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

Boggia, Lorenzo; Pignata, Giuseppe; Sgorbini, Barbara; Colombo, Maria Laura; Marengo, Arianna; Casale, Manuela; Nicola, Silvana; Bicchi, Carlo; Rubiolo, Patrizia. Artemisia umbelliformis Lam. and Génépi 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

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

Lorenzo Boggia[†], Giuseppe Pignata[§], Barbara Sgorbini[†], Maria Laura Colombo[†], Arianna Marengo[‡], Manuela Casale[§], Silvana Nicola[§], Carlo Bicchi[†], Patrizia Rubiolo[†]*

[†] Department of Drug Science and Technology (DSTF), University of Turin, Via Pietro Giuria 9, 10125 Torino, Italy

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

[‡] Department of Life and Environmental Sciences, University of Cagliari, Viale S. Ignazio da Laconi 13, 09124 Cagliari, Italy

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

ABSTRACT

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

Keywords: *Artemisia umbelliformis* Lam., génépi liqueur, food safety

INTRODUCTION

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

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

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

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

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

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

This research studied *A. umbelliformis* of different geographical origins, cultivated experimentally at a single site. Headspace sampling with Solid Phase Microextraction, combined with gas chromatography - mass spectrometry (HS-SPME-GC-MS) or directly with mass spectrometry (HS-SPME-MS), was used as an analytical platform to create the first database suitable for chemometric description and prediction. Linear Discriminant Analysis (LDA) was here adopted to define and predict geographic origin and thujone content of plants and liqueurs

Multiple Headspace- Solid Phase Microextraction (MHS-SPME)[26,28,29] was applied for thujone content determination.

MATERIAL AND METHODS

Chemicals. Thujones standard mixture (mixture of a-thujone and β-thujone, purity = 99.9%) and all other reference compounds were from Sigma-Aldrich (St. Louis, MO). Sabinol was kindly supplied by Robertet SA (Grasse, France), and sabinyl ester homologous series were synthesized in the authors' laboratory[18]. HPLC and analytical grade solvents were from Carlo Erba Reagenti, Rodano, Italy.

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

Plants were dried at 30°C in a dry and aerated room at the Experimental Centre of DISAFA (44°53'11.67"N; 7°41'7.00"E - 231 m a.s.l.) in Tetti Frati, Carmagnola (TO), Italy. The seeds were cleaned to remove extraneous parts, under fan ventilation using a series of sieves of different 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 (1st year),
132 and again the subsequent summer (2nd 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 1st summer and autumn and in
135 the 2nd 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

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

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

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

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

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

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

Data were processed with Agilent MSD ChemStation D.03.00.611 (Agilent Technologies). Components were identified by comparing their linear retention indices (I^Ts) (calculated *versus* a C_9-C_{25} hydrocarbon mixture) and their mass spectra to those of authentic standards.

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

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

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

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

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

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

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

Quantitative calibration was carried out by MHS-SPME using seven concentration levels of α- and β-thujone standard mixture (92.6% α-thujone, 7.4% β-thujone) between 0.1 g/L and 10 g/L in cyclohexane (two replicates of 10 µL for each level, three HS-SPME samplings for each replicate)[31].

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

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

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

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

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

RESULTS AND DISCUSSION

The study comprised three main steps: i) to describe statistically-significant differences relating to 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 2nd 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 2nd 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 showa 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.

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

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

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

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

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

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

290 concentration levels. (α-thujone: y = 2205.2x - 88987.3; R^2 = 0.9996; β-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[13]).
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 génépi 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.

ACKNOWLEDGMENT

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

ASSOCIATED CONTENT

Supporting Information

Supplementary Figure 1.

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

Supplementary Table 1.

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

Supplementary Table 2

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

Supplementary Table 3

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

Supplementary Table 4

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

Supplementary Table 5

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

REFERENCES

(1) Funk, V. A.; Susanna, A.; Stuessy, T. F.; Robinson, H. Classification of Compositae. In *Systematics, Evolution, and Biogeography of Compositae*; International Association for Plant Taxonomy, Ed.; Vienna, 2009.

(2) Hayat, M. Q.; Ashraf, M.; Khan, M. A.; Mahmood, T.; Ahmad, M.; Jabeen, S. Phylogeny of Artemisia L.: Recent developments. *African J. Biotechnol.* **2009**, *8* (11), 2423–2428.

(3) Klayman, D. L. Qinghaosu (Artemisinin): An antimalarial drug from China. *Science* **1985**, *228* (4703), 1049–1055.

(4) Xie, D.-Y. Artemisia annua, artemisinin, and the Nobel Prize: beauty of natural products and educational significance. *Sci. Bull.* **2016**, *61* (1), 42–44.

(5) Molyneux, D. H.; Ward, S. A. Reflections on the Nobel Prize for Medicine 2015 - The Public Health Legacy and Impact of Avermectin and Artemisinin. *Trends Parasitol.* **2015**, *31* (12), 605–607.

(6) Vouillamoz, J. F.; Carlen, C.; Taglialatela-Scafati, O.; Pollastro, F.; Appendino, G. The génépi Artemisia species. Ethnopharmacology, cultivation, phytochemistry, and bioactivity. *Fitoterapia* **2015**, *106*, 231–241.

(7) Albert-Puleo, M. Mythobotany, Pharmacology, and Chemistry of Thujone-Containing Plants and Derivatives. *Econ. Bot.* **1978**, *32* (1), 65–74.

(8) Delahaye, M.-C. *Le Génépi*; Equinoxe: Saint-Rémy de Provence, France, 2008.

(9) Blagojević, P.; Radulović, N.; Palić, R.; Stojanović, G. Chemical Composition of the Essential Oils of Serbian Wild-Growing Artemisia absinthium and Artemisia vulgaris. *J. Agric. Food Chem.* **2006**, *54*, 4780–4789.

(10) Lopes-Lutz, D.; Alviano, D. S.; Alviano, C. S.; Kolodziejczyk, P. P. Screening of chemical composition, antimicrobial and antioxidant activities of Artemisia essential oils. *Phytochemistry* **2008**, *69* (8), 1732–1738.

(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.; Bertea, C.; Maffei, M.
413 Chemical and Biomolecular Characterization of Artemisia umbelliformis Lam., an Important
414 Ingredient of the Alpine Liqueur "Genepì." *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.

(29) Sgorbini B., Bicchi C., Cagliero C., Cordero C., Liberto E., Rubiolo R., Herbs and spices: Characterization and quantitation of biologically-active markers for routine quality control by multiple headspace solid-phase microextraction combined with separative or non-separative analysis, J. Chromatogr. A 1376 (2015) 9-17

(30) Regulation (EC) No 94/358 Council Decision of 16 June 1994 accepting, on behalf of the European Community, the Convention on the elaboration of a European Pharmacopoeia. *Off. J. Eur. Communities* **1994**, *L 158*, 17–18.

(31) Markelov M., Guzowski J.P., Matrix independent headspace gas chromatographic analysis. The full evaporation technique. Anal. Chim. Acta 276: 235-45 (1993).

(32) Martinez, A. M.; Kak, A. C. PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23* (2), 228–233.

(33) R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria 2015.

(34) RStudio Team. RStudio: Integrated Development Environment for R. RStudio, Inc.: Boston, Ma, USA 2015.

(35) Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A. *caret: Classification and Regression Training*; 2012.

(36) Venables, W. N.; Ripley, B. D. *Modern Applied Statistics with S*; Springer: New York, 2002.

(37) Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; Chang, C. C.; Lin, C. C. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. Version 1.6-7*; 2015.

Figure captions

Figure 1

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

Figure 2

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

Figure 3

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

Figure 4

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

Figure 5

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

Figure 6

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

Tables

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

Geographic origin	**Code**	**Locations**	**Altitude**	**Geographic coordinates**
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

494

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.

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

495

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.

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

496

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

		Predicted Group (%)	
		< 35 ppm	**> 35 ppm**
Validation	**< 35 ppm**	97.6	2.4
	> 35 ppm	0.0	100.0
Prediction	**< 35 ppm**	95.2	4.8
	> 35 ppm	7.1	92.9

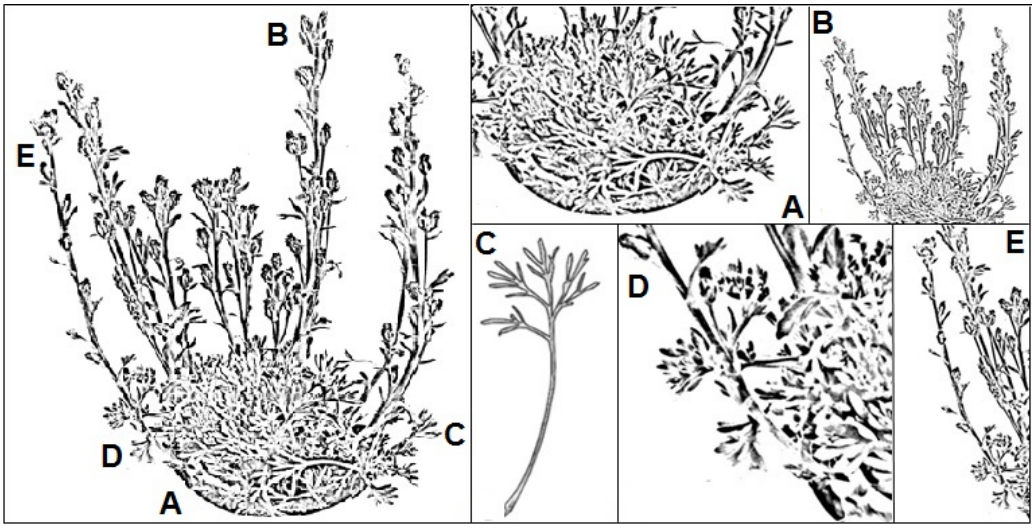
497

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

		Predicted Group (%)	
		< 35 ppm	**> 35 ppm**
Validation	**< 35 ppm**	97.7	2.3
	> 35 ppm	3.6	96.4
Prediction	**< 35 ppm**	88.6	11.4
	> 35 ppm	10.7	89.3

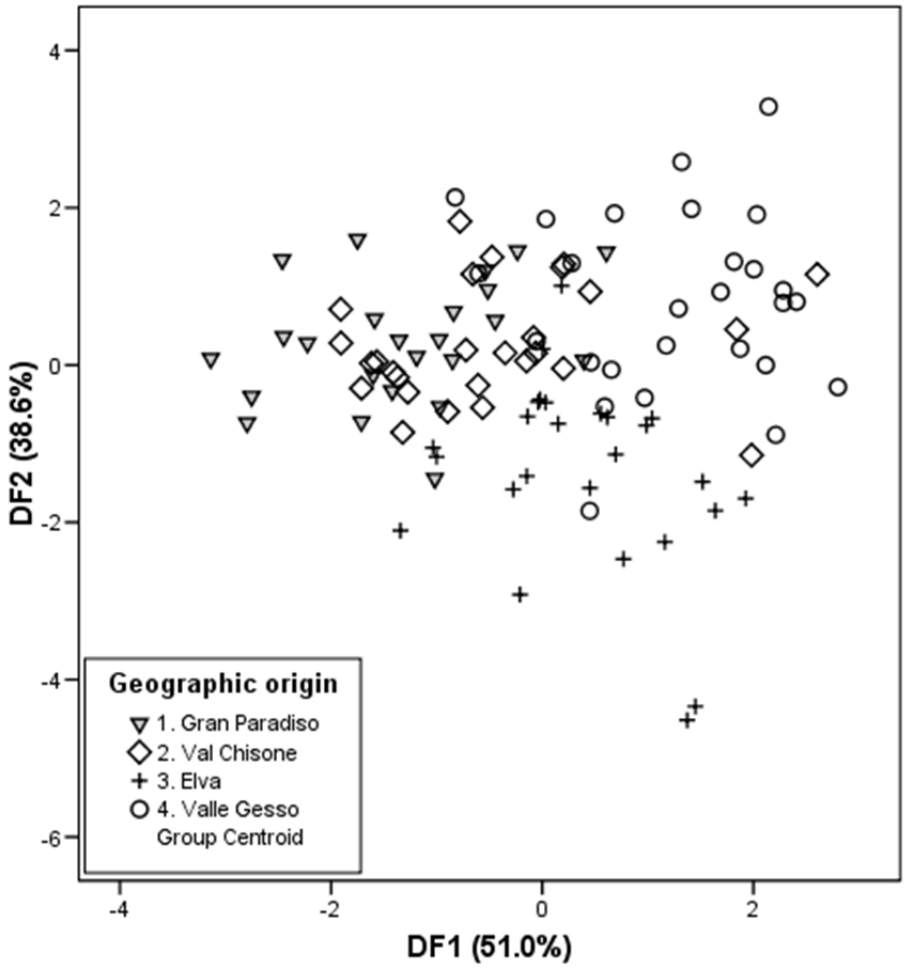
Figures

Figure 1



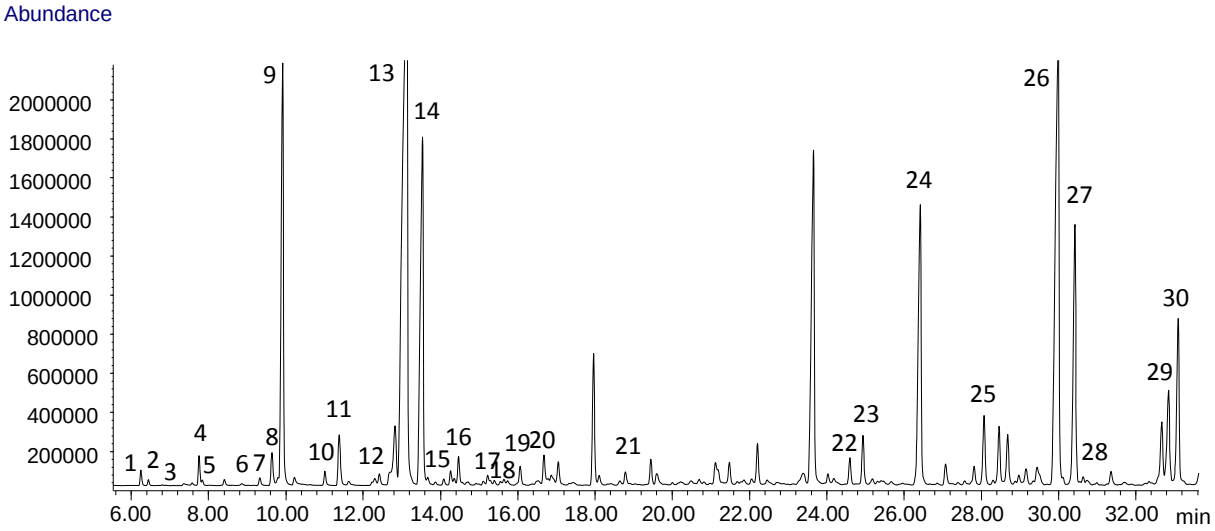
498

Figure 2


4
2
0
-2
-4
-6
DF2 (38.6%)
-4
-2
0
2
DF1 (51.0%)
Geographic origin
1. Gran Paradiso
2. Val Chisone
3. Elva
4. Valle Gesso
Group Centroid

499

Figure 3


Abundance
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
6.00
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
24.00
26.00
28.00
30.00
32.00
min

500

Figure 4

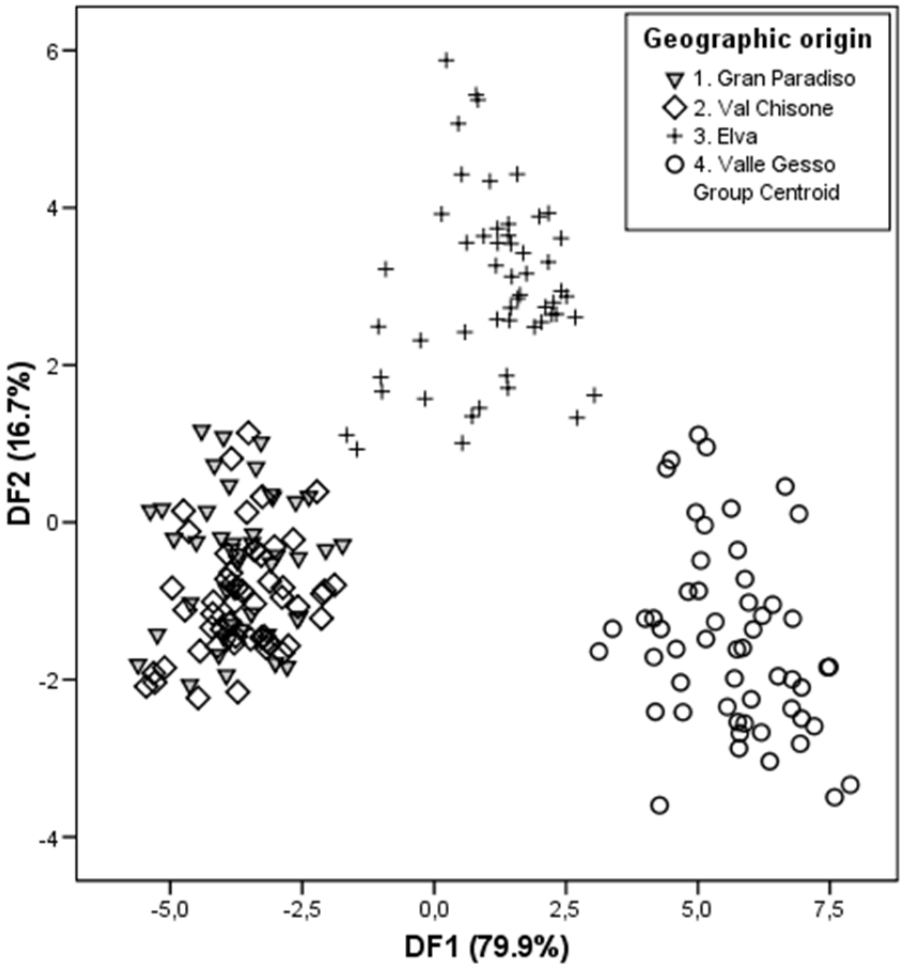
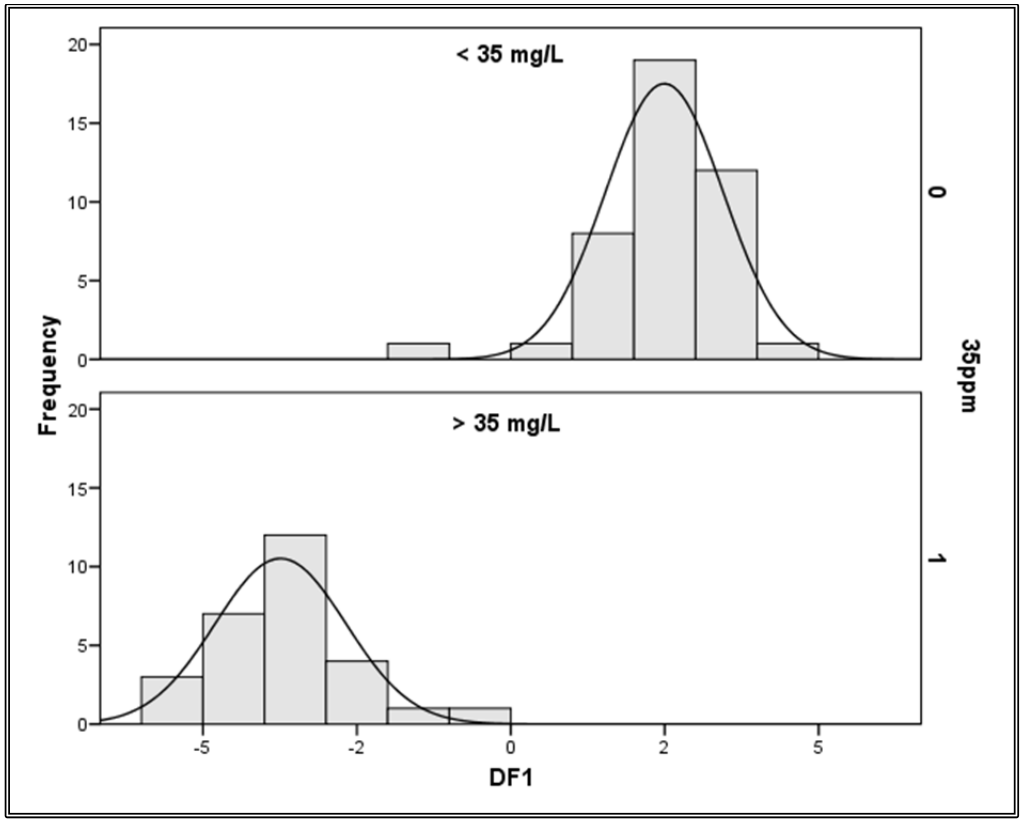
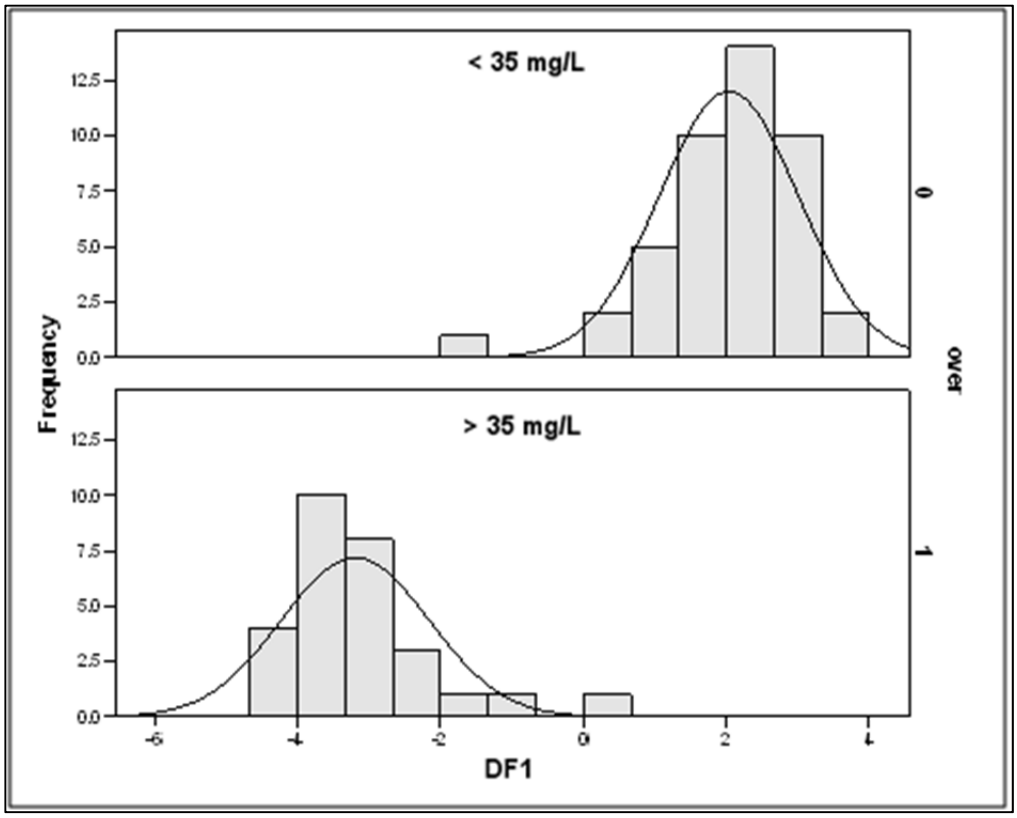

Geographic origin
1. Gran Paradiso
2. Val Chisone
3. Elva
4. Valle Gesso
Group Centroid
DF2 (16.7%)
DF1 (79.9%)
6
4
2
0
-2
-4
-5,0
-2,5
0,0
2,5
5,0
7,5

Figure 5



501

Figure 6



GRAPHIC FOR TABLE OF CONTENTS