Intra-specific variation in the little-known Mediterranean plant Ptilostemon casabonae (L.) Greuter analysed through phytochemical and biomolecular markers

Intra-specific variation in the little-known Mediterranean plant *Ptilostemon casabonae* (L.) Greuter analysed through phytochemical and biomolecular markers

Arianna Marengo[a], Andrea Maxia[b], Cinzia Sanna[b], Manuela Mandrone[c], Cinzia M. Bertea[d*], Carlo Bicchi[a], Barbara Sgorbini[a], Cecilia Cagliero[a], Patrizia Rubiolo[a*]

[a] Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, 10125 Torino, Italy

[b] Dipartimento di Scienze della Vita e dell'Ambiente, sezione di Botanica, Università di Cagliari, Viale Sant'Ignazio da Laconi 13, 09123 Cagliari, Italy

[c] Dipartimento di Farmacia e Biotecnologie, Alma Mater Studiorum - Università di Bologna, Via Irnerio 42, 40126 Bologna, Italy

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

***Corresponding authors**

Abstract

Ptilostemon casabonae (L.) Greuter is a Mediterranean endemism traditionally used for its health-giving properties. Little is known about this species, therefore this study provides additional information about the phytochemical and biomolecular patterns of this plant, to have a combined fingerprint as a taxonomic tool. Several *P. casabona*e specimens were therefore collected from three different sites, two from Sardinia (Italy) and one from Corsica and the hydroalcoholic extracts of their aerial parts were investigated through HPLC-PDA-MS/MS analysis to study the phenolic composition. Quercetin, luteolin, kaempferol, apigenin and diosmetin *O*-glycosides, and caffeoylquinic acid derivatives were found as main components. Samples from the three sites showed similar phenolic profiles, although statistical analyses highlighted some quantitative differences for several compounds.

The biomolecular analysis included amplification and sequencing of *ITS*, *5S-rRNA-NTS* and *psbA* regions. No difference was found in the nucleotides among the *P. casabonae* samples from different geographical origins; however, a comparison with other *Ptilostemon* species sequences from Genbank, revealed an inter-species variability of *ITS* and *psbA* regions.

The combination of the results of the phytochemical and biomolecular studies provide information on *P. casabonae* useful to depict this little-known plant, which can also be applied for future

investigations and to obtain a fingerprint of it. Moreover, the stability of the phenolic profile within the species affords to identify a set of specialised metabolites useful for its chemotaxonomic characterization. At the same time, the stability of the biomolecular profile of *P. casabonae*, and the identification of sequences specific for this species, enables to identify useful biomolecular markers to distinguish it unequivocally.

Keywords: *Ptilostemon casabonae* (L.) Greuter, Compositae, HPLC-PDS-MS/MS, PCR, sequencing, polyphenols, *ITS* sequence, *psbA* sequence, *5S-rRNA-NTS* sequence

1. Introduction

Ptilostemon casabonae (L.) Greuter. (Compositae) (www.theplantlist.org) belongs to the genus *Ptilostemon* Cass., which is part of the larger informal "*Cynara* group", included in the *Carduinae* subtribe (*Cardueae* tribe) (Susanna et al., 2006; Susanna and Garcia-Jacas, 2009; Vilatersana et al., 2010; Barres et al., 2013). The *Ptilostemon* genus is distributed throughout the Mediterranean area, from Crimea and Turkey to the Iberian Peninsula and Morocco. The genus includes 15 species, among them *Ptilostemon casabonae* (L.) Greuter, a central Mediterranean endemism mainly localized in Sardinia (Italy), Corsica and the Hyères islands (France). The limited geographic distribution of *P. casabonae* is peculiar, since usually "thistles" are widespread colonizing weeds (Médail and Verlaque, 1997; Vilatersana et al., 2010; Marengo et al., 2015). *P. casabonae* is a monocarpic perennial spiny herb with sparsely arachnoid-hairy to glabrescent stem and lanceolate to linear-lanceolate leaves, entire or very slightly sinuate (Tutin et al., 1968; Marengo et al., 2015). It is largely spread in Sardinia and Corsica, where it grows spontaneously, preferring dry and open habitats, with riverbed gravel, screes or mining dumps as secondary habitats (Tutin et al., 1968; Ferrauto et al., 2016). *P. casabonae* is the only species of the *Ptilostemon* genus growing in Sardinia and Corsica (Pignatti, 1982). The aerial parts of this species are traditionally used in Sardinia as intestinal antispasmodic. Additionally, the boiled sprouts and immature flower heads are considered edible (Atzei, 2003).

To the best of the authors' knowledge, no phytochemical studies on *P. casabonae* have to date been reported. The literature gives acetylenes, triterpenes, sesquiterpenes and lignans as the main chemical constituents of the *Ptilostemon* species, but no information is available concerning the flavonoid and phenolic acid composition (Bohlmann et al., 1974; Bohlmann and Ziesche, 1980; Bruno et al., 2001; Janaćković et al., 2002; Djordjevic et al., 2008; Di Stefano and Pitonzo, 2012). Conversely, molecular phylogenetic analyses on the *Cardueae* tribe are available, and provide some DNA barcoding sequences of species belonging to the *Ptilostemon* genus including *P. casabonae* from Southern Corsica (Garcia-Jacas et al., 2002; Susanna et al., 2006; Vilatersana et al., 2010; Barres et al., 2013). The aim of the present study is to contribute to improving knowledge of this little-studied species, by combining phytochemical and biomolecular approaches in order to obtain a representative fingerprint of the investigated plant that can be adopted as a taxonomic tool.

Samples of the aerial parts of *P. casabonae* from three different sites, two in Sardinia (Gennargentu and Iglesias) and one in Southern Corsica (Bocca di Tana) were therefore collected and submitted to phytochemical and biomolecular investigation. The combination of these two different approaches

can provide more information on this species and a reliable fingerprint of the plant useful for the unequivocal identification of *P. casabonae* and its distinction from closely related species. Similarities and dissimilarities in the chemical and molecular patterns in the samples from the three harvesting sites were therefore investigated and the variability/stability of the investigated DNA regions might give useful molecular markers of *P. casabonae.*

The hydroalcoholic extracts from this species were analysed by HPLC-PDA-MS/MS to obtain useful information about the non-yet-investigated polyphenolic fraction, whose composition can be used to characterize this endemic species and its intraspecific variations with geographical origin.

At the same time, three specific DNA regions were amplified, sequenced and compared to obtain diagnostic molecular markers for *P. casabonae* and to verify the variability/stability of samples from different geographical origins. The nuclear internal transcribed spacer (*ITS*) and *5S-rRNA-NTS* genes and the chloroplast photosystem II protein D1 (*psbA-trnH*) genes were analysed. With the exception of *5S-rRNA-NTS* region, these sequences are DNA barcoding genes, used to discriminate plant species (Hebert et al., 2003; Galimberti et al., 2013; Dong et al., 2014). Conversely, the *5S-rRNA-NTS* sequence is used for plant DNA fingerprinting, although it is not yet considered a barcoding gene (Gnavi et al., 2010; Kress, 2017; Marengo et al., 2017b). The *PsbA-trnH* and *5S-rRNA-NTS* genes of *P. casabonae* are sequenced here for the first time.

2. Results and discussion

2.1. Phytochemical analysis

2.1.1. Chemical composition by HPLC-PDA-MS/MS

The phytochemical investigation of the hydroalcoholic extracts of *Ptilostemon casabonae* aerial parts from Sardinia (Gennargentu and Iglesias) and Corsica (Bocca di Tana) was carried out by HPLC-PDA-MS/MS and revealed a remarkable abundance of phenolic compounds, and in particular a complex polyphenolic fraction. Moreover, all the investigated extracts show a similar chemical composition (Fig. 1 and Fig. S1).

In agreement with existing reports on the polyphenols of related species, flavonoids are the most representative group, in particular *O*-glycosides of quercetin, luteolin, kaempferol, apigenin and diosmetin, and caffeoylquinic acid derivatives have been reported (Jordon-Thaden and Louda, 2003; González-Tejero et al., 2008; Li et al., 2014; Ha et al., 2015; Ma et al., 2016; Marengo et al., 2017a).

Chromatographic profiles, UV spectra, and mass spectral information, acquired for 10 samples per site, enabled 25 informative compounds to be detected. The UV spectra obtained for each peak provided a first indication of the compound class. In addition, the molecular weight of each peak was obtained from their mass spectral pattern, through the complementary correspondence between the positive and negative pseudomolecular ions in ESI^+ and ESI^-. The Product Ion Scan analysis of the pseudomolecular ions under investigation (in both ESI^+ and ESI^- modes) provided diagnostic fragments for each molecule. Lastly, the identity of 15 compounds in the extract was confirmed by the co-injection of authentic standards, whereas 10 peaks were putatively identified by comparison with literature data (Table 1).

In particular, compounds for which reference standards were not available, were tentatively identified through their tandem mass spectrometry fragmentation pattern, as reported in Marengo et al., 2017a. This approach also provides further structural information on unknown compounds. As an example, all the phenolic acid derivatives (mainly caffeoylquinic acid derivatives), compounds **5**, **8**, **14**, **15**, **18**, **20**, were tentatively identified by their UV spectral data (UV_{max}=330nm) and pseudomolecular diagnostic fragments at 163 m/z and 191 m/z in ESI^+ and ESI^- ionization modes (Marengo et al., 2017a; Martini et al., 2017). *p*-coumaric acid containing chlorogenic and quinic acid derivatives (compounds **17**, **19**) were putatively identified on the basis of their UV maximum absorption at about 310 nm, and pseudomolecular diagnostic fragments in accordance with those reported by Clifford and co-workers (Clifford et al., 2006). Finally, peaks **8** and **12** were tentatively attributed to quercetin *O*-glycosides derivatives, since their UV spectrum and MS/MS fragmentation pattern are similar to those of quercetin derivatives (Marengo et al., 2017a).

The detected compounds have been found in other "thistle" species (Jordon-Thaden and Louda, 2003; Li et al., 2014; Marengo et al., 2017a, 2018), but, to the best of the authors' knowledge this is the first time that these specialised metabolites have been identified in *P. casabonae* and in general in the *Ptilostemon* genus (Bohlmann et al., 1974; Bohlmann and Ziesche, 1980; Bruno et al., 2001; Janaćković et al., 2002; Djordjevic et al., 2008; Di Stefano and Pitonzo, 2012). The analysis of the composition of these specialised metabolites can be very useful in the taxonomic characterization of *P. casabonae*, to discriminate it from other closely related species and to define a representative chemical pattern for the species. Phenolic compounds, in particular flavonoids, have a wide structural diversity, diffuse occurrence, chemical stability and have been already isolated in large scale from Compositae species and used as taxonomic markers at lower hierarchical levels (Emerenciano et al., 2001; Míka et al., 2005). Moreover, polyphenols have been shown to be diagnostic markers to characterize species within *Hieracium rohacsense* Kit. ex Kit. group (Švehlíková et al., 2002), *Rosa*

L. species belonging to the section *Gallicanae* (Sarangowa et al., 2014), species for the genus *Drosera* L. (Braunberger et al., 2015), *Salvia* L. species (Navaz Kharazian, 2014), *Rhodiola* L. species (Liu et al., 2013) , *Camellia sinensis* (L.) Kuntze hybrids (Li et al., 2010). Phenolic compounds are not only of interest for their taxonomical value but also for their biological activity. The literature describes their anti-inflammatory, antibacterial, antioxidant, and enzyme inhibitory activities (Jordon-Thaden and Louda, 2003; Jeong et al., 2008; Kuklev et al., 2013; Marrelli et al., 2013; Al-Shammari et al., 2015; Tacchini et al., 2015; Marengo et al., 2018).

2.1.2. Statistical analysis

Since the *P. casabonae* samples investigated showed very similar polyphenolic profiles, Principal Component Analysis (PCA), an unsupervised multivariate data analysis method, was applied to evaluate the potential for discrimination among samples from different sites, using the 25 detected compounds listed in Table 1 as variables (Marengo et al., 2017a).

Figure 2 reports score and loading plots. The first component (PC1), which explained 30.4% of the variation, individually separated samples from Corsica from those from Sardinia, although samples from Corsica were not clustered in a homogeneous group (Fig. 2A). Discrimination between samples from the two Sardinian sites (Gennargentu and Iglesias) was less evident. Figure 2A shows that the second component (26.5% of the variation) barely separated samples from Gennargentu and Iglesias. The loading plot illustrates the influence of the variables in sample distribution (Fig. 2B). The PCA results showed that twelve compounds with higher and lower weights in PC1 and PC2 (highlighted in fig. 2B with black circles and squares, respectively) could be selected, quantified in each individual, and submitted to ANOVA analysis, to determine whether there were statistically significant differences among *P. casabonae* samples from Sardinia and Corsica. The variables for which co-elution occurred were quantified by SRM (Selected Reaction Monitoring) acquisition, based on specific transitions in ESI^+ for each compound. The SRM quantification showed that some compounds were present in trace amounts in the extracts (tr<0.05 w/v); luteolin and diosmetin were therefore not included in the following ANOVA statistical analysis.

Among the variables with lower weights in PC1, compounds **12** (quercetin hexoside), **24** (kaempferol) and **22** (quercetin) were significantly more abundant in statistical terms in samples from Corsica than in those from Sardinia, while compound **11** (apigenin 7-O-glucoside) was more abundant in Sardinian samples (in particular from Gennargentu) (Fig. 3). At the same time, compounds **14** (succinyl-dicaffeqyoilquinic acid), **18 (**succinyl-dicaffeoylquinic acid), and **19** (succinyl-*p*-coumaroyl-caffeoylquinic acid/succinyl-caffeoyl-*p*-coumaroylquinic acid), positively correlated with PC1, were more abundant in samples from Sardinia. Although samples from Iglesias

and Gennargentu were not easily discriminated, variables with positive and negative correlations with PC2 were selected, in order to determine similarities or dissimilarities between the two Sardinian sites. Among the three compounds positively correlated with PC2, compounds **9** (1,5 dicaffeoylquinic acid) and **4** (quercetin 3-*O*-glucoside) had significantly higher concentrations in samples from Gennargentu than in those from Iglesias; the same occurred for compound **5** (dicaffeoylquinic acid), although the concentration difference was not statistically significant. At the same time, compounds **15** (succinyl-dicaffeoylquinic acid), **23** (apigenin) and **20** (succinyl-succinyl dicaffeoylquinic acid), which were negatively correlated to PC2, were significantly more abundant in statistical terms in samples from Iglesias than in those from Gennargentu. These results indicate that, although *P. casabonae* samples from different sites have similar chemical compositions, there are some differences in the relative abundance of common compounds, in particular between samples from Sardinia and Corsica. The quantitative differences in Sardinian samples are slight, and thus not sufficient to state that samples from the two Sardinian sites can clearly be discriminated. These few intra-species quantitative differences may depend on the geographical origin and on environmental factors such as temperature, humidity, etc. (Labra et al., 2004; Rahali et al., 2016). A biomolecular analysis was then performed, to seek specific DNA regions not influenced by the environment as markers for this species, to reliably evaluate the stability of *P. casabonae*.

2.2. Sequence analysis of *P. casabonae* deriving from the three sites

In this study, two nuclear regions (*ITS*, *5S-rRNA-NTS*) and one mitochondrial gene (*psbA-trnH*) were amplified and sequenced for each *P. casabonae* site. *ITS* and *PsbA* regions are usually used as barcoding genes to discriminate and identify plant samples. The *ITS* region includes the partial sequences of the *18S* and *28S* ribosomal RNA genes, and the complete sequences of the internal transcribed spacers 1 and 2 and the *5.8S* ribosomal RNA gene, while *psbA-trnH* consists of the noncoding intergenic spacer of the photosystem II protein D1 (Wu et al., 2011; Liu et al., 2014). The 5S-*rRNA-NTS* region is not considered a barcoding gene, but it has recently been used to evaluate the phylogenetic relationships of some higher plant species. It refers to the non-transcribed spacer of the gene linked to the *5S-rRNA*, a ribosome component (Rubiolo et al., 2009; Gnavi et al., 2010; Marengo et al., 2017b).

The data obtained in this study may be useful to integrate missing information in the GenBank database, in order to obtain biomolecular markers able to discriminate the species and to verify the intra-species stability or variability of these sequences.

The nucleotide composition and the alignment of all the resulting sequences is shown in Fig.S2. *ITS*, *5s-rRNA-NTS* and *PsbA* sequences are 729, 324, 504 bp, respectively. The alignment of the three site sequences suggests a stability of these biomolecular markers for *P. casabonae* species, since no nucleotide variation was detected among sequences originating from the sites on the two islands where the species is considered endemic (Sardinia and Corsica).

A *P. casabonae ITS* sequence from Corsica (GU907728.1) is already available in GenBank. It is thus interesting to evaluate similarities or differences with those from Sardinia and from a different Corsican locality. The alignment of the *ITS* sequences obtained in this study with those from Southern Corsica: Solaro (Musaei Parisiensis) (BC) shows 98% of query cover and 100% of identity with the present data; this confirms the stability of this region for *P. casabonae* (Zhang et al., 2000; Vilatersana et al., 2010).

A comparison with *ITS* sequences belonging to other *Ptilostemon* species, taken from GenBank, was also made. The *P. casabonae ITS* sequence appears more similar (about 7-8 nucleotide position variations, 99% of identity) to those of *P. abylensis* (Maire) Greuter, *P. dyricola* (Maire) Greuter, *P. rhiphaeus* (Pau & Font Quer) Greuter, and *P. strictus* (Ten.) Greuter. The other *Ptilostemon* available sequences show lower percentages of identity (Vilatersana et al., 2010).

Comparison of the *P. casabonae PsbA* sequence to that of the only *Ptilostemon PsbA* sequence (*P. afer* AF129850.1) available in the database shows two variable sites at positions 296 and 335. A thymine and an adenine are present in the *P. afer* (Jacq.) Greuter sequence, substituted, at the same positions, by a guanine and a cytosine in the samples studied here. A gap is also present at the far end of the sequence.

These data confirm that the *ITS* region could be considered a good molecular marker for *P. casabonae*, since it shows inter-species variability and an intra-species stability. *PsbA*, likewise, presents sequence stability among samples from the same species, and nucleotide variation among different species; however, few sequences are present in the database, and it would be interesting to increase the dataset to confirm the potential of *PsbA* as barcoding gene for *Ptilostemon* species.

Unfortunately, no other *5S-rRNA-NTS* sequences are available in GenBank, making it impossible to compare them to the nucleotide composition of this region from other species. This could be of interest for further investigations on this gene, since this is a documented variable region, used also to distinguish species varieties (Rubiolo et al., 2009; Gnavi et al., 2010; Sun et al., 2014).

3. Conclusions

The study reports the first phytochemical investigation of the hydroalcoholic extract of the aerial parts of the Mediterranean endemic species *P. casabonae*. Interesting compounds were detected in terms of polyphenols, chiefly flavonoids and phenolic acid derivatives and their presence may support the traditional use of this herb. The study also provides useful biomolecular markers for this plant, enabling *P. casabonae* to be discriminated from other *Ptilostemon* species.

Despite some differences in abundance, the compounds present in extracts from the three sites were qualitatively the same. Nor was any difference found in the nucleotide composition of the investigated *P. casabonae* DNA regions from Sardinia and Corsica, confirming that the compound abundance variability is due to environmental factors.

The combination of chemical and biomolecular patterns was found to be extremely useful to provide a reliable *P. casabonae* fingerprint that can be used as a taxonomic tool, thanks to the intra-specific geographical stability of the species. Moreover, the similar chemical pattern of samples from different sites may be a positive feature for future applications of this plant in the food and well-being fields, although further investigations will be necessary.

4. Experimental section

4.1. Chemicals

HPLC-grade acetonitrile (LC-MS grade), formic acid (> 98% purity), quercetin, apigenin, kaempferol and rutin 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 apigenin 7-*O*-glucoside were from Extrasynthese (Genay Cedex, France). Chlorogenic acid, cynarin, 1,5-dicaffeoylquinic acid, 3,5-dicaffeoylquinic acid, 4,5-dicaffeoylquinic acid, luteolin 7-*O*-glucoside, luteolin 7-*O*-glucuronide, apigenin 7-*O*-glucuronide and diosmetin were from Phytolab (Vestenbergsgreuth, Germany).

4.2. Plant material

P. casabonae (L.) Greuter (Compositae) aerial parts were collected from three natural populations: two sites in Sardinia (Italy) and one site in Corsica (France). In detail, samples from Sardinia were collected from Gennargentu (OG, 39°53'54.9"N –9°26'27.9"E) and Iglesias (SU, 39°21'45.8"N - 8°32'24.0"E) in June 2017. *P. casabonae* from Corsica was kindly collected by Prof. Felix Tomi (Université de Corse Pascal Paoli) from Bocca di Tana (Southern Corsica, 41°45.791’N ; 009°02.300’E) in June 2017. Ten specimens were collected from each site. The fresh plant material was dried at 40°C to constant weight. The plants were identified at the Department of Life and

Environmental Sciences, University of Cagliari, Italy, where a voucher specimen was deposited (CAG-796).

4.3. Extraction method

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), thrice, for 10 min each. The resulting three extracts were combined and centrifuged at 4000 rpm for 10 min. The supernatant was brought to a volume of 30 mL and filtered through a 13 mm diameter, 0.22 μm pore diameter, hydrophilic PTFE syringe filter, prior to HPLC-PDA-MS/MS analysis. Each extract was analysed three times and the repeatability of extract composition was evaluated on at least three plants per site.

4.4. HPLC-PDA-MS/MS analysis and quantification

The chromatographic analysis was conducted on a Shimadzu Nexera ×2 system equipped with a SPD-M20A photodiode detector in series to a triple quadrupole Shimadzu LCMS-8040 system provided with electrospray ionization (ESI) source (Shimadzu, Dusseldorf Germany) as reported by Marengo et al., 2017a, without any modification. For the UV quantification, apigenin (λ=340nm) and quercetin 3-*O*-glucoside (λ=350nm) were used for quantification of the relative compounds. 1,5 dicaffeoylquinic acid (λ=330nm) and p-coumaric acid (λ=310nm) were used for quantification of dicaffeoylquinic acid derivatives and coumaroyl-caffeoylquinic acid derivatives, respectively. Only when co-elution was present, quantification by Selected Reaction Monitoring (SRM) acquisition in ESI+ (collision energy: - 35.0 V for ESI+, dwell time: 20) on specific ion products derived from precursor ion fragmentation was performed. The transitions, used for the quantification, were based on the Product Ion Scan (PIS) analysis of each standard (Fig. S3). Apigenin 7-*O*-glucoside (433→271), luteolin (287→153), quercetin (303→153), kaempferol (287→153) and diosmetin (301→258) were used to quantify the relative compounds, quercetin 3-*O*-glucoside (465→303) was used to quantify quercetin hexoside since the two compounds showed a similar fragmentation pattern (Fig. S3). Each standard solution and the extracts were analysed in three replicates. The calibration curve of 1,5 dicaffeoylquinic acid was prepared with seven different concentrations in the range of 0.1–200 μg/mL. Five different concentrations were prepared for all other compounds, in the range of 0.1–25 μg/mL for all of them with the exception of *p*-coumaric acid/diosmetin and apigenin 7-*O*-glucoside/luteolin whose concentration range was; 0.1–10 μg/mL and 0.1-5 μg/mL, respectively. The determination coefficients were in the range of 0.996-1. The calibration curve equations were as follows: apigenin (UV): y = 69736x - 1125,9; quercetin 3-*O*-glucoside (UV): y = 23785x + 1206,7;

1,5 dicaffeoylquinic acid (UV): y = 56041x – 137821; *p*-coumaric acid (UV): y = 190076x + 113336; apigenin 7-*O*-glucoside (SRM): y = 3E+06x + 762083; quercetin 3-*O*-glucoside (SRM): y = 1E+06x + 163544; luteolin (SRM): y = 408509x + 268263; quercetin (SRM): y = 202793x - 6021,7; kaempferol (SRM): y = 253441x + 116146; diosmetin (SRM): y = 2E+06x + 542654.

4.5. DNA extraction, PCR amplification and sequencing

Ten milligrams of leaves from each specimen were ground to a fine powder, with the addition of approximately 5 mg polyvinylpolypyrrolidone (PVPP, Sigma Aldrich, Bellefonte, USA). Genomic DNA was isolated from the ground powder using the Nucleospin Plant II Kit (Macherey Nagel, Düren, Germany) following the manufacturer's instructions. The quantitative and qualitative analyses of the isolated genomic DNA were assessed by spectrophotometry using a 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 and reverse primers specific for the amplification of the *ITS*, *5S-rRNA-NTS* and *PsbA* regions (Table S1). Amplification was performed in a 25 μL reaction mixture containing 2.5 μL of 10× 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, Jena, Germany). Cycling conditions consisted of an initial 4 min at 94 °C, followed by 30 s denaturing at 94 °C, 45 s annealing, respectively at 53°, 52°, and 56°C for *ITS*, *5s-rRNA-NTS*, and *PsbA*, and 45 s elongation 72 °C, repeated for 35 cycles and with 10 min 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 employed as a template for sequencing (Eurofins Genomics, Vimodrone (MI), Italy). Both DNA strands were sequenced.

4.6. Statistical analysis

Peak areas (integrated at λ=330nm and scaled to unit variance) obtained from the chemical analysis were used for the Principal Component Analysis (PCA), to reduce the multivariate space over which objects were distributed.

Abundance differences of characteristic compounds identified by PCA, among samples, were determined by analysis of variance (ANOVA). One-way ANOVA and post hoc multiple means comparison (Tukey's range test) were carried out and statistical significance was accepted at $p<0.05$. Both PCA and ANOVA were evaluated using SPSS 15.0 (IBM Corporation) software.

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 site, were then aligned by modifying the Gap Opening and Gap Extension Cost values to 15 and 1, respectively.

Acknowledgments

The authors are indebted with Prof. Felix Tomi (Université de Corse Pascal Paoli) for providing *P. casabonae* from Corsica. The authors also acknowledge support from the Ricerca Locale (Ex 60% 2018) of the University of Turin, Turin (Italy).

Figure captions

Figure 1. Representative profiles of *P. casabonae* hydroalcoholic extracts from different sites (λ=330nm, intensity range: 0-800000 mAU). For compound identities see Table 1. (A) *P. casabonae* from Gennargentu (OG); (B) *P. casabonae* from San Benedetto (SU); (C) *P. casabonae* from Bocca di Tana (Southern Corsica)

Figure 2. Principal Component Analysis (PCA) of 30 samples from different *P. casabonae* sites based on the 25 informative compounds as variables. (A) Score plot of samples from the three *P. casabonae* sites; (B) loading plot of the variables (compound numbers refer to Table1). Black circle and squares indicate potential discriminating variables for the three sites. (C) Score plot legend.

Figure 3. Concentration (w/v) of discriminating compounds selected by PCA analysis, in extracts from each site. Bars indicate standard deviation of at least two technical replicates of 10 individuals per site. Different letters indicate significant differences at $p<0.05$ for each compound (Tukey range test).

Table1. List of identified and putatively-identified compounds in *P. casabonae* extracts. For each compound the relative retention time, UV spectrum, pseudomolecular ions, fragment ions obtained by Product Ion Scan mode (PIS) and identified or tentatively-identified compound names are given. Identification Confidence values and references are also included.

N°	RT	λ max (nm)	Mol. weight g/mol	MS^2+ m/z	MS^2- m/z	Compound name	IC[b]	References
1[a]	8.691	325	354	163	191	5-*O*-Caffeoylquinic acid (Chlorogenic Acid)	1	(Li et al., 2014; Dimitrova-dyulgerova et al., 2015)
2[a]	15.068	329	516	163	191	1,3 Dicaffeoylquinic acid (Cynarin)	1	(Li et al., 2014; Dimitrova-dyulgerova et al., 2015)
3[a]	21.207	254 352	610	303	301	Quercetin 3-*O*-rutinoside (Rutin)	1	(Li et al., 2014; Dimitrova-dyulgerova et al., 2015)
4[a]	21.726	260 352	464	303	301	Quercetin 3-*O*-glucoside	1	(Li et al., 2014)
5	23.512	325	516	163	191	Dicaffeoylquinic acid	3	(Marengo et al., 2017a)
6	23.707	253 350	550	303 159 127 109	336 301 271 135 121	Quercetin malonylhexoside	2	(Marengo et al., 2017a)
7[a]	23.932	265 344	594	287	285	Kaempferol 3-*O*-rutinoside	1	(Marengo et al., 2017a)
8	23.932	330	632	163 145	335 191 133	Phenolic acid derivative	3	(Clifford et al., 2006)
9[a]	24.318	329	516	163	191	1,5 Dicaffeoylquinic acid	1	(Li et al., 2014; Marengo et al., 2018)
10[a]	24.588	329	516	163	191	3,5 Dicaffeoylquinic acid	1	(Li et al., 2014; Marengo et al., 2018)
11[a]	25.042	337 266	432	271	269	Apigenin 7-*O*-glucoside	1	(Li et al., 2014)
12	25.042	255 351	464	303	301	Quercetin hexoside	3	(Marengo et al., 2017a)

13[a]	25.322	335 266	446	271	269	Apigenin 7-*O*-glucuronide	1	(Li et al., 2014)
14	25.736	330	616	163	191	Succinyl-dicaffeoylquinic acid	2	(Marengo et al., 2018)
15	26.165	330	616	163	191	Succinyl-dicaffeoylquinic acid	2	(Marengo et al., 2018)
16[a]	26.373	329	516	163	191	4,5 Dicaffeoylquinic acid	1	(Li et al., 2014; Marengo et al., 2018)
17	26.965	313	500	163;145;135;119	337 191 163	*p*-coumaroyl-caffeoylquinic acid/caffeoyl-*p*-coumaroylquinic acid	2	(Clifford et al., 2006; Li et al., 2014)
18	27.807	330	616	163	191	Succinyl-dicaffeoylquinic acid	2	(Marengo et al., 2018)
19	28.331	319	600	163;147	337 191 173	Succinyl-*p*-coumaroyl-caffeoylquinic acid/Succinyl-caffeoyl-*p*-coumaroylquinic acid	2	(Clifford et al., 2006; Li et al., 2014; Marengo et al., 2018)
20	29.188	325	716	163	191	Succinyl-succinyl Dicaffeoylquinic acid	2	(Marengo et al., 2018)
21[a]	30.321	349 252 266	286	135;137;153;161		Luteolin	1	(Li et al., 2014; Dimitrova-dyulgerova et al., 2015)
22[a]	30.321	369 254	302	127;137;153;165		Quercetin	1	(Li et al., 2014)
23[a]	33.178	336 267	270	119;153;163		Apigenin	1	(Li et al., 2014; Dimitrova-dyulgerova et al., 2015)
24[a]	33.565	365 265	286	121;137;153;165		Kaempferol	1	(Li et al., 2014; Dimitrova-dyulgerova et al., 2015)
25[a]	33.565	345 267	300	258;286	256 284	Diosmetin	1	(Li et al., 2014; Marengo et al., 2017a)

[a] Compounds identified by comparison with reference standards.

[b] 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

References

Al-Shammari, L.A., Hassan, W.H.B., Al-Youssef, H.M., 2015. Phytochemical and biological studies of *Carduus pycnocephalus* L. J. Saudi Chem. Soc. 19, 410–416.

Atzei, A.D., 2003. Le piante nella tradizione popolare della Sardegna. Carlo Delfino Ed.

Barres, L., Sanmartín, I., Anderson, C.L., Susanna, A., Buerki, S., Galbany-Casals, M., Vilatersana, R., 2013. Reconstructing the evolution and biogeographic history of tribe *Cardueae* (compositae). Am. J. Bot. 100, 867–882.

Bohlmann, F., Rao, N., Schwarz, H., 1974. Über zwei neue Sesquiterpene aus *Ptilostemon chamaepeuce* (L.) Less. Chem. Ber. 107, 650–653.

Bohlmann, F., Ziesche, J., 1980. Neue guajanolide und acetylenverbindungen aus *Ptilostemon*-*arten*. Phytochemistry 19, 692–696.

Braunberger, C., Zehl, M., Conrad, J., Wawrosch, C., Strohbach, J., Beifuss, U., Krenn, L., 2015. Flavonoids as chemotaxonomic markers in the genus *Drosera*. Phytochemistry 118, 74–82.

Bruno, M., Maggio, A., Paternostro, M.P., Rosselli, S., Arnold, N.A., Herz, W., 2001. Sesquiterpene lactones and other constituents of three *Cardueae* from cyprus. Biochem. Syst. Ecol. 29, 433–435.

Clifford, M.N., Marks, S., Knight, S., Kuhnert, N., 2006. Characterization by LC-MS^n of Four New Classes of p -Coumaric Acid-Containing Diacyl Chlorogenic Acids in Green Coffee Beans. J. Agric. Food Chem. 54, 4095–4101.

Di Stefano, V., Pitonzo, R., 2012. Phytochemical Studies on *Ptilostemon greuteri* Raimondo & Domina (Compositae). Rec. Nat. Prod. 6, 390–393.

Dimitrova-dyulgerova, I., Zhelev, I., Mihaylova, D., 2015. Phenolic profile and in vitro antioxidant activity of endemic Bulgarian *Carduus* species. Pharmacogn. Mag. 11, 575–579.

Djordjevic, I., Tesevic, V., Milosavljevic, S., Menkovic, N., Vajs, V., Jadranin, M., Jovanovic, A., Djokovic, D., 2008. Composition of the essential oil of *Ptilostemon gnaphaloides* 44, 538–539.

Dong, W., Cheng, T., Li, C., Xu, C., Long, P., Chen, C., Zhou, S., 2014. Discriminating plants using the DNA barcode rbcLb : an appraisal based on a large data set. Mol. Ecol. Resour. 14, 336–343.

Emerenciano, V.P., Militão, J.S.L.T., Campos, C.C., Romoff, P., Kaplan, M.A.C., Zambon, M., Brant, A.J.C., 2001. Flavonoids as chemotaxonomic markers for Asteraceae. Biochem. Syst. Ecol. 29, 947–957.

Ferrauto, G., Costa, R., Pavone, P., 2016. Pollen diversity in the genus *Ptilostemon* (Asteraceae, Cardueae) from Italy and its taxonomic and palynoecological implications. Plant Biosyst. - An Int. J. Deal. with all Asp. Plant Biol. 3504, 1–15.

Galimberti, A., Mattia, F. De, Losa, A., Bruni, I., Federici, S., Casiraghi, M., Martellos, S., Labra, M., 2013. DNA barcoding as a new tool for food traceability. Food Res. Int. 50, 55–63.

Garcia-Jacas, N., Garnatje, T., Susanna, A., Vilatersana, R., 2002. Tribal and subtribal delimitation and phylogeny of the *Cardueae* (Asteraceae): a combined nuclear and chloroplast DNA analysis. Mol. Phylogenet. Evol. 22, 51–64.

Gnavi, G., Bertea, C.M., Maffei, M.E., 2010. PCR , sequencing and PCR –RFLP of the 5S-rRNA-NTS region as a tool for the DNA fingerprinting of medicinal and aromatic plants. Flavour Fragr. J. 25, 132–137.

González-Tejero, M.R., Casares-Porcel, M., Sánchez-Rojas, C.P., Ramiro-Gutiérrez, J.M., Molero-Mesa, J., Pieroni, A., Giusti, M.E., Censorii, E., de Pasquale, C., Della, A., Paraskeva-Hadijchambi, D., Hadjichambis, A., Houmani, Z., El-Demerdash, M., El-Zayat, M., Hmamouchi, M., ElJohrig, S., 2008. Medicinal plants in the Mediterranean area: Synthesis of the results of the project Rubia. J. Ethnopharmacol. 116, 341–357.

Ha, I.J., Lee, M.Y., Kwon, Y.-K., Jung, Y., Kim, H.K., Hwang, G.-S., 2015. Metabolite profiling to discriminate different species and genus from thistles in Korea using liquid chromatography with quadrupole time-of-flight mass spectrometry. J. Sep. Sci. 38, 502–510.

Hebert, P.D.N., Cywinska, A., Ball, S.L., Jeremy, R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. London 270, 313–321.

Janaćković, P., Tešević, V., Marin, P.D., Milosavljević, S.M., Petković, B., Soković, M., 2002. Polyacetylenes and a sesquiterpene lactone from *Ptilostemon strictus*. Biochem. Syst. Ecol. 30, 69–71.

Jeong, D.M., Jung, H.A., Choi, J.S., 2008. Comparative antioxidant activity and HPLC profiles of some selected Korean thistles. Arch. Pharm. Res. 31, 28–33.

Jordon-Thaden, I.E., Louda, S.M., 2003. Chemistry of *Cirsium* and *Carduus*: A role in ecological risk assessment for biological control of weeds? Biochem. Syst. Ecol. 31, 1353–1396.

Kress, W.J., 2017. Plant DNA barcodes: Applications today and in the future. J. Syst. Evol. 55, 291–307.

Kuklev, D. V., Domb, A.J., Dembitsky, V.M., 2013. Bioactive acetylenic metabolites. Phytomedicine 20, 1145–1159.

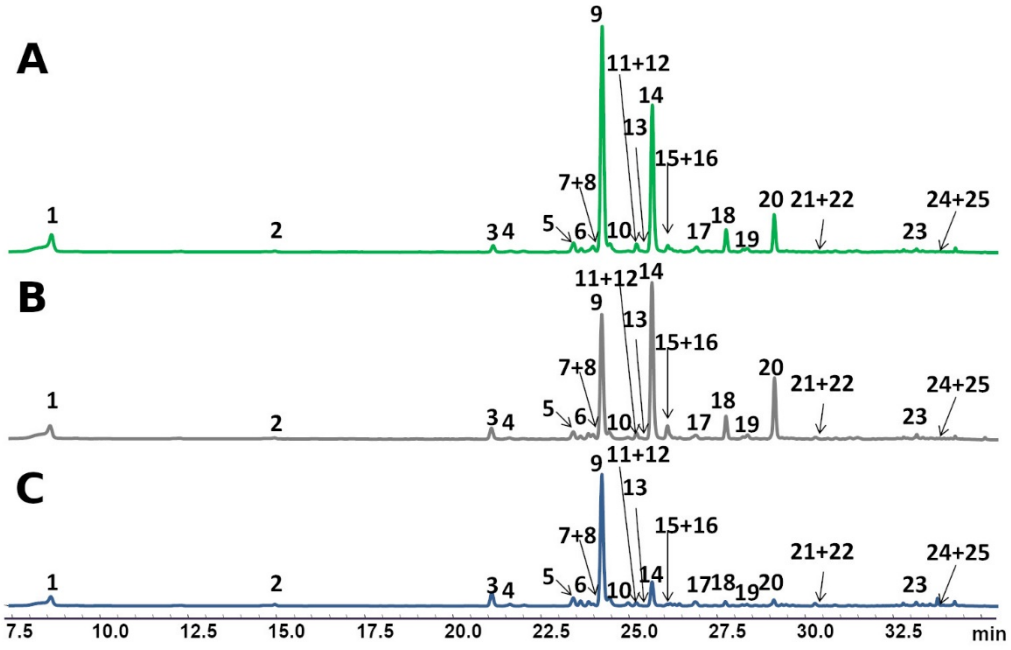
Labra, M., Miele, M., Ledda, B., Grassi, F., Mazzei, M., Sala, F., 2004. Morphological

442 characterization , essential oil composition and DNA genotyping of *Ocimum basilicum* L .
443 cultivars 167, 725–731.
444 Li, J.-H., Nesumi, A., Shimizu, K., Sakata, Y., Liang, M.-Z., He, Q.-Y., Zhou, H.-J., Hashimoto, F.,
445 2010. Chemosystematics of tea trees based on tea leaf polyphenols as phenetic markers.
446 Phytochemistry 71, 1342–1349.
447 Li, R., Liu, S., Song, W., Wang, Y., Li, Y., Qiao, X., Liang, H., Ye, M., 2014. Chemical analysis of
448 the Tibetan herbal medicine *Carduus acanthoides* by UPLC/DAD/qTOF-MS and simultaneous
449 determination of nine major compounds. Anal. Methods 6, 7181.
450 Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., Zhou, X., Meng, F., Downie, S.R., 2014.
451 Identification of species in the angiosperm family Apiaceae using DNA barcodes. Mol. Ecol.
452 Resour. 14, 1231–1238.
453 Liu, Z., Liu, Y., Liu, C., Song, Z., Li, Q., Zha, Q., Lu, C., Wang, C., Ning, Z., Zhang, Y., Tian, C.,
454 Lu, A., 2013. The chemotaxonomic classification of *Rhodiola* plants and its correlation with
455 morphological characteristics and genetic taxonomy. Chem. Cent. J. 7, 118.
456 Ma, Q., Wang, L.-H., Jiang, J.-G., 2016. Hepatoprotective effect of flavonoids from *Cirsium*
457 *japonicum* DC on hepatotoxicity in comparison with silymarin. Food Funct. 7, 2179–2184.
458 Marengo, A., Fenu, G., Gennai, M., Cogoni, D., Fois, M., Bacchetta, G., 2015. Schede per una Lista
459 Rossa della Flora vascolare e crittogamica italiana. *Ptilostemon casabonae* (L.) Greuter. Inf.
460 Bot. Ital. 47, 245–289.
461 Marengo, A., Fumagalli, M., Sanna, C., Maxia, A., Piazza, S., Cagliero, C., Rubiolo, P.,
462 Sangiovanni, E., Dell'Agli, M., 2018. The hydro-alcoholic extracts of Sardinian wild thistles
463 (*Onopordum* spp.) inhibit TNFα-induced IL-8 secretion and NF-κB pathway in human gastric
464 epithelial AGS cells. J. Ethnopharmacol. 210, 469–476.
465 Marengo, A., Maxia, A., Sanna, C., Bertea, C.M., Bicchi, C., Ballero, M., Cagliero, C., Rubiolo, P.,
466 2017a. Characterization of four wild edible *Carduus* species from the Mediterranean region via
467 phytochemical and biomolecular analyses. Food Res. Int. 100, 822–831.
468 Marengo, A., Piras, A., Falconieri, D., Porcedda, S., Caboni, P., Cortis, P., Foddis, C., Loi, C.,
469 Gonçalves, M.J., Salgueiro, L., Maxia, A., 2017b. Chemical and biomolecular analyses to
470 discriminate three taxa of *Pistacia* genus from Sardinia Island (Italy) and their antifungal
471 activity. Nat. Prod. Res. 6419, 1–9
472 Marrelli, M., Loizzo, M.R., Nicoletti, M., Menichini, F., Conforti, F., 2013. Inhibition of Key
473 Enzymes Linked to Obesity by Preparations From Mediterranean Dietary Plants: Effects on α-
474 Amylase and Pancreatic Lipase Activities. Plant Foods Hum. Nutr. 68, 340–346.

475 Martini, S., Conte, A., Tagliazucchi, D., 2017. Phenolic compounds pro fi le and antioxidant
476 properties of six sweet cherry (*Prunus avium*) cultivars. Food Res. Int. 97, 15–26.
477 Médail, F., Verlaque, R., 1997. Ecological characteristics and rarity of endemic plants from
478 Southeast France and Corsica: Implications for biodiversity conservation. Biol. Conserv. 80,
479 269–281.
480 Míka, V., Kubáň, V., Klejdus, B., Odstrčilová, V., Nerušil, P., 2005. Phenolic compounds as
481 chemical markers of low taxonomic levels in the family Poaceae. Plant, Soil Environ. 51, 506–
482 512.
483 Navaz Kharazian, 2014. Chemotaxonomy and flavonoid diversity of *Salvia* L. (Lamiaceae) in Iran.
484 Acta Bot. Brasilica 28, 281–292.
485 Pignatti, S., 1982. La flora d'Italia. Edagricole.
486 Rahali, F.Z., Lamine, M., Gargouri, M., Rebey, I.B., Hammami, M., Sellami, I.H., 2016. Metabolite
487 profiles of essential oils and molecular markers analysis to explore the biodiversity of *Ferula*
488 *communis*: Towards conservation of the endemic giant fennel. Phytochemistry 124, 58–67.
489 Rubiolo, P., Matteodo, M., Bicchi, C., Gnavi, G., Bertea, C., Maffei, M., 2009. Chemical and
490 Biomolecular Characterization of *Artemisia umbelliformis* Lam., an Important Ingredient of
491 the Alpine Liqueur " Genepi#." J. Agric. Food Chem. 57, 3436–3443.
492 Sarangowa, O., Kanazawa, T., Nishizawa, M., Myoda, T., Bai, C., Yamagishi, T., 2014. Flavonol
493 glycosides in the petal of *Rosa* species as chemotaxonomic markers. Phytochemistry 107, 61–
494 68.
495 Sun, Y., Kang, H., Kim, Y., Baek, J., Zheng, S., Xiang, J., Hong, S.-K., 2014. Tomato (*Solanum*
496 *lycopersicum*) variety discrimination and hybridization analysis based on the 5S rRNA region.
497 Biotechnol. Biotechnol. Equip. 28:3, 431–437.
498 Susanna, A., Garcia-Jacas, N., 2009. *Cardueae* (Carduoideae). In: Funk VA, Susanna A, Stuessy
499 TF, Bayer RJ, editors. Systematics, evolution and biogeography of the Compositae. Vienna
500 IAPT 293–313.
501 Susanna, A., Hidalgo, O., Vilatersana, R., Ciencia, M. De, 2006. The *Carduae* (Compositae)
502 revisited: insights from a combined ITS, trnL-trnF and matK nuclear and chloroplast DNA
503 analysis. Annu. Missouri Bot. Gard. 93, 150–171.
504 Švehlíková, V., Mráz, P., Piacente, S., Marhold, K., 2002. Chemotaxonomic significance of
505 flavonoids and phenolic acids in the *Hieracium rohacsense* group (Hieracium sect. Alpina;
506 Lactuceae, Compositae). Biochem. Syst. Ecol. 30, 1037–1049.
507 Tacchini, M., Spagnoletti, A., Marieschi, M., Caligiani, A., Bruni, R., Efferth, T., Sacchetti, G.,

508 Guerrini, A., 2015. Phytochemical profile and bioactivity of traditional ayurvedic decoctions
509 and hydro-alcoholic macerations of *Boerhaavia diffusa* L. and *Curculigo orchioides* Gaertn.
510 Nat. Prod. Res. 29, 2071–2079.
511 Tutin, T.G., Heywood, V.H., Burges, N.A., Valentine, D.H., Walters, S.M., Webb, D.A., 1968.
512 Flora europaea. Cambridge Univ. Press.
513 Vilatersana, R., Garcia-Jacas, N., Garnatje, T., Molero, J., Sonnante, G., Susanna, A., 2010.
514 Molecular Phylogeny of the Genus *Ptilostemon*(Compositae: *Cardueae*) and its Relationships
515 with *Cynara* and *Lamyropsis*. Syst. Bot. 35, 907–917.
516 https://doi.org/10.1600/036364410X539952
517 Wu, P.K., Tai, W.C.S., Choi, R.C.Y., Tsim, K.W.K., Zhou, H., Liu, X., Jiang, Z.H., Hsiao, W.L.W.,
518 2011. Chemical and DNA authentication of taste variants of *Gynostemma pentaphyllum* herbal
519 tea. Food Chem. 128, 70–80.
520 Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA
521 sequences. J. Comput. Biol. 7, 203–214.
522

523 Figure 1

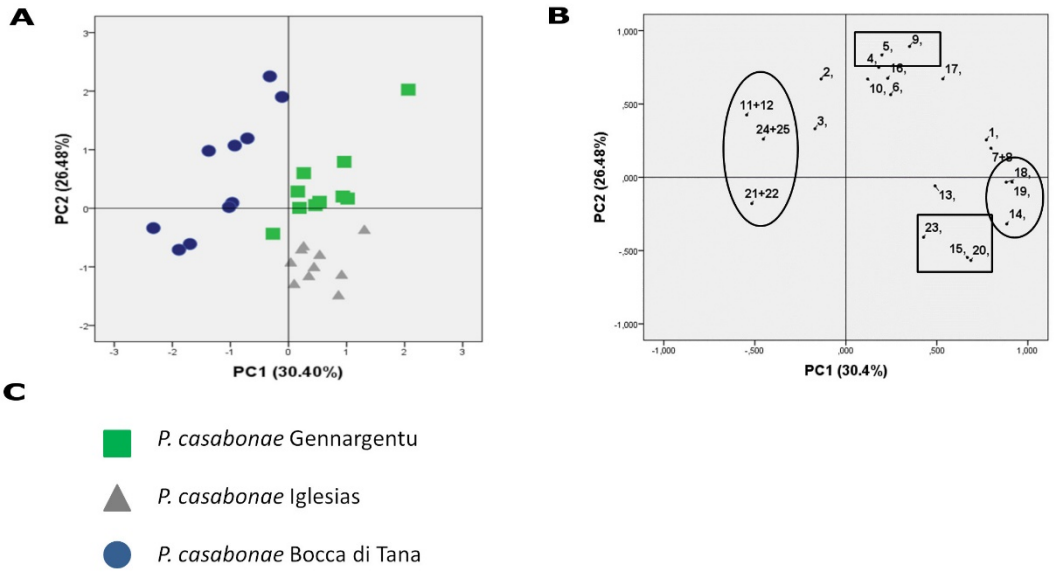


524

525

526

527 Figure 2



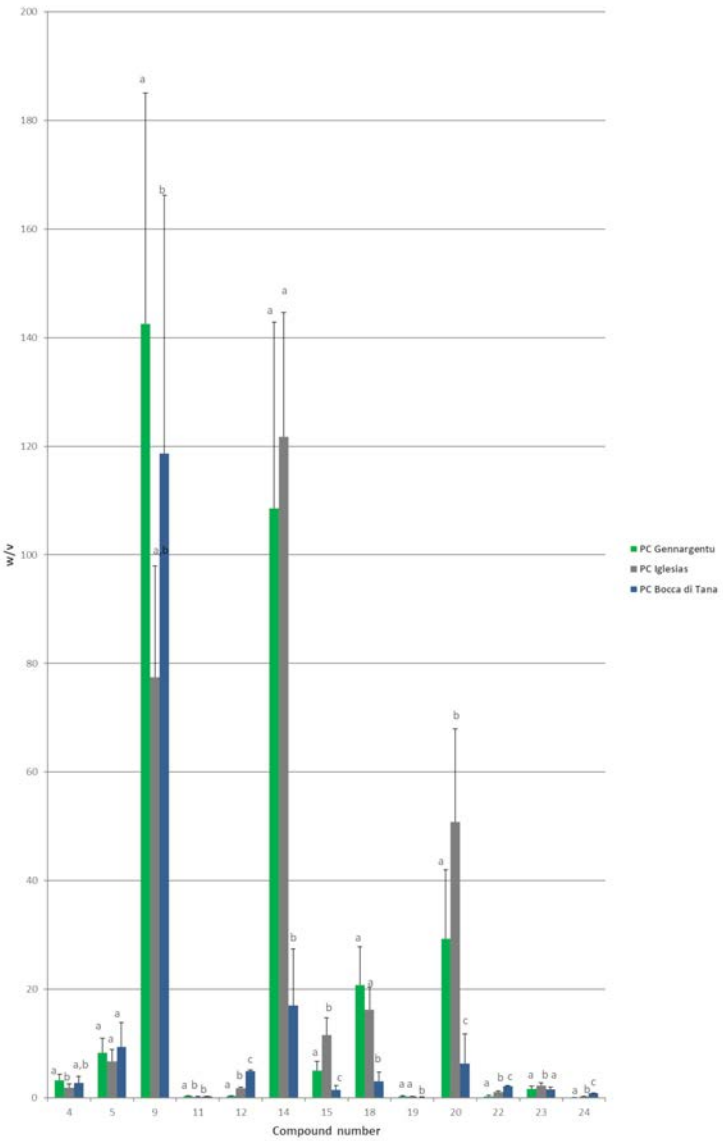
528

529

530

531 Figure 3

532



533

534

535

536

537