



HAL
open science

Characterization and classification of Pinus oleoresin samples according to Pinus species, tapping method, and geographical origin based on chemical composition and chemometrics

Morandise Rubini, Armand Clopeau, Jakub Sandak, Stephane Dumarcay, Anna Sandak, Philippe Gerardin, Bertrand Charrier

► To cite this version:

Morandise Rubini, Armand Clopeau, Jakub Sandak, Stephane Dumarcay, Anna Sandak, et al.. Characterization and classification of Pinus oleoresin samples according to Pinus species, tapping method, and geographical origin based on chemical composition and chemometrics. *Biocatalysis and Agricultural Biotechnology*, 2022, 42, pp.102340. 10.1016/j.bcab.2022.102340 . hal-03647572

HAL Id: hal-03647572

<https://univ-pau.hal.science/hal-03647572v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **1 Page information**

2 Title: Characterization and Classification of *Pinus* oleoresin Samples According to *Pinus*
3 species, Tapping method, and Geographical origin Based on Chemical composition and
4 Chemometrics

5

6 **Article Type:** Research Paper

7

8 **First Author:** Morandise Rubini

9 **Corresponding Author:** Morandise Rubini, morandise.rubini@univ-pau.fr

10 **Order of Authors:** Morandise Rubini¹, Armand Clapeau², Jakub Sandak³, Stephane
11 Dumarcay⁴, Anna Sandak³, Philippe Gerardin⁴, Bertrand Charrier¹

12

13 ¹: University of Pau and the Adour Region, E2S UPPA, CNRS, Institute of Analytical and
14 Physicochemical Sciences for Environment and Materials (IPREM-UMR 5254), 371 Rue du
15 Ruisseau, 40004, Mont de Marsan, France

16 ²: Institut Européen de la Forêt Cultivée, 69 route d'Arcachon, 33612 Cestas, France

17 ³: InnoRenew CoE, Livade 6, 6310 Izola, Slovenia

18 ⁴: Laboratoire d'études et de recherche sur le matériau bois, EA 4370, Université de Lorraine,
19 Faculté des Sciences et Technologies, BP 70239, 54506, Vandoeuvre-lès-Nancy, France

20

21 **2 Highlights**

- 22 • Oleoresin samples collected from three *Pinus* species, three oleoresin tapping
23 methods, and four geographical locations were analyzed using gas chromatography
24 (GC) methods
- 25 • Statistical analyzes were used to create classification models according to the specific
26 sample traits (*Pinus* species, tapping method and geographical origin)
- 27 • Furthermore, characteristic biomarkers of the sample traits were highlighted

28

29 3 Abstract

30 Oleoresin samples collected directly from living-trees by three different tapping methods, and
31 from four geographical origins were analyzed using two gas chromatography (GC) methods.
32 The GC was coupled to a flame ionization detector (GC-FID) for quantification, and to a mass
33 spectrometer (GC-MS) for identifying the chemical composition. Twenty-eight chemical
34 components were detected and quantified. The proportions of each chemical component
35 varied exceedingly between different samples, and other associated factors. The specific
36 sample traits, including *Pinus* species, tapping method and geographical origin differentiated
37 the sample batches. Notwithstanding, the main chemical components present in all the
38 characterized samples are α -pinene and β -pinene. Statistical analysis demonstrated that the
39 majority of molecules are of significant importance to sample traits. Moreover, the statistical
40 analysis allows for the identification of the biomarkers associated with the sample traits.
41 Additionally, Linear Discriminant Analysis models have shown very good performance in
42 classifying samples based on the sample traits. Furthermore, the biomarkers allowing the
43 establishment of differences between geographical origins are sativene, camphene, limonene,
44 isopimaric acid and pimarinal, whereas the differentiation between tapping methods is
45 established by sativene, pimaric acid, β -phellandrene, isopimaric acid, retinol and camphene,
46 and lastly biomarkers allowing the differentiation between *Pinus* species are palustric acid,
47 limonene, β -pinene and sativene.

48

49 *Keywords: Chemical composition, Pinus elliotii, Pinus pinaster, Pinus tropicalis, Statistical*
50 *analysis*

51 **4 Introduction**

52 Oleoresin is a complex mixture of turpentine (volatile fraction), and rosin (non-volatile
53 fraction) (Neis *et al.*, 2019). Turpentine is mainly composed of monoterpenes, and
54 sesquiterpenes (Silvestre and Gandini, 2008a), while rosin is mainly composed of diterpenes
55 (Silvestre and Gandini, 2008b). These components are considered as secondary metabolites,
56 since they are biosynthesized as a part of the defense mechanism of the trees against external
57 factors such as insect attacks or pathogens (Bohlmann and Keeling, 2008; Celedon and
58 Bohlmann, 2019), as well as mechanical wounds, or hormonal and chemical stimulants
59 (Rodríguez-García *et al.*, 2016).

60 In France, oleoresin tapping was an important activity from the beginning of the industrial era
61 until its total disappearance in the 1990s due to the rising cost of labor and the lack of
62 competitiveness against countries such as China (Rubini *et al.*, 2021). Subsequently, tapping
63 moved to the Iberian Peninsula and to Greece due to higher profitability (Génova, Caminero
64 and Dochao, 2014; Soliño *et al.*, 2018). Nowadays, China, Brazil and Indonesia are
65 considered as principal producers of the oleoresin consumed worldwide (Soliño *et al.*, 2018).
66 Oleoresin tapping in France has recently experienced a resurgence thanks to the emergence of
67 niche markets, where high cost of labor justifies the intensive research in mechanized
68 harvesting with the purpose of assessing economic viability. The BioGemme and Borehole
69 methods were developed with the objective of maximizing the productivity of oleoresin
70 tapping. In addition, these tapping methods allow for the collection of oleoresins in a closed
71 cup. Such an approach minimizes the content of impurities and maximizes the turpentine
72 content within the collected oleoresin (Rubini *et al.*, 2021). These properties are highly
73 desired, as turpentine-rich resins are commonly accepted as a higher quality resource. Indeed,
74 the specific chemical composition of the oleoresin defines their market value and determines

75 down-stream conversion streams (Rodrigues-Corrêa, de Lima and Fett-Neto, 2012; Celedon
76 and Bohlmann, 2019).

77 The industrial scale tapping in Spain and Portugal has never experienced any interruptions.
78 Even so, it remains fragile due to increasing international competition. Diverse mechanized
79 oleoresin harvesting techniques have been recently studied but are limited to large-scale
80 applications. These technologies are suitable, therefore, for production chains adapted to high-
81 throughput packaging solutions of oleoresin in barrels. For this reason, most of the pine resins
82 are tapped with non-mechanized methods (Cunningham, 2012).

83 The properties of oleoresins are determined by genetic factors which vary for each pine
84 species. *Pinus massoniana* is predominantly tapped in China, *Pinus elliottii* in Brazil and
85 United-States, *Pinus mercurii* in Indonesia, *Pinus roxburghii* in India, *Pinus halepensis* in
86 Greece, *Pinus pinaster* in France, Spain, and Portugal (Lekha and Sharma, 2005).
87 Environmental factors, including seasonality, geographical origin, soil nutrition, drought,
88 flooding, extreme temperatures, or fire are other key aspects (Ayres and Lombardero, 2000;
89 Turtola *et al.*, 2003; Gaspar *et al.*, 2008; Rodrigues-Corrêa and Fett-Neto, 2013; Rodríguez-
90 García *et al.*, 2018; Kopaczyk, Warguła and Jelonek, 2020; Rissanen *et al.*, 2021).

91 Tapping operations consist of repeated tree wounding processes. Thus, oleoresin productivity
92 can vary depending on the tapping technique adopted (Cunningham, 2012; Clopeau, Soares
93 and Orazio, 2021). Diverse tapping methods have been developed through the centuries. Each
94 region/country uses one or more of these methods depending on tradition, topology of local
95 forest as well as the overall economic context (Cunningham, 2012). In this study oleoresin
96 samples were tapped using three different methods: Pica de corteza, BioGemme, and
97 Borehole.

98 The “Pica de corteza” method is similar to the “American” method (Cunningham, 2012;
99 Pérez, Cueva-Gálvez and Mora, 2016). The process starts by manually making a narrow
100 horizontal incision to remove the bark until the cambium is reached, leaving the oleoresin
101 ducts uncovered. The minimum diameter of the tree that allows for initiating tapping is
102 determined in order to maximize oleoresin production at the scale of silvicultural rotation, i.e.
103 from plantation to regeneration.

104 In the BioGemme method two large (~ 25 cm) and two shallow holes, at ~ 30 cm from the
105 ground level, are cut using a circular bur. The minimum diameter of the tree recommended
106 for BioGemme tapping depends on the bark thickness. The bark is removed until it is possible
107 to affix the neck of a collection pocket, which allows for the collection of the oleoresin that
108 flows from the tree (Rubini *et al.*, 2021).

109 The Borehole method involves drilling 2 or 3 deep (~ 15 cm) holes at the base of the tree
110 (about 10 cm high) (Lekha and Sharma, 2005). The advantage of the Borehole method is that
111 the hole is drilled close to the ground, avoiding damage to the merchantable part of the tree.
112 An activator is applied in the holes directly after the wounding and before placing a collection
113 system. (Leneveu, 2012; Rodrigues-Corrêa and Fett-Neto, 2013; Füller *et al.*, 2016) The
114 activator contains an organic or mineral acid, and a growth regulator that reduces the healing
115 capacity of the tree and increases the oleoresin yield.

116 In the “Pica de corteza” method, the oleoresin is collected in open-air. Conversely, the air
117 exposure is highly limited in the mechanized methods, including BioGemme and Borehole. In
118 Europe trees are tapped for 4 to 8 months of the year, particularly during the warmest period.
119 It includes summer as well as late spring or early autumn, depending on air temperatures and
120 precipitation. Trees can be tapped throughout the year in tropical countries such as Brazil.
121 Once collected, the crude oleoresin conversion into turpentine and rosin is carried out by

122 steam (Norton, Hamm and Langenheim, 2004; Rezzi *et al.*, 2005) or vacuum (Ariono *et al.*,
123 2020) distillation. Diverse by-products of the distillation are processed afterward for the
124 fabrication of multiple industrial products (Silvestre and Gandini, 2008a, 2008b).
125 Monoterpenes and sesquiterpenes found in the oleoresin are frequently used as food additives,
126 solvents, in fragrances, plasticizers, repellents, insecticides, anti-viral and anti-microbial
127 agents, or in biofuel (Adams, Demyttenaere and De Kimpe, 2003; Macchioni *et al.*, 2003;
128 Glišić *et al.*, 2007; Behr and Johnen, 2009; Mercier, Prost and Prost, 2009; Silva *et al.*, 2012).
129 Diterpenes are commonly utilized as polymer precursors, polymerization emulsifiers for
130 pharmaceutical and cosmetic applications, adhesives, inks, waterproofing materials, or
131 surfactants (Silvestre and Gandini, 2008b).

132 The aim of this study is to identify the relevant biomarkers associated with the *Pinus* species
133 and understand the dependence on the tapping method and geographical origin. Furthermore,
134 this study focuses on the development of the classification model based on the chemical
135 components associated with the sample traits, including the *Pinus* species, the tapping
136 method, and the geographical origin.

137 **5 Material and methods**

138 *5.1 Samples of oleoresins*

139 A total of 70 oleoresin samples were studied. Detailed information on the *Pinus* species, the
140 tapping method used for sample collection, and the geographical origin are summarized in
141 Tab. 1. Samples were collected from three *Pinus* species, including 64 trees of *Pinus pinaster*,
142 3 *Pinus elliottii*, and 3 *Pinus tropicalis*. Three methods were implemented for oleoresin
143 tapping, delivering 14 batches of BioGemme, 44 of Borehole and 12 of Pica de corteza. All
144 experimental samples were collected in four geographical locations, resulting in 58 samples

145 from France, 3 from Spain, 3 from Portugal, and 6 from Brazil. Each oleoresin sample was
146 uniquely marked with three letter codes, determining tree species, tapping method and
147 provenance.

148 Each sample from PBF (*P. pinaster*, BioGemme, France) and POF (*P. pinaster*, Borehole,
149 France) was harvested from a unique and single tree. Conversely, each sample of PHS (*P.*
150 *pinaster*, Pica de corteza, Spain), PHP (*P. pinaster*, Pica de corteza, Portugal), EHB (*P.*
151 *elliottii*, Pica de corteza, Brazil) and THB (*P. tropicalis*, Pica de corteza, Brazil) was a
152 mixture of oleoresin collected from several trees tapped at the same location. In the latter
153 case, resampling was performed to increase the number of samples. The three most tapped
154 pine species in Europe and Latin America were chosen for the investigation. These represent
155 the majority of the oleoresin supply for Europe. The broad portfolio of characterized samples
156 allows for the determination of the effect of secondary metabolites on the oleoresin
157 composition that can be altered by the tapping method and/or the geographical origin of the
158 tapped tree. The goal of this research is, therefore, to systematically assess these biomarkers.

159 5.2 Chemical composition of the oleoresins

160 The qualitative analysis of the oleoresin samples was performed on a Gas Chromatography–
161 Mass Spectrometer (GC–MS) Clarus 680 GC (Perkin Elmer Inc., USA) coupled to the SQ8
162 MS quadrupole mass spectrometer (Perkin Elmer Inc., USA). The injection was performed at
163 250°C in the split mode at a split flow of 20 mL/min. 1.0 µL of the sample was injected into
164 the gas chromatograph. Separation was performed on a (5% phenyl)-95% methylpolysiloxane
165 fused-silica capillary column (DB-5ms, 30 m length × 0.25 mm i.d. × 0.25 µm film thickness,
166 Agilent J&W, USA) with helium as carrier gas at a constant flow of 1 mL/min. The oven
167 temperature program was optimized for the oleoresin samples. The temperature of the oven
168 was stabilized at 50°C for 2 minutes. The sequence started with ramping the temperature to

169 120°C at 10°C/min, then ramping it to 220°C at 10°C/min. The ramping continued at
170 15°C/min till it reached 320°C, to finally hold constant at 320°C for 4 minutes. The total
171 running time was approximately 30 minutes. Helium was used as a carrier gas, with a
172 constant flow of 1.2 mL/min. An electron ionization system in the electron impact mode (70
173 eV ionization energy) was used for the GC–MS detection. The transfer line and source
174 temperatures were set at 250°C and 180°C, respectively. Detection was carried out in the scan
175 mode from m/z 45 to m/z 620 with a scan time of 0.2s and interscan delay of 0.1s. The
176 detector was switched off in the initial 2 min (solvent delay). The identification of chemical
177 components based on the mass spectra comparison with the standard reference database and
178 the mass spectral library published by the NIST (Edition of 2011, US National Institute of
179 Standards and Technology, Gaithersburg, MD, USA).

180 The GC-MS method allows for the identification of chemical compounds by analyzing the
181 specific retention times, whereas the quantitative analysis of experimental samples was
182 performed on a Gas Chromatography–Flame Ionization Detector (FID).

183 The potential biomarkers were quantified using a Clarus 500 GC gas chromatograph equipped
184 with the FID detector (Perkin Elmer Inc., USA). The chromatographic conditions were
185 identical to the GC-MS analysis.

186 All the analyses were performed in triplicate. Each oleoresin sample (~10 mg) was diluted in
187 a mixture of 50:50/Hexane:Ethyl acetate before chromatographic examination.

188 5.3 *Statistical data analysis*

189 The raw chromatograms contained different artifacts, mainly related to time compression
190 (elution time shift, variable alignment). Therefore, all chromatograms were processed before
191 further analysis. First, the GC-FID data in Perkin-Elmer proprietary format (.RAW) were

192 converted to comma separated value (CSV) files by using OpenChrom software (Wenig and
193 Odermatt, 2010). All chromatograms were rearranged before further processing. A data
194 matrix X was created with n rows and p columns, where n corresponds to the number of
195 samples, and p reflects the number of elution time points. Seventy samples in triplicate
196 resulted in $n = 210$. Similarly, the number of elution time points was $p = 22,498$. All
197 chromatograms were subjected in the subsequent step to the baseline correction procedure.
198 Correlation Optimized Warping (COW) alignment algorithm was used to correct the shift of
199 the elution time (Jellema, 2009). X was subjected to baseline correction by applying
200 asymmetric least squares algorithm (Eilers, 2004). A baseline trend was estimated as a
201 second-order polynomial, that was subtracted from the raw chromatogram. The smoothness
202 and asymmetry were optimized and set to 1×10^{-3} and 1×10^6 respectively. The objective
203 function of the COW algorithm was to maximize the correlation coefficient between a
204 reference chromatogram and each single sample chromatogram. It was implemented as
205 chromatograms alignment using piecewise linear stretching and/or compression in
206 combination with interpolation (Azimi and Fatemi, 2018). Three inputs were needed to use
207 COW algorithm: the reference chromatogram, the segment length, and the slack size (also
208 called warping). The reference chromatogram was determined as a maximum cumulative
209 product of correlation coefficients for all sample chromatograms. The segment length and
210 slack size were optimized to reach 12 and 8, respectively. Matlab 2019a (Mathworks, Inc.,
211 USA) environment with PLS toolbox 8.9.1 (Eigenvector Research, Inc., USA) were used for
212 baseline correction. The software code for COW algorithms were obtained from the web
213 depository (http://www.models.life.ku.dk/DTW_COW). The quantitative analysis was
214 performed after baseline and COW corrections by peak identification and included
215 quantification by internal normalization. A new data matrix Z with n rows and k columns was
216 generated as a result of quantifications, with k indicating the number of quantified chemical

217 components. Even if 80 peaks were present in the reference chromatograph, only 28
218 components were formerly identified with the GC-MS. Consequently, only the properly
219 recognized peaks were selected to follow-up multivariate statistical analysis. The resulting Z
220 had $n = 210$, and $k = 28$. Z was analyzed using different statistical techniques, such as,
221 MANOVA, ANOVA, and t-test. Moreover, linear discriminate analysis (LDA) was tested as
222 a suitable classification method.

223 Multivariate analysis of variance (MANOVA) and a *post hoc* Tukey HSD test were applied to
224 study the single and combined effect of multiple categorical independent variables on the
225 quantitative dependent variables. Indeed, with respect to the independent variable
226 (comparison of between-groups and within-group variability), the statistical significance was
227 tested on each dependent variable. Furthermore, multivariate test of significance, as well as
228 Pillai's Trace (V), Wilks' Lambda (Λ), and *p-value* (p) statistics were performed to determine a
229 possible significant effect of the categorical independent variables on the quantitative
230 dependent variables. Such an approach allows for the determination of the effect of specific
231 biomarkers on each of the independent variables (*Pinus* species, tapping method, and
232 geographical origin).

233 For the needs of this study, biomarkers are defined from a statistical point of view. Indeed, the
234 biomarker is considered as a characteristic objectively measured on the samples, and which
235 makes it possible to prove an intrinsic singularity. Thus, molecules that evidence a
236 statistically significant difference are considered as biomarkers.

237 One-way analysis of variances (ANOVA) was applied to compare the effects of each of the
238 categorical independent variables on the quantitative dependent variables. These included
239 *Pinus* species, tapping method, and geographical origin, respectively. Additionally, a series of
240 t-tests with Bonferroni adjustment were performed on samples when a statistical significance

241 difference was found in the quantitative dependent variable. SPSS Statistics (Version
242 28.0.0.0) software was used for statistical analysis (MANOVA, ANOVA and t-test).

243 5.4 Linear Discriminant Analysis (LDA)

244 Selected quantitative dependent variables were used to perform LDA following the
245 MANOVA. The main advantage of LDA is the possibility to select the most relevant
246 quantitative dependent variables. LDA is a supervised method for classification that aims at
247 maximizing the ratio of the between-groups to within-groups variance by finding linear
248 combinations of the original variables. The maximum number of computable discriminant
249 functions is equal to the number of levels in the categorical independent variables minus 1.

250 Subset selection needs to be performed to prepare two independent subsets before building a
251 LDA model. One set is used to calibrate the model (calibration subset), while another is
252 implemented to validate the performance of the previously calibrated model (validation
253 subset). Performance of the calibrated model was assessed by cross- and independent external
254 validations computed on calibration and validation subsets, respectively. Calibration and
255 validation subsets within the same classes of samples were manually selected to provide
256 similar samples variability. For that reason, two-thirds of the samples were placed into the
257 calibration subset and the remaining one-third into the validation subset.

258 A confusion table summarizing relevant classification rates was defined after building the
259 LDA model. The capability of a classification model to correctly recognize samples
260 belonging to the modeled class, is defined as *Sensitivity* or true positive rate (*TPR*). *TPR*
261 values are in the range between 0 and 100, while $TPR = 100$ when number of false negatives
262 (*FN*) is equal to 0. Inversely, the capability of the classification model to correctly reject
263 samples of all other modeled classes is called true negative rate (*TNR*), or *Specificity*. *TNR*

264 values are also in the range between 0 and 100. $TNR = 100$ when the number of false
265 positives (FP) is equal to 0. The positive predictive value (PPV) corresponds to the
266 *Precision*. It represents the proportion of correctly assigned samples compared to the total
267 number of samples in the class. Formulae of these classification parameters are given in Eq.
268 (1) – (3):

$$269 \quad \text{Sensitivity} = TPR = \frac{TP}{TP + FN} \quad (\text{Eq. 1})$$

$$270 \quad \text{Specificity} = TNR = \frac{TN}{TN + FP} \quad (\text{Eq. 2})$$

$$271 \quad \text{Precision} = PPV = \frac{TP}{(TP + FP)} \quad (\text{Eq. 3})$$

272 **6 Results and discussion**

273 *6.1 Descriptive analysis of the data*

274 Eighty peaks were detected on the chromatogram after pre-processing. Of these 80 peaks,
275 only 28 peaks were properly identified with GC-MS and afterward quantified with GC-FID.
276 The results obtained are summarized in Tab. 2. It is evident that oleoresin is a mixture of
277 diverse chemical components belonging to three subclasses of compounds, including
278 monoterpenes (10 atoms of carbon), sesquiterpenes (15 atoms of carbon), and diterpenes (20
279 atoms of carbon). It is in-line with the observations reported in literature (Arrabal *et al.*,
280 2005).

281 The highest monoterpene concentration was noticed for α -pinene (35.7 – 53.2%) and β -pinene
282 (8.0 – 32.5%). The relative ratio between these two compounds varied depending on the
283 samples. Extreme ratios were noticed for THB and EHB, reaching 6:1, and 1:1 respectively.
284 However, it was also observed that, without considering environmental factors, PBF seems to

285 have higher concentrations of α -pinene when compared to POF, where the ratio was 3:1 and
286 2:1, respectively. the α -pinene/ β -pinene ratio varied also due to tree provenance. *P. pinaster*
287 samples tapped in Spain (PHS) had ratios of 5:1 while it was 2:1 for samples from Portugal
288 (PHP). It should be mentioned the latter had the highest relative content of α -pinene. The
289 overall concentrations of sesquiterpenes and diterpenes was relatively weaker, when
290 compared with monoterpenes. Sativene (0.09 – 6.35%) and α -copaene (0.11 – 3.69%) appear
291 to be present at the highest concentrations within the sesquiterpenes fraction. Isopimaric acid
292 was identified as a major compound of diterpenes, detected in the concentration range of
293 0.88% to 2.36%. The chemical composition of resins as determined in this research are
294 similar to those reported in literature for *P. elliottii* from southern Brazil (Rodrigues-Corrêa *et*
295 *al.*, 2011), *P. pinaster* from Portugal and Morocco (Pio and Valente, 1998; Ghanmi *et al.*,
296 2005, 2009), and *P. tropicalis* from Cuba (Valterová *et al.*, 1995). However, it was also
297 reported by several authors that the profile of terpenes composition can substantially vary
298 within different populations and individuals. It is associated with several factors, including
299 genetic, environmental or tapping methodology, among others (Arrabal *et al.*, 2002, 2005;
300 Kopaczyk, Warguła and Jelonek, 2020).

301 6.2 Multivariate test statistics

302 Three-way MANOVA (multivariate analysis of variances) was implemented on the dataset
303 considering three categorical independent variables (*Pinus* species, tapping method, and
304 geographical origin). Twenty-eight dependent variables representing concentrations of
305 identified oleoresins were applied for MANOVA together with 70 triplicated oleoresin
306 samples. The objective of this analysis was to identify the relevant biomarkers that are
307 associated with the specific *Pinus* species, tapping methodology applied and/or unique
308 geographical origin.

309 MANOVA results revealed a statistically significant contribution of each chemical
310 component on the independent variables classification. The *Pinus* species as a model variable
311 contributed the most ($V = 0.941$, $\Lambda = 0.059$, $F = 100.478$, $p = 0.000$), followed by the tapping
312 method ($V = 0.823$, $\Lambda = 0.177$, $F = 29.358$, $p = 0.000$), and the geographical origin ($V =$
313 0.646 , $\Lambda = 0.358$, $F = 11.533$, $p = 0.000$).

314 6.3 Global LDA model for overall classification of oleoresin samples

315 Fig. 1 presents projections of sample points on the plane defined by the first three
316 discriminant functions. The shape and color of marks corresponds to the training (circle) and
317 validation (square) samples. The whole set of molecules was used to discriminate sample
318 characteristics based on species, tapping, and geographical origin, all distinguished by varying
319 colors.

320 A clear separation between samples of *P. pinaster* (PBF, POF, PHP, PHS) at the left end, *P.*
321 *elliottii* (EHB) at the right end, and *P. tropicalis* (THB) in the center can be noticed
322 considering the DF1-DF2 plane. Samples of *P. pinaster* (PBF, POF, PHP, PHS) are grouped
323 together regardless of the tapping methodology or geographical origin. Indeed, they share
324 common characteristics in terms of chemical composition, as can be seen in Tab. 1.
325 Furthermore, compositions of constitutive molecules within *P. elliottii* (EHB), *P. tropicalis*
326 (THB), and *P. pinaster* (PBF, POF, PHP, PHS) differ noticeably from each other. The
327 analysis of loading plots presented in Fig. 2 reveals that (2E.4E)-3.7-dimethyl-2.4-octadiene,
328 and limonene are two main molecules contributing to the DF1 component. The concentrations
329 of (2E.4E)-3.7-dimethyl-2.4-octadiene are relatively lower in *P. elliottii* when compared to *P.*
330 *pinaster*. Conversely the limonene concentrations are typically higher in *P. elliottii* than in *P.*
331 *pinaster*. Clustering of sample points was also observed considering the DF2 component
332 (DF1-DF2). It was associated with the apparent differences between diverse species (*P.*

333 *pinaster*, *P. elliottii*, and *P. tropicalis*). The DF2 scores are negative for *P. elliottii* samples,
334 while these scores are centered in *P. pinaster* and positive for *P. tropicalis*. DF2 loadings
335 indicate that retinol, manoyl oxide, and β -pinene are the main molecules that influence the
336 separation. From the negative to positive direction of the DF2 axis, the concentration of
337 different components seems to also correlate with the DF2 loading values. Indeed, the
338 concentrations are higher for retinol, followed by manoyl oxide, and lower for β -pinene.

339 In summary, the LDA model presented indicates that *Pinus* species samples share several
340 common intrinsic characteristics that are independent of tapping methodologies and
341 geographical origin.

342 6.4 Classification of samples according to the *Pinus* species

343 A separate MANOVA combined with a *post hoc* Tukey HSD study was performed to identify
344 biomarkers differentiating samples according to the tree species. Results including the *p-value*
345 of the *post hoc* Tukey HSD tests are presented in Tab. 3. Twenty-seven of the 28 chemical
346 components were identified as significant biomarkers for differentiating *Pinus* oleoresins ($p <$
347 0.05). The only exception was pimarinal ($p = 0.063$). The most predominant biomarkers
348 according to V analysis is, in order of importance, palustric acid $>$ limonene $>$ β -pinene $>$
349 sativene.

350 All twenty-seven significant biomarkers were subjected to the follow-up LDA modelling for
351 classification of samples according to species. Fig. 3 presents the projection of the training
352 and validation samples in the two discriminant functions (DF1 vs. DF2). The results were in
353 line with that obtained in the global LDA model, confirming a great separation capability of
354 samples according to the *Pinus* species.

355 The loading graph for the LDA model is shown in Fig. 4. The high value (neither positive or
356 negative) of loadings highlights all the variables (or molecules) predominantly involved in the
357 success of classification. Indeed, DF1 loading reveals that the variables that contributed the
358 most are, in order of importance, (2E,4E)-3,7-Dimethyl-2,4-octadiene > limonene > palustric
359 acid > β -myrcene. An analogous sequence derived from DF2 loading corresponds to manoyl
360 oxide > β -pinene > retinol > methyl dehydroabietate. The results of the confusion table are
361 presented in Tab. 4. It is evident that the overall correct classification success rate was 100%
362 for all *Pinus* species, considering either the cross-validation or validation sample sets.

363 Only a few preceding studies concerning the classification of resins collected from various
364 *Pinus* species (based on their chemical composition) were reported in literature references.
365 However, the characterization methodologies implemented by other authors, as well as the
366 form of samples analyzed were different from those tested in this research. Myrcene and
367 noracid 1 enable a correlation between the character of a tree as a potentially high producer
368 (plus tree) and the chemical composition of oleoresin (Arrabal *et al.*, 2002). The level of
369 significance was very high ($p = 0.01$) providing a high confidence for classification. The
370 same task was addressed by analyzing the abienol and dehydroabietic acid concentrations,
371 even though the level of significance was slightly lower (5%). Other studies focused on
372 distinguishing *P. Pinaster* subspecies by analyzing the acid fraction of wood extractives
373 (Arrabal and Cortijo, 1994). It was found that there are significant differences in the
374 quantitative and qualitative composition of extracts enabling classification of tree subspecies.
375 In this case, low percentage of levopimaric acid, ratio of abietic and dehydroabietic acids,
376 relative concentration of abietic and dehydroabietic acids, as well as the presence of
377 neoabietic, palustric and levopimaric acids were identified as the most relevant traits.

378 6.5 *Classification of samples according to the tapping method*

379 According to the results of the *p-value* of the *post hoc* Tukey HSD (from MANOVA),
380 presented in Tab. 5, twenty-six of the 28 chemical components were found to be significant (*p*
381 < 0.05) as biomarkers for the studied tapping methods. In that case, pimara-7,15-dien-3-one
382 (*p* = 0.081), and methyl dehydroabietate (*p* = 0.161) were excluded from the initial
383 component set. The most predominant biomarkers according to *V* values were sativene >
384 pimaric acid > β-phellandrene > isopimaric acid.

385 Twenty-six biomarkers identified as significant were subjected to LDA classification. Fig. 5
386 presents projections of the training and validation samples on the first two discriminant
387 functions. There is a superposition of the Borehole and BioGemme tapping methods on DF1.
388 It evidenced similar chemical composition characteristics for those tapping methods. The
389 variable that contributed the most to the first loading was palustric acid, followed by
390 dehydroabietic acid, camphene, and retinol (Fig. 6). It can be seen on the DF2 chart that there
391 is a clear overlap of the Borehole and Pica de corteza tapping methods. Variables that
392 contributed the most to the second loading were longifolene > palustric acid > α-copaene >
393 isopimaric acid, listed in order of statistical significance. Tab. 6 summarizes all results for the
394 confusion matrix of the LDA model. A total of 87 Borehole samples were correctly classified
395 in a cross-validated model as Borehole, with only 1 misclassified as BioGemme. The same
396 trend was observed in the validation subset where 43 Borehole samples were correctly
397 classified as Borehole with one sample wrongly identified as BioGemme. Consequently, the
398 sensitivity of Pica de corteza and BioGemme classification was 100%, while that of Borehole
399 was of 98.9% (Tab. 6).

400 6.6 *Classification of samples according to geographical origin*

401 Twenty-six of 28 chemical components were found to be significant ($p < 0.05$) geographical
402 origin biomarkers according to the *post hoc* Tukey HSD (from MANOVA) (Tab. 7). Methyl
403 dehydroabietate ($p = 0.751$), and palustric acid ($p = 0.059$) were, therefore, excluded from the
404 variable set used for the LDA modelling. The most predominant biomarkers were sativene >
405 camphene > limonene > isopimaric acid, sorted according to V values.

406 Results of LDA classification are summarized in Fig. 7. There is an evident separation of
407 samples collected in Brazil (EHB, THB) compared to other geographical origins as is
408 observed on the DF1 component (DF1 vs. DF2). Separation of Spanish samples (PHS) was
409 recorded in DF2 (DF1 vs. DF2). All the samples that originated in Portugal (PHP) were
410 differentiated from other locations by means of DF3 (DF1 vs. DF3). The first loading chart
411 (Fig. 8) reveals that biomarkers that contributed the most to the success of classification were
412 limonene > (2E,4E)-3,7-dimethyl-2,4-octadiene > β -myrcene > α -pinene, while the second
413 loading shows high contribution of camphene > longifolene > retinol > terpinolene. Four
414 biomarkers most relevant in the third loading of the LDA model were retinol > isopimaric
415 acid > α -pinene > terpinolene, listed here in the order of importance. The performant
416 quantifiers of the geographical origin classification model are summarized in Tab. 8. In
417 comparison to other LDA models, the accuracy of prediction was very high as the sensitivity
418 was 100% for Brazil, Portugal, and Spain, while it reached 99.1% for samples originating
419 from France.

420 Differentiation of Spanish *P. pinaster* oleoresins from five provenances has been recently
421 studied (Arrabal *et al.*, 2005). Isopimaric acid and α -pinene were identified as two compounds
422 allowing for superior differentiation between provenances when the model was based on

423 classical statistical univariate analysis. However, pimaric acid and noracid 1 contributed the
424 most towards discrimination by means of canonical multivariate analysis model.

425 **7 Conclusion**

426 Different statistical techniques tested in this study confirm, with high statistical significance,
427 that the chemical composition of oleoresin differs due to *Pinus* oleoresin, the tapping method
428 applied, and the geographical origin of the analyzed samples.

429 It was evidenced that *P. tropicalis* and *P. pinaster* tend to produce relatively high amounts of
430 two monoterpenes, namely α -pinene and β -pinene. Furthermore, the same compounds highly
431 affect the preservation of the oleoresin quality, especially when collected by the BioGemme
432 tapping method. Other degradation markers of the rosin fraction, such as pimaric acid, or
433 methyl dehydroabietate also tend to be minimized when oleoresin collection is carried out
434 with a closed-cup method. It confirms the importance of tapping technology to assure superior
435 quality of the collected oleoresin.

436 Results of statistical analysis revealed that palustric acid, limonene, β -pinene and sativene
437 grapes are most pronounced in distinguishing the three *Pinus* species studied. Sativene,
438 pimaric acid, β -phellandrene, isopimaric acid, retinol and camphene are highly useful for the
439 classification of tapping methods tested in this research. Finally, sativene, camphene,
440 limonene, isopimaric acid and pimarinal are chemical components most related to the
441 geographical origin of the investigated oleoresin samples. Three LDA models were developed
442 with the purpose of classifying samples according to the *Pinus* species, tapping methodology
443 as well as sample provenance. The overall performance of these models was optimal,
444 reaching near 100% success according to the sensitivity results. It encourages further
445 development of this research toward in-field testing and enlarging the sample portfolio by
446 adding tree species and forest locations.

447 **8 CRediT author statement**

448 **Morandise Rubini:** Conceptualization, Methodology, Formal analysis, Investigation, Writing
449 - Original Draft, Writing - Review & Editing. **Armand Clapeau:** Resources, Writing -
450 Review & Editing. **Jakub Sandak:** Validation, Writing - Review & Editing. **Stephane**
451 **Dumarcay:** Investigation, Writing - Review & Editing. **Anna Sandak:** Writing - Review &
452 Editing. **Philippe Gerardin:** Investigation. **Bertrand Charrier:** Funding acquisition.

453 **9 Acknowledgements**

454 Morandise Rubini thanks E2S UPPA for the financial support (International doctoral
455 mobility) allowing him to spend three months in Slovenia at *InnoRenew CoE* research
456 laboratory.

457 Morandise Rubini and Charrier Bertrand gratefully acknowledge the financial support from
458 the Nouvelle Aquitaine regional council, the Landes departmental council, the Agence
459 Nationale de la Recherche (National Agency for Research) and Xyloforest (ANR-10-EQPX-
460 16).

461 Jakub and Anna Sandak acknowledge the European Commission for funding the InnoRenew
462 project (grant agreement #739574 under the Horizon2020 Widespread-2-Teaming program)
463 and the Republic of Slovenia (investment funding from the Republic of Slovenia and the
464 European Regional Development Fund).

465 **10 Declaration of competing interest**

466 The authors declare no actual or potential conflicts of interests.

467 **11 References**

- 468 Adams, A., Demyttenaere, J. C. R. and De Kimpe, N. (2003) 'Biotransformation of (R)-(+)-
469 and (S)-(-)-limonene to α -terpineol by *Penicillium digitatum*— investigation of the culture
470 conditions', *Food Chemistry*. Elsevier, 80(4), pp. 525–534. doi: 10.1016/S0308-
471 8146(02)00322-9.
- 472 Ariono, D. *et al.* (2020) 'Fractionation of Turpentine', *IOP Conference Series: Materials
473 Science and Engineering*. IOP Publishing, 742, p. 12029. doi: 10.1088/1757-
474 899x/742/1/012029.
- 475 Arrabal, C. *et al.* (2002) 'Pinus pinaster Oleoresin in Plus Trees'. *De Gruyter*, 56(3), pp. 261–
476 266. doi: 10.1515/HF.2002.043.
- 477 Arrabal, C. *et al.* (2005) 'Differentiation among five Spanish Pinus pinaster provenances
478 based on its oleoresin terpenic composition', *Biochemical Systematics and Ecology*.
479 Pergamon, 33(10), pp. 1007–1016. doi: 10.1016/J.BSE.2005.03.003.
- 480 Arrabal, C. and Cortijo, M. (1994) 'Fatty and Resin Acids of Spanish Pinus pinaster Ait.
481 Subspecies'.
- 482 Ayres, M. P. and Lombardero, M. J. (2000) 'Assessing the consequences of global change for
483 forest disturbance from herbivores and pathogens', *Science of the Total Environment*, 262(3),
484 pp. 263–286. doi: 10.1016/S0048-9697(00)00528-3.
- 485 Azimi, F. and Fatemi, M. H. (2018) 'Multivariate curve resolution-correlation optimized
486 warping applied to the complex GC-MS signals; toward comparative study of peel chemical
487 variability of Citrus aurantium L. varieties', *Microchemical Journal*. Elsevier, 143, pp. 99–
488 109. doi: 10.1016/j.microc.2018.07.041.
- 489 Behr, A. and Johnen, L. (2009) 'Myrcene as a natural base chemical in sustainable chemistry:
490 A critical review', *ChemSusChem*. Wiley-VCH Verlag, 2(12), pp. 1072–1095. doi:
491 10.1002/CSSC.200900186.

492 Bohlmann, J. and Keeling, C. I. (2008) 'Terpenoid biomaterials', *Plant Journal*, 54(4), pp.
493 656–669. doi: 10.1111/J.1365-313X.2008.03449.X.

494 Celedon, J. M. and Bohlmann, J. (2019) 'Oleoresin defenses in conifers: chemical diversity,
495 terpene synthases and limitations of oleoresin defense under climate change', *New*
496 *Phytologist*. Blackwell Publishing Ltd, 224(4), pp. 1444–1463. doi: 10.1111/nph.15984.

497 Clopeau, A., Soares, R. and Orazio, C. (2021) *Marché De La Gemme De Pin En Contexte De*
498 *Crise Sanitaire*. Available at: [http://www.iefc.net/fr/marche-de-la-gemme-de-pin-en-contexte-](http://www.iefc.net/fr/marche-de-la-gemme-de-pin-en-contexte-de-crise-sanitaire/)
499 [de-crise-sanitaire/](http://www.iefc.net/fr/marche-de-la-gemme-de-pin-en-contexte-de-crise-sanitaire/) (Accessed: 30 August 2021).

500 Cunningham, A. (2012) 'Pine resin tapping techniques used around the world', *Pine resin:*
501 *biology, chemistry and applications*, 661(2), pp. 1–8.

502 Eilers, P. H. C. (2004) 'Parametric Time Warping', *Analytical Chemistry*. American
503 Chemical Society, 76(2), pp. 404–411. doi: 10.1021/ac034800e.

504 Füller, T. N. *et al.* (2016) 'Stimulant paste preparation and bark streak tapping technique for
505 pine oleoresin extraction', in *Methods in Molecular Biology*. Methods Mol Biol, pp. 19–26.
506 doi: 10.1007/978-1-4939-3393-8_2.

507 Gaspar, M. J. *et al.* (2008) 'Age trends in genetic parameters of wood density components in
508 46 half-sibling families of *Pinus pinaster*', *Canadian Journal of Forest Research*, 38(6), pp.
509 1470–1477. doi: 10.1139/X08-013.

510 Génova, M., Caminero, L. and Dochao, J. (2014) 'Resin tapping in *Pinus pinaster*: Effects on
511 growth and response function to climate', *European Journal of Forest Research*, 133(2), pp.
512 323–333. doi: 10.1007/s10342-013-0764-4.

513 Ghanmi, M. *et al.* (2005) 'Étude du rendement et de la composition de l'essence de
514 térébenthine du Maroc: Cas du Pin maritime (*Pinus pinaster*) et du Pin d'Alep (*Pinus*
515 *halepensis*)', *Acta Botanica Gallica*. Taylor & Francis Group , 152(1), pp. 3–10. doi:
516 10.1080/12538078.2005.10515450.

517 Ghanmi, M. *et al.* (2009) 'Évaluation de la qualité de la colophane du pin maritime (Pinus
518 pinaster) et du pin d'Alep (Pinus halepensis) du Maroc Évaluation de la qualité de la
519 colophane du pin maritime (Pinus pinaster) et du pin d'Alep (Pinus halepensis) du Maroc',
520 *Acta Bot. Gallica*, 156(3), pp. 427–435. doi: 10.1080/12538078.2009.10516168.

521 Glišić, S. B. *et al.* (2007) 'Antimicrobial activity of the essential oil and different fractions of
522 Juniperus communis L. and a comparison with some commercial antibiotics', *Journal of the*
523 *Serbian Chemical Society*. Serbian Chemical Society, 72(4), pp. 311–320. doi:
524 10.2298/JSC0704311G.

525 Jellema, R. H. (2009) 'Variable Shift and Alignment', *Comprehensive Chemometrics*.
526 Elsevier, 2, pp. 85–108. doi: 10.1016/B978-044452701-1.00104-6.

527 Kopaczyk, J. M., Warguła, J. and Jelonek, T. (2020) 'The variability of terpenes in conifers
528 under developmental and environmental stimuli', *Environmental and Experimental Botany*.
529 Elsevier, 180, p. 104197. doi: 10.1016/J.ENVEXPBOT.2020.104197.

530 Lekha, C. and Sharma, K. R. (2005) 'Borehole method of oleoresin tapping chir pine (pinus
531 roxburghii sargent)', *Forest Chemicals Review*, 115(3), pp. 11–17.

532 Leneveu, L. (2012) 'Procédé pour favoriser l'exsudation de l'oleorésine et composition pour
533 mettre en œuvre ce procédé'.

534 Macchioni, F. *et al.* (2003) 'Chemical composition of essential oils from needles, branches
535 and cones of Pinus pinea, P. halepensis, P. pinaster and P. nigra from central Italy', *Flavour*
536 *and Fragrance Journal*. John Wiley & Sons, Ltd, 18(2), pp. 139–143. doi: 10.1002/FFJ.1178.

537 Mercier, B., Prost, J. and Prost, M. (2009) 'The essential oil of turpentine and its major
538 volatile fraction (α - and β -pinenes): A review', *International Journal of Occupational*
539 *Medicine and Environmental Health*. Nofer Institute of Occupational Medicine, 22(4), pp.
540 331–342. doi: 10.2478/V10001-009-0032-5.

541 Neis, F. A. *et al.* (2019) 'Resin exudation profile, chemical composition, and secretory canal

542 characterization in contrasting yield phenotypes of *Pinus elliottii* Engelm', *Industrial Crops*
543 *and Products*. Elsevier B.V., 132, pp. 76–83. doi: 10.1016/j.indcrop.2019.02.013.

544 Norton, R. E., Hamm, J. and Langenheim, J. H. (2004) 'Plant Resins: Chemistry, Evolution,
545 Ecology, and Ethnobotany', *Journal of the American Institute for Conservation*, 43(3), p. 285.
546 doi: 10.2307/4129644.

547 Pérez, C. M. O., Cueva-Gálvez, G. E. and Mora, H. E. G. (2016) 'Caracterización de la
548 oleoresina de *Pinus caribaeae* obtenido por sistema de pica de corteza con ácido sulfúrico',
549 *Revista Forestal del Perú*. Universidad Nacional Agraria la Molina, 31(2), pp. 58–68. doi:
550 10.21704/RFP.V31I2.1027.

551 Pio, C. A. and Valente, A. A. (1998) 'Atmospheric fluxes and concentrations of
552 monoterpenes in resin-tapped pine forests', *Atmospheric Environment*. Pergamon, 32(4), pp.
553 683–691. doi: 10.1016/S1352-2310(97)00327-0.

554 Rezzi, S. *et al.* (2005) 'Composition and chemical variability of the oleoresin of *Pinus nigra*
555 ssp. *laricio* from Corsica', *Industrial Crops and Products*. Elsevier, 21(1), pp. 71–79. doi:
556 10.1016/J.INDCROP.2003.12.008.

557 Rissanen, K. *et al.* (2021) 'Drought effects on carbon allocation to resin defences and on resin
558 dynamics in old-grown Scots pine', *Environmental and Experimental Botany*. Institute for
559 Atmospheric and Earth System Research / Forest Sciences, Faculty of Agriculture and
560 Forestry, University of Helsinki, Latokartanonkaari 7, Helsinki, Finland: Elsevier B.V., 185.
561 doi: 10.1016/j.envexpbot.2021.104410.

562 Rodrigues-Corrêa, K. C. da S. *et al.* (2011) 'Efficient oleoresin biomass production in pines
563 using low cost metal containing stimulant paste', *Biomass and Bioenergy*, 35(10), pp. 4442–
564 4448. doi: 10.1016/j.biombioe.2011.08.021.

565 Rodrigues-Corrêa, K. C. da S. and Fett-Neto, A. G. (2013) 'Seasonality and chemical
566 elicitation of defense oleoresin production in field-grown slash pine under subtropical

567 climate', *Theoretical and Experimental Plant Physiology*, 25(1), pp. 56–61. doi:
568 10.1590/s2197-00252013000100007.

569 Rodrigues-Corrêa, K. C. da S., de Lima, J. C. and Fett-Neto, A. G. (2012) 'Pine oleoresin:
570 Tapping green chemicals, biofuels, food protection, and carbon sequestration from
571 multipurpose trees', *Food and Energy Security*. John Wiley & Sons, Ltd, pp. 81–93. doi:
572 10.1002/fes3.13.

573 Rodríguez-García, A. *et al.* (2016) 'Effect of four tapping methods on anatomical traits and
574 resin yield in Maritime pine (*Pinus pinaster* Ait.)', *Industrial Crops and Products*. Elsevier
575 B.V., 86, pp. 143–154. doi: 10.1016/j.indcrop.2016.03.033.

576 Rodríguez-García, A. *et al.* (2018) 'Can prescribed burning improve resin yield in a tapped
577 *Pinus pinaster* stand?', *Industrial Crops and Products*. Elsevier B.V., 124, pp. 91–98. doi:
578 10.1016/J.INDCROP.2018.07.049.

579 Rubini, M. *et al.* (2021) 'Comparison of the performances of handheld and benchtop near
580 infrared spectrometers: Application on the quantification of chemical components in maritime
581 pine (*Pinus Pinaster*) resin', *Talanta*, 221. doi: 10.1016/j.talanta.2020.121454.

582 Silva, A. C. R. da *et al.* (2012) 'Biological Activities of α -Pinene and β -Pinene Enantiomers',
583 *Molecules*, 17(6), pp. 6305–6316. doi: 10.3390/molecules17066305.

584 Silvestre, A. J. D. and Gandini, A. (2008a) 'Chapter 2 - Terpenes: Major Sources, Properties
585 and Applications', in. Elsevier, pp. 17–38. doi: [https://doi.org/10.1016/B978-0-08-045316-](https://doi.org/10.1016/B978-0-08-045316-3.00002-8)
586 [3.00002-8](https://doi.org/10.1016/B978-0-08-045316-3.00002-8).

587 Silvestre, A. J. D. and Gandini, A. (2008b) 'Chapter 4 - Rosin: Major Sources, Properties and
588 Applications', in. Elsevier, pp. 67–88. doi: [https://doi.org/10.1016/B978-0-08-045316-](https://doi.org/10.1016/B978-0-08-045316-3.00004-1)
589 [3.00004-1](https://doi.org/10.1016/B978-0-08-045316-3.00004-1).

590 Soliño, M. *et al.* (2018) 'Resin-tapped pine forests in Spain: Ecological diversity and
591 economic valuation', *Science of the Total Environment*. Elsevier B.V., 625, pp. 1146–1155.

592 doi: 10.1016/J.SCITOTENV.2018.01.027.

593 Turtola, S. *et al.* (2003) 'Drought stress alters the concentration of wood terpenoids in Scots
594 pine and Norway spruce seedlings', *Journal of Chemical Ecology*, 29(9), pp. 1981–1995. doi:
595 10.1023/A:1025674116183.

596 Valterová, I. *et al.* (1995) 'Contents and enantiomeric compositions of monoterpene
597 hydrocarbons in xylem oleoresins from four *Pinus* species growing in Cuba. Comparison of
598 trees unattacked and attacked by *Dioryctria horneana*', *Biochemical Systematics and Ecology*,
599 23(1), pp. 1–15. doi: 10.1016/0305-1978(94)00072-O.

600 Wenig, P. and Odermatt, J. (2010) 'OpenChrom: A cross-platform open source software for
601 the mass spectrometric analysis of chromatographic data', *BMC Bioinformatics*. Department
602 of Wood Science, University of Hamburg, Leuschnerstraße 91, 21031 Hamburg, Germany,
603 11. doi: 10.1186/1471-2105-11-405.

604

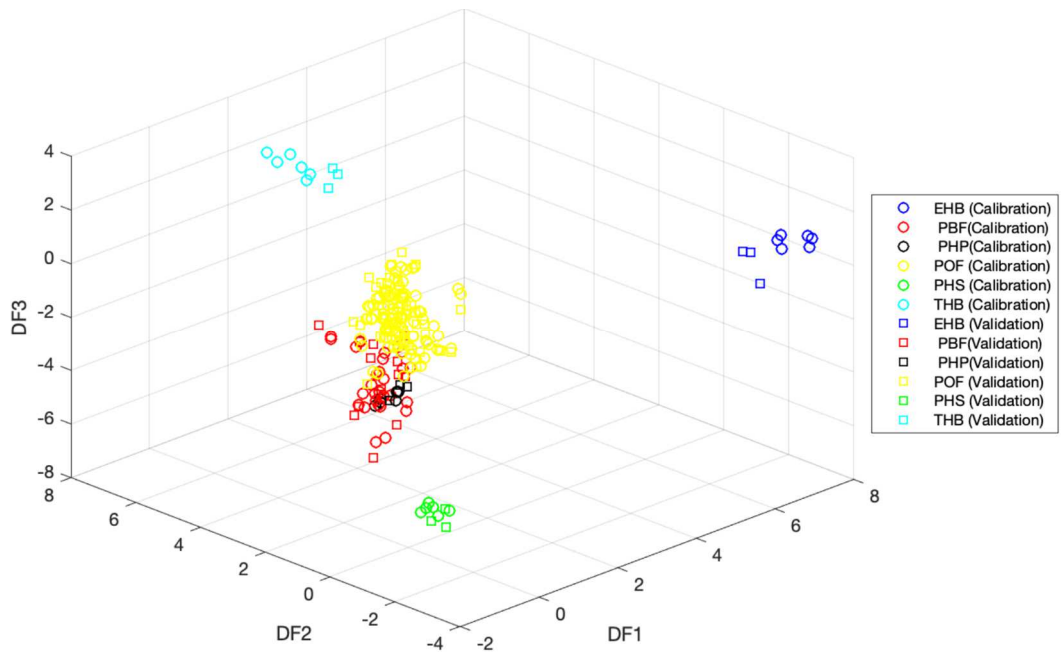


Fig. 1 Discriminant scores plot for the projection of training (circles) & validation (squares) data representing chemical components of all oleoresin samples investigated

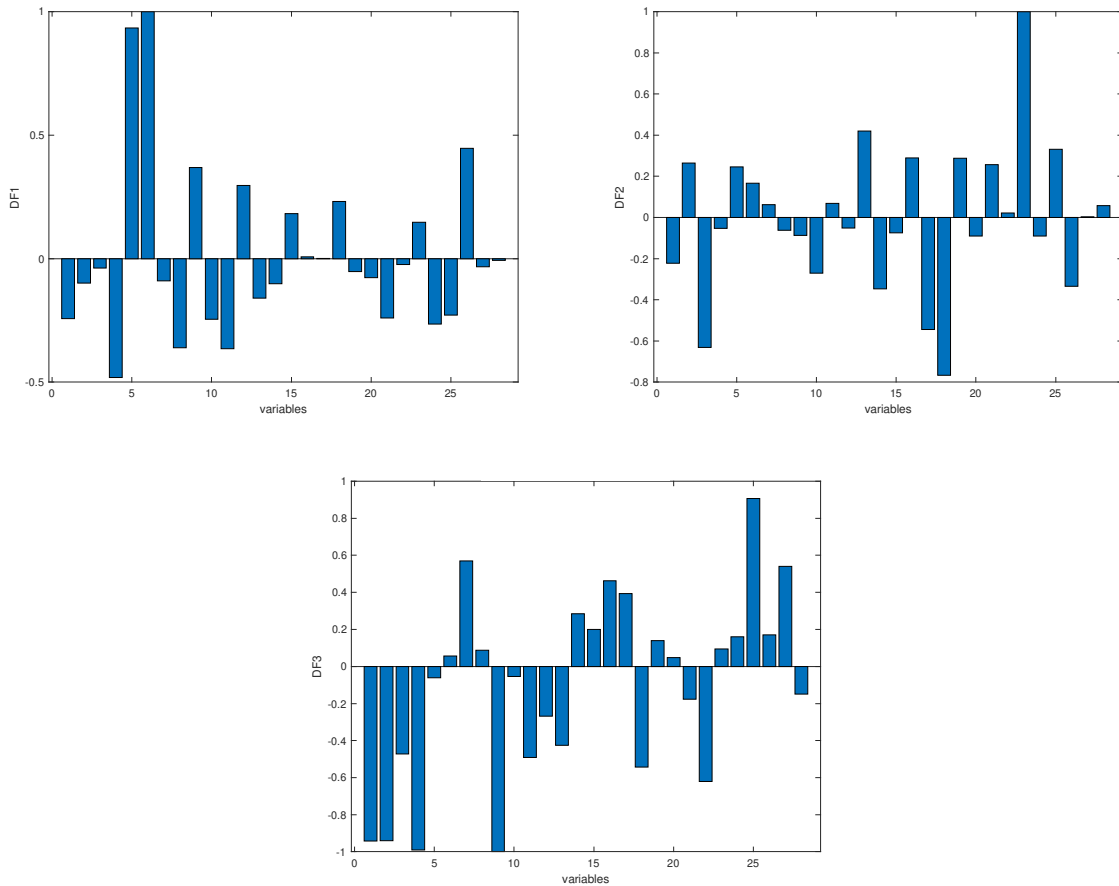


Fig. 2 Loading plots for the LDA model discriminating all oleoresin sample classes

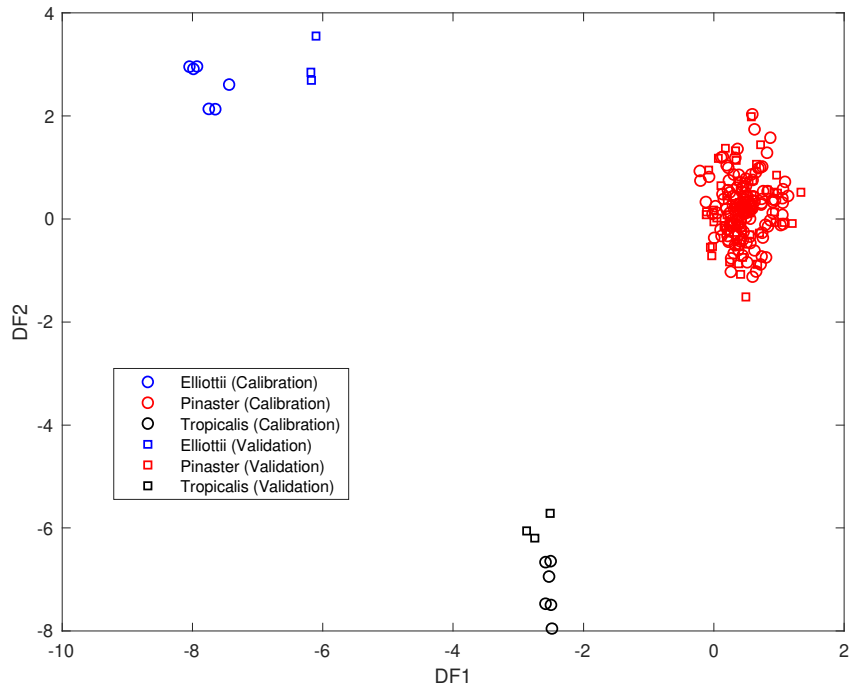


Fig. 3 Discriminant scores plot for the projection of training (circles) & validation (squares) data representing chemical components of different *Pinus* species

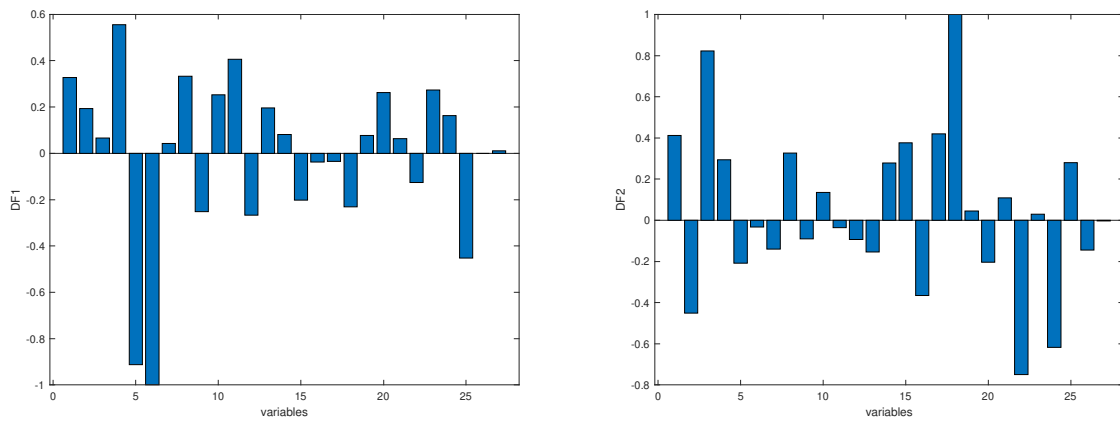


Fig. 4 Loading plots for the LDA model discriminating classes of different *Pinus* species investigated

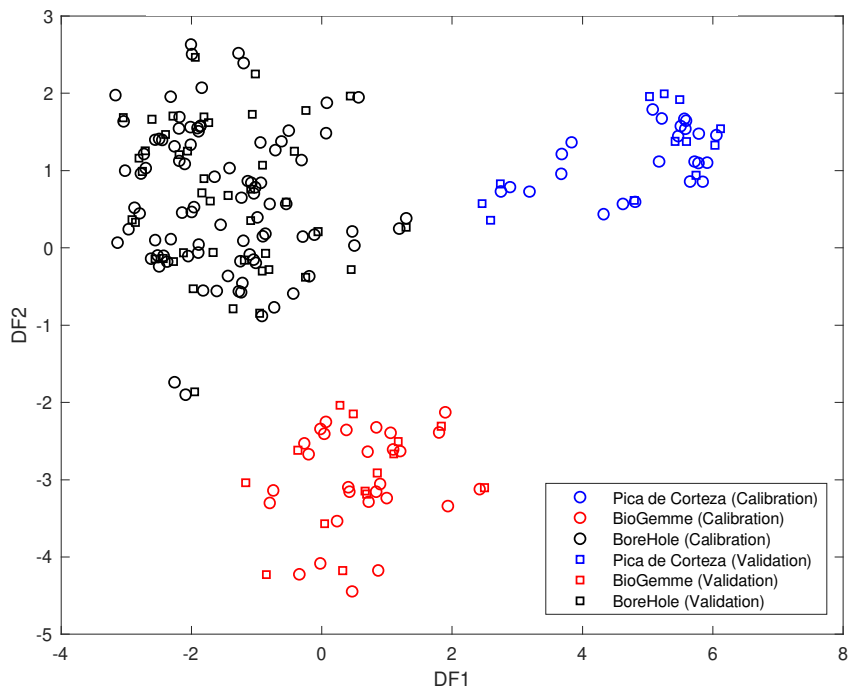


Fig. 5 Discriminant scores plot for the projection of training (circles) & validation (squares) data representing chemical components related to studied tapping methods

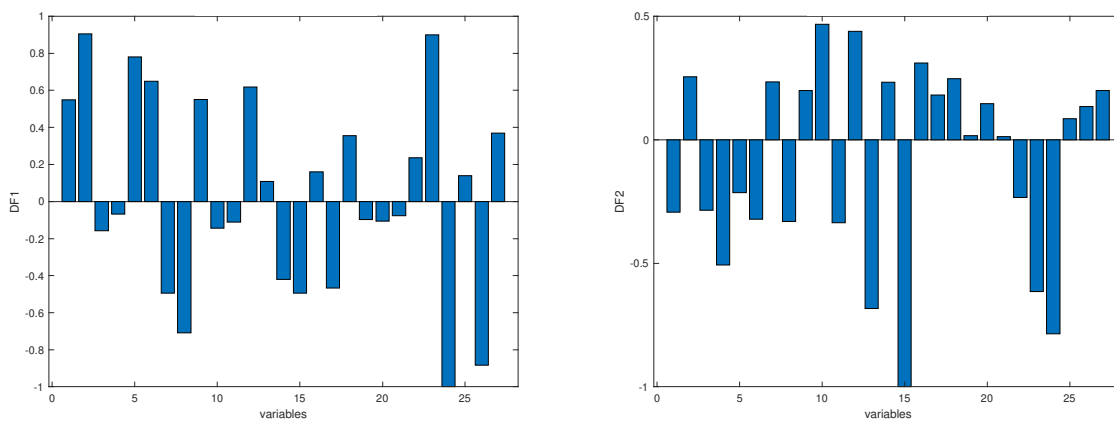


Fig. 6 Loading plots for the LDA model discriminating classes of different tapping methods

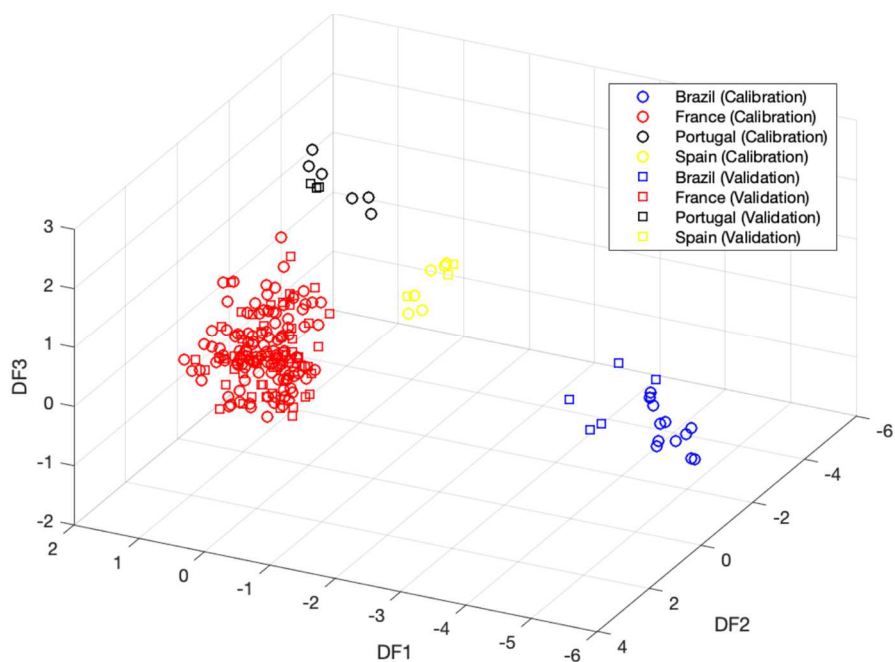


Fig. 7 Discriminant scores plot for the projection of training (circles) & validation (squares) data representing chemical components related to the geographical origin of samples

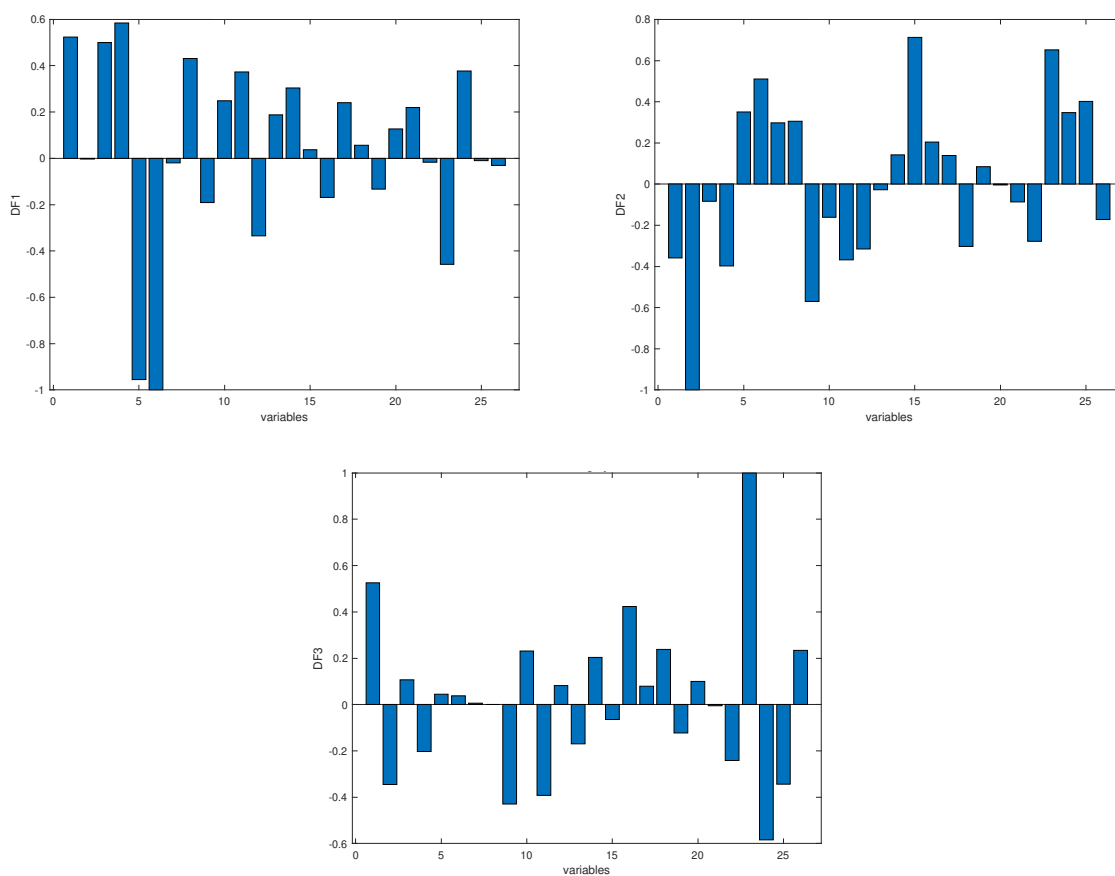


Fig. 8 Loading plots for the LDA model discriminating classes of the samples' geographical origin

Tab. 1 List of samples analyzed in this study

Coded name	Number Of samples	Species	Tapping	Geographical origins
PBF	14	<i>P. pinaster</i>	BioGemme	France
POF	44	<i>P. pinaster</i>	Borehole	France
PHS	3 (mixture)	<i>P. pinaster</i>	Pica de corteza	Spain
PHP	3 (mixture)	<i>P. pinaster</i>	Pica de corteza	Portugal
EHB	3 (mixture)	<i>P. elliottii</i>	Pica de corteza	Brazil
THB	3 (mixture)	<i>P. tropicalis</i>	Pica de corteza	Brazil

Tab. 2 Oleoresin composition of the samples analyzed in this study

Molecule	EHB	PBF	PHP	PHS	POF	THB
α -pinene	35.72 \pm 2.49	46.54 \pm 4.51	53.15 \pm 1.14	48.06 \pm 2.03	41.28 \pm 4.50	44.44 \pm 1.51
Camphene	0.43 \pm 0.04	0.57 \pm 0.14	0.55 \pm 0.10	0.92 \pm 0.06	0.46 \pm 0.07	0.66 \pm 0.05
β -pinene	32.53 \pm 1.62	15.71 \pm 3.38	10.55 \pm 0.40	17.93 \pm 1.58	18.07 \pm 3.52	8.03 \pm 1.21
β -myrcene	1.08 \pm 0.04	0.87 \pm 0.18	0.62 \pm 0.05	1.14 \pm 0.06	0.91 \pm 0.19	0.37 \pm 0.02
(2E,4E)-3,7-dimethyl-2,4-octadiene	1.84 \pm 1.76	2.34 \pm 0.44	1.78 \pm 0.08	2.85 \pm 0.08	2.00 \pm 0.35	2.75 \pm 0.04
Limonene	2.05 \pm 1.14	0.44 \pm 0.08	0.36 \pm 0.01	0.50 \pm 0.03	0.35 \pm 0.05	0.48 \pm 0.03
β -phellandrene	0.07 \pm 0.00	0.02 \pm 0.02	0.03 \pm 0.00	0.01 \pm 0.00	0.07 \pm 0.03	0.04 \pm 0.00
γ -terpinene	0.37 \pm 0.05	0.7 \pm 0.16	0.72 \pm 0.03	0.76 \pm 0.05	0.63 \pm 0.18	0.78 \pm 0.01
Terpinolene	0.07 \pm 0.01	0.12 \pm 0.06	0.22 \pm 0.01	0.12 \pm 0.01	0.19 \pm 0.23	0.31 \pm 0.03
α -cubebene	0.02 \pm 0.01	0.34 \pm 0.25	0.33 \pm 0.02	0.33 \pm 0.01	0.26 \pm 0.16	0.32 \pm 0.02
α -longipinene	0.02 \pm 0.01	0.41 \pm 0.2	0.28 \pm 0.01	0.27 \pm 0.05	0.29 \pm 0.09	0.22 \pm 0.01
Cyclosativene	0.11 \pm 0.02	0.10 \pm 0.07	0.12 \pm 0.01	0.09 \pm 0.01	0.09 \pm 0.05	0.18 \pm 0.01
α -copaene	0.11 \pm 0.01	3.68 \pm 2.77	3.48 \pm 0.26	3.47 \pm 0.49	2.78 \pm 1.98	3.69 \pm 0.20
Sativene	0.09 \pm 0.02	5.68 \pm 1.39	5.40 \pm 0.46	3.98 \pm 0.43	6.35 \pm 1.82	0.30 \pm 0.01
Longifolene	0.04 \pm 0.01	0.17 \pm 0.07	0.11 \pm 0.01	0.11 \pm 0.00	0.11 \pm 0.03	0.06 \pm 0.00
epi-bicyclosesquiphellandrene	0.01 \pm 0.00	0.20 \pm 0.14	0.16 \pm 0.03	0.18 \pm 0.01	0.14 \pm 0.07	0.04 \pm 0.00
γ -muurolene	0.01 \pm 0.00	0.04 \pm 0.04	0.04 \pm 0.01	0.03 \pm 0.00	0.04 \pm 0.01	0.03 \pm 0.01
Manoyl oxide	0.03 \pm 0.01	0.12 \pm 0.08	0.23 \pm 0.01	0.03 \pm 0.01	0.21 \pm 0.25	0.03 \pm 0.00
Pimarinal	0.06 \pm 0.01	0.06 \pm 0.02	0.05 \pm 0.02	0.04 \pm 0.00	0.06 \pm 0.01	0.11 \pm 0.02
Pimara-7,15-dien-3-one	0.06 \pm 0.00	0.16 \pm 0.04	0.22 \pm 0.02	0.25 \pm 0.03	0.26 \pm 0.12	0.14 \pm 0.01
Dehydroabietinal	0.06 \pm 0.00	0.13 \pm 0.09	0.15 \pm 0.01	0.1 \pm 0.01	0.15 \pm 0.04	0.24 \pm 0.05
Pimaric acid	0.05 \pm 0.01	0.14 \pm 0.04	0.09 \pm 0.01	0.05 \pm 0.00	0.09 \pm 0.02	0.11 \pm 0.01
Retinol	0.34 \pm 0.02	0.78 \pm 0.2	0.76 \pm 0.02	0.24 \pm 0.01	0.64 \pm 0.17	1.13 \pm 0.14
Isopimaric acid	1.00 \pm 0.19	2.36 \pm 0.51	1.61 \pm 0.14	0.88 \pm 0.07	1.88 \pm 0.42	2.03 \pm 0.11
Methyl dehydroabietate	0.09 \pm 0.00	0.09 \pm 0.03	0.06 \pm 0.00	0.05 \pm 0.00	0.12 \pm 0.06	0.21 \pm 0.00
Palustric acid	3.59 \pm 0.15	1.29 \pm 0.47	1.09 \pm 0.10	0.65 \pm 0.02	1.15 \pm 0.28	2.19 \pm 0.17
Dehydroabietic acid	0.10 \pm 0.02	0.47 \pm 0.12	0.30 \pm 0.01	0.21 \pm 0.01	0.62 \pm 0.21	0.84 \pm 0.06
Abietic acid	0.39 \pm 0.22	0.58 \pm 0.28	0.64 \pm 0.03	0.27 \pm 0.02	0.72 \pm 0.30	1.04 \pm 0.36

Tab. 3 Oleoresin composition regarding the *Pinus* species associated with their *p*-value of post hoc Tukey HSD test

Molecule	<i>Pinus elliotii</i> (n = 9)	<i>Pinus pinaster</i> (n = 192)	<i>Pinus tropicalis</i> (n = 9)	<i>p</i> -value of post hoc Tukey HSD test
α -pinene	35.72 \pm 2.49 (b)	43.31 \pm 5.41 (a)	44.44 \pm 1.51 (a)	< 0.001
Camphene	0.43 \pm 0.04 (c)	0.51 \pm 0.13 (b)	0.66 \pm 0.05 (a)	< 0.001
β -pinene	32.53 \pm 1.62 (a)	17.19 \pm 3.77 (b)	8.03 \pm 1.21 (c)	< 0.001
β -myrcene	1.08 \pm 0.04 (a)	0.90 \pm 0.2 (b)	0.37 \pm 0.02 (c)	< 0.001
(2E,4E)-3,7-dimethyl-2,4-octadiene	1.84 \pm 1.76 (b)	2.10 \pm 0.42 (b)	2.75 \pm 0.04 (a)	< 0.001
Limonene	2.05 \pm 1.14 (a)	0.37 \pm 0.07 (b)	0.48 \pm 0.03 (b)	< 0.001
β -phellandrene	0.07 \pm 0.00 (a)	0.05 \pm 0.03 (a, b)	0.04 \pm 0.00 (b)	< 0.001
γ -terpinene	0.37 \pm 0.05 (b)	0.65 \pm 0.17 (a)	0.78 \pm 0.01 (a)	< 0.001
Terpinolene	0.07 \pm 0.01 (b)	0.17 \pm 0.2 (a, b)	0.31 \pm 0.03 (a)	< 0.001
α -cubebene	0.02 \pm 0.01 (b)	0.28 \pm 0.18 (a)	0.32 \pm 0.02 (a)	< 0.001
α -longipinene	0.02 \pm 0.01 (b)	0.31 \pm 0.13 (a)	0.22 \pm 0.01 (a)	< 0.001
Cyclosativene	0.11 \pm 0.02 (b)	0.10 \pm 0.05 (b)	0.18 \pm 0.01 (a)	< 0.001
α -copaene	0.11 \pm 0.01 (b)	3.05 \pm 2.12 (a)	3.69 \pm 0.20 (a)	< 0.001
Sativene	0.09 \pm 0.02 (b)	6.05 \pm 1.74 (a)	0.30 \pm 0.01 (b)	< 0.001
Longifolene	0.04 \pm 0.01 (b)	0.12 \pm 0.05 (a)	0.06 \pm 0.00 (b)	< 0.001
epi-bicyclosesquiphellandrene	0.01 \pm 0.00 (b)	0.16 \pm 0.09 (a)	0.04 \pm 0.00 (b)	< 0.001
γ -muurolene	0.01 \pm 0.00 (b)	0.04 \pm 0.02 (a)	0.03 \pm 0.01 (a, b)	< 0.001
Manoyl oxide	0.03 \pm 0.01 (a)	0.18 \pm 0.21 (a)	0.03 \pm 0.00 (a)	< 0.001
Pimarinal	0.06 \pm 0.01 (a)	0.06 \pm 0.02 (a)	0.11 \pm 0.02 (a)	0.063
Pimara-7,15-dien-3-one	0.06 \pm 0.00 (b)	0.24 \pm 0.11 (a)	0.14 \pm 0.01 (b)	< 0.001
Dehydroabietinal	0.06 \pm 0.00 (c)	0.15 \pm 0.06 (b)	0.24 \pm 0.05 (a)	< 0.001
Pimaric acid	0.05 \pm 0.01 (b)	0.1 \pm 0.04 (a)	0.11 \pm 0.01 (a)	< 0.001
Retinol	0.34 \pm 0.02 (c)	0.66 \pm 0.2 (b)	1.13 \pm 0.14 (a)	< 0.001
Isopimaric acid	1.00 \pm 0.19 (b)	1.93 \pm 0.52 (a)	2.03 \pm 0.11 (a)	< 0.001
Methyl dehydroabietate	0.09 \pm 0.00 (a)	0.11 \pm 0.06 (a)	0.21 \pm 0.00 (b)	0.029
Palustric acid	3.59 \pm 0.15 (a)	1.16 \pm 0.34 (c)	2.19 \pm 0.17 (b)	< 0.001
Dehydroabietic acid	0.10 \pm 0.02 (c)	0.56 \pm 0.22 (b)	0.84 \pm 0.06 (a)	< 0.001
Abietic acid	0.39 \pm 0.22 (c)	0.66 \pm 0.30 (b)	1.04 \pm 0.36 (a)	< 0.001

Lower case letters inside the brackets represent the values of the pairwise comparisons (statistically different results)

Tab. 4 Confusion table of the LDA model based on the *Pinus* species

Class:	<i>Sensitivity</i>		<i>Specificity</i>		<i>Precision</i>	
	Cross-Validation	Prediction	Cross-Validation	Prediction	Cross-Validation	Prediction
<i>P. elliotii</i>	100.0	100.0	100.0	100.0	100.0	100.0
<i>P. pinaster</i>	100.0	100.0	100.0	100.0	100.0	100.0
<i>P. tropicalis</i>	100.0	100.0	100.0	100.0	100.0	100.0

Tab. 5 Oleoresin composition and tapping methods associated with their *p-value of post hoc Tukey HSD test*

Molecule	Biogemme (n = 42)	Borehole (n = 132)	Pica de corteza (n = 36)	<i>p-value of post hoc Tukey HSD test</i>
α -pinene	46.54 \pm 4.51 (a)	41.28 \pm 4.50 (b)	45.34 \pm 6.69 (a)	< 0.001
Camphene	0.57 \pm 0.14 (b)	0.46 \pm 0.07 (c)	0.64 \pm 0.19 (a)	< 0.001
β -pinene	15.71 \pm 3.38 (b)	18.07 \pm 3.52 (a)	17.26 \pm 9.75 (a, b)	< 0.001
β -myrcene	0.87 \pm 0.18 (a, b)	0.91 \pm 0.19 (a)	0.80 \pm 0.33 (b)	< 0.001
(2E,4E)-3.7-dimethyl-2.4-octadiene	2.34 \pm 0.44 (a)	2.00 \pm 0.35 (b)	2.30 \pm 0.98 (a)	< 0.001
Limonene	0.44 \pm 0.08 (b)	0.35 \pm 0.05 (b)	0.85 \pm 0.89 (a)	< 0.001
β -phellandrene	0.02 \pm 0.02 (b)	0.07 \pm 0.03 (a)	0.04 \pm 0.02 (b)	0.035
γ -terpinene	0.7 \pm 0.16 (a)	0.63 \pm 0.18 (b)	0.66 \pm 0.17 (a, b)	0.033
Terpinolene	0.12 \pm 0.06 (a)	0.19 \pm 0.23 (a)	0.18 \pm 0.09 (a)	< 0.001
α -cubebene	0.34 \pm 0.25 (a)	0.26 \pm 0.16 (b)	0.25 \pm 0.14 (a, b)	< 0.001
α -longipinene	0.41 \pm 0.20 (a)	0.29 \pm 0.09 (b)	0.20 \pm 0.11 (c)	< 0.001
Cyclosativene	0.10 \pm 0.07 (a, b)	0.09 \pm 0.05 (b)	0.12 \pm 0.03 (a)	< 0.001
α -copaene	3.68 \pm 2.77 (a)	2.78 \pm 1.98 (b)	2.69 \pm 1.54 (a, b)	< 0.001
Sativene	5.68 \pm 1.39 (a)	6.35 \pm 1.82 (a)	2.44 \pm 2.36 (b)	< 0.001
Longifolene	0.17 \pm 0.07 (a)	0.11 \pm 0.03 (b)	0.08 \pm 0.03 (c)	< 0.001
epi-bicyclosesquiphellandrene	0.20 \pm 0.14 (a)	0.14 \pm 0.07 (b)	0.10 \pm 0.07 (c)	< 0.001
γ -muurolene	0.04 \pm 0.04 (a)	0.04 \pm 0.01 (a, b)	0.03 \pm 0.01 (b)	< 0.001
Manoyl oxide	0.12 \pm 0.08 (b)	0.21 \pm 0.25 (a)	0.08 \pm 0.09 (b)	< 0.001
Pimarinal	0.06 \pm 0.02 (b)	0.06 \pm 0.01 (b)	0.07 \pm 0.03 (a)	< 0.001
Pimara-7,15-dien-3-one	0.16 \pm 0.04 (a)	0.26 \pm 0.12 (a)	0.17 \pm 0.07 (a)	0.081
Dehydroabietinal	0.13 \pm 0.09 (a)	0.15 \pm 0.04 (a)	0.14 \pm 0.07 (a)	< 0.001
Pimaric acid	0.14 \pm 0.04 (a)	0.09 \pm 0.02 (b)	0.07 \pm 0.03 (c)	< 0.001
Retinol	0.78 \pm 0.20 (a)	0.64 \pm 0.17 (b)	0.62 \pm 0.37 (b)	< 0.001
Isopimaric acid	2.36 \pm 0.51 (a)	1.88 \pm 0.42 (b)	1.38 \pm 0.49 (c)	< 0.001
Methyl dehydroabietate	0.09 \pm 0.03 (a)	0.12 \pm 0.06 (a)	0.10 \pm 0.06 (a)	0.161
Palustric acid	1.29 \pm 0.47 (b)	1.15 \pm 0.28 (c)	1.88 \pm 1.16 (a)	0.024
Dehydroabietic acid	0.47 \pm 0.12 (b)	0.62 \pm 0.21 (a)	0.36 \pm 0.29 (c)	< 0.001
Abietic acid	0.58 \pm 0.28 (b)	0.72 \pm 0.30 (a)	0.59 \pm 0.36 (b)	< 0.001
Lower case letters inside the brackets represent the values of the pairwise comparisons (statistically different results)				

Tab. 6 Confusion table of the LDA model based the tapping methods

Class:	<i>Sensitivity</i>		<i>Specificity</i>		<i>Precision</i>	
	Cross-Validation	Prediction	Cross-Validation	Prediction	Cross-Validation	Prediction
Hugues	100.0	100.0	100.0	100.0	100.0	100.0
BioGemme	100.0	100.0	100.0	100.0	100.0	100.0
BoreHole	98.9	97.7	99.1	98.2	96.6	93.3

Tab. 7 Oleoresin composition and geographical origin associated with their *p*-value of post hoc Tukey HSD test

Molecule	Brazil (n = 18)	France (n = 174)	Portugal (n = 9)	Spain (n = 9)	<i>p</i> -value of post hoc Tukey HSD test
α -pinene	40.08 \pm 4.91 (c)	42.55 \pm 5.03 (c)	53.15 \pm 1.14 (a)	48.06 \pm 2.03 (b)	< 0.001
Camphene	0.55 \pm 0.12 (b)	0.49 \pm 0.10 (b)	0.55 \pm 0.10 (b)	0.92 \pm 0.06 (a)	< 0.001
β -pinene	20.28 \pm 12.68 (a)	17.50 \pm 3.62 (b)	10.55 \pm 0.40 (c)	17.93 \pm 1.58 (a, b)	< 0.001
β -myrcene	0.72 \pm 0.37 (c)	0.90 \pm 0.19 (b)	0.62 \pm 0.05 (c)	1.14 \pm 0.06 (a)	< 0.001
(2E,4E)-3.7-dimethyl-2.4-octadiene	2.30 \pm 1.30 (b)	2.08 \pm 0.40 (b)	1.78 \pm 0.08 (b)	2.85 \pm 0.08 (a)	< 0.001
Limonene	1.27 \pm 1.12 (a)	0.37 \pm 0.07 (b)	0.36 \pm 0.01 (b)	0.50 \pm 0.03 (b)	< 0.001
β -phellandrene	0.05 \pm 0.02 (a, b)	0.06 \pm 0.03 (a)	0.03 \pm 0.00 (b, c)	0.01 \pm 0.00 (c)	< 0.001
γ -terpinene	0.58 \pm 0.21 (b, c)	0.65 \pm 0.18 (a, c)	0.72 \pm 0.03 (a, b)	0.76 \pm 0.05 (a)	< 0.001
Terpinolene	0.19 \pm 0.12 (a)	0.17 \pm 0.21 (a)	0.22 \pm 0.01 (a)	0.12 \pm 0.01 (a)	< 0.001
α -cubebene	0.17 \pm 0.16 (a)	0.28 \pm 0.19 (a)	0.33 \pm 0.02 (a)	0.33 \pm 0.01 (a)	< 0.001
α -longipinene	0.12 \pm 0.10 (b)	0.32 \pm 0.13 (a)	0.28 \pm 0.01 (a)	0.27 \pm 0.05 (a)	< 0.001
Cyclosativene	0.14 \pm 0.04 (a)	0.10 \pm 0.06 (b, c)	0.12 \pm 0.01 (a, b)	0.09 \pm 0.01 (a, c)	< 0.001
α -copaene	1.9 \pm 1.85 (a)	3.00 \pm 2.22 (a)	3.48 \pm 0.26 (a)	3.47 \pm 0.49 (a)	< 0.001
Sativene	0.19 \pm 0.11 (c)	6.19 \pm 1.75 (a)	5.4 \pm 0.46 (a, b)	3.98 \pm 0.43 (b)	< 0.001
Longifolene	0.05 \pm 0.01 (b)	0.12 \pm 0.05 (a)	0.11 \pm 0.01 (a)	0.11 \pm 0.00 (a)	< 0.001
epi-bicyclosesquiphellandrene	0.03 \pm 0.02 (b)	0.15 \pm 0.10 (a)	0.16 \pm 0.03 (a)	0.18 \pm 0.01 (a)	< 0.001
γ -muurolene	0.02 \pm 0.01 (b, c)	0.04 \pm 0.02 (a)	0.04 \pm 0.01 (a, b)	0.03 \pm 0.00 (a, c)	< 0.001
Manoyl oxide	0.03 \pm 0.01 (a)	0.19 \pm 0.22 (a, b)	0.23 \pm 0.01 (a, b)	0.03 \pm 0.01 (a, b)	< 0.001
Pimarinal	0.09 \pm 0.03 (a)	0.06 \pm 0.02 (b)	0.05 \pm 0.02 (b, c)	0.04 \pm 0.00 (c)	< 0.001
Pimara-7,15-dien-3-one	0.1 \pm 0.04 (b)	0.24 \pm 0.11 (a)	0.22 \pm 0.02 (a)	0.25 \pm 0.03 (a)	0.044
Dehydroabietinal	0.15 \pm 0.10 (a)	0.15 \pm 0.06 (a)	0.15 \pm 0.01 (a)	0.10 \pm 0.01 (a)	< 0.001
Pimaric acid	0.08 \pm 0.03 (b, c)	0.10 \pm 0.04 (a)	0.09 \pm 0.01 (a, b)	0.05 \pm 0.00 (c)	< 0.001
Retinol	0.74 \pm 0.42 (a)	0.68 \pm 0.19 (a)	0.76 \pm 0.02 (a)	0.24 \pm 0.01 (b)	< 0.001
Isopimaric acid	1.51 \pm 0.55 (b)	2.00 \pm 0.48 (a)	1.61 \pm 0.14 (b)	0.88 \pm 0.07 (c)	< 0.001
Methyl dehydroabietate	0.15 \pm 0.06 (a)	0.11 \pm 0.06 (a)	0.06 \pm 0.00 (a)	0.05 \pm 0.00 (a)	0.751
Palustric acid	2.89 \pm 0.74 (a)	1.19 \pm 0.34 (a)	1.09 \pm 0.10 (a)	0.65 \pm 0.02 (a)	0.059
Dehydroabietic acid	0.47 \pm 0.38 (a, b)	0.59 \pm 0.21 (a)	0.3 \pm 0.01 (b, c)	0.21 \pm 0.01 (c)	< 0.001
Abietic acid	0.72 \pm 0.44 (a)	0.68 \pm 0.30 (a)	0.64 \pm 0.03 (a)	0.27 \pm 0.02 (b)	< 0.001

Lower case letters inside the brackets represent the values of the pairwise comparisons (statistically different results)

Tab. 8 Confusion table of the LDA model based on the geographical origin

Class:	Sensitivity		Specificity		Precision	
	Cross-Validation	Prediction	Cross-Validation	Prediction	Cross-Validation	Prediction
Brazil	100.0	100.0	100.0	100.0	100.0	100.0
France	99.1	98.3	99.3	98.5	85.7	75.0
Portugal	100.0	100.0	100.0	100.0	100.0	100.0
Spain	100.0	100.0	100.0	100.0	100.0	100.0