Characterization of four wild edible Carduus species from the Mediterranean region via phytochemical and biomolecular analyses

This is the author's manuscript

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/1657772 since 2018-03-09T12:47:05Z

Published version:
DOI:10.1016/j.foodres.2017.07.071

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
This is the author's final version of the contribution published as:


The publisher's version is available at:

When citing, please refer to the published version.

Link to this full text:
[http://hdl.handle.net/2318/1657772]

This full text was downloaded from iris-Aperto: https://iris.unito.it/
Characterization of four wild edible Carduus species from the Mediterranean region via phytochemical and biomolecular analyses

Arianna Marengo⁹, Andrea Maxia⁹, Cinzia Sanna⁹, Cinzia M. Bertea⁷*, Carlo Bicchi⁹, Mauro Ballero⁹, Cecilia Caglieri⁹, Patrizia Rubiolo⁹*

⁹Dipartimento di Scienze della Vita e dell’Ambiente, sezione di Botanica, Università di Cagliari, Viale Sant’Ignazio da Laconi 13, 09123 Cagliari, Italy

⁷Dipartimento di Scienze della Vita e Biologia dei Sistemi, Unità di Fisiologia Vegetale, Università di Torino, via Quarello 15/A, 10135 Torino, Italy

⁹Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, I-10125 Torino, Italy

*Corresponding authors

Prof. Dr. Patrizia Rubiolo
Dipartimento di Scienze della Vita e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, I-10125 Torino, Italy, e-mail: patrizia.rubiolo@unito.it, tel:+39 011 670 7173 fax: +39 011 236 7661

Prof. Dr. Cinzia M Bertea
Unità di Fisiologia Vegetale, Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, via Quarello 15/A, 10135 Torino, Italy,
e-mail: cinzia.bertea@unito.it, tel: +39 011 670 6360 fax: +39 011 236 6360
Abstract

*Carduus* species (Compositae) are widely distributed in the Mediterranean area, and traditionally used for both food and medicinal purposes.

The hydroalcoholic extracts of four wild edible *Carduus* species collected in Sardinia (*Carduus argyroa* Biv., *Carduus nutans* subsp. *macrocephalus* (Desf.) Nyman, *Carduus pycnocephalus* L., *Carduus cephalanthus* Viv.) were analyzed and characterized by HPLC-PDA-MS/MS and PCR-RFLP of the nrDNA internal transcribed spacer (ITS).

Flavonoids and caffeoylquinic acid derivatives were the predominant classes of secondary metabolites characterizing the extracts. The ITS region was sequenced in parallel, and a PCR-RFLP method was applied with three selective restriction enzymes. Statistical analyses, on both chemical and biomolecular results, revealed that individuals clustered according to their taxonomic classification.

The combination of the two techniques discriminates the four species within the genus, giving further information on these little-investigated plants, traditionally used in the Mediterranean area and in Sardinia.

**Keywords:** *Carduus* spp., flavonoids, caffeoylquinic acids, HPLC-PDA-MS/MS, PCR-RFLP, ITS sequence
1. Introduction

Wild edible species are traditionally consumed mainly for their taste, as well as for their healthy and nutritional properties. Many popular dishes prepared with wild plants are still consumed, nowadays increasingly so with the return to traditions, with the primary aim of finding healthy alternatives to commercial foods (Guarrera & Savo, 2016). This is part of a trend aiming at re-discovering local products, often offered as culinary specialities in fairs and markets. Further, several edible plants are traditionally used as depuratives or to treat trivial illnesses. These species may potentially play an important role as functional foods, thanks to the great variety of physiologically-active components providing health benefits (Guarrera & Savo, 2016; Lentini & Venza, 2007; Pardo-de-Santayana et al., 2007; Ranfa, Maurizi, Romano, & Bodesmo, 2014).

Species from the genus *Carduus* (Compositae family), known in English as thistles, are traditionally consumed for their taste and biological effects. They are annual or perennial plants, 0.5 -2 m high, with lance-shaped, spiny-toothed leaves, spiny-winged stems and white-to-purple flowers. The genus includes approximately 100 species worldwide, which are widely distributed over the Mediterranean area (Al-Shammari, Hassan, & Al-Youssef, 2015; Dimitrova-dyulgerova, Zhelev, & Mihaylova, 2015; Lahaye et al., 2008; Thao et al., 2011; Tutin et al., 1968). They are consumed as raw or cooked, and are used as medicinal plants for the treatment of liver disorders or, more in general, for their diuretic and digestive properties (Al-Shammari et al., 2015; Atzei, 2003; Dimitrova-dyulgerova et al., 2015; Guarrera & Savo, 2016; Lentini & Venza, 2007; Rinchen & Pant, 2014; Signorini, Piredda, & Bruschi, 2009; Tardío, Pardo-de-santayana, & Morales, 2006). Several classes of secondary metabolites, chiefly flavonoids, phenolic acids, lignans, coumarins, alkaloids, sterols, and triterpenes, have been found in these species (Al-Shammari et al., 2015; Cardona, García, José R., & Pérez, 1992; Dimitrova-dyulgerova et al., 2015; Fernández, Garcia, Pedro, & Varea, 1991; Jordon-Thaden & Louda, 2003). The presence of these compounds may be associated
to the documented wide range of biological and nutraceutical properties that are associated to *Carduus* species: liver tonicity, anti-inflammatory, antioxidant, antispasmodic, anticancer, antiviral, and antibacterial activity (Al-Shammari et al., 2015; Jordon-Thaden & Louda, 2003; Koc et al., 2015; Slavov, Mihayloiva, & Dimitrova-dyulgerova, 2014).

This study aims to verify similarities and dissimilarities in the chemical and biomolecular profiles of four wild edible *Carduus* species (*Carduus argyroa* Biv., *Carduus nutans* subsp. *macrocephalus* (Desf.) Nyman, *Carduus pycnocephalus* L., *Carduus cephalanthus* Viv) growing in the Mediterranean area and in particular in Sardinia, where these species are traditionally consumed. Little has been published on these species, in particular on those growing in Sardinia. Some data on the characteristic compounds (mainly flavonoids) of *C. pycnocephalus* and *C. nutans* L. (Al-Shammari et al., 2015; Bain & Desrochers, 1988; Jordon-Thaden & Louda, 2003; Marrelli, Loizzo, Nicoletti, Menichini, & Conforti, 2013) and on the polyacetylenes of *C. argyroa* extracts are available (Harborne, Baxter, & Moss, 1999; Jordon-Thaden & Louda, 2003). To the best of the authors’ knowledge, no information is available on *C. cephalanthus* and *C. nutans* subsp. *macrocephalus*. At the same time, because of their relevance as traditional foods and remedies, it is of interest to learn more about these species, to verify the presence of compounds with nutraceutical properties, and to identify them in their extracts. The identification and discrimination of these closely-related species were approached by combining high performance liquid chromatography with diode array and mass spectrometry detectors (HPLC-PDA-MS/MS) and Restriction Fragment Length Polymorphism (PCR-RFLP) analysis of the nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) sequences, together with unsupervised multivariate data analysis and cluster analysis (PCA, HCA, Neighbor Joining, UPGMA). ITS gene as a DNA barcode marker is a useful tool to authenticate raw herbal materials, and in particular when (dried or processed) closely related species show similar chemical compositions. DNA barcoding has
successfully been applied to authenticate plant and animal samples, also in terms of food safety and quality control. However, a limit of this method is that reference sequences of uncommon plants are usually lacking in databases (Ali et al., 2014; Galimberti et al., 2013; Ha et al., 2015; Hebert, Cywinska, Ball, & Jeremy, 2003).

2. Materials and methods

2.1. Plant material

Aerial parts of the four wild species belonging to the genus *Carduus* were collected from different sites in Sardinia, in May and June 2015 (Table S1). They were identified at the Department of Life and Environmental Sciences, University of Cagliari, Italy, where a voucher specimen for each species was deposited. In total 10 specimens of *C. argyroa*, 6 of *C. cephalanthus*, 13 of *C. nutans* subsp. *macrocephalus* and 10 of *C. pycnocephalus* were collected. All plants growing at each site were separated by 1–50 m from one another, and were collected randomly. The fresh materials were dried at 40°C to constant weight.

2.2. Chemicals

HPLC-grade acetonitrile (LC-MS grade), formic acid (>98% purity), chlorogenic acid, rutin, apigenin 7-O-glucoside, apigenin, diosmin and kaempferol were from Sigma Aldrich (Bellefonte, USA). De-ionized water (18.2 MΩ cm) was obtained from a Milli-Q purification system (Millipore, Bedford, MA, USA). Luteolin, quercetin 3-O-glucoside, kaempferol 3-O-rutinoside and kaempferol 3-O-glucoside were from Extrasynthese (Genay Cedex, France). Cryptochlorogenic acid, 1,5-dicaffeoylquinic acid, 3,5-dicaffeoylquinic acid, 4,5-dicaffeoylquinic acid, luteolin 7-O-glucoside, luteolin 7-O-glucuronide, apigenin 7-O-glucuronide, kaempferol 3-O-rhamnoside, diosmetin and tricin were from Phytolab (Vestenbergsgreuth, Germany).
2.3. Sample preparation and HPLC-PDA-MS/MS analysis

Five hundred mg of each dried and ground aerial part were submitted to ultrasonic extraction with 10 mL of methanol/water (70:30, v/v) three times for 10 min. The extracts were then combined and centrifuged at 4000 rpm for 10 min. The supernatant was brought to a volume of 30 ml and filtered with a 13 mm diameter, 0.22 μm pore diameter hydrophilic PTFE syringe filter, before the HPLC-PDA-MS/MS analysis.

2.4. HPLC-PDA-MS/MS analysis

Each extract (5 μl) was analyzed in duplicate with a Shimadzu Nexera X2 system equipped with a photodiode detector SPD-M20A in series to a triple quadrupole Shimadzu LCMS-8040 system provided with electrospray ionization (ESI) source (Shimadzu, Dusseldorf Germany). Samples were analyzed on an Ascentis Express C18 column (15 cm x 2.1 mm, 2.7μm, Supelco, Bellefonte, USA) using water/formic acid (999:1, v/v) and acetonitrile/formic acid (999:1, v/v) as mobile phases A and B, respectively. The flow rate was 0.4 mL/min and the column temperature was maintained at 30°C. The gradient program was as follows: 5% B for 3 min, 5-15% B in 17 min, 15-25% B in 10 min, 25-75% B in 12 min, 75-100% B in 10 min, 100% B for 1 min. Total pre-running and post-running time was 60 min. UV spectra were acquired over the 220-450 nm wavelength range and the resulting chromatograms were integrated at 330 nm. MS operative conditions were as follows: heat block temperature: 200°C; desolvation line (DL) temperature: 250°C; nebulizer gas flow rate: 3 L/min drying gas flow rate: 15 L/min. Mass spectra were acquired both in positive and in negative full-scan mode over the range 100-1000 m/z, event time 0.5 sec. Product Ion Scan mode (collision energy: -35.0 V for ESI⁺ and 35.0 V for ESI⁻, event time: 0.2 sec) was applied to compounds for which a correspondence between the pseudomolecular ions [M+H]⁺ in ESI⁺ and [M-H]⁻ in ESI⁻ had been confirmed. Multiple Reaction Monitoring acquisition (collision energy: -35.0 V for ESI⁺ and 35.0 V for ESI⁻, dwell time: 20) was carried out on specific product ions derived
from precursor ions fragmentation (Table S2). Some of the main components were identified by comparing their retention times, UV and MS spectra to those of authentic standards (chlorogenic acid, cryptochlorogenic acid, dicafeoylquinic acids, rutin, quercetin 3-O-glucoside, luteolin 7-O-glucoside, luteolin 7-O-glucuronide, kaempferol 3-O-rutinoside, kaempferol 3-O-glucoside, kaempferol 7-O-rhamnoside, apigenin 7-O-glucoside, apigenin 7-O-glucuronide, diosmin, apigenin, luteolin, kaempferol, tricin, diosmetin). The other components were tentatively identified on the basis of their UV spectra and mass spectral information, compared to those given in the literature. Data were processed using LabSolution software (Shimadzu, Dusseldorf Germany).

2.5. DNA extraction, PCR amplification and sequencing

Ten milligrams of the same material employed for chemical analyses were ground to a fine powder, with the addition of approximately 5 mg of polyvinylpolypyrrolidone (PVPP, Sigma Aldrich, Bellefonte, USA). Genomic DNA was extracted from the ground powder using the Eurogold Plant DNA Mini Kit (Euroclone, Pero, Italy) following the manufacturer’s instructions. The quantitative and qualitative analyses of the isolated genomic DNA were assessed by spectrophotometry using the Nanophotometer (Implen GmbH, Munich, Germany) and by gel electrophoresis. Approximately 20 ng of genomic DNA were used as a template for PCR amplification with forward primer ITS1 (5’-TCCGTAGGTGAACCTGCGG-3’) and reverse primer ITS4 (5’-TCCTCCGCTTATGATATGC-3’) (White, Bruns, Lee, & Taylor, 1990). The amplification was carried out in a 25 μl reaction mixture containing 2.5 μl of 10X PCR buffer (Thermo-Scientific, Waltham, MA USA), 0.2 mM deoxynucleotide triphosphates (dNTPs), 20 pmol of forward and reverse primers, and 0.5 U of Taq DNA polymerase (Thermo-Scientific, Waltham, MA USA). PCR reactions were carried out in a T-Gradient Thermalcycler (Biometra GmbH, Göttingen, Germany). Cycling conditions consisted of an initial 4 min at 94°C, followed by 30 s of denaturing at 94°C, 45 s of annealing at 53°C and 45 s of elongation at 72°C, repeated for 35 cycles and with 10 min of final
extension at 72°C. PCR products were separated by 1.5% (w/v) agarose gel electrophoresis and visualized by ethidium bromide staining under UV. PCR products were purified using the Agencourt® AMPure® kit following the manufacturer’s protocol, and employed as a template for sequencing (IGA Technology Services, Udine, Italy). Both DNA strands were sequenced.

2.6. PCR-RFLP

The PCR products of the ITS gene of the four species were digested in separate reactions with 10 U of NdeI at 37°C for 2 h, or with 10U of Stul or of ApaLI, at 37°C for 1 h (NEB, New England Biolabs, Ipswich, AM, USA). One microliter of each digestion reaction was analyzed by capillary gel electrophoresis (CGE) using the Agilent 2100 Bioanalyzer (Agilent Technologies) and the DNA 1000 LabChip Kit (Agilent Technologies) following the manufacturer’s instructions. The DNA 1000 LabChip Kit provides sizing and quantitation of dsDNA fragments ranging from 25 to 1000 bp.

2.7. Statistical elaboration

Data obtained from the chemical analyses were processed through Principal Component Analysis (PCA), to reduce the multivariate space in which objects were distributed, and through hierarchical cluster analysis with quadratic Euclidean distance and Ward linkage, using SPSS 15.0 (IBM Corporation) software. Before data treatment, all variables obtained from HPLC-PDA-MS/MS analysis were scaled to unit variance.

Gene sequences were aligned with CLC sequence viewer software using default parameters to check the integrity of each sample sequence. Consensus sequences, obtained by aligning the individual sequences of each species, were then aligned using MEGA7 software (ClustalW) by modifying the Gap Opening and Gap Extension Cost values to 15 and 1, respectively. From this last alignment, a cluster analysis was made. Neighbor Joining and UPGMA statistical methods were
selected and relationships were tested with 1000 Bootstrap replicates, considering gaps in the Partial Deletion option.

3. Results and discussion

3.1. Phytochemical analyses

The extracts of *Carduus argyroa, Carduus cephalanthus, Carduus nutans* subsp. *macrocephalus* and *Carduus pycnocephalus* were analyzed by HPLC-PDA-MS/MS, to obtain their chromatographic profiles, and to obtain UV spectra, and mass spectral information concerning their components. MS data were acquired for a significant number of plants per species to identify the secondary metabolites characterizing them; all plants were analyzed twice. The repeatability of extract composition was evaluated on at least three plants per species. Fig.1 shows a representative chromatogram for each species. Approximately sixty peaks were detected through the HPLC-PDA-MS/MS untargeted metabolite analysis. Accordingly to the UV and MS collected information, 31 informative peaks were identified or tentatively identified and selected as target compounds for the statistical analysis. Among them, 20 were identified by comparing their UV and MS spectra to those of reference standards. The other 11 components were tentatively identified from their UV, MS and MS/MS spectra by comparison to literature data (Table 1). In agreement with the existing literature, flavonoids were the most representative compounds, and in particular quercetin, luteolin, kaempferol, diosmetin, and apigenin O-glycosides, resulting in 18 flavones and 7 flavonols; in addition, caffeoylquinic acids were also found (Al-Shammari et al., 2015; Ha et al., 2015; Jordon-Thaden & Louda, 2003; Li et al., 2014; Thao et al., 2011). Tandem mass spectrometry fragmentation provides further structural information on the compounds for which authentic standards are not available. The following compounds are given as illustrative examples of how the fragmentation patterns were used for identification. Compounds 4, 5, 6, 9, 10 and 16 are
characterized by maximum UV absorptions at approximately 348, 253 and 265 nm, which are typical for flavones (Li et al., 2014; Pandino, Lombardo, Mauromicale, & Williamson, 2011). For each peak, the MS/MS fragmentation gave a potential aglycone of 286 g/mol, and the positive ion mode MS/MS fragmentation yielded $^{1,3}\text{B}^+ (m/z \ 135)$ and $^{1,3}\text{A}^+ (m/z \ 153)$ fragments due to the $X^{1,3}$ cleavage of the C-ring, $^{0,2}\text{B}^+ (m/z \ 137)$ and $^{0,4}\text{B}^+ - \text{H}_2\text{O} (m/z \ 161)$ fragments due to the $X^{0,2}$ and $X^{0,4}$ cleavage of the C-ring, respectively. Since this fragmentation pattern is consistent with that of luteolin standard (compound 26), compounds 4, 5, 6, 16 were tentatively identified as luteolin O-glycosides. Compounds 9 and 10 were identified as luteolin glucuronide and luteolin glucoside, respectively, by comparison with the authentic standards. (Cuyckens & Claeys, 2004). Apigenin (17, 18, 20), kaempferol (12, 14, 24), diosmetin (19, 21, 22, 25, 27), quercetin (3, 7, 8), and their glycosides were identified or tentatively identified in the extracts with the same approach. The pseudomolecular [M – H]$^+$ diagnostic fragments of the aglycones were as follows: apigenin: MS$^2$ at $m/z$ 119, 153, 163; kaempferol: MS$^2$ at $m/z$ 121, 137, 153, 165; diosmetin: MS$^2$ at $m/z$ 258, 286; quercetin: MS$^2$ at $m/z$ 121, 127, 137, 153, 161, 165; tricin: MS$^2$ at $m/z$ 153, 203, 315. Six caffeoylquinic acid derivatives were also identified (1, 2, 11, 13, 15, 23), by comparison with reference standards. The only exception was compound 11, for which the MS/MS fragments ($m/z$ 163 and 191 in the positive and negative ionization mode, respectively) were consistent with those diagnostic for the dicafeoylquinic acid derivatives (Li et al., 2014; Martini, Conte, & Tagliazucchi, 2017).

From these results, interspecific differences can be associated to the presence/absence of specific compounds (Table 1). Some compounds, such as chlorogenic acid (1), cryptochlorogenic acid (2), dicafeoylquinic acids (13 and 14), kaempferol 3-O-glucoside (15), kaempferol 3-O-rhamnoside (24), luteolin (26), apigenin (28), kaempferol (29), diosmetin (30) and tricin (31), were present in all samples. Diosmetin derivatives (19, 21, 22) and 4,5 dicafeoylquinic acid (23) were only detected in
*C. cephalanthus* and *C. pycnocephalus*, while luteolin O-arabinosyl-glucoside (5), apigenin O-
rhamnosyl-glucoside (17), and apigenin 7-O-glucoside (18) were only present in *C. argyroa* and *C.
nutans* subsp. *macrocephalus* samples. Moreover, some peaks, detectable only in one of the four
species, can be taken as markers. In particular, luteolin O-arabinosyl-glucoside isomer (6) is
present only in *C. argyroa*, dicafeoylquinic acid (11) in *C. cephalanthus*, luteolin diglucoside (4),
luteolin acetyl diglycoside (16) and diosmetin acetyl glycosides (25, 27) in *C. nutans* subsp.
*macrocephalus*, and a quercetin diglycoside (3) in *C. pycnocephalus*. In spite of the same retention
time (19.898 min) and pseudomolecular ions (611 in ESI⁺ and 609 in ESI⁻, i.e. supposed molecular
weight 610 g/mol), compounds 3 and 4 fragmented differently in MS/MS, giving potential
aglycones at 302 g/mol and 286 g/mol, respectively.

A survey of the literature confirmed some of the results, viz. the presence of chlorogenic acid,
rutin, kaempferol, apigenin, luteolin and the related O-glycosides in *C. pycnocephalus* and *C.
nutans* (Al-Shammari et al., 2015; Bain & Desrochers, 1988; Jordon-Thaden & Louda, 2003;
Marrelli et al., 2013). The other (putatively) identified compounds, namely dicafeoylquinic acids,
diosmetin, quercetin, and tricin, were found in other *Carduus* species (Dimitrova-dylgerova et al.,
2015; Jordon-Thaden & Louda, 2003; Li et al., 2014). No data are available on the flavonoid
composition of *C. argyroa*, *C. cephalanthus* and *C. nutans* subsp. *macrocephalus*.

It is worth noting that the compounds detected in the species investigated are known to have
several nutraceutical properties. For instance, the antioxidant, hepatoprotective, antibacterial,
anticarcinogenic, and anti-inflammatory activities of most of the flavonoids and caffeineoylquinic acid
derivatives identified are documented (de Falco, Incerti, Amato, & Lanzotti, 2015; Nijveldt et al.,
2001).

3.2. Statistical data treatment
Principal Component Analysis (PCA) and hierarchical cluster analysis (HCA), were applied to discriminate samples from the species investigated. PCA is an unsupervised multivariate data analysis method with which the multivariate space in which objects are distributed can be reduced, so as to visualize similarities and/or differences within multivariate data of secondary metabolite composition (Zheng, Jiang, Wu, Wang, & Huang, 2014). HCA is an unsupervised pattern recognition method to detect similarities (Ha et al., 2015). The dataset of the 39 samples investigated was thus submitted to PCA and HCA, in a targeted approach based on the 31 previously characterized compounds listed in Table 1. The best result was obtained by scaling all variables to unit variance, so that all components had a standard deviation of one (Berg, Hoefsloot, Westerhuis, Smilde, & Werf, 2006).

Both score and loading plots were built with the aim of discriminating the four species and verifying the presence of discriminating variables. As shown in Fig. 2A, the first component (PC1) that explains 42.97% of the variation, individually separates *C. argyroa* and *C. nutans* subsp. *macrocephalus* from *C. pycnocephalus* and *C. cephalanthus*. Conversely, PC2 (20.98% of the variation) discriminates between *C. argyroa* and *C. nutans* subsp. *macrocephalus*. The related loading plot shows the influence of the variables in the distribution of the samples in the loading plot (Fig. 2B). From the results obtained in this plot, it is interesting to note that, among the compounds that negatively explain PC1, 19, 21, 22, 23, are present only in *C. pycnocephalus* and *C. cephalanthus* samples, while compounds 5, 17, and 18 are positively correlated with PC1 and are characteristic of *C. argyroa* and *C. nutans* subsp *macrocephalus* species (Table 1). Conversely, among the PC2 positively correlated variables, compounds 4, 25 and 27 are characteristic of *C. nutans* subsp. *macrocephalus* while 6, which is negatively correlated with PC2, is only present in *C. argyroa* (Fig. 2B). The third component (PC3) explains 10.30% of the variation and separates *C. cephalanthus* from *C. pycnocephalus* (Fig.2C). Fig. 2D highlights the positive correlation of 3 and
the negative correlation of 11 with PC3; these compounds are characteristic of *C. pycnocephalus* and *C. cephalanthus* species, respectively (Table 1).

These results suggest that some of these compounds can be discriminating for species identification purposes, although all components must be considered together for better species discrimination.

A three-dimensional PCA was also carried out, since the combination of the first two components is not sufficient to separate all individuals belonging to each species. The first three components, which explain 74.25% of the variation, give four distinct clusters according to species (Fig. S1).

In the next step, a dendrogram generated by hierarchic cluster analysis was constructed. Fig. 2E shows that all samples belonging to the same species are clustered together. This dendrogram confirms the results obtained with the PCA, and provides more information on the relationships among the four species. In particular, the chemical composition of *C. argyroa* appears to be more closely correlated to that of *C. nutans* subsp. *macrocephalus*, while *C. pycnocephalus* and *C. cephalanthus* cluster together.

### 3.3 ITS sequence analysis

For the PCR amplification, primers flanking the ITS-1 and ITS-2 regions were employed (Fig. 3A). It has been shown that the ITS regions provide better discrimination at the species level, for species belonging to Compositae family (Choi & Thines, 2015). This sequence comprises the rDNAs transcription units, which are well conserved in the higher plants, and the ITS-1 and ITS-2 regions, which vary widely in closely related species (Wu et al., 2011). The nucleotide composition of the resulting sequences is shown in Fig. 3B. The sequences are approximately 785-811 bp long, and the alignment of the four species' sequences shows that 89% of the sites are conserved. Data suggest the presence of some differences in the nucleotide composition of the sequences belonging to each species. In particular, of the 10.3% that are variable sites, 0.6% provide little
information and 9.3% are singleton sites. In agreement with the literature, most of these variable sites are located in the ITS-1 and ITS-2 regions (Wu et al., 2011). Analysis of the individual ITS sequences indicates the lack of intraspecific nucleotide variation (data not shown). A consensus sequence was built for each species.

The Neighbor Joining (NJ) and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) trees were built from the ITS region sequence alignment (Fig. 4 A, B). Cluster analyses showed a cluster linking C. argyroa to C. nutans subsp. macrocephalus. The C. cephalanthus nucleotide composition appeared to vary more significantly than the sequences of the other species. These results are in agreement with the dendrogram based on chemical data (Fig. 2E).

Analysis of the sequences obtained through pairwise sequence alignment with the software BLASTN 2.4.0+ gave the percentage of identity between each pair of sequences (Table S3) (Zhang, Schwartz, Wagner, & Miller, 2000). From the resulting percentages of identity, it emerged that the C. nutans subsp. macrocephalus sequence is more similar to those of C. argyroa and C. pycnocephalus (95% and 94%, respectively). C. argyroa and C. pycnocephalus showed 94% nucleotide identity, while the C. cephalanthus sequence was more variable than the sequences of the other species (92-93% of identity vs. the other species). These data are in accordance with the results of the hierarchical cluster analysis (Fig. 4).

A further interesting aspect is that each sequence is species-specific, and can be used as a barcoding gene. ITS sequences of C. pycnocephalus, C. nutans and C. nutans subsp. leiophyllus (Petrović) Stoj. & Stef. and other Carduus species are deposited in GenBank (Kelch & Baldwin, 2003; Robba, Carine, Russell, & Raimondo, 2005; Susanna, Hidalgo, Vilatersana, & Ciencia, 2006). The Carduus database ITS sequences, which include the ITS-1, 5.8 rRNA gene, and ITS-2 complete sequences, and 18S and 28S ribosomal RNA gene partial sequences, are generally shorter (from 686 to 737 bp) than those reported here. Blast analysis shows that the C. pycnocephalus sequence
obtained here is identical to those present in the database (EF123105.1, KT363916.1).

Additionally, the *C. nutans* subsp. *macrocephalus* sequence is almost identical to those of *C. nutans* and *C. nutans* subsp. *leiophyllus* (98-99% identity) (AF443678.1, HQ540426.1, EF543521.1, KT249753.1, KC603920.1, KT363914.1, JX867642.1, KKX167785.1). In particular, the *C. nutans* subsp. *leiophyllus* ITS sequence shows only one nucleotide variation versus the *C. nutans* ITS sequence. Conversely, in positions 70, 298 and 627, the *C. nutans* subsp. *macrocephalus* sequence has an adenine, while the *C. nutans* sequences already reported have a guanine, a cytosine, and a thymine, in the respective positions. No sequences belonging to the species *C. argyroa* and *C. cephalanthus* are present in GenBank, but BLAST alignment indicates the similarity of these sequences to those belonging to other *Carduus* species. This reveals that the ITS sequence has some conserved regions throughout the genus *Carduus*, even if variable nucleotide sites can discriminate between the individual species. The fact that the *C. nutans* subsp. *macrocephalus* sequence is almost identical to those of *C. nutans* and *C. nutans* subsp. *leiophyllus* may suggest that, in this case, this gene is not discriminatory at the subspecies level.

### 3.4. PCR-RFLP analysis

To better discriminate among the species, PCR-RFLP analyses were carried out. Three restriction enzymes were used to selectively cleave the resulting amplicons (Fig. 3C). The first four lanes of Fig. 5 show the undigested PCR products of the four *Carduus* species. From the RFLP analysis, it is possible to note that *NdeI*, ineffective on the other *Carduus* sequences, is able to discriminate *C. nutans* subsp. *macrocephalus* by giving two distinct fragments (512 and 295 bp, respectively). Conversely, *Stul* selectively cleaves only *C. argyroa* and *C. pycnocephalus* sequences, giving two diagnostic fragments (623 and 188 bp for *C. argyroa*, and 620 and 178 bp for *C. pycnocephalus*). Since the *C. argyroa* and *C. pycnocephalus* sequences are cleaved by *Stul* at the same sites, *ApoI* was employed to discriminate between these two species. This enzyme is able to cleave only *C.
pycnocephalus, releasing two diagnostic fragments of 508 and 290 bp, respectively. Lastly, the C. cephalanthus sequences show no cleavage sites for the three enzymes analyzed. These results show that it is possible to differentiate among the four species investigated with a combination of these three different restriction enzymes.

4. Conclusion

This study reports the first investigation of the phenolic acid and flavonoid profiles, and the ITS sequences and PCR-RFLP, of C. argyroa, C. cephalanthus, C. pycnocephalus and C. nutans subsp. macrocephalus from Sardinia. The differences in secondary metabolite profiles, defined by HPLC-PDA-MS/MS analysis, together with Principal Components and Hierarchical Clustering Analysis, afforded their unequivocal discrimination. Similar results were obtained with biomolecular analysis, through ITS sequence and PCR-RFLP analyses. ITS sequences of C. pycnocephalus, C. nutans and C. nutans subsp. leiophyllus and other Carduus species are available in the database (Kelch & Baldwin, 2003; Robba, Carine, Russell, & Raimondo, 2005; Susanna, Hidalgo, Vilatersana, & Ciencia, 2006) while the ITS genes of C. argyroa, C. cephalanthus and C. nutans subsp. macrocephalus were here sequenced for the first time and deposited in GenBank.

The complementary combination of the genetic and chemical approaches provides important information on little-investigated, but traditionally widely-used, plants and offers reliable discrimination of four morphologically-similar species belonging to the same genus. Moreover, chemical analysis detected interesting compounds with nutraceutical properties in the extracts investigated. This might support the traditional consumption and medicinal uses of these plants, and open new perspectives for further investigation of the compound(s) potentially responsible for the properties attributed to them. These findings may also promote the consumption of these herbs, including through the development of food supplements and functional foods. Moreover,
this approach can successfully be used for quality control of the species, e.g. in thistle-based commercial food products or traditional herbal remedy, for which their botanical and chemical composition must be quoted to assess origin, or because of regulatory requirements (e.g. food supplements).

**Funding**

This work was in part supported by Local Research Grants from the University of Turin, Italy (to CMB).

**References**


Choi, Y. J., & Thines, M. (2015). Host jumps and radiation, not co-Divergence drives diversification

https://doi.org/10.1371/journal.pone.0133655


https://doi.org/10.1371/journal.pone.0098061
Figure captions

Fig. 1. Representative profiles of the four Carduus species (λ=330nm). Compounds are numbered as in Table 1. (A) Carduus argyroa; (B) Carduus cephalanthus; (C) Carduus pycnocephalus; (D) Carduus nutans subsp. macrocephalus

Fig. 2. Statistical analysis of the 39 samples from the four Carduus species based on the 31 Carduus target compounds as variables. (A, B, C, D) PCA score and loading plots of samples from the four Carduus species. (E) Hierarchical cluster analysis.

Fig. 3. Comparison of ITS sequences among the four Carduus species. (A) Structure of the rRNA gene cluster, arrows indicates the position of the primers (ITS1 and ITS4). (B, C) Carduus ITS sequence alignment is shown. Variations in the nucleotide composition among the species are shaded. Ndel, Stul and ApaLI sites are indicated by squared solid box, dotted box, and dashed box, respectively.

Fig. 4. Hierarchical cluster analysis of Carduus species (numbers at the node indicate bootstrap values). (A) Neighbor Joining tree. (B) UPGMA tree.

Fig. 5. Capillary gel electrophoresis analysis of PCR products of the four Carduus species ITS regions and the relative fragments produced by the Ndel, Stul and ApaLI restriction enzymes. Lanes: 1, a single product of about 811 bp is produced by C. argyroa; 2 a single product of about 807 bp is produced by C. nutans subsp. macrocephalus; 3, a single product of about 798 bp is produced by C. pycnocephalus; 4, a single product of about 785 bp is produced by C. cephalanthus; 5, C. argyroa PCR products are not digested by Ndel; 6, C. nutans subsp. macrocephalus PCR products digested
by \textit{Nde}I give two major fragments of 512 and 295 bp, respectively; 7, \textit{C. pycnocephalus} PCR products are not digested by \textit{Nde}I; 8, \textit{C. cephalanthus} PCR products are not digested by \textit{Nde}I; 9, \textit{Stu}I cleaves \textit{C. argyroa} ITS region, giving two fragments 623 and 188 bp long; 10, \textit{C. nutans} subsp. \textit{macrocephalus} PCR products are not digested by \textit{Stu}I; 11, \textit{Stu}I cleaves \textit{C. pycnocephalus} ITS region, giving two fragments, 620 and 178 bp long; 12, \textit{C. cephalanthus} PCR products are not digested by \textit{Stu}I; 13, \textit{C. argyroa} PCR products are not digested by \textit{Apa}LI; 14, \textit{C. nutans} subsp. \textit{macrocephalus} PCR products are not digested by \textit{Apa}LI; 15, PCR products from \textit{C. pycnocephalus} give two fragments, of 508 and 290 bp, when digested by \textit{Apa}LI; 16, \textit{C. cephalanthus} PCR products are not digested by \textit{Apa}LI.

\textbf{Supplementary material}

\textbf{Table S1.} Sites and date of collection, voucher specimens, and GenBank accession numbers of the four \textit{Carduus} species

\textbf{Table S2.} Multiple Reaction Monitoring acquisitions

\textbf{Table S3.} Sequence homologies of ITS sequences between each pair of \textit{Carduus} sequences

\textbf{Fig. S1.} Three dimensional PCA score plot of samples of the four \textit{Carduus} species.
Fig. 1
Fig. 2
Fig. 3

CATATG = NdeI site

AGGCCCT = Stul site

G-TGCAC = ApaLI site
Table 1. List of identified and putatively-identified compounds in the four Carduus species. Each compound is quoted through its relative retention time, UV spectrum, molecular formula, pseudomolecular ions, molecular weight fragments obtained by Product Ion Scan mode (PIS), identified or tentatively-identified compound names. The identification confidence value, the presence of the compounds in the Carduus species, and the references are also given.

<table>
<thead>
<tr>
<th>No</th>
<th>RT (min)</th>
<th>λmax (nm)</th>
<th>Putative molecular formula</th>
<th>[M – H]+ m/z</th>
<th>[M – H]– m/z</th>
<th>Mol. weight g/mol</th>
<th>MS2+ m/z</th>
<th>MS2– m/z</th>
<th>Compound Name</th>
<th>ICb</th>
<th>Presence in Carduus species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1°</td>
<td>7.282</td>
<td>325</td>
<td>C16H18O9</td>
<td>355</td>
<td>353</td>
<td>354</td>
<td>163</td>
<td>191</td>
<td>5-O-Caffeoylquinic acid (chlorogenic acid)</td>
<td>1</td>
<td>A, C, M, P</td>
<td>(Dimitrova-dyulgerova et al., 2015; Li et al., 2014; Marrelli et al., 2013)</td>
</tr>
<tr>
<td>2°</td>
<td>9.075</td>
<td>325</td>
<td>C16H18O9</td>
<td>355</td>
<td>353</td>
<td>354</td>
<td>163</td>
<td>191</td>
<td>4-O-Caffeoylquinic acid (Cryptochlorogenic acid)</td>
<td>1</td>
<td>A, C, M, P</td>
<td>(Li et al., 2014)</td>
</tr>
<tr>
<td>3</td>
<td>19.88</td>
<td>255 353</td>
<td>C27H30O16</td>
<td>611</td>
<td>609</td>
<td>610</td>
<td>303, 121, 127, 135, 153, 161</td>
<td>301</td>
<td>Quercetin diglycoside</td>
<td>2</td>
<td>P</td>
<td>(Dimitrova-dyulgerova et al., 2015; Jordon-Thaden &amp; Louda, 2003; Li et al., 2014)</td>
</tr>
<tr>
<td>5</td>
<td>20.617</td>
<td>348 253 265</td>
<td>C26H30O16</td>
<td>581</td>
<td>579</td>
<td>580</td>
<td>287, 135, 137, 153</td>
<td>285</td>
<td>Luteolin O-arabinosyl-glucoside</td>
<td>2</td>
<td>A, M</td>
<td>(Cuyckens &amp; Claeyes, 2004; Li et al., 2014)</td>
</tr>
<tr>
<td>6</td>
<td>20.815</td>
<td>348 253 265</td>
<td>C26H30O16</td>
<td>581</td>
<td>579</td>
<td>580</td>
<td>287, 135, 137, 153</td>
<td>285</td>
<td>Luteolin O-arabinosyl-glucoside isomer</td>
<td>2</td>
<td>A</td>
<td>(Cuyckens &amp; Claeyes, 2004; Li et al., 2014)</td>
</tr>
<tr>
<td>7°</td>
<td>20.823</td>
<td>255 351</td>
<td>C27H30O16</td>
<td>611</td>
<td>609</td>
<td>610</td>
<td>303</td>
<td>301</td>
<td>Quercetin 3-O-rutinoside (rutin)</td>
<td>1</td>
<td>M, P</td>
<td>(Dimitrova-dyulgerova et al., 2015; Li et al., 2014; Marrelli et al., 2013)</td>
</tr>
<tr>
<td>8°</td>
<td>21.351</td>
<td>255 351</td>
<td>C22H20O12</td>
<td>465</td>
<td>463</td>
<td>464</td>
<td>303</td>
<td>301</td>
<td>Quercetin 3-O-glucoside</td>
<td>1</td>
<td>A, M, P</td>
<td>(Li et al., 2014)</td>
</tr>
<tr>
<td>No.</td>
<td>n°</td>
<td>m/z</td>
<td>C_{x}H_{y}O_{z}</td>
<td>Retention time</td>
<td>Species</td>
<td>A, B, C, D</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------------------------</td>
<td>------------</td>
<td>------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>21.758</td>
<td>349.253 265</td>
<td>C_{21}H_{20}O_{11}</td>
<td>449</td>
<td>447 448 287 135 137 153 161</td>
<td>285</td>
<td>Luteolin 7-O-glucoside 1 A, C, M, P (Cuyckens &amp; Claeys, 2004; Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>23.233</td>
<td>329.253 265</td>
<td>C_{25}H_{24}O_{12}</td>
<td>517</td>
<td>515 516 163 191</td>
<td>Dicaffeoylquinic acid 3 C (Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>23.702</td>
<td>265.344 265</td>
<td>C_{27}H_{30}O_{15}</td>
<td>595</td>
<td>593 594 287 285</td>
<td>Kaempferol 3-O-rutinoside 1 M, P (Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>24.109</td>
<td>329.253 265</td>
<td>C_{25}H_{24}O_{12}</td>
<td>517</td>
<td>515 516 163 191</td>
<td>1,5-dicaffeoylquinic acid 1 A, C, M, P (Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>24.231</td>
<td>265.344 344</td>
<td>C_{21}H_{20}O_{11}</td>
<td>517</td>
<td>515 516 163 191</td>
<td>Kaempferol-3-O-glucoside 1 A, C, M, P (Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>24.257</td>
<td>329.253 344</td>
<td>C_{25}H_{24}O_{12}</td>
<td>517</td>
<td>515 516 163 191</td>
<td>3,5-dicaffeoylquinic acid 1 A, C, M, P (Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>24.257</td>
<td>348.253 266</td>
<td>C_{27}H_{30}O_{16}</td>
<td>653</td>
<td>651 652 449 329 287 135 137 153 161</td>
<td>Luteolin acetyl diglycoside 3 M (Cuyckens &amp; Claeys, 2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>24.7</td>
<td>337.266 266</td>
<td>C_{27}H_{30}O_{16}</td>
<td>593</td>
<td>591 592 271 119 153 163</td>
<td>Apigenin O-rhamnosylglucuronide 2 A, M (Jordon-Thaden &amp; Louda, 2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>24.836</td>
<td>337.266 266</td>
<td>C_{27}H_{30}O_{10}</td>
<td>433</td>
<td>431 432 271</td>
<td>Apigenin 7-O-glucoside 1 A, M (Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>24.916</td>
<td>344.252 344</td>
<td>C_{27}H_{34}O_{17}</td>
<td>595</td>
<td>593 594 301</td>
<td>Diosmetin arabinosylglucoside 2 C, E (Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Value</td>
<td>C, H, O</td>
<td>Retention Time</td>
<td>Mass</td>
<td>Compound Name</td>
<td>Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>-----------</td>
<td>----------------</td>
<td>--------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25.074</td>
<td>C_{21}H_{20}O_{10}</td>
<td>447</td>
<td>269</td>
<td>Apigenin 7-O-glucuronide</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>25.221</td>
<td>C_{27}H_{34}O_{17}</td>
<td>595</td>
<td>301</td>
<td>Diosmetin arabinosyl-glucoside</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>25.552</td>
<td>C_{28}H_{32}O_{15}</td>
<td>609</td>
<td>301</td>
<td>Diosmetin O-rhamnosyl-glucoside (diosmin)</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>26.061</td>
<td>C_{21}H_{20}O_{10}</td>
<td>433</td>
<td>287</td>
<td>Kaempferol 3-O-rhamnoside</td>
<td>(Al-Shammari et al., 2015; Jordon-Thaden &amp; Louda, 2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>27.037</td>
<td>C_{21}H_{20}O_{10}</td>
<td>433</td>
<td>285</td>
<td>Diosmetin acetyl diglycoside</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>27.037</td>
<td>C_{21}H_{20}O_{10}</td>
<td>433</td>
<td>285</td>
<td>Diosmetin diacetyl diglycoside</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>30.232</td>
<td>C_{15}H_{10}O_{6}</td>
<td>287</td>
<td>286</td>
<td>Luteolin</td>
<td>(Bain &amp; Desrochers, 1988; Dimitrova-dyulgerova et al., 2015; Jordon-Thaden &amp; Louda, 2003; Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>32.678</td>
<td>C_{15}H_{10}O_{6}</td>
<td>287</td>
<td>286</td>
<td>Diosmetin diacetyl diglycoside</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>33.168</td>
<td>C_{15}H_{10}O_{5}</td>
<td>271</td>
<td>270</td>
<td>Apigenin</td>
<td>(Al-Shammari et al., 2015; Dimitrova-dyulgerova et al., 2015; Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>33.569</td>
<td>C_{15}H_{10}O_{6}</td>
<td>287</td>
<td>286</td>
<td>Kaempferol</td>
<td>(Al-Shammari et al., 2015; Dimitrova-dyulgerova et al., 2015; Jordon-Thaden &amp; Louda, 2003; Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>33.64</td>
<td>C_{15}H_{10}O_{6}</td>
<td>301</td>
<td>299</td>
<td>Diosmetin</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>33.64</td>
<td>C_{15}H_{10}O_{6}</td>
<td>331</td>
<td>330</td>
<td>Tricin</td>
<td>(Li et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Compounds identified by comparison with reference standards.

The Identification Confidence value is in agreement with the CAWG (2007) guidelines and indicates: Level 1: Identified compound (A minimum of two independent orthogonal data (such as retention time and mass spectrum) compared directly to an authentic reference standard; Level 2: Putatively annotated compound (similarity of chromatographic and spectral data to published data); Level 3: Putatively characterized class of compounds.

A= C. argyroa; C= C. cephalanthus; M= C. nutans subsp. macrocephalus; P= C. pycnocephalus.
## Supplementary material

**Table S1.** Sites and date of collection, voucher specimens, and GenBank accession numbers of the four *Carduus* species

<table>
<thead>
<tr>
<th>Species</th>
<th>Sites and date of collection</th>
<th>Coordinates</th>
<th>Voucher specimen</th>
<th>GenBank Access. No.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Carduus argyroa</em></td>
<td>Decimomannu, 27 May 2015</td>
<td>(39°17'47.96&quot;N - 8°58'14.95&quot;E)</td>
<td>CAG-803</td>
<td>KY242483</td>
</tr>
<tr>
<td><em>Carduus cephalanthus</em></td>
<td>Capo Testa, 12 June 2015</td>
<td>(41°14'33.80&quot;N – 9°8'49.25&quot;E)</td>
<td>CAG-807</td>
<td>KY242486</td>
</tr>
<tr>
<td><em>Carduus nutans subsp. macrocephalus</em></td>
<td>Gennargentu, 18 June 2015</td>
<td>(39°57'35.77&quot;N - 9°19'12.46&quot;E)</td>
<td>CAG-802</td>
<td>KY242485</td>
</tr>
<tr>
<td><em>Carduus pycnocephalus</em></td>
<td>Monte dei Sette Fratelli, 21 May 2015</td>
<td>(39°20'43.60&quot;N – 9°17'43.74&quot;E)</td>
<td>CAG-805</td>
<td>KY242484</td>
</tr>
</tbody>
</table>

**Table S2.** Multiple Reaction Monitoring acquisitions

<table>
<thead>
<tr>
<th>Compound</th>
<th>MRM (m/z) ESI⁺</th>
<th>MRM (m/z) ESI⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorogenic acid</td>
<td>355 → 163</td>
<td>353 → 191</td>
</tr>
<tr>
<td>Dicaffeoylquinic acid</td>
<td>517 → 163</td>
<td>515 → 191</td>
</tr>
<tr>
<td>Apigenin</td>
<td>271 → 119</td>
<td>271 → 153</td>
</tr>
<tr>
<td></td>
<td>271 → 163</td>
<td></td>
</tr>
<tr>
<td>Luteolin</td>
<td>287 → 127</td>
<td>287 → 135</td>
</tr>
<tr>
<td></td>
<td>287 → 137</td>
<td>287 → 153</td>
</tr>
<tr>
<td></td>
<td>287 → 161</td>
<td></td>
</tr>
<tr>
<td>Kaempferol</td>
<td>287 → 121</td>
<td>287 → 137</td>
</tr>
<tr>
<td></td>
<td>287 → 153</td>
<td>287 → 165</td>
</tr>
<tr>
<td>Quercetin</td>
<td>303 → 121</td>
<td>303 → 127</td>
</tr>
<tr>
<td></td>
<td>303 → 137</td>
<td>303 → 153</td>
</tr>
<tr>
<td></td>
<td>303 → 161</td>
<td>303 → 165</td>
</tr>
<tr>
<td>Diosmetin</td>
<td>301 → 258</td>
<td>299 → 256</td>
</tr>
<tr>
<td></td>
<td>301 → 286</td>
<td>299 → 284</td>
</tr>
</tbody>
</table>
Table S3. Sequence homologies of ITS sequences between each pair of *Carduus* sequences are shown

<table>
<thead>
<tr>
<th></th>
<th><em>C. argyroa</em></th>
<th><em>C.cephalanthus</em></th>
<th><em>C. nutans subsp. macrocephalus</em></th>
<th><em>C. pycnocephalus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. argyroa</em></td>
<td>100%</td>
<td>93%</td>
<td>95%</td>
<td>94%</td>
</tr>
<tr>
<td><em>C. cephalanthus</em></td>
<td>93%</td>
<td>100%</td>
<td>92%</td>
<td>92%</td>
</tr>
<tr>
<td><em>C. nutans subsp. macrocephalus</em></td>
<td>95%</td>
<td>92%</td>
<td>100%</td>
<td>94%</td>
</tr>
<tr>
<td><em>C. pycnocephalus</em></td>
<td>94%</td>
<td>92%</td>
<td>94%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Fig. S1. Three dimensional PCA plot of samples of the four *Carduus* species.