaction (PCR) amplification using primers trnK685F and trnK2R* according to the protocol of Wojciechowski et al. (2004). For the trnL-F chloroplast spacer, we used the primers TabF and TabC and the protocol of Shaw et al. (2005).

PCR products were purified and sequenced using Applied Biosystems sequencers ABI3700 and ABI3730XL at Macrogen Inc. (Seoul, Korea), using the primers described previously, plus the internal primers matK4LaF, matK832R, matK1100L and matk1932Ra (Wojciechowski et al. 2004) for the *matK* gene. Electropherograms of the sequenced products were visualised using FinchTV1.4.0 (Geospiza Inc., Seattle, WA, USA) and the resulting sequences were exported as FASTA text files to the alignment software.

Phylogenetic analyses

Previously available and newly generated DNA sequences were aligned using ClustalX (Thompson et al. 1997) and the alignment editor Se-Al v2.0a11 (Rambaut 1996). After removal of ambiguous alignment, we obtained a 1614 bp matrix for the Fabaceae containing all 17 genera, and a 924 bp matrix containing 21 of the 23 genera of the Solanaceae present in the high Andes. The evolutionary model that best fitted the data was determined using Modeltest (Posada & Crandall 1998) based on the Akaike information criterion (AIC) (Posada & Buckley 2004).

MrBayes 3.1 (Ronquist & Huelsenbeck 2003) was used to perform the Bayesian phylogenetic analyses. Posteriors on the model parameters were estimated from the data, using the default priors. The analyses were carried out

using two million generations, sampling the Markov chain every 100 generations, and using four independent chains running simultaneously. Stationarity of the Markov chain was inferred using the following indicators: (1) a stable value of the log likelihood of the cold chain in two separate runs; (2) a value approaching zero for the standard deviation between runs; and (3) a value approaching 1.0 for the potential scale reduction factor (PSRF) for each parameter in the model.

Calculation of PD

PD per Andean sector was calculated as the sum of the branches of the subtree formed by the taxa present in each area. This was expressed as a percentage of total PD in the full phylogeny, which in turn was calculated as the sum of all branches in the phylogeny. PD per area was calculated using a script in R (R-Team DCT 2011) with an adaptation of the matrix-based method of Rodrigues & Gaston (2002). Their Imatrix was used as a PD vector in which the columns contain each tree returned by the Bayesian analysis and used to calculate the confidence intervals, and the rows contain the length of each branch segment.

Variability in PD calculation

Confidence intervals were calculated for PD on 35 equally likely trees obtained at random from the set of trees generated by the Bayesian analysis after removing 25% to account for stabilisation of the likelihood values. For the Fabaceae, the trees were identical topologically but varied in their branch lengths. For the Solanaceae, equally likely trees had an assortment of different topologies. An ANOVA was performed to compare the relative PDs among Andean areas. Arcsine transformation (Sokal & Rohlf 1981) was used to normalise the percentage data.

Finally, taxon richness was calculated as the percentage contribution of genera in each high Andean area in relation to the total number of genera in Fabaceae or Solanaceae in the three areas.

Results

Phylogenies

Fig. 1 shows the phylogeny obtained for the Fabaceae. Congruence of the clades was compared with the published Fabaceae tree (Wojciechowski et al. 2004). For the Solanaceae, alternative topologies were equally supported by the Bayesian analysis (Fig. 2).

Measurements of relative PD

Fig. 3 shows variation in relative PD among the three Andean areas, and how this compares with taxon richness. Patterns of distribution in PD vary in both families. In the Fabaceae, PD and richness showed a similar pattern, and significant differences in PD were observed among Andean areas ($F = 35264$, $P < 2.2e^{-16}$). However, for the Solanaceae, this was not the case. Even though PUN had the highest proportional taxon richness, when calculating PD, PAR was statistically similar to PUN because their confidence intervals overlapped. The SAS, which had higher taxon richness than the PAR, showed a statistically lower proportion of PD ($F = 402.23$, $P < 2.2e^{-16}$).

Discussion

In this study, we present an improvement to traditional methods of PD calculation, by increasing the number of trees used and generating variability that can be tested statistically. This is relevant given that measures of PD are based on branch lengths, and these are variable among the phylogenies obtained from an analysis. Some previous studies have explored the effect of variability on measures of PD at regional scales (Thuiller et al. 2011) or at the level of communities (Cadotte et al. 2008). An R script MultiPD (developed by Marc Cadotte, University of Toronto) also makes use of several trees to calculate PD. This is not,

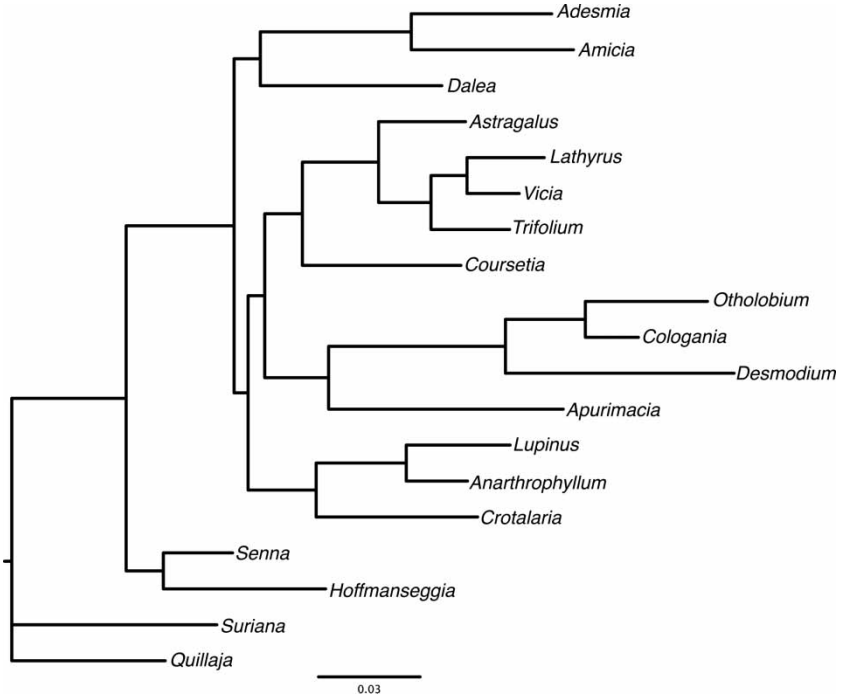

Figure 1 Phylogram showing the consensus phylogeny of the Fabaceae obtained by Bayesian analysis.

however, the general trend in most of the studies of PD.

A significant illustration of the importance of considering variability was seen in this study with the Solanaceae, in which equally likely phylogenies showed different topologies. Table 1 shows relative PDs for the three areas calculated for each one of the four topologies (A–D) shown in Fig. 2. In all of them, PUN showed the higher proportion, followed by PAR and SAS. However, the values obtained for each topology were different enough so that when several trees were considered, PUN and PAR were statistically equal in their contribution to PD, overlapping each other's PD distribution. This is important for fine discrimination given that we frequently observe more than one topology in a set of trees, as well as when phylogenies are built using different sets of genes. The effect of tree resolution on PD calculation has been explored previously (Swenson, 2009), noting that despite observing a larger effect with unresolved terminal nodes, the correlation between fully bifurcating and highly polytomous trees on PD is high. In fact, for our different topologies the differences in PD only became evident after analysing the confidence intervals.

When the results of PD calculations are used as indices to rank areas for conservation purposes, fine discrimination will often be needed. The use of variability allows us to statistically distinguish one area from another in terms of its PD and thus in terms of its true evolutionary conservation value.

PD versus richness

The combination of genera that colonise a biome and the shape of their diversification in terms of branches in a phylogenetic tree can be expected to impact PD differently than the main ecological factors that drive taxon richness. In relatively new and heterogeneous

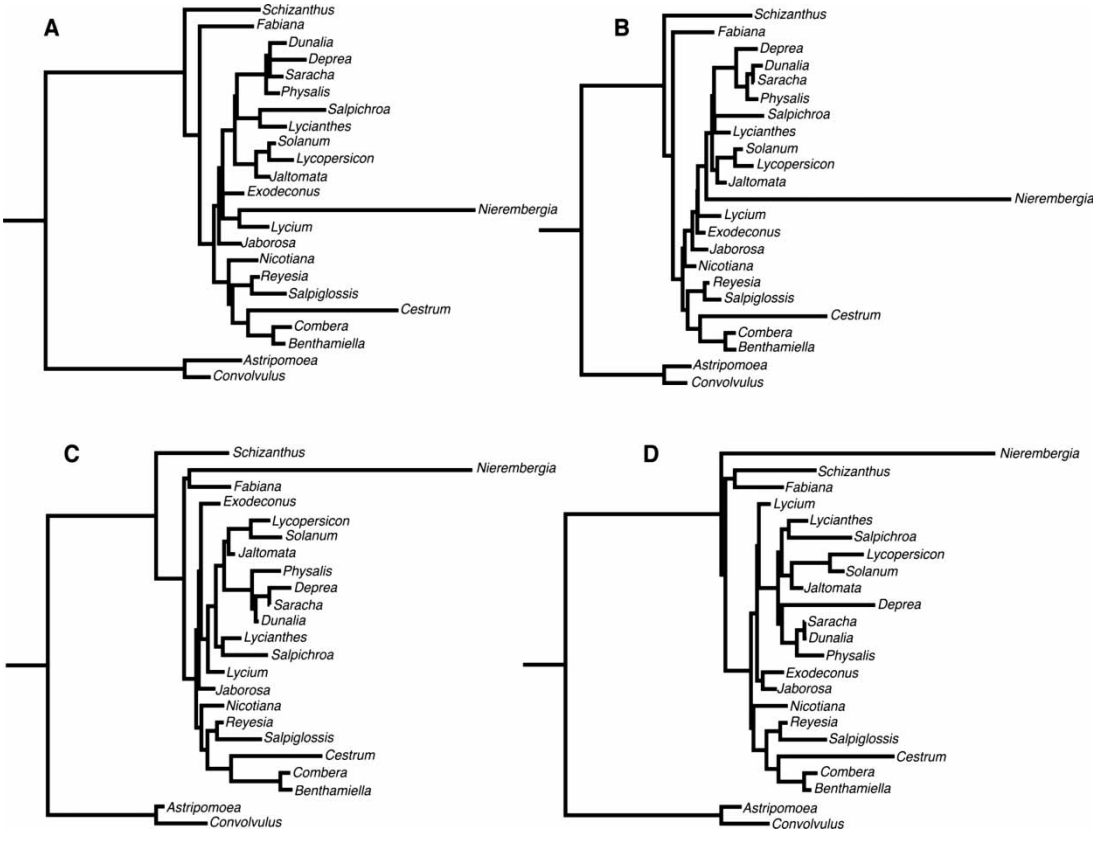

Figure 2 A–D, Four different equally likely topologies of the Solanaceae obtained by Bayesian analyses.

environments such as the high Andes, the constituent genera will be a function of those genera found in lower elevations, coupled with long distance dispersal, and the patterns of diversification within individual genera. When a particular genus tends to 'take over' in a flora, it may be expected that fewer additional taxa are able to colonise in. Thus, areas that support high taxon richness will not necessarily be the richest in terms of PD.

Studies comparing taxon richness and phylogenetic diversity have often concluded that one measure does not necessarily predict the other (Diniz-Filho et al. 2004; Forest et al. 2007; Pio et al. 2011). In our study, we saw that for the Fabaceae, the Andean sector that showed the highest generic richness (PUN) also had the highest PD. However, for the Solanaceae, the structure appears to be more complex. The use of PD suggests that genera of Solanaceae in the northern area might be on longer branches, and that generic diversity underestimates evolutionary diversity.

The use of variability in the calculation of PD made a difference in our study in this respect. In our calculations for the Solanaceae, for example, the proportion of PD contributed by the PUN is somewhat higher than for the PAR, which would coincide with the trend observed for genus richness. However, with incorporation of confidence intervals the contribution of each was not statistically different.

Clearly, conclusions regarding the relative amount of PD in the three Andean sectors must await more representative sampling of the flora. We are now in the process of studying other

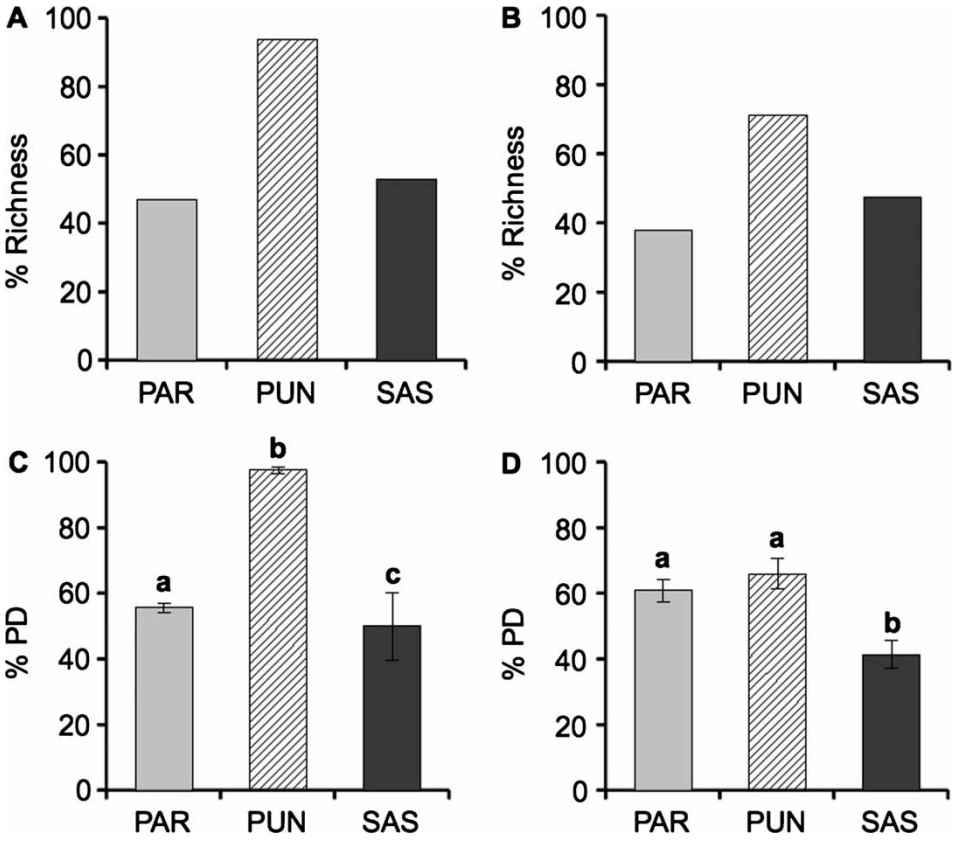

Figure 3 Relative generic and evolutionary richness (PD) of genera in the high Andes of South America for two angiosperm families, measured as a percentage of the total richness or phylogenetic diversity per area (PAR, PUN, SAS). Bars represent the standard deviation of the measurements obtained on a set of 35 Bayesian trees. Bars with different letters are statistically different ($P < 0.05$). **A,** Relative generic richness of the Fabaceae. **B,** Relative generic richness of the Solanaceae. **C,** Relative PD measured in branch lengths for the Fabaceae. **D,** Relative PD measured in branch lengths for the Solanaceae.

representative families and refining our taxonomic lists in order to obtain more precise information with respect to the inclusion of taxa in PAR, PUN and SAS. The present study was designed primarily to address methodological considerations in the calculation and interpretation of the PD index.

Table 1 Percentage of phylogenetic diversity (PD) per area calculated for each one of the four representative topologies obtained for the Solanaceae depicted in Fig. 2.

	% PD			
	Tree A	Tree B	Tree C	Tree D
PAR	48.90	47.60	50.18	51.79
PUN	53.35	53.57	53.71	58.56
SAS	32.44	33.18	33.07	35.50

Acknowledgements

We thank Martin F. Wojciechowski, for providing unpublished DNA sequences and helpful advice, and Michelle McMahon for providing herbarium material for *Apurimacia*. We are also grateful to Michael Sanderson and Michael Crisp for helpful advice during the study, and the reviewers for helpful

comments on the submitted manuscript. This study was funded by Fondecyt 3085004 grant to RS, ICM PO5-002 FICM, and PFB-23 Conicyt.

References

Allen B, Kon M, Bar-Yam Y 2009. A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. American Naturalist 174: 236–243.

Brako L, Zarruchi JL 1993. Catalogue of the flowering plants and gymnosperms of Peru. St Louis, MO, Missouri Botanical Garden. 1986 pp.

Cadotte MW, Cardinale BJ, Oakley TH 2008. Evolutionary history and the effect of biodiversity on plant productivity. Proceedings of the National Academy of Sciences of the United States of America 105: 17012–17017.

Diniz-Filho JAF, Rangel T, Hawkins BA 2004. A test of multiple hypotheses for the species richness gradient of South American owls. Oecologia 140: 633–638.

Faith DP 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61: 1–10.

Faith DP, Baker AM 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evolutionary Bioinformatics Online 2: 70–77.

Forest F, Grenyer R, Rouget M, Davies JT, Cowling RM, Faith DP, Balmford A, Manning JC, Proches S, van der Bank M et al. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. Nature 445: 757–760.

Luteyn JL 1999. Páramos: A checklist of plant diversity, geographical distribution and botanical literature. New York, New York Botanical Garden Press. 278 p.

McGoogan K, Kivell T, Hutchison M, Young H, Blanchard S, Keeth M, Lehman SM 2007. Phylogenetic diversity and the conservation biogeography of African primates. Journal of Biogeography 34: 1962–1974.

Pio VD, Broennimann O, Reeves G, Barraclough TG, Rebelo AG, Thuiller W, Guisan A, Salamin N 2011. Spatial predictions of phylogenetic diversity in conservation decision making. Conservation Biology 25 (6): 1229–1239.

Posada D, Buckley T 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53: 793–808.

Posada D, Crandall KA 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817–818.

Potter KM 2008. From genes to ecosystems: measuring evolutionary diversity and community structure with forest inventory and analysis (FIA) data. USDA Forest Service Proceedings 56: 49–64.

Purvis A, Hector A 2000. Getting the measure of biodiveristy. Nature 405: 212–219.

R-Team DCT 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Rambaut A 1996. Se-Al: Sequence Alignment Editor. Version 2.0a1. Oxford, UK, University of Oxford.

Rodrigues ASL, Gaston KJ 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. Biological Conservation 105: 103–111.

Ronquist F, Huelsenbeck JP 2003. Mr.Bayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92: 142–166.

Sokal R, Rohlf J 1981. Biometry. Second Edition. New York, WH Freeman. 859 p.

Swenson NG 2009. Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. PloS ONE 4: e4390. doi: 10.1371/journal.pone.0004390

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876–4882.

Thuiller W, Lavergne S, Roquet C, Boulangeat I, Lafourcade B, Araujo MB 2011. Consequences of climate change on the tree of life in Europe. Nature 470: 531–534.

Torres NM, Diniz-Filho JAF 2004. Phylogenetic autocorrelation and evolutionary diversity of Carnivora (Mammalia) in conservation units of the New World. Genetics and Molecular Biology 27: 511–516.

Ulloa C, Zarruchi JL, León B 2004. Diez años de adiciones a la flora del Perú 1993–2003. Arnaldoa Edición Especial Noviembre 2004: 1–242.

Vane-Wright RI, Humphries CJ, Williams PH 1991. What to protect – systematics and the agony of choice. Biological Conservation 55: 235–254.

Wojciechowski MF, Lavin M, Sanderson MJ 2004. A phylogeny of legumes (Leguminosae) based on analyses of the plastid *matK* gene resolves many well-supported subclades within the family. American Journal of Botany 91: 1846–1862.

Zuloaga FO, Morrone O, Belgran MJ 2008. Catálogo de las plantas vasculares del cono Sur (Argentina, Sur de Brasil, Paraguay y Uruguay). Volumen 2. Dicotyledoneae: Acanthaceae-Fabaceae (Abarema-Schizolobium). Monographs in Systematic Botany. St Louis, MO, Missouri Botanical Garden. Pp. 1–2286.

Appendix

Table 1 Species used in the study, with their GenBank accession numbers.

Species name	GenBank accession number
Fabaceae	
Adesmia volckmannii Phil.	AF142690
Amicia glandulosa H.B.&K.	AF203583
Anarthrophyllum desideratum Neger	AY386923
Astragalus lonchocarpus Torr.	AF142736
Apurimacia dolichocarpa (Griseb.) Burkart	FJ968527
Cologania hintoniorum B. L. Turner	AY583013, AY583014
Coursetia glandulosa A. Gray	AF543852
Crotalaria pumila Orteg.	AY386867
Dalea pulchra Gentry	AY386860
Desmodium psilocarpum A. Gray	AY386896
Hoffmannseggia glauca (Ortega) Eifer	EU361969
Lathyrus latifolius L.	AF522085
Lupinus argenteus Pursh.	AY386956
Otholobium bracteolatum (Eckl. & Zeyh.) C.H. Stirt.	EF550005
Senna candolleana (Vogel) Irwin & Barneby (syn. = *Cassia closiana* Phil.).	AY386848
Trifolium beckwithii Brewer ex S.Wats.	AY386946
Vicia ludoviciana Nutt.	AF5220158
Outgroups Fabaceae	
Suriana maritima L.	AY386950
Quillaja saponaria Molina.	AY386843
Solanaceae	
Benthamiella azorella (Skottsb.) A. Soriano	JF907505
Cestrum violaceum Urb.	DQ508649
Combera paradoxa Sandwith	EU580980
Deprea glabra (Standl.) Hunz.	AF212027
Dunalia solanacea Kunth	EU580988
Exodeconus flavus (I.M. Johnst.) Axelius & D'Arcy	JF923767
Fabiana imbricata Ruiz & Pav.	EU580992
Jaborosa integrifolia Lam.	DQ124558
Jaltomata procumbens (Cav.) J.L. Gentry	AY098695
Lycianthes inaequilatera (Rusby) Bitte	EU581018
Lycium vimineum Miers.	DQ124614
Lycopersicon esculentum Mill.	DQ180450
Nicotiana paniculata Goodsp.	AY098701
Nierembergia scoparia Sendtn.	AY772882
Physalis peruviana L.	EU581044

Table 1 (*Continued*)

Species name	GenBank accession number
Reyesia parviflora (Phil.) Hunz.	JF907506
Saracha punctata Ruiz & Pav.	EU581053
Salpichroa origanifolia (Lam.) Thell.	EU581052
Salpiglossis sinuata Ruiz & Pav.	AY206730
Schizanthus grahamii Gill. ex Hooker	EU581054
Solanum prinophyllum Dunal	DQ180407
Outgroups Solanaceae	
Astripomoea malvacea (Klotzsch) A. Meeuse	AY101074
Convolvulus arvensis L.	AY101102