



HAL
open science

New insights on spatial genetic structure and diversity of *Coffea canephora* (Rubiaceae) in Upper Guinea based on old herbaria

Jean-Pierre Labouisse, Philippe Cubry, Frédéric Austerlitz, Ronan Rivallan, Hong Anh Nguyen

► To cite this version:

Jean-Pierre Labouisse, Philippe Cubry, Frédéric Austerlitz, Ronan Rivallan, Hong Anh Nguyen. New insights on spatial genetic structure and diversity of *Coffea canephora* (Rubiaceae) in Upper Guinea based on old herbaria. *Plant Ecology and Evolution*, 2020, 153 (1), pp.82-100. 10.5091/plecevo.2020.1584 . hal-02546041v1

HAL Id: hal-02546041

<https://hal.science/hal-02546041v1>

Submitted on 12 Nov 2020 (v1), last revised 11 Mar 2022 (v2)

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 4.0 International License

1 Running title: Distribution and diversity of *Coffea canephora* in Upper Guinea

2

3 Title: **New insights on spatial genetic structure and diversity of *Coffea canephora***
4 **(Rubiaceae) in Upper Guinea based on old herbaria**

5

6 **Jean-Pierre Labouisse^{1,2,3,*}, Philippe Cubry⁴, Frédéric Austerlitz⁵, Ronan Rivallan^{1,2,3} &**
7 **Hong Anh Nguyen^{1,2,3}**

8

9 ¹CIRAD, UMR AGAP, FR-34398 Montpellier, France

10 ²AGAP, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

11 ³Postal address: CIRAD, UMR AGAP, Avenue Agropolis, TA A-108/03, FR-34398 Montpellier Cedex 5, France

12 ⁴DIADE, Université Montpellier, IRD, FR-34394 Montpellier, France

13 ⁵UMR 7206 Éco-Anthropologie, CNRS, MNHN, Université Paris Diderot, FR-75116 Paris, France

14

15 *Corresponding author: jean-pierre.labouisse@cirad.fr

16

17 CIRAD: Centre de coopération Internationale en Recherche Agronomique pour le Développement; AGAP: Amélioration
18 Génétique et Adaptation des Plantes tropicales et méditerranéennes; CNRS: Centre National de la Recherche Scientifique;
19 INRA: Institut National de la Recherche Agronomique; DIADE: DIversité - Adaptation - Développement des plantes; IRD:
20 Institut de Recherche pour le Développement

21 MNHN: Muséum National d'Histoire Naturelle; UMR: Unité Mixte de Recherche

22

23

24

25

26

27

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

ABSTRACT

Backgrounds and aims –

Previous studies showed that robusta coffee (*Coffea canephora* Pierre ex A.Froehner), one of the two cultivated coffee species worldwide, can be classified in two genetic groups: the Guinean group originating in Upper Guinea and the Congolese group in Lower Guinea and Congolia. Although *C. canephora* of the Guinean group is an important resource for genetic improvement of robusta coffee, its germplasm is under-represented in ex situ gene banks and its genetic diversity and population structure have not yet been investigated.

Methods – To overcome the limitations of living collections, we explored old herbarium specimens collected in Guinea and Côte d’Ivoire and conserved at the *Muséum National d’Histoire Naturelle*, Paris. First, we reviewed the history of collection missions in both countries and how the *C. canephora* herbaria from the *Muséum* were assembled. Then, using 23 nuclear microsatellite markers, factorial and model-based Bayesian analyses, we investigated the genetic diversity of 126 specimens and 36 controls, analysed their distribution among the Congolese and Guinean groups, and estimated admixture proportions for each individual.

Key results – For the first time, we detected population genetic structure within the Guinean group of *C. canephora*. The Guinean genotypes can be assigned to five sub-groups with distinct geographic distribution, especially in Guinea where two sub-groups (Maclaudii and Gamé) are characterized by a low level of admixture due to geographical isolation.

Conclusions –

We showed how combining a literature review and genetic data from old herbarium specimens can shed light on previous observations made by botanists and guide further actions to better preserve native coffee plants in forest remnants of West Africa.

Key words – *Coffea canephora*; Côte d’Ivoire; genetic diversity; Guinea; herbarium; population genetic structure; robusta coffee; Rubiaceae; SSR markers; West African forest.

59 Coffee is a popular beverage enjoyed by millions of consumers throughout the world and a
60 major export commodity, with about nine million tons of green coffee produced yearly (ICO
61 2018: 7). Among the 124 species of the genus *Coffea* L. (Davis et al. 2006; Davis 2011), only
62 two are widely cultivated: *C. arabica* L. (arabica coffee) and *C. canephora* Pierre ex
63 A.Froehner (robusta coffee). A third species is much more rarely in cultivation, *Coffea*
64 *liberica* W.Bull. ex Hiern (liberica coffee). While arabica remains the main traded species,
65 robusta's share has gradually increased since the beginning of the twentieth century and now
66 accounts for almost 40% of the world's coffee market (ICO 2018: 7). Robusta production has
67 also been boosted by the demand for instant coffee in emerging countries and by the wide
68 price gap between arabica and robusta (Wallengren 2017). While there are vast tropical
69 lowland reserves suitable for the production of *C. canephora*, it is likely that, in a few
70 decades, optimum cultivation requirements will become increasingly difficult to reach in
71 many arabica coffee producing regions due to global warming (Davis et al. 2012). Arabica
72 producers in South and Central America are already experiencing a devastating outbreak of
73 coffee leaf rust (CLR), caused by *Hemileia vastatrix* Berk. & Broome (Avelino et al. 2015).
74 *Coffea canephora* is more productive, less sensitive to high temperatures, and more resistant
75 to CLR than *C. arabica*. Although robusta beverage is less prized than arabica, there is a high
76 variability in quality traits within this species and the presence of consistent quantitative trait
77 loci provides breeders with promising tools to improve the cup quality of robusta coffee
78 (Leroy et al. 2011).

79 *Coffea canephora* is a perennial forest shrub with the largest natural distribution range for a
80 coffee tree species, stretching from Guinea to Uganda with a discontinuity in the Togo-Benin
81 region (Dahomey Gap) (Cubry et al. 2013b). The type specimen of the species *C. canephora*
82 was collected in Gabon in 1895 by the Reverend Théophile Klaine and sent to the botanist
83 Jean Baptiste Louis Pierre at the *Muséum National d'Histoire Naturelle* (MNHN) in Paris,
84 France under the name "*Café indigène des Ishira*" [specimen *Klaine 247 (P)*]. Pierre named
85 the species *Coffea canephora*; Pierre's name remained unpublished, however, and was
86 validated by Froehner (1897).

87 *Coffea canephora* is a diploid, allogamous, and self-incompatible species (Devreux et al.
88 1959). Its geographical range and mating system, as well as the impact from past climates,
89 have resulted in wide genetic variability within the species (Gomez et al. 2009). Another

90 consequence of its mating system is that the majority of cultivated *C. canephora* is still made
91 up of unselected populations obtained from open-pollinated seeds (Eskes & Leroy 2012:82).

92 A major breakthrough in our understanding of the genetics of *C. canephora* was made in the
93 1980s by Berthaud (1984) who identified two main diversity groups by analysing isozyme
94 polymorphism. Following the chorological system established by White (1979), Berthaud
95 proposed to classify the wild populations growing in Upper Guinea as “Guinean”, and those
96 from Lower Guinea and Congolia as “Congolese”. Further studies based on the use of DNA
97 markers (RFLP and SSR) validated this classification in two major genetic groups and led to
98 the recognition of five sub-groups (SG1, SG2, B, C, and UW) within the Congolese group
99 (Dussert et al 2003; Gomez et al. 2009; Musoli et al. 2009; Cubry et al. 2013b; Leroy et al.
100 2014).

101 In Upper Guinea, wild *C. canephora* was recorded for the first time in forests of Guinea and
102 Côte d’Ivoire by Auguste Chevalier, in 1905 and 1909 respectively (Chevalier 1905, 1920:
103 336). In both countries, the cultivation of robusta began in the second decade of the twentieth
104 century, mainly on the initiative of a few European planters (Portères 1939, 1962; Fréchou
105 1955; Cordier 1961). Although these planters may have used seeds and seedlings taken from
106 the forests close to their plantations, they mainly imported robusta seeds from Gabon and the
107 former Belgian Congo (now the Democratic Republic of Congo) to compensate for a shortage
108 of plants and, from 1921, for losses caused by attacks by the coffee berry borer *Hypothenemus*
109 *hampei* (Ferrari) (Portères 1939). Massive quantities of robusta seeds belonging to the
110 Congolese group were again introduced after the outbreak of tracheomycosis, a disease
111 caused by *Gibberella xylarioides* R.Heim & Saccas, which affected West African plantations
112 from 1948 onwards (Cordier 1961). Thus, *C. canephora* plants of the Congolese group
113 coexisted with local *C. canephora* in the plantations. In Côte d’Ivoire, Berthaud (1984)
114 analysed the genetic diversity of several clones previously selected for their outstanding
115 agronomic value. He demonstrated that natural hybridization occurred between the two
116 gene pools and showed that a majority of these selected clones were Guinean-Congolese F₁
117 hybrids. In the following years, a breeding programme based on reciprocal recurrent selection
118 was launched in Côte d’Ivoire to create hybrids from the germplasm of the two genetic groups
119 (Leroy et al. 1993, 1997).

120 The genetic resources conserved in ex situ gene banks were the basis of that breeding
121 programme. The largest gene bank of *C. canephora* is located at the *Centre National de*

122 *Recherche Agronomique* (CNRA) research station near Divo, Côte d'Ivoire, where about
123 1900 clonal accessions are conserved in the field. Many of these accessions have been
124 genetically characterized over the past 30 years using isozymes, RFLP, or SSR markers, e.g.
125 by Montagnon et al. (1992a, 1993), Dussert et al. (2003), Gomez et al. (2009), and Cubry et
126 al. (2013b). Only 205 accessions (about 10%) of *C. canephora* germplasm in the CNRA gene
127 bank belong to the Guinean group. In Guinea itself, a field gene bank of 101 accessions is
128 currently maintained by the *Institut de Recherche Agronomique de Guinée* (IRAG) at the
129 Sérédou research station near Macenta, Guinea Forest Region. It consists only of clones and
130 progeny of planting material imported in the past from the former Belgian Congo. Other
131 worldwide collections of robusta germplasm contain mainly accessions belonging to the
132 Congolese group (Anthony et al. 2007). Only two French institutes, the *Institut de Recherches*
133 *pour le Développement* (IRD) and the *Centre de Coopération Internationale en Recherche*
134 *Agronomique pour le Développement* (CIRAD) conserve a few accessions of the Guinean
135 group in their biological resource centres in Réunion and in French Guiana respectively.

136 Consequently, although *C. canephora* native to Upper Guinea may be an important source of
137 diversity for the genetic improvement of the species, its germplasm is underrepresented in ex
138 situ gene banks. While some populations from Côte d'Ivoire have been recently collected and
139 characterized, very little is known about germplasm from Guinea. Since the mid-20th century,
140 the natural forests of Côte d'Ivoire and Guinea have been highly threatened by deforestation
141 (Brou et al. 2000), reinforcing the need to assess the existence of potentially valuable and
142 largely untapped genetic resources for current and future breeding schemes.

143 To overcome the limitations of living collections and expand the geographical and temporal
144 range of data on *C. canephora* belonging to the Guinean group, we managed to get access to
145 old herbarium specimens. Recent advances in genomic technology have revealed the potential
146 of herbarium specimens for population genetics and phylogeographic studies (Lister et al.
147 2010). Genotyping studies on herbarium samples using SSR markers have been successfully
148 conducted on various species, e.g. on 100 year-old emmer wheat (Lister et al. 2008) and on
149 sweet potato samples dating back from the eighteenth century (Roullier et al. 2013). To carry
150 out our study, we used the resources of two herbaria located in the MNHN in Paris. The Paris
151 vascular plant herbarium, registered in the *Index Herbariorum* as P (Thiers 2019), conserves
152 91 herbarium sheets of *C. canephora* collected in Côte d'Ivoire and Guinea corresponding to
153 ca. 50 distinct plant specimens, including the oldest specimens collected in 1905 by Chevalier
154 (1905). Roland Portères was another prolific collector of African cultivated plants (Leroy

155 1974). The second herbarium (acronym PAT for *Paris, Agronomie Tropicale* according to the
156 *Index*), currently managed by the *Laboratoire d'Éco-Anthropologie*
157 (<http://ecoanthropologie.cnrs.fr>), conserves a thousand specimens of *Coffea* spp. collected by
158 Portères, a quarter of which are wild and cultivated *C. canephora* collected in Guinea and
159 Côte d'Ivoire between 1929 and 1961. Our study is restricted to these two countries, which
160 represent only a part of the Upper Guinea subcentre, but cover most of its west – east
161 extension. Large herbarium and germplasm collection programmes were conducted in these
162 former French colonies, where coffee cultivation was particularly developed with active
163 administration support to research and extension services.

164

165 The aim of our study is to provide an updated overview of the geographical distribution and
166 genetic diversity of *C. canephora* originating in Guinea and Côte d'Ivoire. First, we provided
167 a detailed overview about the history of collection missions in both countries and how the
168 coffee herbaria were assembled. Then, we investigated the pattern of genetic diversity of 126
169 herbarium specimens, supplemented by 36 genotypes used as controls, with a set of 23
170 polymorphic nuclear markers (SSRs) and a combination of principal coordinates analysis
171 (PCoA), neighbor-joining (NJ) analysis, and model-based Bayesian analysis. We discussed
172 the results in the light of historical documentation. Lastly, we gave an overview of
173 deforestation trends and of other factors that could hamper further collections of native *C.*
174 *canephora* in both countries.

175

176

REVIEW OF HERBARIUM DATA

177 **Guinea**

178 About three thousand kilometres from the type locality in Gabon, the first botanical record of
179 wild *C. canephora* of West Africa was made by Chevalier in 1905 in Guinea, on the southern
180 edge of the Fouta Djallon plateau (fig. 1) near the village of Bilima, 20 km west of Mamou
181 (Chevalier 1905). Chevalier named it *Coffea maclaudii* (as *Maclaudi*) after Charles Maclaud,
182 a colonial military doctor, who had reported the occurrence of coffee plants in the upper
183 valley of the Konkouré River a few years before (Maclaud 1899). With Octave Caille, a head
184 gardener of the *Jardin des Plantes* in Paris, Chevalier (1914) collected several samples from
185 these coffee trees and sent them to P [Apr.1905: *Chevalier* 12181, 12330 (type specimen),
186 12331 (specimen lost), 12332, 12333; May 1905: 12332bis; 8 Sep. 1905: 14893]. Caille made

187 an additional collection in June 1913. From his monograph (*Les Caféiers du Globe*) onwards,
188 Chevalier named it *C. canephora* var. *maclaudii* (Chevalier 1929: 83).

189 Another important contributor to the P herbarium was Pierre Barthe, an agronomist who was
190 head of the agricultural extension service of the *cercle* of Macenta (a *cercle* was the smallest
191 unit of administration in French Colonial Africa), Guinea Forest Region from 1934 to 1936
192 (Anon. 1934). In 1936, he collected coffee in forests and plantations in the region, established
193 a collection plot at the Macenta farm school and sent samples to P.

194 In 1939, the Sérédou experimental station for coffee and *Cinchona officinalis* L. was
195 established near the village of Sérédou in the Zياما massif, 35 km southeast of Macenta, with
196 Portères as director (Tourte 2005a: 185). Portères planted several collection plots mainly with
197 clones or seeds transferred from Côte d'Ivoire, but originally imported from the Lula and
198 Yangambi research stations of the *Institut National pour l'Étude Agronomique du Congo*
199 *belge* (INEAC). He also collected material from local plantations and from Barthe's
200 collection (Sérédou centre annual report 1958, unpublished report). Portères sent some
201 samples from his collections to PAT in 1958 and 1959.

202 In 1962, Portères published a review of *Coffea* species cultivated in Guinea, which remains
203 the most detailed and comprehensive work on the topic to date (Portères 1962). He classified
204 the *C. canephora* specimens that “grew spontaneously” in Guinea in three groups:

205 - (i) var. *maclaudii* A.Chev.: from Bilima near Mamou.

206 - (ii) cultivar Gamé (*sensu stricto*): from the village of Bambaradou, a few kilometers from
207 Macenta, and named after Gamé Guilavogui, the chief of the *canton* (a territorial subdivision
208 gathering several villages) of Kolibirima Toma in the early 1930s (JOGF 1935) who
209 promoted Gamé cultivation with the support of the colonial administration. Portères reported
210 that 50% of coffee area in the *cercle* of Macenta was planted with Gamé cultivar in the late
211 1950s. He considered this cultivar more productive than any other material introduced in
212 Guinea and its cup quality as “excellent”. In addition to Gamé s.s., he defined a “Gamé
213 geographical-racial complex” (or *Géo-Gamé*) as a group of coffee populations that he thought
214 native and growing over a vast area that extends beyond the border of the *cercle* of Macenta,
215 from Kissidougou (to the north-west) to Nzérékoré and Beyla (to the east).

216 - (iii) cultivar Gouecké: originally from the vicinity of Gouecké in the north of the *cercle* of
217 Nzérékoré.

218 When Guinea became independent in 1958, the Sérédou coffee collection contained coffee
219 trees belonging to the Congolese group and only eight clones of Gamé (Sérédou centre annual
220 report 1958, unpublished report). In 1990, IRAG research staff started to set up a new

221 collection of clonal material with remaining plants from Portères' collection plots and a few
222 clones collected from plantations. At the time of writing (June 2018), the IRAG gene bank
223 comprised 101 accessions of *C. canephora*, most of which are clones or progenies of the
224 Congolese group. No cv. Gamé or var. *maclaudii* genotypes remain.

225

226

227 Côte d'Ivoire

228 Chevalier made the first record of native *C. canephora* in Côte d'Ivoire in 1909, in the forests
229 near Assikasso in the Indénié Region, 45 km northeast of Abengourou, in the far east of the
230 country [specimens: 18 Dec. 1909, *Chevalier* 22589, 22590, and 22602 (P)]. Based on
231 morphological similarities with samples collected in Bilima (Guinea), more than 1000 km to
232 the west, he identified it as conspecific with his *Coffea maclaudii* A.Chev. from Guinea
233 (Chevalier 1920: 336). In 1912, Arsène Dellabonin, an employee of the Assikasso agricultural
234 station, collected similar populations in the same area, near the village of Kongodia, and sent
235 samples to P (Portères 1937b). This coffee was widely cultivated from 1920 in eastern Côte
236 d'Ivoire under the name of *Café Petit Indénié*, in comparison with the species *C. liberica*
237 found by Chevalier in the same area and called *Café Gros Indénié* because of the large size of
238 the bean (Portères 1939).

239 In 1912, while hunting near Koro, northeast of Touba (Western Côte d'Ivoire), the planter
240 Jules Landré found wild *C. canephora* in gallery forests along the Irama-ba River and two
241 tributaries of the Sassandra River: the FéréDougou-ba and the Boa-ba. This coffee, and other
242 populations gathered in the same area by Le Campion in 1913, were called *Café Landré* or
243 *Café Touba* (Portères 1937b). In 1915, it was introduced into the collection of the agricultural
244 station of Bingerville near Abidjan (Portères 1934) and later duplicated and sent to the station
245 of the *Institut d'Enseignement et de Recherches Tropicales* (IDERT) near Adiopodoumé,
246 from where Portères sampled it for PAT.

247 In 1929, Portères founded the agricultural station of Man, in western Côte d'Ivoire, and set up
248 collection plots and comparative trials of various species of *Coffea*, including native and
249 imported *C. canephora* (Poupart 1938; Tourte 2005a: 193). From 1929 to 1931, Portères
250 collected samples around Man, Touba, and Béoumi. In 1937, he published a first review of
251 *Coffea* spp. in Côte d'Ivoire with their geographical distribution and ecology, the maps of the
252 collection sites in the above-mentioned areas, and the main morphological characteristics of
253 the coffee plants (Portères 1937a, 1937b, 1937c).

254 Later, in 1958 and again in 1962, Portères and Louis Cordier, a geneticist at the *Centre de*
255 *Recherches Agronomiques* (CRA) in Bingerville, assembled the largest coffee herbarium
256 collection ever made in Côte d’Ivoire. These collections included coffee plants from various
257 origin: local or imported, from forests, smallholders’ plantations, estates, and agricultural
258 research stations. Duplicates were sent to PAT. There is no published inventory of this
259 herbarium but handwritten documents by Portères give detailed information on the location,
260 origin and morphological traits of the specimens.

261 Beginning in 1957, the coffee germplasm preserved in the first research stations (Bingerville,
262 Akandjé, Abengourou) and agricultural stations (Man, Gagnoa) was gradually transferred to a
263 new site, near the town of Divo (fig. 1), which became the main research station of the *Institut*
264 *Français du Café et du Cacao* (IFCC 1960; Tourte 2005b: 221). This station and its coffee
265 gene bank are now managed by the CNRA.

266 When Côte d’Ivoire became independent, Portères (1959) and Cordier (1961) traced the
267 history of coffee cultivation and made an inventory of the native and imported coffee material
268 in the country. On the basis of morphological traits, they classified the populations of *C.*
269 *canephora* into two groups: “Kouilou” and “Robusta”. On average, “Kouilou” seeds, flowers,
270 leaves, and the tree itself are smaller than those of “Robusta” (Portères 1959). Portères and
271 Cordier classified most of the populations of *C. canephora* native to Côte d’Ivoire (Petit
272 Indénié, Touba, Bandama, Tos, Kouibly, Agbo, Dianlé, etc.) as belonging to the “Kouilou”
273 group. Only the Ébobo population, which Portères thought originated near Aboisso in the
274 extreme southeast of Côte d’Ivoire, was classified as “Robusta”. In fact, there was already a
275 *Café du Kouilou* (or *Café Kouilou*) that was named after the Kouilou River in the Republic of
276 the Congo, along which it was discovered in the 1880s (Chevalier 1929: 85–86). Based on
277 similar morphological characteristics, Chevalier used the term Kouilou in 1909 to describe the
278 *C. maclaudii* that he had found in Guinea in 1905 (Chevalier 1909: 260). Kouilou is also
279 spelled *Kouillou*, *Quillou*, or *Kwilu*. By orthographic distortion, the term became *Conilon* in
280 Brazil and is still applied to the first *C. canephora* populations introduced at the beginning of
281 the twentieth century in that country (Ferrão 2007). The use of the terms Kouilou and Robusta
282 has caused some confusion in the classification of *C. canephora*, an allogamous species with
283 great variability in morphological traits, even within a single population. Berthaud’s genetic
284 works clarified the nomenclature. He found that: (i) all populations of “Robusta” (including
285 the Ébobo population), as well as the true *Café du Kouilou* originating in the Republic of the
286 Congo, belong to the Congolese group and (ii) all the “Kouilou” populations in Côte d’Ivoire
287 belong to the Guinean group (Berthaud 1984: 111).

288 Following Berthaud’s findings, major germplasm collecting efforts resumed from 1975,
289 mainly in Côte d’Ivoire (Berthaud 1984: 12; Le Pierrès et al. 1989; Couturon & Montagnon
290 1991). Twenty-one populations of *C. canephora* were collected from forest sites in Côte
291 d’Ivoire, and one population from the forest of Piné in Guinea. These represent the largest
292 part of the Guinean coffee germplasm currently preserved in the CNRA gene bank (Dussert et
293 al. 2003: 242). Among these populations, Cubry (2008) and Cubry et al. (2013b) genotyped
294 seven populations using microsatellite markers with fresh samples provided by the CNRA
295 gene bank. We used DNA of a few genotypes taken from these populations as controls for the
296 diversity study.

297

298

299

MATERIAL AND METHODS

300 **Material**

301 We genotyped 162 samples of which 126 were herbarium specimens and 36 were obtained
302 from previous works and used as controls (table 1). More details on the material are available
303 in supplementary file 1 (worksheets A1 to A4).

304 We focused on herbarium specimens reported to be “spontaneous”, or “wild”, or “local” by
305 the collectors. The collection dates range from 1905 to 1993. The collection points (red dots
306 in fig. 1) are rather regularly distributed along a north-east – south-west axis from Bilima near
307 Mamou, Guinea to Assikasso near Abengourou, Côte d’Ivoire, two locations 1000 km apart,
308 with a gap in the Faranah Region, Guinea. In total, 40 herbarium specimens were collected
309 from 15 different sites in Guinea and 86 herbarium specimens from 34 sites in Côte d’Ivoire.

310 As controls (purple triangles in fig. 1), we used DNA of *C. canephora* already extracted,
311 characterized in previous studies (Cubry et al. 2013a, 2013b; Leroy et al. 2014), and stored at
312 -20°C at CIRAD, Montpellier, France. We added leaves of a few clones maintained in the
313 CIRAD coffee field gene bank in French Guiana (information available at:

314 <http://florilege.arcad-project.org/fr/collections/collection-cafeiers-guyane>), collected in 2014,
315 and stored in silica gel. The material used as controls (36 individuals) can be briefly

316 characterized as follows:

317 Guinean group:

- 318 • Côte d’Ivoire (locations: Daloa, Divo, Fourougbankoro, Gohitafla, Ira, Kaloufla,
319 Logbonou, Mouniandougou, and Pelezi): wild material collected from the end of the
320 1970s to 1991 (Berthaud 1984:12, Le Pierrès et al. 1989, Couturon & Montagnon
321 1991).

- 322 • Guinea (location: Piné): population collected in 1987 in the forest of Piné on the
323 Guinea-Côte d'Ivoire border (Le Pierrès et al. 1987).

324 Congolese group (not shown on the map):

- 325 • INEAC: population originally imported from the INEAC Lula and Yangambi research
326 stations, Democratic Republic of the Congo. INEAC population was massively
327 imported into Côte d'Ivoire from 1935 on and to Guinea from 1952 on (Portères 1959,
328 1962). It belongs to the SG2 sub-group.
- 329 • Luki: genotypes collected in forests and plantations in the Mayumbe region,
330 Democratic Republic of the Congo. They belong to the SG1 sub-group.
- 331 • Niaouli: population mainly cultivated in Togo and Benin, but originating from the
332 Atlantic Coast of Gabon (Adibolo & Bertrand 1988). It was introduced into Côte
333 d'Ivoire in 1914 (Portères 1934), and later to Guinea (Portères 1962). Niaouli coffee
334 belongs to the SG1 sub-group.

335

336 **Methods**

337 **DNA extraction** – All the samples were processed at the *Grand Plateau Technique Régional*
338 (GPTR) facility in Montpellier, France (<http://www.gptr-lr-genotypage.com>). Genomic DNA
339 was extracted from approximately 20 mg of dried leaf tissue according to the MATAB
340 (Mixed Alkyl Trimethylammonium Bromide) protocol described by Risterucci et al. (2000).
341 The DNA concentration was estimated with a Fluoroskan Ascent Microplate Fluorimeter
342 (ThermoFisher Scientific, Waltham, Massachusetts, USA) with a bisbenzimidazole DNA
343 intercalator (Hoechst 33258) and by comparison with known standards of DNA.

344

345 **Microsatellite markers** – We used SSR markers defined from *C. canephora* and *C. arabica*
346 genomes and already described (Poncet et al. 2004, Leroy et al. 2005, Poncet et al. 2007,
347 Cubry et al. 2008, Cubry et al. 2013a). Because DNA is highly fragmented in herbarium
348 specimens, we selected short size loci (less than 200 bp) as recommended by Särkinen et al.
349 (2012) in order to increase the PCR success rates. Thirty-three markers were initially tested,
350 but 10 were removed from the final analysis due to a high level of missing data or low
351 genotyping quality (table 2).

352

353 **PCR amplification** – PCR reactions were performed as simplex experiments in a volume of
354 10µl containing 1µL of reaction buffer 10X (100 mM Tris-HCl, pH 9,0 ; 100 mM KCl 80
355 mM (NH₄)₂SO₄; X-100 1% Triton), 1 µL of 2 mM dNTP (Jena Bioscience GmbH, Jena,

356 Germany), 0.3 μL of 50 mM MgCl_2 , 0.08 μL of 10 μM forward primer with a M13 tail at the
357 5'-end, 0.1 μL of 10 μM reverse primer, 0.1 μL of fluorescently labelled M13-tail (6-FAM,
358 NED, VIC or PET from Applied Biosystems, Foster City, California, USA), 0.12 μL of
359 5U/ μL Taq DNA Polymerase (Taq'Ozyme OZYA001 from Ozyme, Montigny-le-
360 Bretonneux, France), 5 μL of template DNA (1 ng/ μL), and 2.3 μL of water. A touchdown
361 cycling program was used as follows: an initial denaturation at 94°C for 5 min, followed by
362 10 cycles at 94°C for 30 s, 55°C for 60 s (0.5°C decrease at each cycle), 72°C for 1 min,
363 followed by 25 cycles at 94°C for 45 s, 50°C for 1 min, 72°C for 1 min and a final extension
364 at 72°C for 30 min.

365 Fluorescently labelled PCR products were then organized in five pools for electrophoresis,
366 using respectively 2 μL of products labelled with 6-FAM, 2 μL of those with VIC, 2.5 μL of
367 those with NED and 3.5 μL of those with PET, and completed at 20 μl with high purity water.
368 Table 2 shows which fluorochrome was used to label each marker and how PCR pools were
369 composed. 2 μL of this solution was taken and added to 10 μL of Hi-Di formamide and 0.12
370 μL of GeneScan 600 LIZ size standard (Applied Biosystems). Migration of PCR products
371 was made on an ABI 3500xL Genetic Analyzer (Life Technologies, Carlsbad, California,
372 USA). Alleles were scored using GeneMapper v4.1 software (Applied Biosystems).

373

374 For some low-quality samples resulting in poor PCR amplification, we used a modified
375 version of this protocol. After extraction, genomic DNAs were purified with Agencourt
376 AMPure XP (Beckman Coulter, Brea, California, USA) magnetic beads with 1 volume of
377 DNA for 1.8 volume of beads, then standardized at 0.5 or 1 ng/ μl with high purity water.
378 Then, we performed PCR in 10 μl reaction volume using Qiagen Type-it Microsatellite PCR
379 Kits with 5 μL of PCR MasterMix (HotStarTaq Plus DNA polymerase, PCR buffer, dNTP
380 Mix), 2 μl of purified DNA (0.5-1 ng), 0.2 μL of primer mix (0.1 μL of 10 μM forward
381 primer with an M13 tail at the 5'-end, 0.1 μL of 10 μM reverse primer), 0.2 μL of
382 fluorescently labelled M13-tail (6-FAM, NED, VIC or PET from Applied Biosystems), and
383 2.7 μL of high purity water. The PCR amplification was conducted with an initial
384 denaturation at 95°C for 5 min, followed by 10 cycles of 95°C for 30 s, 55°C for 1 min 30s
385 (0.5°C decrease at each cycle), 72°C for 30 s, followed by 25 cycles at 95°C for 30 s, 50°C
386 for 1 min 30 s, 72°C for 30 s and a final extension at 60°C for 30 min. Dilution, migration of
387 PCR products, and allele scoring were made as previously described.

388

389 **Structure analysis** – To identify the pair-wise genetic relationships between the individual
390 genotypes, we computed a genetic dissimilarity matrix using simple matching index with
391 DARwin v.6 software (Perrier & Jacquemoud-Collet 2006). An overall representation of the
392 structure of genetic diversity was obtained by a factorial analysis (Principal Coordinates
393 Analysis, PCoA) using distance matrices, while genetic relationships between individuals
394 were assessed using the neighbor-joining (NJ) method (Saitou & Nei 1987), as implemented
395 in DARwin v.6.

396 In order to test for sample clustering and to estimate admixture proportions for each
397 individual, we used the model-based Bayesian approach implemented in STRUCTURE
398 v.2.3.4, (Pritchard et al. 2000). The parameters were set to a burn-in period of 100,000 with
399 200,000 iterations. We performed 20 independent runs for each K, K varying from 1 to 10.
400 We used the online servers CLUMPAK (Kopelman et al. 2015) and STRUCTURE
401 HARVESTER (Earl & vonHoldt 2012) to analyse the STRUCTURE outputs and the method
402 of Evanno et al. (2005) to identify the number of genetic clusters corresponding to the
403 uppermost hierarchical level of genetic partitioning between populations.

404 Firstly, we analysed the data of all the samples of both countries, assigned each sample to
405 Congolese or Guinean groups and estimated admixture rates. Then, we removed all the
406 Congolese samples as well as admixed ones (we chose a threshold of 95% for membership
407 coefficient) and performed the same analyses in order to reveal a possible structure within the
408 Guinean group.

409

410 **Diversity parameters** – We calculated several descriptors of genetic diversity: allele number
411 per marker, observed heterozygosity (H_o), and expected heterozygosity (H_e), and performed a
412 test of Hardy-Weinberg (HW) equilibrium, using the R package STRATAG (Archer et al.
413 2017). These statistics were computed for different sets of genotypes defined according to the
414 structure analysis. We used the R package HIERFSTAT (Goudet 2005) to calculate allelic
415 richness, corrected for the different sample size of populations, using a rarefaction method as
416 recommended by El Mousadik & Petit (1996). We also computed the number of private
417 alleles for the different sets of genotypes. Lastly, three different measures of population
418 differentiation (Weir's F_{st} , G_{st} , and Jost's D), as well as their significance, were calculated
419 between each genetic group or sub-groups using the R package STRATAG: Congolese group
420 vs. Guinean group and between all Guinean sub-groups as defined by the structure analysis.

421

422

RESULTS

Structure analysis

Whole sample – The genotyping data matrix of the 162 samples may be found in supplementary file 1 (worksheet B). The results of the factorial analyses of the SSR-based dissimilarity matrix are represented in fig. 2. The first axis enabled a clear separation of the Guinean and Congolese groups with 19.68% of the global inertia. The second axis (5.88% inertia) and the third axis (4.75% inertia) separated the SG1 and SG2 Congolese sub-groups. Most of the herbarium specimens were distributed in the Guinean group. However, some herbarium specimens were grouped with the Congolese controls, and others occupied an intermediate position between Guinean and Congolese groups, which allowed us to classify them as admixed genotypes. The NJ tree (appendix 2) confirmed that classification. Similarly, Bayesian analysis using STRUCTURE clearly differentiated two clusters corresponding to Guinean and Congolese groups (fig. 3 and supplementary file 3).

Examining these results in detail and by collecting country (supplementary files 2 & 3), we found that:

(i) in Guinea, eight herbarium specimens (numbers 11 to 18) collected in the villages around Kissidougou grouped together with the Congolese genotypes (SG1 or SG2 sub-groups) used as controls, while all other samples were assigned to the Guinean group (membership coefficient $Q > 95\%$);

(ii) in Côte d'Ivoire, three herbarium specimens from Ébobo (numbers 151, 152, and 153) collected by Portères near Aboisso grouped together with the Niaouli and Luki populations (SG1 Congolese sub-group);

(iii) one herbarium specimen (number 142), collected in 1930 near Abengourou and stated to be “spontaneous” by Chevalier, clearly belongs to the SG1 Congolese sub-group ($Q > 99.5\%$).

(iv) unlike the herbarium specimens collected in Guinea, some of those collected in Côte d'Ivoire are Guinean-Congolese admixed genotypes. The results of STRUCTURE program ($K = 2$) showed that 12 herbarium specimens have a membership coefficient that varies from $Q = 47.9\%$ (probable F_1 hybrid between Congolese and Guinean genotypes) to $Q = 95\%$. For example, Portères collected six plants “in a plantation established from local seeds and seedlings collected in the forest of Tos” near Daloa (Centre West of Côte d'Ivoire). Among them, we found three pure Congolese ($Q > 99\%$ for numbers 101, 102, and 103), one putative F_1 hybrid ($Q = 52\%$ for number 104), and two admixed ($Q = 10.8\%$ and 35.7% for numbers 99 and 100 respectively).

457

458 **Within Guinean group** – Once we removed the Congolese, putative Guinean-Congolese F₁
459 hybrids, and other admixed genotypes ($Q < 95\%$), there were 126 samples of probable
460 Guinean origin left (99 herbarium specimens and 27 controls) that we reanalysed with the
461 same tools. The results of factorial analysis of the matrix of the Guinean group samples did
462 not reveal a strong structure (fig. 4). However, the first axis (9.98% of the global inertia)
463 enabled the separation of all the specimens of *C. canephora* var. *maclaudii* near Bilima and
464 Gamé cultivar near Macenta from more eastern origins. The second axis (5.37%) separated
465 *maclaudii* specimens from Gamé specimens. In accordance with the factorial analysis, the
466 results of the Bayesian assignment for $K = 2$ (best K according to Evanno's test) reflected the
467 separation of the *maclaudii* and the Gamé specimens from more eastern origins (fig. 5 and
468 supplementary file 4). From $K = 5$, the *maclaudii* and the Gamé specimens were assigned to
469 two distinct ancestry sub-groups with little admixture. The other specimens collected from
470 South East Guinea to East Côte d'Ivoire were distributed into three other clusters with a
471 significant level of admixture and no strong geographical support.

472

473 **Genetic diversity of the Guinean and Congolese groups**

474 Once we removed the admixed individuals based on STRUCTURE assignment threshold
475 defined above (i.e. removal of individuals having $Q < 95\%$), we calculated the diversity
476 parameters of the 126 individuals of the Guinean group and the 24 individuals of the
477 Congolese group. We found the Congolese group more diverse than the Guinean one with
478 higher expected and observed heterozygosities (mean $He = 0.67$ and mean $Ho = 0.51$ for
479 Congolese while mean $He = 0.48$ and mean $Ho = 0.34$ for Guinean, supplementary file 5,
480 worksheet 2) and higher rarefied allelic richness per marker (mean of 6.87 for Congolese and
481 mean of 4.21 for Guinean, supplementary file 5, worksheet 10). Several markers were found
482 not to be at HW equilibrium within both groups (16 and 21 out of 23 markers for Congolese
483 and Guinean groups respectively). We also found a high number of private alleles
484 differentiating the two groups, with 79 alleles found within the Congolese group and 54
485 within the Guinean group (supplementary file 5, worksheet 9). The three complementary
486 differentiation statistics computed were consistently high and highly significant between the
487 two groups (supplementary file 5, worksheet 11).

488

489 **Genetic diversity and measure of population differentiation between STRUCTURE-**
490 **defined Guinean genetic sub-groups**

491 We assigned genotypes to sub-groups based on within Guinean group STRUCTURE analysis
492 at $K = 5$ by considering a threshold of 75% of ancestry. This left us with 91 samples (out of
493 126) clustered into five different Guinean sub-groups (sgG1 to sgG5) of 10, 9, 18, 29, and 25
494 genotypes respectively, the remaining 35 samples having admixed ancestry (supplementary
495 file 5, worksheet 3). Mean number of alleles per marker ranged from 1.61 (sgG1) up to 4.04
496 (sgG3). Observed heterozygosity and expected heterozygosity remained rather high, with
497 values ranging from 0.16 to 0.41 (H_o) and from 0.16 to 0.5 (H_e) depending on sub-groups.
498 The number of markers found not to be at HW equilibrium within sub-groups ranged from 1
499 to 11. The number of private alleles detected was moderate, ranging from 2 to 19 depending
500 on the sub-group considered (supplementary file 5, worksheets 4 to 9). Genetic diversity
501 assessed through rarefied allelic richness ranged from 1.64 to 3.38, with the lowest value for
502 sgG1 as compared to the other groups (supplementary file 5, worksheet 10). As for the
503 measures of genetic divergence between sub-groups, they were rather high for F_{st} and G_{st}
504 and moderate for Jost's D but all were highly significant, whichever the pairwise comparison
505 considered (supplementary file 5, worksheet 11).

506 Based on the results of the STRUCTURE analysis of the Guinean group detailed above, we
507 represented the geographic distribution of the five Guinean sub-groups on a map (fig. 6). The
508 first sub-group sgG1, located in the south-west of Fouta Djallon, corresponds to *Coffea*
509 *canephora* var. *maclaudii* with 10 specimens grouped into a single cluster with very little
510 admixture from other groups. The second sub-group sgG2 corresponds to the Gamé cultivar
511 located near Macenta. The last sub-groups, less differentiated, are located in the south of Côte
512 d'Ivoire near the towns of Divo and Tiassalé (sgG3), and in the west and centre west of Côte
513 d'Ivoire (sgG4 and sgG5, respectively).

514

515 **DISCUSSION**

516 The most recent comprehensive reviews of *C. canephora* in Guinea and Côte d'Ivoire, by
517 Portères, date back to the end of the 1950s (Portères 1959, 1962). He classified coffee plants
518 as local or introduced based on historical records (e.g. gene bank entry books), habitat
519 description, phenotypic observations, and by questioning planters and villagers during his
520 collection missions. By genotyping old herbarium specimens with nuclear microsatellites and
521 appropriate controls, we were able to determine whether each specimen was native or

522 introduced, and, if applicable, to estimate the level of admixture. In the light of our results,
523 Portères' observations appear to be remarkably precise and correct. Two notable exceptions
524 are specimens collected around Kissidougou in Guinea and those collected near Aboisso
525 (Ébobo population) in Côte d'Ivoire, which both grouped with Congolese genotypes whereas
526 their collectors considered them native to the country.

527

528 **Pattern of diversity of genotypes belonging to the Guinean group**

529 Within the Guinean group, our analyses revealed five sub-groups distributed over the two
530 countries (fig. 6) and variable levels of admixture between them.

531 In Guinea, two sub-groups were clearly distinguished from the others, suggesting restricted
532 gene flow probably due to their geographical isolation. The first sub-group (sgG1)
533 corresponds to *Coffea canephora* var. *maclaudii* determined by Chevalier in 1905 near
534 Bilima. The 10 analysed specimens (numbers 1 to 10), collected between 1905 and 1993,
535 grouped into a single cluster with very little admixture from other groups. This is the most
536 western and the most northern natural habitat of *C. canephora* ever found in Africa. The high
537 plateaus of the Fouta Djallon form a natural barrier separating the Southern Rivers basin from
538 the upper basin of Bafing-Sénégal (Boulvert 2003: 43). Bilima, located west of Fouta Djallon,
539 is exposed to rains (about 2000 mm per year) coming from the coast and has a climate that is
540 suitable for coffee. The only known site where coffee was observed is a forest island,
541 approximately 1400 m long and 350 m wide, which stretches down over the western slope of
542 the Bilima-Hénééré plateau, south of Bilima village, and is surrounded by cultivated fields and
543 fallow land. This diversity group also had the lowest allelic richness observed in our study,
544 suggesting both its peculiarity and vulnerability.

545 A second Guinean sub-group (sgG2) corresponds to the Gamé cultivar described by Portères
546 (1962). Bilima and Macenta are separated by the southern prolongation of the Fouta Djallon
547 to Mamou and by a forest-savanna transition zone around Faranah and Kissidougou over a
548 distance of about 300 km. According to Portères, the area of origin of Gamé is limited to the
549 plain near Macenta and the surrounding hills, west of the Ziama massif. The only herbarium
550 specimen (number 19) taken from the sub-group's natural habitat was collected by Barthe in
551 1936 in Bambaradou, a village located 3 km from Macenta, west of the Ziama massif, and
552 which was considered by Portères as the birth place of Gamé. The other specimens were
553 collected by Portères in 1958 either from Barthe's collection established in Macenta (number
554 20) or from the plantation that Fora Camara, an assistant of Portères, set up near the village of

555 Sérédou, east of the Ziamá massif (8 genotypes of original collection whose location is
556 unknown).

557 Separated from the Gamé sub-group in the west by the Ziamá massif, the other samples are
558 distributed over a vast area extending from the south of the prefecture of Kérouane, Guinea to
559 the far east of Côte d'Ivoire. The factorial and Bayesian analyses showed low differentiation
560 and weak grouping among these individuals. Several factors could account for this situation.
561 In terms of ecology and topography, there is no significant natural barrier that could restrict
562 gene flow between coffee populations in that vast area. Moreover, from the first quarter of the
563 twentieth century, there was a dramatic expansion of the plantations and intense movements
564 of planting material throughout both French colonies from eastern Côte d'Ivoire to the Guinea
565 Forest Region. This movement may have included coffee with a wide range of admixture
566 levels in locations characterized by plantations and forests situated in proximity. However, the
567 geographical distribution of the sub-groups shown in fig. 6 suggests that one sub-group
568 (sgG3) could be located in the south of Côte d'Ivoire near the towns of Divo and Tiassalé.
569 The last two less differentiated groups (sgG4 and sgG5) are located in the west and centre
570 west of Côte d'Ivoire.

571
572 For the very first time, despite the limited number of markers (23) we used, compared with
573 the number used in previous studies, e.g. 108 SSR markers in Cubry et al. (2013a), we found
574 significant evidence for population genetic structure within the Guinean group of *C.*
575 *canephora*. The structure detected is well organized spatially, especially in Guinea, and may
576 reflect diverse local adaptation or a long history of isolation of the populations. Further
577 investigations are needed to confirm these results and to elucidate the precise origin of the
578 structure (isolation by distance, gene flow barriers, past climatic history, etc.).

579

580 **Admixture between Guinean and Congolese groups**

581 We focused on herbarium specimens reported to be “spontaneous”, or “wild”, or “local” by
582 their collectors. However, we found several Guinean-Congolese admixed genotypes, mainly
583 among the specimens collected by Portères in Côte d'Ivoire at the end of the 1950s, e.g. the
584 ones collected from Tos plantation (see supplementary file 3). At that time, Congolese
585 germplasm was predominant in the plantations after massive imports from the 1930s. The
586 results of our genotyping study underline the interest of the large herbaria collected by
587 Portères to validate information related to the history of introductions and movements of
588 coffee in West Africa we retrieved in our literature review.

589

590 **The case of two populations of the Congolese group**

591 **Robusta Kissidougou** – Passing the barrier of Fouta Djallon and progressing to the southeast,
592 coffee is cultivated in the prefecture of Kissidougou in the forest-savanna transition zone,
593 mainly in the island forests that surround the villages. We found that the eight herbarium
594 specimens of our study collected in 1959 from four villages around Kissidougou (Boundou,
595 Makolo, Wendékéré, Massakoundou) belong to the Congolese group (SG1 and SG2 sub-
596 groups), although Portères considered them to derive from local material. According to a
597 handwritten note kept at MNHN, he located the discovery of a local *C. canephora* by Sagna
598 Camara, an agriculture field assistant, between 1927 and 1930, in a forest near Boundou and
599 the dissemination of that material only from the end of the 1940s. However, our study showed
600 that the coffee plants sampled by Portères most probably belong either to the population
601 Niaouli (Massakoundou specimens) or to the population that he called *Robusta Congo Belge*.
602 The latter was introduced for the first time in 1914 in Côte d'Ivoire from the botanical garden
603 of Eala (Democratic Republic of the Congo). This population, which is genetically close to
604 the INEAC population, was widely disseminated in Côte d'Ivoire from 1922 and since the
605 end of the 1920s in Guinea, especially in the Kissidougou region (Portères 1934, 1959, 1962,
606 Cordier 1961). According to Fairhead & Leach (1996) “coffee became more central to
607 Kissidougou's economy from the 1930s. In the context of the high coffee prices of the late
608 1940s and 1950s farmers increased their coffee production dramatically, filling virtually all
609 available forest island space with plantations and extending forest islands to harbour new
610 plantations where possible.” Therefore, the existence of coffee of the Guinean group
611 originating in the region of Kissidougou, as stated by Portères, remains to be demonstrated.

612

613 **Robusta Ébobo** – In the extreme south east of Côte d'Ivoire, the *Plantation Ébobo* was
614 created by the *Société des Plantations de la Tanoé* in 1925 near the village of the same name
615 on the Abi lagoon. According to Portères' handwritten notes conserved at MNHN, this
616 plantation sent seeds to the *Plantation de La Bia*, an estate belonging to the *Société des*
617 *Plantations d'Elima* (SPE) located at the mouth of the La Bia River, south of the town of
618 Aboisso (Lemery 1937). In 1958, Portères collected 20 herbarium specimens of robusta
619 Ébobo in the *Plantation de La Bia* and sent them to PAT. According to Portères (1959) and
620 Cordier (1961), the robusta Ébobo may derive from plants native to the Montagnes Bleues (70
621 to 100 m a.s.l.) located between the villages of Ébobo and Elima (Laplante & Rougerie 1949).
622 In another handwritten note, Portères mentioned a plot of robusta Ébobo established in 1935

623 on the *Plantation Ébobo* and inferred 1933–1934 as the likely date of discovery of its natural
624 habitat. We genotyped three herbarium specimens (numbers 151, 152, and 153) and found
625 them to belong to the Congolese group. They are genetically close to the Niaouli and Luki
626 populations that belong to the SG1 sub-group. These results are in accordance with those of
627 previous studies on several clones of robusta Ébobo by Berthaud (1984), Adibolo & Bertrand
628 (1988), Montagnon et al. (1992b), and Dussert et al. (2003), Dussert using the term “Aboisso”
629 instead of “Ébobo”. Therefore, there is little evidence to support Portères’ hypothesis
630 concerning a local origin of robusta Ébobo. This coffee may have been introduced earlier
631 from Lower Guinea, possibly by Beynis, a European planter who established his plantation
632 near Aboisso using seeds of Niaouli and *Café du Kouilou* imported from Gabon around 1910
633 (Cordier 1961). Bodard (c. 1960, CRA Bingerville, Côte d’Ivoire, unpubl. res.) mentioned
634 other introductions of *Café du Kouilou* from Gabon between 1917 and 1927 that were mainly
635 disseminated in the coastal region, particularly to the *Plantation de La Bia*.
636 The same early imports could explain why the herbarium sample number 142, collected in
637 1930 by Chevalier near Abengourou, although labelled as “spontaneous”, is actually a
638 specimen of the Niaouli population.

639

640 **Potential sites for further collection of Guinean germplasm**

641 As shown in fig. 6, most of the native *C. canephora* populations collected in the past in Côte
642 d’Ivoire and Guinea were found in the vegetation zone (Z2) characterized by semi-deciduous
643 moist forests, according to the classification introduced during the Conference of Yangambi
644 in 1956 (Aubréville 1957, Guillaumet & Adjanooun, 1971: 190). As already mentioned by
645 Portères (Portères 1937a, 1937b) and later by Berthaud (1984: 140), a number of populations
646 were also found in a transition zone (Z3) where forest and savanna co-exist in a complex
647 mosaic of forest islands and gallery forests along the rivers.

648 However, since the middle of the twentieth century, population pressure and immigration,
649 illegal or excessive logging, land encroachment and extension of coffee and cocoa
650 plantations, and inefficient public forest administration have led to the destruction or
651 fragmentation of large forest areas. This historical trend has mainly been documented in Côte
652 d’Ivoire (Brou et al. 1998, Brou et al. 2005b). According to the conservative estimate by
653 Fairhead & Leach (1998: 40), in 1900, Côte d’Ivoire may have had a forest cover of 7 to 8
654 million hectares that remained stable until 1955, but only 2.7 million in 1990. The U.S.
655 Geological Survey Earth Resources Observation and Science (USGS EROS) Centre

656 (<http://eros.usgs.gov/westafrica/>) has mapped the land cover of West Africa for three periods
657 in time (1975, 2000, and 2013) using many hundreds of Landsat images. By 2013, Côte
658 d’Ivoire had lost nearly 60% of the 3.7 million hectares of dense tropical forests that existed
659 in 1975. Similarly, degraded forest decreased by 28% and woodland area declined by 48%. In
660 Guinea, during the same period, dense forest areas decreased by about 33% to only 0.4
661 million hectares. Recent studies have started to document the impact of the political-military
662 crisis faced by Côte d’Ivoire between 2001 and 2013, on the 182 protected forests (*forêts*
663 *classées* in French) of that country. Before the conflict, the forest of Haut-Sassandra, in which
664 a coffee population (Pélézi, numbers 79 to 82) was collected in 1986, was one of the best
665 protected forest reserves in Côte d’Ivoire with 93% of forest cover. In 2015, the same forest
666 cover has decreased to less than 28%, mainly due to cocoa planting and tree logging that
667 followed the withdrawal of the forest ranger force (Barima et al. 2016).

668 Both countries have also been affected by periods of severe drought since the end of the
669 1960s, in step with what has been observed in the Sahel. The trend intensified during the
670 1980s and 1990s before a remission in the 2000s (Lubès et al. 1995; Brou et al. 2000, 2005a).
671 Moreover, recent studies investigated the potential impact of global climate change on the
672 natural habitats and cultivation areas of several species of *Coffea* (Davis et al. 2012, 2019;
673 Bunn et al. 2015). In a few decades, changes in climate will reduce the global area suitable for
674 coffee, leading to severe stress and a high risk of extinction. Even if the species *C. canephora*
675 (as a whole, including the Congolese group) is not considered as critically endangered, the
676 populations of the Guinean group growing in the transition zone between savanna and forest
677 (Z3) in Guinea and Côte d’Ivoire will likely be the first affected. Lastly, as important as forest
678 degradation, the dilution of the Guinean genetic pool resulting from hybridization with coffee
679 belonging to the Congolese group, which constitutes the major part of the coffee plantations
680 in Guinea and Côte d’Ivoire, could also hamper new collections of native coffee plants in both
681 countries.

682

683 Despite those factors, a few populations of coffee trees belonging to the Guinean group were
684 successfully collected in the last decades of the twentieth century in Côte d’Ivoire and Guinea
685 (Le Pierrès et al. 1989; Couturon & Montagnon 1991; Montagnon et al. 1993). Berthaud
686 (1984: 140) mentioned the presence of wild coffee in “sacred forests” that are preserved to
687 host initiation ceremonies. These sites often contain vast biodiversity with abundant fauna and
688 flora (Yao et al. 2013). Recent studies on farming systems in the Guinea Forest Region have

689 highlighted the expansion of agroforests surrounding the villages (Fairhead & Leach 1996;
690 Camara et al. 2009; Correia et al. 2010). In these agroforests, coffee, cocoa and kola are
691 associated with various forest tree species. However, the genetic identity of the coffee
692 populations found in these agroforests remains unknown.

693 Our findings on the genetic structure of *C. canephora* in Guinea and Côte d'Ivoire will guide
694 further surveys in both countries. The use of recent aerial and satellite images will provide
695 preliminary information on the current state of vegetation at past collection points. On the
696 spot, in addition to coffee sampling, it is advisable to record detailed information about the
697 ecology and associated flora, as well as a description of the coffee populations (age structure,
698 sanitary status, etc.) and of the environmental and social factors that could affect the future of
699 the forest. For sampling, priority should be given to germplasm that is not or insufficiently
700 represented in ex situ gene banks i.e. var. *maclaudii*, cultivar Gamé, and populations from
701 South-East Guinea, and, in Côte d'Ivoire, forest remnants in the south (Toumodi, Kassa, and
702 Mopri) and in the east (Abengourou and Agnibilekrou). The gene banks belonging to CNRA
703 and IRAG, once supplemented with new material from the Guinean group, will be major
704 resources for the reactivation of coffee breeding activities, especially given the increasing
705 need for adapted germplasm in the context of climate change. As stated by Davis et al.
706 (2019), wild variants of coffee species will be of primary importance in the future, especially
707 with the increasing incidence and duration of drought and the emergence or spread of diseases
708 and pests. This is particularly the case for populations located at the margins of the
709 environmental range of the species, like var. *maclaudii* in the extreme northwest and
710 populations located in the transition zone (Z3). These particular locations might be associated
711 with specific local adaptations of potential interest for breeding purposes. A more in depth
712 analysis including phenotypic evaluation of the genotypes from these locations should be
713 undertaken to detect particular traits associated with their specific history and ecological
714 growing conditions.

715

716 CONCLUSIONS AND FURTHER STUDIES

717 The analysis of genetic data of herbarium specimens, combined with information extracted
718 from historical records, herbarium documentation and scientific literature, allowed us to
719 validate (or, in a very few cases, invalidate) previous observations made by botanists
720 regarding the history of introductions and movements of robusta coffee germplasm in Guinea
721 and Côte d'Ivoire.

722 We explored the genetic diversity of *C. canephora* in these two countries by genotyping old
723 herbarium specimens with 23 nuclear microsatellites and appropriate controls. We were able
724 to determine the genetic group (Congolese or Guinean) of each specimen, and, if applicable,
725 to estimate the level of admixture. We identified several genotypes belonging to the
726 Congolese group, introduced with the development of commercial coffee plantations from
727 1920, and diverse Guinean-Congolese hybrids resulting from natural hybridization between
728 local and introduced genetic stocks. The results of factorial and Bayesian analyses provided
729 significant evidence for population genetic structure within the Guinean group of *C.*
730 *canephora*. The Guinean genotypes can be assigned to five sub-groups with distinct spatial
731 distribution, especially in Guinea where two sub-groups, corresponding to the var. *maclaudii*
732 and the cultivar Gamé, are characterized by a low level of admixture, which suggests
733 restricted gene flow probably due to geographical distances and the natural barriers formed by
734 the Fouta-Djallon and the Ziama massifs.

735 Because of past and current trends in forest degradation, there is an urgent need to describe
736 and protect potential collection sites in West Africa and to evaluate the genetic diversity of
737 the coffee populations they contain. Collecting *C. canephora* genotypes of the Guinean group
738 will serve as a valuable resource for future genetic improvement of robusta coffee. Due to its
739 limited genetic diversity compared to other Guinean sub-groups, the population of *C.*
740 *canephora* var. *maclaudii* near Bilima may be considered vulnerable and a priority target for
741 conservation.

742 The following step could be to extend genetic analysis to a larger number of specimens taken
743 from MNHN herbaria (P and PAT) and other European herbaria (e.g. herbarium LISU,
744 Lisbon, Portugal and BR, Meise Botanic Garden, Belgium). With recent DNA technologies
745 like next generation sequencing methods (chloroplast sequencing, gene capture, etc.), already
746 successfully applied to herbarium specimens (e.g. Staats et al. 2013, Hart et al. 2016, Suchan
747 et al. 2016), it is now possible to conduct in-depth investigation of diversity within a species,
748 at reasonable cost. Provided the herbarium specimens are accurately documented, a detailed
749 and comprehensive study of the species *C. canephora*, including Congolese and Guinean
750 groups and their sub-groups, becomes an achievable goal.

751

752

753

754

SUPPLEMENTARY FILES

755 Five supplementary files are associated to this paper:

756 Supplementary file 1: Detailed information on material used and genotyping data matrix of
757 the whole sample

758 Supplementary file 2: Neighbor-joining tree of the whole sample

759 Supplementary file 3: Results of STRUCTURE analysis of the whole sample

760 Supplementary file 4: Results of the STRUCTURE analysis of the Guinean group samples

761 Supplementary file 5: Genetic diversity parameters and measures of population differentiation

762

763

ACKNOWLEDGMENTS

764 The authors thank the *Muséum National d'Histoire Naturelle*, Paris and Myriam Gaudeul for
765 providing the specimen samples from the P herbarium, and Serge Bahuchet, Julien Blanc, and
766 Simon Juraver (UMR 7206 *Eco-Anthropologie*, <http://ecoanthropologie.cnrs.fr>) for giving
767 access to Portères' coffee specimens and documentation at the PAT herbarium. We also thank
768 Pierre Charmetant for providing samples from Cirad gene bank in French Guiana. We are
769 most grateful to Pierre Mournet and Hélène Vignes at the GPTR platform in Montpellier,
770 France for help in genotyping and to Xavier Perrier and Fabien De Bellis in data analyses. We
771 also thank the reviewers and the editors for their thorough editing and their helpful comments.
772 H.A. Nguyen's scholarship was supported by a grant from the CULTIVAR project funded by
773 Agropolis Fondation (ID 1504-007) through the Investissements d'Avenir program (Labex
774 Agro:ANR-10-LABX-0001-01) under the frame of I-SITE MUSE (ANR-16-IDEX-0006).

775

776

AUTHOR CONTRIBUTIONS

777 J.-P. Labouisse and P. Cubry. conceived and designed the research. R. Rivallan., H.A.
778 Nguyen., and J.-P. Labouisse performed the laboratory works. J.-P. Labouisse, P. Cubry, and
779 F. Austerlitz analysed and interpreted the data. J.-P. Labouisse, P. Cubry, F. Austerlitz, and R.
780 Rivallan wrote the manuscript.

781

- 783 Adibolo Y., Bertrand B. (1988) Étude de l'origine de la variété Niaouli au Togo et Bénin.
784 *Café Cacao Thé* 32: 293–298.
- 785 Anonymous (1934) Nouvelles de l'école et des anciens élèves. Nominations et mutations.
786 *L'Agronomie coloniale*. 197: 154.
- 787 Anthony F., Dussert S., Dulloo E. (2007) Coffee genetic resources. In: Engelmann F., Dulloo
788 M. E, Astorga C., Dussert S., Anthony F. (eds) Conserving coffee genetic resources :
789 complementary strategies for ex situ conservation of coffee (*Coffea arabica* L.) genetic
790 resources : a case study in CATIE, Costa Rica, 12–22. Rome, Bioversity International.
- 791 Archer F.I., Adams P.E., Schneiders B.B. (2017) StrataG: An R package for manipulating,
792 summarizing and analysing population genetic data. *Molecular Ecology Resources* 17:
793 5–11. <https://doi.org/10.1111/1755-0998.12559>
- 794 Aubréville A. (1957) Accord à Yangambi sur la nomenclature des types africains de
795 végétation. *Bois et Forêts des Tropiques* 51: 23–27.
- 796 Avelino J., Cristancho M., Georgiou S., Imbach P., Aguilar L., Bornemann G., Läderach P.,
797 Azueto F., Hruska A.J., Morales C. (2015) The coffee rust crises in Colombia and
798 Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food*
799 *Security* 7: 303–321. <https://doi.org/10.1007/s12571-015-0446-9>
- 800 Barima Y.S.S., Kouakou A.T.M., Bamba I., Sangne Y.C., Godron M., Andrieu J., Bogaert J.
801 (2016) Cocoa crops are destroying the forest reserves of the classified forest of Haut-
802 Sassandra (Ivory Coast). *Global Ecology and Conservation* 8: 85–98.
803 <https://doi.org/10.1016/j.gecco.2016.08.009>
- 804 Berthaud J. (1984) Les ressources génétiques pour l'amélioration des caféiers africains
805 diploïdes : évaluation de la richesse génétique des populations sylvestres et de ses
806 mécanismes organisateurs. Conséquences pour l'application. PhD thesis, Université de
807 Paris-Sud, Orsay, France. Available from
808 <http://www.documentation.ird.fr/hor/fdi:16623> [accessed 7 November 2017].

- 809 Boulvert Y. (2003) Carte morphopédologique de la République de Guinée à 1:500 000 en
810 feuilles Ouest et Est. Paris, IRD.
- 811 Brou Y.T., Akindès F., Bigot S. (2005a) La variabilité climatique en Côte d’Ivoire : entre
812 perceptions sociales et réponses agricoles. *Cahiers Agricultures* 14: 533–540.
813 Available from: [http://revues.cirad.fr/index.php/cahiers-](http://revues.cirad.fr/index.php/cahiers-agricultures/article/view/3054816623)
814 [agricultures/article/view/3054816623](http://revues.cirad.fr/index.php/cahiers-agricultures/article/view/3054816623) [accessed 7 June 2018]
- 815 Brou Y.T., Oszwald J., Bigot S., Servat E. (2005b) Risques de déforestation dans le domaine
816 permanent de l’Etat en Côte d’Ivoire : quel avenir pour ces derniers massifs forestiers?
817 *Téledétection* 5: 236–275.
- 818 Brou Y.T., Servat E., Paturel J.-E. (1998) Activités humaines et variabilité climatique : cas du
819 sud forestier ivoirien. In: Servat E., Hughes D., Fritsch J.-M., Hulme M. (eds) Water
820 resources variability in Africa during the 20th century. International Conference,
821 Abidjan (CIV), 1998/11/16-19, Publication - AISH, 365–373. Wallingford, UK,
822 AISH.
- 823 Brou Y.T., Servat E., Paturel J.-E. (2000) Évolution du couvert forestier ivoirien sur la
824 période 1950-1990, en relation avec la variabilité du climat et les activités
825 anthropiques. In : Servat M., Servat-Vildary S. (eds) Dynamique à long terme des
826 écosystèmes forestiers intertropicaux, 57–62. Paris, IRD, UNESCO.
- 827 Bunn, C., Läderach P., Ovalle Rivera O., Kirschke D. (2015) A bitter cup: climate change
828 profile of global production of Arabica and Robusta coffee. *Climatic Change* 129: 89–
829 101. <https://doi.org/10.1007/s10584-014-1306-x>
- 830 Camara A., Dugué P., Cheylan J.-P., Kalms J.-M. (2009) De la forêt naturelle aux agroforêts
831 en Guinée forestière. *Agriculture* 5: 425–432. <https://doi.org/10.1684/agr.2009.0325>
- 832 Chevalier A. (1905) Les caféiers sauvages de la Guinée française. Séance du 29 mai 1905.
833 *Comptes rendus hebdomadaires des séances de l’Académie des Sciences Paris* 140:
834 1472–1475.
- 835 Chevalier A. (1909) Les Hauts plateaux du Fouta Djallon. *Annales de Géographie* 28: 253–
836 261.

- 837 Chevalier A. (1914) Sudania. Énumération des plantes récoltées en Afrique tropicale et
838 équatoriale par M. Aug. Chevalier de 1905 à 1907. Tome II (N° 12001 à 20214).
839 Paris, Auguste Challamel,
- 840 Chevalier A. (1920) Exploration botanique de l’Afrique occidentale française. Tome I:
841 Énumération des plantes récoltées. Paris, Lechevalier.
- 842 Chevalier A. (1929) Les caféiers du Globe. Fascicule I. Généralités sur les caféiers. Paris,
843 Lechevalier.
- 844 Cordier L. (1961) Les objectifs de la sélection caféière en Côte d’Ivoire. *Café Cacao Thé* 5:
845 147–159.
- 846 Correia M., Diabaté M., Beavogui P., Guilavogui K., Lamanda N., de Foresta, H. (2010)
847 Conserving forest tree diversity in Guinée Forestière (Guinea, West Africa): the role
848 of coffee-based agroforests. *Biodiversity and Conservation* 19: 1725–1747.
849 <https://doi.org/10.1007/s10531-010-9800-6>
- 850 Couturon E., Montagnon C. (1991) Rapport de prospection de caféiers *C. canephora*
851 spontanés dans la région de Touba en Côte d’Ivoire du 4/11/1991 au 7/11/1991. IRCC
852 Côte d’Ivoire.
- 853 Cubry P. (2008) Structuration de la diversité génétique et analyse des patrons de déséquilibre
854 de liaison de l’espèce *Coffea canephora* Pierre ex. Froehner. PhD thesis, Université
855 Montpellier 2, France.
- 856 Cubry P., De Bellis F., Avia K., Bouchet S., Pot D., Dufour M., Legnate H., Leroy T. (2013a)
857 An initial assessment of linkage disequilibrium (LD) in coffee trees: LD patterns in
858 groups of *Coffea canephora* Pierre using microsatellite analysis. *BMC Genomics* 14:
859 <https://doi.org/10.1186/1471-2164-14-10>
- 860 Cubry P., De Bellis F., Pot D., Musoli P., Leroy T. (2013b) Global analysis of *Coffea*
861 *canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals
862 impacts from climatic refuges and migration effects. *Genetic Resources and Crop*
863 *Evolution* 60: 483–501. <https://doi.org/10.1007/s10722-012-9851-5>

864 Cubry P., Musoli P., Legnaté H., Pot D., De Bellis F., Poncet V., Anthony F., Dufour M.,
865 Leroy T. (2008). Diversity in coffee assessed with SSR markers: structure of the genus
866 *Coffea* and perspectives for breeding. *Genome* 51: 50–63. [https://doi.org/10.1139/g07-](https://doi.org/10.1139/g07-096)
867 096

868 Davis A.P. (2011) *Psilanthus mannii*, the type species of *Psilanthus*, transferred to *Coffea*.
869 *Nordic Journal of Botany* 29: 471–472. [https://doi.org/10.1111/j.1756-](https://doi.org/10.1111/j.1756-1051.2011.01113.x)
870 1051.2011.01113.x

871

872 Davis A.P., Chadburn H., Moat J., O’Sullivan R., Hargreaves S., Lughadha E.N. (2019) High
873 extinction risk for wild coffee species and implications for coffee sector sustainability.
874 *Science Advances* 5: eaav3473. <https://doi.org/10.1126/sciadv.aav3473>

875 Davis, A.P., Gole T.W., Baena S., Moat J. (2012) The impact of climate change on
876 indigenous arabica coffee (*Coffea arabica*): Predicting future trends and identifying
877 priorities. *PLoS ONE* 7(11): e47981. <https://doi.org/10.1371/journal.pone.0047981>

878 Davis A.P., Govaerts R., Bridson D.M., Stoffelen P. (2006) An annotated taxonomic
879 conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society*
880 152: 465–512. <https://doi.org/10.1111/j.1095-8339.2006.00584.x>

881 Devreux M., Vallaëys G., Pochet P., Gilles A. (1959) Recherches sur l’autostérilité du caféier
882 robusta (*Coffea canephora* Pierre). Bruxelles, Publications de l’Institut National pour
883 l’Étude Agronomique du Congo Belge (INEAC).

884 Dussert S., Lashermes P., Anthony F., Montagnon C., Trouslot P., Combes M.-C., Berthaud
885 J., Noirot M., Hamon S. (2003). Coffee (*Coffea canephora*). In Hamon P., Seguin M.,
886 Perrier X., Glaszmann J.-C. (eds) Genetic diversity of cultivated tropical plants, 239–
887 258. Enfield (NH), USA, Montpellier, France, Science Publishers, Inc. and CIRAD.

888 Earl D.A., vonHoldt B.M. (2012) STRUCTURE HARVESTER: a website and program for
889 visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
890 *Genetics Resources* 4: 359–361. <https://doi.org/10.1007/s12686-011-9548-7>

891 El Mousadik A., Petit R.J. (1996) High level of genetic differentiation for allelic richness
892 among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to

- 893 Morocco. *Theoretical and Applied Genetics* 92: 832–839.
894 <https://doi.org/10.1007/BF00221895>
- 895 Eskes A.B., Leroy T. (2012). Coffee selection and breeding. In: Wintgens J.N. (ed.) Coffee:
896 growing, processing, sustainable production: a guidebook for growers, processors,
897 traders, and researchers, 61–90. Weinheim, Germany, Wiley-VCH Verlag GmbH &
898 Co. KGaA.
- 899 Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals
900 using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620.
901 <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- 902 Fairhead J., Leach M. (1996) Enriching the landscape: social history and the management of
903 transition ecology in the forest–savanna mosaic of the Republic of Guinea. *Africa* 66:
904 14–36.
- 905 Fairhead J., Leach M. (1998) Reframing deforestation: global analyses and local realities with
906 studies in West Africa. London, Routledge.
- 907 Ferrão R.G. (2007) Café Conilon. Vitória, Brazil, Incaper-Instituto Capixaba de Pesquisa,
908 Assistência Técnica e Extensão Rural.
- 909 Fréchou H. (1955) Les plantations européennes en Côte d’Ivoire. *Cahiers d’Outre-Mer* 8: 56–
910 83.
- 911 Froehner A. (1897) Übersicht über die Arten der Gattung *Coffea*. *C. canephora* Pierre.
912 *Notizblatt des Königl. Botanischen Gartens und Museums zu Berlin* 7:237–238.
913 Available from
914 <https://www.biodiversitylibrary.org/page/28795736#page/279/mode/1up> [accessed 27
915 April 2019]
- 916 Gomez C., Dussert S., Hamon P., Hamon S., de Kochko A., Poncet V. (2009) Current genetic
917 differentiation of *Coffea canephora* Pierre ex A.Froehn in the Guineo-Congolian
918 African zone: cumulative impact of ancient climatic changes and recent human
919 activities. *BMC Evolutionary Biology* 9: 167. <https://doi.org/10.1186/1471-2148-9-167>

920 Goudet, J. (2005) HIERFSTAT, a package for r to compute and test hierarchical F-statistics.
921 *Molecular Ecology Notes* 5: 184–186. <https://doi.org/10.1111/j.1471->
922 8286.2004.00828.x

923 Guillaumet J.-L., Adjanohoun E. (1971) La végétation de la Côte d’Ivoire. In : Le milieu
924 naturel de la Côte d’Ivoire, Mémoires ORSTOM, 161–263. Paris, ORSTOM.
925 Available from <http://www.documentation.ird.fr/hor/fdi:16372> [accessed 13
926 September 2018]

927 Hart M.L., Forrest L.L., Nicholls J.A., Kidner C.A. (2016) Retrieval of hundreds of nuclear
928 loci from herbarium specimens. *Taxon* 65: 1081–1092. <https://doi.org/10.12705/655.9>

929 ICO (2018) Coffee market report. October 2018. International Coffee Organization. Available
930 from <http://www.ico.org/documents/cy2018-19/cmr-1018-e.pdf> [accessed 28
931 November 2018].

932 IFCC (1960) Institut Français du Café et du Cacao. Annual report 1960. Paris.

933 JOGF (1935) Actes du gouvernement local. Décision du Lieutenant-Gouverneur fixant les
934 soldes annuelles des chefs de canton du cercle de Macenta. *Journal Officiel de la*
935 *Guinée française. 1er septembre 1935*. Available from
936 <http://gallica.bnf.fr/ark:/12148/bpt6k9776901t> [accessed 13 June 2018].

937 Kopelman N.M., Mayzel J., Jakobsson M., Rosenberg N.A., Mayrose I. (2015) Clumpak: a
938 program for identifying clustering modes and packaging population structure
939 inferences across K. *Molecular Ecology Resources* 15: 1179–1191.
940 <https://doi.org/10.1111/1755-0998.12387>

941 Laplante A.C., Rougerie G. (1949) Étude et carte pédologique provisoire de la région littorale
942 située à l’Est de la Côte d’Ivoire. *Journal d’agriculture traditionnelle et de botanique*
943 *appliquée* 29: 24–33. <https://doi.org/10.3406/jatba.1949.6201>

944 Le Pierrès D., Charmetant P., Gnesio T., Bamba A., Konate M., Yapo A. (1987) Les
945 ressources génétiques naturelles des caféiers en Guinée : rapport sur la mission de
946 prospection effectuée par une équipe ORSTOM/IRCC du 1er février au 2 mars 1987.
947 Adiopodoumé, Côte d’Ivoire, ORSTOM. Available from:
948 <http://www.documentation.ird.fr/hor/fdi:27766> [accessed 22 February 2017].

949 Le Pierrès D., Charmetant P., Yapo A., Leroy T., Couturon E., Bontems S., Téhé H. (1989)
950 Les caféiers sauvages de Côte d'Ivoire et de Guinée : bilan des missions de
951 prospection effectuées de 1984 à 1987. In : Treizième colloque scientifique
952 international sur le café, Païpa, Colombie, 1989/08/21-25, 420–428. Paris, ASIC.

953 Lemery A. (ed) (1937). Le café en Côte d'Ivoire. La Société des Plantations d'Elima. La
954 Revue des ambassades et des questions diplomatiques et coloniales. Janvier 1937: 36.
955 Available from [https://bibliotheque-](https://bibliotheque-numerique.diplomatie.gouv.fr/ark:/12148/bpt6k6473705h/f29.image.texteImage)
956 [numerique.diplomatie.gouv.fr/ark:/12148/bpt6k6473705h/f29.image.texteImage](https://bibliotheque-numerique.diplomatie.gouv.fr/ark:/12148/bpt6k6473705h/f29.image.texteImage)
957 [accessed 22 January 2019]

958 Leroy J.-F. (1974) Roland Portères : Professeur au Muséum National d'Histoire Naturelle
959 (1906-1974). *Journal d'agriculture tropicale et de botanique appliquée* 21: 384–384.
960 Available from www.persee.fr/doc/jatba_0021-7662_1974_num_21_10_6948
961 [Accessed 10 January 2018]

962 Leroy T., De Bellis F., Legnate H., Kananura E., Gonzales G., Pereira L.F., Andrade A.C.,
963 Charmetant P., Montagnon, C., Cubry P., Marraccini P., Pot D, de Kochko, A. (2011).
964 Improving the quality of African robustas: QTLs for yield-and quality-related traits in
965 *Coffea canephora*. *Tree Genetics & Genomes* 7: 781–798.
966 <https://doi.org/10.1007/s11295-011-0374-6>

967 Leroy T., De Bellis F., Legnate H., Musoli P., Kalonji A., Loor Solórzano R.G., Cubry P.
968 (2014) Developing core collections to optimize the management and the exploitation
969 of diversity of the coffee *Coffea canephora*. *Genetica* 142: 185–199.
970 <https://doi.org/10.1007/s10709-014-9766-5>

971 Leroy T., Marraccini P., Dufour M., Montagnon C., Lashermes P., Sabau X., Ferreira L.P.,
972 Jourdan I., Pot D., Andrade A.C., Glaszmann J.-C., Vieira L.G.E., Piffanelli P. (2005)
973 Construction and characterization of a *Coffea canephora* BAC library to study the
974 organization of sucrose biosynthesis genes. *Theoretical and Applied Genetics* 111:
975 1032–1041. <https://doi.org/10.1007/s00122-005-0018-z>

976 Leroy T., Montagnon C., Charrier A., Eskes A.B. (1993) Reciprocal recurrent selection
977 applied to *Coffea canephora* Pierre. I. Characterization and evaluation of breeding

- 978 populations and value of intergroup hybrids. *Euphytica* 67: 113–125.
979 <https://doi.org/10.1007/BF00022734>
- 980 Leroy T., Montagnon C., Cilas C., Yapo A., Charmetant P., Eskes A.B. (1997) Reciprocal
981 recurrent selection applied to *Coffea canephora* Pierre. III. Genetic gains and results
982 of first cycle intergroup crosses. *Euphytica* 95: 347–354.
983 <https://doi.org/10.1023/A:1003074716379>
- 984 Lister D.L., Bower M.A., Howe C.J., Jones M.K. (2008) Extraction and amplification of
985 nuclear DNA from herbarium specimens of emmer wheat: a method for assessing
986 DNA preservation by maximum amplicon length recovery. *Taxon* 57: 254–258.
- 987 Lister D.L., Bower M.A., Jones M.K. (2010) Herbarium specimens expand the geographical
988 and temporal range of germplasm data in phylogeographic studies. *Taxon* 59: 1321–
989 1323.
- 990
- 991 Lubès H., Paturel J.-E., Servat E., Masson J.M., Kouamé B. (1995) Analysis of annual rainfall
992 series in Ivory Coast in order to detect climatic variations. In: Sixth international
993 meeting on statistical climatology, 455–458. Galway, Ireland, University College.
994 Available from: <http://www.documentation.ird.fr/hor/fdi:010060227> [accessed 12
995 December 2018]
- 996 Maclaud C. (1899) À travers la Guinée et le Fouta-Djallon. La mission Maclaud. *Bulletin du*
997 *Comité de l'Afrique française*. Septembre 1899: 293–298.
- 998 Montagnon C., Leroy T., Yapo A. (1992a) Diversité génotypique et phénotypique des
999 quelques groupes de caféiers (*Coffea canephora* P.) en collection. Conséquence sur
1000 leur utilisation en sélection. *Café Cacao Thé* 36: 187–198.
- 1001 Montagnon C., Leroy T., Yapo A. (1992b) Étude complémentaire de la diversité génotypique
1002 et phénotypique des caféiers de l'espèce *Coffea canephora* en collection en Côte
1003 d'Ivoire. In : Quatorzième colloque scientifique international sur le café, San
1004 Francisco, USA, 1991/07/14-19, Paris, ASIC.
- 1005 Montagnon C., Leroy T., Yapo A. (1993) Caractérisation et évaluation des caféiers *Coffea*
1006 *canephora* prospectés dans les plantations de Côte d'Ivoire. *Café Cacao Thé* 37: 115–
1007 119.

- 1008 Musoli P., Cubry P., Aluka P., Billot C., Dufour M., De Bellis F., Pot D., Bieysse D., Charrier
1009 A., Leroy T. (2009) Genetic differentiation of wild and cultivated populations:
1010 diversity of *Coffea canephora* Pierre in Uganda. *Genome* 52: 634–646.
1011 <https://doi.org/10.1139/g09-037>
- 1012 Perrier X., Jacquemoud-Collet J.-P. (2006) DARwin software Version 6. Available from:
1013 <http://darwin.cirad.fr/darwin> [accessed 20 November 2018].
- 1014 Poncet V., Dufour M., Hamon P., Hamon S., de Kochko A., Leroy T. (2007) Development of
1015 genomic microsatellite markers in *Coffea canephora* and their transferability to other
1016 coffee species. *Genome* 50: 1156–1161. <https://doi.org/10.1139/g07-073>
- 1017 Poncet V., Hamon P., Minier J., Carasco C., Hamon S., Noirot M. (2004) SSR cross-
1018 amplification and variation within coffee trees (*Coffea* spp.). *Genome* 47: 1071–1081.
1019 <https://doi.org/10.1139/g04-064>
- 1020 Portères R. (1934) Observations et sélection sur les caféiers à la station de Bingerville. I.
1021 Conduite de la sélection du caféier à la station de Bingerville. Bingerville, Côte
1022 d’Ivoire. Service de l’agriculture. Gouvernement de Côte d’Ivoire.
- 1023 Portères R. (1937a) Étude sur les caféiers spontanés de la section des *Eucoffea*. Partie 1 :
1024 Répartition et habitat. *Annales Agricoles de l’Afrique Occidentale Française et*
1025 *étrangère* 1: 68–91.
- 1026 Portères R. (1937b) Étude sur les caféiers spontanés de la section des *Eucoffea*. Partie 2 :
1027 Espèces, variétés et formes. *Annales Agricoles de l’Afrique Occidentale Française et*
1028 *étrangère* 1: 219–263.
- 1029 Portères R. (1937c) Étude sur les caféiers spontanés de la section des *Eucoffea*. Partie 3 : Mise
1030 en culture et sélection. *Annales Agricoles de l’Afrique Occidentale Française et*
1031 *étrangère* 1: 407–439.
- 1032 Portères R. (1939) Le choix des variétés et l’amélioration des caféiers en Côte d’Ivoire. *Revue*
1033 *de botanique appliquée et d’agriculture coloniale* 19: 18–29.
- 1034 Portères R. (1959) Valeur agronomique des caféiers des types Kouilou et Robusta cultivés en
1035 Côte d’Ivoire. *Café Cacao Thé* 3: 3–13.

- 1036 Portères R. (1962) Caféiers de la République de Guinée. *Café Cacao Thé* 6: 3–18.
- 1037 Poupart Y. (1938) La culture du *Coffea Robusta* dans l’Ouest de la Cote d’Ivoire.
1038 *L’Agronomie Coloniale* 252: 173–180.
- 1039 Pritchard J.K., Stephens M., Donnelly P. (2000). Inference of population structure using
1040 multilocus genotype data. *Genetics* 155: 945–959.
- 1041 Risterucci A.M., Grivet L., N’Goran J.A.K., Pieretti I., Flament M.H., Lanaud C. (2000). A
1042 high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*
1043 101: 948–955. <https://doi.org/10.1007/s001220051566>
- 1044 Roullier C., Benoit L., McKey D.B., Lebot V. (2013) Historical collections reveal patterns of
1045 diffusion of sweet potato in Oceania obscured by modern plant movements and
1046 recombination. *Proceedings of the National Academy of Sciences* 110: 2205–2210.
1047 <https://doi.org/10.1073/pnas.1211049110>
- 1048 Saitou N., Nei M. (1987) The neighbor-joining method: a new method for reconstructing
1049 phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.
1050 <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- 1051 Särkinen T., Staats M., Richardson J.E., Cowan R.S., Bakker F.T. (2012) How to open the
1052 treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE*
1053 7(8): e43808. <https://doi.org/10.1371/journal.pone.0043808>
- 1054 Staats M., Erkens R.H.J., van de Vossen B., Wieringa J.J., Kraaijeveld K., Stielow B.,
1055 Geml J., Richardson J., Bakker F.T. (2013) Genomic treasure troves: complete
1056 genome sequencing of herbarium and insect museum specimens. *PLoS ONE* 8(7):
1057 e69189. <https://doi.org/10.1371/journal.pone.0069189>
- 1058 Suchan T., Pitteloud C., Gerasimova N.S., Kostikova A., Schmid S., Arrigo N., Pajkovic M.,
1059 Ronikier M., Alvarez N. (2016) Hybridization capture using RAD probes (hyRAD), a
1060 new tool for performing genomic analyses on collection specimens. *PLoS ONE* 11(3):
1061 e0151651. <https://doi.org/10.1371/journal.pone.0151651>

- 1062 Thiers, B. (2019) Index Herbariorum: A worldwide index of 3,400 herbaria and associated
1063 staff. New York Botanical Garden Steere Herbarium (continuously updated).
1064 Available from <http://sweetgum.nybg.org/science/ih/> [accessed 27 April 2019].
- 1065 Tourte R. (2005a) Histoire de la recherche agricole en Afrique tropicale francophone. Volume
1066 V. Le temps des stations et de la mise en valeur 1918–1940/1945. Rome, FAO.
- 1067 Tourte R. (2005b) Histoire de la recherche agricole en Afrique tropicale francophone. Volume
1068 VI. De l’empire colonial à l’Afrique indépendante. 1945–1960. La recherche prépare
1069 le développement. Rome, FAO.
- 1070 Wallengren M. (2017) The arabica-robusta balance. Does it really matter? A special report by
1071 SpillingTheBeans for the Global Coffee Review’s. Jan. 2013 issue. Available from
1072 [http://spilling-the-beans.net/special-report-the-arabica-robusta-balance-does-it-really-](http://spilling-the-beans.net/special-report-the-arabica-robusta-balance-does-it-really-matter/)
1073 [matter/](http://spilling-the-beans.net/special-report-the-arabica-robusta-balance-does-it-really-matter/) [accessed 2 December 2018].
- 1074 White F. (1979) The Guineo-Congolian region and its relationships to other phytochoria.
1075 *Bulletin du Jardin botanique national de Belgique / Bulletin van de Nationale*
1076 *Plantentuin van België* 49: 11–55.
- 1077 Yao A., Kpangui K., Kouao K., Adou L., Vroh B., N’Guessan K. (2013) Diversité floristique
1078 et valeur de la forêt sacrée Bokasso (Est de la Côte d’Ivoire) pour la conservation.
1079 *VertigO - la revue électronique en sciences de l’environnement*. Available from:
1080 <https://vertigo.revues.org/13500> [accessed 18 November 2017].

1081 **Figure 1**– Collection points of the *C. canephora* herbarium specimens (red dots) and
1082 genotypes used as controls (purple triangles), main towns (black dots), and IRAG and CNRA
1083 research stations mentioned in the text (green stars). Two main geographic features of Guinea
1084 (elevation>800m a.s.l.), the Fouta Djallon plateau (A) and the Ziama massif with its northern
1085 extension (B), are shown in brown.

1086

1087 **Figure 2** – Factorial analysis (PCoA) of the simple matching distance matrix of the 162
1088 samples of *C. canephora*: 126 herbarium specimens collected in Guinea and Côte d’Ivoire
1089 and 36 genotypes used as controls. Representation of the first three axes: axis2/axis1 (A);
1090 axis3/axis1 (B).

1091

1092 **Figure 3** – Results of STRUCTURE analysis for the 162 samples of *C. canephora*: 126
1093 herbarium specimens collected in Guinea and Côte d’Ivoire and 36 genotypes from previous
1094 studies used as controls. Population assignment for $K = 2$. Each individual is represented by a
1095 thin vertical line, which is partitioned into two coloured segments that represent the
1096 individual's estimated membership fractions in the two clusters.

1097

1098 **Figure 4** – Factorial analysis (PCoA) of the simple matching distance matrix of 126 samples
1099 (99 herbarium specimens and 27 controls) of *C. canephora* collected in Guinea and Côte
1100 d’Ivoire and belonging to the Guinean group. Representation of the first two axes.

1101

1102 **Figure 5** – Results of STRUCTURE analysis for 126 samples of *C. canephora* belonging to
1103 the Guinean group. Estimated population structure for $K = 2$ and $K = 5$. Each individual is
1104 represented by a thin vertical line, which is partitioned into K coloured segments that
1105 represent the individual's estimated membership fractions in K clusters.

1106

1107 **Figure 6** – Geographical distribution of Guinean sub-groups of *C. canephora* (pie charts).
1108 Assignment in STRUCTURE analysis at $K = 5$ assumed groups ($Q \geq 75\%$ membership). The
1109 four main vegetation zones are delimited by coloured lines: Guinean domain (Z1: moist
1110 evergreen forest, widely degraded; Z2: moist semi-deciduous forest); Z3: transition zone; Z4:
1111 Sudanian domain. The lines for vegetation zones are drawn after distribution data presented in
1112 maps by Guillaumet & Adjanohoun (1971) for Côte d’Ivoire and Boulvert (2003) for Guinea.

1113

1114

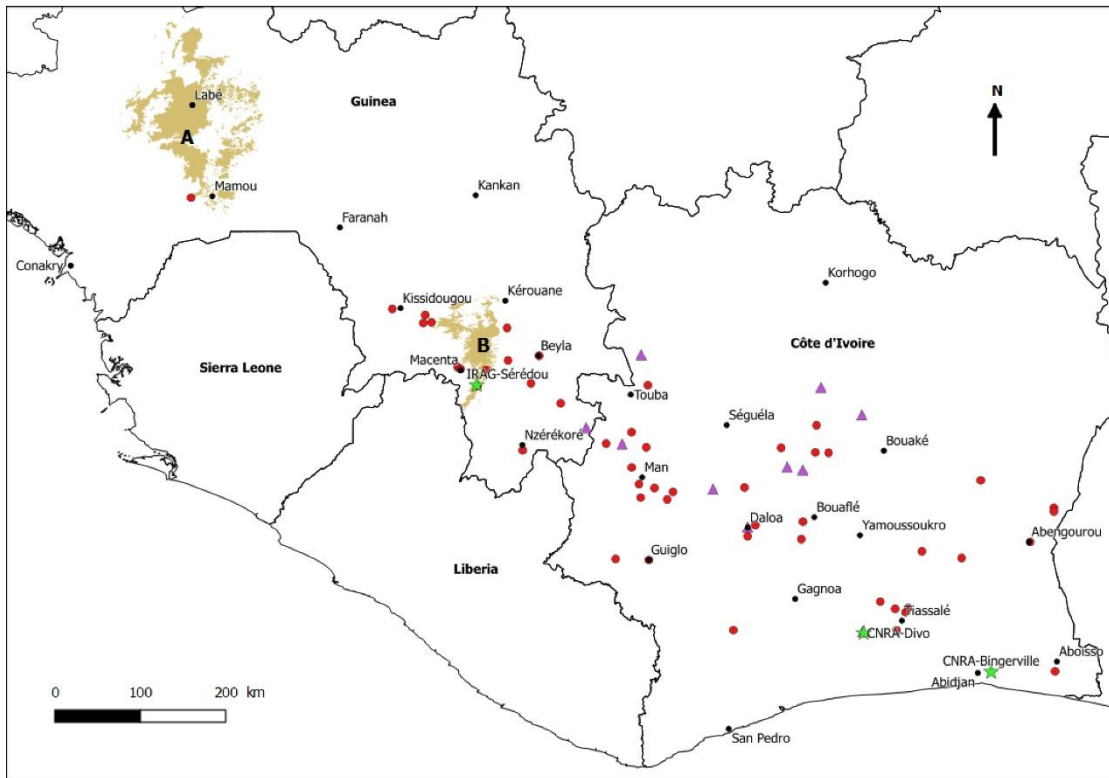
1115 **Table 1 – List of the collection points, geographic coordinates and number of samples of**
1116 ***Coffea canephora* genotyped in the study.**

1117 The samples are ordered according to the longitude coordinate (from the West to the East).
1118 Details are in supplementary file 1. (*) Exact collection point unknown (coordinates
1119 estimated); (**) Gamé cultivar collected in Sérédou gene bank by Portères (exact origin
1120 unknown).

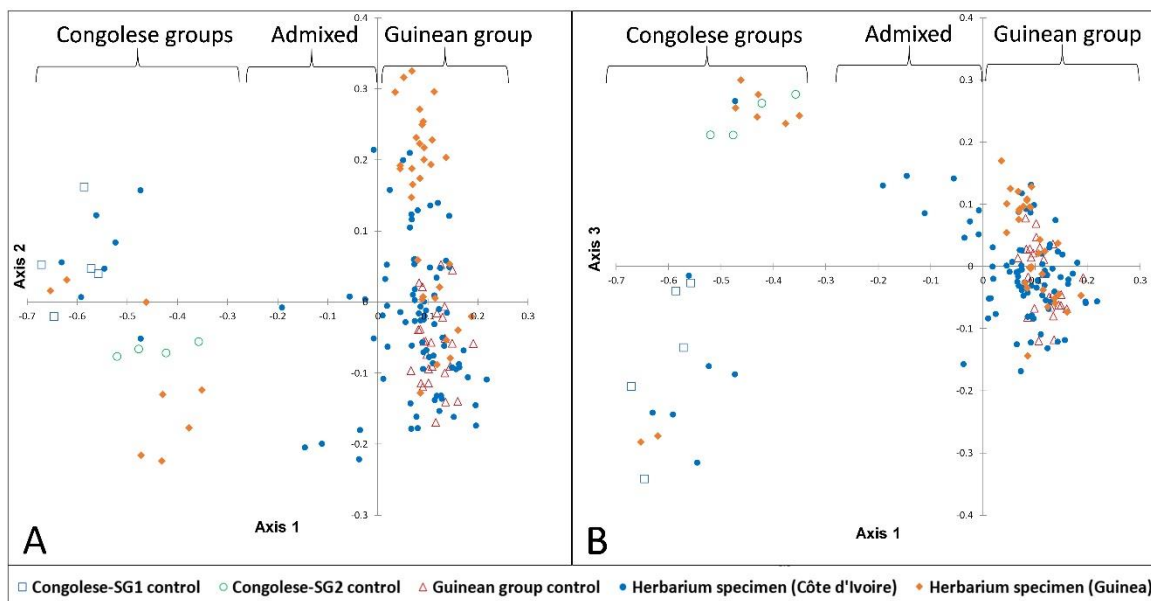
1121

1122 **Table 2 – Main characteristics and references of the 23 SSRs markers used in the study.**

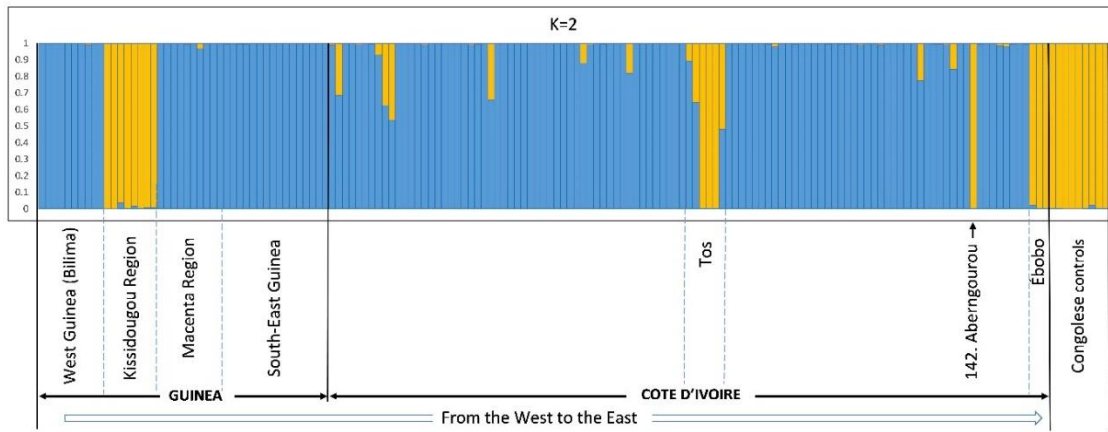
1123 (*) accession number in EMBL-EBI database (European Bioinformatics Institute,
1124 <http://www.ebi.ac.uk>); (**) in SGN database (Sol Genomics Network,
1125 <http://solgenomics.net>).



1
 2 **Figure 1**– Collection points of the *C. canephora* herbarium specimens (red dots) and
 3 genotypes used as controls (purple triangles), main towns (black dots), and IRAG and CNRA
 4 research stations mentioned in the text (green stars). Two main geographic features of Guinea
 5 (elevation>800m a.s.l.), the Fouta Djallon plateau (A) and the Ziama massif with its northern
 6 extension (B), are shown in brown.
 7



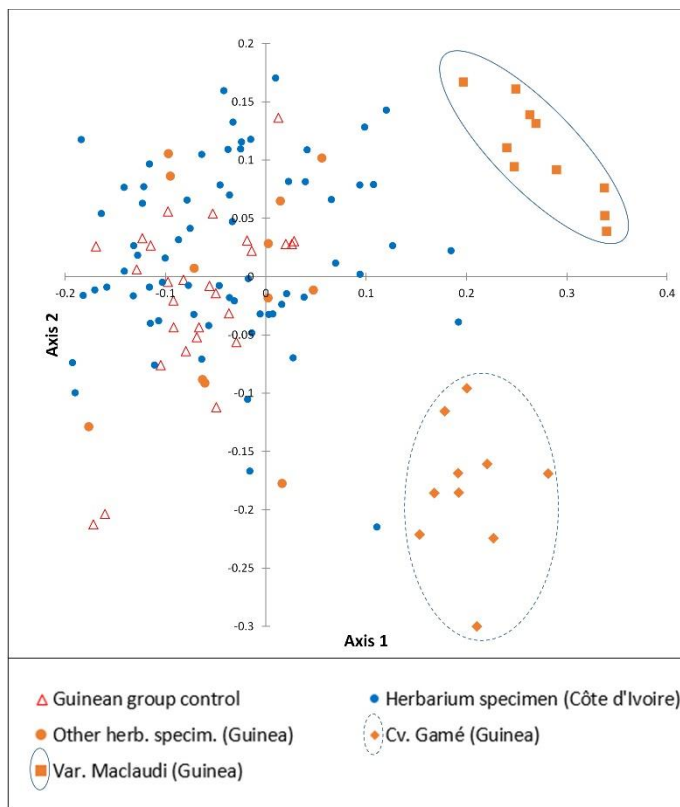
8
 9 **Figure 2** – Factorial analysis (PCoA) of the simple matching distance matrix of the 162 samples
 10 of *C. canephora*: 126 herbarium specimens collected in Guinea and Côte d’Ivoire and 36
 11 genotypes used as controls. Representation of the first three axes: A, axis2/axis1; B, axis3/axis1.
 12



13

14 **Figure 3** – Results of STRUCTURE analysis for the 162 samples of *C. canephora*: 126
 15 herbarium specimens collected in Guinea and Côte d’Ivoire and 36 genotypes from previous
 16 studies used as controls. Population assignment for K=2. Each individual is represented by a
 17 thin vertical line, which is partitioned into two colored segments that represent the individual's
 18 estimated membership fractions in the two clusters.

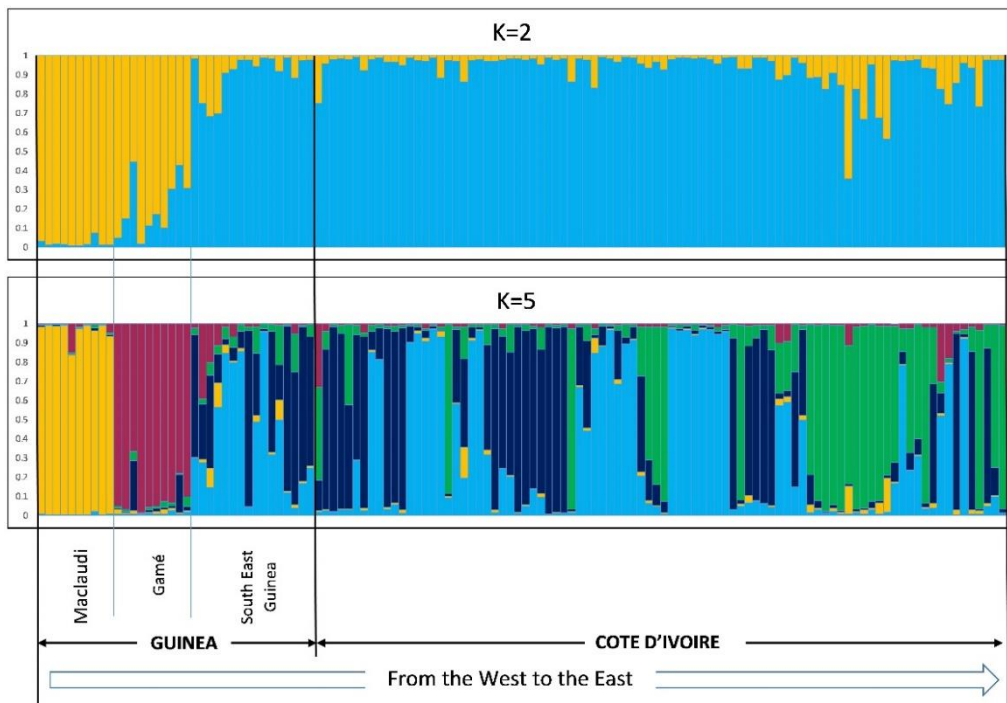
19



20

21 **Figure 4** – Factorial analysis (PCoA) of the simple matching distance matrix of 126 samples
 22 (99 herbarium specimens and 27 controls) of *C. canephora* collected in Guinea and Côte
 23 d’Ivoire and belonging to the Guinean group. Representation of the first two axes.

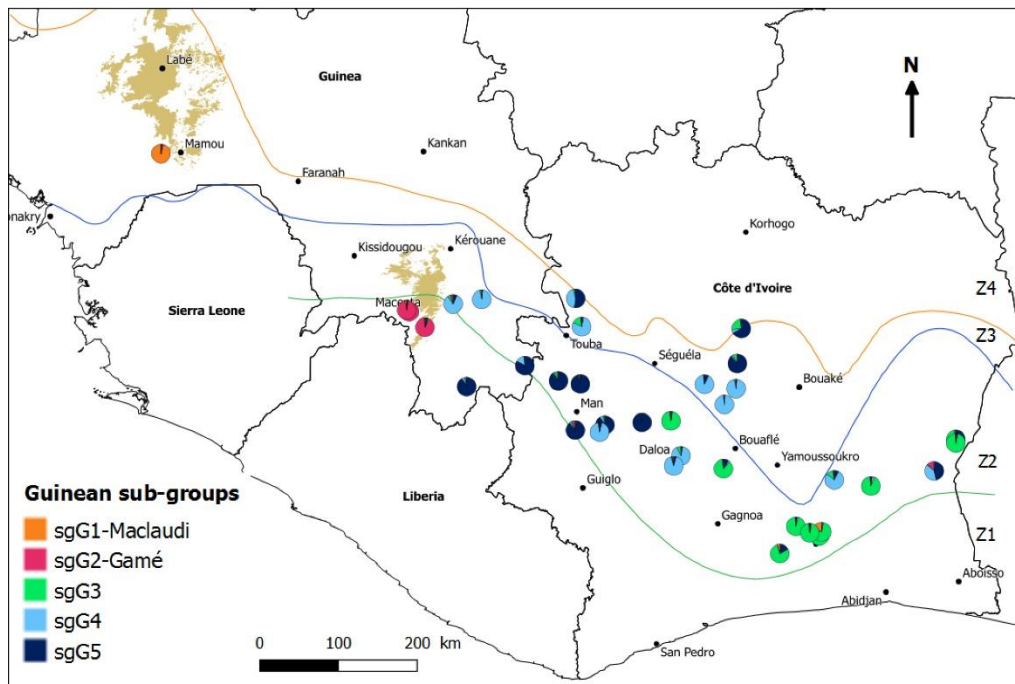
24



25

26 **Figure 5** – Results of STRUCTURE analysis for 126 samples of *C. canephora* belonging to
 27 the Guinean group. Estimated population structure for K=2 and K=5. Each individual is
 28 represented by a thin vertical line, which is partitioned into K colored segments that represent
 29 the individual's estimated membership fractions in K clusters.

30



31

32 **Figure 6** -- Geographical distribution of Guinean sub-groups of *C. canephora* (pie charts).
 33 Assignment in STRUCTURE analysis at K=5 assumed groups ($Q \geq 75\%$ membership). The
 34 four main vegetation zones are delimited by colored lines: Guinean domain (Z1: moist
 35 evergreen forest, widely degraded; Z2: moist semi-deciduous forest); Z3: transition zone; Z4:

36 Sudanian domain. The lines for vegetation zones are drawn after distribution data presented in
37 maps by Guillaumet & Adjanohoun (1971) for Côte d'Ivoire and Boulvert (2003) for Guinea.

38

39 **Table 1 -- List of the collection points, geographic coordinates and number of samples of**
 40 ***C. canephora* genotyped in the study.** The samples are ordered according to the longitude
 41 coordinate (from the West to the East). Details are in Appendix S1.

Collection points	Number of samples	Sample ID	Latitude	Longitude
GRAND TOTAL	162			
<i>Herbarium specimens</i>	<i>126</i>			
Guinea	40			
Bilima	10	1 - 10	10.35	-12.31
Massakoundou	2	11-12	9.18	-10.19
Makolo	2	13-14	9.04	-9.86
Wondékéré	2	15-16	9.12	-9.84
Boundo	2	17-18	9.04	-9.78
Bambaradou	1	19	8.57	-9.50
Macenta	1	20	8.55	-9.47
Sérédou_Gamé (**)	8	21-28	8.37	-9.29
Cercle_Macenta (*)	1	29	8.54	-9.20
Milo	1	30	8.98	-8.98
Koréla	5	31-35	8.64	-8.97
Nzérékoré	1	36	7.69	-8.82
Yarankadougou	1	37	8.40	-8.73
Beyla	2	38-39	8.69	-8.64
Foumbadougou	1	40	8.19	-8.42
Côte d'Ivoire	86			
Bangopolé	1	45	7.77	-7.94
Gueya	1	46	6.55	-7.84
Bafing	1	51	7.89	-7.67
Kiélé	2	52-53	7.51	-7.67
Dianlé	1	54	7.34	-7.59
Kodopleu	1	55	7.20	-7.58
Digoualé	1	60	7.73	-7.52
Touba	7	61-67	8.38	-7.50
Guiglo	1	68	6.54	-7.49
Kékobly	1	69	7.30	-7.43
Bahibly	2	70-71	7.18	-7.30
Kouibly	7	72-78	7.26	-7.24
Soubéré	1	83	5.80	-6.60
Déma	1	84	7.30	-6.48
Gbétitapia	6	86-91	6.79	-6.45
Zaprégouhé	1	92	6.91	-6.37
Toubalo	1	93	7.72	-6.10
Prozibanfla	3	95-97	6.76	-5.89
Tos	6	99-104	6.94	-5.87
Asséyakro	7	105-111	7.67	-5.74
Totokro	1	112	7.96	-5.73
Kongonouassou	2	118-119	7.67	-5.60

Goudy	2	123-124	6.10	-5.06
Mbrimbo	2	125-126	6.03	-4.90
Mopri	1	127	5.80	-4.89
Kassa I	4	128-131	5.99	-4.79
Kassa II	2	132-133	6.04	-4.76
Sagassou	2	134-135	6.63	-4.62
Agbo	4	136-139	6.56	-4.20
Nzinzibelekro	2	140-141	7.38	-4.00
Abengourou	3	142-144	6.73	-3.48
Assikasso	4	145-148	7.09	-3.23
Kongodia	2	149-150	7.05	-3.23
Ébobo	3	151-153	5.37	-3.22
Controls		36		
A. Guinean group				
Guinea		4		
Piné	4	41-44	7.93	-8.15
Côte d'Ivoire		23		
Ira	4	47-50	7.76	-7.77
Mouniandougou	4	56-59	8.70	-7.57
Pélézi	4	79-82	7.29	-6.82
Daloa	1	85	6.89	-6.45
Kaloufla	1	94	7.52	-6.04
Gohitafla	1	98	7.49	-5.87
Fourougbankoro	5	113-117	8.36	-5.68
Logbonou	2	120-121	8.07	-5.25
Divo	1	122	5.79	-5.24
B. Congolese group				
DR Congo		7		
Luki (sub-group SG1)	3	154-156	N/A	N/A
INEAC (sub-group SG2)	4	159-162	N/A	N/A
Gabon via Togo/Bénin		2		
Niaouli (sub-group SG1)	2	157-158	N/A	N/A

42 (*) Exact collection point unknown (coordinates estimated); (**) Gamé cultivar collected in
43 Sérédou gene bank by Portères (exact origin unknown).

44

Table 2 -- Main characteristics and references of the 23 SSRs markers used in the study.

TropGeneDB ID	Forward primer sequence	Reverse primer sequence	PCR product size	Repeat motif	Fluoro-chrome	PCR pool	Accession number	Reference
Me_BE_gA14_DL024	CGAGCGTAGAGGAACA	ACATGATACAAGAATGGAGA	133	(AAT)4	VIC	M2	AJ871903*	Leroy et al. 2005
Me_ssr120291	GCTGCTAAAGCATCCGAAAC	CTTGGTAAACCTCAACCCGA	159	(AAG)8	6-FAM	M3	SGN-U614007**	Cubry et al. 2008
Me_ssr120568	TTGAGATTTTCAGCCGAAGG	AAACTGAATCCAACGCATCC	117	(GA)16	6-FAM	M4	SGN-U619826**	Cubry et al. 2008
Me_ssr123106	GAGACGTGGTTCGTGCTGTA	GTAATCGCAGGCTAAAGCG	185	(AGA)5	NED	M4	SGN-U619140**	Cubry et al. 2008
Me_ssr123682	TCTCTCCAATTTCCAATGGC	TCCATAGTCCCAAAAAGCTG	126	(GAA)7	PET	M1	SGN-U613225**	Cubry et al. 2008
Mg_CMA0041	GCCTCTGTCGATCTTTCTA	GAGTTCTGTCCCGACCTTC	117	(ATG)8	NED	M3	AJ308802*	Poncet et al. 2004
Mg_M257	GACCATTACATTTACACAC	GCATTTTGTGTCACACTGTA	103	(AC)8	6-FAM	M1	AJ250257*	Poncet et al. 2004
Mg_M312	TTGCGTAAAGAGAACGAGCA	GTGTGAAAATTCACCTGAGC	106	(TG)9 (AAAT)4	VIC	M5	AM408740*	Poncet et al. 2007
Mg_M344	CTGAGATGAAAAGGAAAAGG	TGAAAAAGAGCCAATCAAGT	132	(GT)9	6-FAM	M2	AM408766*	Poncet et al. 2007
Mg_M348	TGGCTATGGTGTGAAGAACT	GATGCGATCATATTCCATTGT	92	(AC)10	NED	M5	AM408770*	Poncet et al. 2007
Mg_M356	TGAAGTCAACCTGAATACCAGA	ACGCACGCACGAATG	175	(TG)14	NED	M5	AM231553*	Poncet et al. 2007
Mg_M364	AGAAGAATGAAGACGAAACACA	TAACGCCCTGCCATCG	90	(A)21	6-FAM	M1	AM231556*	Poncet et al. 2007
Mg_M370	ATTGCCAGAGCATTGAAA	TGCAGTTGTAGTTAATCTCTCCT	93	(CA)8(GA)5	NED	M2	AM408601*	Poncet et al. 2007
Mg_M441	TTAGCAATACAAATAGCCTTCC	GCATTTAGCCAGTTCGTGT	189	(AC)9	6-FAM	M2	AM408664*	Poncet et al. 2007
Mg_M449	GGTGGAACCGAAGATTGAG	CTTTGCCTCTAAAGTTCTTGA AAC	99	(AC)8 (CA)9	VIC	M1	AM408670*	Poncet et al. 2007
Mg_M486	GGTATCGGTGGAGAAAGTATG	CCAGCTATCAATGAAGACGA	136	(AC)8 (TA)5	NED	M2	AM408700*	Poncet et al. 2007
Mg_M494	ACTTTGTGGAGTAATTGACAGG	CCATTTTGGATTACATACGG	102	(AC)8	NED	M1	AM408708*	Poncet et al. 2007
Mg_M497	TGTTTGA AAAATCCCATATCATAGA	TGGGCTCACATACGTTAATTG	101	(TG)7	6-FAM	M3	AM408710*	Poncet et al. 2007
Mg_M524	TCATGTCTGAATTTCTATTCTCTT	ACTTCTTTCCCTCTTATCACC	115	(TG)10	NED	M4	AM408735*	Poncet et al. 2007
Mg_M742	GGCTTCTGGGTGTCTGTGT	CCATTGGCTTTGTATTTCTGG	110	(TG)7(AG)6	VIC	M3	AJ308742*	Poncet et al. 2004
Mg_M779	TCCCCATCTTTTCTTTCC	GGGAGTGTTTTGTGTTGCTT	116	(TG)17	6-FAM	M5	AJ308779*	Poncet et al. 2004
Mg_M783	CCAACCTCGTATGGTTGTCTG	TGATAGGAGGCACTTGACACA	121	(TA)5(TG)17	PET	M3	AJ308783*	Poncet et al. 2004
Mg_M850	AAGCCGTTTCAAGCCATA	GCATTTTCTGCTTCTGTCTC	159	(TA)6(TG)11	6-FAM	M4	AJ308850*	Poncet et al. 2004

(*) accession number in EMBL-EBI database (European Bioinformatics Institute, <http://www.ebi.ac.uk>); (**) in SGN database (Sol Genomics Network, <http://solgenomics.net>)