

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Artemisia umbelliformis Lam. and Génépi Liqueur: Volatile Profile as Diagnostic Marker for Geographic Origin and to Predict Liqueur Safety**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1633099> since 2017-05-16T15:44:37Z

*Published version:*

DOI:10.1021/acs.jafc.6b03394

*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:

Boggia, Lorenzo; Pignata, Giuseppe; Sgorbini, Barbara; Colombo, Maria Laura; Marengo, Arianna; Casale, Manuela; Nicola, Silvana; Bicchi, Carlo; Rubiolo, Patrizia. *Artemisia umbelliformis* Lam. and G n pi Liqueur: Volatile Profile as Diagnostic Marker for Geographic Origin and to Predict Liqueur Safety. *JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY*. 65 (13) pp: 2849-2856.  
DOI: 10.1021/acs.jafc.6b03394

The publisher's version is available at:

<http://pubs.acs.org/doi/pdf/10.1021/acs.jafc.6b03394>

When citing, please refer to the published version.

Link to this full text:

<http://hdl.handle.net/2318/1633099>

1     ***Artemisia umbelliformis* Lam. and génépi liqueur: volatile profile as diagnostic marker for**  
2                                   **geographic origin and to predict liqueur safety**

3     Lorenzo Boggia<sup>†</sup>, Giuseppe Pignata<sup>§</sup>, Barbara Sgorbini<sup>†</sup>, Maria Laura Colombo<sup>†</sup>, Arianna  
4     Marengo<sup>‡</sup>, Manuela Casale<sup>§</sup>, Silvana Nicola<sup>§</sup>, Carlo Bicchi<sup>†</sup>, Patrizia Rubiolo<sup>†\*</sup>

5     <sup>†</sup> Department of Drug Science and Technology (DSTF), University of Turin, Via Pietro Giuria 9,  
6     10125 Torino, Italy

7     <sup>§</sup> Vegetable Crops & Medicinal and Aromatic Plants, VEGMAP; Department of Agricultural,  
8     Forest and Food Sciences, DISAFA; University of Turin. Via Leonardo da Vinci, 44 - Largo Paolo  
9     Braccini, 2, 10095 Grugliasco (TO), Italy

10    <sup>‡</sup> Department of Life and Environmental Sciences, University of Cagliari, Viale S. Ignazio da  
11    Laconi 13, 09124 Cagliari, Italy

12  
13    \* Phone: +39 011 6707662. Fax: +39 011 2367661. E-mail: patrizia.rubiolo@unito.it.

14

15 **ABSTRACT**

16 *A. umbelliformis*, commonly known as "white g n pi", is characterized by a volatile fraction rich in  
17  $\alpha$ - and  $\beta$ - thujone, two monoterpenoids; under EU regulations these are limited to 35 mg/L in  
18 *Artemisia*-based beverages , because of their recognized activity on human central nervous system.  
19 This study reports the results of an investigation to define the geographical origin and thujone  
20 content of individual plants of *A. umbelliformis* from different geographical sites, cultivated  
21 experimentally at a single site, and to predict the thujone content in the resulting liqueurs, through  
22 their volatile fraction. Headspace Solid Phase Microextraction (HS-SPME) combined with Gas  
23 Chromatography-Mass Spectrometry (GC-MS) and non-separative HS-SPME-MS were used as  
24 analytical platforms to create a database suitable for chemometric description and prediction  
25 through Linear Discriminant Analysis (LDA). HS-SPME-MS was applied to shorten analysis time.  
26 With both approaches, a diagnostic prediction of: i) plant geographical origin, and ii) thujone  
27 content of plant-related liqueurs could be made.

28

29 Keywords: *Artemisia umbelliformis* Lam., g n pi liqueur, food safety

30

31

32

## 33 INTRODUCTION

34 *Artemisia* L., with its 500 specific and sub-specific *taxa*, is the largest genus in the tribe *Anthemidae*  
35 of the *Asteraceae* family<sup>1,2</sup>. Its *taxa* grow worldwide at different latitudes, altitudes, and  
36 environments, and have traditionally been used for their biological activity<sup>1</sup>, even before recent  
37 pharmacological developments<sup>3-5</sup>.

38 *Artemisia* species were used historically to prepare infusions, wines, and liqueurs<sup>6</sup>. The best-known  
39 species is *A. absinthium*, traditionally known as wormwood and in widespread use since Roman  
40 times as the base for aromatic wines and liqueurs<sup>6,7</sup>. Wormwood has been studied in depth because  
41 of its diffusion and ready accessibility<sup>6</sup>, but several other *Artemisia* species are also used in  
42 beverages. A group of herbs used for liqueur production, traditionally known as "génépi", consists  
43 of five rare species growing at high altitudes in the alpine area: *A. eriantha* Ten., *A. genipi* Weber,  
44 *A. glacialis* L., *A. nivalis* Br.-Bl. and *A. umbelliformis* Lam.. Génépi species traditionally have less  
45 economic importance than wormwood, being almost exclusively used within their production and  
46 harvesting areas; however, over the last 40 years, liqueur production have markedly increased, not  
47 least thanks to a number of small and medium-sized producers<sup>6</sup>. Génépi species have historically  
48 been used in folk and traditional medicine because of their bioactivity<sup>6</sup>. They were known and used  
49 as thermogenic agents against the common cold , in infusions against fever, and in aromatic wines  
50 and liqueurs, to stimulate the appetite and the digestion,<sup>8-10</sup> respectively. The extensive ethno-  
51 pharmacological knowledge, documented from the second half of the eighteenth century, contrasts  
52 with the paucity of scientific studies of génépi species<sup>6</sup>.

53 Liqueur production on a larger scale is not possible if the raw material is limited to the wild plant  
54 harvest. Génépi cultivation, which is restricted by the plants' pedoclimatic needs, facilitates  
55 harvesting operations and protects wild génépi species<sup>6</sup>. The five génépi species have a common  
56 environmental distribution, growing at around 1500-2000 m a.s.l.<sup>11</sup>. The altitude requirement is due  
57 to specific soils in which these species face neither competition from weeds nor fungal pathogens,  
58 which become predominant at lower altitudes<sup>6,12</sup>.

59 *A. umbelliformis*, commonly known as "white génépi", is the génépi species that has been cultivated  
60 more successfully, above 1600 m a.s.l.<sup>6,11,12</sup>; this species has therefore been selected for wider-scale  
61 cultivation. Independently of its domestication and production, the supply chain of *A. umbelliformis*  
62 faces two issues linked to the food industry: product adulteration and product safety. The need for  
63 an authentication process emerges, since it is one of the few Italian infusion liqueurs whose  
64 geographic limits are defined in the EU by Reg. EC 110/2008<sup>13</sup>. This regulation introduced three  
65 categories for génépi, distinguishing the most important production areas situated in North-West  
66 Italy and France: "Genepi della Valle d'Aosta (Italy)", "Genepi del Piemonte (Italy)", and "Génépi  
67 des Alpes/Genepi delle Alpi (France, Italy)".

68 Thujones are natural monoterpenoids widely present in *Artemisia* species.  $\alpha$ - and  $\beta$ - thujone are the  
69 most abundant volatiles in génépi species<sup>14-16</sup>, and are characterized by recognized activity on the  
70 human central nervous system<sup>11,17,18</sup>. For this reason, the maximum content of thujone in *Artemisia*-  
71 based beverages is limited to 35 mg/L in the EU (Reg. EC 110/2008<sup>6,13,18</sup>). *A. umbelliformis* is the  
72 génépi species with the highest content of thujones<sup>19</sup>. The variability of the plant's chemical  
73 composition (i.e. the variability of thujone content) is a crucial point for génépi liqueur producers.  
74 To overcome this problem, some agronomical selection has been done to create thujone-free  
75 chemotypes<sup>11,18</sup>. However, chemical comparisons between thujone-free and thujone-containing  
76 chemotypes have found a marked difference in the content of both aromatic and bitter compounds,  
77 responsible for the typical sensory characteristics of the liqueur<sup>18</sup>.

78 Chemical analysis combined with chemometrics can be useful to deal with the problems emerging  
79 from large-scale production<sup>20-25</sup>. In the present study, Linear Discriminant Analysis (LDA)<sup>26</sup> was  
80 adopted for its ability to perform class discrimination, instead of Principal Component Analysis  
81 (PCA), which is widely used to describe large datasets. LDA is known to elucidate the underlying  
82 data structure, a key point when creating a prediction model<sup>24,27</sup>. Some PCA elaborations were,  
83 however, carried out and a comparison of the results to those of LDA confirm that the latter  
84 technique has a better prediction capacity (data not shown).

85 This research studied *A. umbelliformis* of different geographical origins, cultivated experimentally  
86 at a single site. Headspace sampling with Solid Phase Microextraction, combined with gas  
87 chromatography - mass spectrometry (HS-SPME-GC-MS) or directly with mass spectrometry (HS-  
88 SPME-MS), was used as an analytical platform to create the first database suitable for chemometric  
89 description and prediction. Linear Discriminant Analysis (LDA) was here adopted to define and  
90 predict geographic origin and thujone content of plants and liqueurs  
91 Multiple Headspace- Solid Phase Microextraction (MHS-SPME)<sup>26,28,29</sup> was applied for thujone  
92 content determination.

93

## 94 **MATERIAL AND METHODS**

95 **Chemicals.** Thujones standard mixture (mixture of  $\alpha$ -thujone and  $\beta$ -thujone, purity = 99.9%) and all  
96 other reference compounds were from Sigma-Aldrich (St. Louis, MO). Sabinol was kindly supplied  
97 by Robertet SA (Grasse, France), and sabinyl ester homologous series were synthesized in the  
98 authors' laboratory<sup>18</sup>. HPLC and analytical grade solvents were from Carlo Erba Reagenti, Rodano,  
99 Italy.

100 **Plant material and experimental site preparation.** Growing fields of *A. umbelliformis* were  
101 monitored in spring and summer at various alpine sites in Piedmont in North-West Italy (**Table 1**).  
102 The sites are located in the South-West Alps within the area indicated for the liqueur "Genepi del  
103 Piemonte"<sup>13</sup>. At the end of the blossoming stage, four *A. umbelliformis* geographic origins were  
104 selected (**Table 1**). Plant growing on the verge of the raised beds were excluded, to avoid the edge  
105 effect. Mother plants were chosen based on the phenological profile (data not shown) and then  
106 carefully harvested to preserve inflorescences and seeds.

107 Plants were dried at 30°C in a dry and aerated room at the Experimental Centre of DISAFA  
108 (44°53'11.67"N; 7°41'7.00"E - 231 m a.s.l.) in Tetti Frati, Carmagnola (TO), Italy. The seeds were  
109 cleaned to remove extraneous parts, under fan ventilation using a series of sieves of different  
110 meshes, by Semina S.r.l. (La Loggia (TO), Italy), and then stored at 6°C in a seed room. The seeds

111 of the four different geographic groups were sown the following summer in trays of a specific peat-  
112 based horticultural medium (Neuhaus Huminsubstrat N17; Klasmann-Deilmann® GmbH, Geeste,  
113 Niedersachsen, Germany). The trays were placed in an air-conditioned greenhouse providing the  
114 necessary supply of water until the seeds germinated. After cotyledon expansion, *A. umbelliformis*  
115 was transplanted into 104-cell plastic trays (0.522m × 0.322m × 0.042m; 0.034m upper diameter  
116 cells), filled with the above peat medium. Trays were placed in an air-conditioned plastic  
117 greenhouse and seedlings were irrigated overhead as required. After secondary leaf development,  
118 trays were placed under a plastic film tunnel covered with black shading systems (50%-shade  
119 cloth). Before transplanting the *A. umbelliformis* seedlings into the soil, the shade cloth was  
120 removed to harden the plants and acclimatize them to outdoor environmental conditions. The count  
121 of dead plants was performed weekly (data not shown). Plants were transplanted in the autumn at  
122 the experimental site of Pragelato (TO), Italy (45°00'36.41"N; 6°56'14.56"E - 1,500 m a.s.l.). The  
123 site is characterized by a soil with a moderate presence of skeleton, absence of slope, and optimal  
124 sun exposition. After minimum tillage, a white/black (upper/lower side) plastic mulching film  
125 provided with holes arranged in a quincunx formation (0.05 m hole diameter, 0.20 m distance  
126 between holes) was used to set 6 mulched raised beds subdivided into 12 plots (4 geographic groups  
127 (Gran Paradiso, GP; Val Chisone; Elva; Valle Gesso, VG) ×3 blocks) for a total of ca. 1,350 *A.*  
128 *umbelliformis* plants.

129 **Production, phenological and morphological parameters.** The site was inspected to check *A.*  
130 *umbelliformis* stand establishment with a ≈ 30 days frequency. The number of dead plants was  
131 monitored and recorded 30 days after transplanting, the following summer and autumn (1<sup>st</sup> year),  
132 and again the subsequent summer (2<sup>nd</sup> year). The percentage of dead plants was calculated and used  
133 as an indicator of stand establishment and the plants' adaptation to the site.

134 Phenological and morphological parameters were measured in the 1<sup>st</sup> summer and autumn and in  
135 the 2<sup>nd</sup> summer after transplanting, on 12 plants per geographic origin and per block, randomly  
136 chosen at the beginning of the experiment. Inflorescence parameters were only measured during the



137 2<sup>nd</sup> summer of experimentation. The phenological parameters measured on the canopy were: canopy  
138 diameter, canopy height (**Figure 1A**); on the inflorescences: inflorescence height, inflorescence  
139 number (**Figure 1B**). The morphological parameters measured on the canopy leaves were: number  
140 of segments, leaf length, leaf segment length and width (**Figure 1C**); on the inflorescence leaves:  
141 cauline leaf length, cauline leaf petiole length (**Figure 1D**); on the inflorescences: footstalk length,  
142 number of glomerules (at first, second and third internode), number of plant glomerules, number of  
143 head glomerules, total number of glomerules (**Figure 1E**).

144 Harvesting took place in the 2<sup>nd</sup> summer after transplanting, during the blossoming stage. Fresh  
145 weight production per plant was measured, and weight after drying at 30°C until constant weight,  
146 dry matter percentage was calculated as in the European Pharmacopoeia<sup>30</sup>.

147 **Headspace Solid Phase Microextraction (HS-SPME) sampling and analysis conditions:**  
148 **qualitative profiling of plants.**

149 **Separative analysis.** The 104 selected plants surviving at the end of the 2<sup>nd</sup> summer were used to  
150 analyze the volatile fraction. Portions of 25 mg were sampled, as the minimum plant weight  
151 required to have both an adequate informative profile and good repeatability with headspace solid-  
152 phase microextraction (HS-SPME).

153 Each plant was sampled in a 20 mL headspace vial (2 replicates of 25 mg) using a 2 cm  
154 DVB/CAR/PDMS fiber (Supelco, Bellefonte, PA, USA). Sampling was carried out with a MPS2  
155 automatic sampling system (Gerstel, Mülheim a/d Ruhr, Germany) at T=50 °C (15 min of pre-  
156 equilibrium, 15 min of sampling). The fiber was automatically transferred to the GC injector of a  
157 7890 GC coupled to a 5975C MS (Agilent, Little Falls, DE, USA) from where the sampled analytes  
158 were recovered directly by thermal desorption and transferred to the GC column for analysis.

159 *GC-MS conditions.* Inlet T=250°C, split injection (5 min, 1/10 split ratio); Helium was used as  
160 carrier gas at a flow rate of 1 mL/min. Column: MEGA5MS (30 m × 0.25 mm i.d. × 0.25 µm;  
161 MEGA, Legnano, MI, Italy). Temperature program: 50 °C (1 min)-3 °C/min-250 °C (5 min). MS  
162 operated in EI mode at 70 eV with a mass range from 35 to 350 amu in full scan mode.

163 Data were processed with Agilent MSD ChemStation D.03.00.611 (Agilent Technologies).  
164 Components were identified by comparing their linear retention indices ( $I^T$ s) (calculated *versus* a  
165 C<sub>9</sub>-C<sub>25</sub> hydrocarbon mixture) and their mass spectra to those of authentic standards.

166 The HS-SPME-GC-MS method repeatability was tested on ten samples in five replicates for the 30  
167 selected compounds, reaching an overall value of relative standard deviation (RSD%) of 9.3%. As  
168 reported in **Supplementary Table 1**, the minimum RSD% was detected with sabinyl valerianate  
169 (2.5%) and the maximum RSD% with sabinene (29.3%).

170 **Non-separative analysis.** The same plant samples analyzed by HS-SPME-GC-MS were also  
171 analyzed without performing GC separation. Sampling was carried out on 2 replicates of 150 mg of  
172 each plant weighted in a 20 mL headspace vial using a 2 cm DVB/CAR/PDMS fiber (Supelco,  
173 Bellefonte, PA, USA) at 100 °C for 5 min. Sampling and injection were managed by a Gerstel  
174 MPS2 automatic system (Gerstel, Mülheim a/d Ruhr, Germany). The fiber was automatically  
175 transferred to the GC injector of a 7890 GC coupled to a 5975C MS (Agilent, Little Falls, DE,  
176 USA) and the sampled analytes thermally desorbed directly into the empty tubing for MS analysis.

177 *MS analysis conditions.* Inlet T:250°C, split injection (5 min, 1/20 split ratio); carrier gas: helium,  
178 flow rate: 1 mL/min. MS transfer column: deactivated fused silica empty tubing (8 m × 0.25 mm  
179 i.d.; MEGA, Legnano, MI, Italy). Oven temperature: 250 °C (5 min). MS operated in EI mode at 70  
180 eV with a mass range from 35 to 250 amu in full scan mode.

181 MS data were processed with Agilent MSD ChemStation D.03.00.611 (Agilent Technologies) and  
182 Pirouette 4.0.2 (Infometrix Inc., Bothell, WA, USA).

183 The HS-SPME-MS method repeatability was tested on ten samples in five replicates for the 30 most  
184 abundant fragment ions, reaching an overall relative standard deviation (RSD%) of 1.8%. As  
185 reported in **Supplementary Table 2**, the minimum RSD% was found for  $m/z=43$  (0.6%) and the  
186 maximum for  $m/z=107$  (2.5%).

187 **Multiple headspace solid phase microextraction (MHS-SPME-GC-MS): quantitation of**  
188 **thujones in plants and liqueurs.** Thirty six samples of *A. umbelliformis* among the 104 samples  
189 employed in the previous steps were selected to produce experimental liqueurs. For each plant, one  
190 individual liqueur was prepared, following the guidelines for liqueur preparation given in Annex II  
191 (32) of EC Regulation No 110/2008<sup>13</sup>. MHS-SPME was then applied to the liqueurs (three  
192 replicates of 10 µL for each sample introduced in a 20 mL headspace vial, three HS-SPME  
193 samplings for each replicate).

194 Beside liqueur analysis, each plant sample was also analyzed by MHS-SPME-GC-MS (three  
195 replicates of 3 mg for each sample weighted in a 20 mL headspace vial, three HS-SPME samplings  
196 for each replicate).

197 Quantitative calibration was carried out by MHS-SPME using seven concentration levels of  $\alpha$ - and  
198  $\beta$ -thujone standard mixture (92.6%  $\alpha$ -thujone, 7.4%  $\beta$ -thujone) between 0.1 g/L and 10 g/L in  
199 cyclohexane (two replicates of 10 µL for each level, three HS-SPME samplings for each  
200 replicate)<sup>31</sup>.

201 *Sampling conditions.* MHS-SPME analyses were carried out using a 2 cm  
202 divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber (Supelco, Bellefonte, PA,  
203 USA) at 60 °C for 30 min. Sampling and injection were managed by a Gerstel Multipurpose  
204 Sampler 2 (MPS2) (Gerstel, Mülheim a/d Ruhr, Germany). The fiber was automatically transferred  
205 to a 7890 GC unit coupled to a 5975C MS to analyze the sampled analytes (Agilent, Little Falls,  
206 DE, USA). During MHS-SPME sampling, the sample was not heated between sampling, but placed  
207 back into the sample tray after each sampling.

208 *GC-MS conditions.* Inlet T:250°C, split injection (5 min, 1/20 split ratio); carrier gas: helium, flow  
209 rate: 1 mL/min. Column: MEGA-SE52 (30 m × 0.25 mm i.d. × 0.25 µm; Mega, Legnano, MI,  
210 Italy). Temperature program: 50°C-3°C/min-105°C-20°C/min-250°C (1 min). MS operated in EI  
211 mode at 70 eV with a mass range from 35 to 350 amu in full scan mode. Data were processed with  
212 Agilent MSD ChemStation D.03.00.611 (Agilent Technologies).

213 **Statistical data treatment.** Preliminary descriptive, significance analyses of variance (ANOVA)  
214 and linear regression attempts were run on MHS-SPME-GC-MS data. Linear Discriminant Analysis  
215 (LDA)<sup>27,32</sup> was applied to correlate MHS-SPME-GC-MS data with HS-SPME-GC-MS data and  
216 HS-SPME-MS data. Statistical analyses were performed by using SPSS 15.0 (IBM Corporation) or  
217 R<sup>33</sup>; RStudio<sup>34</sup> was used to write a LDA cycling script. Other R packages used were: caret<sup>35</sup>,  
218 MASS<sup>36</sup> and e1071<sup>37</sup>.

219 LDA describes and predicts multivariate data matrices in which samples are described by one  
220 categorical variable and  $n$  continuous variables. The LDA model calculates Discriminant Functions  
221 (DFs) which are used first to describe the group distribution in the analyzed dataset and then to  
222 predict the groups of samples. DFs are calculated on the basis of the relationships between the  
223 continuous variables describing the dataset and the categorical variables defining the group of each  
224 sample. DFs can be considered similar to the Principal Components in PCA, but they differ for  
225 some characteristics, such as their meaning (they maximize distances between the groups of the  
226 dataset instead of maximizing the dataset explained variance), and for their number (PC number =  
227 variable number; DF number = group number *minus* one)<sup>27</sup>.

228 LDA was carried out on four different matrices based on different data: i) plant geographic origin as  
229 categorical variable and plant morphological and phenological data as continuous variables; ii) plant  
230 geographic origin as categorical variable and plant HS-SPME-GC-MS compound abundances as  
231 continuous variables; iii) liqueur MHS-SPME-GC-MS thujone level as categorical variable and  
232 plant HS-SPME-GC-MS compound abundances as continuous variables; iv) liqueur MHS-SPME-  
233 GC-MS thujones level as categorical variable and plant HS-SPME-MS fragment ions abundances as  
234 continuous variables.

235

## 236 **RESULTS AND DISCUSSION**

237 The study comprised three main steps: i) to describe statistically-significant differences relating to  
238 the geographical distribution of cultivated *A. umbelliformis* plants; ii) to develop and apply

239 analytical methods to rapidly quantify thujone content of *A. umbelliformis*-based liqueur (génépi);  
240 iii) to create multivariate statistical models to predict the above two points.

241 The morphological/phenological dataset and the chemical dataset were processed with LDA .

242 **Origin discrimination through phenological and morphological data.** At the end of the 2<sup>nd</sup> year  
243 of field tests, 61.8% of the *A. umbelliformis* plants survived without showing significant differences  
244 between plants from different geographical origins (data not shown). Morphological data collected  
245 in the 2<sup>nd</sup> year included both vegetative and reproductive parts of the 104 plant samples (**Figure 1,**  
246 **Supplementary Table 3**).

247 ANOVA was carried out, when possible ( $p < 0.05$ ), together with the Tukey Honestly Significant  
248 Difference (HSD) test. The measurements related to the length of footstalks and to the number of  
249 glomerules were found to be the most significant parameters in differentiating plant individuals  
250 originating from different geographical areas (**Supplementary Table 4**).

251 LDA was used to describe the morphological differentiation of the plant groups and was carried out  
252 in two steps: cross validation (CV) and prediction. The dataset was divided into two subsets.  
253 Approximately 60% of the samples were used to create the prediction model based on three  
254 Discriminant Functions (DF). Subsequently, the residual 40% of the samples were predicted by the  
255 three Discriminant Functions, in order to calculate the predictive power of the LDA model.

256 LDA was carried out considering phenological and morphological parameters, and the three DFs  
257 explain the total variance (DF1 = 51.0%; DF2 = 38.6%; DF3 = 10.4%). The scatter plot reported in  
258 **Figure 2** does not show a clear distribution of the samples in function of their geographical origin. .

259 However Elva and Valle Gesso are separated along DF2 DF1 and DF2 showed higher correlations  
260 with the variables related to the length of the footstalk and the number of glomerules. The variables  
261 related to the canopy and the leaves showed a higher correlation with DF3, which, however, did not  
262 provide significant results (**Supplementary Table 5**).

263 Discriminant Functions can be used not only to describe the phenological and morphological  
264 differences, but also to predict the origin of the plant samples.

265 LDA usually reports the success of a prediction model with a confusion matrix that describes the  
266 recognition success of the group prediction performed on the basis of DFs. Cross validation and  
267 prediction results report the percentage of the positive prediction divided by group.

268 **Table 2** shows that the LDA model recognizes samples' geographical origin in the range of 66.7-  
269 77.8% in cross validation and 29.6%-51.9% in prediction. This result shows that the use of LDA  
270 with morphological and phenological data does not give satisfactory results.

271 **Differentiation of plant origin by HS-SPME-GC-MS. Figure 3** reports the HS-SPME-GC-MS  
272 profile of an *A. umbelliformis* sample with thirty identified compounds used to create the  
273 multivariate data matrix. Similarly to the morphological data, LDA was applied to find differences  
274 between plants with different geographic origins.

275 Each of the three Discriminant Functions calculated was significant, unlike the morphological  
276 dataset. The scatter plot of the plant samples presents a distribution similar to the one obtained with  
277 the morphological data along DF1 (**Figure 4**). Groups are better separated, especially Elva (EL) and  
278 Valle Gesso (VG), while Gran Paradiso (GP) and Val Chisone still significantly overlap. DF3 helps  
279 to differentiate GP from VC (**Supplementary Table 5**).

280 The improvement with HS-SPME-GC-MS data was considerable, especially in terms of prediction  
281 rates obtained with the LDA model. As reported in the related confusion matrix (**Table 3**),  
282 classification rates were calculated to be above 90% (mean value) both in validation (min=88.9%;  
283 max=100.0%; mean=93.7%) and in prediction (min=82.6%; max=96.3%; mean=90.3%).

284 **Quantitation of thujones in plants and liqueurs by MHS-SPME-GC-MS.** MHS-SPME-GC-MS  
285 analysis was used to quantify  $\alpha$ - and  $\beta$ -thujone in plants and liqueurs and, then, to find a correlation  
286 between the total thujone amount in plants and liqueurs. Thirty-six plant and liqueur samples were  
287 analyzed in three replicates, by performing three subsequent HS-SPME-GC-MS analyses on each  
288 replicate. Total  $\alpha$ - and  $\beta$ -thujone GC peak area estimation of each replicate was quantified by MHS-  
289 SPME-GC-MS via external calibration, analyzing  $\alpha$ - and  $\beta$ -thujone standard mixture at seven

290 concentration levels. ( $\alpha$ -thujone:  $y = 2205.2x - 88987.3$ ;  $R^2 = 0.9996$ ;  $\beta$ -thujone:  $y = 1977.6x -$   
291  $250263$ ;  $R^2 = 0.9995$ ).

292 The thujone amount in liqueur and plant material was used to calculate a linear regression, which  
293 could then be used to predict the thujone content in liqueurs on the basis of the related plant  
294 material (i.e. before the liqueur preparation, in agreement with the EC protocol<sup>13</sup>).

295 The linear regression model between thujone quantities in plants and liqueurs was not satisfactory  
296 for prediction (**Supplementary Figure 1**). As a consequence, the possibility of predicting the  
297 thujone level in liqueurs had to be investigated by a different approach. LDA was then adopted to  
298 predict the probability of a liqueur to enter (or not) within the thujone EC limit on the basis of the  
299 thujone content in the plant used to produce it.

300 **Thujone level prediction in liqueurs.** A new matrix was created using the total thujone amount as  
301 categorical variable and HS-SPME-GC-MS compound abundances as continuous describing  
302 variables. The samples were divided into two groups based on the total thujone concentration, using  
303 the 35 mg/L EC limit to discriminate the high-thujone group from the low-thujone group.

304 With LDA, only one DF was calculated to discriminate the two groups, because the categorical  
305 response variable was represented by only two possibilities (above and below the EC limit).

306 As reported in **Figure 5**, a clear discrimination between the two groups was possible using HS-  
307 SPME-GC-MS data. This result shows that the samples above the limit have the lowest values in  
308 DF scores, while the samples below the limit have the highest DF scores. The result is clear  
309 considering that the two group distributions did not overlap, giving the model a strong possibility of  
310 recognizing the response group. The model efficiency is supported by the confusion matrix (**Table**  
311 **4**) that reports high validation (<35 ppm=97.6%; >35 ppm=100.0%) and prediction (<35  
312 ppm=95.2%; >35 ppm=92.9%) rates.

313 The combination of HS-SPME-GC-MS data with LDA modeling achieved satisfactory prediction  
314 rates. However, the analytical process (HS-SPME-GC-MS) is still too time-consuming (70  
315 min/sample). The method proposed is reliable in qualitative and quantitative analyses when it is

316 necessary to consider each compound, but it should be made easier and quicker to fit better with the  
317 aim of this study. The analysis can significantly be speeded up by eliminating the GC separation  
318 step and reducing the sampling time to five minutes, thus reducing the total analysis time to eight  
319 minutes.

320 The scan range of the non-separative HS-SPME-MS method has been fixed between 35 and 250  
321 *m/z*. As in the previous LDA model, a matrix was created based on the total thujone amount as  
322 categorical variable, but using the 215 fragment ions as continuous describing variables of each  
323 sample.

324 As reported in **Figure 6** and in **Table 5**, the LDA results on non-separative MS data were slightly  
325 less predictive than those obtained with GC-MS data, but they were still in a significant range of  
326 significance and prediction. The model efficiency for non-separative approach is supported by the  
327 confusion matrix (**Table 5**) that reports good validation (<35 ppm=97.7%; >35 ppm=96.4%;  
328 mean=97.1%) and prediction (<35 ppm=88.6%; >35 ppm=89.3%; mean=89.0%) rates.

329 In conclusion, the combination of chromatographic techniques (HS-SPME-GC-MS, HS-SPME-MS  
330 and MHS-SPME-GC-MS) with a chemometric tool (LDA) has successfully been used to create  
331 prediction models that can help to solve two important issues concerning *A. umbelliformis* and its  
332 related g n pi liqueur. HS-SPME-GC-MS data recovered from a plant sample were found to be  
333 good predictive markers for: i) plant geographical origin, and ii) thujone level in the plant-related  
334 liqueur. Furthermore, a non-separative methods (HS-SPME-MS) was also developed to markedly  
335 shorten the analysis time, while at the same time maintaining its capacity as prediction method for  
336 thujone level in the related liqueur.



337 **ACKNOWLEDGMENT**

338 This study was carried out within the project “Studio di composti biologicamente attivi del  
339 metabolismo secondario di matrici di origine vegetale” financially supported by the Ricerca Locale  
340 (Ex 60% 2015) of the University of Turin, Turin (Italy).

341

342 **ASSOCIATED CONTENT**

343 Supporting Information

344 **Supplementary Figure 1.**

345 Linear regression model between total thujone content of plant and liqueur. A, Correlation between  
346 plant and liqueur MHS-SPME-GC-MS data. B, Correlation between MHS-SPME-GC-MS liqueur  
347 data and plant-based predicted liqueur values

348 **Supplementary Table 1.**

349 Repeatability of HS-SPME-GC-MS analysis of all compounds detected.

350 **Supplementary Table 2**

351 Repeatability of HS-SPME -MS analysis of the 30 most abundant MS fragments.

352 **Supplementary Table 3**

353 ANOVA and Tukey Honestly Significant Difference (HSD) test on morphological parameters.

354 **Supplementary Table 4**

355 Results of LDA performed with morphological parameters. Discriminant Functions are analyzed  
356 testing their significance in discriminating the different geographical origin in the Wilks' Lambda  
357 Table. The second table reports the correlation between the three Discriminant Functions and the  
358 morphological parameters.

359 **Supplementary Table 5**

360 Results of LDA performed with compounds detected by HS-SPME-GC-MS. Discriminant  
361 Functions are analyzed testing their significance in discriminating the different geographical origin  
362 in the Wilks' Lambda Table. The second table reports the DF value for each geographical origin.  
363 The third table reports the correlation between the three Discriminant Functions and the HS-SPME-  
364 GC-MS compounds.

365

366 **REFERENCES**

- 367 (1) Funk, V. A.; Susanna, A.; Stuessy, T. F.; Robinson, H. Classification of Compositae. In  
368 *Systematics, Evolution, and Biogeography of Compositae*; International Association for Plant  
369 Taxonomy, Ed.; Vienna, 2009.
- 370 (2) Hayat, M. Q.; Ashraf, M.; Khan, M. A.; Mahmood, T.; Ahmad, M.; Jabeen, S. Phylogeny of  
371 *Artemisia* L.: Recent developments. *African J. Biotechnol.* **2009**, 8 (11), 2423–2428.
- 372 (3) Klayman, D. L. Qinghaosu (Artemisinin): An antimalarial drug from China. *Science* **1985**,  
373 228 (4703), 1049–1055.
- 374 (4) Xie, D.-Y. *Artemisia annua*, artemisinin, and the Nobel Prize: beauty of natural products and  
375 educational significance. *Sci. Bull.* **2016**, 61 (1), 42–44.
- 376 (5) Molyneux, D. H.; Ward, S. A. Reflections on the Nobel Prize for Medicine 2015 - The Public  
377 Health Legacy and Impact of Avermectin and Artemisinin. *Trends Parasitol.* **2015**, 31 (12),  
378 605–607.
- 379 (6) Vouillamoz, J. F.; Carlen, C.; Tagliatalata-Scafati, O.; Pollastro, F.; Appendino, G. The  
380 g n pi *Artemisia* species. Ethnopharmacology, cultivation, phytochemistry, and bioactivity.  
381 *Fitoterapia* **2015**, 106, 231–241.
- 382 (7) Albert-Puleo, M. Mythobotany, Pharmacology, and Chemistry of Thujone-Containing Plants  
383 and Derivatives. *Econ. Bot.* **1978**, 32 (1), 65–74.
- 384 (8) Delahaye, M.-C. *Le G n pi*; Equinoxe: Saint-R my de Provence, France, 2008.
- 385 (9) Blagojevi , P.; Radulovi , N.; Pali , R.; Stojanovi , G. Chemical Composition of the  
386 Essential Oils of Serbian Wild-Growing *Artemisia absinthium* and *Artemisia vulgaris*. *J.*  
387 *Agric. Food Chem.* **2006**, 54, 4780–4789.
- 388 (10) Lopes-Lutz, D.; Alviano, D. S.; Alviano, C. S.; Kolodziejczyk, P. P. Screening of chemical  
389 composition, antimicrobial and antioxidant activities of *Artemisia* essential oils.  
390 *Phytochemistry* **2008**, 69 (8), 1732–1738.
- 391 (11) Comino, C.; Pignata, G.; Portis, E.; Dolzhenko, Y.; Casale, M.; Nicola, S. Selection in

- 392 Artemisia umbelliformis Lam. Piedmont ecotypes to improve cultivation in alpine  
393 environment. *Genet. Resour. Crop Evol.* **2015**, *62*, 567–577.
- 394 (12) Rey, C.; Slacanin, I. Domestication du genépi blanc (*Artemisia umbelliformis* Lam.). *Rev.*  
395 *suisse Vitic. Arboric. Hortic.* **1997**, *29* (3), 1–8.
- 396 (13) Regulation (EC) No 110/2008 of the European Parliament and of the Council, on the  
397 definition, description, presentation, labelling and the protection of geographical indications  
398 of spirit drinks and repealing Council Regulation (EEC) No 1576/89. *Off. J. Eur. Union*  
399 **2008**, 16–54.
- 400 (14) Appendino, G.; Belliardo, F.; Nano, G. M.; Stefenelli, S. Sesquiterpene lactones from  
401 *Artemisia genipi* Weber: isolation and determination in plant material and in liqueurs. *J.*  
402 *Agric. Food Chem.* **1982**, *30* (3), 518–521.
- 403 (15) Bicchi, C.; D’Amato, A.; Nano, G. M.; Frattini, C. Capillary GLC controls of some alpine  
404 *Artemisiae* and of the related liqueurs. *Chromatographia* **1984**, *18* (10), 560–566.
- 405 (16) Bicchi, C.; Nano, G. M.; Frattini, C. On the composition of the essential oils of *Artemisia*  
406 *genipi* Weber and *Artemisia umbelliformis* Lam. *Z. Lebensm. Unters. Forsch.* **1982**, *175* (3),  
407 182–185.
- 408 (17) Lachenmeier, D. W.; Nathan-Maister, D.; Breaux, T. A.; Sohnius, E.-M.; Schoeberl, K.;  
409 Kuballa, T. Chemical Composition of Vintage Preban Absinthe with Special Reference to  
410 Thujone, Fenchone, Pinocamphone, Methanol, Copper, and Antimony Concentrations. *J.*  
411 *Agric. Food Chem.* **2008**, *56*, 3073–3081.
- 412 (18) Rubiolo, P.; Matteodo, M.; Bicchi, C.; Appendino, G.; Gnavi, G.; Berteà, C.; Maffei, M.  
413 Chemical and Biomolecular Characterization of *Artemisia umbelliformis* Lam., an Important  
414 Ingredient of the Alpine Liqueur “Genépi.” *J. Agric. Food Chem.* **2009**, *57*, 3436–3443.
- 415 (19) Lê, C. L. Culture in vitro du genépi blanc (*Artemisia umbelliformis* Lam.). *Rev. suisse Vitic.*  
416 *Arboric. Hortic.* **1998**, *30* (3), 153–156.
- 417 (20) Berna, A. Z.; Trowell, S.; Clifford, D.; Cynkar, W.; Cozzolino, D. Geographical origin of

- 418 Sauvignon Blanc wines predicted by mass spectrometry and metal oxide based electronic  
419 nose. *Anal. Chim. Acta* **2009**, *648*, 146–152.
- 420 (21) Fudge, A. L.; Wilkinson, K. L.; Ristic, R.; Cozzolino, D. Classification of Smoke Tainted  
421 Wines Using Mid-Infrared Spectroscopy and Chemometrics. *J. Agric. Food Chem.* **2011**, *60*,  
422 52–59.
- 423 (22) Falqué, E.; Fernández, E.; Dubourdieu, D. Volatile Components of Loureira, Dona Branca,  
424 and Treixadura Wines. *J. Agric. Food Chem.* **2002**, *50*, 538–543.
- 425 (23) Nicolotti, L.; Cordero, C.; Bicchi, C.; Rubiolo, P.; Sgorbini, B.; Liberto, E. Volatile profiling  
426 of high quality hazelnuts (*Corylus avellana* L.): Chemical indices of roasting. *Food Chem.*  
427 **2013**, *138* (2-3), 1723–1733.
- 428 (24) Wu, Q. J.; Dong, Q. H.; Sun, W. J.; Huang, Y.; Wang, Q. Q.; Zhou, W. L. Discrimination of  
429 Chinese Teas with Different Fermentation Degrees by Stepwise Linear Discriminant  
430 Analysis (S-LDA) of the Chemical Compounds. *J. Agric. Food Chem.* **2014**, *62*, 9336–9344.
- 431 (25) Bicchi, C.; Ruosi, M. R.; Cagliero, C.; Cordero, C.; Liberto, E.; Rubiolo, P.; Sgorbini, B.  
432 Quantitative analysis of volatiles from solid matrices of vegetable origin by high  
433 concentration capacity headspace techniques: determination of furan in roasted coffee. *J.*  
434 *Chromatogr. A* **2011**, *1218* (6), 753–762.
- 435 (26) Serrano E., Beltràn J., Hernandez F., Application of multiple headspace-solid-phase  
436 microextraction followed by gas chromatography-mass spectrometry to quantitative analysis  
437 of tomato aroma components, *J. Chromatogr. A* 1216 (2009) 127-133.
- 438 (27) Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Hum. Genet.*  
439 **1936**, *7* (2), 179–188.
- 440 (28) Ezquerro O., Ortiz G., Pons B., Tena M. T., Determination of benzene, toluene, ethylbenzene  
441 and xylenes in soils by multiple headspace solid-phase microextraction, *J. Chromatogr. A*  
442 1035 (2004) 17-22.

- 443 (29) Sgorbini B., Bicchi C., Cagliero C., Cordero C., Liberto E., Rubiolo R., Herbs and spices:  
444 Characterization and quantitation of biologically-active markers for routine quality control by  
445 multiple headspace solid-phase microextraction combined with separative or non-separative  
446 analysis, *J. Chromatogr. A* 1376 (2015) 9-17
- 447 (30) Regulation (EC) No 94/358 Council Decision of 16 June 1994 accepting, on behalf of the  
448 European Community, the Convention on the elaboration of a European Pharmacopoeia. *Off.*  
449 *J. Eur. Communities* **1994**, L 158, 17–18.
- 450 (31) Markelov M., Guzowski J.P., Matrix independent headspace gas chromatographic analysis.  
451 The full evaporation technique. *Anal. Chim. Acta* 276: 235-45 (1993).
- 452 (32) Martinez, A. M.; Kak, A. C. PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.*  
453 **2001**, 23 (2), 228–233.
- 454 (33) R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for  
455 Statistical Computing: Vienna, Austria 2015.
- 456 (34) RStudio Team. RStudio: Integrated Development Environment for R. RStudio, Inc.: Boston,  
457 Ma, USA 2015.
- 458 (35) Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A. *caret*:  
459 *Classification and Regression Training*; 2012.
- 460 (36) Venables, W. N.; Ripley, B. D. *Modern Applied Statistics with S*; Springer: New York, 2002.
- 461 (37) Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; Chang, C. C.; Lin, C. C.  
462 *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly:*  
463 *E1071), TU Wien. Version 1.6-7*; 2015.

464

465

466

467 **Figure captions**

468 **Figure 1**

469 Génépi illustrated to highlight phenological (A, B) and morphological (C, D, E) parameters  
470 measured (Drawing: Karla Alejandra Palacios Antenucci).

471 **Figure 2**

472 Scatter plot of the first two Discriminant Functions of morphological data LDA. The geographic  
473 origin is here reported by numbers: 1, Gran Paradiso; 2, Val Chisone; 3 Elva; 4, Valle Gesso.

474 **Figure 3**

475 *A. umbelliformis* Lam. HS-SPME-GC-MS chromatographic profile. Legend: 1)  $\alpha$ -thujene, 2)  $\alpha$ -  
476 pinene, 3) camphene, 4) sabinene, 5)  $\beta$ -pinene, 6)  $\alpha$ -phellandrene, 7)  $\alpha$ -terpinene, 8) *p*-cymene, 9)  
477 1.8-cineole, 10)  $\gamma$ -terpinene, 11) *cis*-sabinene hydrate, 12)  $\alpha$ -terpinolene, 13)  $\alpha$ -thujone, 14)  $\beta$ -  
478 thujone, 15) *trans*-pinocarveol, 16) sabinol, 17) pinocarvone, 18) borneol, 19) terpinen-4-ol, 20)  $\alpha$ -  
479 terpineol, 21) cuminaldehyde, 22)  $\alpha$ -copaene, 23) sabinyl propionate, 24) sabinyl isobutyrate, 25)  
480 *trans*- $\beta$ -farnesene, 26) sabinyl isovalerianate, 27) sabinyl valerianate, 28)  $\delta$ -cadinene, 29)  
481 caryophyllene oxide, 30) neryl isovalerianate

482 **Figure 4**

483 Scatter plot of the first two Discriminant Functions of HS-SPME-GC-MS data LDA. The  
484 geographic origin is here reported by numbers: 1, Gran Paradiso; 2, Val Chisone; 3 Elva; 4, Valle  
485 Gesso.

486 **Figure 5**

487 Histograms representing the frequencies of discriminant function scores of the analyzed samples by  
488 HS-SPME-GC-MS. The upper panel shows samples below the limit, while the lower panel shows  
489 samples above the limit

490 **Figure 6**

491 Histograms representing the frequencies of discriminant function scores of the analyzed samples by  
492 HS-SPME-MS. The upper panel shows samples below the limit, while the lower panel shows  
493 samples above the limit.



## Tables

**Table 1** - Geographic origin of *A. umbelliformis* used in the experiment as mother plants.

<b>Geographic origin</b>	<b>Code</b>	<b>Locations</b>	<b>Altitude</b>	<b>Geographic coordinates</b>
Gran Paradiso	GP	Campiglia (Torino)	1,400 m a.s.l.	45°32'27.00"N, 7°32'17.78"E
Val Chisone	VC	Pragelato (Torino)	1,500 m a.s.l.	45°00'36.41"N, 6°56'14.56"E
Elva	EL	Elva (Cuneo)	1,600 m a.s.l.	44°32'16.31"N, 7°05'33.87"E
Valle Gesso	VG	Castelmagno (Cuneo)	1,900 m a.s.l.	44°24'13.10"N, 7°09'53.98"E

**Table 2** - Validation and prediction confusion matrices in Gran Paradiso (GP); Val Chisone (VC); Elva (EL); Valle Gesso (VG). The recognition rates of the model are not successful.

		<b>Predicted Group Membership (%)</b>			
		<b>GP</b>	<b>VC</b>	<b>EL</b>	<b>VG</b>
<b>Validation</b>	<b>GP</b>	66.7	25.0	0.0	8.3
	<b>VC</b>	18.5	66.7	3.7	11.1
	<b>EL</b>	3.8	19.2	76.9	0.0
	<b>VG</b>	7.4	7.4	7.4	77.8
<b>Prediction</b>	<b>GP</b>	50.0	33.3	4.2	12.5
	<b>VC</b>	51.9	29.6	3.7	14.8
	<b>EL</b>	7.7	42.3	42.3	7.7
	<b>VG</b>	11.1	14.8	22.2	51.9

**Table 3** - Confusion matrices in validation and prediction in Gran Paradiso (GP); Val Chisone (VC); Elva (EL); Valle Gesso (VG) using LDA model created with HS-SPME-GC-MS data.

		<b>Predicted Group Membership (%)</b>			
		<b>GP</b>	<b>VC</b>	<b>EL</b>	<b>VG</b>
<b>Validation</b>	<b>GP</b>	89.1	10.9	0.0	0.0
	<b>VC</b>	11.1	88.9	0.0	0.0
	<b>EL</b>	1.9	1.9	96.2	0.0
	<b>VG</b>	0.0	0.0	0.0	100.0
<b>Prediction</b>	<b>GP</b>	82.6	17.4	0.0	0.0
	<b>VC</b>	13.0	87.0	0.0	0.0
	<b>EL</b>	1.9	1.9	94.3	1.9
	<b>VG</b>	0.0	0.0	3.7	96.3

**Table 4** - Confusion matrix for liqueur thujone level prediction with HS-SPME-GC-MS data.

		<b>Predicted Group (%)</b>	
		<b>&lt; 35 ppm</b>	<b>&gt; 35 ppm</b>
<b>Validation</b>	<b>&lt; 35 ppm</b>	97.6	2.4
	<b>&gt; 35 ppm</b>	0.0	100.0
<b>Prediction</b>	<b>&lt; 35 ppm</b>	95.2	4.8
	<b>&gt; 35 ppm</b>	7.1	92.9

**Table 5** - Confusion matrix for liqueur thujone level prediction with HS-SPME-MS data.

		<b>Predicted Group (%)</b>	
		<b>&lt; 35 ppm</b>	<b>&gt; 35 ppm</b>
<b>Validation</b>	<b>&lt; 35 ppm</b>	97.7	2.3
	<b>&gt; 35 ppm</b>	3.6	96.4
<b>Prediction</b>	<b>&lt; 35 ppm</b>	88.6	11.4
	<b>&gt; 35 ppm</b>	10.7	89.3

Figures

Figure 1

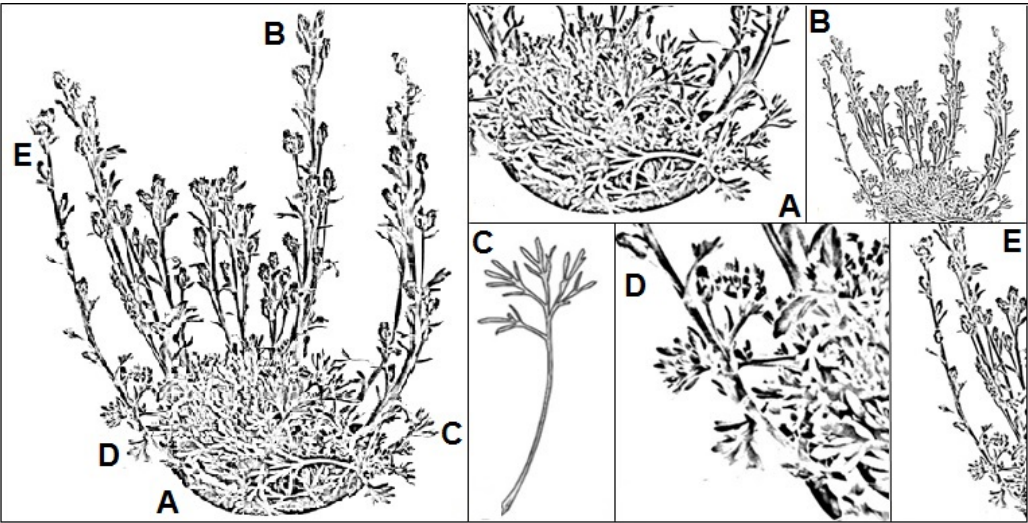
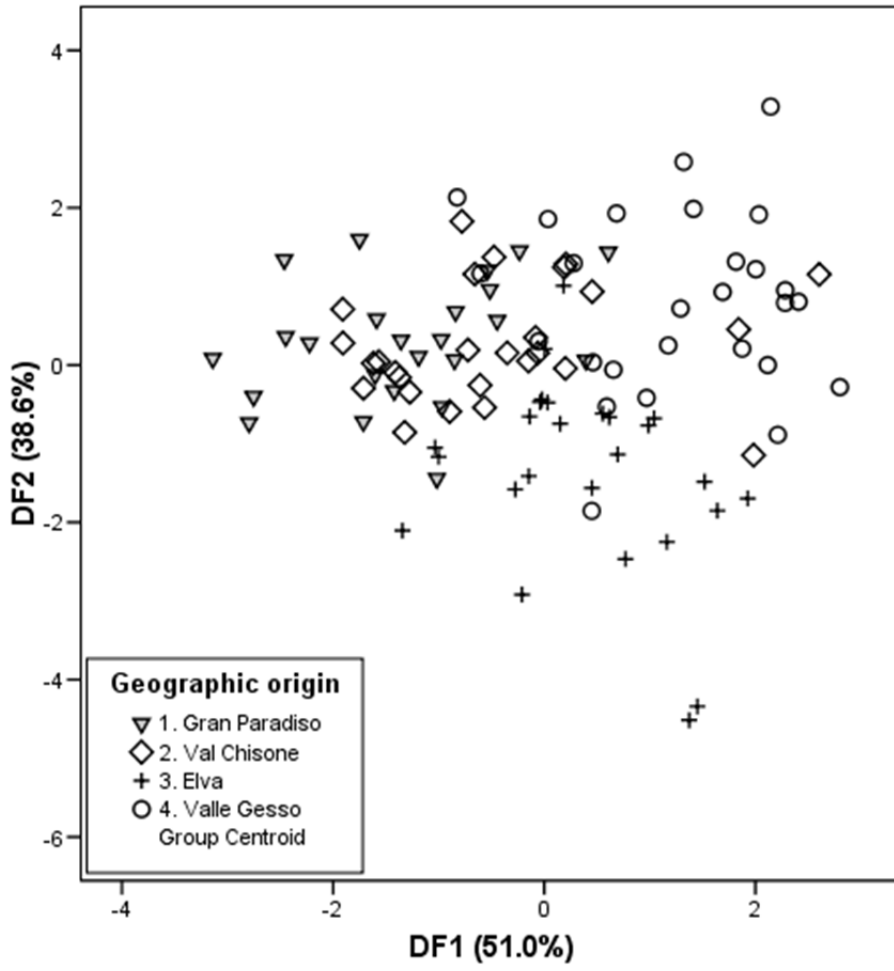


Figure 2



**Figure 3**

Abundance

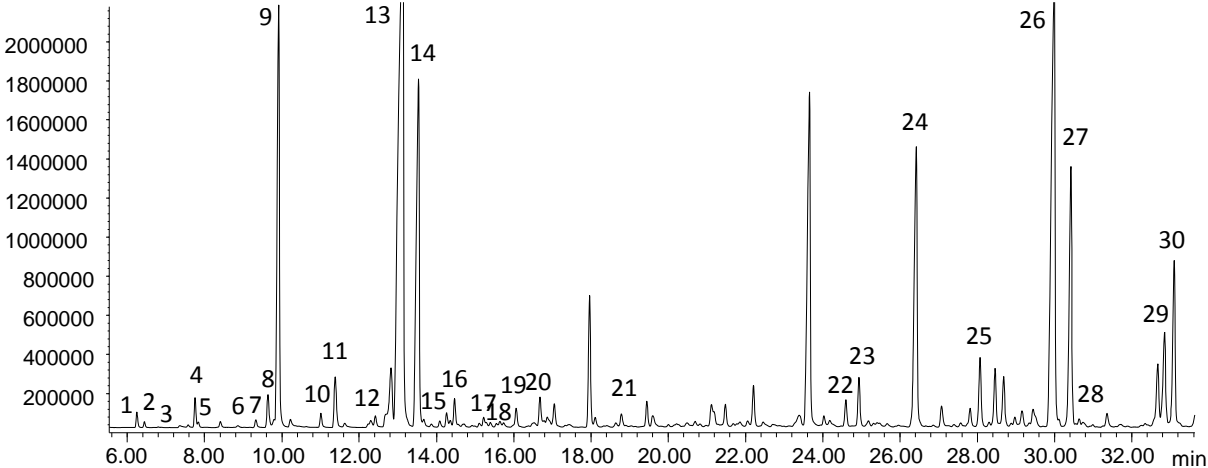




Figure 4

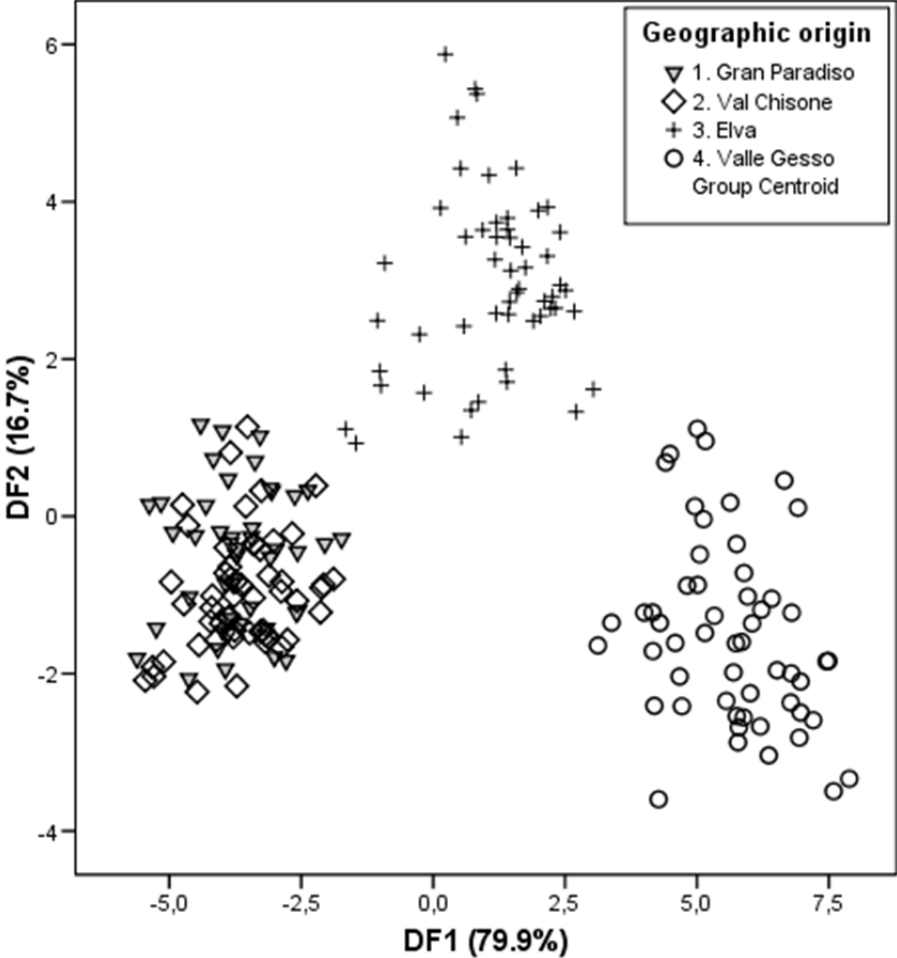


Figure 5

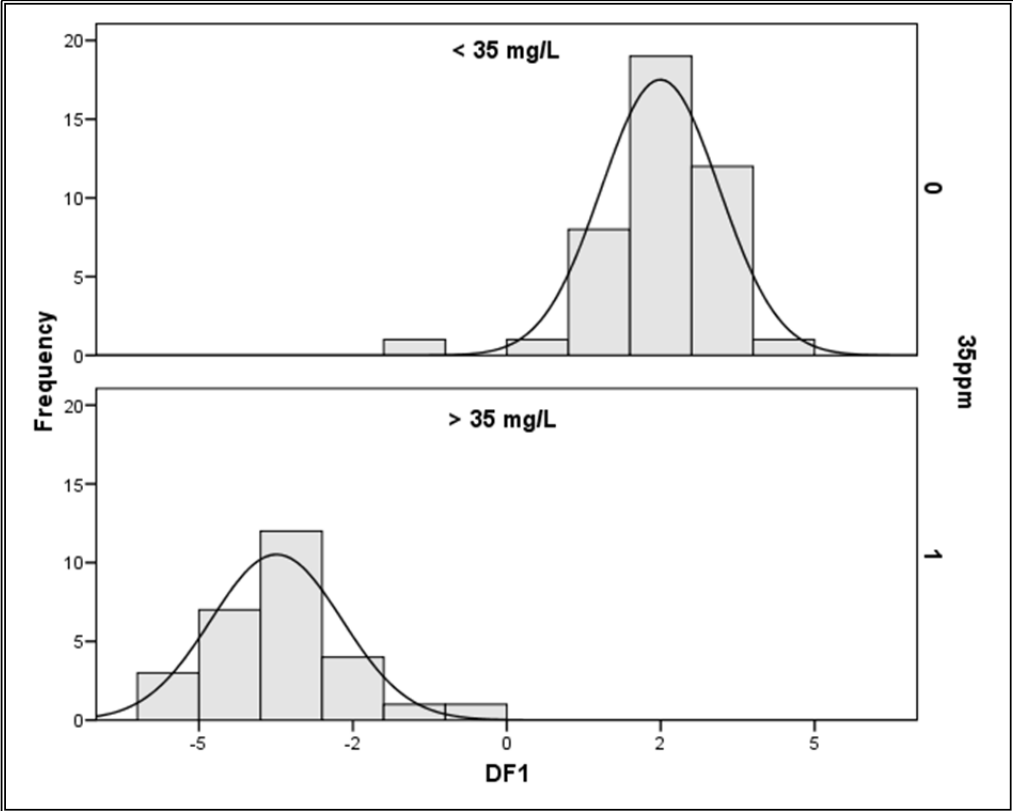
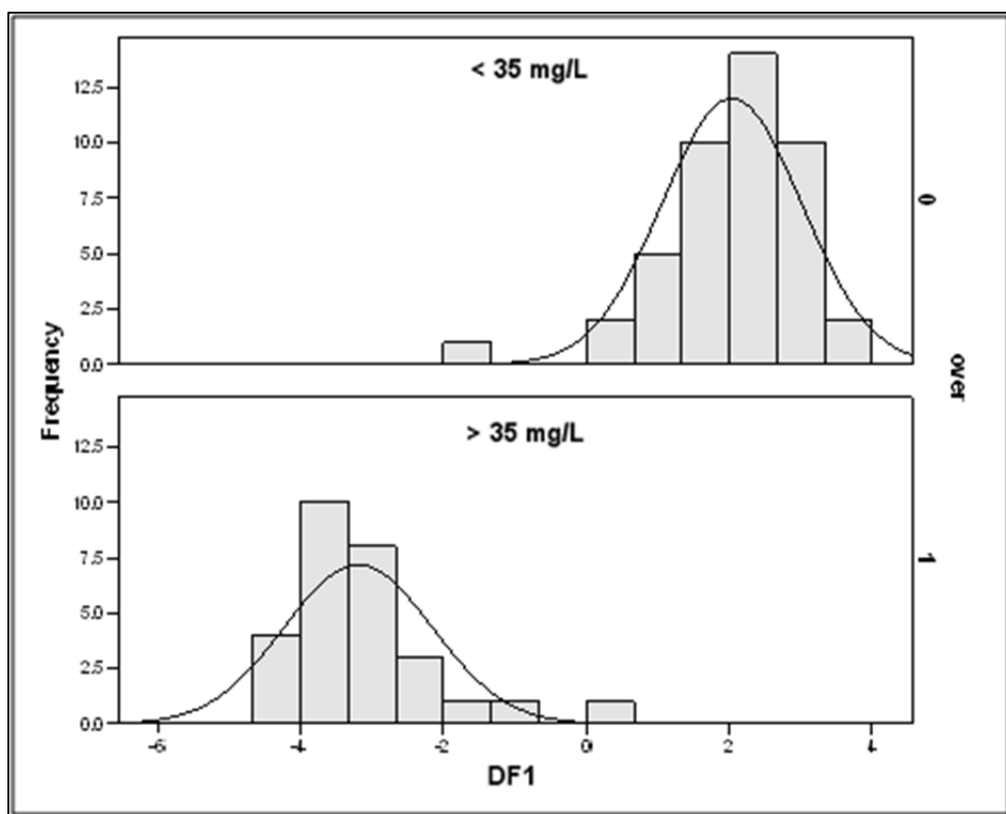


Figure 6



## GRAPHIC FOR TABLE OF CONTENTS

