Building a Robust, Densely-Sampled Spider Tree of Life for Ecosystem Research

Nuria Macías-Hernández [1,2,3,†,*], Marc Domènech [4,†], Pedro Cardoso [1], Brent C. Emerson [2], Paulo Alexandre Vieira Borges [5], Jesús Lozano-Fernandez [4,6], Octávio S. Paulo [7], Ana Vieira [7], Alba Enguídanos [4], François Rigal [5,8], Isabel R. Amorim [5] and Miquel A. Arnedo [4]

[1] Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History LUOMUS, University of Helsinki, 00014 Helsinki, Finland; pedro.cardoso@helsinki.fi

[2] Island Ecology and Evolution Research Group, IPNA-CSIC, La Laguna, Tenerife, 38206 Canary Islands, Spain; bemerson@ipna.csic.es

[3] Department of Animal Biology, Edaphology and Geology, University of Laguna, La Laguna, Tenerife, 38206 Canary Islands, Spain

[4] Department of Evolutionary Biology, Ecology and Environmental Sciences, & Biodiversity Research institute (IRBio), Universitat de Barcelona, 08028 Barcelona, Spain; mdomenan@gmail.com (M.D.); jesus.lozanof@gmail.com (J.L.-F.); albaengarcia@gmail.com (A.E.); marnedo@gmail.com (M.A.A.)

[5] Centre for Ecology, Evolution and Environmental Changes (cE3c)/Azorean Biodiversity Group and University of the Azores–Faculty of Agriculture and Environment, PT-9700-042 Angra do Heroísmo, Portugal; paulo.av.borges@uac.pt (P.A.V.B.); frantz.rigal@hotmail.fr (F.R.); isabelr@sapo.pt (I.R.A.)

[6] Institute of Evolutionary Biology (CSIC-UPF), 08003 Barcelona, Spain

[7] Centre for Ecology, Evolution and Environmental Changes (cE3c), Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal; ofpaulo@fc.ul.pt (O.S.P.); yanavieira1@gmail.com (A.V.)

[8] Institute of Analytical Sciences and Physico-Chemistry for Environment and Materials UMR5254, E2S UPPA, CNRS, Université de Pau et des Pays de l'Adour, 64000 Pau, France

* Correspondence: nemaciash@gmail.com

† These authors contributed equally to this work.

check for updates

Abstract: Phylogenetic relatedness is a key diversity measure for the analysis and understanding of how species and communities evolve across time and space. Understanding the nonrandom loss of species with respect to phylogeny is also essential for better-informed conservation decisions. However, several factors are known to influence phylogenetic reconstruction and, ultimately, phylogenetic diversity metrics. In this study, we empirically tested how some of these factors (topological constraint, taxon sampling, genetic markers and calibration) affect phylogenetic resolution and uncertainty. We built a densely sampled, species-level phylogenetic tree for spiders, combining Sanger sequencing of species from local communities of two biogeographical regions (Iberian Peninsula and Macaronesia) with a taxon-rich backbone matrix of Genbank sequences and a topological constraint derived from recent phylogenomic studies. The resulting tree constitutes the most complete spider phylogeny to date, both in terms of terminals and background information, and may serve as a standard reference for the analysis of phylogenetic diversity patterns at the community level. We then used this tree to investigate how partial data affect phylogenetic reconstruction, phylogenetic diversity estimates and their rankings, and, ultimately, the ecological processes inferred for each community. We found that the incorporation of a single slowly evolving marker (28S) to the DNA barcode sequences from local communities, had the highest impact on tree topology, closely followed by the use of a backbone matrix. The increase in missing data resulting from combining partial sequences from local communities only had a moderate impact on the resulting trees, similar to the difference observed when using topological constraints. Our study further revealed substantial differences in both the phylogenetic structure and diversity rankings of the analyzed communities estimated

from the different phylogenetic treatments, especially when using non-ultrametric trees (phylograms) instead of time-stamped trees (chronograms). Finally, we provide some recommendations on reconstructing phylogenetic trees to infer phylogenetic diversity within ecological studies.

1. Introduction

Understanding how communities assemble, and the evolutionary and ecological processes involved, is a key objective within the study of biodiversity [1,2]. The incorporation of phylogenetic approaches to the study of community ecology (phylogenetic community ecology) has advanced our understanding of how species pools evolve across time and space, and has highlighted the importance of integrating ecological and evolutionary processes to test mechanisms of community assembly [1–8]. Recent studies demonstrate the importance of quantifying phylogenetic relatedness between species to understand how evolutionary history and colonization dynamics influence diversity within communities [9,10]. Phylogenetic diversity (PD) also plays an important role in conservation [11], for better prioritizing phylogenetically diverse areas, as well as for preserving ecosystem function and associated ecosystem services [12].

The increase in phylogenetic information for many taxa has promoted the development of a plethora of PD metrics [6,11,13,14] to quantify different aspects of PD at the community level [7,8]. Selection of one metric over another may depend upon the nature of the research question, the hypothesis being tested, the available data (occurrence or abundance data) and the scale of the study [15,16]. One of the more commonly used PD metrics is the phylogenetic diversity index [17], a measure of richness defined as the total branch length of the minimum spanning tree linking all species represented in a given community.

Choice of phylogenetic reconstruction method, together with assumptions and decisions introduced during the tree-building process—taxon sampling, backbone constraints, genetic markers, model of evolution, among others—may have an impact on the estimation of phylogenetic diversity metrics. Factors influencing phylogenetic reconstruction and their implications for community phylogenetic analyses have been the subject of investigation, including taxonomic resolution (e.g., family level vs. species level) of the angiosperm tree flora of North America [18], proportion of missing data performed with empirical (plants DNA) and simulated data sets [19], the inclusion or not of nuclear genes in a study of North American desert bats [20] and the use of a tree specifically generated for a set of target taxa (several plant taxa from USA, France and Florida) vs. available published supertrees (Phylomatic and Open Tree of Life) [21], all revealing relatively low impacts on the estimation of phylogenetic diversity metrics. In contrast, resolution among basal nodes in a study of Mediterranean plants and simulated communities [22,23], and the use of well resolved topologies were proven fundamental for PD estimation [19,24]. The use of a chronogram (a time-stamped tree with branch lengths in units of evolutionary time) or a phylogram (trees with branch lengths in units of substitutions per site) has been shown to strongly influence PD values, in a recent study with vascular plants in Florida [25] with consequential effects for the inference of ecological patterns [26]. The use of model-based phylogenetic methods and a backbone phylogeny has been found to improve estimations of phylogenetic community metrics in a study of stonefly communities from Canada [27]. Finally, including multigene data (nuclear and mitochondrial genes) vs. solely the mitochondrial cytochrome c oxidase I gene (COI), barcode region [28] has been explored, concluding that phylogenies derived from the COI barcode region alone do not show significant differences in PD compared to multigene approaches [27].

A variety of methods to build phylogenies from which to calculate PD have been described. The simplest approach uses taxonomic hierarchy as a surrogate of PD, assigning increasing values

to distances among species, genera, families, etc. [29,30]. Pre-existing trees containing target species can also be used [31], or super-trees developed by pruning and grafting taxa from different available phylogenetic trees using tools such as those included in the software Phylomatic [32]. These approaches facilitate community phylogenetic studies, but are not without limitations and drawbacks [32]. A combination of supertrees together with DNA barcode data for a subset of taxa has also been explored, e.g., [33,34]. Some phylogenies have been developed by combining multiple, non-dated supertrees and setting all branch lengths to one [3]. Finally, phylogenetic trees have also been assembled either from genetic markers downloaded from GenBank [9] or from next-generation sequencing (NGS) data [12].

Spiders are among the most diverse and abundant invertebrate predators in terrestrial habitats worldwide, and they play a key role in ecosystem functioning as top-down regulators of above-ground food webs [35]. However, while many studies have addressed community patterns of taxonomic and functional diversity both at global [36] and regional scales [29,37,38], to date, only two studies have included phylogenetic information in the analysis of spider communities [39,40].

Here, we take advantage of the wealth of spider DNA sequence information available in public databases to investigate the effects of (i) taxon sampling, (ii) multi-gene data, (iii) topological constraints and (iv) time information on estimates of PD at the community level. We focus on communities of spiders from two distinct biogeographical regions: the Iberian Peninsula and Macaronesia (Azores, Madeira and Canary Islands). We first constructed species-level phylogenies of the mitochondrial cytochrome c oxidase subunit I (COI), which has been shown to be very effective at identifying spiders at the species level [41,42]. We used the COI both on its own and also in concatenation with sequences of the nuclear 28S rRNA (28S) for the 491 species collected in the local communities (local pool). We then inferred phylogenetic trees by adding approximately 1000 additional species representing most spider families, sampled for the same two gene regions, together with four additional genes [43]. This was performed both with and without topological constraints provided by the most complete phylotranscriptomic study on spiders published to date [44], which also provided the calibration points to infer time-stamped phylogenies. Our first objective was to explore how alpha phylogenetic diversity (PDalpha), its respective community rank and the ecological inferences for each plot site vary with the different phylogenetic trees. Secondly, we aimed to investigate the relative influence of several factors (i.e., adding nuclear information, backbone matrices, topological constraints, calibration and their combinations) on the inferred tree topology and branch lengths. Finally, we propose a workflow for obtaining a phylogeny and estimating PD metrics at the community level, explaining the possible troubleshooting to face during the generation of the final phylogeny. We also provide best practice recommendations for the reconstruction of phylogenetic trees to calculate PD metrics in community ecology studies.

2. Materials and Methods

2.1. Field Collection

Spiders were collected as part of different projects (see further details in the funding section) using the standardized sampling protocol COBRA (Conservation Oriented Biodiversity Rapid Assessment) [38]. In the Iberian Peninsula, spider communities were sampled from woodlands of different species of oak (namely Quercus petraea (Matts.) Liebl., Quercus faginea Lam., Quercus subpyrenaica Villar and Quercus pubescens Willd.) across the Spanish National Parks Network: four plots in Picos de Europa, two plots in Ordesa and Monte Perdido, two plots in Aigüestortes i Estany de Sant Maurici, two plots in Monfragüe, four plots in Cabañeros and two plots in Sierra Nevada [45]. In the Macaronesian archipelagos, spiders were collected from natural laurel forests distributed as follows: six plots in Pico and 10 plots in Terceira (Azores) [46], 12 plots in Madeira [47], 10 plots in Tenerife and seven plots in La Gomera (Canary Islands).

2.2. Species Identification and Community Composition

All spiders were morphologically identified at the species level (or genus level when not possible) using a stereomicroscope (see Crespo et al. [45] and Malumbres-Olarte et al. [46,47] for further details) and were stored at the collections of the CRBA (Animal Biodiversity Resource Center of the University of Barcelona, Barcelona, Spain); DZUL (Department of Zoology, University of La Laguna, San Cristóbal de La Laguna, Spain); IPNA-CSIC (Instituto de Productos Naturales-CSIC, San Cristóbal de La Laguna, Spain); MZB (Museum of Zoology, Barcelona, Spain); and UAc (University of Azores, Angra do Heroísmo, Portugal). Only species with adult specimens were included. Cryptic species identified by DNA as different lineages (e.g., see Crespo et al. [45] and Emerson et al. [48]) were assigned morphospecies numbers and were also included in the analyses as distinct entities. A community data matrix with species abundances collected for each plot (61 plots (rows) × 490 species (columns)) of the Iberian Peninsula [45] and Macaronesia [46,47] was constructed to calculate taxonomic diversity (TD) and PD.

2.3. Laboratory Protocols and Molecular Data

Sequencing of COI and 28S markers was attempted for all sampled species. For Macaronesian species, DNA extraction and amplification were carried out as explained in Emerson et al. [48]. For Iberian spiders, legs were used for DNA extraction and the rest of the individual was kept as a voucher, although for small species, the entire prosoma or opisthosoma was used. Total genomic DNA was extracted using the SpeedTool Tissue DNA Extraction Kit (Biotools), following the manufacturer's protocol. For COI, LCOI1490 and HCOI2198 [49] were the default primer combination, while Nancy [50] was used as a replacement for HCOI2198, and Ron [50] as replacement for LCOI1490 for amplifications that failed with the default primer pair. For 28S rRNA, the default primer combination was 28S-O and 28S-B [51,52], while 28S-A [53] and 28SrD1a [54] were used as replacements for 28S-O, and 28S-C [51] as a replacement for 28S-B for failed amplifications. All reactions were carried out with 5 μL of MyTaq RedMix from Bioline (London, UK), 0.2 mM of each primer, 2 μL of DNA template and adding ultrapure, distilled water up to a total reaction volume of 20 μL. PCR conditions for COI amplification were as follows: initial denaturation step at 95 °C for 5 min, 35 amplification cycles (94 °C for 30 s, 45 °C for 35 s, 72 °C for 45 s) and a final extension step at 72 °C for 5 min. PCR conditions for 28S amplification were as follows: initial denaturation step at 94 °C for 5 min, 35 amplification cycles (94 °C for 30 s, 48–52 °C for 35 s, 72 °C for 1 min) and a final extension step at 72 °C for 10 min. PCR products were cycle-sequenced in both directions at Macrogen Inc. (Amsterdam, Netherlands). If the amplification of one of the genes was unsuccessful, a sequence for the same fragment of the same species was downloaded from GenBank when possible (see supplementary Table S1 for details in taxon sampling, including GenBank accession numbers). All sequence chromatograms were assembled, visualized, error-checked, edited and aligned in Geneious 11.1.5 [55]. Assembled contigs were checked against the online NCBI BLAST or BOLD (for COI barcode region) databases to account for possible contaminations.

2.4. Phylogenetic Analyses

2.4.1. Ibercoding and NetBiome-MacDiv Matrices

One sequence per taxon was used to assemble the mitochondrial COI barcode region and the nuclear 28S matrices of the species collected in the local community surveys. The Ibercoding project (from the Iberian Peninsula, see further details in the funding section) contributed 372 taxa for the COI (658 bp) and 358 taxa for the 28S matrices. The NetBiome and the MacDiv projects (both from Macaronesia, see further details in the funding section) contributed 117 taxa for the COI (658 bp) and 115 taxa for the 28S matrices. The COI sequences of the three projects (local communities) were combined in a single matrix, hereafter referred to as Lc1 (489 taxa, 658 bp), to assess the effect of analyzing only the COI partition on tree topology and phylogenetic diversity indices of local communities. In addition,

we assembled a second matrix to also include the 28S sequences, hereafter referred to as Lc128 (492 taxa, 1958 bp) to explore the impact of this additional nuclear locus (Table 1).

Table 1. Attributes of the matrices assembled in this study with information on their taxa, number of genes and missing data. Matrices names are given according to the data included, as follows: (B) taxa from backbone matrix [43,44] and (L) taxa from local communities (Iberian Peninsula and Macaronesia) (= Taxa); genes added for local species, (c1): mitochondrial COI barcode gene, (28): nuclear 28S rRNA gene (= Genes); (tc): topological constraint (= Topology); (cal): time-calibrated trees (= Time). Terminals: number of terminals; % missing: percentage of missing cells in the matrix; COI, 28S, 12S, 16S, 18S, H3: numbers of positions (aligned) for the respective genes.

Matrix Name	Taxa	Genes	Topology	Time	Terminals	%missing	COI	28S	12S	16S	18S	H3
B	Backbone matrix	NA	Unconst.	no	964	58.88	1079	4369	769	1031	2608	318
B_tc	Backbone matrix	NA	Const.	no	964	58.88	1079	4369	769	1031	2608	318
BLc1_tc	Backbone matrix + local species	c1	Const.	no	1453	70.62	1079	4369	769	1031	2608	318
BLc1_tc_cal	Backbone matrix + local species	c1	Const.	yes	1453	70.62	1079	4369	769	1031	2608	318
BLc1	Backbone matrix + local species	c1	Unconst.	no	1453	70.62	1079	4369	769	1031	2608	318
BLc1_cal	Backbone matrix + local species	c1	Unconst.	yes	1453	70.62	1079	4369	769	1031	2608	318
BLc128_tc	Backbone matrix + local species	c1 + 28s	Const.	no	1456	68.36	1079	4364	769	1031	2608	318
BLc128_tc_cal	Backbone matrix + local species	c1 + 28s	Const.	yes	1456	68.36	1079	4364	769	1031	2608	318
BLc128	Backbone matrix + local species	c1 + 28s	Unconst.	no	1456	68.36	1079	4364	769	1031	2608	318
BLc128_cal	Backbone matrix + local species	c1 + 28s	Unconst.	yes	1456	68.36	1079	4364	769	1031	2608	318
Lc1	Local species	c1	Unconst.	no	489	3.61	658					
Lc128	Local species	c1 + 28s	Unconst.	no	492	32.23	658	1300				

Two recent analyses of the spider tree of life, one including the most thorough taxon sample to date (932 species in 115 families) for 6 Sanger sequenced genes [43], and one with the largest transcriptome representation (ca. 2500 genes), were used to assess the effects of (i) topological constraints, and, by using a backbone matrix, (ii) number of markers and (iii) taxonomic sampling on estimates of PD for a more modest taxa representation (159 species) [44].

2.4.2. Topological Constraint

The phylogenomic study of Fernández et al. [44] was used to construct a topologically constrained tree, hereafter referred to as the F_topological_constraint (Table 1). The information of the preferred tree of Fernández et al. [44] was slightly modified, such that some misspelled taxon names were corrected according to the World Spider Catalog [56] or modified to match our taxon labelling (see details in supplementary Table S2). Additionally, outgroup taxa and terminals that were not present in our final matrix were pruned, and nodes with bootstrap support < 95 were collapsed. Tree manipulation was made with the software TreeGraph2 [57] and R v3.6.1 [58].

2.4.3. Backbone Matrix

The concatenated gene matrix of Wheeler et al. [43], which included three mitochondrial (12S rRNA, 16S rRNA and COI) and three nuclear (histone H3, 18S rRNA and 28S rRNA) genes, was used as a backbone matrix for our analyses (see Supplementary Table S3 for details of taxa included). Taxa from the phylotranscriptomic tree of Fernández et al. [44] not present in the Wheeler et al. [43] matrix were included to maximize the number of constrained nodes. Sequences from the missing taxa or, if unavailable, from close relatives within the same genus (see Figure 1 for further details) were either (1) downloaded from GenBank, or (2) bioinformatically extracted from the original transcriptome data in Fernández et al. [44] (see Supplementary Table S2 for GenBank transcriptomes accession numbers). For retrieving the target genes from available spider transcriptomes from Fernández et al. [44], we downloaded the raw data from NCBI (see Supplementary Table S2). Transcriptome assembly was carried out using Trinity version 2.0.3 [59,60] under default parameters and employing Trimmomatic (default parameters, as part of the Trinity package) for quality control. Relevant molecular markers were retrieved from the resulting assembled transcriptome using BLASTn [61] with an *E*-value cut-off of less than 10^{-20} to account for potential orthologs. A total of 32 transcriptomes were analyzed. Extracted sequences were checked by comparing against the online NCBI BLAST databases and by building single-gene phylogenies to screen for possible contaminants. The topological position of each newly added taxon was first checked, and in the case of the obvious misplacing of the specific gene partition, or the whole taxon if all partitions were involved, was deleted.
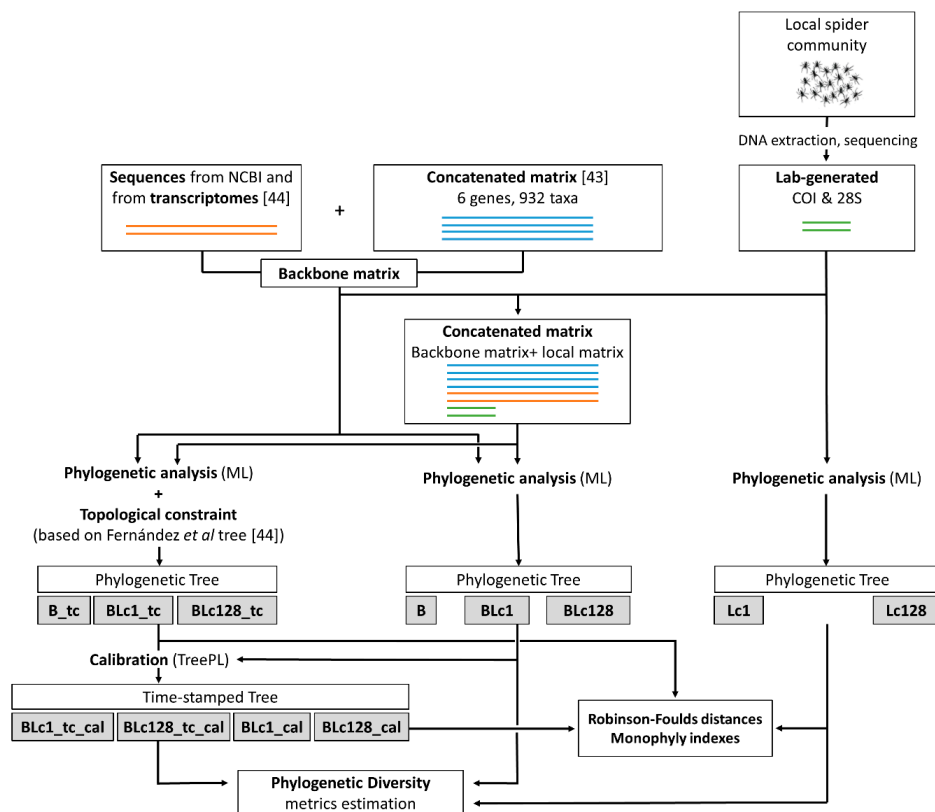

Figure 1. Workflow followed for inferring phylogenetic trees and estimating community phylogenetic metrics.

A total of 167 new sequences of the six markers (12S, 16S, COI, H3, 18S and 28S) (126 sequences extracted from transcriptomic data and 41 sequences downloaded from GenBank) were added to the Wheeler et al. [43] matrix after the removal of non-spider outgroups. The final backbone matrix, hereafter referred to as the B matrix, contained 964 terminals (Table 1, see Supplementary Table S2 for the new sequences included). With the inclusion of the species *Cepheia longiseta*, a representative

of the family Synaphridae, the B matrix included 118 of the 120 (98.3%) spider families currently recognized [56]. The Lc1 matrix (containing COI sequences of the Ibercoding and NetBiome-MacDiv projects) was added to the B matrix to generate a new matrix hereafter referred to as BLc1 (Table 1). Similarly, the Lc128 matrix (containing both COI and 28S sequences of the former projects) was added to the B matrix, hereafter referred to as the BLc128 matrix (Table 1). Therefore, the influence of taxon sampling (backbone matrix) was tested by running the analyses including the B matrix (all matrices but Lc1 and Lc128), or without including the B matrix (only the species collected from our local communities; matrices Lc1 and Lc128).

Ribosomal gene sequences were automatically aligned with the online version of MAFFT v. 7 [62] using the algorithm FFT-NS-I with default settings. Alignments of the protein-coding COI and H3 were trivial since no indel mutations were observed.

2.4.4. Maximum Likelihood Analyses

Phylogenetic trees, for the different matrices constructed, were inferred using Maximum Likelihood (ML) as implemented in the program RAxML-HPC v. 8.2.12 [63], remotely run on XSEDE at the CIPRES Science Gateway [64]. The concatenated matrices were partitioned by gene, and an unlinked GTR + CAT substitution model was assigned to each gene. The ML tree was inferred out of 10 iterations of randomized stepwise addition Maximum Parsimony starting trees (-N 10). The effect of incorporating a topological constraint in the phylogenetic reconstruction and on PD estimations was assessed by conducting maximum likelihood analyses on each concatenated data matrix with and without topological constraints (-g), the latter was defined by the F_topological_constraint tree (Table 1). Trees were rooted at the node splitting representatives of the suborder Mesothelae (family Liphistiidae) from the reminder species [43,44]. The matrices including species only from the local communities, where no Liphistiidae were present, were rooted at the node splitting the single Mygalomorphae representative (*Nemesia*) from the remaining species, assuming the sister group relationship between Mygalomorphae and Araneomorphae [43,44].

2.4.5. Inferring Ultrametric Trees

All resulting ML trees (except Lc1 and Lc128) were made ultrametric using the congruify.phylo function in the R package GEIGER [65]. The dated phylogeny generated by Fernández et al. [44] was used as a reference to provide fixed calibration points for the target trees (i.e., each of the ML trees inferred for the different data matrices with and without topological constraints). The resulting table of calibration points was used to date the target trees using the program treePL [66]. Recently, Magalhaes et al. (2020) [67] have published a time stamped tree of life of spiders based on a reexamination of the available fossil data. Although the location of some of the calibration points differed from those used by Fernández et al. 2018 [44], the inferred confidence intervals broadly overlapped. A first random cross validation analysis was run to determine the preferred smoothing parameter values for each tree. Figure 1 summarizes the workflow followed for inferring phylogenetic trees and estimating community phylogenetic metrics.

2.4.6. Tree Similarity and Monophyly Levels

The level of similarity among trees inferred from the assembled matrices was assessed with the R package phangorn [68]. Pairwise tree similarities were estimated using Robinson–Foulds (RF) distances [69], both absolute and normalized to the maximum possible distance. Low values of RF indicate high similarity between trees. For these analyses, we used all non-ultrametric trees that were pruned to ensure that both trees shared the same labels. The number of monophyletic families and genera recovered in each tree was explored with the R package MonoPhy [70]. Levels of monophyly between trees were compared by calculating the ratio of monophyletic families/genera taking into account either the total taxa of the corresponding taxonomic level, or excluding the monotypic taxa.

2.5. Phylogenetic Community Metrics and Statistical Analyses

Phylogenetic diversity for each community was calculated as the Faith's phylogenetic diversity (PD) [14,17]. This index measures phylogenetic community richness as the total branch length of the minimum spanning tree linking all species represented in a community. It was calculated with the R package BAT [71] for each sampling plot, using each phylogenetic tree.

All statistical analyses were conducted under the assumption that the time-stamped tree obtained with the full data set including topological constraints (BLc128_tc_cal) conveys the most accurate representation of the underlying evolutionary relationships of the sampled taxa, as it maximizes its explanatory power by including all available sources of evidence [72]. This tree is hereafter referred to as the "reference tree" (See Figure 2 and supplementary Figure S1). We further explored how PD varied across the trees obtained under the different conditions explored, and if the different inferred trees affected the rank of alpha diversity comparisons of the communities.
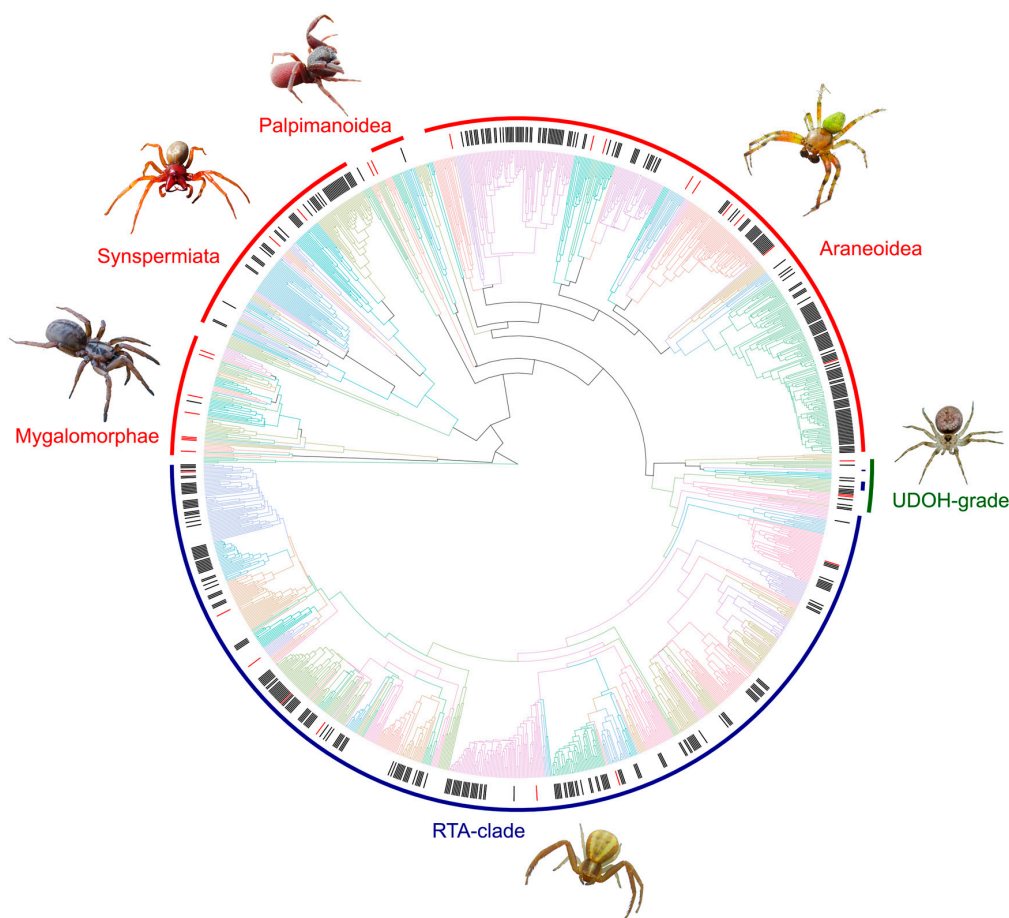

Figure 2. General representation of the final reference tree BLc128_tc_cal (time-calibrated tree built with a backbone and a topological constraint using COI and 28S information). Branch colors represent different families. Outer lines delimit the main spider clades.

To compare how the estimates of PD varied between the different trees used, the proportional difference between PD values in each plot was measured. The values of PD per plot obtained with each tree were compared to the reference tree—BLc128_tc_cal for comparing dated trees and BLc128_tc for comparing non-ultrametric trees—as follows: (PDa-PDr)/PDr*100; where "r" denotes the reference tree (BLc128_tc_cal or BLc128_tc) and "a" the alternative tree [32]. A value of zero indicates that no differences exist between the PD values obtained with the two trees tested. In order to determine if the rank of the PD of the communities of the reference tree (BLc128_tc_cal) was correlated with

the ranks obtained with the other inferred trees, a linear regression of the rank of the reference tree against the remaining trees was applied using the "lme4" library in R [73]. A high R2 would indicate that the rank of the communities with the two phylogenetic trees tested were similar. Additionally, a series of generalized linear mixed-effect models (GLMM) were constructed to determine how the conditions under which the different trees were inferred (i.e., adding nuclear information, backbone matrices, topological constraints, calibration and their combinations) affected the rank of the PD of the communities. GLMMs were constructed using the lme function in the nlme library in R [74]. The general form of the GLMM for each combination of trees constructed is as follows:

<div align="center">Rank community using tree A~Rank community using tree B + (1|Region)</div>

where trees A and B are trees inferred under the different conditions tested (i.e., backbone matrices vs. not backbone, topological constraint vs. unconstrained, time-stamped vs. non-ultrametric, 28S and COI barcode region vs. COI barcode region-only) (see details in Table 1). For example, testing the influence of using topological constraints was accomplished by defining Rank of communities obtained with tree BLc128_tc_cal~Rank of communities obtained with tree BLc128_cal + (1|Region). The fixed factor was one of the trees inferred under specific conditions. The region pool (1: Iberian Peninsula; 0: Macaronesia) was included as a random effect to remove the influence of the geographic origin of the plot (see Table 2 for details of GLMM analyses). The proportion of variance explained by the fixed and the random factor together (conditional R2) [75,76] was calculated using the function r.squaredGLMM in the library MuMIn in R [75]. The variable importance for each GLMM accomplished was obtained as 1-R2. For each of the factors tested, the mean of the variable importance was calculated to determine which factor had a larger effect on the rank of the PD (see Table 2). The true skill statistic (TSS) test [77] was applied to check the accuracy of each phylogenetic tree in inferring the same ecological processes (PD overdispersion or PD clustering) as the reference tree. TSS accounts for the accuracy in detecting false negatives and positives, in opposition to true positives and negatives, and it ranges from −1 to + 1, where + 1 indicates perfect agreement, and values of zero or less indicate a performance no better than random. All statistical analyses were performed with the software R [58] and the TSS test conducted by applying the formula from [77] in Excel.

Table 2. Details on the generalized linear mixed-effect models (GLMM) analyses performed. Factor tested: Topology (constrained /unconstrained); Time (time stamped trees; yes/no); Genes (c1/c1 + 28s); Taxa (set: local species/all: backbone matrix + local species). varimp: Variable importance ($1-R^2$). Mean varimp: mean of variance importance for each factor tested. The factor tested in each set of analyses is in bold.

GLMM	Topology	Time	Genes	Taxa	R^2	Varimp	Mean Varimp
BLc1_tc_cal~BLc1_cal	**const vs. unconst**	yes	c1	all	0.889	0.111	
BLc1_tc~BLc1	**const vs. unconst**	no	c1	all	0.889	0.111	0.113
BLc128_tc_cal~BLc128_cal	**const vs. unconst**	yes	c1 + 28s	all	0.886	0.114	
BLc128_tc~BLc128	**const vs. unconst**	no	c1 + 28 s	all	0.886	0.114	
BLc1_tc_cal~BLc1_tc	constrained	**yes vs. no**	c1	all	0.909	0.091	
BLc1_cal~BLc1	unconstrained	**yes vs. no**	c1	all	0.901	0.099	0.093
BLc128_tc_cal~BLc128_tc	constrained	**yes vs. no**	c1 + 28 s	all	0.910	0.090	
BLc128_cal~BLc128	unconstrained	**yes vs. no**	c1 + 28 s	all	0.907	0.093	
BLc128_tc_cal~BLc1_tc_cal	constrained	yes	**c1 vs. c1 + 28 s**	all	0.891	0.109	
BLc128_cal~BLc1_cal	unconstrained	yes	**c1 vs. c1 + 28 s**	all	0.891	0.109	0.102
BLc128_tc~BLc1_tc	constrained	no	**c1 vs. c1 + 28 s**	all	0.913	0.087	
BLc128~BLc1_cal	unconstrained	no	**c1 vs. c1 + 28 s**	all	0.896	0.104	
BLc1~Lc1	unconstrained	no	c1	**all vs. set**	0.898	0.102	
BLc128~Lc128	unconstrained	no	c1 + 28s	**all vs. set**	0.896	0.104	0.103
Lc128~Lc1	unconstrained	no	**c1 vs. c1 + 28 s**	set	0.915	0.085	0.085

2.6. Null Models and Ecological Patterns

Null models were applied to test if species within communities were phylogenetically more distant (phylogenetic overdispersion) or close (phylogenetic clustering) to each other than expected from a random draw of species from the pool, while keeping species richness constant in each local community [15,78]. The motivation of null models is to quantify the effect of the main mechanism being tested (in this case, species phylogenetic relationship) by excluding it from the analyses. Higher values of PD than expected by chance indicate phylogenetic overdispersion, while lower values indicate clustering. The motivation of null models is to quantify the effect of the main mechanism being tested (in this case, species phylogenetic relationship) by excluding it from the analyses. An analysis was run with 999 randomizations to generate a distribution of PD under null expectations. For each community and tree, the Standard Effect Size (SES) was calculated as the difference between the observed PD and the mean PD of communities from null models divided by its standard deviation.

$$SES = (PDobs- mean\ PDnull) / sd\ PDnull$$

We further computed a two-tailed test, with SES values greater than 1.96 or less than −1.96considered to be significantly higher or lower than expected, respectively. Therefore, when values of observed PD are significantly higher than expected PD (PDnull), this indicates phylogenetic overdispersion, and an observed PD significantly lower than expected indicates phylogenetic clustering. Phylogenetic trees were evaluated for inference consistency (PD overdispersion vs. PD clustering) within each plot.

3. Results

3.1. COI and 28S Sequences

The COI and 28S amplicons of 375 and 360 species, respectively, from the Iberian Peninsula, plus the COI and 28S amplicons of 117 and 115 species, respectively, from Macaronesia, were successfully amplified and sequenced, thus accounting for ca. 25% of spider species from the two regions (492 out of 2180 species in total) (see supplementary Table S1).

3.2. Phylogenetic Trees

Differences between the trees inferred under the distinct treatments and their respective monophyly indexes are summarized in Tables 3 and 4. The new taxa added to increase the number of constrained nodes (B) resulted in a tree with some differences compared to the legacy tree reported by Fernández et al. [44], which had been obtained by analyzing the Wheeler et al. [43] matrix constraining well-supported nodes of their preferred transcriptomic tree. Most of the basal relationships at the family level were identical, but some families were not recovered as monophyletic. For example, in the tree without topological constraint, the nemesids *Damarchus* sp. and *Pionothele* sp. and the theraphosid *Trichopelma* sp. did not cluster together with other taxa of their respective families, whereas in the constrained trees only *Damarchus* sp. did not. Surprisingly, although the topologically constrained tree (B_tc) had a higher number of monophyletic families (e.g., 8 families were recovered as monophyletic only in the constrained tree), the families Psechridae and Thomisidae were monophyletic only in the unconstrained analysis (B).

Adding the 28S sequences of the local communities (Lc128, BLc128) resulted in trees with the highest normalized RF distances among trees, when compared to trees without those sequences (Lc1, BLc1) (see Table 3 and Figure 1 for further details). Similarly, high distances were recovered when the COI barcode matrix (Lc1) was analyzed in combination with the backbone matrix (BLc1), but distances were much lower when the 28S sequences (Lc128) were included. Lower distances were observed between trees with and without topological constraints. Adding species from the local communities, either COI barcode only (Lc1) or COI + 28S (Lc128) to the backbone matrix (BLc1, BLc128)

resulted in similar trees, although distances were higher when topological constraints were enforced (Table 3).

Table 3. Robinson–Foulds distances (normalized RF) among unrooted trees inferred for each matrix/treatment. NA: Not Applicable.

	B_tc	BLc1_tc	BLc1	BLc128_tc	BLc128	Lc1	Lc128
B	400 (0.208)	402 (0.209)	204 (0.106)	552 (0.287)	462 (0.222)	NA	NA
B_tc		364 (0.189)	404 (0.211)	414 (0.215)	516 (0.268)	NA	NA
BLc1_tc			782 (0.27)	1240 (0.427)	1262 (0.435)	454 (0.467)	516 (0.531)
BLc1				1266 (0.436)	1158 (0.399)	434 (0.446)	518 (0.533)
BLc128_tc					544 (0.187)	508 (0.522)	284 (0.290)
BLc128						526 (0.541)	266 (0.272)
Lc1							508 (0.522)

Table 4. Monophyly indexes at family and genus levels for the trees inferred from the matrices/treatments analyzed. See Table 1 for matrix names and information. Indexes and number of taxa are shown for both complete trees and trees with either local or non-local taxa removed (notice the number of species). Number of taxa: number of terminals in each tree; monophyletic/total: ratio of monophyletic families/genera per total number of families/genera; monophyletic/total (excluding monotypic): same as before but excluding monotypic families/genera.

Number of Taxa	B	B_tc	BLc1_tc			BLc1			BLc128_tc			BLc128			Lc1	Lc128
Species	964	964	1453	964	489	1453	964	489	1456	964	492	1456	964	492	489	492
Families	117	117	117	117	39	117	117	39	117	117	39	117	117	39	39	39
Genera	761	761	884	761	215	884	761	215	884	761	215	884	761	215	215	215
Monophyletic/Total																
Families	0.58	0.62	0.49	0.59	0.54	0.47	0.57	0.46	0.59	0.62	0.67	0.58	0.61	0.64	0.41	0.54
Genera	0.16	0.15	0.20	0.15	0.34	0.20	0.16	0.33	0.21	0.16	0.35	0.21	0.15	0.34	0.34	0.35
Monophyletic/Total (excluding monotypic)																
Families	0.67	0.71	0.56	0.68	0.70	0.54	0.66	0.60	0.68	0.71	0.87	0.67	0.70	0.83	0.53	0.70
Genera	0.79	0.78	0.67	0.79	0.73	0.66	0.80	0.70	0.72	0.79	0.74	0.71	0.79	0.73	0.72	0.74

Adding 28S information of the species from local communities had the largest impact on the monophyly indexes of the resulting trees, which showed an increase in the number of monophyletic groups (e.g., 0.67 versus 0.54 and 0.71 versus 0.66—monotypic excluded—between BLc128 and BLc1, for families and genera, respectively) (Table 4). This effect was more pronounced in families than in genera, and this observation holds true across all comparisons (Table 4). Adding backbone matrices to the species from local communities also had a beneficial effect in terms of increasing the number of monophyletic groups, although less pronounced that in the case of adding 28S information (Table 4). The use of topological constraints, on the other hand, had the least impact in terms of increasing the number of monophyletic groups (Table 4). The increase in the percentage of missing data as a result of adding the sequences from local taxa to the backbone matrix had little, if any, effect on the number of monophyletic groups recovered (Table 4).

3.3. Phylogenetic Community Metrics

Overall, plots from the Iberian Peninsula showed a higher species richness (alpha taxonomic diversity, TD) and alpha phylogenetic diversity (PD) than Macaronesian plots (see Supplementary Table S4, statistical significance not assessed). The rank of the PD calculated from the reference tree (BLc128_tc_cal) indicated that plots Ordesa_2 (*Quercus subpyrenaica* forest), Picos_4 (*Q. faginea*) and Cabañeros_4 (*Q. faginea*) showed the highest PD values, whereas Picos_1 (*Q. petraea*), Monfragüe_2

(*Q. faginea*) and Picos_3 (*Q. petraea*) showed the lowest PD values within the Iberian Peninsula. Within the Macaronesian archipelagos, plots from the Canary Islands (La Gomera and Tenerife) and Madeira yielded the highest PD values, while plots from the Azores (Pico and Terceira) ranked the lowest.

Both the estimated PD within plots and their ranking varied depending on the tree used (see Supplementary Tables S4 and S5). With the time-stamped trees and averaging over all communities, the percentages of difference in estimated PD between the reference and the alternative trees were higher (range from 6.5% to 14%) than in the non-ultrametric trees (range from 2% to 10.2%). Among the time-stamped trees, the unconstrained tree including only the COI barcode data of the local communities (BLc1_cal) performed the worst (i.e., yielded the most different PD values from the reference tree). Among the non-ultrametric trees, the tree built from COI-only data from local taxa (Lc1) was the worst (see supplementary Table S4). The PD values were on average 10% higher when using phylogenetic trees inferred from species from the local communities alone than when incorporating the backbone matrix. Similarly, unconstrained trees resulted in higher differences in PD values (7%).

Most importantly, none of the PD indices estimated from the alternative phylogenetic trees recovered the same ranking as the reference tree (BLc128_tc_cal) (see supplementary Table S5). The correlation (R^2) found between the PD ranking of communities based on the reference tree (BLc128_tc_cal) and the remaining trees varied between 0.78 and 0.49 (see Table 5). The highest correlation ($R^2 = 0.78$) was obtained using the same conditions as the reference tree, but without topological constraints (BLc128_cal). The correlation of the PD community rank obtained from the non-ultrametric trees showed the lowest correlation ($R^2 < 0.57$). When comparing non-ultrametric trees, the lowest values of R^2 (< 0.47) were obtained with the trees built using only COI barcode data (see Table 5).

Table 5. Results of linear regression of the phylogenetic diversity (PD) ranking of communities of the reference tree (BLc128_tc_cal) against the ranks obtained with the remaining trees. When comparing non-ultrametric trees alone, the reference tree is BLc128_tc.

BLc128_tc_cal against	R^2	BLc128_tc Against	R^2
BLc128_cal	0.786	BLc128	0.810
BLc1_tc_cal	0.683	Lc128	0.519
BLc1_cal	0.660	BLc1_tc	0.475
Lc1	0.575	BLc1	0.459
BLc128	0.545	Lc1	0.422
Lc128	0.521		
BLc1	0.503		
BLc128_tc	0.502		
BLc1_tc	0.493		

Fifteen generalized linear mixed-effect models (GLMM) were constructed for testing the effect of different factors on the rank of community plots (view Table 2 for further details). All factors tested yielded similar values of variable importance, i.e., using or not using topological constraint (0.112); time-stamped or non-ultrametric trees (0.093); adding or not adding 28S data (0.10); and with or without a backbone matrix (0.10) (see Table 2), indicating that all factors had a similar effect in failing to recover the same ranking as the reference tree.

3.4. Null Models and Ecological Patterns

Phylogenetic diversity patterns (PD overdispersion or PD clustering) inferred for each community plot varied depending on the phylogenetic tree used (see supplementary Table S6). The values of TSS indicated that the accuracy of each phylogenetic tree to infer the same ecological processes as the reference tree decreased when using non-ultrametric trees (TSS < 0.61) and no backbone matrix (TSS < 0.42). The Lc1 tree (COI barcode-only tree built with local taxa) was the least accurate (TSS < 0.38), while time-stamped trees were the most accurate ones (TSS > 0.86) (see Table 6). In general, phylogenetic trees built using only COI barcode data (both time-stamped and non-ultrametric) tended to detect more false positives (significant PD clustering) compared to the reference tree (see supplementary Table S6). The main phylogenetic pattern inferred for most plots was PD clustering (in 17 out of 61 plots if we consider all trees), while PD overdispersion was inferred in five plots. There was congruence in the ecological patterns inferred for each plot regardless of the phylogenetic tree used. No plot shifted from significant PD clustering to significant overdispersion, or vice versa, when using alternative trees. Furthermore, for few plots, the same phylogenetic structure was inferred regardless of the tree used (e.g., significant clustering and overdispersion was always recovered for Ordesa_1 and Acebiños, respectively).

Table 6. Results of the true skill statistic (TSS) test when comparing each tree with the reference tree (BLc128_tc_cal). See Table 1 for further details in matrix codes.

Phylogenetic Tree	TSS
BLc128_cal	1.000
BLc1_tc_cal	0.918
BLc1_cal	0.870
BLc1_tc	0.603
BLc128_tc	0.536
BLc128	0.519
BLc1	0.511
Lc128	0.417
Lc1	0.374

4. Discussion

We have inferred a densely sampled, multi-locus, constrained species-level phylogeny of spiders to empirically test the influence of different factors (topological constraints, backbone matrices, genetic markers and time calibration) on phylogenetic reconstruction and how the alternative trees inferred affect phylogenetic diversity estimates. We included 375 COI barcode sequences from plots in the Iberian Peninsula [45] and 117 COI barcode sequences from plots in the Macaronesian archipelagos [46,47], as well as 475 new 28S sequences from both regions. The wealth of genetic data on spiders available through recently published phylogenetic studies allowed us to incorporate both a backbone matrix, composed of a concatenated matrix of six genes from a rich sample of spider species [43], and topological constraints, derived from a phylotranscriptomic analysis of more than 2500 genes [44]. Node calibration information was also provided from previous phylogenetic work to estimate divergence times [44]. The results of the simultaneous analyses of the sequences sampled from the local communities, combined with the backbone matrix and the topological constraint (Figure 2 and supplementary Figure S1), constitute the most complete phylogeny of spiders in terms of terminals and background information to date. The newly generated phylogenetic tree can serve as a standard reference for ongoing and future studies focused on deciphering phylogenetic diversity patterns at the community level.

4.1. Tree Topology

Tree topologies were influenced by all evaluated factors, namely the addition of a backbone matrix, the enforcement of topological constraints and the inclusion of a slow-evolving nuclear gene (28S) to the mitochondrial COI barcode sequences of local communities, but some factors had a stronger effect than others. The reference tree (BLc128_tc_cal) outperformed the trees based either on only the set of species from local communities or on the combination of the backbone matrix and the COI barcode sequences of local communities, in terms of monophyly recovery. Increasing taxon sampling improves phylogenetic accuracy and reduces phylogenetic error [79,80]. However, the combination of sequences of species from local communities with backbone matrices that include a larger sampling of genes also increases the proportion of missing data, particularly when only COI barcode sequences from local taxa are available. However, our results revealed that, although in a few cases the value of the monophyly indexes slightly decreased when increasing the proportion of missing data, it was circumscribed to a family level comparison and when adding only COI barcode data.

As expected, phylogenetic trees reconstructed using topological constraints provided better monophyly recovery than unconstrained trees, thus influencing PD estimations. The use of topological constraints derived from highly supported nodes from genome scale analyses is a simple and efficient way of incorporating robust information within matrices with limited DNA sequence data. It has additionally been shown that the incorporation of topological constraints can provide more accurate estimates of branch lengths [25,27,32] and potentially improve the estimation of phylogenetic community metrics [23,81]. The monophyly of many families was better recovered, both in constrained and unconstrained phylogenetic trees, when species sampled from local communities contained both COI barcode and 28S sequences, compared with COI barcode data alone. This is in agreement with other studies that suggest that multilocus approaches outperform single locus ones in phylogenetic reconstruction [20,81]. The mitochondrial gene COI has a higher mutation rate [82] and smaller effective population sizes compared to nuclear genes, which allows for effective sorting at recent lineage splits but becomes rapidly saturated (i.e., multiple hits on the same positions) proving little resolution at deeper nodes [83]. Thus, the combination of genes with different substitution rates helps to resolve relationships at different hierarchical levels, from deeper nodes to shallow relationships [84]. The former observation is well exemplified by the remarkable increase in monophyletic families reported when including the 28S data (Table 4), while the number of monophyletic genera remained similar.

4.2. Phylogenetic Diversity (PD)

Our results suggest that trees inferred from limited data affect not only the estimation of phylogenetic diversity (PD), but also the ranking of PD and the ecological processes inferred for local communities. In our analyses, we have assumed that the use of family-level topological constraints [27], combined with multilocus data [81] and the use of a backbone matrix ([19,25]; but see [32]), could provide the most accurate representation of the underlying evolutionary relationships of the sampled taxa, and thus a more accurate estimation of phylogenetic community structure. Tree topology and branch lengths can influence estimates of phylogenetic community metrics in many ways [22,24,25,85,86], for example, underestimating phylogenetic diversity and reducing the power to detect non-random community phylogenetic structure [23]. Our results suggest that phylogenetic trees constructed only using the species pool from our communities or without topological constraint tend to overestimate PD compared to the reference tree. Other studies using simulations have shown that community phylogenies that only include the species found in the sampled site or species pool of interest also produce lower estimates of PD relative to pruned trees [19].

The trees based on more limited amounts of data failed to recover the same ranking as the reference tree. The R2 values of the linear regression models revealed a large variation in the ranks obtained with the different trees, with the non-ultrametric and COI barcode sequence-only trees being the least correlated (R2 < 0.57). The GLMM analyses showed that all the factors tested (topological constraints, backbone matrices, additional genetic markers and time calibration) had a proportionally similar

effect in failing to recover the same ranking as the reference tree. However, the lowest R2 values of the linear regression models suggested that time information and the use of COI barcode data alone have the highest impact on correctly ranking the community plots according to their PD. Sensitivity of the community plot ranking to the tree topology has also been observed in other studies [19].

It is generally assumed that larger amounts of data result in more resolved and better supported tree inferences. However, little is known about how trees inferred from different amounts of data affect PD estimations. Some studies have found that PD indices estimated from COI barcode sequence-only phylogenies do not significantly differ from those estimated from multigene trees [27], while others suggest that multilocus trees outperform single locus ones in estimating phylogenetic diversity [81]. Our results support this last view, as many trees based on COI barcode data alone tended to overestimate both PD and its extent. Topological constraints, on the other hand, seem to have a relatively low impact on PD values and phylogenetic structure.

4.3. Use of Time-Stamped vs. Non-Ultrametric Trees in PD Estimation

The branch lengths of non-ultrametric trees indicate divergence in characters (genes) used for reconstructing the trees, while chronograms with time-calibrated branch lengths indicate time since species diverged [26,87]. The relative branch lengths between two species in a non-ultrametric tree can differ from those in a time-stamped tree, because during the smoothing process of dating the tree, the long branches are shortened while short branches are lengthened. There are some species from our communities (such as *Marilynia bicolor, Iberina montana, Pirata piraticus, Titanoeca schineri, Steatoda* sp., *Walckenaeria* n.sp. and *Dictyna pusilla*) that have long branches in the non-ultrametric trees (available at the Zenodo Digital Repository at 10.5281/zenodo.3941712), thus increasing the PD distances between species compared to the time-stamped trees, on which long and short branches are accordingly shortened or lengthened during the smoothing process of time-calibration. Other studies have also shown a low correlation and significantly different results in phylogenetic diversity indices when using time-stamped versus non-ultrametric trees [22,25,26,87].

Although all factors tested had a similar effect in failing to correctly rank the community plots according to their PD, most non-ultrametric trees were the least correlated with the rank values of the best resolved tree. Other studies have also shown differences in PD estimations and rank of community plots with different dating methods [19,26]. The calibration or not of a phylogenetic tree has an important effect on the ecological patterns inferred, thus affecting the interpretation of PD [26,88]. In our case, the non-ultrametric trees inferred significant PD clustering and PD overdispersion in 3 more plots than the time-stamped trees and failed to detect significant PD clustering in another 3 plots. This contrasts with other studies using simulations which suggest that phylogenetic diversity measures are not sensitive to branch lengths of the phylogeny, but to topology [24].

As the branch lengths of non-ultrametric trees are proportional to the number of character changes, they have been suggested to be correlated with ecological and phenotypic traits (feature diversity) [88,89]. Thus, non-ultrametric trees have been used for comparing PD with trait diversity within a community [25,90]. On the other hand, time-stamped trees, in which branch lengths indicate time since species divergence, reveal not only the topology but also the absolute time shared by species, providing a more accurate record of the evolutionary processes acting on the ecosystems. It has been stated that both types of methods are plausible depending on the question being asked [25], arguing that the interpretation of differences in ecological patterns inferred in the community are related to time or to the amount of trait changes (calibrated or non-ultrametric trees, respectively) [26].

There are other considerations to account for when reconstructing phylogenetic trees for community phylogenetics studies. For example, the different methods of phylogenetic tree generation, such as purpose-built phylogenies (based on the reconstruction of a phylogenetic tree for a set of target taxa and sequence data) vs. synthesis-based phylogenies (phylogenies based on available synthesis trees), affect phylogenetic diversity (PD) metrics for community phylogenetic analyses [21].

Considering the implicit effort of reconstructing a phylogeny from source (taxon sampling, taxon ID, DNA extraction, amplification and sequencing, plus reconstruction of the phylogenetic tree), some studies [21] suggest the use of synthesis-based phylogenies for calculating phylogenetic diversity metrics, because the PD estimations were similar with the two approaches. For some organisms, the use of synthesis-based trees for community phylogenetic analyses is a good approach because there are good available resolved phylogenies (e.g., Open Tree of Life, OTL [91] or Phylomatic [92] for plants). However, in the case of specific communities such as those included in the present study (Iberian Peninsula and Macaronesia), or in communities that contain undescribed and cryptic species, the sequencing of the taxa present in the community is fully necessary [21]. The taxonomic resolution of the phylogenetic trees (family/species-level phylogenies) for inferring community phylogenetic patterns also depends on the scale of study. For macroecological studies, a supertree approach can be enough for understanding phylogenetic structure on a broad spatial scale [18]; but for small or fine regional scales, species-level phylogeny including the species sampled from the local communities are the most reliable method for understanding phylogenetic community assembly patterns.

4.4. Recommendations and Suggested Workflow

One of our main goals was to provide guidelines to help researchers in making informed decisions about the best strategy to infer a tree for estimating phylogenetic diversity (see Figure 3 for the suggested workflow). Our recommendations are by no means restricted to spiders, but to any other taxa for which the kind of information here discussed can be either compiled or generated. Specifically, we interrogated how the following factors affect phylogenetic diversity rankings and inferred phylogenetic structure for local communities: (1) using DNA barcode sequences, commonly generated in community studies for accelerating species identification; (2) combining DNA barcode sequences with more slowly evolving markers (i.e., 28S); (3) incorporating molecular data from local communities into larger matrices (including more extensive taxa and genes of the target lineage diversity); (4) using topological constraints derived from more extensive phylogenetic studies; and (5) transforming branch lengths to represent absolute time. Incorporating all, some or none of the treatments/data listed above will depend on either data availability or tradeoffs between time and resource costs for generating data and the potential downstream benefits.

A major improvement both in terms of phylogenetic relationships and estimates of PD was accomplished by simply combining the DNA barcodes with a more conserved marker, in our case 28S. The amount of improvement was similar or even greater than that obtained by combining DNA barcode sequences with a backbone matrix. Although generating an additional marker increases cost, our results indicate this to be a worthy investment. Assembling a backbone matrix, on the other hand, may be a financially more economical alternative, or even complement, if relevant sequences are available within public databases. In this regard, and contrary to previous suggestions, recent studies have shown that taxonomic inaccuracy in public databases is of limited concern, at least for non-parasitic macroscopic animals [93], and several programs are available that automatize the assembly of backbone matrices from public databases (e.g., [94–96]). Interestingly, our results revealed that an increase in missing data introduced by combining barcode sequences with a backbone matrix had little effect on the topology of the inferred tree. Surprisingly, the use of topological constraints yielded trees very similar to those from unconstrained searches. Constrained trees slightly increased monophyly indexes for higher taxa but had little effect on patterns of community phylogenetic diversity. Thus, the use of topological constraints may be less important when either additional markers or backbone matrices are available, although they may help to transform trees into chronograms (see below).
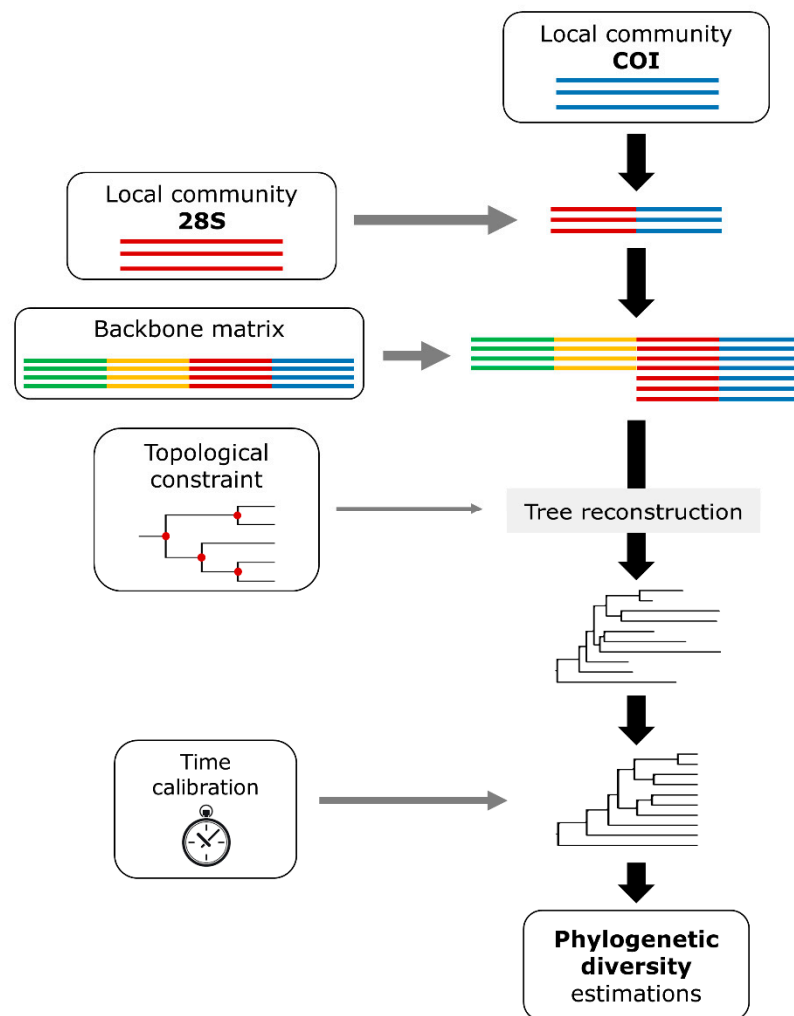
Figure 3. Suggested workflow for reconstructing phylogenetic trees for community-level diversity estimates. Arrow thickness indicates the importance of each factor.

Consistent with findings for other organisms (e.g., vascular plants [22,25,26,87] or reptiles [97]), our study demonstrates substantial differences in community phylogenetic structure inferred from non-ultrametric trees (phylograms) compared to their time-stamped counterparts (chronograms). It has been argued that both tree types may be relevant, as they reflect different aspects of evolutionary history and trait change [25]. We argue that there is little basis for assuming that genetic divergence correlates with the amount of phenotypic change, as suggested by former authors [88,89]. Moreover, phylogram branch lengths not only convey substitution rate, but also time, making it impossible to consider one without the influence of the other. Therefore, where possible, we would recommend the use of time-stamped trees for estimating phylogenetic diversity, as they provide a more robust representation of evolutionary time. The limited representativeness of the overall biodiversity of a lineage contained within a given community may limit the options (i.e., calibration information) for estimating divergence times. In these situations, the incorporation of backbone matrices will increase the chance of finding calibration points or, as in our case, the combination with topological constraints will offer the possibility of transferring time estimates of former studies into our community estimates.

5. Conclusions and Perspectives

Spiders are well represented in public repositories of DNA sequences [43,44], providing excellent opportunities for generating phylogenetic information for community level studies, especially when

combined with additional samples generated from local inventories. Here, we built a densely sampled, species-level tree of spiders combining Sanger sequencing and topological constraints derived from recent phylogenomic studies that can be used to investigate patterns of phylogenetic diversity and structure across local communities. We further used this tree to investigate how the availability of multiple molecular markers, taxonomic sampling completeness, use of topological constraints and incorporation of time information may ultimately influence inferences of PD and community phylogenetic structure. While the most robust trees that we infer provide a scaffold for examining phylogenetic diversity in spider communities worldwide, our results should be generalizable to other taxonomic groups.

References

1. Cavender-Bares, J.; Kozak, K.H.; Fine, P.V.A.; Kembel, S.W. The merging of community ecology and phylogenetic biology. *Ecol. Lett.* **2009**, *12*, 693–715. [CrossRef] [PubMed]
2. Mittelbach, G.G.; Schemske, D.W. Ecological and evolutionary perspectives on community assembly. *Trends Ecol. Evol.* **2015**, *30*, 241–247. [CrossRef] [PubMed]
3. Devictor, V.; Mouillot, D.; Meynard, C.; Jiguet, F.; Thuiller, W.; Mouquet, N. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: The need for integrative conservation strategies in a changing world. *Ecol. Lett.* **2010**, *13*, 1030–1040. [CrossRef] [PubMed]

4.   Morlon, H. Phylogenetic approaches for studying diversification. *Ecol. Lett.* **2014**, *17*, 508–525. [CrossRef] [PubMed]

5.   Mouquet, N.; Devictor, V.; Meynard, C.N.; Munoz, F.; Bersier, L.F.; Chave, J.; Couteron, P.; Dalecky, A.; Fontaine, C.; Gravel, D.; et al. Ecophylogenetics: Advances and perspectives. *Biol. Rev.* **2012**, *87*, 769–785. [CrossRef]

6.   Vamosi, S.M.; Heard, S.B.; Vamosi, J.C.; Webb, C.O. Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol. Ecol.* **2009**, *18*, 572–592. [CrossRef]

7.   Gerhold, P.; Cahill, J.F.; Winter, M.; Bartish, I.V.; Prinzing, A. Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Funct. Ecol.* **2015**, *29*, 600–614. [CrossRef]

8.   Narwani, A.; Matthews, B.; Fox, J.; Venail, P. Using phylogenetics in community assembly and ecosystem functioning research. *Funct. Ecol.* **2015**, *29*, 589–591. [CrossRef]

9.   Speed, J.D.M.; Skjelbred, I.Å.; Barrio, I.C.; Martin, M.D.; Berteaux, D.; Bueno, C.G.; Christie, K.S.; Forbes, B.C.; Forbey, J.; Fortin, D.; et al. Trophic interactions and abiotic factors drive functional and phylogenetic structure of vertebrate herbivore communities across the Arctic tundra biome. *Ecography* **2019**, *42*, 1152–1163. [CrossRef]

10. Zhang, L.; Wang, F.; Qiao, L.; Dietrich, C.H.; Matsumura, M.; Qin, D. Population structure and genetic differentiation of tea green leafhopper, *Empoasca (Matsumurasca) onukii*, in China based on microsatellite markers. *Sci. Rep.* **2019**, *9*, 1202. [CrossRef]

11. Winter, M.; Devictor, V.; Schweiger, O. Phylogenetic diversity and nature conservation: Where are we? *Trends Ecol. Evol.* **2013**, *28*, 199–204. [CrossRef] [PubMed]

12. Grab, H.; Branstetter, M.G.; Amon, N.; Urban-Mead, K.R.; Park, M.G.; Gibbs, J.; Blitzer, E.J.; Poveda, K.; Loeb, G.; Danforth, B.N. Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. *Science* **2019**, *363*, 282–284. [CrossRef] [PubMed]

13. Kembel, S.W. Disentangling niche and neutral influences on community assembly: Assessing the performance of community phylogenetic structure tests. *Ecol. Lett.* **2009**, *12*, 949–960. [CrossRef] [PubMed]

14. Schweiger, O.; Klotz, S.; Durka, W.; Kühn, I. A comparative test of phylogenetic diversity indices. *Oecologia* **2008**, *157*, 485–495. [CrossRef]

15. Miller, E.T.; Farine, D.R.; Trisos, C.H. Phylogenetic community structure metrics and null models: A review with new methods and software. *Ecography* **2017**, *40*, 461–477. [CrossRef]

16. Tucker, C.M.; Cadotte, M.W.; Carvalho, S.B.; Jonathan Davies, T.; Ferrier, S.; Fritz, S.A.; Grenyer, R.; Helmus, M.R.; Jin, L.S.; Mooers, A.O.; et al. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol. Rev.* **2017**, *92*, 698–715. [CrossRef]

17. Faith, D.P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **1992**, *61*, 1–10. [CrossRef]

18. Qian, H.; Zhang, J. Are phylogenies derived from family-level supertrees robust for studies on macroecological patterns along environmental gradients? *J. Syst. Evol.* **2016**, *54*, 29–36. [CrossRef]

19. Park, D.S.; Worthington, S.; Xi, Z. Taxon sampling effects on the quantification and comparison of community phylogenetic diversity. *Mol. Ecol.* **2018**, *27*, 1296–1308. [CrossRef]

20. Patrick, L.E.; Stevens, R.D. Investigating sensitivity of phylogenetic community structure metrics using North American desert bats. *J. Mammal.* **2014**, *95*, 1240–1253. [CrossRef]

21. Li, D.; Trotta, L.; Marx, H.E.; Allen, J.M.; Sun, M.; Soltis, D.E.; Soltis, P.S.; Guralnick, R.P.; Baiser, B. For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology* **2019**, *100*, e02788. [CrossRef]

22. Molina-Venegas, R.; Roquet, C. Directional biases in phylogenetic structure quantification: A Mediterranean case study. *Ecography* **2014**, *37*, 572–580. [CrossRef]

23. Swenson, N.G. Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. *PLoS ONE* **2009**, *4*, e4390. [CrossRef]

24. Cadotte, M.W. Phylogenetic diversity-ecosystem function relationships are insensitive to phylogenetic edge lengths. *Funct. Ecol.* **2015**, *29*, 718–723. [CrossRef]

25. Jantzen, J.R.; Whitten, W.M.; Neubig, K.M.; Majure, L.C.; Soltis, D.E.; Soltis, P.S. Effects of taxon sampling and tree reconstruction methods on phylodiversity metrics. *Ecol. Evol.* **2019**, *9*, 9479–9499. [CrossRef]

26. Allen, J.M.; Germain-Aubrey, C.C.; Barve, N.; Neubig, K.M.; Majure, L.C.; Laffan, S.W.; Mishler, B.D.; Owens, H.L.; Smith, S.A.; Whitten, W.M.; et al. Spatial Phylogenetics of Florida Vascular Plants: The Effects of Calibration and Uncertainty on Diversity Estimates. *iScience* **2019**, *11*, 57–70. [CrossRef]

27. Boyle, E.E.; Adamowicz, S.J. Community phylogenetics: Assessing tree reconstruction methods and the utility of DNA barcodes. *PLoS ONE* **2015**, *10*, e0126662. [CrossRef]

28. Hebert, P.D.N.; Ratnasingham, S.; DeWaard, J.R. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B Biol. Sci.* **2003**, *270* (Suppl. 1), S96–S99. [CrossRef]

29. Cardoso, P. Diversity and community assembly patterns of epigean vs. troglobiont spiders in the Iberian Peninsula. *Int. J. Speleol.* **2012**, *41*, 83–94. [CrossRef]

30. Crozier, R.H.; Dunnett, L.J.; Agapow, P.-M. Phylogenetic Biodiversity Assessment Based on Systematic Nomenclature. *Evol. Bioinforma.* **2005**, *1*, 11–36. [CrossRef]

31. Purschke, O.; Schmid, B.C.; Sykes, M.T.; Poschlod, P.; Michalski, S.G.; Durka, W.; Kühn, I.; Winter, M.; Prentice, H.C. Contrasting changes in taxonomic, phylogenetic and functional diversity during a long-term succession: Insights into assembly processes. *J. Ecol.* **2013**, *101*, 857–866. [CrossRef]

32. Erickson, D.L.; Jones, F.A.; Swenson, N.G.; Pei, N.; Bourg, N.; Chen, W.; Davies, S.J.; Ge, X.; Hao, Z.; Howe, R.W.; et al. Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: A mega-phylogeny approach. *Front. Genet.* **2014**, *5*, 358. [CrossRef] [PubMed]

33. Kress, W.J.; Erickson, D.L.; Swenson, N.G.; Thompson, J.; Uriarte, M.; Zimmerman, J.K. Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a puerto rican forest dynamics plot. *PLoS ONE* **2010**, *5*, e15409. [CrossRef]

34. Uriarte, M.; Swenson, N.G.; Chazdon, R.L.; Comita, L.S.; John Kress, W.; Erickson, D.; Forero-Montaña, J.; Zimmerman, J.K.; Thompson, J. Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: Implications for community assembly. *Ecol. Lett.* **2010**, *13*, 1503–1514. [CrossRef]

35. Schmitz, O.J.; Hambäck, P.A.; Beckerman, A.P. Trophic cascades in terrestrial systems: A review of the effects of carnivore removals on plants. *Am. Nat.* **2000**, *155*, 141–153. [CrossRef]

36. Cardoso, P.; Pekár, S.; Jocqué, R.; Coddington, J.A. Global patterns of guild composition and functional diversity of spiders. *PLoS ONE* **2011**, *6*, e21710. [CrossRef]

37. Carvalho, J.C.; Cardoso, P.; Crespo, L.C.; Henriques, S.; Carvalho, R.; Gomes, P. Determinants of beta diversity of spiders in coastal dunes along a gradient of mediterraneity. *Divers. Distrib.* **2011**, *17*, 225–234. [CrossRef]

38. Carvalho, J.C.; Malumbres-Olarte, J.; Arnedo, M.A.; Crespo, L.C.; Domènech, M.; Cardoso, P. Taxonomic divergence and functional convergence in Iberian spider forest communities: Insights from beta diversity partitioning. *J. Biogeogr.* **2019**, *47*, 288–300. [CrossRef]

39. Ulrich, W.; Hajdamowicz, I.; Zalewski, M.; Stańska, M.; Ciurzycki, W.; Tykarski, P. Species assortment or habitat filtering: A case study of spider communities on lake islands. *Ecol. Res.* **2010**, *25*, 375–381. [CrossRef]

40. Ridel, A.; Lafage, D.; Devogel, P.; Lacoue-labarthe, T.; Petillion, J. Habitat filtering differentially modulates phylogenetic vs functional diversity relationships between dominant ground-dwelling arthropods in salt marshes. *bioRxiv* 2020. [CrossRef]

41. Barrett, R.D.H.; Hebert, P. Identifying spiders through DNA barcodes. *Can. J. Zool.* **2005**, *83*, 481–491. [CrossRef]

42. Blagoev, G.A.; DeWaard, J.R.; Ratnasingham, S.; DeWaard, S.L.; Lu, L.; Robertson, J.; Telfer, A.C.; Hebert, P. Untangling taxonomy: A DNA barcode reference library for Canadian spiders. *Mol. Ecol. Resour.* **2016**, *16*, 325–341. [CrossRef]

43. Wheeler, W.C.; Coddington, J.A.; Crowley, L.M.; Dimitrov, D.; Goloboff, P.A.; Griswold, C.E.; Hormiga, G.; Prendini, L.; Ramírez, M.J.; Sierwald, P.; et al. The spider tree of life: Phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics* **2016**, *33*, 574–616. [CrossRef]

44. Fernández, R.; Kallal, R.J.; Dimitrov, D.; Ballesteros, J.A.; Arnedo, M.A.; Giribet, G.; Hormiga, G. Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. *Curr. Biol.* **2018**, *28*, 1489–1497.e5. [CrossRef]

45. Crespo, L.C.; Domènech, M.; Enguídanos, A.; Malumbres-Olarte, J.; Moya-Laraño, J.; Frías-López, C.; Macías-Hernández, N.; De Mas, E.; Mazzuca, P.; Mora, E.; et al. A DNA barcode-assisted annotated checklist of the spider (Arachnida, Araneae) communities associated to white oak woodlands in Spanish National Parks. *Biodivers. Data J.* **2018**, *6*, e29443. [CrossRef]

46. Malumbres-olarte, J.; Cardoso, P.; Crespo, L.C. Standardised inventories of spiders (*Arachnida, Araneae*) of Macaronesia I: The native forests of the Azores (Pico and Terceira islands). *Biodivers. Data J.* **2019**, *7*, e32625. [CrossRef]

47. Malumbres-Olarte, J.; Boieiro, M.; Cardoso, P.; Carvalho, R.; Crespo, L.C.F.; Gabriel, R.; Hernández, N.M.; Paulo, O.S.; Pereira, F.; Rego, C.; et al. Standardised inventories of spiders (*Arachnida, Araneae*) of Macaronesia II: The native forests and dry habitats of Madeira archipelago (Madeira and Porto Santo islands). *Biodivers. Data J.* **2020**, *8*, e47502. [CrossRef]

48. Emerson, B.C.; Casquet, J.; López, H.; Cardoso, P.; Borges, P.A.V.; Mollaret, N.; Oromí, P.; Strasberg, D.; Thébaud, C. A combined field survey and molecular identification protocol for comparing forest arthropod biodiversity across spatial scales. *Mol. Ecol. Resour.* **2016**, *17*, 694–707. [CrossRef]

49. Folmer, O.; Black, M.; Hoeh, W.; Lutz, R.; Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **1994**, *3*, 294–299. [CrossRef]

50. Simon, C.; Frati, F.; Beckenbach, A.; Crespi, B.; Liu, H.; Flook, P. Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. *Ann. Entomol. Soc. Am.* **1994**, *87*, 651–701. [CrossRef]

51. Hedin, M.C.; Maddison, W.P. A combined molecular approach to phylogeny of the jumping spider subfamily Dendryphantinae (Araneae: Salticidae). *Mol. Phylogenet. Evol.* **2001**, *18*, 386–403. [CrossRef] [PubMed]

52. Whiting, M.F.; Carpenter, J.C.; Wheeler, Q.D.; Wheeler, W.C. The strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* **1997**, *46*, 1–68. [CrossRef] [PubMed]

53. Giribet, G.; Rambla, M.; Carranza, S.; Baguñà, J.; Riutort, M.; Ribera, C. Phylogeny of the Arachnid Order Opiliones (Arthropoda) Inferred from a Combined Approach of Complete 18S and Partial 28S Ribosomal DNA Sequences and Morphology. *Mol. Phylogenet. Evol.* **1999**, *11*, 296–307. [CrossRef] [PubMed]

54. Edgecombe, G.D.; Giribet, G. A century later-A total evidence re-evaluation of the phylogeny of scutigeromorph centipedes (Myriapoda: Chilopoda). *Invertebr. Syst.* **2006**, *20*, 503–525. [CrossRef]

55. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]

56. World Spider Catalog. Version 21.0. Natural History Museum Bern. Available online: http://wsc.nmbe.ch (accessed on 3 March 2020). [CrossRef]

57. Stöver, B.C.; Müller, K.F. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinform.* **2010**, *11*, 7. [CrossRef]

58. R Core Research Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2004; Volume 2, ISBN 3-900051-07-0.

59. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2013**, *29*, 644–652. [CrossRef]

60. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef]

61. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

62. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2018**, *20*, 1160–1166. [CrossRef]

63. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]

64. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*; IEEE: Piscataway, NJ, USA, 2010. [CrossRef]

65. Harmon, L.J.; Weir, J.T.; Brock, C.D.; Glor, R.E.; Challenger, W. GEIGER: Investigating evolutionary radiations. *Bioinformatics* **2008**, *24*, 129–131. [CrossRef] [PubMed]

66. Smith, S.A.; O'Meara, B.C. treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **2012**, *28*, 2689–2690. [CrossRef] [PubMed]

67. Magalhaes, I.L.F.; Azevedo, G.H.F.; Michalik, P.; Ramírez, M.J. The fossil record of spiders revisited: Implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic. *Biol. Rev.* **2020**, *95*, 184–217. [CrossRef]

68. Schliep, K.P. Phangorn: Phylogenetic analysis in R. *Bioinformatics* **2011**, *27*, 592–593. [CrossRef]

69. Robinson, D.F.; Foulds, L.R. Comparison of phylogenetic trees. *Math. Biosci.* **1981**, *53*, 131–147. [CrossRef]

70. Schwery, O.; O'Meara, B.C. MonoPhy: A simple R package to find and visualize monophyly issues. *PeerJ Comput. Sci.* **2016**, *2*, e56. [CrossRef]

71. Cardoso, P.; Rigal, F.; Carvalho, J.C. BAT-Biodiversity Assessment Tools, an R package for the measurement and estimation of alpha and beta taxon, phylogenetic and functional diversity. *Methods Ecol. Evol.* **2015**, *6*, 232–236. [CrossRef]

72. De Queiroz, A.; Gatesy, J. The supermatrix approach to systematics. *Trends Ecol. Evol.* **2007**, *22*, 34–41. [CrossRef]

73. Bates, D.; Mächler, M.; Bolker, B.M.; Walker, S.C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]

74. Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; Team, R.C. nlme: Linear and Nonlinear Mixed Effects Models. *R Package Version* **2020**, *3*, 111.

75. Barton, K. MuMIn: Multi-model inference. *R Package Version* **2020**, *1*, 18.

76. Nakagawa, S.; Schielzeth, H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **2013**, *4*, 133–142. [CrossRef]

77. Allouche, O.; Tsoar, A.; Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **2006**, *43*, 1223–1232. [CrossRef]

78. Hardy, O.J. Testing the spatial phylogenetic structure of local communities: Statistical performances of different null models and test statistics on a locally neutral community. *J. Ecol.* **2008**, *96*, 914–926. [CrossRef]

79. Heath, T.A.; Hedtke, S.M.; Hillis, D.M. Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* **2008**, *46*, 239–257. [CrossRef]

80. Zwickl, D.J.; Hillis, D.M. Increased Taxon Sampling Greatly Reduces Phylogenetic Error. *Oxford Journals* **2020**, *51*, 588–598. [CrossRef]

81. Liu, J.; Liu, J.; Shan, Y.X.; Ge, X.J.; Burgess, K.S. The use of DNA barcodes to estimate phylogenetic diversity in forest communities of southern China. *Ecol. Evol.* **2019**, *9*, 5372–5379. [CrossRef]

82. Brown, W.M.; George, M.; Wilson, A.C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 1967–1971. [CrossRef]

83. Bidegaray-Batista, L.; Arnedo, M.A. Gone with the plate: The opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evol. Biol.* **2011**, *11*, 317. [CrossRef]

84. Hillis, D.M. Molecular versus morphological approaches to systematics. *Annu. Rev. Ecol. Syst.* **1987**, *18*, 23–42. [CrossRef]

85. Davies Jonathan, T.; Buckley, L.B. Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philos. Trans. R. Soc. B Biol. Sci.* **2011**, *366*, 2414–2425. [CrossRef] [PubMed]

86. Mazel, F.; Davies, T.J.; Gallien, L.; Renaud, J.; Groussin, M.; Münkemüller, T.; Thuiller, W. Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography* **2016**, *39*, 913–920. [CrossRef]

87. Elliott, M.J.; Knerr, N.J.; Schmidt-Lebuhn, A.N. Choice between phylogram and chronogram can have a dramatic impact on the location of phylogenetic diversity hotspots. *J. Biogeogr.* **2018**, *45*, 2190–2201. [CrossRef]

88. Thornhill, A.H.; Baldwin, B.G.; Freyman, W.A.; Nosratinia, S.; Kling, M.M.; Morueta-Holme, N.; Madsen, T.P.; Ackerly, D.D.; Mishler, B.D. Spatial phylogenetics of the native California flora. *BMC Biol.* **2017**, *15*, 96. [CrossRef] [PubMed]

89. Faith, D.P.; Lozupone, C.A.; Nipperess, D.; Knight, R. The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's "phylogenetic beta diversity" framework. *Int. J. Mol. Sci.* **2009**, *10*, 4723–4741. [CrossRef] [PubMed]

90. Anderson, T.M.; Shaw, J.; Olff, H. Ecology's cruel dilemma, phylogenetic trait evolution and the assembly of Serengeti plant communities. *J. Ecol.* **2011**, *99*, 797–806. [CrossRef]

91. Hinchliff, C.E.; Smith, S.A.; Allman, J.F.; Burleigh, J.G.; Chaudhary, R.; Coghill, L.M.; Crandall, K.A.; Deng, J.; Drew, B.T.; Gazis, R.; et al. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12764–12769. [CrossRef]

92. Webb, C.O.; Donoghue, M.J. Phylomatic: Tree assembly for applied phylogenetics. *Mol. Ecol. Notes* **2005**, *5*, 181–183. [CrossRef]

93. Leray, M.; Knowlton, N.; Ho, S.L.; Nguyen, B.N.; Machida, R.J. GenBank is a reliable resource for 21st century biodiversity research. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22651–22656. [CrossRef]

94. Bennett, D.J.; Hettling, H.; Silvestro, D.; Zizka, A.; Bacon, C.D.; Faurby, S.; Vos, R.A.; Antonelli, A. Phylotar: An automated pipeline for retrieving orthologous DNA sequences from GenBank in R. *Life* **2018**, *8*, 20. [CrossRef] [PubMed]

95. Xu, X.; Dimitrov, D.; Rahbek, C.; Wang, Z. NCBIminer: Sequences harvest from Genbank. *Ecography* **2015**, *38*, 426–430. [CrossRef]

96. Smith, S.A.; Walker, J.F. PyPHLAWD: A python tool for phylogenetic dataset construction. *Methods Ecol. Evol.* **2019**, *10*, 104–108. [CrossRef]

97. Fenker, J.; Tedeschi, L.G.; Pyron, R.A.; Nogueira, C. de C. Phylogenetic diversity, habitat loss and conservation in South American pitvipers (Crotalinae: *Bothrops and Bothrocophias*). *Divers. Distrib.* **2014**, *20*, 1108–1119. [CrossRef]