

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A metabolomic approach to quality determination and authentication of raw plant material in the fragrance field. Iris rhizomes: A case study

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/154425> since 2016-07-14T12:43:05Z

Published version:

DOI:10.1016/j.chroma.2014.09.076

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)



UNIVERSITÀ DEGLI STUDI DI TORINO

This Accepted Author Manuscript (AAM) is copyrighted and published by Elsevier. It is posted here by agreement between Elsevier and the University of Turin. Changes resulting from the publishing process - such as editing, corrections, structural formatting, and other quality control mechanisms - may not be reflected in this version of the text. The definitive version of the text was subsequently published in [*Journal of Chromatography A Volume 1368, 14 November 2014, Pages 143–154 DOI: 10.1016/j.chroma.2014.09.076*].

You may download, copy and otherwise use the AAM for non-commercial purposes provided that your license is limited by the following restrictions:

- (1) You may use this AAM for non-commercial purposes only under the terms of the CC-BY-NC-ND license.
- (2) The integrity of the work and identification of the author, copyright owner, and publisher must be preserved in any copy.
- (3) You must attribute this AAM in the following format: Creative Commons BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>), [<http://dx.doi.org/10.1016/j.chroma.2014.09.076>]

1

2 **A metabolomic approach to quality determination and authentication of raw**
3 **plant material in the fragrance field. Orris roots: a case study**

4

5 Jerome Masson^{1,2}, Erica Liberto², Hugues Brevard¹, Carlo Bicchi², Patrizia Rubiolo^{2*}

6

7 ¹ Robertet SA, Research Division, 37 Avenue Sidi Brahim, F-06130 Grasse, France

8 ² Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, I-
9 10125 Torino, Italy

10

11

12 Corresponding author:

13 Prof. Dr. Patrizia Rubiolo

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

15 *10125 Torino, Italy; Tel: +39 011 6707662, fax: +39 011 6707687;*

16 *e-mail: patrizia.rubiolo@unito.it*

17

18 **Abstract**

19 This study aimed to discriminate twenty-two samples of commercial *Iris* rhizomes (orris root)
20 by species and origin (*Iris germanica* (Morocco), *I. albicans* (Morocco), *I. pallida* (Morocco), *I.*
21 *pallida* (China), *I. pallida* (Italy)) by applying a strategy derived from those adopted in
22 metabolomics. The specimens' fingerprints from conventional analysis methods (LC-UV
23 and/or LC-MS) were unable to provide clear discrimination. A strategy combining
24 UHPLC/TOF-HRMS, in positive and negative modes, with multivariate statistical methods was
25 therefore applied. Exact Mass/Retention Time (EMRT) pairs obtained by UHPLC-TOF/HRMS
26 were successfully submitted to statistical processing by Principal Component Analysis (PCA),
27 Partial Least Square Discriminant Analysis (PLS-DA), and then Orthogonal Partial Least
28 Square-Discriminant Analysis (OPLS-DA), to extract the discriminating EMRT pairs through
29 their trend views. 146 EMRT pairs were selected on the basis of their trend views, because
30 they significantly varied, and 104 of them were included to discriminate between species
31 and origins. 32 of them were tentatively identified as discriminating markers (flavonoids,
32 isoflavonoids, triterpenoids, benzophenone derivatives and related glycosides ...) from the
33 reference database created on the basis of *Iris* genus components reported in the literature:
34 eight of them specific for *I. albicans*, four for *I. germanica*, five for *I. pallida* (Italy), five for *I.*
35 *pallida* (China), and ten for *I. pallida* (Morocco). The reliability of this strategy was confirmed
36 by identifying species and origin of two unknown samples submitted to the same analytical
37 procedure.

38

39 **Keywords:**

40 *Iris* species and origins; Metabolomic strategy, Phytochemical analysis; UHPLC/TOF-MS;
41 statistical data treatment, discriminant markers

42 1. Introduction

43 Raw plant materials used in the flavor and fragrance field often differ in terms of species and
44 origins and, as a consequence, they present variable chemical compositions, resulting in
45 different characteristic organoleptic properties. The most recent developments in analytical
46 technologies have greatly contributed to an improved knowledge of the chemical
47 composition of vegetable matrices, thus affording more precise definition of their identity
48 and origin.

49 Metabolomics is one of the most recent “omics” sciences. It was introduced at the end of
50 the 20th century, and first defined by Oliver et al. in 1998 [1] as “*the systematic study of the*
51 *unique chemical metabolite fingerprints (the metabolome) resulting from specific cellular*
52 *processes*”. This discipline entails defining a matrix fingerprint or profile, mainly consisting of
53 small molecules, that is representative of the biological system investigated. Several
54 practical approaches have been developed to meet the requirements of metabolomics,
55 some being of great interest for quality evaluation and authentication of raw plant materials.
56 The most promising approaches in this connection are *fingerprinting* and *profiling*, which
57 differ in terms of results achieved and methodology adopted. *Fingerprinting* is an *untargeted*
58 *method*, in which samples are compared on the basis of their profiles, to detect quali- and
59 quantitative differences between them, or to classify samples within a data set.
60 *Fingerprinting* comprises general and rapid high-throughput screening, with the aim of
61 discriminating and classifying samples. As a consequence, it can also be achieved through a
62 non-separative approach (HS-MS, direct infusion-MS, NMR, FT-IR (MIR, NIR, etc.) combined
63 with appropriate statistical data processing. *Profiling* is a *targeted method*, in which samples
64 are compared on the basis of the distribution of known target analytes (markers). *Profiling*
65 produces a detailed analytical profile of the sample, by combining separation with a
66 spectroscopic technique (*e.g.* GC/MS, LC/MS, CE/MS etc.) with the aim of identifying and
67 quantifying a selected number of diagnostic components [2].

68 Recent advances in analytical technologies, in particular the on-line combination of UHPLC
69 with TOF-MS, make it possible to acquire rapidly a full set of exact masses present in
70 different matrices, although the resulting huge amount of data may be complex to manage.
71 The use of this approach, in combination with advanced statistical methods [3] deriving from
72 metabolomics, may be a further decisive option in the area of phytochemical analysis. It can

73 successfully be used to define specific markers, and to establish identity profiles for
74 unequivocal identification and authentication of raw plant materials.

75 *Iris* rhizome essential oil, commonly known as “Orris butter”, is one of the most important,
76 and most expensive, raw materials used in the fragrance field [4]. It is valued for its unique
77 violet-like scent, and is obtained by hydrodistillation of *Iris* rhizomes. Its distinctive odor is
78 due to irones (**Figure 1**), first described by Tiemann and Krüger in 1893 [5]; their structure
79 was defined by Naves [6] and Ruzicka [7], and their stereochemistry elucidated by
80 Rautenstrauch and Ohloff [8, 9].

81 Several studies on species belonging to the *Iris* genus have dealt with their biological activity
82 [10-13], with the formation of irones and identification of their precursors [14-25], and have
83 lead to the isolation of different classes of secondary metabolites, namely triterpenoids,
84 flavonoids and phenols [25-27]. In 1983, Krick [14] showed that iripallidal and iriflorental are
85 the irone precursors (Fig. 1). Subsequently, Marner’s group [15, 16] isolated and described
86 several mono-cyclic [17-19] and bicyclic iridal structures [20], for a total of more than forty
87 triterpenic compounds, and also clarified their biosynthetic pathway [21-24]. In 1993, Bicchi
88 and Rubiolo [25] showed that iridals were also present as esters of myristic acid.

89 The phenolic fraction of *Iris* rhizomes has also been investigated in depth. In 1983, Ali [26]
90 determined the structure of several flavonoids and isoflavonoids (**Figure 1**) in *I. germanica*
91 rhizomes, including irigenin, irisolidon, and iridin, all known for their biological activities.
92 Bicchi and Rubiolo [25] identified irisflorentin, iristectogenin B, and iriflophenone in *I.*
93 *pallida*. More recently, Roger et al. [27] isolated, identified, and quantified 11 flavonoids in *I.*
94 *pallida* and *I. germanica* rhizomes, and characterized a benzophenone derivative typical of *I.*
95 *pallida* (2,6,4’-trihydroxy-4-methoxybenzophenone).

96 Over the last 30 years, the geographical and botanical origins of “orris butter” have been the
97 objective of several studies, mostly based on the isomeric abundances of irones and iridals
98 [28, 29], and more recently also on the flavonoid fraction, whose composition was used to
99 discriminate *I. germanica* from *I. pallida* from Morocco [27].

100 This study aimed to apply a metabolomic approach, based on both UHPLC/TOF-MS analyses
101 and statistical elaboration, to authenticate species and origins of a set of samples of *Iris*
102 rhizomes, thanks to the definition of specific discriminating markers. The reliability of the
103 strategy applied was confirmed by controlling the authenticity of unknown samples.

104

105 **2. Experimental**

106 2.1 Chemicals and reagents

107 Acetonitrile (ACN), methanol (MeOH), orthophosphoric acid, formic acid, and ammonium
108 formate (UHPLC/MS grade) were purchased from Biosolve (Dieuze, France). Distilled water
109 was purified 'in-house' using an ELGA MilliQ system VeoliaWater STI (Le Plessis-Robinson,
110 France). CH₂Cl₂ was purchased from Carlo Erba (Val de Reuil, France). Leucine-Enkephalin
111 was from Waters (Milford, USA)

112

113 2.2 *Plant material*

114 Five groups of certified samples of *Iris* rhizomes rhizomes available from Robertet (Grasse,
115 France) for a total of 24 samples were used for this study. They comprised *Iris germanica* (3,
116 origin: Morocco) (*g,MA*), *I. albicans* (4, Morocco) (*a,MA*), *I. pallida* (3, Morocco) (*p,MA*), *I.*
117 *pallida* (7, China) (*p,CN*), *I. pallida* (5, Italy) (*p,IT*) and 2 unknown samples.

118

119 2.3 *Sample preparation*

120 Dried plant material was crushed and extracted in a Soxtec system (Foss, France), with the
121 aim of developing an exhaustive and, at the same time, simple and automated sample
122 preparation technique compatible with demands for applicability to a routine quality control
123 process. Different experiments were carried out to optimize the extraction parameters
124 (extraction and rinsing times, extraction solvents, etc.). Five different solvent mixtures
125 (methanol 100%; methanol/methylene chloride 70/30, 50/50; methanol/chloroform 70/30,
126 30/70) were tested. The extraction yield was evaluated through the relative areas of five
127 components, representing different chemical classes already identified in the *Iris* genus
128 (irigenin; iripallidal, iriflorental, iridin, iriflophenone) analyzed by UHPLC/TOF-MS at two
129 different collision capillary voltages (1kV and 3kV). On the basis of the results of these
130 preliminary tests, the following method was applied thereafter: plant material (25g) was
131 submitted to an exhaustive extraction with MeOH:CH₂Cl₂ (70:30) (100mL) for two hours at
132 boiling temperature, and to rinsing for another two and a half hours. The resulting extracts
133 were then analyzed by HPLC/DAD-UV and UHPLC/TOF-MS in duplicate.

134

135

136

137 2.4 *HPLC/DAD-UV analysis*

138 HPLC analyses were run on an Agilent 1200 system. A Nucleodur HTec 5µm 4.6 x 250mm C18
139 column (Macherey-Nagel, France) at 45°C was used for chromatographic separations.
140 Solvent A was water with orthophosphoric acid (0.5%), solvent B was methanol (MeOH), and
141 solvent C was acetonitrile (ACN). The gradient profile was: 2.5% of solvent B and solvent C
142 for 5 min in isocratic mode; then from 2.5% to 47.5% of solvent B and Solvent C over 40 min,
143 hold for 5 min; the system was then re-equilibrated for 10 min at the initial conditions. The
144 flow rate was 1 ml/min, and the injection volume 10µL. Analyses were run at a wavelength
145 of 230nm, with a bandwidth of 4nm.

146

147 2.5 *UHPLC/TOF-MS analysis*

148 UHPLC analyses were carried out on a Waters ACQUITY® UPLC - Class system.
149 Chromatographic separation was achieved on an ACQUITY® UPLC HSS, 1.8µm, 2.1 x 100mm
150 C18 stationary phase (Waters, Milford, USA), which was maintained at 45°C; mobile phase:
151 eluent A: 5mM ammonium formate solution adjusted to pH=3.8 with formic acid; eluent B:
152 methanol/acetonitrile (50:50) with ammonium formate (5mM) and 0.1% formic acid. Mobile
153 phase gradient: from 5% B (1 min) to 95% of B in 10 min (3 min); re-equilibration at initial
154 condition: 2 min; flow rate: 0.45 ml/min; injection volume: 1µL.

155 UHPLC/TOF-MS analyses were run with a Waters XEVO® G2 TOF system. ESI source
156 parameters: capillary voltage: 1kV, cone voltage: 30V, extraction cone: 4V, source
157 temperature: 120°C, desolvation temperature: 400°C, gas flow (nitrogen): 10L/h, desolvation
158 gas flow: 1200L/h. The analyses were performed in positive and negative modes. The mass
159 axis was calibrated with sodium formate in the range 50 to 1200 Da; mass lockspray: Leucine
160 Enkephalin was used to correct masses; two masses were checked in both modes (positive
161 mode: 278.1141; 556.2771 and negative mode: 236.1035; 554.2615).

162 Each analysis was also carried out at a high capillary voltage (3kV) in both positive and
163 negative modes. The acquisitions were performed in MS^E mode, using argon as collision gas.

164

165 2.6 *Data processing*

166 The resulting data set was processed through the package Markerlynx; (Waters, Milfors USA)
167 based on the determination of Exact Mass Retention Time pairs (EMRT).

168 The EMRT pairs were first processed by Principal Component Analysis (PCA), followed by
169 Partial Least Square-Discriminant Analysis (PLS-DA) to improve group separation and classify
170 components. Then, specific data treatment, known as Orthogonal Partial Least Square-
171 Discriminant Analysis (OPLS-DA), was used to extract the EMRT pairs responsible for the
172 group separation; OPLS-DA entails a two-by-two comparison for either species or origin
173 discrimination.

174

175 2.7 *Building up the reference database*

176 A specific database containing 250 compounds identified in the investigated *Iris* species was
177 built up in the Masslynx Software package (Waters, Milfors USA) with the data retrieved
178 from the free data base KNApSACK [30], and from the original articles in which each
179 compound had first been described. The resulting database included exact mass, molecular
180 formula, RN from CAS, and structure, for each component included, and was then used to
181 identify discriminant markers on the basis of their molecular formula or exact mass.

182 The following procedure was adopted for tentative identification:

183 a) the EMRT pair of each eluting component obtained from UHPLC-TOF-MS was matched
184 with the *Iris* database, by its exact mass or elemental composition; the former was preferred
185 because it is less variable. Mass accuracy was used as diagnostic parameter for
186 identification. An output corresponding to components having similar exact mass (or
187 elemental composition) in both ionization modes (positive or negative) was obtained.

188 b) The resulting tentative identification was verified by comparing the exact masses of the
189 experimental fragments of each EMRT pair, obtained at high collision energy, to those of the
190 theoretical fragmentation, obtained via Mass Fragment software (Waters, Milfors USA). The
191 probability of a tentative identification being correct therefore increases markedly if, and
192 only if, experimental and theoretical fragmentations agree.

193 The reported identifications from analyses in the MS^E mode, in combination with the above
194 mentioned database, may be considered as tentative/putative, in agreement with the
195 recommendations of both the Chemical Analysis Working Group (CAWG) of the
196 Metabolomics Standards Initiative (MSI) [31], [32], and Schymanski et al. for the
197 environmental field [33].

198 In any case, the definitive identification must always be confirmed with the injection of
199 authentic standards, although, in most cases, they are not available commercially.

200 3. Results and discussion

201 The strategy adopted in this study is closely linked to the potential offered by new and
202 effective technologies, which can support comprehensive metabolite studies in
203 discriminating complex matrices of vegetable origin. Thus an untargeted approach was first
204 applied to the investigated samples, to obtain sample fingerprints for comparison, to select
205 components without considering peak identification. A targeted approach was then applied,
206 to validate discriminant markers from each species and origin, for differentiation of the
207 samples. This section comprises a critical step-by-step discussion of the results obtained
208 from the application of this strategy to the authentication of *Iris* rhizomes, in terms of
209 species and origin.

210

211 3.1 Untargeted approach

212 3.1.1 HPLC-DAD-UV analysis versus UHPLC-TOF-MS analysis

213 **Figure 2** reports the HPLC-DAD-UV chromatographic profiles of all samples. It is clear that,
214 under the analytical conditions adopted, the differences are very small. The most significant
215 difference concerns two components (peaks at retention time (Rt) = 23.35 min and at Rt =
216 30.16 min respectively) that are present in all extracts although in different abundances.
217 Conventional HPLC-DAD-UV at 230 nm analyses took about 50 minutes; analyses were also
218 carried out at different wavelengths (250, 300, and 350 nm) but the resulting
219 chromatographic patterns were not very significant (data not shown).

220 The same samples were then submitted to UHPLC/TOF-MS to obtain more informative
221 results in a shorter time. **Figures 3A and 3B** give the UHPLC/TOF-MS patterns of the extracts
222 of the investigated species and origins, in both positive and negative ESI ionization modes.
223 Thanks to the high efficiency of the analytical system (i.e. UHPLC, sub 2µm columns and a
224 TOF-MS detector) analysis time was reduced to 14 min.

225 Although, due to the TOF-MS detection, the chromatographic patterns of the investigated
226 *Iris* samples varied, again, only minor differences in composition were seen, either in positive
227 or in negative ionization modes. The main variations between samples in positive mode
228 (**Figure 3A**) concerned the different abundance of some components (markers). Moreover,
229 all *I. pallida* extracts showed very similar chromatographic profiles, making it impossible to
230 discriminate them in terms of origin. The fingerprints of the extract of all samples were very
231 similar, with the exception of *I. albicans*, which presented a characteristic component eluting

232 at $R_t = 8.97$ min with an exact mass m/z of 388.1148 Da. Similar results were obtained in
233 negative mode (**Figure 3B**), in which the differences mainly concerned the abundance of
234 some components within the set of samples. For instance, *I. pallida* samples from all origins
235 showed a peak at $R_t = 6.20$ min with an exact mass $m/z = 260.0685$ Da; its abundance was
236 higher than those detected in the other two species. *I. albicans* and *I. germanica* samples
237 showed a common abundant peak at $R_t = 8.81$ min with an exact mass of $m/z = 314.0788$
238 Da. However, the differences detected in this first set of experiments, in either ionization
239 mode, were not sufficient to make an unequivocal and correct assignment of species and
240 origin.

241

242 3.1.2 Metabolite fingerprinting

243 The main aim of this study was to differentiate orris root samples in terms of origin and
244 species, on the basis of the composition of their extracts. Meaningful information from the
245 very large number of data produced by modern analytical platforms (UHPLC-TOF-MS) can
246 only be obtained by adopting suitable strategies supported by dedicated data processing.

247 The software used in this study automatically converted each data point into an Exact Mass
248 Retention Time (EMRT) pair of data; about 39000 variables were extracted from the analysis
249 in positive ionization mode, and tabulated into a 2D matrix, and 5000 variables were
250 obtained from the negative mode, and processed independently.

251 The data were first processed by Principal Component Analysis (PCA), *i.e.* an unsupervised
252 clustering method able to reduce the dimensionality of the multivariate data while keeping
253 the intrinsic variability of the data set [3]. PCA processing was preceded by UV-scaling pre-
254 processing (Unit Variance scaling); this is one of the scaling systems most widely adopted
255 with PCA but, in this case, it suffered from considerable statistical noise (72% in positive
256 mode), (data not reported). This noise may decrease the meaningfulness of the resulting
257 information to below 50%. Other scaling procedures were therefore applied (center scaling
258 and Pareto scaling) to minimize the noise level, reducing it from 72% to 39%. **Figure 4** shows
259 the PCA discrimination of UHPLC/positive-TOF-MS data on the investigated *Iris* rhizome
260 species and origins, using Pareto-scaling. The PCA score plot showed a satisfactory
261 separation between samples of *I. pallida* from different origins and other species, achieving
262 an explained variance of 34 % (PC1 vs. PC2), despite the dispersion of the samples within the
263 same group (with the exception of the samples of *I. germanica* from Morocco (g, MA)).

264 Moreover, *I. pallida* from Morocco (p, MA) and Italy (p, IT) were not fully discriminated.
265 These results increase the difficulty of establishing which EMRT pairs could clearly
266 differentiate samples or groups.

267 Partial Least Square-Discriminant Analysis (PLS-DA) was then applied, to increase the
268 separation between the different groups, and to decrease the dispersion of samples within
269 groups. This model uses a fixed variable, in this case the species and/or origin of a group.
270 PLS-DA data treatment was also combined with different scaling approaches (UV-scaling,
271 center scaling, and Pareto scaling). Pareto scaling was chosen because it offered the lowest
272 level of statistical noise in both ionization modes (**Figures 5A and 5B**) resulting in an
273 explained variance of 35% in positive mode and 41% in negative mode; this difference is due
274 to the much greater number of variables in positive mode (39000 vs. 5000). This statistical
275 tool provided good separation of all groups on the first component, especially with the
276 negative ionization data sets; in particular, *I. pallida* samples were clearly separated from *I.*
277 *albicans* and *I. germanica*, and within *I. pallida* samples, the origins Italy, Morocco and China
278 were discriminated in negative ionization mode, showing the ability of the model to describe
279 the distribution of samples also in function of their geographical origin.

280 OPLS-DA statistical data treatment was then applied to clarify which components
281 (discriminating markers) made the biggest contribution to differentiating species and origins
282 and, as a consequence, to reducing the huge number of variables. The main benefit of
283 introducing an OSC (orthogonal signal correction) filter is the possibility of applying OPLS-DA
284 statistical treatment, to better separate predictive from non-predictive (orthogonal)
285 variations. This treatment enables the specificity of each group to be determined by pair
286 comparison. The models were internally fully cross-validated (by the leave-one-out method)
287 and sensitivity and specificity were calculated, being respectively 95% and 100% in negative
288 mode, and 100% and 100% in positive mode. The sensitivity of a model indicates the
289 percentage of acceptance relative to its own objects, while its specificity indicates the
290 percentage of the objects of another class that are rejected by the model. This over-fitting
291 was unsurprising, because of a) the high level of statistical noise, in particular in positive
292 mode, and b) the statistically limited number of samples analyzed, in comparison to the
293 number of variables considered. Through a scatter plot (S-Plot) based on OPLS-DA, the
294 discriminant markers that differentiate two groups of samples can be determined. In the S-
295 Plot, each point represents a unique EMRT pair. The x-axis shows the contributions of each

296 EMRT to discrimination between the two sample groups: the farther is the data point from
297 zero, the greater the contribution it gives to sample variance. The y-axis shows sample
298 correlations within a single sample group. As a consequence, the EMRT pairs at both ends of
299 the S-shaped curve represent the EMRT pairs (potential discriminating markers) contributing
300 the most to discrimination, with the highest level of confidence, from each sample group.

301 Twelve different comparisons by OPLS-DA were made, in positive and negative modes: three
302 in each mode between the different geographical origins within the same species (*I. pallida*),
303 and the other three pair-wise among the three different species from the same geographical
304 origin. **Figure 6** gives an example of an OPLS-DA diagram and its corresponding S-Plot
305 diagram, obtained by comparing *I. pallida* Italy vs. *I. pallida* China samples, in positive mode
306 (39000 variables). The EMRT pairs at both extremities of the diagrams were selected, and
307 imported into Markerlynx software for further detailed investigations. In this example, (i.e.
308 *p*,IT vs. *p*,CN), 83 potential discriminant markers were extracted.

309 The twelve OPLS-DA treatments gave approximately 850 potential markers, extracted in
310 both positive and negative ionization modes. This first set of data could be reduced by
311 eliminating the non-significant EMRT pairs (e.g. those close to the noise level). From these
312 850 EMRT pairs, the true discriminating markers were defined after visualizing the whole set
313 of data, through the so-called trend views, i.e. the results of the comparison of the
314 respective abundance (normalized area) of an EMRT pair in all investigated samples.
315 Through visualization of the trend views of each EMRT, potential markers can be selected
316 arbitrarily, by comparing their abundance in the different samples. In a first step, a total of
317 200 components were selected as potential discriminating markers, from the trend views of
318 each EMRT for all investigated samples. As an example, **Figure 7** shows the trend view in
319 positive mode of the EMRT at 6.82 min with an exact mass at 375.1075 DA. This compound
320 can be taken as discriminant marker, because of its abundance in *I. pallida* samples from
321 China and its near absence in all other specimens.

322 Conversely, irigenin, an isoflavonoid characteristic of the *Iris* genus, cannot be taken as a
323 discriminating marker, because its abundance does not vary significantly in the set of
324 samples investigated.

325

326

327

328 3.2 Targeted approach

329 The 200 EMRT can be considered potential diagnostic statistical variables for each species
330 and origin. The possibility to associate one or more ions obtained in positive and negative
331 ionization modes univocally to a single substance and, when possible, to identify those ions,
332 was investigated by comparing the measured exact masses to those in the home-made
333 ~~laboratory-made?~~ orris root database. The exact masses for each component were
334 correlated to the protonated and deprotonated ions $[M+H]^+$, $[M-H]^-$ or all other ionized
335 adducts *e.g.* $[M+NH_4]^+$ or $[M-H+HCOOH]$. 146 compounds were characterized, (table 1
336 Supplementary Material, 1SM) on the basis of their trend views, and used to reprocess the
337 data by the above multivariate analyses. PLS-DA data re-processing with the discriminant
338 markers gave better separations between *Iris* species and origins than those obtained with
339 the whole set of variables. The explained variances increased significantly (by 45% for the
340 positive mode and 58% for the negative mode) while the noise level decreased to
341 approximately 3% (data not shown).

342 UHPLC-TOF-MS analyses were also run in MS^E mode to obtain complementary spectral
343 information. With MS^E acquisition, high resolution spectra at low (parent ions) and high
344 (daughter ions) collision energies can be collected simultaneously. Daughter ions were
345 obtained through the deconvolution of each potential marker, and were tentatively
346 associated to the theoretical fragmentations obtained from the reference compounds in the
347 data bank, thus increasing the reliability of the tentative identification. A number of
348 discriminating markers were then characterized and, where possible, tentatively identified
349 by matching experimental and reference spectra.

350 104 components were found to be specific of a species or an origin, and 42 non-specific. In
351 particular, 15 discriminating components were found to be characteristic of *I. albicans* from
352 Morocco, 12 of *I. germanica* from Morocco, 19 of *I. pallida* originating from Italy, 28 from
353 China, and 30 from Morocco. Common compounds but with significantly different
354 abundances can also be taken as diagnostic for discrimination between species or origins,
355 *e.g.* compound 7 is present in *I. pallida* of all origins, while the triterpenoids 132 ~~132~~ and 137
356 ~~137~~ are present in all species originating from Morocco.

357 ~~52~~ 53 compounds on 104 were tentatively identified as discriminant markers and, on the
358 basis of their trend view profiles, 32 of them were taken as specifically discriminative both
359 for species and for *I. pallida* origins, meaning that their abundance in one or more species or

360 origins is decidedly higher than in the others: they may be present in other species or
361 origins, but in significantly smaller amounts. Table 1 reports the compounds tentatively
362 identified in the methanol/dichloromethane extracts by UHPLC-ESI-TOF-MS in both positive
363 and negative ionization modes that are discriminant for each *Iris* species and origins. The
364 tentatively identified compounds are grouped by species and origins; the peak number was
365 assigned according to the overall elution order in the UHPLC-MS pattern resulting from the
366 reference database (table 1SM). Table 1 also includes retention time (min), exact mass (Da),
367 ionization mode, elemental composition, mass accuracy (ppm), tentative identification of
368 the component with the considered elemental formula, references, and fragments in
369 positive and negative modes, and where available, daughter ions. Most of the tentatively
370 identified components have already been reported in *Iris* genus extracts. These components
371 therefore discriminate the investigated *Iris* species and origins on the basis of the adopted
372 statistical strategy, and may be different from those identified through conventional
373 phytochemical studies and usually used for *Iris* species characterization. Two main classes of
374 secondary metabolites have been detected as discriminating for species and origins within
375 the samples investigated, i.e. phenolic compounds and terpenoids. In particular they
376 respectively include four subclasses of phenolic compounds (flavonoids, isoflavonoids,
377 xanthone and benzophenone derivatives and related glycosides) and a subgroup of
378 triterpenoids, the iridals (i.e. the irone precursors).

379 The following paragraphs list the tentatively identified discriminant markers that can be used
380 to distinguish between different *Iris* species (*I. pallida*, *I. germanica* and *I. albicans*) and/or
381 different origins of *I. pallida* (Italy, China and Morocco).

382 Eight components specifically discriminant for *I. albicans* were tentatively identified, in
383 particular an irisxanthone (18), iriflogenin (61), an irigenin isomer (67), 5-hydroxy-3',4',5'-
384 trimethoxy-6,7-ethylenedioxyisoflavone (71), irisjaponin B isomer 2 (78) and isomer 3 (82),
385 an irisolidone isomer (86) and belamcandin (90).

386 Four discriminating components for *I. germanica* from Morocco were (tentatively) identified
387 as discriminant markers through their trend view profiles, in particular 4'-hydroxy-5-
388 methoxy-6,7-methylenedioxyisoflavone 4'-O-glucoside (43), 4'-hydroxy-5,3'-dimethoxy-6,7-
389 methylene- dioxyisoflavone 4'-O-glucoside (45), a trihydroxy-7-methoxy flavone isomer (66)
390 and an iriflogenin analogue (80).

391 Five components were identified and found to be specific of *I. pallida* from Italy. Two of
392 them belong to the class of triterpenoids, i.e. iriflorental (121) and iripallidal (122), and three
393 are benzophenone derivatives and isoflavonoids, i.e. 2,6,4'-trihydroxy-4-
394 methoxybenzophenone (42) and a 2,6,4'-trihydroxy-4-methoxybenzophenone isomer (51),
395 and a homotectorigenin-7-glucoside isomer (32). The high quality of *I. pallida* from Italy is
396 undoubtedly linked to the greater abundance of iriflorental (121) and iripallidal (122), from
397 which irones are formed during the maturation steps.

398 Five specific discriminating markers were also found for *I. pallida* from China, in particular
399 mangiferin (13), an irisxanthone isomer (23), irisjaponin B isomer 1 (58), and two
400 triterpenoids, i.e. trieneiridal (107) and trieneiridal isomer (115).

401 Ten discriminant components for *I. pallida* from Morocco were found, in particular two
402 iridals, spiroiridal-3 and spiroiridal-3 isomer (98, 101) and two trihydroxy-trimethoxy
403 isoflavone 7-O-(6''-glucosylglucoside) isomers (21, 25), iso-swertiajaponin (26), three iridin
404 isomers (28, 33, 46), a homotectorigenin-7-glucoside isomer (29), and an irigenin isomer
405 (47).

406

407 3.3 Confirmation of the reliability of the metabolomic approach

408 Two unknown samples were submitted to the same analytical procedure as the whole set of
409 previous samples. A predictive statistical data treatment was carried out in order to classify
410 these unknown samples within the different groups. The PLS-DA prediction obtained in
411 positive mode classified the species, but not their origins unequivocally, while that in
412 negative mode clearly associated the two samples to Italian origin. **Figures 8A and 8B** report
413 the PLS-DA diagrams of the total set of samples (including unknowns) obtained using the
414 discriminating markers listed in Table 1 ~~100~~, using EMRT pairs in both positive and negative
415 ionization modes.

416 This predictive approach was confirmed, on the one hand, by comparing the trend views of
417 the 19 discriminating markers of *I. pallida* from Italy (in particular of those tentatively
418 identified *i.e.* 32, 42, 51, 121, 122) in the unknown samples and, on the other, by comparing
419 those identified as characteristic of *I. pallida* from Morocco (*i.e.* 21, 25, 26, 28, 29, 33, 46, 47,
420 98, 101, 105). The profiles of the marker trend views of the unknown samples were found to
421 be very similar to those of reference samples of *I. pallida* from Italy, and completely different

422 from those from Morocco, thus unequivocally classifying origin and species of the unknown
423 samples.

424

425 **4. Conclusions**

426 Modern HPLC and MS instrumentation (UHPLC/TOF-HRMS) in combination with advanced
427 statistical tools open up the field of phytochemical analysis to the use of strategies adopted
428 in metabolomics. The above approaches made it possible to authenticate a natural raw
429 material that is important for the fragrance field, authentication that is difficult with
430 conventional methods: i.e. *Iris* rhizomes and their extracts. The strategy used here (EMRT
431 pairs combined with an advanced statistical strategy) enabled three *Iris* species *I. albicans*, *I.*
432 *germanica*, and *I. pallida*, the latter from three different origins (Italy, China and Morocco),
433 to be discriminated. This discrimination was achieved by detecting 104 discriminating
434 markers; 53 of them were tentatively identified and 32 of them were found to be specific for
435 species and origins. These results can successfully be used to define trend views of specific
436 discriminating markers, to automatically authenticate species and origins, in view of the
437 method's application to routine analysis. Further studies are under way concerning this
438 latter aspect.

439

440 **Acknowledgment**

441 The authors thank Dr. Anne Chartogne from Waters Company for her support and her
442 expertise, and Stéphanie Millo-Lamote from Robertet SA. for her technical support
443 JM, CB, EL, PR are indebted with the project "*Studio di metaboliti secondari biologicamente*
444 *attivi da matrici di origine vegetale*" financially supported by the Ricerca Locale (Ex60%2012)
445 of University of Turin, Turin (Italy).

446

447 **References**

448

449 [1] S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Systematic functional analysis of the yeast
450 genome, Trends Biotechnol. 16 (1998) 373-378.

451 [2] J.M. Halket, D. Waterman, A.M. Przyborowska, R.K.P. Patel, P.D. Fraser, P.M. Bramley,
452 Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and
453 LC/MS/MS, J. Experim. Botany 56 (2005) 219-243.

454 [3] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikstrom, S. Wold, Multi- and
455 Megavariate Data Analysis. Part I. Basic principles and applications (2nd revised and enlarged
456 edition),- Umetrics Academy, Sweden, 2006.

457 [4] J.C. Ellena, M.C. Grasse, L. Peyron, Un jour, une plante : l'Iris dans tous ses états. MIP (Editor),
458 Grasse, 2005.

459 [5] F. Tiemann, P. Kruger, Zum nachweis von ionon und iron, Ber. Dtsch. Chem. Ges. 28 (1895)
460 1754-1757.

461 [6] Y.R. Naves, A.V. Grampoloff, P. Bachmann, Etudes sur les matières végétales volatiles. Etudes
462 dans les méthyl-3-linalols, des méthyl-3-citrals et des méthyl-6-ionones, Helv. Chim. Acta 30
463 (1947) 1599.

464 [7] L. Ruzicka, C.F. Seidel, H. Schinz, M. Pfeiffer, Die konstitution des irons, Helv. Chim. Acta 30
465 (1947) 1807-1810.

466 [8] V. Rautenstrauch, G. Ohloff, Die stereochemie der irone, Helv. Chim. Acta 54 (1971) 1776-
467 1778.

468 [9] V. Rautenstrauch, B. Willhalm, W. Thommen, G. Ohloff, On the stereochemistry of the irones,
469 Helv. Chim. Acta 67 (1984) 325-331.

470 [10] E. Wollenweber, J.F. Stevens, K. Klimo, J. Knauff, N. Frank, C. Gerhauser, Cancer
471 chemopreventive in vitro activities of isoflavones isolated from *Iris germanica*, Planta Med.
472 69 (2003) 15-20.

473 [11] D. Rigano, C. Formisano, A. Grassia, G. Grassia, A. Perrone, S. Piacente, M.L. Vuotto, F.
474 Senatore, Antioxidant flavonoids and isoflavonoids from rhizomes of *Iris pseudopumila*,
475 Planta Med. 73 (2007) 93-96.

476 [12] Atta-ur-Rahman, S. Nasim, I. Baig, S. Jalil, I. Orhan, B. Sener, M.I. Choudhary, Anti-
477 inflammatory isoflavonoids from the rhizomes of *Iris germanica*, J. Ethnopharmacol. 86
478 (2003) 177-180.

479 [13] J.J. Paufigue, Principe actif riche en isoflavones, son procédé d'extraction et ses applications
480 cosmétiques, FR Patent 2 815 539, 2000

481 [14] W. Krick, F.J. Marner, L. Jaenicke, Isolation and structure determination of the precursors of
482 α - and γ -irone and homologous compounds from *Iris pallida* and *Iris florentina*, Z.
483 Naturforsch. 38C (1983) 179-184.

- 484 [15] F.J. Marner, L. Jaenicke, Die absolute konfiguration der iridale und cycloiridale
485 unterschiedlicher herkunft, *Helv. Chim. Acta* 72 (1989) 287-294.
- 486 [16] F.J. Marner, B. Hanisch, Hoogianal, a β -irone precursor from *Iris hoogiana* Dykes (Iridaceae),
487 *Helv. Chim. Acta* 84 (2001) 933-938.
- 488 [17] F.J. Marner, W. Krick, B. Gellrich, L. Jaenicke, W. Winter, Irigermanal and iridogermanal: two
489 new triterpenoids from rhizomes of *Iris germanica* L, *J. Org. Chem.* 47 (1982) 2531-2536.
- 490 [18] F.J. Marner, I. Longerich, Isolation and structure determination of new iridals from *Iris sibirica*
491 and *Iris versicolor*, *Liebigs Ann. Chem.* (1992) 269-272.
- 492 [19] F.J. Marner, K. Simic, B. Scholz, B. Kuster, Isolation and structure determination of a new
493 iridal from *Iris sibirica*, *J. Nat. Prod.* 58 (1995) 299-301.
- 494 [20] F.J. Marner, A. Littek, R. Arold, K. Seferiadis, L. Jaenicke, Isolation and structure
495 determination of new spiro-bicyclic triterpenoids from *Iris pseudacorus*, *Liebigs Ann. Chem.*
496 (1990) 563-567.
- 497 [21] F.J. Marner, D. Gladtko, L. Jaenicke, Studies on the biosynthesis of iridals and cycloiridals,
498 *Helv. Chim. Acta* 71 (1988) 1331-1338.
- 499 [22] F.J. Marner, B. Kerp, Composition of iridals, unusual triterpenoids from sword-lilies, and the
500 seasonal dependence of their content in various parts of different Iris species, *Z. Naturforsch.*
501 47c (1992) 21-25.
- 502 [23] F.J. Marner, I. Ritzdorf, G. Johnen, Metabolism of iridals in plant cell cultures of *Iris pallida*,
503 *Phytochemistry* 33 (1993) 573-579.
- 504 [24] F.J. Marner, T. Kasel, Biomimetic synthesis of the iridal skeleton, *J. Nat. Prod.* 58 (1995) 319-
505 323.
- 506 [25] C. Bicchi, P. Rubiolo, Analysis of constituents of Iris rhizomes. Part I. High performance liquid
507 chromatographic-particle beam-mass spectral analysis of iridals from *Iris pallida* rhizomes,
508 *Phytochem. Anal. analysis* 4 (1993) 171-177.
- 509 [26] A.A. Ali, N.A. El-Emary, M.A. El-Moghazi, F.M. Darwish, A.W. Frahm, Three isoflavonoids from
510 *Iris germanica*, *Phytochemistry* 22 (1983) 2061-2063.
- 511 [27] B. Roger, V. Jeannot, X. Fernandez, S. Cerantola, J. Chahboun, Characterisation and
512 quantification of flavonoids in *Iris germanica* L. and *Iris pallida* Lam. resinoids from Morocco,
513 *Phytochem. Anal. analysis* 23 (2012) 450-455.
- 514 [28] A. Guenet, Comparaison des beurres d'iris de diverses origines botaniques et géographiques,
515 12èmes Journées Internationales Huiles Essentielles (Digne, 1993) (1993) 492-492.
- 516 [29] M. Remy, J.C. Bayle, M. Derbesy, R. Uzio, H. Casabianca, J.B. Graff, Les beurres d'iris. Dosage
517 des irones par chromatographie gazeuse sur phase chirale, vérification des origines, Journées
518 Internationales Huiles Essentielles 12èmes (Digne, 1993) (1993) 506-506.
- 519 [30] Y. Nakamura, H. Asahi, Md. Altaf-Ul-Amin, K. Kurokawa, S. Kanaya, KNApSACK,
520 <http://kanaya.aist-nara.ac.jp/KNApSACK/> (2013).

- 521 [31] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.M. Fan, O. Fiehn,
522 R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane,
523 J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum
524 reporting standards for chemical analysis Chemical Analysis Working Group (CAWG)
525 Metabolomics Standards Initiative (MSI), *Metabolomics* 3 (2007) 211-221.
- 526 [32] D.J. Creek, W.B. Dunn, O. Fiehn, J.L. Griffin, R.D. Hall, Z. Lei, R. Mistrik, S. Neumann, E.L.
527 Schymanski, L.W. Sumner, Metabolite identification: are you sure? And how do your peers
528 gauge your confidence, *Metabolomics* 10 (2014) 350-353.
- 529 [33] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying
530 small molecules via high resolution mass spectrometry: Communicating confidence, *Environ.*
531 *Sci. Technol.* (2014) 2097-2098.
- 532 [34] M. Arisawa, N. Morita, Y. Kondo, T. Takemoto, Studies on constituents of genus *Iris*. VII. The
533 constituents of *Iris unguicularis* Poir, *Chem. Pharm. Bull.* 21 (1973) 2323-2328.
- 534 [35] M. Fujita, T. Inoue, Studies on the constituents of *Iris florentina* L. II. C-Glycosides of
535 xanthenes and flavones from the leaves, *Chem. Pharm. Bull.* 30 (1982) 2342-2348.
- 536 [36] J.B. Harborne, H. Baxter, G.P. Moss, Xanthenes in J.B. Harborne, H. Baxter, G.P. Moss (Eds),
537 *Phytochemical Dictionary. A Handbook of Bioactive Compounds from Plants* (2nd ed.), Taylor
538 & Francis, Padstow, 1999, p. 582-590.
- 539 [37] H. Minami, A. Okubo, M. Kodama, Y. Fukuyama, Highly oxygenated isoflavones from *Iris*
540 *japonica*, *Phytochemistry* 41 (1996) 1219-1222.
- 541 [38] A. Littek, F.J. Marner, Structure determination of new iridals from *Iris pallida* and *Iris*
542 *foetidissima*, *Helv. Chim. Acta* 74 (1991) 2035-2037.
- 543 [39] Atta-ur-Rahman, S. Nasim, I. Baig, I. Orhan, B. Sener, F. Ayanoglu, M.I. Choudhary,
544 Isoflavonoid glycosides from the rhizomes of *Iris germanica*, *Helv. Chim. Acta* 86 (2003) 3354-
545 3362.
- 546 [40] J.W. Wallace, G. Morris, C-glycosylflavones in *Gnetum gnemon*, *Phytochemistry* 17 (1978)
547 1809-1810.
- 548 [41] J.B. Harborne, H. Baxter, C-Glycosylflavones in J.B. Harborne, H. Baxter (Editors), *The*
549 *Handbook of Natural Flavonoids* (Vol. 1), Wiley, Chichester, 1999, p. 549-644..
- 550 [42] P. Kassak, Secondary metabolites of the chosen genus *Iris* species, *Acta Univ. Agricult.*
551 *Silvicult. Mendalianae Brunensis* 60 (2012) 269-280.
- 552 [43] J.B. Harborne, H. Baxter, G.P. Moss, Isoflavonoids and neoflavonoids in J.B. Harborne,
553 H. Baxter, G.P. Moss (Editors), *Phytochemical Dictionary - A Handbook of Bioactive*
554 *Compounds from Plants* (2nd ed.). Taylor & Francis, Padstow, 1999, p. 458-477.
- 555 [44] H. Ito, S. Onoue, T. Yoshida, Isoflavonoids from *Belamcanda chinensis*, *Chem. Pharm. Bull.* 49
556 (2001) 1229-1231.
- 557 [45] Atta-ur-Rahman, S. Nasim, I. Baig, I.A. Jahan, B. Sener, I. Orhan, M.I. Choudhary, Isoflavonoid
558 glycosides from the rhizomes of *Iris germanica*, *Chem. Pharm. Bull.* 50 (2002) 1100-1102.

- 559 [46] J.B. Harborne, H. Baxter, Isoflavonoids and neoflavonoids in J.B.Harborne, H.Baxter (Editors),
560 The Handbook of Natural Flavonoids (Vol. 2). Wiley, Chichester, 1999, p. 517-708.
- 561 [47] A. Kawase, N. Ohta, K. Yagishita, On the chemical structure of a new isoflavone glucoside,
562 homotectoridin, isolated together with tectoridin from the rhizomes of *Iris germanica*
563 Linnaeus, Agric. Biol. Chem. 37 (1973) 145-147.
- 564 [48] M. Arisawa, N. Morita, Studies on constituents of *Iris* genus plants. IV. The constituents of *Iris*
565 *florentina* L. (2), Chem. Pharm. Bull. 24 (1976) 815-817.
- 566 [49] M. Moriyasu, Y. Igi, M. Ichimaru, K. Iwasa, J. Kobayakawa, F. Sato-Nishimori, Y. Matsukawa,
567 C. Nagaze, New isoflavones from *Belamcandae* rhizome, J. Nat. Med. 61 (2007) 329-333.
- 568 [50] M.I. Choudhary, M. Nur-e-Alam, I. Baig, F. Akhtar, A. Khan, P. Ndogonii, T. Badarchiin, G.
569 Purevsuren, N. Nahar, Atta-ur-Rahman, Four new flavones and a new isoflavone from *Iris*
570 *bungei*, J. Nat. Prod. 64 (2001) 857-860.
- 571 [51] K.L. Dhar, A.K. Kalla, Isoflavones of *Iris kumaonensis* and *Iris germanica*, Phytochemistry 11
572 (1972) 3097-3098.
- 573 [52] U. Mahmood, V.K. Kaul, L. Jirovetz, Alkylated benzoquinones from *Iris kumaonensis*,
574 Phytochemistry 61 (2002) 923-926.
- 575 [53] Y. Miyake, H. Ito, T. Yoshida, Identification of iridals as piscicidal components of iradaceous
576 plants and their conformations associated with CD spectra, Canadian J. Chem. 75 (1997) 734-
577 741.
- 578 [54] F. Abe, R.F. Chen, T. Yamauchi, Iridals from *Belamcanda chinensis* and *Iris japonica*,
579 Phytochemistry 30 (1991) 3379-3382.
- 580 [55] L. Jaenicke, F.J. Marner, On irones, dihydroirones and their precursor, Topics in Flavour
581 Research (Proc. International conference Freising-Weihestephan), (1985), p. 253-253.
- 582 [56] W. Krick, F.J. Marner, L. Jaenicke, On the stereochemistry of natural irones, dihydroirones
583 and their precursors, Helv. Chim. Acta 67 (1984) 318-324.
- 584 [57] O. Potterat, C. Herzog, M. Raith, S.N. Ebrahimi, M. Hamburger, Irigermanone, a noriridal with
585 unprecedented methylketone function, from *Iris germanica*, Helv. Chim. Acta 97 (2014) 32-
586 38.
- 587 [58] W.S. Woo, E.H. Woo, An isoflavone noririsflorentin from *Belamcanda chinensis*,
588 Phytochemistry 33 (1993) 939-940.
- 589 [59] J.B. Harborne, H.Baxter, Flavonones and dihydroflavonols in J.B.Harborne, H.Baxter (Editors),
590 The Handbook of Natural Flavonoids (Vol. 2). Wiley, Chichester, 1999, p. 218-354.
591
592

593 Captions to figures

594 Figure 1 Main skeletons of secondary metabolites identified in the *Iris* genus.

595 Figure 2 HPLC/DAD-UV chromatographic profile of orris extracts (for analysis conditions see
596 text).

597 Figure 3 UHPLC/TOF-MS chromatographic profiles of the investigated orris root extracts: a)
598 in positive mode, b) in negative mode. For analysis conditions see text.

599 Figure 4 PCA score plot with pareto scaling of the investigated orris root samples analyzed by
600 UHPLC-TOF-MS in positive mode.

601 Figure 5: PLS-DA score plot with pareto scaling of the investigated orris root samples
602 analyzed by UHPLC-TOF-MS a) in positive mode, b) in negative mode.

603 Figure 6 OLPS-DA and S-plot diagram from the comparison between *Iris pallida* Italy vs. *Iris*
604 *pallida* China samples.

605 Figure 7 Trend view of a selected discriminating marker (EMRT 6.82, 375.1075) in the
606 investigated samples.

607 Figure 8: PLS-DA prediction of unknown orris root samples with pareto scaling: a) with
608 dataset obtained by UHPLC-TOF-MS in positive mode, b) with dataset obtained by UHPLC-
609 TOF-MS in negative mode.

610

611

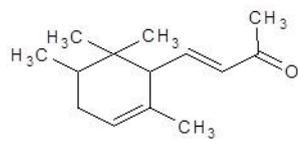
Table 1 List of the tentatively identified discriminant markers in agreement with species and origin, and UHPLC retention times. Legend: *I. pallida* China (p, CN), *I. pallida* Morocco (p, MA), *I. pallida* Italy (p, IT), *I. albicans* Morocco (a, MA), *I. germanica* Morocco (g, MA).

N°	Rt (min)	Exact mass (Da)	Ioniz. mode	Elemental compos.	Mass accuracy (ppm)	Tentative identification	References	Species and/or origin	Fragments in pos. mode			Fragments in neg. mode			
									Exact mass (Da)	Element. compos.	Mass accuracy (ppm)	Exact mass (Da)	Element. compos.	Mass accuracy (ppm)	
1	13	4.38	422.0838	±	C ₁₉ H ₁₈ O ₁₁	2.63	Mangiferin	[34], [35]	p, CN	327.0578 303.0516 273.0416	C ₁₇ H ₁₁ O ₇ C ₁₅ H ₁₁ O ₇ C ₁₄ H ₉ O ₆	22.39 3.703 6.18	331.0446 301.0338	C ₆ H ₁₁ O ₈ C ₁₅ H ₉ O ₇	2.39 3.41
2	23	5.30	436.1004	-	C ₂₀ H ₂₀ O ₁₁	0.37	Irisxanthone isomer	[36]	p, CN				315.0505 272.0344	C ₁₆ H ₁₁ O ₇ C ₁₄ H ₈ O ₆	0.07 8.49
3	58	6.82	374.0997	±	C ₁₉ H ₁₈ O ₈	1.25	irisjaponin B isomer 1	[37]	p, CN	360.0849 342.0740	C ₁₉ H ₁₆ O ₈ C ₁₈ H ₁₄ O ₇	1.06 0.14			
4	107	10.75	456.3595	±	C ₃₀ H ₄₈ O ₃	1.85	Trieneiridal isomer 1	[24], [38]	p, CN	439.3567 421.3433	C ₃₀ H ₄₇ O ₂ C ₃₀ H ₄₆ O	2.06 8.80			
5	115	11.49	456.3599	±	C ₃₀ H ₄₈ O ₃	0.98	Trieneiridal isomer 2	[24], [38]	p, CN	439.3584 421.3474	C ₃₀ H ₄₇ O ₂ C ₃₀ H ₄₅ O	1.81 0.85			
6	21	5.12	684.1897	±	C ₃₀ H ₃₆ O ₁₈	0.68	5,7,3'-Trihydroxy-6,4',5'-trimethoxy isoflavone 7-O-(6"-glucosyl glucoside) isomer 1	[39]	p, MA	523.1453 361.0920 331.0825	C ₂₄ H ₂₇ O ₁₃ C ₁₈ H ₁₇ O ₈ C ₁₇ H ₁₅ O ₇	0.26 0.95 2.18	521.1276 359.0747 329.0656	C ₂₄ H ₂₅ O ₁₃ C ₁₈ H ₁₅ O ₈ C ₁₇ H ₁₃ O ₇	3.67 5.55 1.60
7	25	5.41	684.1897	±	C ₃₀ H ₃₆ O ₁₈	0.68	5,7,3'-Trihydroxy-6,4',5'-trimethoxy isoflavone 7-O-(6"-glucosyl glucoside) isomer 2	[39]	p, MA	523.1453 361.0914	C ₂₄ H ₂₇ O ₁₃ C ₁₈ H ₁₇ O ₈	0.26 2.61	359.0751	C ₁₈ H ₁₅ O ₈	4.44
8	26	5.46	462.1162	±	C ₂₂ H ₂₂ O ₁₁	0.03	Iso-swertiajaponin	[40], [41]	p, MA	301.0711 286.0474	C ₁₆ H ₁₃ O ₆ C ₁₅ H ₁₀ O ₆	0.38 1.18	299.0559 284.0325	C ₁₆ H ₁₁ O ₆ C ₁₅ H ₈ O ₆	1.13 1.45

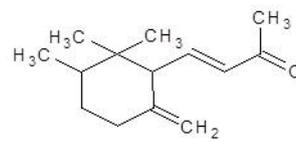
N°	Rt (min)	Exact mass (Da)	Ioniz. mode	Elemental compos.	Mass accuracy (ppm)	Tentative identification	References	Species and/or origin	Fragments in pos. mode			Fragments in neg. mode			
									Exact mass (Da)	Element. compos.	Mass accuracy (ppm)	Exact mass (Da)	Element. compos.	Mass accuracy (ppm)	
9	<u>28</u>	<u>5.56</u>	<u>522.1367</u>	±	<u>C₂₄H₂₆O₁₃</u>	<u>1.23</u>	<u>Iridin isomer 1</u>	<u>[25], [26], [42], [43], [44], [45]</u>	<u>p, MA</u>	<u>361.0921</u>	<u>C₁₈H₁₇O₈</u>	<u>0.67</u>			
10	<u>29</u>	<u>5.60</u>	<u>492.1262</u>	±	<u>C₂₃H₂₄O₁₂</u>	<u>1.17</u>	<u>Homotectorigenin-7-O-glucoside isomer</u>	<u>[46], [47]</u>	<u>p, MA</u>	<u>331.0806</u> <u>316.0572</u>	<u>C₁₇H₁₅O₇</u> <u>C₁₆H₁₂O₇</u>	<u>3.56</u> <u>3.49</u>	<u>329.0668</u> <u>314.0424</u>	<u>C₁₇H₁₃O₇</u> <u>C₁₆H₁₀O₇</u>	<u>2.04</u> <u>0.81</u>
11	<u>33</u>	<u>5.89</u>	<u>522.1367</u>	-	<u>C₂₄H₂₆O₁₃</u>	<u>1.23</u>	<u>Iridin isomer 2</u>	<u>[25], [26], [42], [43], [44], [45]</u>	<u>p, MA</u>				<u>506.1063</u> <u>359.0762</u> <u>344.0526</u>	<u>C₂₃H₂₂O₁₃</u> <u>C₁₈H₁₅O₈</u> <u>C₁₇H₁₂O₈</u>	<u>0.51</u> <u>1.37</u> <u>1.79</u>
12	<u>46</u>	<u>6.40</u>	<u>522.1368</u>	-	<u>C₂₄H₂₆O₁₃</u>	<u>1.04</u>	<u>Iridin</u>	<u>[25], [26], [42], [43], [44], [45]</u>	<u>p, MA</u>				<u>506.1054</u> <u>359.0762</u> <u>344.0529</u>	<u>C₂₃H₂₂O₁₃</u> <u>C₁₈H₁₅O₈</u> <u>C₁₇H₁₂O₈</u>	<u>1.27</u> <u>1.37</u> <u>0.92</u>
13	<u>47</u>	<u>6.41</u>	<u>360.0842</u>	±	<u>C₁₈H₁₆O₈</u>	<u>0.88</u>	<u>Irigenin isomer</u>	<u>[25], [26], [30], [42], [48], [49], [50], [45], [51], [52]</u>	<u>p, MA</u>	<u>346.0695</u>	<u>C₁₇H₁₄O₈</u>	<u>1.83</u>			
14	<u>98</u>	<u>9.98</u>	<u>470.3390</u>	±	<u>C₃₀H₄₆O₄</u>	<u>1.30</u>	<u>Spiroiridal-3</u>	<u>[22], [23], [25], [38], [53], [54]</u>	<u>p, MA</u>	<u>453.3354</u> <u>219.1380</u>	<u>C₃₀H₄₅O₃</u> <u>C₁₄H₁₉O₂</u>	<u>3.24</u> <u>2.30</u>			
15	<u>101</u>	<u>10.22</u>	<u>470.3388</u>	±	<u>C₃₀H₄₆O₄</u>	<u>1.72</u>	<u>Spiroiridal-3 isomer</u>	<u>[22], [23], [25], [38], [53], [54]</u>	<u>p, MA</u>	<u>453.3353</u> <u>219.1357</u>	<u>C₃₀H₄₅O₃</u> <u>C₁₄H₁₉O₂</u>	<u>3.46</u> <u>12.80</u>			
16	<u>32</u>	<u>5.84</u>	<u>492.1261</u>	±	<u>C₂₃H₂₄O₁₂</u>	<u>1.37</u>	<u>Homotectorigenin-7-O-glucoside isomer</u>	<u>[46], [47]</u>	<u>p, IT</u>	<u>331.0818</u> <u>316.0587</u>	<u>C₁₇H₁₅O₇</u> <u>C₁₆H₁₂O₇</u>	<u>0.24</u> <u>1.26</u>	<u>329.0652</u> <u>314.0411</u>	<u>C₁₇H₁₃O₇</u> <u>C₁₆H₁₀O₇</u>	<u>2.81</u> <u>4.94</u>
17	<u>42</u>	<u>6.21</u>	<u>260.0684</u>	±	<u>C₁₄H₁₂O₅</u>	<u>0.28</u>	<u>2,6,4'-trihydrox-4-methoxybenzo phenone</u>	<u>[27]</u>	<u>p, IT</u>	<u>167.0355</u> <u>121.0301</u> <u>93.0345</u>	<u>C₈H₇O₄</u> <u>C₇H₅O₂</u> <u>C₆H₅O</u>	<u>6.38</u> <u>9.46</u> <u>4.95</u>	<u>165.0188</u> <u>121.0309</u> <u>93.0344</u>	<u>C₈H₅O₄</u> <u>C₇H₅O₂</u> <u>C₆H₅O</u>	<u>0.11</u> <u>6.07</u> <u>3.87</u>
18	<u>51</u>	<u>6.45</u>	<u>260.0682</u>	-	<u>C₁₄H₁₂O₅</u>	<u>1.05</u>	<u>2,6,4'-trihydrox-4-methoxybenzo-phenone isomer</u>	<u>[27]</u>	<u>p, IT</u>				<u>243.0303</u> <u>215.0727</u> <u>151.0034</u>	<u>C₁₃H₇O₅</u> <u>C₁₈H₅O₂</u> <u>C₇H₃O₄</u>	<u>3.92</u> <u>5.32</u> <u>1.76</u>
19	<u>121</u>	<u>11.88</u>	<u>486.3704</u>	±	<u>C₃₁H₅₀O₄</u>	<u>1.05</u>	<u>Iriflorentale</u>	<u>[14], [23], [25], [55], [56], [57]</u>	<u>p, IT</u>	<u>469.3665</u> <u>451.3565</u> <u>217.1967</u> <u>191.1811</u>	<u>C₃₁H₄₉O₃</u> <u>C₃₁H₄₇O₂</u> <u>C₁₆H₃₅</u> <u>C₁₄H₃₃</u>	<u>3.56</u> <u>2.45</u> <u>6.78</u> <u>5.88</u>	<u>467.3533</u> <u>437.3425</u> <u>237.1497</u>	<u>C₃₁H₄₇O₃</u> <u>C₃₀H₄₅O₂</u> <u>C₁₄H₂₁O₃</u>	<u>1.67</u> <u>1.25</u> <u>2.66</u>

N°	Rt (min)	Exact mass (Da)	Ioniz. mode	Elemental compos.	Mass accuracy (ppm)	Tentative identification	References	Species and/or origin	Fragments in pos. mode			Fragments in neg. mode			
									Exact mass (Da)	Element. compos.	Mass accuracy (ppm)	Exact mass (Da)	Element. compos.	Mass accuracy (ppm)	
<u>20</u>	<u>122</u>	<u>12.06</u>	<u>486.3701</u>	±	<u>C₃₁H₅₀O₄</u>	<u>1.67</u>	<u>Iripallidal</u>	<u>[14], [23], [25], [55], [56], [57]</u>	<u>p, IT</u>	<u>469.3655</u> <u>451.3556</u> <u>255.2309</u> <u>217.1967</u> <u>191.1814</u>	<u>C₃₁H₄₉O₃</u> <u>C₃₁H₄₇O₂</u> <u>C₁₆H₃₁O₂</u> <u>C₁₆H₃₅</u> <u>C₁₄H₃₃</u>	<u>5.69</u> <u>4.44</u> <u>5.89</u> <u>9.09</u> <u>7.48</u>			
<u>21</u>	<u>18</u>	<u>5.06</u>	<u>436.1001</u>	-	<u>C₂₀H₂₀O₁₁</u>	<u>1.06</u>	<u>Irisxanthone</u>	<u>[36]</u>	<u>a, MA</u>				<u>345.0606</u> <u>315.0500</u> <u>272.0319</u>	<u>C₁₇H₁₃O₈</u> <u>C₁₆H₁₁O₇</u> <u>C₁₄H₈O₆</u>	<u>1.28</u> <u>1.52</u> <u>0.69</u>
<u>22</u>	<u>61</u>	<u>6.97</u>	<u>328.0580</u>	-	<u>C₁₇H₁₂O₇</u>	<u>0.92</u>	<u>Iriflogenin</u>	<u>[42], [27]</u>	<u>a, MA</u>				<u>312.0272</u> <u>255.0288</u> <u>240.0428</u>	<u>C₁₆H₈O₇</u> <u>C₁₄H₇O₅</u> <u>C₁₄H₈O₄</u>	<u>0.63</u> <u>2.15</u> <u>2.25</u>
<u>23</u>	<u>67</u>	<u>7.40</u>	<u>360.0840</u>	±	<u>C₁₈H₁₆O₈</u>	<u>1.44</u>	<u>Irigenin isomer</u>	<u>[25], [26], [30], [42], [48], [49], [50], [45], [51], [52]</u>	<u>a, MA</u>	<u>346.0690</u> <u>328.0580</u> <u>313.0346</u>	<u>C₁₇H₁₄O₈</u> <u>C₁₇H₁₂O₇</u> <u>C₁₆H₉O₇</u>	<u>0.38</u> <u>0.92</u> <u>0.73</u>	<u>344.0533</u> <u>329.0302</u>	<u>C₁₇H₁₂O₈</u> <u>C₁₆H₉O₈</u>	<u>0.24</u> <u>1.39</u>
<u>24</u>	<u>71</u>	<u>7.70</u>	<u>372.0838</u>	±	<u>C₁₉H₁₆O₈</u>	<u>1.93</u>	<u>5-hydroxy-3',4',5'-trimethoxy-6,7-methylenedioxy isoflavone</u>	<u>[58]</u>	<u>a, MA</u>	<u>358.0693</u>	<u>C₁₈H₁₄O₈</u>	<u>1.21</u>			
<u>25</u>	<u>78</u>	<u>8.21</u>	<u>374.0997</u>	±	<u>C₁₉H₁₈O₈</u>	<u>1.25</u>	<u>Irisjaponin isomer 2</u>	<u>[37]</u>	<u>a, MA</u>	<u>360.0820</u> <u>345.0623</u> <u>342.0740</u>	<u>C₁₈H₁₆O₈</u> <u>C₁₇H₁₃O₈</u> <u>C₁₈H₁₄O₇</u>	<u>6.99</u> <u>3.65</u> <u>0.14</u>			
<u>26</u>	<u>82</u>	<u>8.36</u>	<u>374.0997</u>	±	<u>C₁₉H₁₈O₈</u>	<u>1.25</u>	<u>Irisjaponin isomer 3</u>	<u>[37]</u>	<u>a, MA</u>	<u>360.0823</u> <u>345.0618</u> <u>342.0741</u>	<u>C₁₈H₁₆O₈</u> <u>C₁₇H₁₃O₈</u> <u>C₁₈H₁₄O₇</u>	<u>6.16</u> <u>2.20</u> <u>0.43</u>			
<u>27</u>	<u>86</u>	<u>8.62</u>	<u>314.0787</u>	-	<u>C₁₇H₁₄O₆</u>	<u>1.08</u>	<u>Irisolidone isomer</u>	<u>[26], [27], [42], [43], [46], [37]</u>	<u>a, MA</u>				<u>298.0474</u> <u>283.0242</u> <u>255.0287</u>	<u>C₁₆H₁₀O₆</u> <u>C₁₅H₇O₆</u> <u>C₁₄H₇O₅</u>	<u>1.13</u> <u>0.22</u> <u>2.54</u>
<u>28</u>	<u>90</u>	<u>8.86</u>	<u>358.1045</u>	±	<u>C₁₉H₁₈O₇</u>	<u>2.10</u>	<u>Belamcandin</u>	<u>[46]</u>	<u>a, MA</u>	<u>344.0893</u> <u>327.0868</u> <u>315.0871</u> <u>300.0637</u>	<u>C₁₈H₁₆O₇</u> <u>C₁₈H₁₅O₆</u> <u>C₁₇H₁₅O₆</u> <u>C₁₆H₁₂O₆</u>	<u>0.88</u> <u>0.19</u> <u>0.75</u> <u>1.04</u>			

	<i>N</i> ^o	<i>Rt</i> (min)	<u>Exact mass</u> (Da)	<u>Ioniz. mode</u>	<u>Elemental compos.</u>	<u>Mass accuracy</u> (ppm)	<u>Tentative identification</u>	<u>References</u>	<u>Species and/or origin</u>	<u>Fragments in pos. mode</u>			<u>Fragments in neg. mode</u>		
										<u>Exact mass</u> (Da)	<u>Element. compos.</u>	<u>Mass accuracy</u> (ppm)	<u>Exact mass</u> (Da)	<u>Element. compos.</u>	<u>Mass accuracy</u> (ppm)
29	43	6.28	474.1160	±	C ₂₃ H ₂₂ O ₁₁	0.45	4'-Hydroxy-5-methoxy-6,7-methylenedioxy isoflavone 4'-O-glucoside	[42], [45]	g, MA	313.0714	C ₁₇ H ₁₃ O ₆	0.6	311.0518	C ₁₇ H ₁₁ O ₆	0.20
30	45	6.39	504.1263	±	C ₂₄ H ₂₄ O ₁₂	0.95	4'-Hydroxy-5,3'-dimethoxy-6,7-methylenedioxy isoflavone 4'-O-glucoside	[42], [45]	g, MA	343.0819 328.0580	C ₁₈ H ₁₅ O ₇ C ₁₇ H ₁₂ O ₇	0.36 0.92			
31	66	7.37	302.0790	-	C ₁₆ H ₁₄ O ₆	0.13	5,8,2'-Trihydroxy-7-methoxyflavone	[59]	g, MA						
32	80	8.30	328.0579	±	C ₁₇ H ₁₂ O ₇	1.23	Iriflogenin isomer	[42], [27]	g, MA	314.0423 297.0398 180.0054	C ₁₆ H ₁₀ O ₇ C ₁₆ H ₉ O ₇ C ₈ H ₄ O ₅	1.12 0.38 2.63			

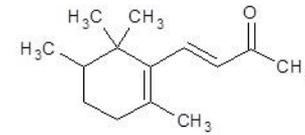


α -Irone

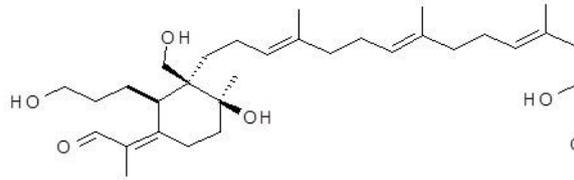


γ -Irone

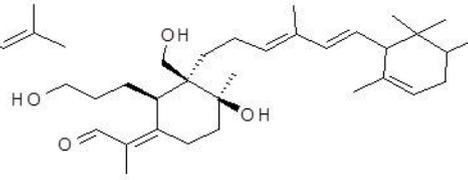
Irone skeletons



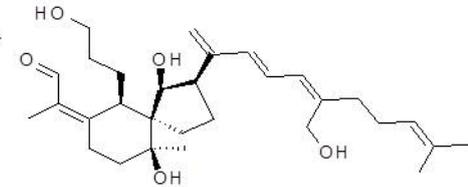
β -Irone



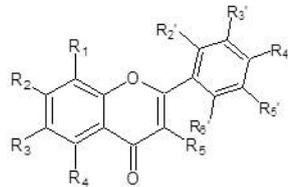
Iridal (mono-cyclic)



Iripallidal (bi-cyclic)
Triterpenoid skeletons



Spiroiridal-1 (spiro-cyclic)



Flavonoid

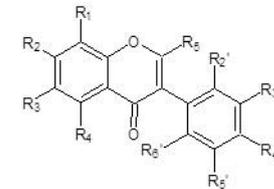
As examples: Iso-swertijaponin

$R_1 = \text{Glu}$

$R_2 = \text{OMe}$

$R_3, R_5, R_2', R_3', R_6' = \text{H}$

$R_4, R_4', R_5' = \text{OH}$



Isoflavonoid

Irigenin

$R_1, R_5, R_2', R_6' = \text{H}$

$R_2, R_4, R_3' = \text{OH}$

$R_3, R_4', R_5' = \text{OMe}$

Flavonoid and isoflavonoid skeletons

Figure 1

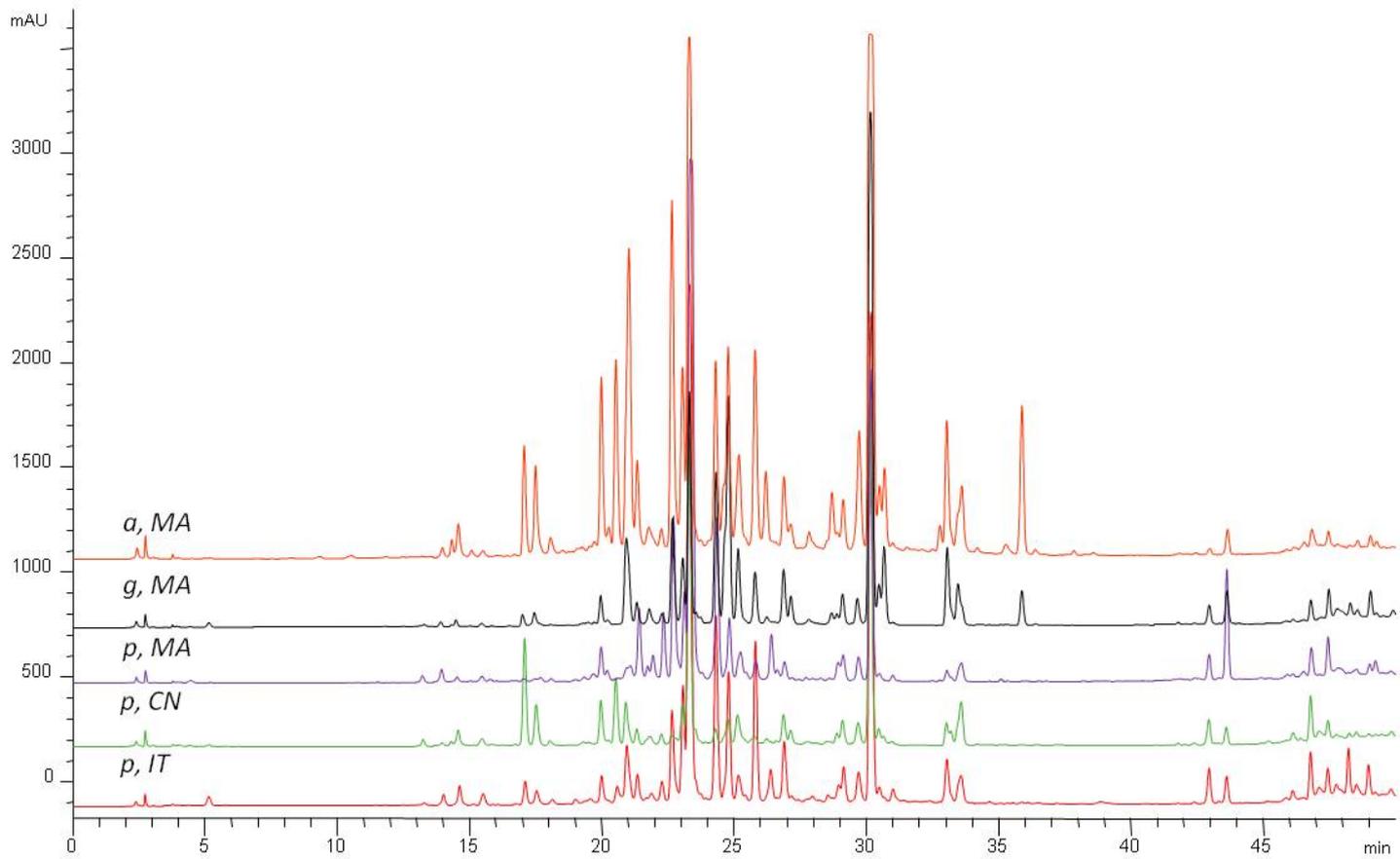
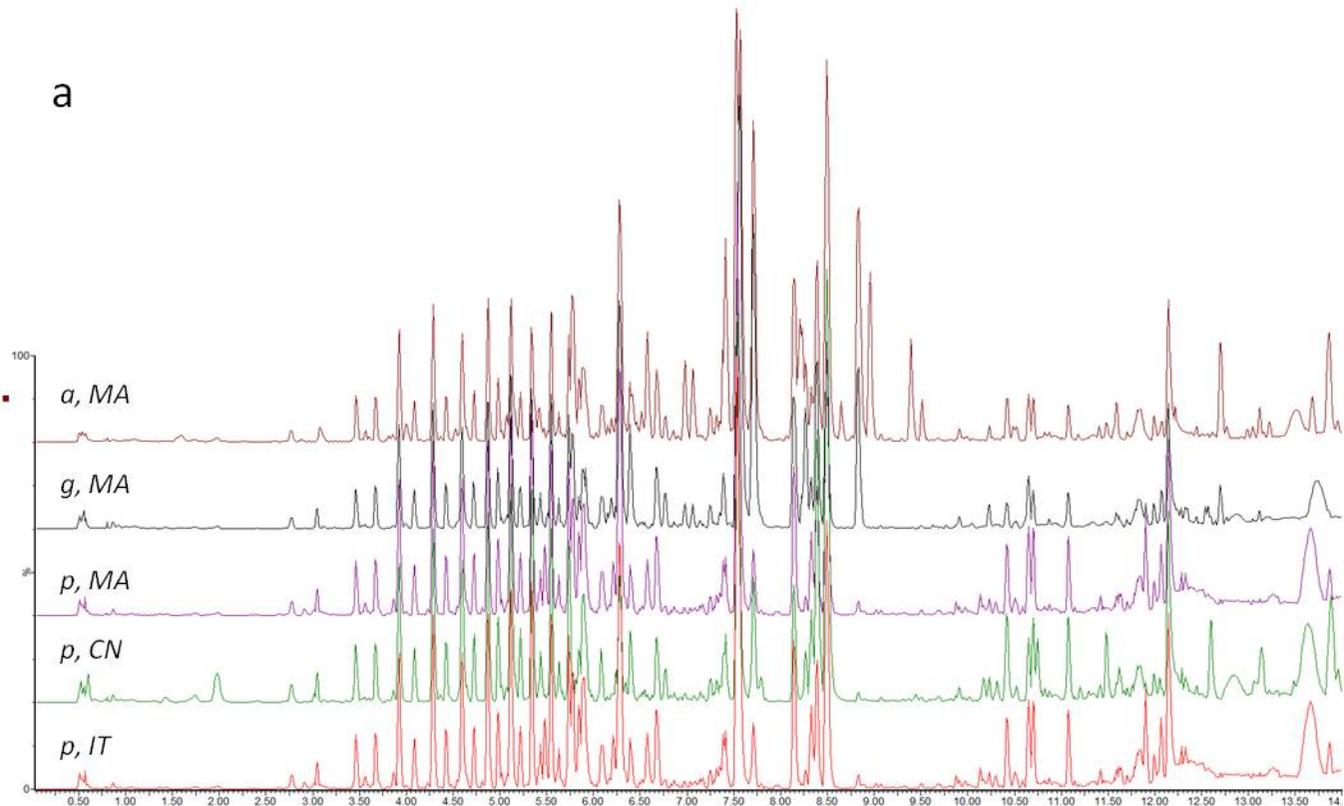
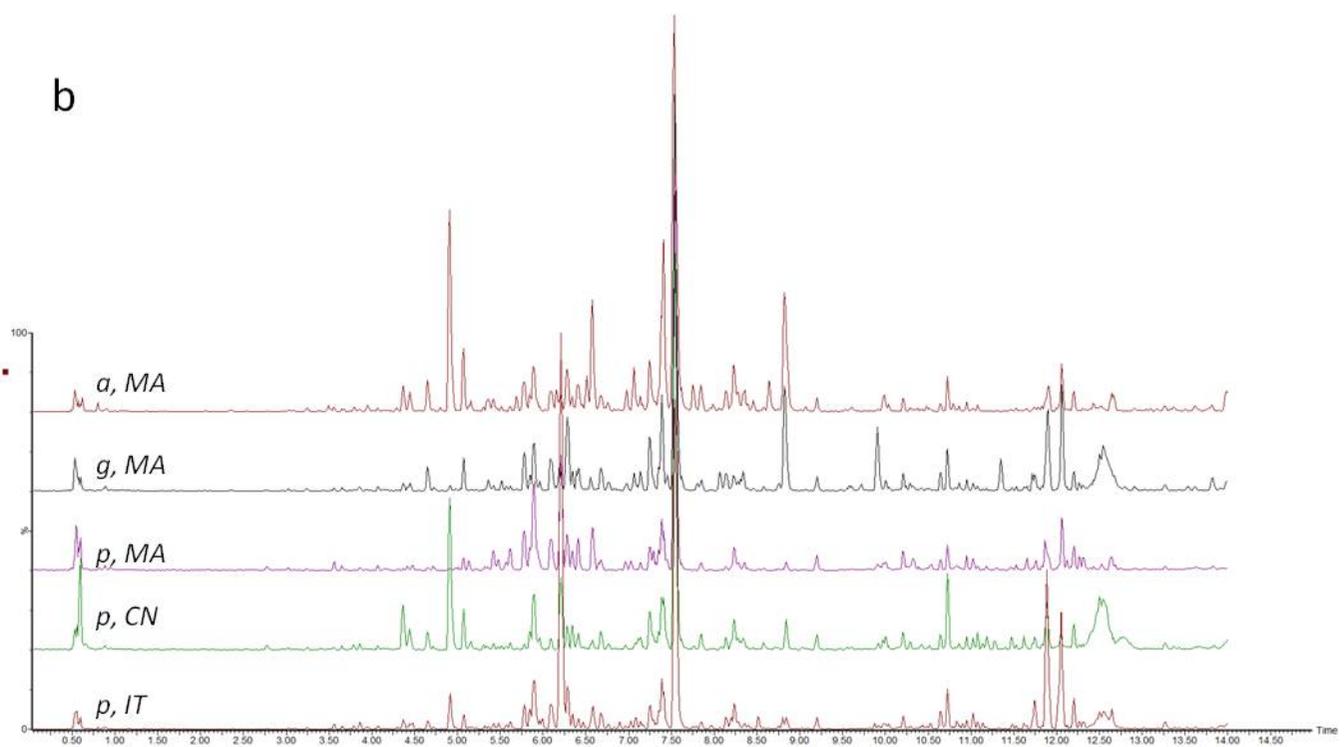


Figure 2

a



b



Figures 3

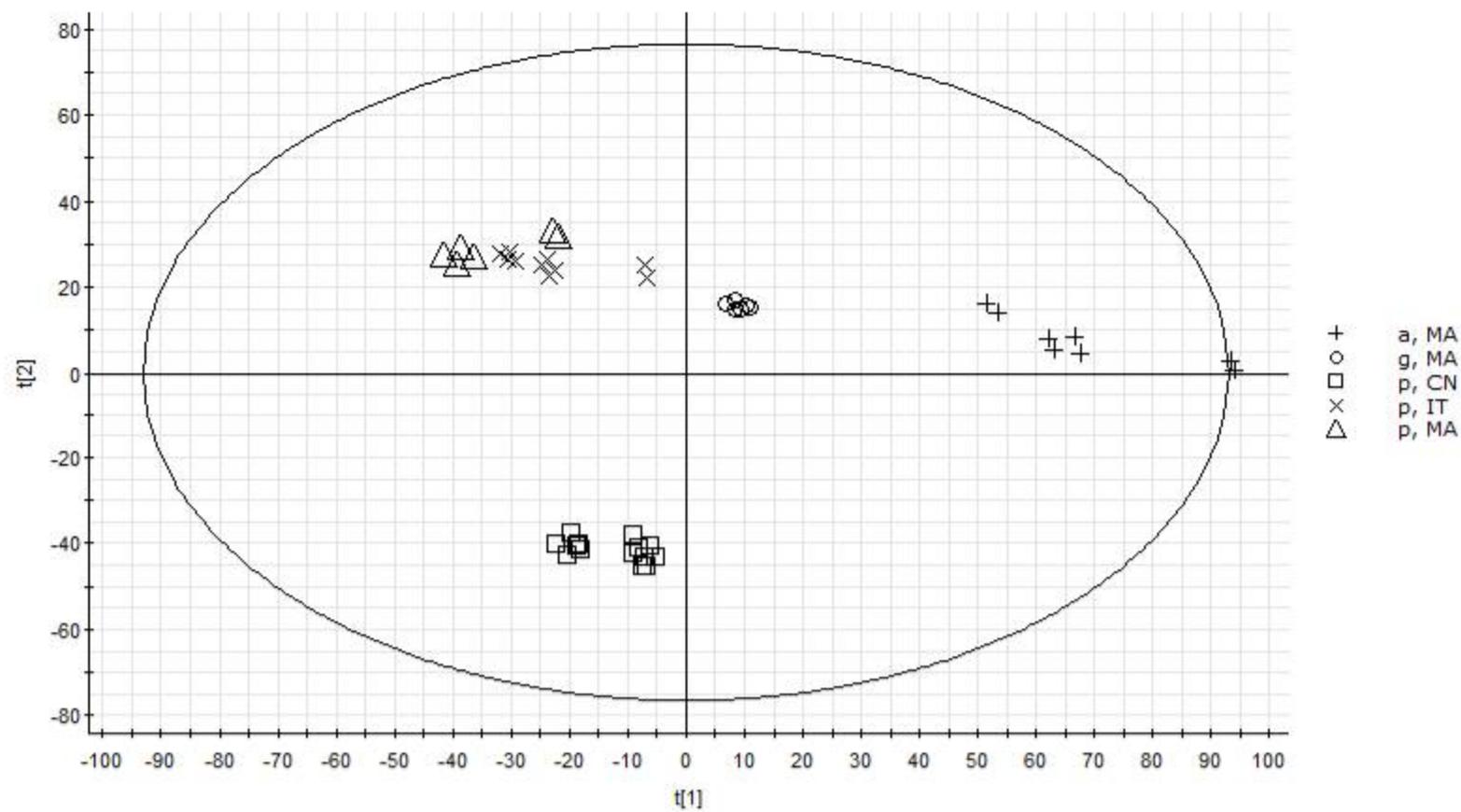
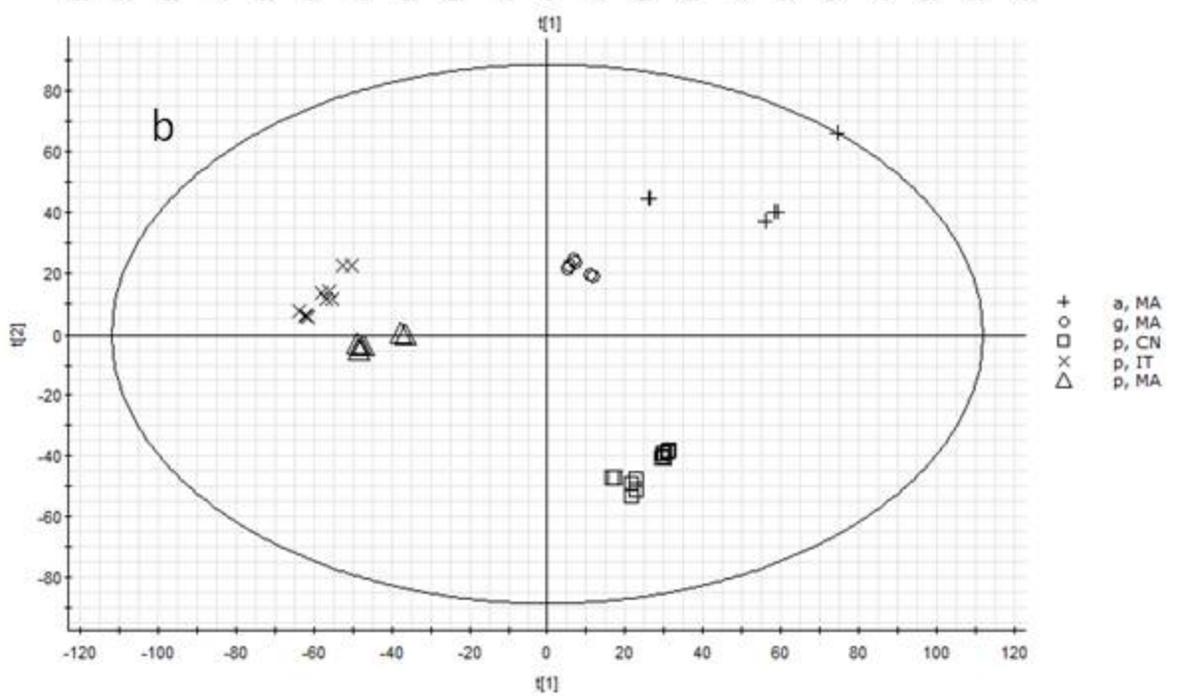
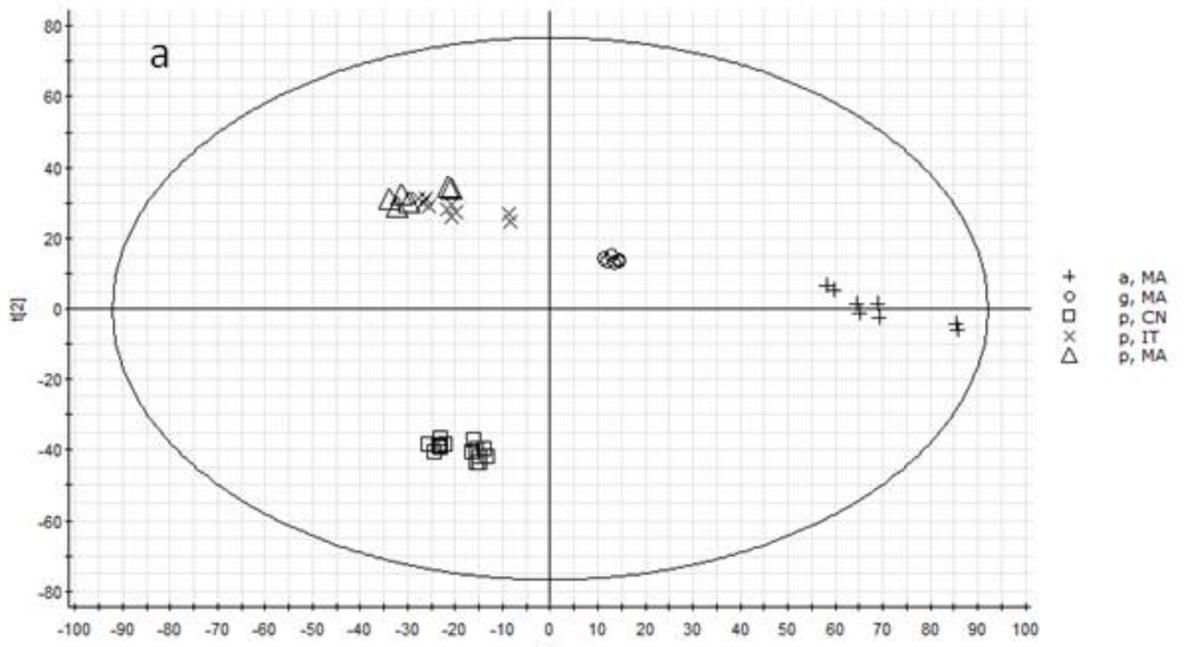


Figure 4



Figures 5

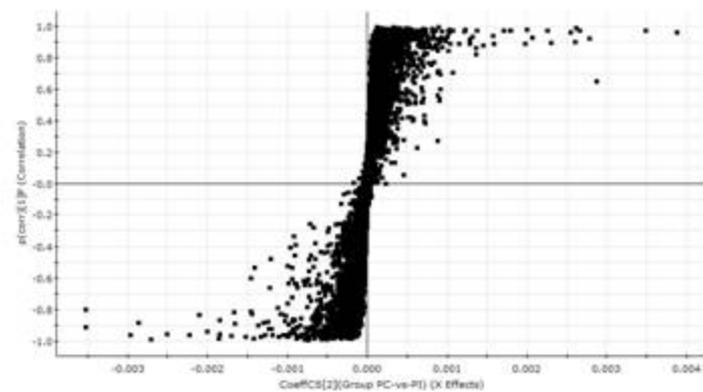
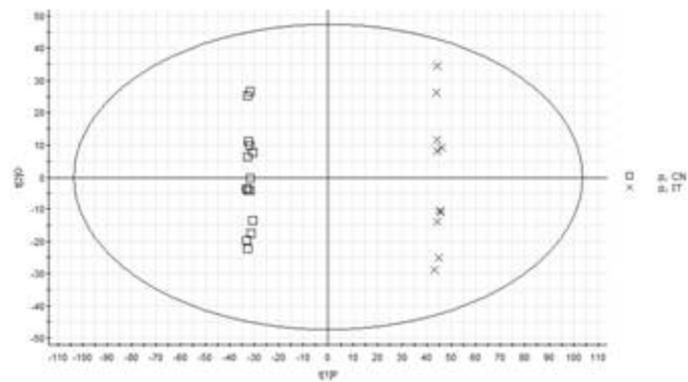


Figure 6

Marker (6.82,375.1075)

6.582e+000

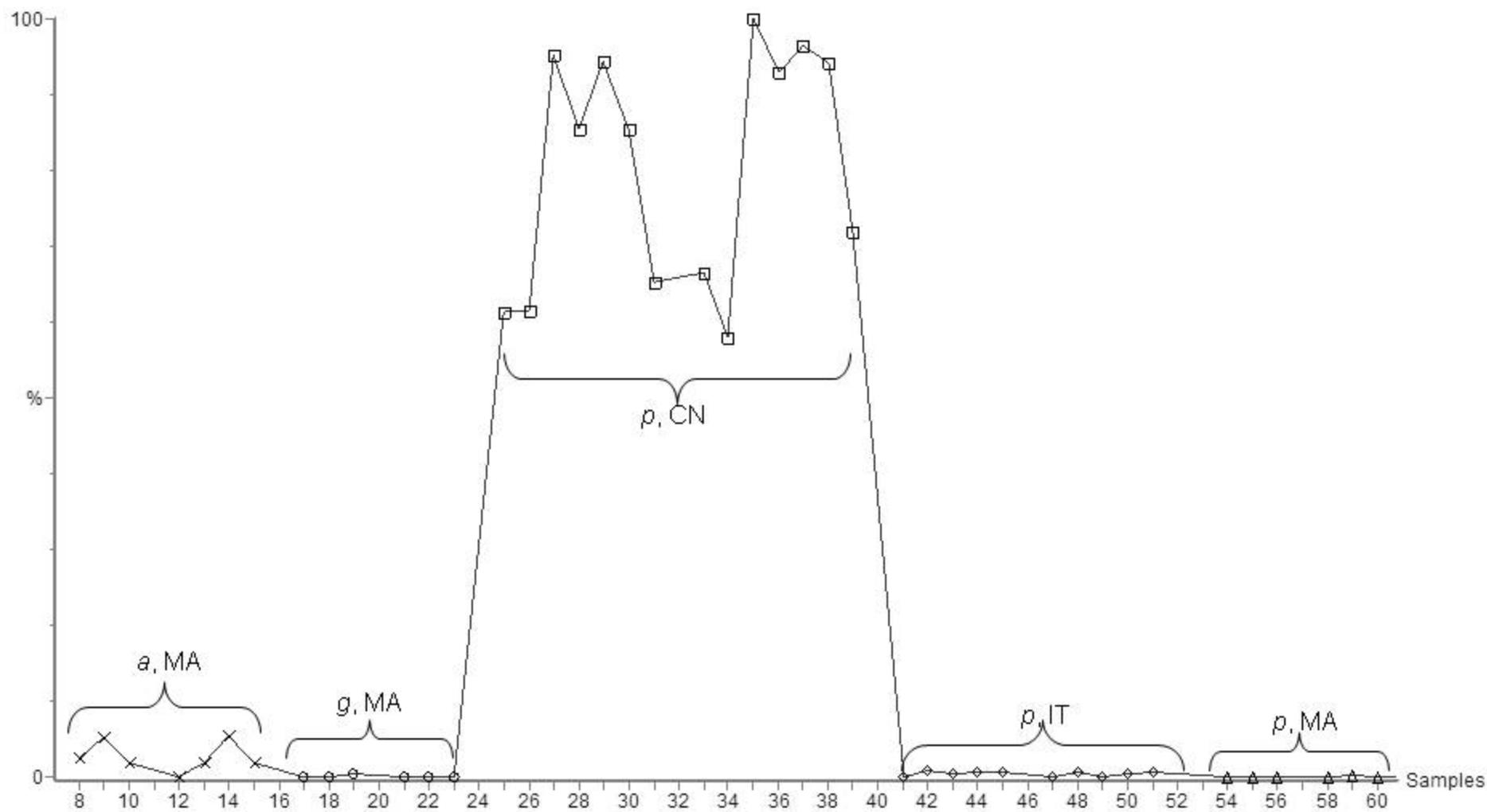


Figure 7

