**Analysis of minor low molecular weight carbohydrates in cocoa beans by
chromatographic techniques coupled to mass spectrometry**

Roberto Megías-Pérez¹, Ana Isabel Ruiz-Matute²,

Marcello Corno³, Nikolai Kuhnert^{1*}

¹*Department of Life Sciences and Chemistry, Jacobs University Bremen, Campus Ring 1,
28759 Bremen, Germany*

²*Instituto de Química Orgánica General (CSIC), Juan de la Cierva, 3, 28028, Madrid, Spain*

³*Barry Callebaut Belgium N.V., Aalstersestraat 122, 9280 Lebbeke-Wieze, Belgium*

*Correspondence to:

Prof. Dr. Nikolai Kuhnert,

Department of Life Sciences and Chemistry,

Jacobs University, Campus Ring 1,

28759 Bremen, Germany.

Phone: +49 421 200 3120,

Fax: +49 421 200 3249.

E-mail: n.kuhnert@jacobs-university.de

22 **Highlights:**

- 23 • *scyllo*-Inositol and kestoses were found for the first time in cocoa beans by GC-MS
- 24 • *scyllo*-Inositol, 1-kestose and galactinol were selected as target carbohydrates
- 25 • HILIC-ESI-TOF MS method was optimized for the analysis of target carbohydrates
- 26 • Target LMWC were quantified in cocoa beans from different origin and fermentation
- 27 status
- 28 • Galactinol could be a potential indicator of fermentation status

29

30 Abstract

31 The low molecular weight carbohydrate (LMWC) profile of cocoa beans has recently been
32 studied using hydrophilic interaction liquid chromatography coupled to electrospray
33 ionization-time of flight mass spectrometry (HILIC-ESI-TOF MS) and HILIC-ESI-tandem
34 mass spectrometry (HILIC-ESI-MSⁿ). However, different LMWC could not be
35 unambiguously identified. Thus, as a first approach in this paper, gas chromatography coupled
36 to mass spectrometry (GC-MS) was used as a complementary analytical technique to
37 characterize LMWC of cocoa beans. Different mono-, di-, tri- and tetrasaccharides, as well as
38 *myo*-inositol, galactinol and a diglycosil glycerol were detected. *scyllo*-Inositol, 1-kestose and
39 6-kestose were identified in unfermented cocoa beans for the first time. Moreover, other
40 minor LMWC were tentatively assigned as fructosyl-fructose, fructosyl-glucose and glucosyl-
41 sucrose. As a second step, in order to evaluate new possible indicators of cocoa bean origin or
42 fermentation status, *scyllo*-inositol, 1-kestose and galactinol were selected as target
43 compounds and a HILIC-ESI-TOF MS method was optimized for their analysis. The
44 optimized conditions, using an acetonitrile:water gradient with 0.05% ammonium hydroxide
45 at 40°C showed narrow peaks (w_h : 0.3-0.5 min) with good resolution values (R_s : 0.83–2.83).
46 The validated HILIC-ESI-TOF MS method was applied to the analysis of 35 cocoa bean
47 samples from different origins and fermentation status. The content of *scyllo*-inositol, 1-
48 kestose and galactinol in unfermented beans (n=21) was in the range of traces-504.9, 36.1-
49 133.5 and traces-1970.4 $\mu\text{g g}^{-1}$ cocoa DM respectively. In fermented beans (n=14), the
50 content of *scyllo*-inositol and 1-kestose was in the range of 15.5-491.9 and traces-115.5 $\mu\text{g g}^{-1}$
51 cocoa DM respectively. Galactinol was absent in fermented beans, indicating that it could be
52 a potential indicator of fermentation status. The methodology proposed could be used for
53 quality control of natural products and other food ingredients containing inositols and
54 oligosaccharides.

55 **Keywords:** *scyllo*-inositol, 1-kestose, galactinol, cocoa bean, HILIC-ESI-TOF MS, GC-MS.

56 1. INTRODUCTION

57 Cocoa beans, the seeds of the cocoa tree *Theobroma cacao* L., are used as raw material in the
58 manufacturing of chocolate and other derived products. They constitute the principal
59 agricultural export commodity for the producing countries and they are also the main source
60 of income for about 6 million smallholders around the world. [1, 2].

61 In recent years, the study of cocoa beans has received much attention from the scientific
62 community. Different research lines have been focused on the characterization of unknown
63 compounds[3], the search for indicators of bean origin or status of fermentation [4], the
64 evaluation of quality parameters to distinguish among hybrid cultivars [5] and the
65 development of a sustainable production of high-quality beans and cocoa products [6], among
66 others. Most of these researches are based on the study of polyphenols [7, 8], volatile
67 compounds [9], lipids [10] or proteins/peptides [11] in cocoa beans and their evolution during
68 fermentation processes and roasting. However, despite the significant role of carbohydrates in
69 the Maillard reaction, which takes place during cocoa bean roasting [12], few studies have
70 been reported regarding low molecular weight carbohydrates (LMWC) in cocoa beans.

71 Among the different analytical techniques that can be used for carbohydrate analysis, gas
72 chromatography (GC) and liquid chromatography (LC) coupled to mass spectrometry (MS)
73 are the most extensively employed [13].

74 GC-MS has been widely used for the analysis of LMWC due to its high-resolution power,
75 sensitivity and potential for structural identification. In GC-MS, the combination of GC
76 retention times (or retention indices) and specific electron impact ionization (EI) mass spectra
77 of derivatives provide valuable information about the chemical structure of a molecule. This
78 information is specifically useful for complex mixtures of carbohydrate isomers, where
79 compounds with the same molecular weight differ only in the configuration of their hydroxyl
80 groups and the position of their glycosidic linkages [14]. GC-MS has been applied for the

81 structural characterization of different types of bioactive carbohydrates such as inositols [15],
82 cyclitol glycosides [16], iminosugars [17] or trisaccharides [18]. However, the main
83 disadvantage of the analysis of LMWC using GC-MS is that a previous derivatization step is
84 required [19], which can be a tedious and expensive task for routine analysis.

85 Conversely, in LC-MS the sample preparation for carbohydrate analysis is usually much
86 simpler and does not require a lengthy time, which is an advantage over GC. Among the
87 different operation modes, hydrophilic interaction liquid chromatography (HILIC) is
88 considered an efficient alternative to reverse-phase liquid chromatography for the analysis of
89 complex oligosaccharide mixtures. This technique provides appropriate resolution between
90 isomers and good peak shapes [20]. Furthermore, HILIC is easily coupled to MS due to the
91 mobile phases used, which consist of water mixtures with a high percentage of organic
92 solvents, enhancing the ionization and increasing sensitivity [20]. Successful application of
93 HILIC-MS to the analysis of carbohydrates from different matrices can be found in the
94 literature, such as the analysis of iminosugars in plants [21] or oligosaccharides in milk [20],
95 among others. However, applications of this technique in the analysis of cocoa bean LMWC
96 are very scarce. We have recently developed a methodology based on HILIC-electrospray
97 ionization-time of flight MS (HILIC-ESI-TOF MS) and HILIC-ESI-tandem mass
98 spectrometry (HILIC-ESI-MSⁿ) to monitor changes in cocoa bean LMWC according to their
99 origin and fermentation status [22]. That study has demonstrated the suitability of the
100 carbohydrate profile as an indicator of fermentation status, fermentation procedure and
101 duration of fermentation. Monosaccharides (fructose, glucose), polyols (*myo*-inositol and
102 mannitol), disaccharides (sucrose and melibiose), trisaccharides (raffinose) and
103 tetrasaccharides (stachyose) were detected. Furthermore, the presence of other minor LMWC,
104 which could not be unambiguously identified, was reported. This lack of assignment of

105 identities is mainly a consequence of the inability to identify co-eluting peaks due to the non-
106 specific fragmentation spectra of the carbohydrates.

107 Consequently, the present paper aims to develop a new methodology which allows to
108 investigate the content of unknown minor LMWC detected in cocoa beans. GC-MS has been
109 used to complement the structural information previously reported using HILIC-ESI-MSⁿ. An
110 appropriate methodology using HILIC-ESI-TOF MS has been developed and validated to
111 analyze the selected target LMWC in different types of cocoa bean samples and therefore, to
112 evaluate these LMWC as putative potential indicators of origin or fermentation status.

113 **2. MATERIALS AND METHODS**

114 **2.1. Chemicals and standards.**

115 Dichloromethane and LC-MS grade acetonitrile (ACN) were supplied by Aplichem Panreac
116 (Darmstadt, Germany). Ammonium hydroxide solutions, Asp-Phe methyl ester (used as
117 internal standard in HILIC-ESI-TOF MS analysis), fructose, glucose, galactose, maltose,
118 melibiose, *myo*-inositol, phenyl- β -D-glucoside (used as internal standard in GC-MS analysis),
119 sucrose, raffinose and stachyose were obtained from Sigma Chemical Co. (St. Louis, USA).
120 *Scyllo*-inositol, 1-kestose and galactinol were purchased at Carbosynth (Compton, UK).

121 **2.2. Cocoa bean samples**

122 Cocoa bean samples (n = 35) were provided and certified with regard to their location and
123 status of fermentation by Barry Callebaut Belgium (Table S1, supplementary information).

124 Unfermented cocoa beans (n = 21) were collected from different geographical origins:
125 Ecuador (n = 5), Malaysia (n = 3), Brazil (n = 3), Indonesia (n = 4) and Ivory Coast (n = 6).
126 Unfermented samples were kept at -80°C from the collection time until the beans were
127 processed for analysis.

128 Fourteen spontaneously fermented cocoa beans collected in Ecuador (n = 3), Malaysia (n = 3),
129 Brazil (n = 2), Indonesia (n = 3) and Ivory Coast (n = 3) were as well analyzed. The
130 spontaneous fermentation was induced by piling up the unfermented beans with the pulp in
131 platforms. The beans were rotated during the first 4 days of fermentation to ensure
132 homogenization. Ivory Coast samples (n=2) were fermented for 7 days. There was no data
133 regarding the fermentation duration from one sample from Ivory Coast. Samples from
134 Indonesia, Malaysia and Brazil were fermented for 6 days and samples from Ecuador for a
135 period ranging from 4 to 5 days. After fermentation, cocoa beans were dried under the sun for
136 a period of 7 to 10 days. The samples were stored in falcon tubes at 4°C until the beans were
137 processed for analysis.

138 Cocoa beans were de-shelled manually and ground using a mechanical grinder Retsch (Haan,
139 Germany). Dry matter (DM) content was calculated as the difference in weight of 2 g of
140 cocoa powder before and after heating at 105 °C for 20 hours.

141 **2.3. Extraction of carbohydrates**

142 For the extraction of LMWC from cocoa, defatted cocoa powder (150 mg) was subjected to
143 two cycles of solid-liquid extraction (SLE) using ethanol 80% as extracting solvent, following
144 the procedure described by Megias-Perez et al. [23].

145 All samples were prepared in duplicate.

146 **2.4. GC-MS analysis**

147 A two-step derivatization procedure (oximation + silylation) of carbohydrates was carried out
148 prior to GC–MS analysis according to the method described elsewhere [16]. Briefly, 0.1 mL
149 of 70% methanolic solution of phenyl- β -D-glucoside (1 mg mL⁻¹; internal standard) was
150 added to 2 mL of cocoa sample extracts and to 0.5 mL of solutions of carbohydrate standards
151 (1 mg mL⁻¹ in methanol: water 70:30, v/v). Thereafter, samples were evaporated under

vacuum and treated with 350 μL of 2.5% hydroxylamine chloride in pyridine at 75 $^{\circ}\text{C}$ for 30 min. Then, 350 μL of hexamethyldisilazane (HMDS) and 35 μL of trifluoroacetic acid (TFA) (both from Sigma Aldrich) were added and the solution was kept at 45 $^{\circ}\text{C}$ for 30 min. Samples were centrifuged at 7000 g at 5 $^{\circ}\text{C}$ for 5 min. The derivatization procedure employed converts carbohydrates to oximes prior to trimethylsilylation, thereby reducing the number of chromatographic peaks for each reducing carbohydrate to two (corresponding to *E*- and *Z*-oxime isomers), and one peak for the non-reducing carbohydrates. Thus, simpler chromatographic profiles are obtained to facilitate the identification of target compounds in complex mixtures [18].

Qualitative analysis of carbohydrates was carried out using a 7890A gas chromatograph coupled to a 5975C quadrupole mass detector (Agilent Technologies, Palo Alto, CA, USA) operating in EI mode at 70 eV. A high-temperature polycarborane–siloxane HT-5 capillary column (25 m $\times$ 0.22 mm i.d. $\times$ 0.1 μm film thickness; SGE, Ringwood, Australia) was used for the qualitative analysis. Helium at 1 mL min⁻¹ was used as carrier gas. The oven temperature was programmed as follows: 180 $^{\circ}\text{C}$ for 10 min, followed by 200 $^{\circ}\text{C}$ at a heating rate of 5 $^{\circ}\text{C}$ min⁻¹, afterwards 270 $^{\circ}\text{C}$ at a heating rate of 15 $^{\circ}\text{C}$ min⁻¹, then 290 $^{\circ}\text{C}$ at 1 $^{\circ}\text{C}$ min⁻¹, 360 $^{\circ}\text{C}$ at 5 $^{\circ}\text{C}$ min⁻¹ and held for 30 min. The transfer line was set at 280 $^{\circ}\text{C}$ and the ionization source at 230 $^{\circ}\text{C}$. Subsequently, 1 μL of sample was injected in the split mode (split ratio of 1:10) at 300 $^{\circ}\text{C}$. Data acquisition was performed using a HPChem Station software (Agilent Technologies). Samples were analyzed in duplicate.

Linear retention indices (I^T) were calculated from the retention times of LMWC trimethylsilyl oxime (TMSO) derivatives and those of suitable n-alkanes (from C₁₇ to C₄₀) as described by Messadi *et al.* (1990) [24].

Carbohydrates were identified by comparison of experimental linear retention indices (I^T) and mass spectra with the available standards. Compounds for which commercial standards were

177 not available were tentatively identified on the basis of their mass spectral information and
178 data from the literature [14, 15, 25].

179 **2.5. HILIC-ESI-TOF MS analysis.**

180 For HILIC-ESI-TOF MS analysis, 1 mL of the sample was filtered through a CHROMAFIL
181 Xtra PTFE-45/25 filter (Macherey-Nagel, (Macherey Nagel, Düren, Germany) and 10 µL of
182 an internal standard solution of 1 mg mL⁻¹ of Asp-Phe methyl ester were added.

183 An Agilent 1100 Series HPLC (Agilent Technologies, Karlsruhe, Germany) was used to
184 perform the chromatographic analysis. A BEH X-Bridge column, with a trifunctionally-
185 bonded amide phase (Waters Company, USA) and the following characteristics: 150 mm ×
186 3.0 mm; 3.5 µm particle size and 135 Å pore size was chosen to perform the chromatographic
187 analysis. Water (solvent A) and acetonitrile (solvent B) with 0.05% ammonium hydroxide
188 were used as mobile phase. Injection volume was set to 3 µL. A flow rate of 0.4 mL min⁻¹ was
189 used.

190 A microTOF mass spectrometer fitted with an ESI ion source (Bruker Daltonics HCT Ultra,
191 Bremen, Germany) operating in positive ion mode in the range of 50-1200 *m/z* was used to
192 perform the identification of the molecular formula of the compounds and the quantitative
193 analysis. Internal calibration of the instrument was carried out by the injection of 0.1 M
194 sodium formate solution before starting the sequence run. Additionally, sodium formate
195 solution was injected automatically through a six-port valve prior to each chromatographic
196 run to perform a posterior calibration. The ESI source parameters were adjusted as follows:
197 spray voltage, 4.5 kV; drying gas (N₂, 99.5% purity); temperature = 220 °C; drying gas flow,
198 12 L min⁻¹; nebulizer (N₂, 99.5% purity) pressure, 1.6 bar. Data acquisition was performed
199 using HyStar 3.2 software (Bruker, Bremen, Germany).

200 Optimization of the chromatographic method was carried out considering different parameters
201 such as chromatographic resolution (R_s), retention time (t_R) and peak width at half height (w_h)
202 for the target compounds. The chromatographic resolution was calculated as $2(t_{R2} - t_{R1})/(w_{b1}$
203 $+ w_{b2})$, where t_{R1} and t_{R2} refer to retention times of two consecutive eluting carbohydrates
204 with the same m/z value and w_b is the peak width at base of each of them. Different binary
205 gradients were evaluated (Table 1). The optimal conditions were achieved using the following
206 gradient: 0–5 min, 17.5 % A; 5–37 min, 17.5–40% A; 37–50 min 17.5% A. The effect of
207 column temperature was also studied (25–55°C).

208 The identification of carbohydrates in cocoa beans was performed by comparing the retention
209 time and exact mass data obtained for each carbohydrate operating in positive ion mode with
210 those of commercial standards. The mass error detected was below 5 ppm.

211 For the quantitative analysis, the areas of the Extracted Ion Chromatogram (EIC) of the
212 sodium adduct $[M+Na]^+$ of *scyllo*-inositol, 1-kestose and galactinol (m/z 203.05, 527.15,
213 365.10 respectively) of standards and samples were normalized with respect to the area of the
214 internal standard. Calibration curves were calculated using the normalized areas of each
215 standard. The calibration curves and linear range of concentration of each standard are shown
216 in Table 2. Results were expressed as $\mu\text{g g}^{-1}$ DM cocoa bean.

217 Matrix effect for each carbohydrate was evaluated as the average of the recovery obtained
218 after the addition of two different amounts of the standard to the carbohydrate extract of cocoa
219 bean. The recovery for each amount of standard was performed in triplicate.

220 The chromatographic precision of the method was measured on the basis of intra-day
221 precision and inter-day precision. Three different samples were selected for the determination
222 of intra-day precision, calculated as the average of the relative standard deviations (RSD, %)
223 in 5 independent measurements from each sample for the concentrations of *scyllo*-inositol, 1-
224 kestose and galactinol. Inter-day precision was determined as the average of RSD of the

225 concentrations of *scyllo*-inositol, 1-kestose and galactinol of three different samples,
226 measured on three different days each.

227 The evaluation of the reproducibility of the entire method (extraction of LMWC and
228 chromatographic separation) was based on the average RSD values of each compound
229 measured on the 35 samples performed in duplicate of sample preparation.

230 Signal to noise ratio (S/N) of three and ten was used as criteria to determine the limit of
231 detection (LOD) and quantification (LOQ).

232 **2.6. Statistical analysis.**

233 GraphPad Prism 7.0 software (San Diego, California, USA) was used for statistical analysis.
234 A one-way analysis of variance (ANOVA) followed by a Fisher test as a post hoc comparison
235 of means was used to determine significant differences ($P < 0.05$) in *scyllo*-inositol, 1-kestose
236 and galactinol content among cocoa beans collected from different countries. A t-test was
237 applied to evaluate the possible differences in the content of each carbohydrate per
238 fermentation status (unfermented versus fermented).

239 **3. RESULTS AND DISCUSSION**

240 **3.1. Qualitative analysis of cocoa bean LMWC**

241 As mentioned above, GC-MS was selected as a powerful tool for the characterization of
242 LMWC of cocoa bean extracts and therefore, complement the structural information
243 previously reported using HILIC-ESI-TOF MS and HILIC-ESI-MSⁿ [22].

244 Different LMWC were detected, including mono-, di-, tri-, tetrasaccharides and cyclitols.
245 Figure 1 shows the GC-MS profile of TMSO derivatives of carbohydrates obtained from an
246 unfermented cocoa bean extract (see Table 3 for peak identification, experimental and
247 literature reported I^T values for the different compounds).

248 The monosaccharide profile previously reported in the literature [22, 26-28], including
249 fructose, glucose and galactose, was corroborated (peaks 1 to 5 in Figure 1) by the
250 comparison of the experimental retention times and mass spectra with the corresponding
251 standards. Peaks 6 and 7 (I^T of 1957 and 2042, respectively) showed typical spectra of free
252 inositols, with a pair of characteristic ion fragments at m/z 305 and 318, which were similar to
253 the intensity of the pair of m/z 191 and 217 [29]. These compounds were identified as *scyllo*-
254 inositol (peak 6) and *myo*-inositol (peak 7). It must be pointed out that this is the first time that
255 *scyllo*-inositol is reported in cocoa beans.

256 Figure 1A shows different peaks eluting in the disaccharide zone. Sucrose and melibiose,
257 which have been previously reported in cocoa beans [22, 26] were identified (peaks 8, 18 and
258 20 respectively). Different peaks with mass spectra compatible with disaccharides were also
259 detected (see Figure 2). Peaks 12 and 13 (I^T 2691 and 2711 respectively) were identified as
260 maltose isomers *E* and *Z*. Peaks 9 (I^T 2566) and 11 (I^T 2675) showed a mass spectrum similar
261 to sucrose, with m/z fragments at 437 and 451. The high intensity of the m/z 217 fragment
262 indicated the presence of fructose as the non-reducing unit of the disaccharide [14]. Moreover,
263 m/z fragments at 361 (characteristic of glycosidic linkages) and at 538 (characteristic of the
264 whole oxime chain) were also observed. Thus, these compounds were tentatively assigned as
265 fructosyl-glucoses. Peak 10 mass spectrum showed ions at m/z 307, (characteristic of
266 disaccharides with a reducing ketose substituted in C1 or C3) and m/z 437 (characteristic of
267 ketohexoses in both pyranose and furanose forms, free and monosubstituted) which is
268 compatible with a fructosyl-fructose [30]. Due to the absence of standards for these
269 compounds, these identifications could only be considered as tentative.

270 Mass spectrum of peak 19 was characterized by the triplet m/z ions 191/204/217,
271 (characteristic of silylated pyranose rings), 305 and 318 (typical of cyclitols) and a low
272 abundance of m/z ion 361 (related to glycosidic linkages). This fragmentation pattern has been

273 described to be typical of cyclitol glycosides [16]. Considering that *myo*-inositol was the most
274 abundant free inositol detected in cocoa beans, this compound could be a glycosyl-*myo*-
275 inositol and it was identified as galactinol (*O*- α -D-galactopyranosyl-(1 $\rightarrow$ 3)-D-*myo*-inositol)
276 by comparing its retention time and mass spectrum with that of its standard. The presence of
277 galactinol has been previously described in cocoa beans by Wang et al. [31], however, this
278 compound has been scarcely studied in this matrix.

279 Peak 22 could be tentatively assigned as diglycosyl-glycerol due to the presence of specific
280 *m/z* ion at 337 [32].

281 The presence of raffinose family oligosaccharides (RFOs) such as raffinose (peak 23) and
282 stachyose (peak 28), previously reported in the literature [22, 26, 27], was confirmed in cocoa
283 bean extracts by comparison with the corresponding standards. Unknown peaks were also
284 detected in the trisaccharide eluting zone (Figure 1, B). Peak 25 could be identified as 1-
285 kestose, while peak 24 and 26 were identified as 6-kestose and planteose respectively by
286 comparing their I^T values and spectra with data reported in the bibliography [14]. Peak 27 (I^T
287 3307) showed a mass spectrum similar to theandrose (α -Glc-(1 $\rightarrow$ 6)- α -Glc-(1 $\rightarrow$ 2)- β -Fru) but
288 with lower retention time, thus this peak was tentatively identified as glucosyl-sucrose (Glc-
289 (X $\rightarrow$ X)-Glc-(1 $\rightarrow$ 2)- β -Fru).

290 Overall, although the metabolic functions of the LMWC have extensively been studied in
291 higher plants [33, 34], few studies can be found regarding the role of the LMWC in *T. cacao*.

292 From a nutritional point of view, the presence of inositols (*scyllo*-inositol and galactinol) and
293 fructooligosaccharides (FOS), as 1-kestose or 6-kestose in cocoa beans, is noteworthy due to
294 their nutraceutical properties. For example, *scyllo*-inositol has been reported as an inhibitor of
295 the amyloid plaque formation in Alzheimer disease [35], while FOS have been described as
296 prebiotics [36].

297 3.2. Quantitative analysis of cocoa bean LMWC

298 Once LMWC composition of unfermented cocoa beans was identified employing GC-MS, to
299 evaluate their potential utility as indicators of cocoa bean origin or fermentation status,
300 quantitative analysis of those carbohydrates scarcely studied (galactinol) or reported for the
301 first time in cocoa bean (*scyllo*-inositol and 1-kestose) were carried out using HILIC-ESI-
302 TOF MS. This technique does not require previous derivatization of the sample and it could
303 be advantageous for routine analysis. However, a previous optimization of the method was
304 required to avoid overlapping of the peaks.

305 3.2.1. Optimization of HILIC method

306 An unfermented cocoa bean extract and carbohydrate standards were used to optimize the
307 HILIC-ESI-TOF MS method. The selection of the optimal chromatographic conditions was
308 based on the R_s , t_R and w_h values determined for the sodiated molecular ion adducts of the
309 target compounds (see Table 1).

310 Different gradient conditions and column temperatures were used to evaluate the
311 chromatographic separation of *scyllo*-inositol, 1-kestose and galactinol from other interferent
312 carbohydrates. In all the conditions, 50 min was considered as an appropriate time for the
313 chromatographic analysis.

314 The acetonitrile mobile phase content was first evaluated at a constant temperature of 40°C.
315 Table 1 shows the different R_s values calculated for the different chromatographic conditions
316 tested. As it can be observed in Table 1, the chromatographic conditions using high water
317 content with 0.05% ammonium hydroxide (26% in 0-5 min, 26–40% in 5-37 min and 26% in
318 37-50 min) resulted in the lowest resolution between *myo*- and *scyllo*-inositol, with an R_s
319 value of 0.25. A higher resolution between both inositols was achieved as the initial
320 acetonitrile content increased. The same tendency was observed for the R_s values considered

321 for the other compounds. Among the different chromatographic conditions tested, the best
322 baseline separation of the target carbohydrates was achieved under the following water
323 gradient: 17.5% in 0-5 min, 17.5–40% in 5-37 min and 17.5% in 37-50 min. In general,
324 narrow peaks were obtained for all the conditions tested (0.2-0.5 min).

325 The effect of the temperature (25, 40 and 55°C) was also evaluated using the gradient
326 mentioned above. According to Table 1, the use of higher temperatures (55°C) did not
327 enhance the R_s and w_h values obtained at 40°C for inositols. However, decreased R_s values
328 were observed at room temperature (25°C). In general, no pronounced differences in the
329 resolution values and peak widths of the other target compounds were observed with the
330 different temperatures. Thus, 40°C was selected as the optimum temperature to perform the
331 chromatographic separation. Figure 3 shows the profile of EIC corresponding to cocoa bean
332 monosaccharides ($[M+Na]^+$, m/z 203.05), disaccharides ($[M+Na]^+$, m/z 365.10) and
333 trisaccharides ($[M+Na]^+$, m/z 527.15) of an unfermented cocoa bean under optimized
334 chromatographic conditions. Under these conditions, t_R of the target carbohydrates was 17.6
335 min for *scyllo*-inositol, 22.6 min for 1-kestose and 26.3 min for galactinol.

336 In comparison with other methods employed in LMWC analysis of cocoa beans previously
337 reported, which employ for sample preparation tedious procedures such as solid phase
338 extraction (SPE), ion exchange or ultrafiltration followed by purification with Sep-pak C18
339 cartridge [22, 27, 37], the proposed method requires a minimum sample preparation, offers a
340 better resolution between inositols (*myo*-, *scyllo*-inositol) and would allow the quantification
341 of an elevated number of LMWC in cocoa beans.

342 3.2.2. Analytical parameters

343 The suitability of the optimized HILIC-ESI-TOF MS method for the quantitation of the target
344 carbohydrates was evaluated. Using reference standards of *scyllo*-inositol, 1-kestose and
345 galactinol, a calibration curve was obtained for each carbohydrate with good Pearson R^2

346 values (higher than 0.98), confirming the linearity of the calibration. After that, the influence
347 of the matrix effect on the quantitation of target compounds was evaluated. The average
348 recovery values of 105.1 %, 102.0%, 104.6 % for *scyllo*-inositol, 1-kestose and galactinol
349 respectively confirmed the absence of matrix effect.

350 Table 2 shows the analytical parameters of the chromatographic analytical method. The
351 lowest LOQ and LOD values were determined for 1-kestose (0.25 and 0.08 $\mu\text{g mL}^{-1}$,
352 respectively) and the highest values were observed for galactinol (0.75 and 0.25 $\mu\text{g mL}^{-1}$,
353 respectively). Regarding repeatability of the method, good values of intraday precision (range
354 4.98 - 7.53%) and interday precision (range 4.83 - 5.81%) were obtained for the target
355 carbohydrates under study.

356 The reproducibility of the whole method (extraction of carbohydrates from the cocoa bean
357 and chromatographic separation) was evaluated based on the average RSD values of each
358 analyte measured on the 35 samples performed in duplicate of sample preparation. The
359 average RSD values for *scyllo*-inositol, 1-kestose and galactinol were 5.41%, 6.07%, 6.37%
360 respectively. These data suggested lower variability related to the extraction method and good
361 reproducibility of the measurements.

362 Overall, all parameters indicated the suitability of the proposed HILIC-ESI-TOF MS method
363 for the quantification of *scyllo*-inositol, 1-kestose and galactinol in cocoa beans.

364 **3.3. Analysis of the content of *scyllo*-inositol, 1-kestose and galactinol in cocoa beans.**

365 Table 4 shows the different values of *scyllo*-inositol, 1-kestose and galactinol content from 21
366 unfermented beans and 14 spontaneously fermented cocoa beans from five different origins.
367 Supplementary information (tables S2, S3 and S4) shows the individual *scyllo*-inositol, 1-
368 kestose and galactinol values of the samples analyzed.

369 **3.3.1. *scyllo*-Inositol.**

370 The content of *scyllo*-inositol was in the range of traces - 504.9 $\mu\text{g g}^{-1}$ cocoa DM in
 371 unfermented and 15.5 - 491.9 $\mu\text{g g}^{-1}$ cocoa DM in spontaneously fermented beans. The values
 372 of *scyllo*-inositol determined in cocoa beans were lower in comparison to the content reported
 373 in other food matrices, such as Apiaceae family (values close to 2 mg g^{-1} dry weight) [38].

374 In unfermented beans, one-way ANOVA test showed significant differences among bean
 375 origin ($p\text{-value} = 0.0450$). Fisher test revealed significant differences in the content of *scyllo*-
 376 inositol ($p < 0.05$) between the following pair of countries: Ecuador-Ivory Coast, Ecuador-
 377 Brazil and Ecuador-Indonesia (individual p -values of these comparisons are shown in Figure
 378 1S, supplementary information). Inositols are a family of compounds crucial for development
 379 and signalling in plants. They function mainly as either metabolic mediators or participating
 380 in various signalling pathways in response to environmental conditions (stress, hormones, and
 381 nutrients) through transcriptional regulation of the stimuli-responsive genes [39]. Thus, the
 382 variability observed in unfermented beans could be attributed to the different environmental
 383 conditions, such as soil pH in the different farming locations of the samples under study [40].

384 Regarding fermented cocoa beans, one-way ANOVA test identified significant differences
 385 among countries ($p = 0.0058$). Fisher test revealed significant differences in the content ($p <$
 386 0.05) between the following pair of countries: Ecuador-Ivory Coast, Ecuador-Brazil, Ecuador-
 387 Indonesia, Malaysia-Ivory Coast, Malaysia-Brazil, Malaysia-Indonesia (individual p values of
 388 these comparisons are shown in Figure 3S, supplementary information).

389 Figure 2S (supplementary information) shows the comparison within each country of the
 390 *scyllo*-inositol concentration in the fermented and unfermented samples. Differences in the
 391 content from Brazil samples were detected using t-test.

392 Further studies need to be carried out to understand the variability observed in *scyllo*-inositol
 393 content according to the origin and to identify the role of *scyllo*-inositol during fermentation.

394 **3.3.2. 1-Kestose**

395 The content of 1-kestose determined in unfermented and fermented beans was in the range of
396 36.1 - 133.5 and traces - 115.5 $\mu\text{g g}^{-1}$ cocoa DM respectively. These values were lower in
397 comparison to the content determined in other food matrices such as green and ripe fruits of
398 ackee, carambola and jun plum (range of 0.1 - 0.5 g/100 g dry weight) [41] or bananas from
399 different cultivars (values range from 297 - 1630 $\mu\text{g g}^{-1}$ DM) [42].

400 One-way ANOVA analysis was not able to detect significant differences in 1-kestose content
401 among the different countries of origin, neither for the unfermented nor the fermented
402 samples.

403 The comparison within each country of 1-kestose concentration in the pair fermented and
404 unfermented samples is shown in the Figure 4S (supplementary information). The t-test
405 analysis showed significant differences in Ivory Coast, Ecuador and Indonesia samples.

406 To date, the mechanism of LMWC degradation during the fermentation of cocoa beans is
407 unknown. The LMWC degradation could be attributed to the elevated temperatures during the
408 fermentation and the diminution of pH in the bean as a consequence of the production of
409 organic acids by the microorganisms.

410 **3.3.3. Galactinol.**

411 The content of galactinol in unfermented beans was in the range of traces -1970.4 $\mu\text{g g}^{-1}$
412 cocoa DM. These values were in the same order of magnitude to the values described in other
413 seeds, e.g., lentils (0.12 % DM) [43]. No galactinol content was determined in spontaneously
414 fermented beans. Thus, this compound might be considered as a potential marker for the
415 fermentation status.

416 To the best of our knowledge, there are not previous studies focusing on the evaluation of
417 galactinol during the spontaneous fermentation of cocoa beans. As commented before, during

418 the fermentation process, a diminution of the pH in the beans consequence of the organic acid
419 production by microbial activities and the relative high temperatures (up to 50°C) could
420 produce the degradation of some compounds present in unfermented beans, which could
421 explain the absence of galactinol in the fermented bean samples.

422 One-way ANOVA analysis did not show any significant differences per country in
423 unfermented beans. The wide variability observed in the content of galactinol might be
424 attributed to the different ripening degree of the unfermented beans collected in different
425 origins [31].

426 **4. CONCLUSIONS**

427 In this study, the information reported about LMWC using GC-MS has been shown to be
428 complementary to the data previously reported employing HILIC-ESI-TOF MS and HILIC-
429 ESI-MSⁿ. Different carbohydrates such as *scyllo*-inositol, 1-kestose and 6-kestose have been
430 reported for the first time in unfermented cocoa beans. Moreover, tentative identifications of
431 other minor LMWC have been proposed.

432 Furthermore, the optimized and validated HILIC-ESI-TOF MS method allowed the separation
433 and quantitation in different cocoa bean samples of *scyllo*-inositol, 1-kestose and galactinol,
434 compounds interesting for their remarkable biological functions. The content of these LMWC
435 in cocoa beans from different locations and different fermentation status has been reported for
436 the first time in this manuscript. The results showed significant differences in *scyllo*-inositol
437 content among the different origins, independently of the fermentation status of the beans.
438 Also, the results indicated that galactinol content seems to be related with the different
439 fermentation status of the bean. However, further studies with a higher number of samples or
440 a detailed study of the changes during the fermentation would be necessary to confirm this
441 affirmation.

442 The proposed methodology has demonstrated to be highly sensitive, with minimum sample
443 preparation, which may be used for the quality control analysis of cocoa beans or other
444 matrixes containing inositols and oligosaccharides.

445 **5. ACKNOWLEDGEMENTS**

446 Authors would like to mention their gratitude to Sabur Badmos for proofreading of the
447 manuscript and Dr Gorka Ruiz de Garibay for the fruitful and valuable discussions on the
448 statistical analysis. Also, the authors would like to acknowledge to Anja Müller for her
449 assistance during the measurements of the samples as well as to Diana Sirbu and Britta
450 Behrends for performing the defatting process of all samples of the study.

451 This work was supported by Barry Callebaut (Belgium), by Ministerio de Economía, Industria
452 y Competitividad of Spain (project AGL2016-80475-R, AEI/FEDER, UE), by Comunidad de
453 Madrid (Spain) and European funding from FEDER program (S2013/ABI-3028
454 AVANSECAL-CM).

455

456 **Figure captions**

457 Figure 1. GC–MS profile of TMSO derivatives of carbohydrates of unfermented cocoa bean
458 extract. For peak identifications see Table 3.

459 Figure 2. Electron impact ionization mass spectra of TMSO derivatives of different unknown
460 carbohydrates

461 Figure 3. Extracted ion chromatograms obtained by HILIC-ESI-TOF MS of unfermented
462 cocoa bean from Malaysia corresponding to: A) 203.05 *m/z* ions, B) 365.10 *m/z* ions and C)
463 527.15 *m/z* ions. Peak numbers: 1) Fructose, 2) Glucose + galactose, 3) Sucrose, 4) *myo*-
464 Inositol, 5) *scyllo*-Inositol, 6) 1-Kestose, 7) Raffinose, 8) Galactinol, 9) Maltose + unknown
465 disaccharide (UD), 10) Melibiose, UT) Unknown trisaccharides.

466

Table 1. HILIC parameters: retention times (t_R), peak widths (w_h) and resolution (R_s) obtained under different chromatographic conditions. UT: unknown trisaccharide; UD: unknown disaccharide. R_s calculated: *myo*-Inositol – *scyllo*-Inositol; UT 1 – 1-Kestose; 1-Kestose – UT 2; UD 2 – Galactinol.

Method conditions (A % water)	T (°C)	<i>myo</i> -Inositol	<i>scyllo</i> -Inositol	UT 1	1-kestose	UT 2	UD 2	Galactinol	
0-5 min, 26%; 5-37 min 26-40%; 37-50 min 26%	40	t_R (min)	8.9	9.0	10.5	11.5	12.6	13.7	15.2
		w_h (min)	0.2	0.2	0.3	0.3	0.5	0.3	0.3
		R_s		0.3		1.3	1.3		2.1
0-5 min, 24%; 5-37 min 24-40%; 37-50 min 24%	40	t_R (min)	10.5	10.8	12.9	14.1	15.6	16.4	18.1
		w_h (min)	0.3	0.2	0.3	0.3	0.5	0.4	0.3
		R_s		0.6		2.0	2.1		2.3
0-5 min, 20%; 5-37 min 20-40%; 37-50 min 20%	40	t_R (min)	14.6	15.0	18.3	19.6	21.1	21.7	23.6
		w_h (min)	0.3	0.4	0.3	0.3	0.4	0.3	0.3
		R_s		0.6		2.2	2.1		2.9
0-5 min, 17.5%; 5-37 min 17.5-40%, 37-50 min 17.5%	40	t_R (min)	17.1	17.6	21.3	22.6	24.0	24.6	26.3
		w_h (min)	0.3	0.3	0.3	0.3	0.5	0.3	0.3
		R_s		0.8		2.4	2.3		2.8
0-5 min, 17.5%; 5-37 min 17.5-40%, 37-50 min 17.5%	25	t_R (min)	19.4	19.8	22.8	23.9	25.4	26.3	27.7
		w_h (min)	0.3	0.3	0.3	0.3	0.3	0.3	0.3
		R_s		0.7		1.7	2.1		2.0
0-5 min, 17.5%; 5-37 min 17.5-40%, 37-50 min 17.5%	55	t_R (min)	15.6	16.1	20.3	21.7	22.9	23.2	25.0
		w_h (min)	0.3	0.3	0.4	0.3	0.5	0.3	0.3
		R_s		0.8		2.3	2.0		3.0

Table 2. Analytical parameters of HPLC-ESI-TOF MS.

	Calibration curve	R ²	Linear working range (µg mL ⁻¹)	L.O.Q (µg mL ⁻¹)	L.O.D (µg mL ⁻¹)	Intra-day Precision (% RSD) (n*=3)	Inter-day Precision (% RSD) (n=3)
<i>scyllo</i> -Inositol	y = 0.0157x + 0.0222	0.9859	0.50 - 40	0.50	0.17	6.61	4.83
1-Kestose	y = 0.013x + 0.0125	0.9955	0.25 -25	0.25	0.08	7.53	5.43
Galactinol	y = 0.0101x + 0.0086	0.9942	0.75 - 40	0.75	0.25	4.98	5.81

n* = number of samples analyzed. For intra-day precision, each sample was injected five times on the same day. For inter-day precision samples were injected in three different days.

1 Table 3. Peak assignation and linear retention indices (I^T) of different carbohydrates detected
2 in unfermented cocoa bean extracts by GC–MS.

Peak number	Compound	Retention time (min)	Experimental I^T	Literature I^T
1	Fructose 1	4.963	1830	1846 [15]
2	Fructose 2	5.187	1844	1858 [15]
3	Galactose <i>E</i>	5.913	1887	1888 [15]
4	Glucose <i>E</i>	6.163	1901	1896 [15]
5	Glucose <i>Z</i> + Galactose <i>Z</i>	6.720	1923	1920 [15]
6	<i>Scyllo</i> -inositol	7.588	1957	1979 [25]
7	<i>Myo</i> -inositol	10.090	2042	2048 [15]
8	Sucrose	18.385	2508	2517 [15]
9	Fructosyl-glucose* + unknown	18.846	2566	-
10	Fructosyl-fructose*	19.124	2602	-
11	Fructosyl-glucose*	19.708	2675	-
12	Maltose <i>E</i>	19.836	2691	2697 [15]
13	Maltose <i>Z</i>	19.999	2711	2715 [15]
14	Unknown	20.162	2729	-
15	Unknown disaccharide	20.230	2737	-
16	Unknown	20.406	2757	-
17	Unknown	20.637	2784	-
18	Melibiose <i>E</i>	20.718	2793	-
19	Galactinol	21.274	2849	2874 [15]
20	Unknown + Melibiose <i>Z</i>	21.539	2875	-
21	Unknown	22.468	2958	-
22	Diglycosyl glycerol	22.766	2983	-
23	Raffinose	25.405	3166	3158[15]
24	6-Kestose	25.622	3179	3170 [14]
25	1-Kestose	25.818	3191	3198 [14]
26	Planteose	27.188	3265	3278 [14]
27	Unknown trisaccharide (Glc-(X→X)-Glc-(1→2)-β-Fru)	28.016	3307	-
28	Stachyose	44.843	3976	3976 [15]

3 * Peaks tentatively identified

4

5

6 Table 4. Mean ± standard deviation (Min-Max) values of each carbohydrate under study (*scyllo*-inositol, 1-kestose and galactinol) per country and status of
7 fermentation (U: unfermented beans, F: fermented beans).

Country group size	Mean ± SD (Min-Max) µg /g DM					
	<i>scyllo</i> -Inositol		1-kestose		Galactinol	
	Unfermented	Fermented	Unfermented	Fermented	Unfermented	Fermented
Ivory Coast U*: n=6 F: n=3	54.9 ± 85.3 (tr** – 219.0)	96.4 ± 80.5 (15.5 – 176.5)	101.2 ± 19.2 (73.2 – 127.9)	54.2 ± 33.1 (16.1 – 75.7)	1014.4 ± 658.0 (361.5 – 1970.4)	0.0 ± 0.0 (0.0 – 0.0)
Indonesia U: n=4 F: n=3	77.3 ± 56 (17.7 – 145.3)	140.6 ± 31.8 (108.2 – 171.7)	74.4 ± 41.7 (36.1 – 133.5)	48.7 ± 59.8 (tr – 115.5)	11.8 ± 3.1 (8.6 – 16.0)	0.0 ± 0.0 (0.0 – 0.0)
Malaysia U: n=3 F: n=3	190.1 ± 57.1 (131.5 – 245.6)	381.2 ± 83.4 (303.6 – 469.4)	88.3 ± 37.9 (48.2 – 123.6)	35.0 ± 36.6 (tr – 73.1)	383.1 ± 351.9 (163.6– 789.1)	0.0 ± 0.0 (0.0 – 0.0)
Brazil U: n=3 F: n=2	32.4 ± 16.3 (13.5 – 42.5)	77.1 ± 19.0 (63.6 – 90.5)	92.9 ± 18.3 (76.1 – 112.4)	14.9 ± 21.1 (tr – 29.9)	56.6 ± 75.1 (tr– 141.8)	0.0 ± 0.0 (0.0 – 0.0)
Ecuador U: n=5 F: n=3	235.2 ± 175.2 (57.1 – 504.9)	338.4 ± 136.4 (231.0 – 491.9)	85.0 ± 18.9 (57.0 – 108.4)	11.0 ± 19.0 (tr– 32.9)	597.2 ± 676.5 (128.7 – 1770.1)	0.0 ± 0.0 (0.0 – 0.0)

8 *U: unfermented; F: fermented; ** tr: traces

9

- 11 [1] J.E. Kongor, M. Hinnah, D.V. de Walle, E.O. Afoakwa, P. Boeckx, K. Dewettinck, Factors influencing
12 quality variation in cocoa (*Theobroma cacao*) bean flavour profile - a review, *Food Res. Int.*, 82 (2016)
13 44-52.
- 14 [2] A.M. Wickramasuriya, J.M. Dunwell, Cacao biotechnology: current status and future prospects,
15 *Plant Biotechnol. J.*, 16 (2018) 4-17.
- 16 [3] M.A. Patras, B.P. Miley, G. Vraricken, N. Kuhnert, Identification of novel cocoa flavonoids from
17 raw fermented cocoa beans by HPLC-MSn, *Food Res. Int.*, 63 (2014) 353-359.
- 18 [4] R.N. D'Souza, S. Grimbs, B. Behrends, H. Bernaert, M.S. Ullrich, N. Kuhnert, Origin-based
19 polyphenolic fingerprinting of *Theobroma cacao* in unfermented and fermented beans, *Food Res.*
20 *Int.*, 99 (2017) 550-559.
- 21 [5] I.M.D.V. Moreira, L.D.F. Vilela, C. Santos, N. Lima, R.F. Schwan, Volatile compounds and protein
22 profiles analyses of fermented cocoa beans and chocolates from different hybrids cultivated in Brazil,
23 *Food Res. Int.*, 109 (2018) 196-203.
- 24 [6] C.N. Quiroz-Reyes, V. Fogliano, Design cocoa processing towards healthy cocoa products: The role
25 of phenolics and melanoidins, *J. Funct. Foods*, 45 (2018) 480-490.
- 26 [7] C. Barnaba, T. Nardin, A. Pierotti, M. Malacarne, R. Larcher, Targeted and untargeted
27 characterisation of free and glycosylated simple phenols in cocoa beans using high resolution-
28 tandem mass spectrometry (Q-Orbitrap), *J. Chromatogr. A*, 1480 (2017) 41-49.
- 29 [8] D. Żyżelewicz, W. Krysiak, J. Oracz, D. Sosnowska, G. Budryn, E. Nebesny, The influence of the
30 roasting process conditions on the polyphenol content in cocoa beans, nibs and chocolates, *Food*
31 *Res. Int.*, 89 (2016) 918-929.
- 32 [9] F. Magagna, E. Liberto, S.E. Reichenbach, Q. Tao, A. Carretta, L. Cobelli, M. Giardina, C. Bicchi, C.
33 Cordero, Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from
34 thermal to differential-flow modulated comprehensive two dimensional gas chromatography, *J.*
35 *Chromatogr. A*, 1536 (2017) 122-136.
- 36 [10] A. Servent, R. Boulanger, F. Davrieux, M.N. Pinot, E. Tardan, N. Forestier-Chiron, C. Hue,
37 Assessment of cocoa (*Theobroma cacao* L.) butter content and composition throughout
38 fermentations, *Food Res. Int.*, 107 (2018) 675-682.
- 39 [11] R.N. D'Souza, A. Grimbs, S. Grimbs, B. Behrends, M. Corno, M.S. Ullrich, N. Kuhnert, Degradation
40 of cocoa proteins into oligopeptides during spontaneous fermentation of cocoa beans, *Food Res. Int.*,
41 109 (2018) 506-516.
- 42 [12] E.S. De Brito, N.H.P. García, M.I. Gallão, A.L. Cortelazzo, P.S. Fevereiro, M.R. Braga, Structural
43 and chemical changes in cocoa (*Theobroma cacao* L) during fermentation, drying and roasting, *J. Sci.*
44 *Food Agric.*, 81 (2001) 281-288.
- 45 [13] A.I.R.-M. M.L. Sanz, N. Corzo, I. Martínez-Castro Analysis of prebiotic oligosaccharides, in:
46 D.C.a.R.A. Rastall (Ed.) *Prebiotics and Probiotics Science and Technology*, Springer, 2009, pp. 465-534.
- 47 [14] A.I. Ruiz-Matute, M. Brokl, A.C. Soria, M.L. Sanz, I. Martínez-Castro, Gas chromatographic-mass
48 spectrometric characterisation of tri- and tetrasaccharides in honey, *Food Chem.*, 120 (2010) 637-
49 642.
- 50 [15] L. Ruiz-Aceituno, S. Rodríguez-Sánchez, A.I. Ruiz-Matute, L. Ramos, A.C. Soria, M.L. Sanz,
51 Optimisation of a biotechnological procedure for selective fractionation of bioactive inositols in
52 edible legume extracts, *J. Sci. Food Agric.*, 93 (2013) 2797-2803.
- 53 [16] L. Ruiz-Aceituno, C. Carrero-Carralero, A.I. Ruiz-Matute, L. Ramos, M.L. Sanz, I. Martínez-Castro,
54 Characterization of cyclitol glycosides by gas chromatography coupled to mass spectrometry, *J.*
55 *Chromatogr. A*, 1484 (2017) 58-64.
- 56 [17] S. Rodríguez-Sánchez, A.I. Ruiz-Matute, M.L. Sanz, A.C. Soria, Characterization of trimethylsilyl
57 ethers of iminosugars by gas chromatography-mass spectrometry, *J. Chromatogr. A*, 1372 (2014)
58 221-227.
- 59 [18] M. Brokl, A.C. Soria, I. Martínez-Castro, M.L. Sanz, A.I. Ruiz-Matute, Characterization of O-
60 trimethylsilyl oximes of trisaccharides by gas chromatography-mass spectrometry, *J. Chromatogr. A*,
61 1216 (2009) 4689-4692.
- 62 [19] A.I. Ruiz-Matute, O. Hernández-Hernández, S. Rodríguez-Sánchez, M.L. Sanz, I. Martínez-Castro,
63 Derivatization of carbohydrates for GC and GC-MS analyses, *J. Chromatogr. B: Anal. Technol. in the*
64 *Biomed. and Life Sci.*, 879 (2011) 1226-1240.
- 65 [20] A. Martín-Ortiz, J. Salcedo, D. Barile, A. Bunyatratchata, F.J. Moreno, I. Martín-García, A.
66 Clemente, M.L. Sanz, A.I. Ruiz-Matute, Characterization of goat colostrum oligosaccharides by nano-
67 liquid chromatography on chip quadrupole time-of-flight mass spectrometry and hydrophilic
68 interaction liquid chromatography-quadrupole mass spectrometry, *J. Chromatogr. A*, 1428 (2016)
69 143-153.
- 70 [21] S. Rodriguez-Sanchez, M.J. Garcia-Sarrio, J.E. Quintanilla-Lopez, A.C. Soria, M.L. Sanz, Analysis of
71 iminosugars and other low molecular weight carbohydrates in *Aglaonema* sp extracts by hydrophilic
72 interaction liquid chromatography coupled to mass spectrometry, *J. Chromatogr. A*, 1423 (2015) 104-
73 110.
- 74 [22] R. Megías-Pérez, S. Grimbs, R.N. D'Souza, H. Bernaert, N. Kuhnert, Profiling, quantification and
75 classification of cocoa beans based on chemometric analysis of carbohydrates using hydrophilic
76 interaction liquid chromatography coupled to mass spectrometry, *Food Chem.*, 258 (2018) 284-294.
- 77 [23] R. Megias-Perez, J. Gamboa-Santos, A.C. Soria, M. Villamiel, A. Montilla, Survey of quality
78 indicators in commercial dehydrated fruits, *Food Chem.*, 150 (2014) 41-48.

79 [24] D. Messadi, F. Helaimia, S. Ali-Mokhnache, M. Boumahraz, Accurate determination of retention
80 indices in programmed temperature gas chromatography, *Chromatogr.*, 29 (1990) 429-434.

81 [25] C. Carrero-Carralero, D. Mansukhani, A.I. Ruiz-Matute, I. Martínez-Castro, L. Ramos, M.L. Sanz,
82 Extraction and characterization of low molecular weight bioactive carbohydrates from mung bean
83 (*Vigna radiata*), *Food Chem.*, 266 (2018) 146-154.

84 [26] J. Cerbulis, Carbohydrates in cacao beans .2. Sugars in Caracas cacao beans, *Arch. Biochem.*
85 *Biophys.*, 58 (1955) 406-413.

86 [27] R.J. Redgwell, V. Trovato, D. Curti, Cocoa bean carbohydrates: roasting-induced changes and
87 polymer interactions, *Food Chem.*, 80 (2003) 511-516.

88 [28] G.A. Reineccius, P.G. Kenney, T.E. Kavanagh, D.A. Andersen, Identification and quantification of
89 free sugars in cocoa beans *J. Agric. Food Chem.*, 20 (1972) 199-202.

90 [29] S. Rodríguez-Sánchez, A.I. Ruiz-Matute, M.E. Alañón, M.S. Pérez-Coello, L.F. de Julio-Torres, R.
91 Morales, I. Martínez-Castro, Analysis of cyclitols in different *Quercus* species by gas chromatography-
92 mass spectrometry, *J. Sci. Food Agric.*, 90 (2010) 1735-1738.

93 [30] A.I. Ruiz-Matute, M. Weiss, D. Sammataro, J. Finely, M.L. Sanz, Carbohydrate composition of
94 high-fructose corn syrups (HFCS) used for bee feeding: Effect on honey composition, *J. Agric. Food*
95 *Chem.*, 58 (2010) 7317-7322.

96 [31] L. Wang, T. Nägele, H. Doerfler, L. Fragner, P. Chaturvedi, E. Nukarinen, A. Bellaire, W. Huber, J.
97 Weiszmman, D. Engelmeier, Z. Ramsak, K. Gruden, W. Weckwerth, System level analysis of cacao seed
98 ripening reveals a sequential interplay of primary and secondary metabolism leading to polyphenol
99 accumulation and preparation of stress resistance, *The Plant J.: for cell and mol. Biol.*, 87 (2016) 318-
100 332.

101 [32] A. Cardelle-Cobas, C. Martínez-Villaluenga, M.L. Sanz, A. Montilla, Gas chromatographic-mass
102 spectrometric analysis of galactosyl derivatives obtained by the action of two different β -
103 galactosidases, *Food Chem.*, 114 (2009) 1099-1105.

104 [33] S. Sengupta, S. Mukherjee, P. Basak, A.L. Majumder, Significance of galactinol and raffinose
105 family oligosaccharide synthesis in plants, *Front. Plant Sci.*, 6 (2015).

106 [34] W. Van den Ende, Multifunctional fructans and raffinose family oligosaccharides, *Front. Plant*
107 *Sci.*, 4 (2013).

108 [35] J. McLaurin, R. Golomb, A. Jurewicz, J.P. Antel, P.E. Fraser, Inositol stereoisomers stabilize an
109 oligomeric aggregate of alzheimer amyloid β peptide and inhibit A β -induced toxicity, *J. Biol. Chem.* ,
110 275 (2000) 18495-18502.

111 [36] D.A. Flores-Maltos, S.I. Mussatto, J.C. Contreras-Esquivel, R. Rodríguez-Herrera, J.A. Teixeira,
112 C.N. Aguilar, Biotechnological production and application of fructooligosaccharides, *Cri. Rev.*
113 *Biotechnol.*, 36 (2016) 259-267.

114 [37] M.M. Ardhana, G.H. Fleet, The microbial ecology of cocoa bean fermentations in Indonesia, *Int.*
115 *J. Food Microbiol.*, 86 (2003) 87-99.

116 [38] A.C. Soria, M.L. Sanz, M. Villamiel, Determination of minor carbohydrates in carrot (*Daucus*
117 *carota* L.) by GC-MS, *Food Chem.*, 114 (2009) 758-762.

118 [39] R. Valluru, W. Van den Ende, *Myo*-inositol and beyond - Emerging networks under stress, *Plant*
119 *Sci.*, 181 (2011) 387-400.

120 [40] N. Kumari, A. Grimbs, R.N. D'Souza, S.K. Verma, M. Corno, N. Kuhnert, M.S. Ullrich, Origin and
121 varietal based proteomic and peptidomic fingerprinting of *Theobroma cacao* in non-fermented and
122 fermented cocoa beans, *Food Res. Int.*, 111 (2018) 137-147.

123 [41] N. Benkeblia, M.G. Lopez, Saccharides and fructooligosaccharides composition of green and ripe
124 *Averrhoa carambola*, *Blighia sapida* and *Spondias dulcis* fruits, *Food Chem.*, 176 (2015) 314-318.

125 [42] R.G. Der Agopian, C.A. Soares, E. Purgatto, B.R. Cordenunsi, F.M. Lajolo, Identification of
126 fructooligosaccharides in different banana cultivars, *J. Agric. Food Chem.*, 56 (2008) 3305-3310.

127 [43] B. Quemener, J.M. Brillouet, Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and
128 white lupin, *Phytochem.*, 22 (1983) 1745-1751.

129

130

Figure 2.

Intensity

Supplementary information of “Analysis of minor low molecular weight carbohydrates in cocoa beans by chromatographic techniques coupled to mass spectrometry”

Contents

Table S1. Characteristics of the samples under study (country, fermentation status, season of collection).	3
Table S2. Summary of the <i>scyllo</i>-inositol content from the different samples analyzed	4
Table S3. Summary of the galactinol content from the different samples analyzed	5
Table S4. Summary of the 1-kestose content from the different samples analyzed	6
Figure 1S. Results of one-way ANOVA analysis followed by a Fisher test as a post hoc comparison of the means of <i>scyllo</i>-inositol content of unfermented beans from different origins. 7	
Figure 2S. Results of t-test and graphic representation of the quantities of <i>scyllo</i>-inositol between unfermented/fermented for each country.	8
Figure 3S. Results of one-way ANOVA analysis followed by a Fisher test as a post hoc comparison of the means of <i>scyllo</i>-inositol content of fermented beans from different origins.	9
Figure 4S. Results of t-test and graphic representation of the quantities of 1-kestose between unfermented/fermented for each country.	10

Table S1. Characteristics of the samples under study (country, fermentation status, season of collection).

Country	Fermentation status	Season collection
Ivory Coast	Unfermented	2014
Ivory Coast	Unfermented	2015
Ivory Coast	Unfermented	2015
Ivory Coast	Unfermented	2014
Ivory Coast	Unfermented	2015
Ivory Coast	Unfermented	2014
Ivory Coast	Fermented	2015
Ivory Coast	Fermented	2014
Ivory Coast	Fermented	2014
Indonesia	Unfermented	2014
Indonesia	Unfermented	2015
Indonesia	Unfermented	2015
Indonesia	Unfermented	2014
Indonesia	Fermented	2015
Indonesia	Fermented	2015
Indonesia	Fermented	2014
Malaysia	Unfermented	2014
Malaysia	Unfermented	2016
Malaysia	Unfermented	2014
Malaysia	Fermented	2014
Malaysia	Fermented	2014
Malaysia	Fermented	2016
Brazil	Unfermented	2015
Brazil	Unfermented	2016
Brazil	Unfermented	2016
Brazil	Fermented	2016
Brazil	Fermented	2015
Ecuador	Unfermented	2014
Ecuador	Unfermented	2015
Ecuador	Unfermented	2015
Ecuador	Unfermented	2015
Ecuador	Unfermented	2015
Ecuador	Fermented	2015
Ecuador	Fermented	2015
Ecuador	Fermented	2015

Table S2. Summary of the *scyllo*-inositol content from the different samples analyzed

		<i>scyllo</i> -Inositol ($\mu\text{g g}^{-1}$ DM)		
		Average	SD	RSD
Ivory Coast	Unfermented	219.01	2.47	1.13
		69.54	6.95	10.00
		traces	0.00	0.00
		40.94	0.83	2.04
		traces	0.00	0.00
		traces	0.00	0.00
	Fermented	97.32	6.69	6.87
		176.51	1.38	0.78
		15.48	0.87	5.59
Indonesia	Unfermented	97.63	0.85	0.87
		17.70	0.42	2.37
		48.67	2.37	4.87
		145.34	4.28	2.94
	Fermented	171.71	5.17	3.01
		141.76	13.48	9.51
		108.22	5.05	4.66
Malaysia	Unfermented	193.04	18.51	9.59
		245.62	23.43	9.54
		131.69	0.32	0.24
	Fermented	469.44	19.93	4.24
		303.63	24.10	7.94
		370.47	20.81	5.62
Brazil	Unfermented	42.51	1.23	2.90
		41.05	3.28	7.99
		13.54	0.96	7.07
	Fermented	90.48	5.52	6.10
		63.65	3.77	5.93
Ecuador	Unfermented	504.92	21.31	4.22
		57.07	3.40	5.95
		244.11	22.66	9.28
		266.26	24.91	9.35
		103.61	4.45	4.30
Ecuador	Fermented	231.04	13.86	6.00
		292.14	1.05	0.36
		491.87	15.67	3.18

Table S3. Summary of the galactinol content from the different samples analyzed

		Galactinol ($\mu\text{g g}^{-1}$ DM)		
		Average	SD	RSD
Ivory Coast	Unfermented	361.52	14.15	3.91
		525.08	47.95	9.13
		1083.57	91.94	8.48
		537.58	38.25	7.12
		1970.40	160.52	8.15
		1608.51	61.03	3.79
	Fermented	0.00	0.00	0.00
		0.00	0.00	0.00
		0.00	0.00	0.00
Indonesia	Unfermented	11.85	0.48	4.01
		8.59	0.69	7.99
		16.00	2.32	14.53
		10.88	0.85	7.78
	Fermented	0.00	0.00	0.00
		0.00	0.00	0.00
		0.00	0.00	0.00
Malaysia	Unfermented	789.05	29.22	3.70
		196.70	17.99	9.14
		163.65	8.38	5.12
	Fermented	0.00	0.00	0.00
		0.00	0.00	0.00
		0.00	0.00	0.00
Brazil	Unfermented	27.98	2.06	7.36
		141.81	7.39	5.21
		traces	0.00	0.00
	Fermented	0.00	0.00	0.00
		0.00	0.00	0.00
Ecuador	Unfermented	150.57	11.61	7.71
		430.15	15.92	3.70
		506.56	9.46	1.87
		128.74	8.81	6.84
		1770.11	31.92	1.80
	Fermented	0.00	0.00	0.00
		0.00	0.00	0.00
		0.00	0.00	0.00

Table S4. Summary of the 1-kestose content from the different samples analyzed

		1-kestose ($\mu\text{g g}^{-1}$ DM)		
		Average	SD	RSD
Ivory Coast	Unfermented	88.49	1.37	1.55
		127.91	5.80	4.53
		103.11	9.07	8.80
		114.82	6.88	5.99
		73.17	5.72	7.82
		99.56	3.98	4.00
	Fermented	70.76	6.04	8.53
		75.75	7.00	9.24
		16.05	1.55	9.64
		133.49	10.88	8.15
Indonesia	Unfermented	68.49	0.24	0.35
		59.60	5.36	9.00
		36.13	3.59	9.94
		115.47	4.55	3.94
	Fermented	30.61	1.66	5.41
		traces	0.00	0.00
		123.60	5.06	4.09
Malaysia	Unfermented	93.01	7.93	8.52
		48.22	4.45	9.23
		32.06	2.89	9.03
	Fermented	traces	0.00	0.00
		73.07	8.18	11.19
		112.35	2.92	2.60
Brazil	Unfermented	76.06	5.29	6.96
		90.27	1.24	1.37
		29.86	0.79	2.63
	Fermented	traces	0.00	0.00
		79.60	1.01	1.27
Ecuador	Unfermented	56.97	1.63	2.87
		87.42	4.78	5.47
		92.56	1.84	1.99
		108.36	9.81	9.06
		traces	0.00	0.00
Ecuador	Fermented	32.95	2.96	8.99
		traces	0.00	0.00
		traces	0.00	0.00

Figure 1S. Results of one-way ANOVA analysis followed by a Fisher test as a post hoc comparison of the means of *scyllo*-inositol content of unfermented beans from different origins.

Figure 2S. Results of t-test and graphic representation of the quantities of *scyllo*-inositol between unfermented/fermented for each country.

Figure 3S. Results of one-way ANOVA analysis followed by a Fisher test as a post hoc comparison of the means of *scyllo*-inositol content of fermented beans from different origins.

Figure 4S. Results of t-test and graphic representation of the quantities of 1-kestose between unfermented/fermented for each country.