<https://helda.helsinki.fi>

Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data : A new subtribal classification and a temporal diversification framework

Cardueae Radiations Grp

2019-08

Cardueae Radiations Grp 2019 , ' Nuclear and plastid DNA phylogeny of tribe Cardueae
(Compositae) with Hyb-Seq data : A new subtribal classification and a temporal
diversification framework ' , Molecular Phylogenetics and Evolution , vol. 137 , pp. 313-332 . <https://doi.org/10.1016/j.ympev.2019.05.001>

<http://hdl.handle.net/10138/318086>

<https://doi.org/10.1016/j.ympev.2019.05.001>

cc_by_nc_nd

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Nuclear and plastid DNA phylogeny of the tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal framework for the origin of the tribe and the subtribes

Sonia Herrando-Moraira^{a,*}, Juan Antonio Calleja^b, Mercè Galbany-Casals^b, Núria Garcia-Jacas^a, Jian-Quan Liu^c, Javier López-Alvarado^b, Jordi López-Pujol^a, Jennifer R. Mandel^d, Noemí Montes-Moreno^a, Cristina Roquet^{b,e}, Llorenç Sáez^b, Alexander Sennikov^f, Alfonso Susanna^a, Roser Vilatersana^a

^a Botanic Institute of Barcelona (IBB, CSIC-ICUB), Pg. del Migdia, s.n., 08038 Barcelona, Spain

^b Systematics and Evolution of Vascular Plants (UAB) – Associated Unit to CSIC, Departament de Biologia Animal, Biologia Vegetal i Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, ES-08193 Bellaterra, Spain

^c Key Laboratory for Bio-Resources and Eco-Environment, College of Life Sciences, Sichuan University, Chengdu, China

^d Department of Biological Sciences, University of Memphis, Memphis, TN 38152, USA

^e Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA (Laboratoire d'Ecologie Alpine), FR-38000 Grenoble, France

^f Botanical Museum, Finnish Museum of Natural History, PO Box 7, FI-00014 University of Helsinki, Finland; and Herbarium, Komarov Botanical Institute of Russian Academy of Sciences, Prof. Popov str. 2, 197376 St. Petersburg, Russia

*Corresponding author at: Botanic Institute of Barcelona (IBB, CSIC-ICUB), Pg. del Migdia, s. n., ES-08038 Barcelona, Spain. E-mail address: sonia.herrando@gmail.com (S. Herrando-Moraira).

Abstract

Classification of the tribe Cardueae in natural subtribes has always been a challenge due to the lack of support of some critical branches in previous phylogenies based on traditional Sanger markers. With the aim to propose a new subtribal delimitation, we applied a Hyb-Seq approach to a set of 76 Cardueae species representing all the subtribes and informal groups defined in the tribe, targeting 1061 nuclear conserved orthology loci (COS) designed for Compositae and obtaining chloroplast coding regions as by-product of off-target reads. For the extraction of target nuclear data, we used two strategies, PHYLUCE and HybPiper, and 776 and 1055 COS loci were recovered with each of them, respectively. Additionally, 87 chloroplast genes were assembled and annotated. With the three datasets, phylogenetic relationships within the tribe were reconstructed under approaches of concatenation (using supermatrices as input for maximum likelihood analysis with RAxML) and coalescence (species tree estimated with ASTRAL based on the individual gene trees of each COS locus). The phylogenetic analyses of the nuclear datasets fully resolved virtually all nodes with very high support. Although nuclear and plastid tree topologies are highly congruent, they still present some incongruences, which are shortly discussed. On the basis of the phylogenies obtained, we propose a new taxonomic scheme of 12 monophyletic and morphologically consistent subtribes: Carlininae, Cardopatiinae, Echinopsinae, Dipteroquinae (new), Xerantheminae (new), Berardiinae (new), Staehelininae (new), Onopordinae (new), Carduinae (redelimited), Arctiinae (new), Saussureinae (new), and Centaureinae. Another main key result of the study was the high resolution recovered at the backbone of the Cardueae tree, which led to obtain better inter-subtribal relationships. Using as tree base the nuclear HybPiper phylogeny, we updated the temporal framework for the origin and diversification of the tribe and subtribes. Overall, the power of Hyb-Seq is demonstrated to solve relationships traditionally suggested by morphology but never recovered with support using Sanger sequencing of a few DNA markers.

Keywords

Asteraceae
COS targets
Subtribes
Systematics
Phylogenomics
Target enrichment

1. Introduction

The Cardueae is one of the largest tribes of the 43 described in the Compositae (Asteraceae; Funk et al., 2009), with almost 10% of the species of the whole family: 2400 species in 73 genera (Susanna and Garcia-Jacas, 2009). The tribe belongs to the subfamily Carduoideae, which is composed by four tribes (Panero and Funk, 2008; Ortiz et al., 2009): Dicomeae (97 spp.), Tarchonantheae (13 spp.), Oldenburgieae (4 spp.), and Cardueae, the latter representing the 95% of the subfamily's diversity (Susanna and Garcia-Jacas, 2009). The Carduoideae is a very successful evolutionary lineage within Compositae, which was estimated by Panero and Crozier (2016) to have the second highest diversification rate in the family and a moderate rate of extinction. The number of species is not uniform across the genera of the tribe Cardueae, being six of them highly diversified (ca. 200–600 spp.) and with a high endemism rate (*Carduus*, *Cirsium*, *Centaurea*, *Cousinia*, *Jurinea*, and *Saussurea*), while, on the other extreme, 22 (ca. 30%) are monotypic. Geographically, the tribe is distributed mainly in the Mediterranean and the Irano-Turanian regions, but it is reported from all continents except the Antarctica. The ecosystems where the Cardueae species inhabit are very variable, e.g. Mediterranean landscapes, steppes, semiarid areas, deserts, alpine meadows, or tropical savannahs (Susanna and Garcia-Jacas, 2009). Indeed, in many aspects the tribe Cardueae is an extremely heterogeneous group in terms of habit (perennial, biennial, monocarpic or annual herbs, shrubs, treelets, often spiny), karyology (high variability in chromosome numbers from $x=6$ to $x=18$, frequent dispolidy), or pollen structure (caveate or ecaveate, smooth, scabrate, or spiny). This complexity summed to its high diversity have greatly contributed to the turbulent taxonomic history of Cardueae at several ranks, e.g. tribal and subtribal (see Table 1 for a historical overview) or even misclassifications of some genera in different tribes or subtribes (see examples in Table 2).

In the first tribal classification of the family, Cassini (1819) divided the present Cardueae into three tribes: Echinopseae, Carlineae, and Cardueae, the latter comprising the subtribes Carduinae and Centaureinae. Years later, Bentham (1873) and Hoffmann (1894) proposed grouping the three former tribes in a single tribe Cardueae, with four subtribes: Carlininae Dumort., Echinopsinae ["Echinopinae"] Dumort., Carduinae Cass., and Centaureinae Dumort. This conservative approach was accepted over long time, until Wagenitz (1976) reinstated the tribe Echinopseae. Soon after, Dittrich (1977) returned to Cassini's earlier views, proposing the restoration of tribes Echinopseae, Carlineae, and Cardueae. Bremer (1994) favored the conservative classification of a single tribe Cardueae, which is nowadays generally accepted (Susanna and Garcia-Jacas, 2007, 2009), with the recent reinstatement of the subtribe Cardopatiinae (Susanna et al., 2006).

Not surprisingly, Carlininae and Echinopsinae, two of the two basalmost subtribes in molecular phylogenies, have been considered at some time as independent tribes, mainly due to their clear diagnostic characters with respect to the rest of the Cardueae assembly (Susanna and Garcia-Jacas, 2009). However, the Cardueae are morphologically consistent as a whole entity at the tribal level, sharing a unique synapomorphic morphological character within the Compositae: style with a papillose-pilose thickening below the branches and the stigmatic areas confined to the inner surface (Susanna and Garcia-Jacas, 2009). Additionally, the tribe Cardueae has also been broadly confirmed empirically as monophyletic, first by cladistic analyses based on morphology (Bremer, 1987, 1994; Karis et al., 1992), and later by molecular phylogenies (Jansen et al., 1990, 1991; Kim et al., 1992; Susanna et al., 1995, 2006; Garcia-Jacas et al., 2002; Barres et al., 2013).

In the most recent and accepted treatments (Susanna et al., 2006; Susanna and Garcia-Jacas, 2007, 2009), the authors pointed out that four subtribes are natural groups with clear limits (Cardopatiinae, Carlininae, Centaureinae, and Echinopsinae); however, subtribe Carduinae is an unnatural, artificial, and problematic group. The Carduinae have been a dumping ground of genera that do not fit in any of the other subtribes and totals 1700 species, near 70% of the whole tribe diversity (Susanna and Garcia-Jacas, 2007). The fact that the group is a questionable and heterogeneous assemblage of genera was also reflected in several phylogenetic studies, which have reported the subtribe as paraphyletic (Susanna et al., 1995, 2006; Häffner and Hellwig, 1999; Garcia-Jacas et al., 2002; Barres et al., 2013). The alternate solution of combining subtribes Carduinae and Centaureinae in one enormous subtribe was discarded, owing to the impractical constitution of a huge subtribe encompassing 2300 species and 90% of the species of the whole tribe (Susanna and Garcia-Jacas, 2007).

With the intention of fragmenting subtribe Carduinae, some informal morphological groups have been described within it (Susanna and Garcia-Jacas, 2007), which could on the long run be considered subtribes for a more natural classification: "[the alternative is] splitting present Carduinae in at least seven new subtribes (many of them presently unsupported): Xerantheminae, Staehelininae, Berardininae, Onopordinae, Carduinae, Arctiinae, and Saussureinae [...]" (Susanna and Garcia-Jacas, 2009). In view of

the lack of support for the monophyly of all segregate subtribes, especially for the most important in terms of species number (core Carduinae, Arctiinae, and Saussureinae), no formal proposal was performed in the past. A well-resolved and supported phylogenetic hypothesis for the major lineages within Carduinae is still lacking, and it would be very useful to confirm the morphological alliances or generic complexes proposed by [Susanna and Garcia-Jacas \(2009\)](#).

An additional taxonomic problem within the tribe Cardueae is that some genera or complex of genera have been classified within one subtribe or another depending on the classification followed (see some of the cases in [Table 2](#)). Moreover, in extreme cases, some genera have been independently classified in tribe Cardueae or in other tribes of Compositae (e.g. *Berardia* in tribe Mutisiae, or *Dipterocome* in tribe Calenduleae). A great proportion of subtribal misplacements were reported for uncertain cases between subtribes Carduinae and Centaureinae, which could be attributed to the inconspicuous morphological differences between both subtribes (based on microcharacters of achene and pappus) that are sometimes lacking in herbarium specimens, or are immature structures during field collections ([Susanna et al., 2006](#)). Even though many cases have been successfully resolved with the aid of the molecular phylogenies obtained with Sanger sequencing (see references in [Table 2](#)), more changes in the adscription of genera could occur in a new subtribal framework for the Cardueae.

Since to date, all previous molecular studies of the tribe have been based on Sanger sequencing data ([Susanna et al., 1995, 2006](#); [Häffner and Hellwig, 1999](#); [Garcia-Jacas et al., 2002](#); [Barres et al., 2013](#)), being the largest datasets constructed with four chloroplast regions (*trnL-trnF*, *matK*, *ndhF*, *rbcL*) and one nuclear ribosomal marker (ITS; [Barres et al., 2013](#)). Although subtribal clades of Carlininae, Echinopsinae, Cardopatiinae, and Centaureinae are well supported, Carduinae remain paraphyletic when Centaureinae are removed, and not all the informal groups have been recovered as monophyletic. Moreover, relationships between subtribes have not been resolved, and the backbone of the phylogenetic tree remains undefined. Accordingly, the available divergence time estimation of Cardueae is based on a partially resolved phylogenetic tree that was performed on a combined dataset of chloroplast markers, and this dating could be significantly improved in terms of the methodological approach and the sequence data used.

In the last decade, next generation sequencing (NGS) has emerged as an important methodological advance for solving phylogenetic problems (see review of [Harrison and Kidner, 2011](#)). Plant phylogenies of historical taxonomically complex groups are becoming resolved at different taxonomic levels, e.g. order (Ranunculales, cf. [Lane et al., 2018](#); or Zingiberales, cf. [Carlsen et al., 2018](#)), family (Goodeniaceae, cf. [Gardner et al., 2016](#)), tribe (Shoreeae, cf. [Heckenhauer et al., 2018](#)), or species complexes as *Claytonia* ([Stoughton et al., 2018](#)) or *Amaranthus* ([Viljoen et al., 2018](#)). For the Compositae, [Mandel et al. \(2014\)](#) designed a probe set that hybridizes with 1061 nuclear conserved orthology loci (hereafter COS), which in combination with genome skimming allowed also to recover other parts of the genome such as chloroplast regions (defined as Hyb-Seq technique; [Weitemier et al., 2014](#)). This methodological workflow has been proved successfully in Compositae-wide studies ([Mandel et al., 2015, 2017](#)) and was tested in a recent research focused on highly radiated genera within Cardueae ([Herrando-Moraira et al., 2018](#)). However, the COS locus set has not been yet applied to taxonomic delimitation within the family.

Thereafter, we here apply the Hyb-Seq method to a sample of 76 species representing all the subtribes and informal suprageneric groups of the Cardueae with the main goals of: (1) to obtain a well-defined phylogeny of the high-level lineages in the tribe with nuclear and chloroplast data; on the basis of this phylogeny, (2) to propose a new subtribal classification, especially focused on testing the splitting of subtribe Carduinae in smaller and more practical natural subtribes; (3) to examine previously unresolved phylogenetic relationships between subtribal lineages at the backbone of the tree; and (4) to update the temporal framework of tribe and subtribes origin and diversification, based on a better resolved phylogeny.

2. Materials and methods

2.1. Taxon sampling

To obtain a complete sampling of the tribe Cardueae, we included representatives of all described subtribes based on the taxonomic treatment of [Susanna and Garcia-Jacas \(2009\)](#): (1) Carlininae, 6 species; 2) Cardopatiinae, 2 species; 3) Echinopsinae, 3 species; 4) Carduinae, 34 species; 5) Centaureinae, 31 species. The sampling strategy was based on maximizing the number of species of unresolved informal groups, especially within Carduinae and Centaureinae, and was proportional to the total number of genera and species included in each group. In total, 76 different species of the tribe Cardueae were included. We

did not include all the genera within the tribe because the adscription of many of the ones with taxonomic problems had already been confirmed in previous works (Table 2). In addition, we incorporated 5 species as representatives of the subfamily Carduoideae outside of the Cardueae; 5 species from other subfamilies within Compositae (2 from the Mutisioideae, 2 from the Barnadesioideae, 1 from the Famatinanthoideae); and finally, *Nastanthus patagonicus* from the Calyceraceae, which has been recognized as the sister family to Compositae (Mandel et al., 2017). Overall, 58 species were newly sequenced for this study, and the data of the remaining 29 were directly obtained from raw reads from Mandel et al. (2014, 2017) or Herrando-Moraira et al. (2018). See Supplementary Table S1 for details of each sampled species.

2.2. DNA extraction, library and capture preparation, and sequencing

Around 10–30 mg of dried plant material per sample was weighed and homogenized with a Mixer Mill MM 301 (Retsch®, Haan, Germany). The DNA was extracted with the DNeasy plant mini kit (Qiagen, Valencia, CA, USA) or the E.N.Z.A SP Plant DNA Mini Kit (Omega Bio-Tek Inc., Norcross, Georgia, USA) according to the manufacturer's protocol. The total genomic DNA quantity was measured with the Qubit™ 3.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). The standardized DNA (1 µg in 70 µl) was sheared in the Genomics Unit of the Centre for Genomic Regulation (CRG, Barcelona, Spain) using a Covaris S2 System (Covaris, Woburn, MA, USA) in microTUBEs with sample volume of 50 µl and a target peak set to 400 bp. Sequencing libraries and subsequent sequence capture were conducted as specified in Herrando-Moraira et al. (2018). Additionally, for the species newly sequenced for this study (see Supplementary Table S1), we conducted a library spiking with the following proportions: 40% of unenriched solution and 60% of enriched libraries. The final spiked library pools were sequenced (pair-end 100 bp) in the DNA Sequencing Core CGRC/ICBR of the University of Florida on one lane of an Illumina HiSeq 3000 (Illumina, USA).

2.3. Raw data processing

Demultiplexed raw sequence reads were checked with FastQC v.0.10.1 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) in order to perform a first quality control step. Subsequently, raw reads were trimmed by a quality based assessment (sliding-window set to 5:20), and were also subjected to an adapter trimming using Trimmomatic v.0.36 (Bolger et al., 2014). In addition, we specified the software to remove those cleaned reads that after trimming were less than 36 bp, and those with a missing forward or reverse pair.

2.4. Nuclear data extraction

Following the methodological workflow proposed in Herrando-Moraira et al. (2018), the 1061 target COS loci were extracted with two orthology-detection pipeline packages: PHYLUCE v.1.5 (Faircloth, 2015) and HybPiper v.1.1 (Johnson et al., 2016). One of main differences between both approaches is the treatment of potential paralog loci (Herrando-Moraira et al., 2018). We extracted the COS loci with the two pipelines to be sure that the distinct paralog processing strategy was not affecting phylogenetic reconstructions in terms of topology and branch support values.

For the PHYLUCE method, the trimmed reads were *de novo* assembled into contigs with the software SPAdes v.3.9.0 (Bankevich et al., 2012), specifying different predefined k-mer lengths (21, 33, 55, and 77). Then, the recovered contigs were mapped back to target references using LASTZ (Harris, 2007), and were extracted following the methodological specifications detailed in Herrando-Moraira et al. (2018). For the HybPiper method, the trimmed reads were firstly mapped to the targets using BWA (Li and Durbin, 2009), and secondly were assembled into contigs also using SPAdes.

The multi-fasta files obtained for each target locus both with PHYLUCE and HybPiper were aligned with MAFFT v.7.266 (Katoh and Standley, 2013) using the *auto* setting mode. The ambiguously aligned regions were removed with trimAl v.1.4 (Capella-Gutiérrez et al., 2009) applying the *automated1* flag. 126 out of the total 902 loci extracted with PHYLUCE, and 1 locus out of the total 1057 loci extracted with HybPiper, were removed from each dataset because they were obtained for less than three species. One additional locus was removed from the HybPiper dataset due to its short length recovered after the alignment trimming (3 bp). Finally, for each target extraction method, two datasets were constructed in order to perform two different phylogenetic inference analyses. One consisted of the trimmed aligned sequences of each locus separately, and the other consisted of a supermatrix obtained by concatenating all trimmed aligned loci with FASconCAT-G v.1.02 (Kück and Longo, 2014). The summary statistics of the supermatrices were calculated with AMAS (Borowiec, 2016).

2.5. Chloroplast data extraction

For the reconstruction of the chloroplast genomes, we used the off-target reads also recovered with the Hyb-Seq approach. The chloroplast extraction was conducted using the pipeline package MITObim v.1.9 (Hahn et al., 2013) with the default conditions. The package has incorporated the module MIRA v.4.0.2 (Chevreux et al., 1999), which is used in mapping mode. In the first step, it identifies the more conserved regions between the total readpool and a phylogenetically related reference genome (in this case, *Cirsium arvense* NCBI accession number NC_036965.1) by mapping the trimmed interleaved reads to this initial reference. Then, the mapped reads are assembled into contigs and a new reference sequence or bait is created. To extend the reference close the gaps, two more steps are performed. The second consists of separating from the total read pool the reads that overlap with the reference, and the third recovers the reads separated to map them back to the bait, thus building a new more extended reference sequence through an assembly process of new mapped reads. These two last steps are iteratively repeated until a stationary stage of the mapped and assembled reads is reached (Hahn et al., 2013).

The chloroplast genomes recovered were annotated with the web tool application GeSeq (Tillich et al., 2017), which uses a customizable reference database to annotate the genomes using BLAT-driven best-match approach. As a database, we selected all the available annotations in NCBI of Cardueae representatives (*Carthamus tinctorius* NC_030783.1, *Centaurea diffusa* NC_024286.1, *Cirsium arvense* NC_036965.1, *Cirsium eriophorum* NC_036966.1, *Cirsium vulgare* NC_036967.1, *Cynara baetica* NC_028005.1, *Cynara cornigera* NC_028006.1, *Cynara humilis* NC_027113.1, *Saussurea chabyoungsanica* NC_036677.1, *Saussurea involucrata* NC_029465.1, *Saussurea polylephis* NC_036490.1, and *Silybum marianum* NC_028027.1). To extract the coding regions of genes (CDS) in separate files from the global multi-fasta matrix generated with GeSeq, we used the script “Phyluce_assembly_explode_get_fastas_file” of PHYLUCES package with slight modifications. The recovered CDS were individually aligned with the MACSE codon aligner (Ranwez et al., 2011). The alignments were visualized in SeaView v.4.7 (Gouy et al., 2009), and 10 species from the initial sampling were excluded due to the high presence of frameshifts and preceding codon stops. To remove the poorly aligned regions from the alignments, we used the program Gblocks v.0.91b (Castresana, 2000) with the option “codon” to trim entire codon sets. A total 87 CDS regions were recovered. We concatenated all CDS in a single supermatrix (chloroplast dataset) with FASconCAT-G v.1.02, assuming that the chloroplast genome is considered to be haploid, nonrecombinant, generally uniparentally inherited, and “single copy” (Small et al., 1998, 2004).

2.6. Phylogenetic inference analyses

The phylogeny reconstruction of the tribe Cardueae was performed under two different approaches for nuclear data: the concatenation approach (using the supermatrix dataset) and the coalescence approach (using the individual gene trees for inferring the species tree). For chloroplast data, the phylogenetic tree inference was only performed under the concatenation approach.

For the concatenation approach, Maximum Likelihood (ML) analyses were run using the software RAxML v.8.2.9 (Stamatakis, 2014), which were conducted on XSEDE in the CIPRES Science Gateway v.3.1 (Miller et al., 2010). We selected the algorithm of simultaneously searching the best ML tree (from 10 randomized maximum parsimony starting trees) and performing a rapid bootstrapping (1000 replicates). Branches were considered as statistically supported for bootstrap (BS) values > 70% (Hillis and Bull, 1993). In relation to the partition scheme, each locus (in the nuclear datasets) or gene (in the chloroplast dataset) was considered a unit of partition, using the evolution model of GTRGAMMA following recommendations of Stamatakis (2006). Output trees were visualized and exported in FigTree v.1.4.3 (Rambaut, 2016).

For the coalescence approach, gene trees were inferred also using RAxML, under the same conditions as for the concatenation based analyses but with a bootstrap resampling of 200 replicates. Resulting unrooted gene trees were inputted into ASTRAL-III v.5.5.3 (Zhang et al., 2018) to infer the species tree. The support values were calculated using local posterior probabilities (LPP; Sayyari and Mirarab, 2016). Branches with LPP > 0.95 were considered as strongly supported.

2.7. Gene trees concordance analyses

Topological conflicts among gene trees and the species tree were explored using the software Phyparts (Smith et al., 2015; <https://bitbucket.org/blackrim/phyparts>) on the nuclear HybPiper dataset. This methodological approach consists of mapping, for each node or bipartition of interest, the level of concordance/discordance between the different individual gene trees on supporting the reference tree

topology, being this the ASTRAL species tree in this case. First, gene and species trees were rooted with the online tool STRAW (Shaw et al., 2013; <http://bioinformatics.publichealth.uga.edu/SpeciesTreeAnalysis/>). The outgroup selection was based on the species divergence rank recovered by the species ML tree inferred under concatenation. When the first preferred outgroup was missing, STRAW searches the next specified species. Then, Phyparts was run, considering only the branches of gene trees with more than 33% BS support. Finally, the script phypartspiecharts.py (<https://github.com/mossmatters/MJPythonNotebooks>) was used to summarize and map the Phyparts output on the ASTRAL species tree. Owing to computing limitations of Phyparts, we could not run the analysis for the entire set of 1055 nuclear gene trees of HybPiper dataset, therefore a subset of 750 random gene trees was selected.

2.8. Divergence time analysis

The divergence time analysis was performed on the phylogenetic tree obtained under concatenation from the nuclear HybPiper dataset. This tree was time-calibrated rescaling the branch lengths using the penalized likelihood method (Sanderson, 2002), implemented in the software treePL (Smith and O'Meara, 2012). This approach is reported to produce similar results than the software BEAST (e.g. Lagomarsino et al., 2016; Stubbs et al., 2018), but, as an advantage, is capable of computing larger datasets in a lower running time. The dating procedure was divided in two main stages, which consisted on: (1) selection of the optimal model parameters; and (2) running the analysis with the optimal parameters selected and, additionally, accounting for the uncertainty in calibration points to obtain confidence intervals (95% CI) in the estimated node ages.

For both stages, five calibration points (CP) were used as node age constraints. One was a secondary dated node corresponding to the origin of the Compositae family (69.56 Myr with a 95% CI 59.02–80.17; coded here as CP1) reported in Panero and Crozier (2016). The other four points were based on fossil records: (CP2) the capitulescence of *Raiguenrayum cura* that was dated at 47.5 Myr (Barreda et al., 2012), which was used to constrain the clade of Subfamily Mutisioideae + Subfamily Carduoideae as in the most recent interpretation by Panero and Crozier (2016), instead of placing it at the crown Compositae as in Barres et al. (2013); (CP3) the achenes identified as belonging to *Cirsium* with an age of 14 Myr (Mai, 1995), which were placed at the stem node of the *Carduus-Cirsium* clade following Barres et al. (2013); (CP4) the achenes assigned to *Arctium* (López-Vinyallonga et al., 2009) and dated at 8 Myr (Mai, 2001), which were placed at the stem node of *Arctium*, correcting the misplacement of this fossil in Barres et al. (2013), where it was used at the split of *A. minus* and *A. lappa*; and (CP5) the pollen of *Centaurea* type *Cyanus* dated at 6 Myr (Wagenitz, 1955; Ivanov et al., 2007), which was placed at the stem node of *Centaurea* subgenus *Cyanus*. This is a new, more precise and better resolved *Centaurea* fossil that the pollen and achene records of Mai (1995) and Popescu (2002) that were used in Barres et al. (2013), where they were placed at stem age of *Centaurea* due to their uncertain placement within the genus.

In the first dating stage, an initial “priming” run (prime command) was carried out to detect the optimal parameter settings (opt = 4, optad = 2, and optcvad = 2). Subsequently, the best smoothing rate was also evaluated through a random subsample and replicate cross-validation procedure (selecting thorough and randomcv commands), allowing varying it from 0.001 (cvstart) to 1000 (cvstop). The best smoothing value resulted in 0.1 after cross-validation analysis, which was selected from the lowest value of a Chi-Square test. The two runs of the first stage were run using the range of 95% CI value for the CP1 (minimum = 59.02 and maximum = 80.17) and the lower bound of fossil age estimations (minimum = fossil age) for CP2-CP5.

In the second stage, we used our own R (R Core Team 2014) script (Appendix A) to obtain 10,000 random values for each CP in order to account for the uncertainty of the node age estimations. Specifically, the random values were generated under a normal distribution for CP1 (based on 95% CI estimations of Panero and Crozier, 2016; mean = 69.56 and sd = 3) and a lognormal distribution for CP2-CP5 (mean = fossil age and sd = 1.1). Once values were extracted, a new data frame was created in a spreadsheet (Appendix B), which was then used to randomly select 100 unique combinations of five numbers, i.e. one for each CP from the random pool of 10,000. Finally, 100 independent treePL analyses, each one with CP constrained to a single value (minimum = maximum), were ran using the same phylogenetic input tree and the optimized parameters found in the first stage. The resultant 100 dated tree files were modified to fit the format of TreeAnnotator v.1.7.5 (Drummond et al., 2012), and were introduced in this software to obtain a maximum clade credibility tree chronogram with median node heights and corresponding CIs.

3. Results and discussion

3.1. Performance of Hyb-Seq at the subtribal level within the tribe Cardueae

The Hyb-Seq NGS technique designed for the Compositae (Mandel et al., 2014) is confirmed here as a powerful tool to reconstruct highly resolved phylogenies at deep taxonomic levels, in line with previous studies more focused on the methods (Mandel et al., 2015, 2017; Herrando-Moraira et al., 2018). As far as we known, this is the first study in Compositae that uses the COS loci workflow to obtain taxonomic conclusions, specifically aimed to clarify the entangled delimitation of subtribes within Cardueae. The transition from traditional Sanger datasets, up to 5 markers, to NGS datasets, with 1142 markers (Table 3), has resulted in obtaining the most helpful dichotomous and confidently supported divergence pattern reported to date for Cardueae (Figs. 1, 2, and 3, Supplementary Figs. S1 and S2).

In average 5,420,504 reads ($sd = \pm 3,461,361$) were sequenced per species (Supplementary Table S2). From the total 1061 COS loci, 776 were finally recovered and used with PHYLUCE and 1055 with HybPiper, which conformed matrices of trimmed aligned sequences of 492,549 bp (189,716 parsimony informative [PI] sites) and 332,260 bp (123,73 PI sites), respectively (Table 4). When parameters related to missing data were compared between PHYLUCE and HybPiper (see parameters 4 and 11 in Table 4), it was remarkable that PHYLUCE is less efficient than HybPiper in the sequence extraction process from on-target reads, as documented in Herrando-Moraira et al. (2018). However, phylogenetic trees did not differ in topology or branch supports between both extraction methods at the subtribal level (Figs. 1 and 3, Supplementary Figs. S1 and S2), even though PHYLUCE uses a more restrictive procedure to remove potential paralogs than HybPiper (see Fig. 1 in Herrando-Moraira et al., 2018). As no differences were observed, results from HybPiper method were analyzed in deep to support taxonomic proposals. Interestingly, this lack of differences could be reflecting that paralogy incidence is relatively low in Cardueae. In agreement, the study of Jones et al. (in prep.), which tests the COS loci applicability for seven tribes of Compositae, found that Cardueae was the tribe with the lowest number of paralogs (average = 140, similar to the value found here, average = 130), in comparison with others like Vernonieae (average = 256). Also in concordance, Huang et al. (2016) did not detect any recent whole genome duplication event in Cardueae, in contrast to other Compositae tribes, which could involve paralogy issues.

As other studies have shown, Hyb-Seq can provide nearly complete datasets of chloroplast genomes (e.g. Weitemier et al., 2014; Folk et al., 2015; Carlsen et al., 2018). Regarding the present study, we were able to retrieve 87 protein coding regions of the chloroplast genome from off-target reads, which after concatenation resulted in a matrix of 78,531 bp (4290 PI sites; Table 4). Indeed, the increase of phylogenetically informative characters is responsible of notably higher topological resolution and clade support in comparison with the previous results reported with Sanger sequencing data for the tribe (Susanna et al., 1995, 2006; Häffner and Hellwig, 1999; Garcia-Jacas et al., 2002; Barres et al., 2013; Table 3). In a similar way, this effect has been already documented for other plant groups (e.g. Pinus, Parks et al., 2009; Inga, Nicholls et al., 2015), which have benefited tremendously from the advent of high-throughput sequencing techniques.

3.2. The new subtribal classification of the tribe Cardueae

We can confidently re-confirm the monophyly of subfamily Carduoideae as currently defined (tribes Dicomeae, Oldenburgieae, Tarchonantheae and Cardueae; cf. Funk et al., 2009). The tribe Cardueae results again a robust taxonomic unit, in agreement with former studies based on morphology and molecular data (Bremer, 1987, 1994; Jansen et al., 1990, 1991; Karis et al., 1992; Kim et al., 1992; Susanna et al., 1995, 2006; Garcia-Jacas et al., 2002; Barres et al., 2013). Specifically, the clade of Cardueae presented high support values in all the analyses (Figs. 1, 2, and 3). The alternative option of rising up some subtribes to a tribal rank (see Table 1) is once again discarded. As suggested by Susanna et al. (2006), it is impractical the fragmentation of a natural group that is strongly phylogenetically consistent and can be easily recognized by macromorphological characters. In this way, the classification criteria for the definition of Compositae tribes are also maintained homogeneously along the family (Funk et al., 2009).

At subtribal level, the high resolution of phylogenies provided by the Hyb-Seq data makes possible the proposal of a new classification. Subtribal limits established here are based on the application of the integrative taxonomy principles (Dayrat, 2005; Schlick-Steiner et al., 2010): firstly, considering the morphological entities to be recognized (Susanna and Garcia-Jacas, 2007, 2009), and then verifying their monophyly with molecular data. Following this model, a total of 12 subtribes are suggested. Five of them had been already previously recognized and are also confirmed with our results (Carlininae, Echinopsinae, Cardopatiinae, Carduinae, and Centaureinae; Susanna et al., 2006); seven new subtribes

result from partitioning the former wide and paraphyletic subtribe Carduinae into new monophyletic subtribes based on previously defined informal morphological groups (Dipterocominae, Xerantheminae, Berardiinae, Staehelininae, Onopordinae, Arctiinae, and Saussureinae). These new subtribes are highly heterogenic in terms of species richness, encompassing from relict single lineages to highly recently radiated groups. For instance, Dipterocominae and Berardiinae are defined as monotypic subtribes, comprising only *Dipterocome pusilla* and *Berardia subacaulis*, respectively. In an opposed way, Arctiinae and Saussureinae harbor ca. 600 and 550 species, respectively (Susanna and Garcia-Jacas, 2007). For a complete description of subtribes see below *Taxonomic proposal*. The new phylogenetic reconstruction of high-level lineages presented here will be used as a classification basis for future botanical studies of Cardueae.

Certainly, the historical overview of the tribe (Table 1) reveals that Carlininae, Echinopsinae, and Centaureinae are the best morpho-molecular defined entities, which have been long recovered with statistical support since the first phylogenetic studies (Table 3). In accordance, we obtained high clade support for these subtribes (BS = 100 and LPP = 1; Figs. 1, 2, and 3). The Carlininae initially showed a monophyletic pattern in Garcia-Jacas et al. (2002). However, the incorporation of *Tugarinovia mongolica* broke the monophyly of the subtribe in Susanna et al. (2006). This was probably caused by the fact that it is a dioecious species, a trait unique in the tribe (Susanna and Garcia-Jacas, 2009) and relatively uncommon in angiosperms (6% of species; Renner and Ricklefs, 1995). It has been suggested that dioecy has the negative impact of decreasing diversification rates (Heilbuth, 2000; Kay et al., 2006) with consequent phylogenetic misplacements (Vamوسي et al., 2003). In Barres et al. (2013) and our present study, the increase in the number of molecular markers aided to recover the monophyly of Carlininae, including *Tugarinovia mongolica* within this subtribe as a sister lineage to the rest of members. The Echinopsinae presented a persistent monophyly across all molecular studies (Table 3), except for the chloroplast-based phylogeny of Garcia-Jacas et al. (2002), in which the limited sequence variation found in chloroplast *matK* gene could have been the responsible for the low branch support of Echinopsinae. Similarly, the monophyly of subtribe Centaureinae has been fully supported in former phylogenetic studies (Table 3), with an also a unique exception of the chloroplast tree obtained with the *matK* gene by Garcia-Jacas et al. (2002).

Although Cardopatiinae is a relatively newly recovered subtribe (Susanna et al., 2006), it has resulted also in a strong-supported clade in recent phylogenies with both nuclear and chloroplast sequence data (Table 3), besides its morphological singularity (Susanna and Garcia-Jacas, 2007, 2009). In agreement, we can confirm here that the two species of Cardopatiinae (*Cardopatum corymbosum* and *Cousiniopsis atractyloides*) constitute a well-defined evolutionary segregation (BS = 100 and LPP = 1; Figs. 1, 2, and 3), therefore its taxonomic rank as subtribe seems appropriate.

Conversely, the new seven subtribes have not always been well delineated in Sanger-based phylogenies, at least with enough confidence to propose a formal subtribal splitting. From the morphological informal groups of Carduinae described in Susanna and Garcia-Jacas (2009), the *Xeranthemum* group is one of the best supported of all datasets and analyses (BS = 100 and LPP = 1; Figs. 1, 2, and 3). Thus, its consideration as subtribe Xerantheminae can be corroborated. Our results are not surprising, given that since first phylogenies the Xerantheminae species were recovered as a monophyletic and statistically highly supported group, both with nuclear and chloroplast data (Table 3).

Related to the former *Xeranthemum* group, we found one unexpected finding, which is the phylogenetic isolated position of *Dipterocome pusilla*. This species had not been incorporated to Cardueae phylogenies until the study of Anderberg et al. (2007), confirmed by Barres et al. (2013), where it was clustered with the rest of Xerantheminae members in the combined cpDNA (*matK*, *ndhF*, *rbcL*, and *trnL-trnF*) tree, but without support in a phylogeny of the ITS nuclear region. However, in the present study we found that in four of the five reconstructed phylogenies (Figs. 2 and 3, Supplementary Figs. S1 and S2), *Dipterocome* is placed in a distinct clade from Xerantheminae. Only in the tree reconstructed under concatenation with the nuclear HybPiper dataset *Dipterocome* was grouped with Xerantheminae (BS = 97; Fig. 1), in agreement with the results of Barres et al. (2013). This may probably be caused by the bias produced by concatenation analysis of NGS datasets with a large number of loci (Kubatko and Degnan, 2007; Edwards et al., 2016), that can result on long branch attraction artefacts (Liu et al., 2015). Morphologically, Susanna and Garcia-Jacas (2009) pointed out that floral characters of *Dipterocome* show close relationship with the *Xeranthemum* complex. Nevertheless, a histological analysis of the capitula structure of *Dipterocome* has shown that the purported “florets” of *Dipterocome* are actually uniflowered capitula (Susanna, pers. comm.) grouped in second-order synflorescences, an extremely unfrequent structure in the tribe shared only by the genus *Echinops*. Overall, it seems that there is enough molecular and morphological evidence to classify *Dipterocome* in a separate monotypic subtribe.

Two other controversial genera considered to belong to the former Carduinae are *Berardia* and *Staehelina*. Although both genera do not present obvious morphological affinities (Susanna and Garcia-

Jacas, 2009), some molecular phylogenies suggested that they could be evolutionary closely related. The genus *Staelina* has long been reported as a genetically isolated group within the tribe, presenting always a monophyletic pattern throughout all phylogenetic reconstructions (Table 3). Regarding its relationships with *Berardia*, the combined nuclear (ITS) plus chloroplast (*trnL-trnF* and *matK*) in Susanna et al. (2006) grouped both genera in an unsupported clade, while only considering nuclear data (ITS), they appeared not directly linked but as adjacent lineages, as detected in nuclear and chloroplast phylogenies of Barres et al. (2013). Here, we also identified the same two phylogenetic signals. Even though, only one of five phylogenetic analyses (chloroplast dataset analyzed under the concatenation approach) resulted in a moderately-high grouping of both genera (BS = 76; Fig. 2), and the rest separated *Berardia* and *Staelina* (Figs. 1 and 3, Supplementary Figs. S1 and S2). As we hypothesized for the case of *Dipterocome*, *Berardia* could be misplaced in some chloroplast phylogenies by a long branch attraction effect in concatenated analysis due to its relict character without any known alive direct congener. Conclusively, *Berardia* and *Staelina* could be considered as separate new subtribes Berardiinae and Staeliniinae based on the new molecular evidences presented.

The *Onopordum* group is another lineage of former Carduinae with a consistent morpho-genetic historical division (Susanna and Garcia-Jacas, 2009). Except for the chloroplast *matK* phylogeny of Garcia-Jacas et al. (2002), probably with limited resolution power as mentioned above for other groups, the rest of molecular studies (Table 3) and our genomic phylogenies perfectly outlined the *Onopordum* assembly (BS = 100 and LPP = 1; Figs. 1, 2, and 3). Hence, it seems reasonably to assign the subtribal rank as Onopordinae.

One of most traditionally difficult alliances to delimit within the tribe, and one of the largest complexes of former Carduinae, is the *Carduus* group (ca. 500 species; Susanna et al., 2006). The main reason for its complicate delimitation was attributed by Susanna et al. (2006) to the fact that the group encompasses a great variety of life strategies; annual or biennial species (e.g. *Carduus*, *Galactites*, *Picnomon*, *Silybum*, or *Tyrimnus*) or perennials (as many *Cirsium*, *Cynara*, *Lamyropsis*, or *Ptilostemon*). Certainly, different generational times that affect mutation rates may hinder a proper phylogenetic comparison among species. The overview of previous phylogenetic studies reflects the difficulty of reconstructing the *Carduus* group as monophyletic, which was only achieved in three from the total eight Sanger-based phylogenies (Table 3). In contrast, we found a strong phylogenetic signal with Hyb-Seq data (BS = 92–100 and LPP = 1; Figs. 1, 2, and 3), which reinforces the morphological hypothesis of Susanna et al. (2006) and favor its proposal as subtribe Carduinae. Susanna and Garcia-Jacas (2009) suggested a posterior division of the *Carduus* group into *Cynara* and *Carduus-Cirsium* complexes, but our phylogenies did not show a strictly bifurcating divergence pattern for these species complexes (Figs. 1, 2, and 3), hindering to formalize their segregation at subtribal rank.

The *Arctium-Cousinia* and *Jurinea-Saussurea* complexes have been other of classically entangled groups of former Carduinae. Although both conform well defined assemblies by morphology (Susanna and Garcia-Jacas, 2009), there has not been any Sanger-based phylogeny able to correctly recover the morphological pattern as monophyletic lineages (Table 3). Only two isolated molecular studies, Häffner and Hellwig (1999) and Susanna et al. (2006), reported *Arctium-Cousinia* and *Jurinea-Saussurea* as monophyletic, respectively. In recent times, the application of Hyb-Seq technique and the COS loci set of Mandel et al. (2014) for Compositae, aided to shed light on phylogenetic connections of these four genera. Phylogenomic trees performed in Herrando-Moraira et al. (2018) showed a clear sister relationship between *Arctium-Cousinia* and *Jurinea-Saussurea*, which also formed two monophyletic supported clades. In agreement, on a wider taxonomic sampling of the tribe, we found the same monophyletic pattern for both complexes under concatenation and coalescence approaches with nuclear data (BS = 100 and LPP = 1; Fig. 1 and 3A). However, the chloroplast phylogenies analyzed under concatenation recovered an unsupported clade of *Saussurea-Cousinia-Arctium* (Fig. 2) with *Jurinea* as a separate lineage, hence leaving the complex *Jurinea-Saussurea* as a paraphyletic assembly (Fig. 2). Similarly, a morphologically incongruent grouping pattern has been also recovered in the Sanger-based chloroplast phylogenies of Barres et al. (2013), where the genus *Saussurea-Jurinea-Cousinia* were clustered together with support and *Arctium* was placed as sister clade.

Overall, it seems that even with massive Hyb-Seq data, a clear cyto-nuclear discordance exists between evolutionary histories inferred for *Arctium-Cousinia* and *Jurinea-Saussurea* groups. Two factors could explain such pattern of incongruent genealogies (Morales-Briones et al., 2018), one more related with tree inference methods (phylogenetic error), and the other with inherent evolutionary history of the genera (incomplete lineage sorting [ILS] or hybridization). Firstly, it is stated that using gene fragments with limited phylogenetic signal may derive in an increasing of both stochastic and systematic errors in reconstructed phylogenies (Lemmon et al., 2009, Lemmon and Lemmon, 2013; Carlsen et al., 2018). In agreement, if we compared our nuclear vs. chloroplast datasets, a large difference exists in terms of phylogenetic informativeness between both, for instance, in PI sites (ca. 38% vs. 5.5%; Table 4) or the

average branch length of trees (0.04 vs. 0.004). Chloroplast data from protein coding regions, in comparison with hundreds of nuclear loci, could be providing limited phylogenetic information, which may explain the differences in the resulting tree topologies.

Secondly, processes like ILS or reticulation events remain as possible alternatives to explain the nuclear-chloroplast conflicting topologies, as well as gene tree incongruences. The short internal branches observed at the base of genera in chloroplast tree from concatenation analysis (Fig. 2B) could match the signals left by ILS, i.e. a scenario of a rapid diversification of genera in a short timeframe (rapid bifurcating cladogenesis) or a simultaneous polychotomous divergence (hard polytomy). Past hybridization events could also have taken place as suggested by the concatenation chloroplast tree topology (Fig. 2). The chloroplast genome of ancestral *Saussurea* lineage could have been lost due to an organelle capture, thus, the currently observed is that retained from an ancestral hybridization with the *Arctium-Cousinia* lineage. Cyto-nuclear discordances due to introgressive hybridization have been suggested in other plant groups like *Ficus* through chloroplast NGS phylogenies (Bruun-Lund et al., 2017). But more explicit methods to address these processes should be performed (e.g. Folk et al., 2017; García et al., 2017; Mitchell et al., 2017; Morales-Briones et al., 2018; Knowles et al., 2018), in addition to a wider species sampling.

Taking into account all shortcomings of chloroplast data, it would be reasonable to give more weight to nuclear results, which were also in agreement with the morphological hypothesis, and accept two new subtribes: Arctiinae (comprising *Arctium-Cousinia*) and Saussureinae (*Saussurea-Jurinea*).

3.3. Phylogenetic relationships among subtribes

Undoubtedly, one of major improvements that we obtained using Hyb-Seq data is the high phylogenetic resolution recovered along the backbone of Cardueae tree, which is very useful to disentangle subtribal relationships (Fig. 4). Although in most recent Sanger-based phylogenies, 8 of the total 12 circumscribed subtribes were already monophyletically well defined (see Table 3; Barres et al., 2013), their interrelationships remained unresolved since to date. In the phylogenetic trees of Barres et al. (2013), those inferred under maximum parsimony were completely undefined for the subtribal nodes, and under Bayesian inference most of the subtribe branches were anchored in polytomic structures, especially in the nuclear ITS phylogeny (Fig. 4A). In contrast, our phylogenetic reconstructions showed a significantly supported bifurcating pattern (Figs. 4C, D, and E).

The subtribes Carlininae, Cardopatiinae, and Echinopsinae were placed as the first diverging lineages within tribe Cardueae near the root of the tree. Unfortunately, we did not find which is the sister subtribe to the rest with confidence, due to a cyto-nuclear discordance. In the nuclear tree inferred under concatenation, the Carlininae were sister to the rest (Fig. 4C); but individual gene trees show conflicting topologies at this part of the tree (Figs. 3 and 4E); and in the chloroplast tree inferred under concatenation the Cardopatiinae were sister to the rest (Fig. 4D). Successive sister clades recovered with all reconstruction methods occur in order: Dipterocominae, Xerantheminae, Berardiinae, Staehelininae, and Onopordinae. Note that, as mentioned above, Dipterocominae appeared nested with Xerantheminae and Berardiinae with Staehelininae in the nuclear and chloroplast concatenated analyses, respectively. The grouping pattern of the rest of subtribes was slightly different on nuclear and plastid phylogenies. In nuclear trees, Carduinae was placed as sister to Arctiinae-Saussureinae-Centaureinae clade (BS = 100 and LPP = 0.99), which was subsequently bifurcated in Arctiinae-Saussureinae (BS = 100 and LPP = 1) and Centaureinae clade (Figs. 4C, E). In the chloroplast tree, Carduinae-Arctiinae-Saussureinae-Centaureinae were clustered in a supported clade (BS = 100) that internally formed an unsupported group of Carduinae-Arctiinae-Saussureinae (BS = 63) with Centaureinae as sister to the rest (Fig. 4D).

Here, despite Hyb-Seq data offered bifurcating reconstructions, the short branch lengths recovered (Figs. 1B, 2B) may confirm the previous suspicion of a rapid subtribal lineage divergence, which would explain the traditional difficulty to resolve this part of the tree. The Phyparts analysis performed allowed us to go one step further and inspect the degree of gene tree discordance along nodes of Cardueae tree. Results of Phyparts indicated that most of gene tree topologies are in conflict with the species tree in internal subtribal nodes (great red proportion; Fig. 3), in contrast to shallow nodes of genera or species relationships that showed more congruent gene tree histories (more proportion of blue; Fig. 3). We also detected that a considerable proportion of gene trees were poorly supported and uninformative (gray color in Fig. 3 for gene tree branches with BS < 30). Considering all results, possible sources of gene tree discordance at the backbone could derive from: (1) insufficient phylogenetic signal in sampled loci (e.g. short loci length, high missing positions or taxa as a result of capture failure, or poorly variable loci; Villaverde et al., 2018); (2) extinction of critical lineages; (3) rapid splitting of major subtribal clades; or (4) gene flow among ancestral lineages or ILS effect (Morales-Briones et al., 2018). It is important to notice that although individual gene tree bipartitions were mainly in conflict at the subtribal level, the

coalescence species tree was, in general, not weakly supported (Fig. 3). Thus, this limitation is not necessary problematic to the aim of subtribal delimitation intended here.

3.4. Dating framework for the tribe Cardueae

Figure 5 shows the divergence time estimates of Cardueae based on the nuclear tree obtained with HybPiper and the concatenation approach (1055 COS loci). Median age values and 95% CI are presented for each node in Supplementary Table S3. Our dating analysis revealed that the tribe Cardueae could have originated between Late Eocene and Early Oligocene 34.12 Myr (29.97–40.04 95% CI). Generally, this time frame is in a middle range of the estimated age origin for the tribe in previous dated phylogenies that included several Cardueae members. On the one hand, the studies conducted by Kim et al. (2005) and Panero and Crozier (2016) reported a younger age of the tribe, 24–29 Myr and 22 Myr (11–33 95% CI), respectively. On the other hand, Barres et al. (2013) and Huang et al. (2016) found an older tribe median age with values of 40.15 Myr (35.80–44.26 95% CI) and ca. 42 Myr, respectively. Taxon sampling, calibration points, and molecular markers used could be responsible of dating differences among studies.

The subtribes originated in two different main time periods. The subtribes Carlininae, Cardopatiinae, Dipterocominae, Xerantheminae, and Berardiinae probably originated during the Oligocene (see Fig. 5 and Supplementary Table S3 for age of subtribes), whereas Echinopsinae, Staehelininae, Onopordinae, Carduinae, Arctiinae, Saussureinae, and Centaureinae originated during Middle-Late Miocene (Fig. 5 and Supplementary Table S3).

If we draw in parallel dating results obtained here and the ones obtained by Barres et al. (2013), which is the most comparable study in terms of taxon sampling, we can observe that tribe and subtribe age estimates were broadly older in the latter than those we found here (see Fig. 6). At tribal level, the different placement of the fossil *Raiguenrayun cura* (dated to 47.5 Myr, placed here to constrain subfamily Mutisioideae + subfamily Carduoideae; CP2 in Fig. 6) does not explain why we obtained a most recent tribal origin than Barres et al. (2013), who placed it in a more basal node as it is the crown node of Compositae family. Probably, the fact that we now have a more complete sampling at the basal part of the tree, including more representatives of subfamilies other than Carduoideae and more representatives of tribes within Carduoideae other than Cardueae, results in greater time distances between the origin of the family Compositae and the tribe Cardueae. At subtribal level, the greatest differences between non-overlapping age ranges were detected for subtribes Onopordinae and Centaureinae, for which the median age differed among studies in 9.32 Myr and 15.38 Myr, respectively. The different number and nature of molecular markers used would be one of main causes of the distinct dating results. We employed a set of 1055 low-copy nuclear loci biparentally inherited that potentially provides more sequence variation, while Barres et al. (2013) used four chloroplast regions that are usually maternally inherited and have low mutation rates, which would explain the older age values reported in their estimations. In addition, it should be also noted that for those nodes constrained with minimum ages based on fossil calibration points, we obtained a better fitting between minimum age assigned and final dating result recovered in comparison with Barres et al. (2013); for instance, our CP5 constrained at 6 Myr resulted in 5.91 Myr (5.21–7.01 95% CI), and in Barres et al. (2013) the CP used at the stem node of *Centaurea* constrained at 5 Myr resulted in 15.54 Myr (12.10–19.96 95% CI).

In a general overview of the dated phylogenetic tree it is noteworthy that CI interval limits are wider and more overlapped among nodes at the base of the tree than those at the tips (Fig. 5). Accordingly, lower precision and resolution was obtained for estimated ages of non-Cardueae and early diverging lineages of the Cardueae (e.g. Carlininae or Cardopatiinae). Certainly, the fact that the standard deviation of our most basal and secondary dated calibration point used (CP1) can be higher than the other fossil-based CPs may cause larger uncertainty at the tree base. Moreover, the number of species included as Cardueae outgroups in our study is relatively poor, particularly compared with Panero and Crozier (2016) who used a wide representation of Compositae subfamilies but few representatives for each tribe. The different taxon sampling used could be the reason for differences in the chronograms obtained, resulting the poorly sampled nodes with a more uncertain dating value. As Linder et al. (2005) suggested, taxon sampling is one of most influential factors in molecular clock dating analysis. For this reason, resulting node estimates at the tree base obtained here should be taken with caution, and undoubtedly, a wider sampling of other family members could improve lineages placement and dating resolution at the base of Cardueae.

3.5. Taxonomic proposal

Subtribe Carlininae Dumort. 1827.

Perennial herbs or shrubs, less often annual plants, monoecious or exceptionally dioecious (*Tugarinovia*). Leaves usually spiny, deeply pinnatisect, rarely entire. Capitula frequently subtended by pectinate leaf-like bracts, homogamous, rarely heterogamous with radiate florets (Fig. 7A). Inner involucre bracts often showy, radiant and coloured. Receptacle densely covered with large scales fused into a honeycombed structure, absent only in *Tugarinovia*. Anther filaments glabrous, appendages long and sericeous. Corolla and style very short. Achenes with parenchymatous pericarp, densely sericeous with twin hairs. Pappus of plumose bristles, often connate into stout scales, persistent or deciduous. Genera included in the subtribe: *Atractylodes* DC., *Atractylis* L., *Carlina* L., *Thevenotia* DC., and *Tugarinovia* Iljin. Geographic distribution (Fig. 8A): *Atractylis*, *Carlina*, and *Thevenotia* grow in Eurasia, North Africa, and the Irano-Turanian region; *Tugarinovia* is endemic to the deserts of Mongolia and NW China; *Atractylodes* is distributed in East Asia (China, Japan, and Korea).

Subtribe *Cardopatiinae* Less. 1832.

Spiny perennial or annual herbs. Leaves spiny-dentate or pinnatisect. Capitula few-flowered, homogamous, densely (*Cardopatum*) or loosely (*Cousiniopsis*) clustered in corymbs (Fig. 7B). Involucre bracts with spiny pectinate-fimbriate appendages. Anther filaments glabrous. Florets bright blue below, deep blue above; petals linear. Style very shortly bilobed. Achenes with parenchymatous pericarp, densely sericeous; pappus double, of two rings of short scales. Genera included in the subtribe: *Cardopatum* Juss. and *Cousiniopsis* Nevski. Geographic distribution (Fig. 8B): disjunct area; *Cardopatum* is endemic in the south of the Mediterranean Region, from Algeria to Greece and Turkey, and *Cousiniopsis* grows in the deserts of Middle Asia.

Subtribe *Echinopsinae* Cass. ex Dumort. 1829.

Perennial herbs, rarely annuals, usually very spiny. Capitula one-flowered, aggregated in spherical (rarely hemispherical) secondary inflorescences (Fig. 7C). Bracts spiny. Anther filaments glabrous, basal appendages short, lacinate. Achenes with parenchymatous pericarp, densely sericeous. Pappus of broad, short scales directly attached to the apical plate. This subtribe includes only the genus *Echinops* L. Geographic distribution (Fig. 8C): the genus grows in the Mediterranean and the Irano-Turanian regions, with the main center of speciation in Iran. It is one of the genera that extends more to the south, reaching tropical Africa.

Subtribe *Dipterocominae* Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Dipterocome* Fisch. & C. A. Mey., Index Seminum Horti Petropolitani 1: 26 1835.

Annual desert plants to 20 cm. Leaves entire, linear-spatulate. Capitula unflowered and unisexual, clustered in second-order synflorescences. Bracts of the synflorescence in several rows, lanceolate, with a membranous, shortly ciliate margin. Female capitula 6–8 in the periphery of the synflorescence; involucre bracts fused into a spiny-horned structure acresent with the achene; corollas shortly bilabiate. Male florets ca. 4 in the center of the synflorescence; bracts of the involucre fused into a basal sheath; corolla tubulose with very short, equal petals; anthers shortly caudate. Comprises only the monotypic genus *Dipterocome* Fisch. & C. A. Mey. Geographic distribution (Fig. 8D): Deserts in the East Mediterranean and Irano-Turanian regions: Middle East (Israel, Jordan, Lebanon and Syria), Transcaucasia (Armenia, Azerbaijan and Georgia) and Iran. We have interpreted the minute heads of *Dipterocome pusilla* as second-order inflorescences, formed by female and male unflowered capitula (Fig. 7D). The bracts of the involucre of the female florets are fused into a spiny-horned structure, and the achenes are enclosed in the arcuated structure at maturity (Fig. 4C).

Subtribe *Xerantheminae* Cass. ex Dumort. 1829.

Unarmed annual herbs (*Chardinia*, *Siebera*, *Xeranthemum*), rarely dwarf shrubs (*Amphoricarpos*), exceptionally rhizomatous perennials (*Shangwua*). Leaves always entire, velvety underneath. Capitula usually heterogamous (with the exception of *Shangwua*); central florets discoid, peripheral florets shortly bilabiate, fertile, bisexual or female. Receptacle with large scarious scales. Anther filaments glabrous, appendages short, lacinate. Corolla lobes very short. Achenes often dimorphic; pappus of long tapering or subulate scales, rarely reduced to a corona in *Chardinia*, sparsely sericeous with twin hairs (Fig. 7E). Comprises five genera: *Amphoricarpos* Vis., *Chardinia* Desf., *Shangwua* Yu J. Wang, Raab-Straube, Susanna & J. Quan Liu, *Siebera* J. Gay and *Xeranthemum* L.

Geographic distribution (Fig. 8E): extremely diverse and complicated. The genus *Xeranthemum* is a widespread in waste places in Eurasia, from Middle Asia to the Iberian Peninsula, more frequently in the Mediterranean region. Two other closely related genera, *Chardinia* and *Sieberta*, occupy similar habitats in the Irano-Turanian region, from the Tian Shan to Turkey. *Amphoricarpos* shows a disjunct distribution in the mountains of the East Mediterranean: two species grow in the Balkans, two more in Anatolia, and one in the Caucasus. *Shangwua* is endemic to the mountains of Middle Asia and the Himalayas, from Tadjikistan to China.

Subtribe Berardiinae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Berardia* Vill., Prosp. Hist. Pl. Dauphiné 27. 1779.

Acaulescent, unarmed perennial herb (Fig. 7F). Leaves rounded, entire or denticulate, densely woolly, with veins prominent beneath, white above. Capitula solitary, sessile, homogamous. Involucral bracts subulate, scarious, woolly, ending in a slender flattened point. Receptacle areolate. Florets yellowish or pinkish. Staminal connective very long, apiculate. Achenes oblong, glabrous, slightly sulcate. Pericarp straw-coloured. Pappus of scabrid cylindric bristles retrorsely twisted, directly attached to the apical plate. Comprises only the monotypic genus *Berardia* Vill.

Geographic distribution (Fig. 8F): *Berardia* is a relict genus endemic to the southern Alps (SE France and NW Italy).

Subtribe Staehelininae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Staehelina* L., Sp. Pl. 2: 840. 1753.

Unarmed dwarf shrubs or subshrubs. Leaves entire or dentate-pinnatifid, linear to obovate, dark green above, white-woolly beneath. Capitula corymbose or rarely solitary, homogamous. Involucral bracts ovate to lanceolate, mucronate, sometimes minutely hirsute. Receptacle with wide, basally connate scales. Florets whitish or pink-purple. Corolla lobes very long. Anther filaments glabrous; basal appendages very long, sericeous. Achenes linear-oblong, glabrous or sericeous, with minute apical coronula. Pappus of bristles basally connate into broader paleae, more or less divided apically into pinnulate fimbriae (into capillary hairs in *Staehelina dubia* L. and *S. baetica* DC.), always overtopping involucre, sometimes deciduous in a ring.

Comprises only the genus *Staehelina* L. (Fig. 7G).

Geographic distribution (Fig. 8G): stony places on limestone or rarely serpentines in the Mediterranean region, from the Iberian Peninsula to the limits between Turkey and Irak (*Staehelina kurdica*). Most diverse in the Eastern Mediterranean region (six species, in contrast with only two species in the Western Mediterranean).

Subtribe Onopordinae Garcia-Jacas & Susanna, **new subtribe**. Type of the subtribe: *Onopordum* L., Sp. Pl. 2: 827. 1753.

Robust biennial or perennial herbs, usually spiny-toothed, rarely weakly unarmed. Leaves lanceolate to oblong, variously dissected. Capitula cupuliform or broadly ovoid, homogamous, discoid, most often solitary or on long peduncles. Involucral bracts usually spiny, innermost sometimes scarious. Receptacular bracts very often absent; receptacle foveolate, margins of foveoles with short scales, rarely setose. Achenes obovoid or fusiform, surface usually transversely rugulose, wrinkled or foveolate, apical plate without caruncle. Insertion areole straight or lateral-abaxial. Pappus bristles in a few rows, differentiated or not, basally connate in a ring, always deciduous as a single piece (Fig. 7H).

Comprises seven genera: *Alfredia* Cass., *Ancathia* DC., *Synurus* Iljin, *Syreitchikovia* Pavlov, *Lamyropappus* Knorring & Tamamsch., *Olgaea* Iljin, *Onopordum* L., and *Xanthopappus* C. Winkl.

Geographic distribution (Fig. 8H): the genus *Onopordum* is native in the Mediterranean and Irano-Turanian regions, from the Iberian Peninsula to Middle Asia, and it has been introduced as a weed in USA (California), Argentina, Chile, South Africa, and Australia. The remaining genera have narrow distributions in Middle Asia, the Tian Shan, and west China, with the exception of *Synurus*, endemic in east China, Korea, and Japan.

Subtribe Carduinae Dumort. (1827).

Perennial, biennial or annual spiny herbs or subshrubs, rarely unarmed. Leaves mostly lanceolate to oblong, entire or variously dissected. Capitula globose or cupuliform, homogamous, discoid, very rarely peripheral florets sterile and radiant (*Galactites*). Involucral bracts usually spiny, innermost exappendiculate or with rudimentary appendages. Receptacle densely setose. Anther filaments papillose, rarely glabrous. Achenes obovoid-fusiform, laterally compressed, with insertion areole straight or lateral-abaxial; apical plate very often slanted, inclined adaxially, usually with caruncle. Pappus inserted on a

parenchymatous ring in the apical plate, with bristles in many weakly differentiated or undifferentiated rows, deciduous as a single piece (Fig. 7I).

Comprises nine genera: *Carduus* L., *Cirsium* Mill., *Cynara* L., *Galactites* Moench, *Lamyropsis* (Kharadze) Dittrich, *Notobasis* Cass., *Ptilostemon* Cass., *Silybum* Vaill., and *Tyrinnus* Cass.

Geographic distribution (Fig. 8I): Widespread in all Eurasia, especially in the Mediterranean, extending southwards to Tropical Africa, introduced elsewhere as very noxious weeds. One genus, *Cirsium*, native in America, from Canada to Guatemala.

Subtribe Saussureinae Garcia-Jacas & Susanna, new subtribe. Type of the subtribe: *Saussurea* DC., Ann. Mus. Natl. Hist. Nat. 16: 156. 1810.

Unarmed perennial herbs or subshrubs, very rarely annual herbs. Leaves entire or pinnatisect, often silver-white below and glabrous above, sometimes hirsute-scabrid. Capitula cylindrical or globose, often paniculate, homogamous discoid. polycarpic or rarely monocarpic, Anther filaments glabrous, rarely papillose. Achenes cylindrical, slightly obconical or obpyramidal, indistinctly ribbed to costate, surface smooth or transversally rugose, very rarely with spines or scales, with or without a crown; apical plate with a persisting style base. Pappus of very long (overtopping involucre bracts), showy, scabrid or plumose, strongly differentiated bristles, the inner ones basally connate in a ring, persistent or detachable as a single piece, outer bristles shorter and freely deciduous (Fig. 7J).

Comprises four genera: *Dolomiaea* DC., *Jurinea* Cass., *Polytaxis* Bunge, and *Saussurea* DC.

Geographic distribution (Fig. 8J): Native in Eurasia, *Jurinea* grows especially in the eastern Mediterranean and West Asia with three species reaching the mountains of the Iberian peninsula and Morocco; *Saussurea* shows its most important centre of speciation in the Himalayas and the Hengduan mountains (more than 300 species); two species in Western Europe (Alps and Pyrenees), six species in North America. *Dolomiaea* is circumscribed to the Himalayas and adjacent mountains. *Polytaxis* is limited to the Tian Shan in Middle Asia.

Subtribe Arctiinae Garcia-Jacas & Susanna, new subtribe. Type of the subtribe: *Arctium* L., Sp. Pl. 2: 816. 1753.

Biennial or perennial herbs, polycarpic or rarely monocarpic, very rarely annuals, spiny or less often unarmed. Leaves entire, lyrate or pinnatisect, usually with arachnoid pubescence. Capitula homogamous, cylindrical or globose, discoid. Outer and middle involucre bracts spiny or hooked, inner bracts ended in a scarious, unarmed or spiny appendage. Receptacle with long, twisted fimbriae. Anther filaments slightly papillose or glabrous. Achenes cylindric to narrowly obovoid, usually laterally compressed or 4-costate, rarely shortly winged, usually longitudinally inconspicuously ribbed and stripped, apex glabrous or coronulate; insertion areole lateral-axial; apical plate truncate, without caruncle. Pappus in three rows of freely deciduous undifferentiated bristles, rarely reduced (Fig. 7K).

Comprises two genera: *Arctium* L. and *Cousinia* Cass.

Geographic distribution (Fig. 8K): some species of *Arctium* are widespread in Eurasia, but most of the species of the genus are endemic to the mountains of Middle Asia, especially the Tian Shan. *Cousinia* is confined to the Irano-Turanian floristic region, where it has radiated explosively (600 species!).

Subtribe Centaureinae Dumort. (1827).

Perennial, biennial or annual unarmed herbs, shrubs or very rarely treelets, rarely spiny. Leaves oblong to linear, entire or variously dissected. Capitula ovoid, cupuliform or cylindric, often heterogamous with sterile radiant florets, less often homogamous, discoid. Involucre bracts usually ended in a diversely scarious, fimbriate, pectinate, spiny or unarmed appendage; innermost bracts always with a scarious appendage. Receptacle setaceous. Anther filaments glabrous or papillose. Achenes obovoid, laterally compressed, with sclerified pericarp, usually glabrous and smooth, less often hirsute, rugose, pitted, or ridged; insertion areole concave, lateral-axial, very rarely (*Crupina*) straight, often with an elaiosome; apical plate straight, without caruncle. Pappus inserted on a parenchymatous ring in the apical plate, biseriate, outer series formed by several rows of long, differently pinnulate bristles or rarely scales, basally connate in a ring or free, deciduous as a whole or separately, often persistent, inner series of some short or longer bristles, sometimes reduced; rarely pappus single or missing by abortion (Fig. 7L).

Comprises 32 genera: *Amberboa* Vaill., *Callicephalus* C. A. Mey. *Carduncellus* Adans., *Carthamus* L., *Centaurea*¹ L., *Centaurodendron* Johow, *Centaurothamnus* Wagenitz & Dittrich, *Cheirolophus* Cass., *Crocodylium* Vaill., *Crupina* (Pers.) DC., *Femeniasia* Susanna, *Goniocaulon* Cass., *Karvandarina* Rech. f., *Klasea* Cass., *Mantisalca* Cass., *Myopordon* Boiss., *Ochrocephala* Dittrich, *Oligochaeta* (DC.) K.

¹ Including *Cyanus* Mill., a genus that is sometimes segregated from *Centaurea*. Our analysis (Fig. 1) confirms that it should be considered a subgenus of *Centaurea*.

Koch, *Phalacrachena* Iljin, *Phonus* Hill, *Plagiobasis* Schrenk, *Plectocephalus* D. Don, *Psephellus* Cass., *Rhaponticoides* Vaill., *Rhaponticum* Vaill., *Russowia* C. Winkl., *Schischkinia* Iljin, *Serratula* L., *Stizolophus* Cass., *Tricholepis* DC., *Volutaria* Cass., and *Zoegea* L.

Geographic distribution (Fig. 8L): mainly Mediterranean and Irano-Turanian regions, with some species of *Centaurea* extended widely in north Eurasia as far as Scotland, and the easternmost representatives reaching Middle Asia; exceptionally, *Tricholepis* has gone beyond the mountains of Middle Asia and reaches Burma; one species of *Rhaponticum* grows in Australia. In the south of the area, a few species grow in subtropical and tropical Africa, from Senegal to Somalia and southwards to Zimbabwe. *Plectocephalus* has a disjunct distribution in Ethiopia, U.S.A., and South America (Brazil, Argentina and Chile). Some relict genera grow far away from the center of the distribution of the subtribe, namely *Centaurodendron* in the Juan Fernández archipelago, *Centaurothamnus* in Yemen, *Goniocaulon* in India, and *Ochrocephala* in Etiopia and Sudan; all of them but *Centaurodendron* are monotypic. Some noxious weeds are widespread in all the globe, especially in regions with a Mediterranean-type climate (Australia, California, Chile, and South Africa): *Carthamus lanatus*, *Centaurea diffusa*, *C. solstitialis*, *C. stoebe*, *Rhaponticum repens*, and *Volutaria tubuliflora*.

Key to subtribes of tribe Cardueae

1. Achenes enclosed in an arcuate, spiny-toothed diaspore; heads minute, unisexual, clustered in the axilles of leaf verticils Dipterocominae
– Achenes free, not enclosed in a spiny-toothed diaspore; heads bisexual, not clustered in the axilles of leaf verticils 2
2. Capitula one-flowered, clustered in second-order spherical or hemispherical compound synflorescences Echinopsinae
– Capitula not one-flowered, sometimes clustered in flat-topped corymbs but never forming second-order heads 3
3. Acaulescent perennials with gray-green, entire leaves; achenes with the pappus bristles retrorsely twisted Berardiinae
– Caulescent or acaulescent annuals or perennials; achenes with pappus of straight bristles 4
4. Leaves always entire. Achenes with pappus of 5-15 subulate, dentate or plumose scales, rarely reduced to a short corona Xerantheminae
– Leaves entire or divided. Achenes with pappus of many bristles 5
5. Capitula usually subtended by pectinate-pinnatisect leaf-like bracts. Corolla lobes 1–3 mm long. Pappus of long, plumose bristles usually connate at the base into broader, robust scales Carlininae
– Capitula subtended by entire or dentate bracts or on leafless pedicels. Corolla lobes longer than 3 mm. Pappus of free bristles or scales not connate at the base into broader scales 6
6. Capitula with 8-12 florets, rarely many-flowered. Pappus of two equal rows of short lanceolate scales. Corollas bright blue below, deep blue above, with linear lobes Cardopatiinae
Capitula many-flowered. Pappus of one or several rows of pinnulate or plumose long scales or bristles. Corollas not bright blue below and deep blue above; lobes not linear 7
7. Achenes rugulose, pitted or faintly velvety, never smooth, without apical caruncle or basal elaiosome 8
– Achenes usually smooth or rarely ridged, either with apical caruncle or basal elaiosome 9
8. Receptacle fimbriate with long, twisted fimbriellae. Achenes longitudinally stripped; pappus bristles individually deciduous Arctiinae
– Receptacle usually naked and foveolate, rarely with straight bristles. Achenes transversally rugulose; pappus bristles basally connate

in a ring, detachable as a single piece	Onopordinae
9. Plants unarmed. Leaves always white-woolly beneath. Pappus very long, often overtopping involucre bracts at anthesis	10
– Plants unarmed or spiny. Leaves woolly or glabrous below. Pappus not overtopping involucre bracts	11
10. Shrubs or subshrubs. Corolla lobes very long, narrowly triangular. Style lobes very short, straight. Achenes linear, faintly sulcate; pappus in one row	Staehelininae
– Perennial herbs, rarely annuals. Corolla lobes short, broadly triangular. Style lobes long, reflexed. Achenes cylindrical or slightly obconical or obpyramidal, smooth, echinate or rugulose; pappus in several rows	Saussureinae
11. Insertion areole of the achenes usually straight, basal or basal-axial. Apical plate of the achenes slanted, with elaiosome (except. <i>Notobasis</i>). Pappus undifferentiated, deciduous as a whole	Carduinae
– Insertion areole of the achenes usually concave, lateral-abaxial, often with an elaiosome. Apical plate straight, without caruncle. Pappus usually differentiated in two rows, persistent or rarely deciduous as a single piece, rarely absent by abortion	Centaureinae

Acknowledgments

Authors thank the herbaria that provided material for the study: BC, DUSH, E, ERE, FRU, GDA, H, LE, MJG, TK, W, and A. Pavlenko for kindly supplying the photograph of *Dipterocome pusilla*. Financial support from the Ministerio de Ciencia e Innovación (Project CGL2015-66703-P and Ph.D. grant to Sonia Herrando-Moraira) and the Catalan government (“Ajuts a grups consolidats” 2017-SGR1116) is gratefully acknowledged.

Supplementary material and Appendices

All alignments and tree files for each dataset are deposited in Mendeley Data repository (<http://dx.doi.org/xxxxx>). Supplementary data associated with this article can be found, in the online version, at <https://doi.org/xxxxx>

References

- Anderberg, A., Ghahremaninejad, F., Källersjö, M., 2007. The enigmatic genus *Dipterocome*. *Compositae Newslett.* 45, 23–36.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin V.M., Nikolenko, S.I., Pham, S., Prjibelski A.D., Pyshkin, A.V., Sirotkin A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pyshkin, A.V., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Barreda, V.D., Palazzesi, L., Katinas, L., Crisci, J.V., Tellería, M.C., Bremer, K., Passala, M.G., Bechis, F., Corsolini, R., 2012. An extinct Eocene taxon of the daisy family (Asteraceae): evolutionary, ecological and biogeographical implications. *Ann. Bot., London*, 109, 127–134. <https://doi.org/10.1093/aob/mcr240>
- Barres, L., Sanmartín, I., Anderson, C.L., Susanna, A., Buerki, S., Galbany-Casals, M., Vilatersana, R., 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *Am. J. Bot.* 100, 867–882. <https://doi.org/10.3732/ajb.1200058>
- Bentham, G., 1873. *Compositae*. In: Bentham, G., Hooker, J.D. (Eds.), *Genera plantarum*, vol. 2. Spottiswoode & Co., London, UK, pp. 163–533.
- Boissier, P.E., 1849. *Diagnoses plantarum orientalium novarum*, series 1, vol. 11. Ducloux, Paris, France.

- Bolger, A.M., Lohse, M., Usadel, B.**, 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Borowiec, M.L.**, 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4, e1660. <https://doi.org/10.7717/peerj.1660>
- Bremer, K.**, 1987. Tribal interrelationships of the Asteraceae. *Cladistics* 3, 210–253. <https://doi.org/10.1111/j.1096-0031.1987.tb00509.x>
- Bremer, K.**, 1994. *Compositae: Cladistics and classification*. Timber Press, Portland, Oregon, USA.
- Bruun-Lund, S., Clement, W. L., Kjellberg, F., Rønsted, N.**, 2017. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). *Mol. Phylogenet. Evol.* 109, 93–104. <https://doi.org/10.1016/j.ympev.2016.12.031>
- Candolle, A.P. de.**, 1838. *Prodromus systematis naturalis regni vegetabilis*, vol. 7. Treuttel & Würtz, Paris, France.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T.**, 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Carlsen, M.M., Fér, T., Schmickl, R., Leong-Škorničková, J., Newman, M., Kress, W.J.**, 2018. Resolving the rapid plant radiation of early diverging lineages in the tropical Zingiberales: pushing the limits of genomic data. *Mol. Phylogenet. Evol.* 128, 55–68. <https://doi.org/10.1016/j.ympev.2018.07.020>
- Cassini, H.**, 1819. [Different articles]. *Dictionnaire de Sciences Naturelles*. Paris. In: King, R., Dawson, H.W. (Eds., 1975), *Cassini on Compositae*. Oriole, New York, New York, USA.
- Castresana, J.**, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Chevreur, B., Wetter, T., Suhai, S.**, 1999. Genome sequence assembly using trace signals and additional sequence information. In: *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, vol. 99, pp. 45–56.
- Dayrat, B.**, 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85, 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- Dittrich, M.**, 1977. Cynareae—systematic review. In: Heywood, V.H., Harborne, J.B., Turner, B.L. (Eds.), *The Biology and Chemistry of the Compositae*. Academic Press, London–New York–San Francisco, UK and USA, pp. 999–1015.
- Dittrich, M.**, 1996a. Bemerkungen zur Tribuszugehörigkeit von *Berardia*. *Ann. Naturhist. Mus. Wien* 99 (B, suppl.), pp. 329–342.
- Dittrich, M.**, 1996b. Die Bedeutung morphologischer und anatomischer Achänen—Merkmale für die Systematik der Tribus Echinopeae Cass. und Carlineae Cass. *Boissiera* 51: 9–102.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A.**, 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. <https://doi.org/10.1093/molbev/mss075>
- Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leache, A.D., Liu, L., Davis, C.C.**, 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Mol. Phylogenet. Evol.* 94, 447–462. <https://doi.org/10.1016/j.ympev.2015.10.027>
- Faircloth, B.C.**, 2015. PHYLUCe is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32, 786–788. <https://doi.org/10.1093/bioinformatics/btv646>
- Fischer, F.E.L. von, Meyer, C.A. von.**, 1835. *Index Seminum Hortus Petropolitanus* 1. St. Petersburg, Russia.
- Folk, R.A., Mandel, J.R., Freudenstein, J.V.**, 2015. A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: A phylogenomic example from *Heuchera* (Saxifragaceae). *Appl. Plant Sci.* 3, 1500039. <https://doi.org/10.3732/apps.1500039>
- Folk, R.A., Mandel, J.R., Freudenstein, J.V.**, 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Syst. Biol.* 66, 320–337. <https://doi.org/10.1093/sysbio/syw083>
- Funk, V. A., Susanna, A., Stuessy, T.F., Robinson, H.**, 2009. Classification of Compositae. In: Funk, V.A., Susanna, A., Stuessy, T.F., Bayer, R.J. (Eds.), 2009. *Systematics, evolution, and biogeography of Compositae*. International Association for Plant Taxonomy, Vienna, pp. 171–189.
- García, N., Folk, R.A., Meerow, A.W., Chamala, S., Gitzendanner, M.A., de Oliveira, R.S., Soltis, D.E., Soltis, P.S.**, 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Mol. Phylogenet. Evol.* 111, 231–247. <https://doi.org/10.1016/j.ympev.2017.04.003>

- Garcia-Jacas, N., Garnatje, T., Susanna, A., Vilatersana, R., 2002. Tribal and subtribal delimitation and phylogeny of the Cardueae (Asteraceae): a combined nuclear and chloroplast DNA analysis. *Mol. Phylogenet. Evol.* 22, 51–64. <https://doi.org/10.1006/mpev.2001.1038>
- Gardner, A.G., Sessa, E.B., Michener, P., Johnson, E., Shepherd, K.A., Howarth, D.G., Jabaily, R.S., 2016. Utilizing next-generation sequencing to resolve the backbone of the Core Goodeniaceae and inform future taxonomic and floral form studies. *Molec. Phylogenet. Evol.* 94, 605–617. <https://doi.org/10.1016/j.ympev.2015.10.003>
- Gouy, M., Guindon, S., Gascuel, O., 2009. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. <https://doi.org/10.1093/molbev/msp259>
- Häffner, E., Hellwig, F.H., 1999. Phylogeny of the tribe Cardueae (Compositae) with emphasis on the subtribe Carduinae: an analysis based on ITS sequence data. *Willdenowia* 29, 27–39. <https://doi.org/10.3372/wi.29.2902>
- Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41, e129. <https://doi.org/10.1093/nar/gkt371>
- Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. PhD Thesis, Pennsylvania State University, Pennsylvania.
- Harrison, N., Kidner, C.A., 2011. Next-generation sequencing and systematics: what can a billion base pairs of DNA sequence data do for you? *Taxon* 60, 1552–1566.
- Heckenhauer, J., Samuel, R., Ashton, P.S., Salim, K.A., Paun, O., 2018. Phylogenomics resolves evolutionary relationships and provides insights into floral evolution in the tribe Shoreaeae (Dipterocarpaceae). *Mol. Phylogenet. Evol.* 127, 1–13. <https://doi.org/10.1016/j.ympev.2018.05.010>
- Heilbuth, J.C., 2000. Lower species richness in dioecious clades. *Am. Nat.* 156, 221–41. <https://doi.org/10.1086/303389>
- Herrando-Moraira, S., The Cardueae Radiations Group (Calleja, J.A., Carnicero-Campmany, Fujikawa, K., Galbany-Casals, M., Garcia-Jacas, N., Im, H.-T., Kim, S.-C., Liu, J.-Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Mehregan, I., Montes-Moreno, N., Pyak, E., Roquet, C., Sáez, L., Sennikov, A., Susanna, A., Vilatersana, R.), 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). *Mol. Phylogenet. Evol.* 128, 69–87. <https://doi.org/10.1016/j.ympev.2018.07.012>
- Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42, 182–192. <https://doi.org/10.1093/sysbio/42.2.182>
- Hoffmann, O., 1894. Compositae. In: Engler, A., Prantl, K. (Eds.), *Die natürlichen Pflanzenfamilien*, vol. 5, part 4. Wilhelm Engelmann, Leipzig, Germany, pp. 24–333.
- Huang, C.H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., Ma, H., 2016. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol. Biol. Evol.* 33, 2820–2835. <https://doi.org/10.1093/molbev/msw157>
- Ivanov, D., Ashraf, A.R., Utescher, T., Mosbrugger, V., Slavomirova, E., 2007. Late Miocene vegetation and climate of the Balkan region, palynology of the Beli Breg Coal Basin sediments. *Geol. Carpath.* 58, 367–381.
- Jansen, R.K., Holsinger, K.E., Michaels, H.J., Palmer, J.D., 1990. Phylogenetic analysis of chloroplast DNA restriction site data at higher taxonomic levels: an example from the Asteraceae. *Evolution* 44, 2089–2105. <https://doi.org/10.1111/j.1558-5646.1990.tb04314.x>
- Jansen, R.K., Michaels, H.J., Palmer, J.D., 1991. Phylogeny and character evolution in the Asteraceae based on chloroplast DNA restriction site mapping. *Syst. Bot.* 16, 98–115. <https://doi.org/10.2307/2418976>
- Jaubert, H.F., Spach, E., 1850. *Illustrationes plantarum orientalium*, vol. 3. Roret, Paris, France.
- Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C., Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* 4, 1600016. <https://doi.org/10.3732/apps.1600016>
- Jones, K.E., Fér, T., Schmickl, R., Dikow, R.B., Funk, V.J., Herrando-Moraira, S., Johnston, P.R., Siniscalchi, C.M., Susanna, A., Slovák, M., Thapa, R., Watson, L., Mandel, J.R., in preparation. An empirical assessment of a single family-wide target capture locus set at multiple evolutionary timescales within the largest family of flowering plants.
- Karis, P.O., Källersjö, M., Bremer, K., 1992. Phylogenetic analysis of the Cichorioideae (Asteraceae), with emphasis on the Mutisieae. *Ann. Mo. Bot. Gard.* 79, 416–427. <https://doi.org/10.2307/2399778>

- Katoh, K., Standley, D.M.**, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kay, K.M., Voelckel, C., Yang, J.Y., Hufford, K.M., Kaska, D.D., Hodges, S.A.**, 2006. Floral characters and species diversification. In: Harder, L.D., Barrett, S.C.H. (Eds.), *Ecology and evolution of flowers*. Oxford University Press, Oxford, UK, pp. 311–325.
- Kim, K.-J., Choi, K.-S., Jansen, R.K.**, 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol. Biol. Ecol.* 22, 1783–1792. <https://doi.org/10.1093/molbev/msi174>
- Kim, K.J., Jansen, R.K., Wallace, R.S., Michaels, H.J., Palmer, J.D.**, 1992. Phylogenetic implications of rbcL sequence variation in the Asteraceae. *Ann. Mo. Bot. Gard.* 428–445. <https://doi.org/10.2307/2399779>
- Knowles, L.L., Huang, H., Sukumaran, J., Smith, S.A.**, 2018. A matter of phylogenetic scale: Distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. *Am. J. Bot.* 105, 376–384. <https://doi.org/10.1002/ajb2.1064>
- Kubatko, L.S., Degnan, J.H.**, 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56, 17–24. <https://doi.org/10.1080/10635150601146041>
- Kück, P., Longo, G.C.**, 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.* 11, 81. <https://doi.org/10.1186/s12983-014-0081-x>
- Lagomarsino, L.P., Condamine, F.L., Antonelli, A., Mulch, A., Davis, C.C.**, 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytol.* 210, 1430–1442. <https://doi.org/10.1111/nph.13920>
- Lane, A.K., Augustin, M.M., Ayyampalayam, S., Plant, A., Gleissberg, S., Di Stilio, V.S., Depamphilis C.W., Wong, G.K.-S., Kutchan, T.M., Leebens-Mack, J.H.**, 2018. Phylogenomic analysis of Ranunculales resolves branching events across the order. *Bot. J. Linn. Soc.* 187, 157–166. <https://doi.org/10.1093/botlinnean/boy015>
- Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M.**, 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* 58, 130–145. <https://doi.org/10.1093/sysbio/syp017>
- Lemmon, E.M., Lemmon, A.R.**, 2013. High-throughput genomic data in systematics and phylogenetics. *Ann. Rev. Ecol. Evol. Syst.* 44, 99–121. <https://doi.org/10.1146/annurev-ecolsys-110512-135822>
- Linder, H.P., Hardy, C.R., Rutschmann, F.**, 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Mol. Phylogenet. Evol.* 35, 569–582. <https://doi.org/10.1016/j.ympev.2004.12.006>
- Liu, L., Xi, Z., Davis, C.C.**, 2015. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. *Mol. Biol. Evol.* 32, 791–805. <https://doi.org/10.1093/molbev/msu331>
- López-Vinyallonga, S., Mehregan, I., Garcia-Jacas, N., Tscherneva, O., Susanna, A., Kadereit, J.K.**, 2009. Phylogeny and evolution of the *Arctium-Cousinia* complex (Compositae, Cardueae-Carduinae). *Taxon* 58, 153–171.
- Mai, D.H.**, 1995. *Tertiäre Vegetationsgeschichte Europas: Methoden und Ergebnisse*. Gustav Fischer Verlag, Jena, Germany.
- Mai, D.H.**, 2001. Die mittelmiozänen und obermiozänen Floren aus der Meuroer und Raunoer Folge in der Lausitz, Teil II, Dicotyledonen. *Palaeontogr. Abt. B* 257, 35–174.
- Mandel, J.R., Barker, M.S., Bayer, R.J., Dikow, R.B., Gao, T., Jones, K.E., Keely, S., Kilian, N., Ma, H., Siniscalchi, C.M., Susanna, A., Thapa, R., Watson, L., Funk, V.A.**, 2017. The Compositae Tree of Life in the age of phylogenomics. *J. System. Evol.* 55, 405–410. <https://doi.org/10.1111/jse.12265>
- Mandel, J.R., Dikow, R.B., Funk, V.A.**, 2015. Using phylogenomics to resolve mega-families: an example from Compositae. *J. System. Evol.* 53, 391–402. <https://doi.org/10.1111/jse.12167>
- Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmores, R.W., Rieseberg, L.H., Burke, J.M.**, 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the Compositae. *Appl. Plant Sci.* 2, 1300085. <https://doi.org/10.3732/apps.1300085>
- Miller, M.A., Pfeiffer, W., Schwartz, T.**, 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 14 Nov. 2010. New Orleans, pp. 1–8. <https://doi.org/10.1109/GCE.2010.5676129>

- Mitchell, N., Lewis, P.O., Lemmon, E.M., Lemmon, A.R., Holsinger, K.E., 2017. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. *Am. J. Bot.* 104, 102–115. <https://doi.org/10.3732/ajb.1600227>
- Morales-Briones, D.F., Liston, A., Tank, D.C., 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytol.* 218, 1668–1684. <https://doi.org/10.1111/nph.15099>
- Nicholls, J.A., Pennington, R.T., Koenen, E.J., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N., Kidner, C.A., 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Front. Plant Sci.* 6, 710. <https://doi.org/10.3389/fpls.2015.00710>.
- Ortiz, S., Bonifacio, J.M., Crisci, J.V., Funk, V.A., Hansen, H.V., Hind, D.J.N., Katinas, L., Roque, N., Sancho, G., Susanna, A., Tellería, M.C., 2009. The basal grade of the Compositae: Mutisieae (sensu Cabrera) and Carduoideae. In: Funk, V.A., Susanna, A., Stuessy, T.F., Bayer, R.J. (Eds.), *Systematics, evolution, and biogeography of Compositae*. International Association for Plant Taxonomy, Vienna, Austria, pp. 193–213.
- Panero, J.L., Crozier, B.S., 2016. Macroevolutionary dynamics in the early diversification of Asteraceae. *Mol. Phylogenet. Evol.* 99, 116–132. <https://doi.org/10.1016/j.ympev.2016.03.007>
- Panero, J.L., Funk, V.A., 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Mol. Phylogenet. Evol.* 47, 757–782.
- Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7, 84. <https://doi.org/10.1186/1741-7007-7-84>
- Petit, D.P., 1988. Le genre *Echinops* L. (Compositae, Cardueae). 1. Position phylétique et interprétation de l'incapitulescence. *Candollea* 43, 467–481.
- Petit, D.P., 1997. Generic interrelationships of the Cardueae (Compositae): a cladistic analysis of morphological data. *Plant Syst. Evol.* 207, 173–203. <https://doi.org/10.1007/BF00984388>
- Petit, D.P., Mathez, J., Qaid, A., 1996. Early differentiation of the Cardueae sensu lato: Morphology and pollen. In: Hind, D.J.N., Beentje, H.J. (Eds.), *Compositae: Systematics. Proceedings of the International Compositae Conference, Kew 1994*, pp. 79–93.
- Popescu, S.M., 2002. Repetitive changes in Early Pliocene vegetation revealed by high-resolution pollen analysis: Revised cyclostratigraphy of southwestern Romania. *Rev. Palaeobot. Palynol.* 120, 181–202. [https://doi.org/10.1016/S0034-6667\(01\)00142-7](https://doi.org/10.1016/S0034-6667(01)00142-7)
- Rambaut, A. 2016. FigTree ver. 1.4.3. Department of Zoology, University of Oxford, Oxford. Available at <http://tree.bio.ed.ac.uk/software/figtree/>
- Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J., 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PloS One* 6: e22594. <https://doi.org/10.1371/journal.pone.0022594>
- Renner, S.S., Ricklefs, R.E., 1995. Dioecy and its correlates in the flowering plants. *Am. J. Bot.* 82, 596–606. <https://doi.org/10.1002/j.1537-2197.1995.tb11504.x>
- Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654–1668. <https://doi.org/10.1093/molbev/msw079>
- Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H., 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. *Ann. Rev. Entomol.* 55, 421–438. <https://doi.org/10.1146/annurev-ento-112408-085432>
- Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: Species TRee Analysis Web server. *Nucleic Acids Res.* 41 (Web Server issue), W238–W241.
- Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Aust. Syst. Bot.* 17: 145–170.
- Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* 85, 1301–1315.
- Smith, S.A., Moore, M.J., Brown, J.W., Yang, Y., 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol. Biol.* 15, 150. <https://doi.org/10.1186/s12862-015-0423-0>
- Smith, S.A., O'Meara, B.C., 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>

- Stamatakis, A.**, 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
- Stamatakis, A.**, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Stoughton, T.R., Kriebel, R., Jolles, D.D., O'Quinn, R.L.**, 2018. Next-generation lineage discovery: A case study of tuberous *Claytonia* L. *Am. J. Bot.* 105, 536–548. <https://doi.org/10.1002/ajb2.1061>
- Stubbs, R.L., Folk, R.A., Xiang, C.L., Soltis, D.E., Cellinese, N.**, 2018. Pseudo-parallel patterns of disjunctions in an Arctic-alpine plant lineage. *Mol. Phylogenet Evol.* 123, 88–100. <https://doi.org/10.1016/j.ympev.2018.02.016>
- Susanna, A.**, 1987. *Femeniasia, novus genus Carduearum*. *Collectanea Botanica* 17, 83–88. <https://doi.org/10.3989/collectbot.1988.v17.158>
- Susanna, A., Garcia-Jacas, N.**, 2009. Cardueae (Carduoideae). In: Funk, V.A., Susanna, A., Stuessy, T.F., Bayer, R.J. (Eds.), *Systematics, evolution, and biogeography of Compositae*. IAPT, Vienna, Austria. pp. 293–313.
- Susanna, A., Garcia-Jacas, N., Hidalgo, O., Vilatersana, R. Garnatje, T.**, 2006. The Cardueae (Compositae) revisited: Insights from ITS, trnL-trnF, and matK nuclear and chloroplast DNA analysis. *Ann. Mo. Bot. Gard.* 93, 150–171. [https://doi.org/10.3417/0026-6493\(2006\)93\[150:TCCRIF\]2.0.CO;2](https://doi.org/10.3417/0026-6493(2006)93[150:TCCRIF]2.0.CO;2)
- Susanna, A., Garcia-Jacas, N., Soltis, D.E., Soltis, P.S.**, 1995. Phylogenetic relationships in tribe Cardueae (Asteraceae) based on ITS sequences. *Am. J. Bot.* 82, 1056–1068. <https://doi.org/10.1002/j.1537-2197.1995.tb11571.x>
- Susanna, A., Garcia-Jacas, N.**, 2007. Tribe Cardueae. In: Kadereit, J.W., Jeffrey, C. (Eds.), *Flowering Plants. Eudicots. Asterales*, vol. 8. In: Kubitzki, K. (Ed.), *The families and genera of vascular plants*. Springer Verlag, Berlin, Germany, pp. 123–146.
- Susanna, A., Garnatje, T., Garcia-Jacas, N., Vilatersana, R.**, 2002. On the correct subtribal placement of the genera *Syreitschikovia* and *Nikitinia* (Asteraceae, Cardueae): Carduinae or Centaureinae? *Bot. J. Linn. Soc.* 140, 313–319. <https://doi.org/10.1046/j.1095-8339.2002.00104.x>
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., Greiner, S.**, 2017. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45(W1), W6–W11. <https://doi.org/10.1093/nar/gkx391>
- Vamosi, J.C., Otto, S.P., Barrett, S.C.**, 2003. Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. *J. Evol. Biol.* 16, 1006–1018. <https://doi.org/10.1046/j.1420-9101.2003.00559.x>
- Viljoen, E., Odeny, D.A., Coetzee, M.P., Berger, D.K., Rees, D.J.**, 2018. Application of chloroplast phylogenomics to resolve species relationships within the plant genus *Amaranthus*. *J. Mol. Evol.* 86, 216–239. <https://doi.org/10.1007/s00239-018-9837-9>
- Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M.G., Gardner, E.M., Wickett, N.J., Molero, J., Riina, R., Sanmartín, I.**, 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytol.* 220, 636–650. <https://doi.org/10.1111/nph.15312>
- Wagenitz, G.**, 1955. Pollenmorphologie und Systematik in der Gattung *Centaurea* L. s.l. *Flora* 142, 213–279.
- Wagenitz, G.**, 1958. Die Gattung *Myopordon* Boiss. (Compositae - Cynareae). *Ber. Deutsch. Bot. Ges.* 71, 271–277.
- Wagenitz, G.**, 1976. Systematics and phylogeny of the Compositae (Asteraceae). *Plant Syst. Evol.* 125, 29–46.
- Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A.** 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *App. Plant Sci.* 2, 1400042. <https://doi.org/10.3732/apps.1400042>
- Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S.**, 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19, 153. <https://doi.org/10.1186/s12859-018-2129-y>

TABLES

Table 1. Historical outline of the classification of the tribe Cardueae.

Classification	Taxonomic treatment references
Tribe Echinopeae Tribe Carlineae Tribe Cardueae Subtribe Carduinae Subtribe Centaureinae	Cassini (1819) , Dittrich (1977)
Tribe Cardueae Subtribe Echinopsinae Subtribe Carlininae Subtribe Carduinae Subtribe Centaureinae	Bentham (1873) , Hoffmann (1894) , Bremer (1994)
Tribe Echinopeae Tribe Cardueae Subtribe Carlininae Subtribe Carduinae-Centaureinae	Wagenitz (1976) , Petit (1988, 1997)
Tribe Cardueae Subtribe Cardopatiinae Subtribe Carlininae Subtribe Echinopsinae Subtribe Carduinae Subtribe Centaureinae	Susanna et al. (2006) , Susanna and Garcia-Jacas (2007, 2009)
Tribe Cardueae Subtribe Carlininae Subtribe Cardopatiinae Subtribe Echinopsinae Subtribe Dipterochiniinae Subtribe Xerantheminae Subtribe Berardiinae Subtribe Staehelininae Subtribe Onopordiinae Subtribe Carduinae Subtribe Arctiinae Subtribe Saussureinae	New treatment proposed here

Subtribe Centaureinae	
-----------------------	--

Table 2. Some cases of genera with historical conflicts of adscription to subtribes (or even tribes) of the tribe Cardueae resolved with phylogenies based on Sanger sequencing data. The tribes and treatment references in bold represent the most recent accepted taxonomic proposal and their corresponding source, respectively. The last column indicates the taxonomic treatment proposed in the present study for the traditional taxonomic conflict cases based on new subtribal delimitation. **Xeranthemum* group, including *Amphoricarpos*, *Chardinia*, and *Siebera*.

Genus	Taxonomic rank of conflict	Included in	References of treatments	Taxonomic treatment proposed here
<i>Berardia</i>	tribal	Mutisiae Cardueae (Carduinae)	Bremer (1994), Dittrich (1996a) Garcia-Jacas et al. (2002)	Cardueae (Berardiinae)
<i>Dipterocome</i>	tribal	Cichorieae Calenduleae Cardueae (Carduinae)	Boissier (1849) Bentham (1873), Hoffmann (1894) Susanna and Garcia-Jacas (2009)	Cardueae (Dipterocominae)
<i>Cardopatum</i>	subtribal	Carlininae Echinopsidinae Cardopatiinae	Dittrich (1977, 1996b), Bremer (1994), Petit (1997) Susanna et al. (2006)	Cardopatiinae
<i>Cousiniopsis</i>	subtribal	Carlininae Echinopsidinae Cardopatiinae	Dittrich (1977, 1996b), Bremer (1994), Petit (1997) Susanna et al. (2006)	Cardopatiinae
<i>Femeniasia</i>	subtribal	Carduinae Centaureinae	Susanna (1987) Susanna et al. (1995)	Centaureinae
<i>Myopordon</i>	subtribal	Carduinae Centaureinae	Wagenitz (1958), Dittrich (1977) Susanna et al. (2006)	Centaureinae
<i>Nikitinia</i>	subtribal	Carduinae Centaureinae	Dittrich (1977), Bremer (1994) Susanna et al. (2002, 2006)	Centaureinae
<i>Staelina</i>	subtribal	Carlininae Carduinae	Bentham (1873), Hoffmann (1894), Dittrich (1977, 1996b), Bremer (1994) Petit (1997), Susanna et al. (2006)	Staeliniinae
<i>Synurus</i>	subtribal	Centaureinae Cardueae	Dittrich (1977), Bremer (1994) Häffner and Hellwig (1999)	Onopordiinae
<i>Syreitschikovia</i>	subtribal	Centaureinae Carduinae	Dittrich (1977), Bremer (1994) Susanna et al. (2002)	Onopordiinae

<i>Xeranthemum</i> (all the group)*	subtribal	Carlininae Carduinae	Dittrich (1977, 1996b), Bremer (1994), Petit (1997), Susanna et al. (2006)	Xerantheminae
--	-----------	--------------------------------	--	---------------

Table 3. Overview and results from previous phylogenetic studies (performed with Sanger sequencing data) and present study (performed with NGS data) focused on tribe Cardueae with emphasis in the subtribal classification. Subtribes are divided following the classification proposed here. In the subtribes rows it is indicated the number of species sampled in the corresponding study. Statistically supported subtribes are highlighted in green, considering supported subtribes those with a monophyletic pattern and bootstrap support values $\geq 70\%$ (in ML and MP analyses) and Bayesian posterior probabilities ≥ 0.95 (in BI analysis). In case that the study used two phylogenetic inference methods, supported subtribes are considered when one or both methods showed a significantly supported monophyly. The nuclear dataset summarized corresponds to nuclear HybPiper dataset. Abbreviations used: BI = Bayesian Inference; C = chloroplast marker; H, H = Häffner and Hellwig; ML = Maximum Likelihood; MP = Maximum Parsimony; N = nuclear marker; N° = number; S = Susanna et al.

Systematic studies of the tribe Cardueae at subtribal level										
	S (1995)	H, H (1999)	Garcia-Jacas et al. (2002)		Susanna et al. (2006)		Barres et al. (2013)		Present study	
N° DNA markers	1	1	1	1	1	1+2	1	4	1055	87
DNA marker type	N	N	N	C	N	N+C	N	C	N	C
Inference method	MP	MP	MP	MP	BI	MP, BI	MP, BI	MP, BI	ML	ML
Total sampling	36	32	58	42	187	108	116	124	76	67
Carlininae	1	1	10	9	11	9	1	9	6	5
Echinopsidinae	1	1	5	3	7	5	7	7	3	2
Cardopatiinae	-	-	2	1	2	2	2	2	2	2
Dipterocominae	-	-	-	-	-	-	1	1	1	1
Berardiinae	-	1	1	1	1	1	1	1	1	1
Staehelininae	-	-	1	-	5	4	3	3	4	4
Xerantheminae	-	1	3	3	8	7	7	7	4	3
Onopordiinae	-	5	3	4	11	9	11	11	4	4
Carduinae	3	14	15	11	19	15	24	24	7	6
Saussureinae	1	3	8	2	25	5	5	5	7	6
Arctiinae	1	2	5	3	28	13	10	10	6	2
Centaureinae	29	4	5	5	70	38	44	44	31	31

Table 4. Summary metrics of sequence extraction performance for nuclear datasets (targeting 1061 COS targets; Mandel et al., 2014) using PHYLUCE (Faircloth, 2015) and HybPiper (Johnson et al., 2016) methods, and chloroplast dataset (targeting coding regions). Parameters from 4 to 11 are calculated based on dataset “3. N° of used loci”. Parameter values were extracted with FASconCAT-G v.1.02 (Kück and Longo, 2014) and AMAS (Borowiec, 2016) programs. Abbreviations used: bp = base pairs; max = maximum; min = minimum; N° = number; sd = standard deviation. *loci recovered for more than three species or higher than 3 bp length (see text for details)

Parameters	Nuclear datasets		Chloroplast dataset
	PHYLUCE method	HybPiper method	
1. N° of species included	87	87	77
2. N° of recovered loci (% respect COS targets)	902 (85.0)	1057 (99.6)	87
3. N° of used loci* (% respect COS targets)	776 (73.1)	1055 (99.4)	87
4. N° of captured loci in $\geq 90\%$ of species (%)	13 (1.7)	729 (69.1)	85 (97.7)
5. Average of recovered loci per species (sd; min-max)	363 (59; 169–455)	924 (174; 21–1021)	84 (2; 72–86)
6. Average of species recovered per loci (sd; min-max)	41 (26; 4–82)	76 (14; 4–86)	74 (12; 4–77)
7. Mean alignment length per locus in bp (sd; min-max)	634 (360; 90–3825)	314 (184; 23–1589)	902 (1331; 89–6851)
8. Length of concatenated matrix in bp	492,549	332,260	78,531
9. N° of variable sites in the concatenated matrix (%)	296,609 (60.2)	169,903 (51.1)	10,067 (12.8)
10. N° of parsimony informative sites in the concatenated matrix (%)	189,716 (38.5)	123,731 (37.2)	4290 (5.5)
11. Proportion of missing data in the concatenated matrix (%)	62.6	16.6	18.2

FIGURE CAPTIONS

Fig. 1. Phylogenetic reconstruction obtained with HybPiper target extraction method and under the concatenation approach (maximum likelihood analysis performed with RAxML), showing the evolutionary relationships among 12 newly defined subtribes within the tribe Cardueae. (A) cladogram, (B) phylogram. Branch labels indicate bootstrap (BS) support values, those < 70% (in red) are considered statistically unsupported.

Fig. 2. Phylogenetic reconstruction inferred from the chloroplast dataset (87 coding regions) under the concatenation approach (maximum likelihood analysis performed with RAxML), showing the evolutionary relationships among 12 newly defined subtribes within the tribe Cardueae. (A) cladogram, (B) phylogram. Branch labels indicate bootstrap (BS) support values, those < 70% (in red) are considered statistically unsupported.

Fig. 3. Phylogenetic trees inferred under coalescence approach (estimated with ASTRAL) for the nuclear HybPiper, composed by 1055 COS loci. Branch labels indicate local posterior probabilities (LPP), values < 0.95 (in red) are considered statistically low supported. Results from PhyParts of conflicting and concordant gene trees are also outlined for the nuclear HybPiper dataset (see text for details). The pie charts represent the proportion of the four categories (blue, green, red, and gray) for each node. For the subtribal clades, the number above and below branches indicate the number of gene trees concordant and in conflict, respectively, with that clade respect the species tree.

Fig. 4. Summary phylogenetic trees recovered in Barres et al. (2013) and present study showing evolutionary relationships among the 12 subtribes of the tribe Cardueae. On the left, we show phylogenies reconstructed with nuclear sequence data (A, C, E), and on right (B, D, F), those with plastid data. Trees redrawn from Barres et al. (2013) were obtained under maximum parsimony (MP) and Bayesian inference (BI) approaches, using a nuclear dataset (A) composed by sequences of internal transcribed spacer marker (ITS) and the chloroplast dataset (B) composed by a combination of sequence data from genes *matK*, *ndhF*, *rbcL* and intergenic spacer *trnL-trnF*. Phylogenetic trees inferred here derived from the nuclear HybPiper dataset composed by 1055 COS loci (C, E) and the chloroplast dataset composed by 87 coding regions (D, F), both datasets analyzed under the concatenation approach (app.; maximum

likelihood [ML] analysis performed with RAxML; C, D) and coalescence approach (estimated with ASTRAL; E). Branch labels indicate: bootstrap support (BS) values in case of MP and ML analysis; posterior probabilities (pp) in BI; and local posterior probabilities (LPP) in case of under coalescence. Values with BS < 70% and pp or LPP < 0.95 (in red) are considered statistically low supported. Note that in trees of Barres et al. (2013), the low supported branches were collapsed in the redrawn version. The genus *Saussurea* and *Jurinea* belong to Saussureinae, and *Arctium* and *Cousinia* to Arctiinae. *not monophyletic subtribe.

Fig. 5. Time-calibrated phylogeny of the 12 newly defined subtribes within the tribe Cardueae. The phylogenetic tree estimated with nuclear HybPiper dataset (1055 COS loci) and under concatenation approach (maximum likelihood analysis performed with RAxML) was used as the input tree for the penalized likelihood dating analysis using treePL. Gray bars on nodes show the 95% of confidence intervals (CI). Black circles on nodes represent the calibration points (CP) used in dating analysis. Number above branches are node codes. See Supplementary Table S3 for numeric details of median estimated ages and Lower 95% CI and Upper 95% CI for each node. In the scale axis, “P” and “Q” correspond to Pleistocene and Quaternary, respectively.

Fig. 6. Comparison of dating results of age node estimates (in Myr) for tribe Cardueae and subtribes obtained in the present study, with 1055 COS loci from nuclear genome (upper line in each subtribe), and those from Barres et al. (2013), with four chloroplast regions: *trnL-trnF*, *matK*, *ndhF*, *rbcL* (lower line in each subtribe). Colored boxes comprise the 95% of confidence interval limits and white dots represent the median age of lineages outlined. Note that subtribes Carduinae, Arctiinae, and Saussureinae are not represented due to their lack of monophyly in dated phylogeny of Barres et al. (2013).

Fig. 7. Representation of capitula diversity for the 12 newly subtribes recognized here within the tribe Cardueae: (A) subtribe Carlininae, *Atractylis humilis* L., Spain; (B) subtribe Cardopatiinae, *Cardopatium corymbosum* Pers., Barcelona Botanical Garden; (C) subtribe Echinopsinae, *Echinops viscosus* Rchb., Turkey; (D) subtribe Dipterocominae, *Dipterocome pusilla* Fisch. & C. A. Mey.; (E) subtribe Xerantheminae, *Sieberta pungens* J. Gay, Turkey; (F) subtribe Berardiinae, *Berardia subacaulis* **falta info Cristina**; (G) subtribe Staehelininae, *Staehelina uniflorescens* Sm., Greece; (H) subtribe Onopordinae, *Onopordum nervosum* Boiss., Spain; (I) subtribe Carduinae, *Cynara baetica* (Spreng.) Pau, Spain; (J) subtribe Arctiinae, *Cousinia lanata* C. Winkl., Tadjikistan; (K) subtribe Saussureinae, *Saussurea* **(JORDI)**, China; and (L) subtribe Centaureinae, *Centaurea cephalariifolia* Willk., Spain. (A) (B) (C) (E) (G) (H) (I) (J) (L) Photos by A. Susanna; (D) Photo by A. Pavlenko; (F) Photo by C. Roquet; (K) Photo by J. López-Pujol.

Fig. 8. Geographical representation of main distribution ranges of the 12 subtribes of the tribe Cardueae defined in the present study.

(B)

(A) **Plastid (concatenation approach, 87 genes)** (B)

Nuclear (coalescence approach, 750 loci)

Nuclear phylogenies

Plastid phylogenies

Barres et al. (2013)

Present study

Two different main time periods of subtribes origin

- 1 **Carlininae**
Cardopatiinae
Dipterocominae
Xerantheminae
Berardiinae
- 2 **Echinopsinae**
Staehelininae
Onopordinae
Carduinae
Arctiinae
Saussureinae
Centaureinae

(A)

Nuclear PHYLUCE (concatenation approach, 776 loci)

(B)

Nuclear PHYLUC (coalescence approach, 776 loci)