The limitations of molecular markers in phylogenetic reconstruction: the case of *Centaurea* section *Phrygia* (Compositae)

Javier López-Alvarado,1,2,4 Llorenç Sáez,3 Rossella Filigheddu,1 Núria Garcia-Jacas,2 and Alfonso Susanna2

1 Dipartimento di Botanica ed Ecologia Vegetale, Facoltà di Scienze Matematiche, Fisiche e Naturali, Università degli Studi di Sassari, Via Piandanna 4, I-07100 Sassari, Italy

2 Botanic Institute of Barcelona (IBB-CSIC-ICUB), Passeig del Migdia s.n., ES-08038 Barcelona, Spain

3 Unitat de Botànica, Facultat de Biociències, Universitat Autònoma de Barcelona, ES-08193 Bellaterra, Spain

4 Author for correspondence: al.loja2@gmail.com

SHORTENED TITLE: *Centaurea* sect. *Phrygia* Phylogeny

KEYWORDS: *Centaurea*, *Lepteranthus*, Molecular Phylogeny, *Phrygia*, Systematics
In this paper we seek to elucidate the phylogeny of *Centaurea* sects. *Jacea* and *Phrygia* (= sect. *Lepteranthus* nom. inval.) which have long been a source of controversy. A molecular phylogenetic approach is used based on nrDNA and plastid markers. The study confirmed incongruence between datasets, which can be explained both by hybridization and the occurrence of shared ancestral polymorphisms. Both factors are critical in limiting the resolution of phylogenetic trees. Despite this, we provide an interpretation of the current European distribution of sect. *Phrygia*, and suggest a probable eastern center of origin for the section. Our results support previous studies in *Centaurea*, suggesting on molecular grounds that sects. *Phrygia* and *Jacea* can not be clearly separated, particularly when *C. nigra* s.l. or *C. jacea* s.l. coexist with other taxa of sects. *Phrygia* and *Jacea*. 

This is an Accepted Manuscript of an article published in Taxon 63(5) :1079-1091 (2014), available online: http://dx.doi.org/10.12705/635.6
INTRODUCTION

Reconstruction of the evolutionary history of angiosperms has greatly improved over recent decades through the application of molecular phylogenetic methods, so that the main clades of angiosperms are now well established (Soltis & Soltis, 2004; APG III, 2009). Nevertheless, at lower taxonomic levels, work is far from being complete because the presence of hybridization, introgression and reticulation increases the difficulty of reconstructing phylogenetic relationships. The genus *Centaurea* L. illustrates this ongoing research process: relationships of the main groups are clear (Garcia-Jacas & al., 2001; Susanna & Garcia-Jacas, 2009), but the circumscription of some infrageneric groups remains unresolved.

*Centaurea* is one of the largest genera of the Compositae, with about 250 species (Susanna & Garcia-Jacas, 2007), and its taxonomic treatment has always been complicated. However, the latest morphological (Wagenitz & Hellwig, 1996) and molecular studies (Garcia-Jacas & al., 2000, 2001, 2006) have provided evidence that allows a precise delineation of the genus. *Centaurea* is currently reduced to three natural, monophyletic groups (Susanna & Garcia-Jacas, 2009): subg. *Lopholoma* (Cass.) Dobrocz., subg. *Cyanus* (Mill.) Hayek, and subg. *Centaurea*, the latter known informally as the Jacea group in Garcia-Jacas & al. (2006).

A comprehensive molecular survey of subg. *Centaurea* revealed that the vast majority of species can be classified into two large complexes: a first group of taxa with mostly spiny involucral appendages (eastern and western Mediterranean clades; cf. Garcia-Jacas & al., 2006), and a second group with unarmed appendages comprising sects. *Centaurea* [formerly sect. *Acrolophus* (Cass.) DC.], *Phalolepis* (Cass.) DC., *Willkommia* G. Blanca, *Jacea* (Mill.) DC., and *Phrygia* Pers. [=sect. *Lepteranthus*]...
(Neck.) DC. nom. inval., cf. Lanjouw & al., 1961]. The classification of these diverse sections with unarmed appendages awaits further work, although some attempts have already been made (Garcia-Jacas & al, 2006; Suárez-Santiago & al., 2007b). More recently, Hilpold & al. (2014) confirmed that sects. Centaurea, Phalolepis, and Willkommia form a clade that appears to be distinct from a second clade formed by sects. Jacea and Phrygia. Circumscription of sects. Centaurea, Phalolepis, and Willkommia is an onerous task owing to intense hybridization (Suárez-Santiago & al., 2007a, b), and the relationships among their species remain unsolved (Suárez-Santiago & al., 2007a, b; Hilpold & al., 2014). This lack of resolution has been suggested to be caused at least in part by hybridization events and reticulate evolution, although other phenomena such as retention of ancestral polymorphisms could also be affecting the molecular phylogenies of this group.

Sections Jacea and Phrygia together constitute a natural group. Centaurea sect. Phrygia includes species with long, linear, pectinate-fimbriate appendages and achenes with pappus. In contrast, sect. Jacea is characterized by broadly ovate or orbicular membranous appendages with entire, sometimes laciniate or denticulate margins and achenes usually without pappus. Ecological requirements of both sections are also different. Most members of sect. Jacea are typical elements of montane and subalpine meadows, influenced by frequent mowing, grazing or avalanches, and often show clonal reproduction. In contrast, most species of sect. Phrygia are Mediterranean taxa growing in xeric conditions, and rarely show clonal reproduction. Species of both sections are outcrossers, as is usually the case in Centaurea due to marked protandry (Jeffrey, 2009).

Historically, different authors have classified the species of these two sections differently. Some proposed a broad sect. Jacea that includes species from both sects.
Jacea and Phrygia. For example, Boissier (1875) proposed sect. Jacea to include different subsections, one of them subsect. Phrygia. Other authors, however, considered the long, linear, pectinate-fimbriate appendages as a key character to distinguish sect. Phrygia: Cassini (1823), De Candolle (1838), Hayek (1901), and Dostál (1976). The only molecular data relevant for this problem were reported by Garcia-Jacas & al. (2006), who concluded that the differences between sects. Phrygia and Jacea were not sufficiently significant to separate them.

In addition to historical disagreements in classification, polyploidy and hybridization have been reported for sects. Jacea and Phrygia (Dostál, 1976; Vanderhoeven & al., 2002; Koutecký, 2007). Both sections have a constant chromosome base number of $x = 11$ (Hellwig, 2004) and the most frequent ploidy levels reported are the diploid and the tetraploid (Dostál, 1976; Koutecký, 2007; Arnelas & Devesa, 2011, 2012). Usually both cytotypes coexist in the same species, but recent studies suggest that there is a high degree of isolation between cytotypes both at reproductive and ecological level whereas taxa of the same ploidy level hybridize more or less freely when in contact (Hardy & al. 2001; Koutecký, 2007; Koutecký & al. 2011). Furthermore, cytotype diversity seems to be associated to the wide-ranging species such as C. nigra L., C. jacea L., and C. Phrygia L..

Despite several reported drawbacks (Álvarez & Wendel, 2003; Nieto Feliner & Rosselló, 2007), nuclear-ribosomal DNA might be useful in reconstructing the phylogenetic relations of sects. Phrygia and Jacea because it has been demonstrated to be at least partly useful for establishing the phylogenetic relationships of many closely related species in Centaurea (Garcia-Jacas & al., 2000, 2006; Suárez-Santiago & al., 2007a, b). Furthermore, in case of putative hybrids it can help in the detection of
parental sequences among cloned PCR products (Baldwin & al., 1995; Sang & al., 1995; Fuertes Aguilar & al., 1999; Fuertes Aguilar & Nieto Feliner, 2003). The addition of non-coding cpDNA regions is useful for increasing phylogenetic signal and also may reveal incongruence attributable to gene flow (Garcia-Jacas & al., 2009 and references therein).

The aims of the present work are to test if sect. Phrygia is a monophyletic group and to elucidate phylogenetic relationships between sect. Phrygia and sect. Jacea, paying particular attention to the value of a molecular phylogenetic approach when dealing with complex groups. We will also propose hypotheses on the biogeographic history of the group and the current pattern of species distribution.

MATERIALS AND METHODS

Plant material. — Sampling for molecular analysis was focused on species of sects. Phrygia and Jacea. Besides Flora Europaea (Dostál, 1976), regional Floras or reviews (Laínz, 1967; Soldano, 1978, 1994; Amich, 1991; Strid & Tan, 1991; Bolòs & Vigo, 1996; Bancheva & Greilhuber, 2006) and the latest nomenclatural proposals by Greuter (2006+) were considered, because Dostál (1976) overlooked some previously described species and because many changes in nomenclature and classification have been made since his treatment. Sampling of sect. Phrygia consisted of 20 species which included almost all known species assigned to the section. Although sect. Phrygia is mainly a European group, there is one extra-European species, the north African endemic C. ali-beyana Font Quer & Pau. Unfortunately, this species was not available for the study despite a thorough but unsuccessful search at the type locality. Furthermore, herbarium material was scarce, and when available it had been treated chemically and DNA extraction was unsuccessful. Regarding sect. Jacea, the treatment
of species is complicated since most of the species accepted by Dóstal (1976) have recently been considered as subspecies or varieties of a broadly defined *C. jacea* (Greuter, 2006+). Moreover, *C. nigra* and other morphologically related species, which were placed in sect. *Phrygia* by Dóstal (1976), are here considered as belonging to sect. *Jacea* because it was demonstrated that they constitute extremes of continuous variation (Vanderhoeven & al., 2002). We sampled most of the section with seven species: *C. debeauxii* Godr. & Gren., *C. exarata* Coss., *C. inexpectata* Wagenitz, *C. nemoralis* Jord., *C. nevadensis* Boiss. & Reut., and a representative sampling of the variation of *C. jacea* L. and *C. nigra* L.. Individuals from these species were labeled as *C. nigra* sensu lato (s.l.) or *C. jacea* s.l. when they could not be unequivocally assigned to a particular subspecies. In addition, one individual of *C. nigrescens*, classified by Dostál (1976) in its own section (*Nigrescentes* (Hayek) Dostál), but with poorly-defined sectional affinities (Boissier, 1875), was included. Outgroups were *Centaurea triumfetti* All. subsp. *stricta* (Waldst. & Kit.) Dostál and *C. napulifera* Rochel subsp. *thirkei* (Sch. Bip.) Dostál of subg. *Cyanus*, which is sister to subg. *Centaurea* (Garcia-Jacas & al., 2001). Two species of sect. *Hierapolitanae* Garcia-Jacas, Hilpold, Susanna & Vilatersana, which is part of a group of sections sister to *Jacea-Phrygia* (Hilpold & al., 2014), were also chosen as outgroups, i.e., *C. hierapolitana* Boiss. and *C. toissensis* Freyn & Sint. ex Freyn. The list of plant material is provided in Appendix 1.

**DNA extraction, amplification and sequencing.** — Genomic DNA was extracted following the 2x CTAB method of Doyle & Dickson (1987) as modified by Cullings (1992) using silica gel-dried leaves collected in the field. In some cases, herbarium material was used. The ITS region was amplified for sequencing using primers 17SE and 26SE (Sun & al., 1994) following amplification profiles described in Garcia-Jacas...
The ETS was amplified with primers ETS1F (Linder & al., 2000) and 18SETS (Baldwin & Markos, 1998) following amplification profiles described in Garcia-Jacas & al. (2009). In addition, double-stranded cpDNA of the \textit{trnL}^{(UAG)}-\textit{rpl32} and \textit{ycf3-trnS} regions was amplified using \textit{rpl32F} as forward primer and \textit{trnL}^{(UAG)} as reverse primer for the \textit{trnL}^{(UAG)}-\textit{rpl32} region (Shaw & al., 2007), and SP43122F as forward primer and SP44097R as reverse primer for the \textit{ycf3-trnS} intergenic spacer region (Hershkovitz, 2006). The profile used for cpDNA amplification included a hot start at 95°C for 3 min. This was followed by 30 amplification cycles carried out under the following conditions: 95°C for 40 s, 54°C for 40 s, and 72°C for 1 min, with an additional extension step of 10 min at 72°C. PCR products were purified with ExoSAP-IT (USB Corp., Cleveland, OH, USA) and sequenced using a BigDye Terminator Cycle Sequencing kit v3.1 (Applied Biosystems, Foster City, CA, USA) at the University of Florida ICBR Core Facility, using the protocol recommended by the manufacturer and an ABI 3730xl DNA analyzer (Applied Biosystems). For nuclear markers, the ITS region was sequenced with 26SE, and the ETS with 18SETS as reverse primers. The cpDNA \textit{trnL}^{(UAG)}-\textit{rpl32} was sequenced with \textit{trnL}^{(UAG)} as reverse primer, and \textit{ycf3-trnS} with SP43122F as forward primer to avoid a poly-A region.

In order to determine possible hybridization events or individual polymorphisms, ITS and ETS PCR products from some species were cloned using a TOPO TA Cloning Kit for Sequencing (Invitrogen, Carlsbad, CA, USA) following the manufacturer’s instructions. Eight to sixteen positive colonies were screened with direct PCR using T7 and M13 universal primers under the following conditions: 3 min denaturation at 95°C followed by 35 cycles of 94°C, denaturation for 45 s, 50°C annealing for 45 s, and 72°C
extension for 1 min, with an additional 1 min at 72°C. Six to twelve PCR products were selected for sequencing in both directions using the same primers.

**Phylogenetic analyses.** — Sequences for phylogenetic analyses were edited using BioEdit v. 7.0.9.0 (Hall, 1999) and aligned visually by sequential pairwise comparison (Swofford & Olsen, 1990). The cloned sequences of ITS and ETS were grouped into different consensus sequences based on similarity, or inserted directly in the matrix for phylogenetic analyses.

Possible incongruence between the nuclear and the nuclear and chloroplast datasets was checked by eye by comparing tree topologies recovered for each individual marker. Additionally, indel coding of the cpDNA matrix was performed with IndelCoder 1.0 (Müller, 2006) using the Modified Complex Indel Coding (MCIC) algorithm. Coded indels were included as additional characters in the final phylogenetic analyses.

Phylogenetic analyses for the combined nrDNA and the cpDNA matrices were carried out using maximum parsimony and Bayesian methods. Analyses were performed using a simplified nuclear matrix excluding a) duplicate taxa when they grouped together; b) some cloned sequences; and c) taxa of suspected hybrid or introgressed origin, as suggested by their intermediate morphology. This was done in order to eliminate phylogenetic noise and improve resolution and statistical support (Vriesendorp & Bakker, 2005 and references therein), and also to provide an easier way to interpretate the trees.

Bayesian inference estimation was carried out using MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). The best-available model of molecular evolution required for Bayesian estimations of phylogeny was selected using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as
implemented in jModeltest 0.1.1 (Guindon & Gascuel, 2003; Posada, 2008), which considers nucleotide substitution models that are currently implemented in MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). The HKY model with variable base frequencies was assumed to follow a discrete gamma distribution (Hasegawa & al., 1985) and was selected as the best-fit model of nucleotide substitution for the ETS dataset. For the ITS alignment, the symmetrical model with equal base frequencies and invariable sites (SYM+I) was selected (Zharkikh, 1994). The cpDNA matrix with indels coded as characters was analyzed using the presence-absence model F81 (Felsenstein, 1981). Bayesian inference analyses were initiated with random starting trees and were run for $10^6$ generations. Four Markov Chains were run using Markov Chain Monte Carlo (MCMC) principle sample trees. We saved one out of every 100 generations, which resulted in 10,000 sample trees. Data from the first 2500 generations were discarded as burn-in after we had confirmed that the likelihood values had stabilized prior to the 2500th generation.

Maximum Parsimony (MP) analyses were heuristic searches conducted with PAUP* version 4.0b10 (Swofford, 2002), using tree-bisection-reconnection (TBR) branch swapping with character states specified as unordered and unweighted. We conducted a heuristic search with 1000 replicates and random taxon addition, saving 500 trees per replicate. Strict consensus trees of all most parsimonious trees were calculated (trees not shown). Bootstrap analysis (Felsenstein, 1985) was performed using a heuristic search with 1000 replicates and random taxon addition, saving a maximum of 10 trees per replicate.

In addition, a network analysis, which allows visualization of character incongruence, was conducted. We carried out a distance network analysis (split graphs)
on the combined nrDNA matrix which includes all sequenced individuals (not the simplified nrDNA matrix) in order to represent simultaneously groupings in the data and evolutionary distance between pairs of taxa. We used the neighbor-net (NN) algorithm (Bryant & Moulton, 2004) as implemented in SplitsTree4 v4.10 software (Huson & Bryant, 2006), set to uncorrected pairwise (p) distances, excluding constant and non-informative characters and gap sites.

RESULTS

The aligned matrix for ETS and ITS consisted of 78 sequences of 1181 bp and 170 parsimony-informative characters. The simplified nrDNA matrix consisted of 51 sequences of 1181 bp and 151 parsimony-informative characters. For the cpDNA matrix (\textit{trnL(UAG)}-\textit{rpl32} and \textit{ycf3-trnS}), 68 sequences of 1766 bp (1727 characters and 39 coded indels) and 61 parsimony informative characters were obtained. Discordances between the number of sequences for nuclear and cpDNA matrices are due to the presence of cloned sequences or consensus sequences of multiple clones in the nuclear alignment. The trees recovered from Bayesian analyses, with Bayesian posterior probabilities (PP) and MP bootstrap values (BS), are shown in Figs. 1–2 and refer to simplified nrDNA and cpDNA, respectively. The MP numerical results are shown in Table 1.

In the absence of significant conflict between datasets, the combination of different datasets usually improves the resolution in phylogenetic reconstructions. Separate analyses did not show support for incongruence between ITS and ETS (BS \(\geq 70\%\) or PP values \(\geq 0.95\)) and both markers yielded similar topologies. Therefore, the two nuclear datasets (ETS and ITS) were combined. Nuclear and cpDNA matrices were not combined because of significantly supported topological conflict (see Results). None of
the phylogenetic analyses recovered sects. *Jacea* and *Phrygia* as clades. Furthermore, the analyzed markers failed to resolve the deep relationships among taxa, and provided statistical support only for outgroup nodes and terminal clades of distinct morphological or geographical affinities.

**Cloning.** — ETS and ITS markers were cloned for some taxa and populations when the presence of multiple copies was suspected, either because of intermediate morphology suggesting hybridization, or because of the presence of multiple bands of PCR products, especially in the case of ETS. The presence of multiple bands in ETS was usually due to a different number of motif repetitions at the 5′ETS end. Most of the cloned copies usually grouped in the same clade. Only for *C. caballeroi*, *C. emigrantis*, *C. janeri* subsp. *janeri*, *C. linifolia*, and *C. montis-borlae* cloning provided evidence of intra-individual polymorphism, in most cases consisting in the presence of two different copies belonging to different clades (Figs. 1, 3).

**Nuclear markers.** — Resolution of deep nodes by the nuclear markers was poor, with only terminal branches being supported statistically, especially in the Bayesian analysis (Fig. 1). Four clades were robust in the non-simplified nrDNA analysis (Fig. 3): the *C. nigra* group (PP = 1.00; BS = 98%), the *C. phrygia* group (PP = 0.98), the *C. uniflora* group (PP = 1.00; BS = 84%), and the *C. parilica* group (PP = 1.00). When performing the analysis using the simplified nrDNA matrix, a fifth group, namely the *C. linifolia* group (Fig. 1) obtained statistical support (PP = 0.95). Some of these major clades include species from both sects. *Jacea* and *Phrygia* as used by Dostál (1976).

The network yielded by the NN analysis (Fig. 3) revealed a great amount of conflict in signal and lack of informative characters, especially within the *C. linifolia*, *C. Phrygia*, and *C. nigra* groups.
Statistically supported clades of the nrDNA tree are:

The *C. parilica* group (Figs. 1, 3) comprises species growing in the Greek and Bulgarian mountains. The Greek accession of *C. nervosa* subsp. *nervosa* (accession *C. nervosa* subsp. *nervosa* C), which should be part of the *C. uniflora* aggr. (Greuter, 2006+), groups here.

The *C. uniflora* group (Figs. 1, 3) is composed of species belonging to the *C. uniflora* aggr. proposed in Greuter (2006+). The only exception is one individual attributed to *C. nigrescens*.

The *C. nigra* group (Figs. 1, 3) comprises individuals mainly attributable to *C. jacea* and *C. nigra* from the Iberian Peninsula. One individual of *C. janeri* subsp. *gallaecica*, and a single cloning consensus copy from an individual of *C. janeri* subsp. *janeri*, are also found in this group (Fig. 3).

The *C. phrygia* group (Figs. 1, 3) has strong Bayesian support, even though it is not supported in the parsimony bootstrap analysis. This clade contains widely distributed taxa, including, some individuals of *C. jacea* from the Alps to the Balkans, *C. phrygia*, and some narrow endemics such as *C. jordaniana* subsp. *verguinii*, *C. pectinata*, and *C. triamularia*. In the case of individual #2 of *C. montis-borlae*, a narrow Italian endemic, one ITS clone fell within this clade (*C. phrygia* group), whereas another grouped with the remainder of the *C. montis-borlae* individuals (Figs. 1, 3).

The *C. linifolia* group (Figs. 1, 3), a credible clade on morphological and geographical grounds, is supported when the analysis is performed using the nrDNA simplified dataset. Species of this clade follow a reticulate pattern because *C. caballeroi*, *C. emigrantis*, and *C. linifolia* show different copies placed in different clades (Figs. 1, 3). Cloning efforts have failed to provide evidence for intra-individual
polymorphism in *C. linifolia*. Instead, sequencing two individuals from different populations revealed these differences.

**cpDNA markers.** — Strong incongruence in topologies was detected between nrDNA and cpDNA trees. Some species recovered as monophyletic in the nrDNA trees appeared nested in separate clades in the cpDNA tree, for example individuals A and B of *C. nervosa* subsp. *nervosa* and individuals of *C. stenolepis* subsp. *razgradensis* or *C. jacea* subsp. *vinylsii* (Figs. 1–2). Furthermore, the putative sister species of the remainder of the study group in the nuclear ETS+ITS Bayesian and MP analyses (and also in Garcia-Jacas & al., 2006), *C. inexpectata*, from Turkey, was unexpectedly found to group with species from the Alps (*C. jordaniana* in Fig. 4) with high support (PP = 1.00; BS = 80%, Fig. 2). Of the remaining taxa, only a large clade containing species of diverse geographical distribution and morphological affinities was supported in the Bayesian analysis (PP = 0.97, Fig. 2).

**DISCUSSION**

Our results reveal the limitations of molecular markers when used to resolve phylogenetic relationships within sect. *Phrygia* and between sect. *Phrygia* and sect. *Jacea*. The main reason for such limitations is related to the conflicting signals between datasets which result in incongruent estimates of phylogeny and may reflect a complex evolutionary history (Figs. 1–2). Lineage sorting of ancestral polymorphisms and/or hybridization are possible explanations for the incongruent phylogenies that are to be expected in recent and rapid radiations.

The limited number of informative characters certainly affects the resolution of deep nodes and is a reflection of the recent divergence of the *Jacea-Phrygia* group, dated to 3.73 (1.96–5.77) million years ago by Hilpold & al. (2014). The recent origin
and the rapid species radiation, suggested by the high number of short branches in the trees (Figs. 1–2), probably is the main explanation for the low number of informative characters. As pointed out by Neigel & Avise (1986), and considering in this context that cpDNA is comparable to mtDNA, “phylogenetic distributions of mtDNA can lack concordance with species boundaries when species are recently separated”. Furthermore, as concluded by Smissen & al. (2004), rapid species radiation implies that there has not been enough time for lineage coalescence in each new species or that there has not been enough time for the evolution of reproductive isolation between the newly generated species, which finally means lack of concordance between gene tree and species tree. Although incomplete lineage sorting is difficult to discern from reticulation (Joly & al., 2009) it could be a valid hypothesis, too, when individuals of seemingly unrelated species appear in the phylogeny as having a shared recent evolutionary history, even though their geographical distribution ranges are so widely separated that gene flow is unlikely, or when individuals of the same species appear as not monophyletic (Funk & Olmand, 2003). However, ancient hybridization can also explain this lack of species monophyly, when copies of nrDNA from other species persist due to incomplete concerted evolution, something that occurs very frequently in *Centaurea* (Garcia-Jacas & al., 2009). Furthermore, the evolutionary processes discussed are not always independent (Nieto Feliner & Rosselló, 2007).

**Hybridization and polyploidy.** — Hybridization has been demonstrated in *Centaurea* sect. *Jacea* and sect. *Phrygia* by several studies (Dostál, 1976; Vanderhoeven & al., 2002; Koutecký, 2007), and its importance in the evolution of the group is also highlighted by our results. A clear example of ongoing hybridization is the Kostenc hybrid (*C. jacea × C. phrygia* from Bulgaria; Fig. 3; *C. phrygia* group) which
is morphologically intermediate between *C. jacea* and *C. phrygia*. Nevertheless, the nrDNA of this hybrid can not be assigned to *C. jacea* or *C. phrygia* since the nrDNA of the three accessions (the two parents and the hybrid) is indistinguishable and therefore extensive introgression is the most probable explanation. Another case is *C. janeri* (*C. janeri* subsp. *gallaeca* and *C. janeri* subsp. *janeri* B CL2 con; Figs. 1, 3), from sect. *Phrygia*, which has nrDNA copies grouping within the *C. nigra* group. In particular, ongoing hybridization was detected *in situ* and confirmed in a cpDNA haplotype survey for *C. janeri* subsp. *janeri* B (López-Alvarado, 2012). *Centaurea nigra* is known for its hybridization potential with species from other sections (Roché & Susanna, 2010) and even other subgenera (*Centaurea ×valdesii-bermejoi* Fern. Casas & Susanna, cf. Fernández Casas & Susanna, 1982). Also *C. nervosa* subsp. *nervosa* C (*C. parilica* group, Figs. 1, 3) displays geographical relationships (Fig. 4) dominating over morphology, since it would be expected to group within the *C. uniflora* group, suggesting that reproductive barriers may be weak and this pattern could be the result of gene flow. However, since Goulimis (1956: 19) classified this Greek population as a different subspecies, further work, including morphology, should address if it could be a different species from the *C. nervosa* type from the Alps.

Ancient hybridization is more difficult to detect, but *C. montis-borlae* from the Apuan Alps (Fig. 4) could be the result of an ancient hybridization event. *Centaurea montis-borlae* #2 shares a nrDNA clone with *C. phrygia* and *C. jacea* from eastern Europe (Fig. 1). Although *C. montis-borlae* is restricted to the Apuan Alps (Apennines), which are isolated from the Alps by the Po valley, range continuity might be possible along the Ligurian mountain corridor to the Maritime Alps (Ansell & al., 2008). However, the Italian accession of *C. jacea* subsp. *angustifolia* (D) appears in the same
group (Fig. 3), which could indicate past or ongoing hybridization with *C. jacea*. However, no evidence for this was detected in situ in the sampled population. Finally, *C. jordaniana*, from the Maritime Alps, and *C. pectinata*, from northeastern Spain and southern France (Fig. 4), could also be potential examples of ancient hybridization. *Centaurea jordaniana* subsp. *verginii* appears to be related to eastern species from the Alps and Balkans in the nrDNA analysis, rather than to its geographical neighbors *C. jordaniana* subsp. *aemilii*, and *C. jordaniana* subsp. *jordaniana*. Since the Alpine and Balkan ranges were once connected (Schmitt, 2009 and references therein), an ancient hybridization event can not be discounted. Furthermore, *C. jordaniana* subsp. *verginii* in the cpDNA tree (Fig. 2) grouped with the Turkish endemic *C. inexpectata*, a species that appears to be sister to all sect. *Jacea* and sect. *Phrygia* species in the nrDNA sequences (Fig. 1). On the other hand, *C. pectinata* shares its nrDNA with the Greek endemic *C. triamularia*, but the cpDNA with the Alpine *C. nervosa* subsp. *nervosa* A (Figs. 1–3). However, since only one copy was detected in the nrDNA markers for *C. jordaniana* and *C. pectinata*, retention of ancestral polymorphisms could be an alternative explanation.

Hybridization and polyploidy are often correlated (Soltis & Soltis, 2009), and sects. *Phrygia* and *Jacea* are well known for containing different ploidy levels, ranging from diploids to hexaploids (Dostál, 1976; Hardy & al., 2001; Koutecký & al., 2011; Arnelas & Devesa, 2011, 2012). Polyploidy can complicate the task of phylogenetic reconstruction. First, allopolyploidy implies hybridization which by itself is a source of incongruence, and second, back-crossing (introgression) is usually ploidy-level dependent and there are barriers between taxa of different ploidy level, whereas taxa of the same ploidy level hybridize more or less freely when in contact (Hardy & al., 2001;
Koutecký & al., 2011). This could be important especially in species such as *C. jacea*, *C. nigra*, and *C. phrygia* in which different ploidy levels are frequent without evident morphological differentiation (Koutecký, 2011; Arnelas & Devesa, 2011, 2012). Since different cytotypes of the same species are reproductively isolated evolutionary units despite being morphologically uniform, they evolve independently. This would explain why *C. nigra*, *C. jacea*, and *C. phrygia* (including *C. stenolepis* and *C. indurata*) appear as non-monophyletic species (Figs. 1, 3). As regards *C. phrygia*, the latest taxonomic treatment (Greuter, 2006+) places *C. indurata* and *C. stenolepis* within *C. phrygia*, whereas Dostál (1976) distinguished *C. stenolepis* from *C. phrygia* based mainly on two characters: “capitula usually solitary” vs. “capitula solitary or clustered at apices of branches” and “appendages mostly covering the bracts” vs. “appendages not completely covering the bracts”. Dostál (1976) also considered *C. indurata* a distinct species but indicated a probable hybrid origin (*C. phrygia* ×*C. stenolepis*). Different ploidy levels have been reported for *C. phrygia*, 2n = 22, 44 (Bancheva & Greilhuber, 2006; Koutecký, 2007), *C. stenolepis*, 2n = 22 (Dostál, 1976; Bancheva & Greilhuber, 2006) and *C. indurata*, 2n = 44 (Bancheva & Greilhuber, 2006), and therefore the lack of monophyly without significant morphological differences can be explained by isolation through different ploidy levels. It is difficult to support or reject new taxonomic proposals by Greuter (2006+) with regard to *C. indurata* and *C. stenolepis*, which inspite of weak morphological differences group in different clades. If they are truly isolated, the retention of the names proposed by Dostál (1976) seems to be the better option. Further work is needed to corroborate the hybrid origin hypothesis for *C. indurata*. 
Finally, besides biological phenomena such as hybridization and polyploidy, the inflation in the number of sections and species can affect the interpretation of results. Inflation in the number of species is particularly important because when we handle over-split taxonomies, the phylogeny will reflect contemporary gene flow between populations of the same species rather than hybridization between different species. Regarding the number of sections, it is interesting to comment on the case of *C. nigrescens*, placed by Dostál (1976) in its own section, sect. *Nigrescentes*, which grouped with the species of the *C. uniflora* group of sect. *Phrygia* (Fig. 1, 3). The suitability of retaining sect. *Nigrescentes* as an independent group will be addressed in future studies on the basis of new molecular evidence; however, as was demonstrated in previous studies of *Centauraea*, the number of sections in Dostál’s revision was inflated (García-Jacas & al, 2006; Mameli, 2008; Hilpold & al., 2009). Regarding the number of species, examples illustrating over-splitting are *C. phrygia*, *C. stenolepis*, and *C. indurata* discussed above, although there could be others such as the *C. linifolia* or *C. uniflora* groups. The behavior of individuals within the *C. linifolia* group suggests hybridization (Figs. 1, 3), especially in accessions of *C. caballeroi*, *C. emigrantis*, and *C. linifolia*. Incomplete lineage sorting due to recent allopatric speciation could also be a possible explanation if different nrDNA copies persist due to incomplete concerted evolution (López-Alvarado, 2012). However, some authors have considered *C. antennata*, *C. caballeroi*, and *C. linifolia* as a single species on account of their morphological similarity (Bolòs & Vigo, 1996). Considering over-splitting in *C. linifolia* as a valid hypothesis, due to morphological and ecological similarities between aforementioned species, probably does not solve the complex relationships within the group but makes the tree topology simpler, and much easier to interpret. The same
could be valid for *C. uniflora* and related taxa, (Fig. 3), considered by Greuter (2006+) as the *C. uniflora* aggr.

**cpDNA markers and biogeographic patterns.** — The tree topology obtained from cpDNA does not agree with morphology and is incongruent with the nrDNA tree. Incongruence between these two datasets could be partly due to the differential substitution rates of nrDNA and cpDNA, which are putatively slower in cpDNA. Since we are dealing with a recently evolved group, the slower substitution rates in cpDNA could explain the high degree of uniformity and intermixing among unrelated taxa (Neigel & Avise, 1986). Only an extensive haplotype survey, sampling several individuals per species, could reveal the real cpDNA haplotype pattern and haplotype sharing across the group. Additionally, some supported clades reveal affinities between species with distribution ranges which currently are not connected (Fig. 4), possibly indicating ancestral polymorphism retention or ancestral gene flow, as in *C. pectinata* and *C. nervosa* subsp. *nervosa* A (PP = 1.00; BS = 85%; Fig. 2, I) or *C. montis-borlae*, *C. jacea* subsp. *weldeniana*, and *C. jacea* subsp. *vinyalsii* (PP = 0.95; Fig. 2; II). Other clades provide geographical information, e.g., the clade formed by *C. parilica*, *C. indurata*, and *C. stenolepis* subsp. *razgradensis* from Greece and Bulgaria (PP = 0.95; Fig. 2, III). Finally, the clade formed by *C. janeri* subsp. *janeri* and *C. jacea* subsp. *vinyalsii* A corroborates that gene flow occurs between *C. jacea* s.l. and *C. janeri* (PP = 1.00; BS = 95%; Fig. 2, IV).

Even though cpDNA does not produce taxonomically consistent signal, when compared with nrDNA it allows detection of biogeographical patterns, in particular the eastern center of origin. As in other groups of *Centaurea*, an eastern origin of the *Jacea-Phrygia* group is most probable (Suárez-Santiago & al. 2007b; Font & al., 2009).
According to the nrDNA tree, *C. inexpectata*, a Turkish endemic, is sister to the rest of sect. *Jacea* and sect. *Phrygia*. The sharing of related cpDNA haplotypes among *C. inexpectata* and some narrowly endemic Alpine taxa, such as *C. jordaniana* subsp. *verguinii* and *C. rhaetica*, might indicate ancient hybridization as previously discussed, but also the retention of ancestral polymorphisms of a widely distributed ancestor which became later fragmented to form several narrowly distributed species by allopatric speciation (Figs. 2, 4). Furthermore, a clade formed by the strikingly similar *C. pectinata* and *C. triamularia* in the nrDNA tree (PP = 0.98; BS = 76%; Fig. 1), would also support this hypothesis. Since *C. pectinata* grows in southern France and northeastern Spain, and *C. triamularia* is restricted to the Pachtourion Mountain in Greece (Fig. 4), allopatric speciation is a suitable hypothesis. The pattern of cpDNA haplotype distribution in the remaining species does not follow either taxonomy or geography, but rather supports a wider ancestral distribution area followed by recent isolation. Moreover, recent speciation would also explain ongoing gene flow detected in some species. The basal position in the cpDNA tree of the Greek *C. pangaea* and *C. triamularia* and the Greek accession of *C. nervosa* subsp. *nervosa* is also interesting (Figs. 2, 4). All of them have cpDNA haplotypes similar to outgroup species by sharing a similar gap structure, which could be indicative of cpDNA capture (Schaal & al., 1998 and references therein) or shared ancestral polymorphism involving species outside the study group.

**Concluding remarks.** — The study show that the markers used did not completely succeed in inferring phylogenetic relationships of sects. *Phrygia* and *Jacea*. Conflicting signal and the lack of differentiation appear to be the main factors responsible for the low resolution of phylogenetic trees. Furthermore, hybridization and incomplete lineage
sorting are consistent with the data presented. Therefore, the complex evolution of the group appears to be the main limiting factor in resolving phylogenetic relationships of species within sect. *Phrygia* and the circumscription of sects. *Phrygia* and *Jacea*. The key morphological characters traditionally used to define infrageneric categories in *Centaurea* are not reflecting phylogenetic relationships of the taxa, as had been pointed out by Hilpold & al. (2011). In order to avoid confusion and misinterpretation of a taxonomically and nomenclaturally complex group such as *Centaurea*, we propose awaiting future studies to take decisions about sectional rearrangement.

ACKNOWLEDGMENTS

This study was supported by the Italian MIUR (“Ministero dell’Istruzione, Università e Ricerca”), by the Spanish “Ministerio de Ciencia e Innovación” (projects CGL2009-13322-C03-03/BOS and CGL2011-28613-C03-01), and by the Catalan Government (“Ajuts a Grups de Recerca Consolidats” 2009-SGR-439), and constitutes part of the PhD program of J. L.-A. We thank C. Aedo, G. Alziar, T. Constantinidis, M.B. Crespo, K. Diadema, A. Figueroa, A. Hilpold, F. Médail, A. Strid and *Flora iberica* project for their help in obtaining plant material. We also thank G. Galasso from the MI herbarium, L. Egildo from the FI herbarium, R. Guardia and J. Vicens from the BCN herbarium, I. Karnefelt from the LD herbarium and M. Velayos from the MA herbarium for providing dried plant material.

LITERATURE CITED


Suárez-Santiago, V.N., Blanca, G., Ruiz-Rejón, M. & Garrido-Ramos, M.A. 2007a. Satellite-DNA evolutionary patterns under a complex evolutionary scenario:


TABLES

<table>
<thead>
<tr>
<th></th>
<th>nrDNA</th>
<th>Simplified nrDNA</th>
<th>cpDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Characters</td>
<td>1181</td>
<td>1181</td>
<td>1766</td>
</tr>
<tr>
<td>Informative char.</td>
<td>170</td>
<td>151</td>
<td>61</td>
</tr>
<tr>
<td>Tree length</td>
<td>361</td>
<td>316</td>
<td>106</td>
</tr>
<tr>
<td>CI</td>
<td>0.516</td>
<td>0.577</td>
<td>0.423</td>
</tr>
<tr>
<td>RI</td>
<td>0.390</td>
<td>0.737</td>
<td>0.713</td>
</tr>
<tr>
<td>HI</td>
<td>0.483</td>
<td>0.425</td>
<td>0.577</td>
</tr>
</tbody>
</table>

**Table 1.** Numerical results of analyses of nrDNA, simplified nrDNA and cpDNA alignments. Tree length, CI, RI and HI were calculated for complete trees and for trees without outgroups. Abbreviations: CI, consistency index; HI, homoplasy index; Informative char., phylogenetically informative characters; RI, retention index; tree, the complete tree including outgroups; ingroup, tree excluding outgroups.
FIGURE LEGENDS

Fig. 1. 50% majority-rule consensus tree obtained from the Bayesian analysis of the simplified nrDNA dataset indicating supported clades. Numbers above branches are posterior probabilities, numbers below branches Bootstrap support values. Names in black are sect. Phrygia species, names in hollow letters are sect. Jacea species name in grey and underlined is Nigrescentes. CL = clone. Con = consensus sequence. s.l. = sensu lato. Country names are given as the ISO standard. Different individuals of the same population are indicated with consecutive numbers (1, 2, etc.). Individuals belonging to different populations are indicated with consecutive capital letters (A, B, etc.). Population codes are shown in the appendix table. Arrows indicate cloned sequences discussed in the text.

Fig. 2. 50% majority-rule consensus tree obtained from the Bayesian analysis of the cpDNA alignment. Numbers above branches are posterior probabilities, numbers below branches Bootstrap support values. Names in black are sect. Phrygia species, names in hollow letters are sect. Jacea species name in grey and underlined is Nigrescentes. s.l. = sensu lato. Country names are given as the ISO standard. Different individuals from the same population are indicated with consecutive numbers (1, 2, etc.). Individuals belonging to different populations are indicated with consecutive capital letters (A, B, etc.). Population codes are shown in the appendix table. Arrows and roman numerals indicate the clades discussed in the text.

Fig. 3. NN split graph based on uncorrected p-distances of the entire nrDNA matrix (non-informative, constant characters and gaps excluded). The five supported groups recovered in the study are indicated. PP and BS values are provided for the clades found.
in the Bayesian and MP analyses of the non-simplified nrDNA matrix. Some taxon names have been removed to increase readability of the network. CL = clone. Con = consensus sequence. s.l. = *sensu lato*. Different individuals of the same population are indicated with consecutive numbers (1, 2, etc.). Individuals belonging to different populations are indicated with consecutive capital letters (A, B, etc.). Population codes are shown in the appendix table. Boxes indicate cloned sequences discussed in the text.

**Fig. 4.** Geographical distribution of most species studied. Black squares: species of sect. *Phrygia*; white circles: species of sect. *Jacea*. 
Appendix 1. — Species sampled for molecular analysis, origin of materials, herbaria where vouchers were deposited and GenBank accession nos. for ITS, ETS, \textit{trnL}^{(UAG)}-\textit{rpl32} and \textit{ycf3-trnS} regions.


\textit{Centaurea caballeroi} Font Quer, population A, Spain, Tarragona, Serra de Montsià, Mas de Comú, 04-VI-2009, Barres & López-Alvarado (BC), KF721073, KF720969, KF721025, KF721101, KF721169.

\textit{Centaurea caballeroi} Font Quer, population B, Spain, Tarragona, Ports de Beceit, Portell de Caro, 02-VII-2008, L. Sáez (LSG pers-BCB), KF721039, KF721040, KF720970, KF721021, KF721102, KF721170.

\textit{Centaurea corcubionensis} M. Laínz, Spain, A Coruña, Carnota, O Pindo, carretera Pindo-Carnota, 10 m, orla de matorral, granito, 09-VI-2002, Iglesias-Louzan (BCN20949), KF721037, KF720971, KF721103, KF721171.

\textit{Centaurea debeauxii} Godr. & Gren., Spain, Huesca, Torla, Ribereta de Arazas, Ordesa y Monte Perdido, 14-IX-2008, Hilpold & Kosinsky (BC), KF721080, KF720972, KF721029, KF721104, KF721172.

\textit{Centaurea emigrantis} Bubani, Spain, Huesca, Castellonroi, Congost de Santa Ana, 26-VI-2008, Roquet & Sáez (LSG pers-BCB), KF721077, KF720973, KF721028, KF721105, KF721173.


\textit{Centaurea hierapolitana} Boiss., Turkey, Afyonkarahisar: Dazkırı, Çıkışı, 870 m, 24-VI-2004, Bağcı YBK-1523 (KNYA), KF721035, KF721104, KF721175.

\textit{Centaurea hyssopifolia} Vahl, Spain, Toledo, near Ontígola, 500 m, 22-VI-1996, García-Jacas, Susanna 1600 & Vilatersana (BC), DQ319119, KF720974, KF721108, KF721176.

\textit{Centaurea indurata} Janka, Greece, Dramas, Rodopi Mts., Place called Megalo Livadi ENE Dipotama, 1450 m, 06-VIII-2005, Strid 55839 (pers. herb.), KF721085, KF720975, KF721109, KF721177.

\textit{Centaurea inexpectata} Wagenitz, Turkey, Antalya, Gevne valley, high of village Kucukluk, 1750 m, 30-VI-2004, Uysal 598 (KNYA), DQ319122, KF720976, KF721110, KF721178.

\textit{Centaurea jacea} L., population A, Bulgaria, Pazardzik District, Velingrad, road margins 10 km from Yundola to Kosteneo, 22-VII-2010, Figueroa & López-Alvarado (BC), KF721052, KF720983, KF721117, KF721185.

\textit{Centaurea jacea} L., population B, Spain, Huesca, 1 Km ENE Refugio Linza, 21-VIII-2009, Hilpold
**Centaurea jacea subsp. angustifolia** (DC.) Gremli. population A, Albania, mountain pass between Tirana and Elbasan, 17-VIII-2009, Garnatje & Sánchez-Jiménez 17 (BC), KF721087, KF720984, KF721118, KF721186 –. 

**Centaurea jacea subsp. angustifolia** (DC.) Gremli. population B, Austria, Wien, Alpengarten im Belvedere, 11-X-2009, Hilpold AH20096014 (BC), KF721067, KF720980, KF721116, KF721184 –. 

**Centaurea jacea subsp. angustifolia** (DC.) Gremli. population C, Austria, Marchegg, between Lange Luss and Schloss Gänserndorf, 10-X-2009, Hilpold AH20093235 (BC), KF721068, KF720982, KF721115, KF721183 –. 

**Centaurea jacea subsp. angustifolia** (DC.) Gremli. population D, Italy, Campania, Caserta, Matese, 2.5 Km ENE Letino, 28-VII-2009, Hilpold AH20094010 (BC), KF721069, KF720977, KF721111, KF721182 –. 

**Centaurea jacea subsp. vinyalsii** (Sennen) O. Bolòs et al., population A, Spain, Lleida, Lliminana, in Barcedana riverbed, near road LV-9121, 01-VII-2010, Hilpold AH4058 & López-Alvarado (BC), KF721086, KF720978, KF721114, KF721179 –. 

**Centaurea jacea subsp. vinyalsii** (Sennen) O. Bolòs et al., population B, Spain, Zaragoza, road from Sigüés to Roncal, 19-VIII-2009, Hilpold AH20095001 & Kosinsky (BC), KF721065, KF720979, KF721112, KF721180 –. 

**Centaurea jacea subsp. vinyalsii** (Sennen) O. Bolòs et al., population C, Spain, Barcelona, La Garrotxa, Pass of Bracons, 1100 m, 04-XI-1995, Garcia-Jacas & Susanna 1593 (BC), DQ319125, KJ569086, KF721119, KF721187 –. 

**Centaurea jacea subsp. vinyalsii** (Sennen) O. Bolòs et al., population C, Spain, Barcelona, La Garrotxa, Pass of Bracons, 1100 m, 04-XI-1995, Garcia-Jacas & Susanna 1593 (BC), DQ319125, KJ569086, KF721119, KF721187 –. 

**Centaurea jacea subsp. weldeniana** (Rchb.) Greuter, Greece, Nomos Karditsis, Eparchia Karditsis, c. 2 km S and E of Messenikolas village, along the road to Karditsa, Garcia-Jacas, Karamplianis & Susanna 2740 (BC), KF721082, KF720985, KF721030, KF721120, KF721188 –. 

**Centaurea janeri Graells subsp. gallecica** M. Laínz, Spain, A Coruña, Toques, Serra do Careón, 700 m, 04-VII-2005, Iglesias-Louzán (BCN38899), KF721044, KF721034, KF721122, KF721190 –. 

**Centaurea janeri Graells subsp. janeri**, population A, Spain, Salamanca, carretera entre el Casarito y El Cabaco, 14-VII-2009, Figueroa & López-Alvarado (BC), KF721072, KF720986, KF721124, KF721191 –. 

**Centaurea janeri Graells subsp. janeri**, population B, Spain, La Rioja, Haro, 0,3 Km W San Felices de Bilbilbo, near service area of AP-68, 02-VII-2009, Hilpold AH20093007, Garcia-Jacas & Vilatersana (BC), KF721060, KF721023, KF721024, KF721121, KF721189 –. 

**Centaurea janeri Graells subsp. babiana** M. Laínz, Spain, León, Sena de Luna, Ermita de Rabanal, 13-VII-2009, Figueroa & López-Alvarado (BC), KF721043, KF720987, KF721125, KF721193 –. 

**Centaurea jordaniana Godr. & Gren. subsp. jordaniana**, France, Provence-Alpes-Côte d'Azur, Alpes-Maritimes, Duranus, Gorges de la Vésubie, 17-VI-2009, Diadema (CBNMEDE), KF721042, KF720990,
Centaraea jordaniana subsp. aemilii (Briq.) Kerguélen, France, Provence-Alpes-Côte d'Azur, Alpes-Maritimes, Toudon, Mont Vial, 17-VI-2009, Diadema (CBNMED), KF721041, KF720988, KF721127, KF721194 – Centarea jordaniana subsp. verguinii (Briq. & Cavill.) Kerguélen, France, Provence-Alpes-Côte d'Azur, Alpes-Maritimes, Thiery, Forêt Domaniale de la Madone, 17-VI-2009, Diadema (CBNMED), KF721081, KF720989, KF721126, KF721196 –.
