

1 **TITLE :**

2 **Population genomics of the widespread African savannah trees *Afzelia africana* and**
3 ***Afzelia quanzensis* (Caesalpinoideae, Fabaceae) reveals no significant past**
4 **fragmentation of their distribution ranges**

5 **AUTHORS :**

6

7 Armel S.L. Donkpegan^{1,2,3*}, Rosalía Piñeiro^{4,5}, Myriam Heuertz⁶, Jérôme Duminil^{2,7,8}, Kasso
8 Daïnou^{1,2,9}, Jean-Louis Doucet¹ and Olivier J. Hardy²

9

10

11 **AFFILIATIONS :**

12

13

14 ¹ University of Liège, Gembloux Agro-Bio Tech, Management of Forest Resources, Tropical
15 Forestry, TERRA, 2 Passage des Déportés, B-5030 Gembloux, Belgium

16

17 ² Evolutionary Biology and Ecology Unit, CP 160/12, Faculté des Sciences, Université Libre de
18 Bruxelles, 50 avenue F. D. Roosevelt, B-1050 Brussels, Belgium

19

20

³ UMR 1332 BFP, INRA, Université de Bordeaux, F-33140, Villenave d'Ornon, France

21

22

⁴ University of Exeter, Geography, College of Life and Environmental Sciences, Stocker road,
EX44QD, Exeter, UK

23

24

25

⁵ Evolutionary Genomics, Centre for Geogenetics - Natural History Museum of Denmark, Øster
Voldgade 5-7, 1350 Copenhagen K, Denmark

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

Corresponding author:

*Armel S.L. Donkpegan: armel.donkpegan@gmail.com

44 **ABSTRACT**

45

46 Few studies have addressed the evolutionary history of tree species from African savannahs at
47 large geographic scales, particularly in the southern hemisphere (Zambezian region). *Afzelia*
48 (Fabaceae: Caesalpinoideae) contains economically important timber species, including two
49 species widely distributed in African savannahs: *A. africana* in the Sudanian region and *A.*
50 *quanzensis* in the Zambezian region. To characterize the population genetic diversity and
51 structure of these two species across their distribution ranges, we used nuclear microsatellites
52 (simple sequence repeats, SSRs) and genotyping-by-sequencing (GBS) markers. Six SSR loci
53 were genotyped in 241 *A. africana* and 113 *A. quanzensis* individuals, while 2,800 and 3,841
54 high-quality single nucleotide polymorphisms (SNPs) were identified in 30 *A. africana* and 12 *A.*
55 *quanzensis* individuals, respectively. Both species appeared to be outcrossing (selfing rate ~
56 0%). The spatial genetic structure was consistent with isolation-by-distance expectations based
57 on both SSR and SNP data, suggesting that gene dispersal is spatially restricted in both species
58 ($b_{Ld(SSR)} = -0.005$ and -0.007 and $b_{Ld(SNP)} = -0.008$ and -0.006 for *A. africana* and *A. quanzensis*,
59 respectively). Bayesian clustering of SSR genotypes failed to identify genetic structure within
60 species. In contrast, SNP data resolved intraspecific genetic clusters in both species, illustrating
61 the higher resolving power of GBS at shallow levels of divergence. However, the clusters
62 identified by SNPs revealed low levels of differentiation and no clear geographical entities.
63 These results suggest that, although gene flow has been restricted over short distances in both
64 species, populations have remained connected throughout the large, continuous Savannah
65 landscapes. The absence of clear phylogeographic discontinuities, also found in a few other
66 African savannah trees, indicates that their distribution ranges have not been significantly
67 fragmented during past climate changes, in contrast to patterns commonly found in African
68 rainforest trees.

69

70 **Keywords:** SSRs; genotyping-by-sequencing; population structure; kinship; IBD; savannah;
71 *Afzelia*

72

73 **1. INTRODUCTION**

74

75 Studies on the population genetic structure of African trees have largely focused on rainforest
76 species (Hardy et al., 2013; Daïnou et al., 2014 & 2016; Duminil et al., 2015; Ikabanga et al.,
77 2017; Demenou et al., 2018; Monthe et al., 2018). In contrast, the evolutionary history of trees
78 from the drier Sudanian and Zambezian regions, situated respectively North and South of the
79 Guineo-Congolian rainforest (APPENDIX 5), is still largely undocumented. In these
80 phytogeographic regions, trees occur in savannah, woodlands, dry forests or gallery forests,
81 thus, in vegetation types that cover a wide range of density in tree cover. Therefore, we can
82 expect that the response to climate change and gene flow in these vegetation types differs from
83 those occurring in the rainforests. The climatic changes of the Pleistocene have had a significant
84 impact on the savannah vegetation; however, they did not necessarily lead to fragmentation as
85 usually assumed for the African rainforests (Maley, 1996). During the dry and cold glacial
86 periods, savannahs expanded in the tropical regions in detriment of rainforest, which survived in
87 fragmented refugia. At extreme latitudes the savannah lost ground to the advances of steppes or
88 desert (Lioubimtseva et al., 1998). Conversely, during the humid interglacial periods, savannahs
89 have been replaced by rainforests in the tropics, but were able to expand northwards and
90 southwards at extreme latitudes (Quézel, 1965; Lézine, 1989; Waller & Salzmann, 1999;
91 Salzmann et al., 2002; Vincens et al., 2006; Watrin et al., 2009). In the absence of evidence of
92 past fragmentation, we may expect that widespread savannah trees exhibit only weak or no
93 genetic discontinuities within species.

94

95 To our knowledge, only five savannah tree species have been genetically investigated in Africa
96 using population genetics approaches at large scales. Three of the species occur in the
97 Sudanian savannah (Northern Hemisphere): the shea tree *Vitellaria paradoxa* (Allal et al., 2011;
98 Logossa et al., 2011), the African Mahogany *Khaya senegalensis* (Sexton et al., 2015), and the
99 Locust Bean *Parkia biglobosa* (Lombo et al., 2018). The other two species exhibit a Sudano-
100 Zambezian distribution (Northern and Southern Hemispheres): the baobab, *Adansonia digitata*
101 (Tsy et al., 2009; Kyndt et al., 2009) and the Arabic gum species *Acacia senegal* (Odee et al.,
102 2012; Lyam et al., 2018). Within the Sudanian savannah, weak genetic structure was detected in
103 *K. senegalensis* and *A. digitata*, while moderate differentiation was found in *A. senegal*. For *V.*
104 *paradoxa* and *P. biglobosa* significant genetic structure was detected in the Sudanian savannah,
105 although in both cases large genetically homogenous clusters were spread in central west Africa
106 (Logossa et al., 2011; Lombo et al., 2018). Within the Zambezian domain, significant population

107 genetic structure was detected for *A. senegal*, but not for *A. digitata*. Regional studies of
108 *Syzygium guineense* (Zigelski et al., 2019; restricted to part of its Zambezian distribution) and *V.*
109 *paradoxa* in Ghana revealed very weak genetic structure (Lovett et. al, 2000).

110
111 *Afzelia* (Fabaceae) is a palaetropical genus represented by seven species in Africa, including
112 two savannah and four rainforest species, as well as one putative species which is currently
113 poorly characterized (Brummit et al., 2007). The genus also harbours four species in South-East
114 Asia (Donkpegan et al., 2014). The two African savannah species are widely distributed in Sub-
115 Saharan Africa and occur in allopatry (Donkpegan et al., 2014): *Afzelia africana* Sm. ex Pers
116 occurs in the Sudanian region (from Senegal to Sudan; Aubréville 1968; Geerling 1982) and
117 *Afzelia quanzensis* Welw. in the Zambezian region (from southern Somalia to northern South
118 Africa). Recently, it has been shown that the two savannah species are diploid, as opposed to
119 the rainforest species, which are tetraploid (Donkpegan et al., 2015). In a recent phylogenetic
120 study of African species of *Afzelia*, the genus was estimated to have emerged in open habitats
121 (woodland and savannah) during the early to mid-Miocene (c. 20 to 14.5 Ma), whereas *A.*
122 *quanzensis* and *A. africana* originated during the mid or late Miocene (c. 14.5 Ma to 8 Ma,
123 Donkpegan et al., 2017). African *Afzelia* species are intensively logged for their timber
124 (Donkpegan et al., 2014). However, the population genetic structure and evolutionary processes
125 within the savannah species have not been investigated at a large geographic scales, despite
126 the fact that genetic information may be useful for the development of sustainable management
127 strategies for conservation and timber production (Lowe and Allendorf, 2010). Nuclear Simple
128 Sequence Repeats (nSSR, also called microsatellites) markers revealed low genetic diversity in
129 populations of *A. quanzensis* at small spatial scales, suggesting a limited evolutionary potential
130 in this species (Jinga et al., 2016; Jinga & Ashley, 2018). In SE Asia large-scale population
131 genetic studies have been performed on *A. xylocarpa* using SSR markers (Pakkad et al., 2009;
132 Pakkad et al., 2014). In addition to genetic diversity, the spatial genetic structure between
133 individuals or populations can inform on the evolutionary processes operating in a species and
134 can thus be of interest for conservation management (Frankham et al., 2017). Among different
135 metrics used to estimate relatedness between individuals (Frankham et al., 2017), the kinship
136 coefficient is the most commonly used in tests for isolation by distance and to estimate the
137 spatial extent of gene flow.

138
139 Population genetics studies in tropical trees have mostly used SSRs. Recent technological
140 advances in high-throughput sequencing allow sequencing large portions of the genome in non-

141 model organisms at a reasonable cost, thus offering increased resolution for the characterization
142 of population genetic patterns and the inference of evolutionary processes (Ekblom & Galindo
143 2011). In this study, we use nuclear microsatellites (SSRs, Donkpegan et al., 2015) and single
144 nucleotide polymorphisms (SNPs) derived from Genotyping by Sequencing (GBS) to investigate
145 the population genomic processes in the two savannah species of *Afzelia* across their
146 distribution ranges. This study addresses the following questions: (1) Does the genetic variation
147 at large geographic scales reveal discrete gene pools and/or a pattern of isolation-by-distance
148 within each species? (2) Do species show contrasting levels of genetic diversity and effective
149 population size, or signatures of demographic change compatible with past bottlenecks and/or
150 population growth? Our main objectives are to: (i) estimate the genetic diversity and population
151 genetic structure of *A. africana* and *A. quanzensis*, using nuclear SSRs and SNPs, (ii)
152 characterize the relatedness pattern between individuals in each species to test for isolation by
153 distance, and (iii) understand the origin of these patterns using methods for demographic
154 inference. Using SNP data on widespread savannah species, this paper is one of the first
155 population genomic studies of tropical woodland trees distributed across western and southern
156 Africa.

157

158 2. MATERIALS AND METHODS

159

160 2.1. Study species

161 *Afzelia africana* (Detarioideae, Fabaceae) occurs in the Sudanian region both in dry savannah
162 and in dry forests (Aubréville, 1959, Ahouangonou et al., 1995; Gerard & Louppe, 2011). It can
163 also occur in semi-deciduous forests, but at very low densities (Satabié, 1994). It has a wide
164 ecological amplitude but it prefers areas with > 900 mm annual rainfall and grows at elevations
165 of up to 1400 m and can reach up to 20 m in height. The fruiting period lasts six to eight months
166 and fruits may persist on trees for the following six months (Bationo et al., 2001; Ouédraogo-
167 Koné et al., 2008). *Afzelia quanzensis* occurs in the savannahs of Zambezian region, from
168 Somalia to Angola and the north of South Africa. It has been reported in semi-deciduous coastal
169 forests in Kenya (Brummitt et al., 2007) but also in dry forests, usually in deep sandy soils and
170 also on rocky ridges (Jacana, 1997). The species is drought resistant but frost sensitive. It is a
171 deciduous, medium to large-sized tree, 12-15 m high (reaching 35 m under ideal conditions,
172 Coates-Palgrave, 2002). *Afzelia* species are hermaphrodite and pollinated by insects (e.g. bees,
173 Kato et al. 2008; Ariwaodo and Harry-Asobara 2015). They have large dehiscent woody pods
174 containing characteristic black and red seeds (Jacana, 1997; Gerhardt and Todd, 2009).

175 Squirrels predate the seeds while monkeys, rodents (*Proechimys spp.*), and birds (mainly
176 hornbills) act as dispersers (Van Wyk & Van Wyk, 1997; Gathua, 2000; Bationo et al., 2001;
177 Gerard & Louppe, 2011).

178

179 **2.2. Sampling and DNA extraction**

180 Plant tissue samples were collected directly in the field or in herbaria (National Herbarium of the
181 Netherlands (herbarium code WAG of the Index Herbariorum), the Botanical Garden of Meise
182 (BR) and Université Libre de Bruxelles (BRLU) in Belgium), recording the geographic
183 coordinates of individual sampling locations. Our sampling is representative of the known
184 distribution ranges of the two species, in the Sudanian and Congolian biogeographic regions for
185 *A. africana* and in the Somalian, Zambezian and South African regions for *A. quanzensis*
186 (APPENDIX 5). We sampled 241 *A. africana* individuals from 41 West and Central African
187 locations and 113 *A. quanzensis* individuals from 24 East African locations (APPENDIX 1 & 2).
188 Fresh cambium or leaves were silica-dried in the field to avoid DNA fragmentation. Total DNA
189 was extracted using the NucleoSpin plant kit (Macherey-Nagel, Düren, Germany) or the DNeasy
190 96 Plant Kit (QIAGEN, GMbH, Germany) for the fresh material. For herbarium material, a CTAB
191 protocol was used (Doyle and Doyle, 1987).

192 **2.3. Genotyping of SSRs and SNPs**

193 Six microsatellite markers isolated from *A. bipindensis* were amplified in two PCR multiplexes in
194 all samples according to a previously published protocol (Donkpegan et al. 2015). Amplified
195 fragments were separated on an ABI 3730 sequencer (Applied Biosystems, Lennik, The
196 Netherlands) and sized using the Genemapper software in comparison with the SYBR Safe
197 (Invitrogen, Merelbeke, Belgium) size standard.

198

199 Genotyping by Sequencing (GBS) was performed for a subset of individuals (39 *A. africana* and
200 14 *A. quanzensis* individuals) at the Institute for Genomic Diversity and Computational Biology
201 Service Unit at Cornell University (Ithaca, NY) according to a published protocol (Elshire et al.
202 2011). To select the best enzyme for the GBS protocol, one microgram of DNA of *Afzelia*
203 *bipindensis* was used to build test libraries using three different enzymes: ApeKI (4.5-base
204 cutter), EcoT22I and PstI (both 6-base cutters). Libraries were checked for appropriate fragment
205 sizes (<500bp) and distribution on an Experion automatic electrophoresis system (Bio-Rad,
206 Laboratories, CA, USA). The enzyme EcoT22I, giving appropriate fragment sizes (<500bp) was
207 selected. To limit the risk of uneven coverage across loci and samples when applying GBS data

208 to organisms with large genome sizes, we built and sequenced two independent libraries per
209 individual, or pooled several DNA extracts per individual. Before library construction, DNA
210 extracts were purified using a ZR-96 DNA Clean up kit (Zymo Research, Orange, CA), DNA
211 quality was checked on a 1.5% agarose gel and DNA quantity was measured with Qbit HS
212 (Invitrogen, USA). Overall, 95 GBS libraries were built corresponding to 154 DNA extractions
213 obtained from 53 individuals of *Afzelia*. All libraries were sequenced together on one Illumina
214 lane (HiSeq2000 San Diego, CA, USA), using 100-bp Single Read chemistry.

215
216 We used Sabre (<https://github.com/najoshi/sabre>) to demultiplex barcoded reads. After
217 demultiplexing, sequence quality was evaluated with FastQC version 0.11.15 (Andrews, 2010).
218 Low quality bases and adapter contamination were removed with TRIMOMATIC version 0.33
219 (Bolger et al., 2014) with the following options: ILLUMINACLIP 2:30:10, LEADING 3, TRAILING
220 3, SLIDINGWINDOW 4:15, MINLEN 36.

221 First, a *de novo* assembly of GBS reads was carried out (including sequence reads of tetraploid
222 African *Afzelia* species *A. bella*, *A. pachyloba* and *A. bipindensis*) using pyRAD v.3.0.2 software
223 (Eaton & Ree, 2013) to produce a catalogue of GBS loci (3749 contigs, approximate length of
224 100bp per contig). This catalogue was used as a reference for mapping the reads of all
225 individuals using BWA 0.7.5a-r405 (Li & Durbin 2009). The resulting alignments were converted
226 to BAM format and reads were realigned around indels using SAMtools 0.1.17 (Li et al. 2009).
227 The resulting BAM files were used as input for HaplotypeCaller algorithm of Genome Analysis
228 Toolkit (GATK) v3.7 with standard parameters, to detect polymorphisms in each sample into a
229 VCF format including SNPs and INDELs (DePristo et al. 2011). VCFtools
230 (<http://vcftools.sourceforge.net/>) was used to remove indel variation, and retain only biallelic
231 variants (SNPs) with $\leq 60\%$ missing data within each species.

232 **2.4. Data analysis**

233 **Population genetics parameters at geographic populations level:** In order to characterize
234 the diversity within each species at SSRs, we computed the allelic richness (N_a), the effective
235 number of alleles (N_{ae}) following Nielsen et al. (2003), the observed heterozygosity (H_o), the
236 expected heterozygosity (H_e), the inbreeding coefficient (F) and the genetic differentiation based
237 on allele identity with the statistic F_{ST} using SPAGeDi 1.5a (Hardy, 2015). Permutation tests
238 were used to test whether F or F_{ST} deviated from expectations of panmixia in SPAGeDi 1.5a
239 (Hardy, 2015). For these analyses, we considered for both species, only populations sampled for

240 a minimum of 5 individuals (Table 1). Null allele frequencies were estimated with INEST 1.0
241 (Chybicki & Burczyk, 2009), which also provided a corrected estimation of the inbreeding
242 coefficient F . The selfing rate (S) was estimated in local populations with the largest sample
243 sizes (samples ≥ 25 individuals – Table 1), based on the standardized identity disequilibrium
244 assuming a mixed mating model (i.e. a proportion s of selfing and $1-s$ of random outcrossing)
245 with standard errors (SE) estimated by jackknifing over loci (Hardy, 2015; David et al., 2007).

246 In order to characterize genomic diversity for each species for the GBS data, we computed
247 nucleotide diversity π (π), corresponding to the average number of nucleotide differences per
248 SNP site between pairs of sequences (Nei 1987), using DnaSP v. 5.10.01 software (Librado &
249 Rozas 2009).

250 **Population genetic structure:** For SSR data, we used the Bayesian clustering method
251 implemented in STRUCTURE 2.3.1 (Falush et al., 2003) to detect any putative genetic
252 discontinuities within *A. africana* and *A. quanzensis* separately. We ran STRUCTURE 10 times
253 for each number K of genetic clusters, using $K=1-5$. We ran 1,000,000 iterations after a burn-in
254 period of 100,000 iterations, using the admixture model with independent allele frequencies
255 between clusters, without considering the population of origin of each individual. We estimated
256 $\text{LnP}(K)$ and ΔK using the Evanno method (Evanno et al., 2005) implemented in STRUCTURE
257 HARVESTER (Earl and vonHoldt 2012) to obtain the most likely value of K . We also used an
258 alternative genetic clustering method implemented in the R package *tess3r* (Caye et al., 2016),
259 which takes into account spatial information (the sampling location of each individual) to derive
260 individual ancestry estimates. The default values of the program were used and each run
261 ($K=1-5$) was replicated 10 times. The optimal value of K was defined by the minimum of the
262 cross-entropy criterion.

263 For GBS-derived SNP data, we performed genetic clustering analysis using the sparse non-
264 negative matrix factorization (sNMF) software, implemented in the R package LEA (Fritchot et al.,
265 2014). We also computed a genetic covariance matrix for each species to perform principal
266 components analysis (PCA) using SMARTPCA (Patterson et al., 2006; Price et al., 2006)
267 implemented in the SNPRelate package (Zheng et al., 2012).

268 **Isolation by distance:** Under Wright's isolation-by-distance (IBD) model, the relatedness
269 between individuals and/or populations is expected to decay linearly with the logarithm of their
270 geographic distance on a two-dimensional scale (Hardy & Vekemans, 1999). To detect IBD
271 within each species at large scales for SSR and SNP data, we calculated the kinship coefficient

272 F_{ij} between individuals i and j using the estimator of Loiselle et al. (1995) implemented in
273 SPAGeDi (Hardy and Vekemans, 2002). Positive and negative F_{ij} values indicate whether
274 individuals are more related or less related than the average of two sampled individuals.
275 Pairwise F_{ij} values were regressed on the logarithm of pairwise geographic distance, $\ln(d_{ij})$, and
276 IBD was tested by comparing the regression slope b_{log} to its distribution obtained from 10,000
277 permutations of the spatial locations of individuals. To illustrate IBD patterns, F_{ij} values were
278 averaged over a set of distance classes (d) according to a geometric progression of eleven
279 boundaries (0–1, 1–2, 2–5, 5–10, 10–50, 50–100, 100–200, 200–500, 500–1000, 1000–2000
280 and >2000 km) for *A. africana* and five (2–5, 5–10, 10–300, 300–500, >500 km) for *A.*
281 *quanzensis* giving $F(d)$. We used the *Sp*-statistic (Vekemans & Hardy, 2004) to quantify the
282 strength of the spatial genetic structure: $Sp = -b_{log}/(1 - F_1)$, where F_1 is the mean F_{ij} between
283 neighbouring individuals [approximated by $F(d < 1 \text{ km})$ for the first distance class].

284 **Demographic inference:** Using SSR data, the demographic history of each species was
285 assessed with the ‘bottleneck’ statistic $T2$ implemented in BOTTLENECK 1.2.02 (Piry et al.,
286 1999). This statistic represents an average across loci of the deviation of the actual gene
287 diversity He from the gene diversity expected from the number of alleles in the population
288 assuming mutation-drift equilibrium in a population of constant size. If $T2 > 0$, the gene diversity
289 excess reflects a loss of rare alleles possibly caused by recent founder events (bottlenecks),
290 whereas population expansions almost always cause heterozygosity deficiency ($T2 < 0$, Cornuet
291 & Luikart 1996). Simulations of the coalescent process were performed under three different
292 mutation models: the infinite allele model (IAM), the stepwise mutation model (SMM), and the
293 two-phase model with 70% of single-step mutations (TPM). The two latter models are
294 considered to be more appropriate for SSR data. One thousand simulations were performed.
295 Significant deviation from equilibrium gene diversity was determined using the Wilcoxon signed
296 rank test, which is the most appropriate test when only few polymorphic loci are analysed (Piry
297 et al., 1999).

298 For the SNPs data, to test for departure from the standard neutral model (SNM) the mean value
299 of Tajima’s D (Tajima 1989) over loci was computed and compared with the distribution of mean
300 values from coalescent simulations using DnaSP v.5.10.1 (Librado & Rozas 2009). Tajima’s D
301 statistic measures the standardized difference between nucleotide diversity π and the Watterson
302 estimator θ per site (Watterson 1975). D is expected to be close to zero under the standard
303 neutral model of population evolution, e.g., under a constant size population. High values of
304 Tajima’s D suggest an excess of common variants, which can be consistent with balancing

305 selection at the locus level, or with population contraction when detected at the genome level.
306 Negative values of Tajima's D, on the other hand, indicate an excess of rare variation, consistent
307 with population growth when detected at the genome level, or with positive selection at the locus
308 level (Tajima1989).

309

310 **3. RESULTS**

311 **SSR-based genetic diversity within each species and selfing rate:** A total of 67 alleles were
312 detected over all six loci for *A. africana* and the mean number of alleles per locus was 11.17 and
313 ranged from 4 to 26 alleles. Observed and expected heterozygosity estimates per population
314 ranged from $H_O = 0.43$ to 0.69 and from $H_E = 0.44$ to 0.64, respectively (Table 1). For *A.*
315 *quanzensis*, a total of 42 alleles were detected over all six loci and the mean number of alleles
316 per locus was 7.0 and ranged from 2 to 23 alleles. Observed and expected heterozygosity
317 ranged from $H_O = 0.28$ to 0.41 and from $H_E = 0.46$ to 0.52. Inbreeding coefficients were not
318 significantly different from zero in all populations ($F = 0$) after correcting for null alleles using
319 INEST (APPENDIX 4). The estimated selfing rates S for three populations of *A. africana* (Lama,
320 Penessoulou1 and Pendjari) and two of *A. quanzensis* (Gede and Witu) were close to zero
321 (Table 1), except for Lama population (33%). F_{ST} statistics revealed low but statistically
322 significant differentiation among populations, with weaker genetic structure in *A. africana*, $F_{ST} =$
323 0.045 ($P < 0.01$) than in *A. quanzensis*, $F_{ST} = 0.078$ ($P < 0.01$).

324 **GBS-based SNP data:** After filtering to retain only biallelic SNPs, we obtained VCF files with
325 8541 SNPs for *A. africana* and 8730 SNPs for *A. quanzensis* using the GBS catalogue produced
326 for the genus *Afzelia*. These files were then filtered to retain polymorphic SNPs within each
327 species to remove SNPs and individuals with $\geq 60\%$ missing data. After applying all filters, we
328 removed nine individuals in *A. africana* and two in *A. quanzensis* and obtained VCF files
329 containing 2800 polymorphic SNPs and 30 individuals in *A. africana* and 3841 polymorphic
330 SNPs and 12 individuals in *A. quanzensis*. The final set of *A. africana* genotypes had an average
331 missing data rate of 13.64% per sample with a mean depth of 40X. For *A. quanzensis*, the
332 missing data rate was 28.48% per sample and 34X for mean depth. Total nucleotide diversity
333 was $\pi=0.00420$ and $\theta=0.01124$ in *A. africana*; $\pi=0.03326$ and $\theta=0.05094$ in *A. quanzensis*.

334
335 **Population genetic structure:** The STRUCTURE analyses of SSR data failed to detect
336 population genetic structure at the intraspecific level. For both species, $K=1$ received the

337 strongest support (APPENDIX 3). Runs assuming $K=2$ to $K=5$ revealed admixed ancestry
338 of individuals with similar contributions of genetic clusters. The inclusion of geographic prior
339 information using *tess3r* showed similar results, although *A. quanzensis* displayed somewhat
340 uneven contributions of genetic clusters suggesting weak population substructure (APPENDIX
341 6). Conversely, the two SNPs data showed some evidence of genetic structure. The number of
342 genetic clusters that best described the data was $K=3$ in *A. africana*, based on the criterion of
343 minimum cross entropy (Figure 1, APPENDIX 7a). In *A. africana* only two gene pools occurred
344 widespread across West Africa, without clear geographic pattern and many admixed individuals
345 between these gene pools. The third gene pool was centred on Nigeria. The PCA shows low
346 levels of genetic differentiation (variance explained by PC1 and PC2 are 7.10% and 5.90%,
347 respectively) and highlights the divergence of the Nigeria cluster (Figure 1C). In *A. quanzensis*,
348 cross entropy values decreased with increasing K up to the maximum number of K tested
349 ($K=10$) suggesting a stronger population genetic structure in this species (APPENDIX 7b). This
350 is confirmed by the PCA (PC1 and PC2 explain 18.32% and 12.79% of the variance,
351 respectively, Figure 1D). We chose to retain $K=2$ to represent the highest hierarchical level of
352 genetic structure; higher values of K revealed evidence for additional genetic structuring in the
353 sample. One cluster covered the north-eastern part of the sample range and was mostly
354 represented in coastal Kenya whereas the other one was widespread across the sample range
355 (Fig 1).

356 **Patterns of isolation by distance (IBD):** Pairwise kinship declined with increasing geographic
357 distance for both types of markers (Figure 2). In both species, kinship for the first distance class
358 (ca. 1000 m for *A. africana* and 5000 m for *A. quanzensis*) ranged around 0.05 for SSRs and
359 0.06 for SNPs, and quickly dropped with distance, indicating a signature of isolation by distance.
360 The regression slope b_{Ld} was significantly negative in *A. africana* ($b_{Ld} = -0.005$ for SSRs and -
361 0.008 for SNPs; both $P < 0.001$) and *A. quanzensis* ($b_{Ld} = -0.007$ for SSRs and -0.006 for SNPs
362 respectively; both $P < 0.001$). Overall, similar patterns of IBD were detected for both types of
363 markers in both species.

364
365 **Demographic inference in each species:** With SSRs, under the three models implemented in
366 BOTTLENECK (IAM, TPM and SMM), both species showed a negative value of T2 and a
367 significant heterozygosity deficiency ($P < 0.01$) for *A. africana* and for *A. quanzensis* after
368 Wilcoxon tests (Table 2). These results suggest absence of a recent bottleneck at the species
369 level for both species.

370 For GBS data, mean Tajima's D estimates were negative in both species, with values of -2.017
371 ($P < 0.05$) for *A. africana* and -1.598 (not significant) for *A. quanzensis*. These results are in
372 agreement with a signature of population expansion in *A. africana* at the species level, whereas
373 in *A. quanzensis*, the standard neutral model of constant species-level population size could not
374 be rejected (Table 2).

375

376 4. DISCUSSION

377

378 **Large-scale population structure:** Our results reveal a pattern of isolation by distance in the
379 two savannah representatives of the genus *Afzelia* in Africa -*A. africana* and *A. quanzensis*, i.e.,
380 the kinship between individuals decreased with spatial distance. It is worth noting that SSRs and
381 SNPs gave very similar IBD patterns despite large differences in the number of loci and
382 sampling strategies, as observed in previous studies (Yang et al., 2011). The IBD observed
383 probably reflects the limited movement of pollen and seeds. However, the mechanisms of
384 pollination and seed dispersal are not well known in *Afzelia*. The local movement of seeds would
385 be expected given the fact that the seeds of *Afzelia* are heavy and also given the observation
386 that small rodents act as dispersers (*Cricetomys emini*, *Epixerus wilsoni*, *Protoxerus stangeri*,
387 Bationo et al., 2001; Evrard, 2015). However, long-distance seed dispersers such as monkeys
388 (*Cercopithecus albogularis*) and birds -mainly hornbills- (Van Wyk & Van Wyk, 1997; Gathua,
389 2000) have been also reported. The pollination mechanism is even less studied. While large
390 *Xylocopa* bees (Kato et al., 2008) act as pollination agents in Asian *Afzelia*, their African
391 congeners *Apis mellifera scutellata* would not be able to transfer pollen beyond 3.2 km (Dick et
392 al., 2003).

393 While SSRs could not retrieve distinct genetic clusters across the natural range of *A. africana* in
394 the Sudanian savannah and *A. quanzensis* in the Zambezian savannah, SNP data revealed
395 genetic groups within species (particularly more pronounced *in the latter*). However, the genetic
396 clusters identified by SNPs exhibit high levels of admixture and do not correspond to any clearly
397 delimited geographic entities. This structure might thus reflect solely the trend of IBD rather than
398 a history of past population fragmentation. These observations suggest that gene flow has been
399 restricted but populations have remained connected throughout the large, continuous Sudanian
400 or Zambezian savannahs. The higher discriminating power of SNPs over SSRs for detecting
401 genetic clusters has also been reported previously (e.g. Liu et al 2005; Fischer et al., 2017).

402

403 Different scenarios were tested to reconstruct the demographic history of each species. SSR
404 and SNP data were again congruent in detecting signatures of population expansion. However,
405 our data were not powerful enough to identify if these signatures reflect range expansions (and
406 from which source) or only a demographic expansion without change of distribution. In any case,
407 populations of both savannah species apparently did not experience major disturbances leading
408 to their fragmentation as has been suggested in some other savannah species (Bryja et al.
409 2010; Odee et al. 2012, Sexton et al., 2015).

410
411 **Comparison with other tropical trees in Africa:** our results reveal no cut clear genetic
412 discontinuities over large distances in the Sudanian and the Zambezian savannahs for *A.*
413 *africana* and *A. quanzensis*, respectively, and are consistent with those observed in other
414 savannah tree species, namely *Adansonia digitata* and *Khaya senegalensis*, which showed no
415 geographic discontinuities of the genetic variation, and the moderate levels of differentiation
416 found in *Acacia senegal*. These results suggest that the African savannahs have not
417 experienced major upheavals leading to their fragmentation (Salzmann et al., 2002; Vincens et
418 al., 2006; Watrin et al., 2009) in contrast to the major fluctuations of the rainforest cover over
419 time (Maley, 1996). The cases of *Vitellaria paradoxa* (Allal et al., 2011; Logossa et al., 2011) and
420 *Parkia biglobosa* (Lombo et al., 2018), which show different geographic and genetic clusters in
421 the Sudanian region (but include large genetically homogenous clusters in central west Africa)
422 might be due to their high socio-economic importance in agroforestry systems in savannah
423 parklands because they all produce seeds that are marketed and widely used in human food.
424 Whether their genetic structures have been influenced by human activities remains an open
425 question.

426 In the last few years population genetic data have accumulated for a number of African
427 rainforest trees, indicating strong differentiation of the tree populations in Central and West
428 African rainforests for most of the tree species (Hardy, Born et al. 2013, Heuertz, Duminil et al.
429 2014). This genetic structuring cannot be explained by current geographic barriers such as the
430 main mountain chains (Cameroonian Volcanic Line, Cristal Mountains, and Chaillu massif) or
431 major rivers in the region (Sanaga, Dja, and Oougué river). Molecular dating suggests historical
432 isolation of the tree populations, probably led by rainforest fragmentation, during the cold and dry
433 Ice-Age periods of the Pleistocene (<2.58 Mya). These results contrast with the genetic
434 connectivity found for the *Afzelia* and other savannah tree species over large Sudanian and
435 Zambezian ranges.

436

437 **Local-scale genetic diversity with SSRs:** Inbreeding and selfing rates remain very low in adult
438 populations of *A. africana* and *A. quanzensis*. Gene diversity parameters for SSRs markers
439 showed a large range of local genetic diversity in our study (*A. africana* : H_E = 0.46–0.66 and *A.*
440 *quanzensis*: H_E = 0.40–0.66) and in other population-level studies of *A. quanzensis* from
441 Zimbabwe (H_E = 0.41–0.51; Jinga & Ashley, 2018), *A. africana* from Benin (H_E = 0.09–0.88;
442 Houehanou et al., 2019), and the Asian congener *A. xylocarpa* (H_E = 0.47–0.66; Packkad et al.,
443 2014). Comparable genetic diversity ranges were documented in the investigated African
444 savannah tree species *Khaya senegalensis* (H_E = 0.44–0.71; Sexton et al., 2015), *Vitellaria*
445 *paradoxa* (H_E = 0.42–0.62; Allal et al., 2011), *Acacia senegal* (H_E = 0.63–0.70; Omondi, et al.,
446 2010) and *Parkia biglobosa* (H_E = 0.61–0.82; Lombo et al., 2018). Much lower levels were
447 documented in *Adansonia digitata* (H_E = 0.27–0.35; Kyndt et al., 2009). Despite the high
448 influence of past climate changes and signature of forest fragmentation on rainforest tree
449 species no remarkably lower population genetic diversity is observed: *Aucoumea klaineana* (H_E
450 = 0.38–0.55; Born et al., 2008), *Milicia excelsa* (H_E = 0.53–0.56; Bizoux et al., 2009), *Baillonella*
451 *toxisperma* (H_E = 0.56–0.58; Ndiade-Bourobou et al., 2010), *Distemonanthus benthamianus* (H_E
452 = 0.47–0.58; Debout et al., 2011), *Greenwayodendron suaveolens* (H_E = 0.7–0.8; Piñeiro et al.,
453 2017), *Scorodophloeus zenkeri* (H_E = 0.50–0.60; Piñeiro et al., 2017), *Terminalia superba* (H_E =
454 0.51–0.81; Demenou et al., 2018).

455

456 **5. CONCLUSION**

457 The SSR and SNP-based data analyses of *Afzelia* species from the African savannahs have
458 shown that both species did not exhibit strong geographic barriers to genetic connectivity across
459 their Sudanian and Zambezian ranges, although isolation by distance patterns indicate restricted
460 gene flow. In this study, both markers provided overall congruent results, although SNP had
461 more resolution power than SSRs for population genetic structure analyses. Demographic
462 analyses with both SNPs and SSRs data suggested demographic expansion. Collectively, these
463 data demonstrate the strong influence that savannah ranges exert on genomic diversity, within
464 across their population range. Thus, there is consistent evidence for the signature of population
465 expansion beginning to accumulate in the genome of these savannah species; in contrast to
466 forest species, which show a long history of fragmentation in most of studied species in Guineo-
467 Congolian rainforest (Hardy et al., 2013).

468

469

470

471 **LITERATURE CITED**

472 Ahouangonou S. & Bris B. (1995) *Afzelia africana* Sm. *Flamboyant*, **42**, 7-10

473 Allal F., Sanou H., Millet L., Vaillant A., Camus-Kulandaivelu L., Logossa Z.A., Lefevre F. &
474 Bouvet J-M. (2011). Past climate changes explain the phylogeography of *Vitellaria*
475 *paradoxa* over Africa. *Heredity*, **107**, 174-186

476 Andrews S. (2010) FastQC: a quality control tool for high throughput sequence data.
477 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

478 Ariwaodo J.O. & Harry-Asobara J.L. (2015) Preliminary investigation on flowering and fruiting
479 pattern in a plantation grown *Afzelia africana* Sm stand in Umuahia, Nigeria. *Am. J. Plant.*
480 *Sci.*, **6**(01):219–227

481 Aubréville A. (1968) Légumineuses. Césalpinioidées. Flore du Gabon. Vol. 15. Paris : Museum
482 National d'Histoire Naturelle, 111-118

483 Aubréville A. (1959) *La flore forestière de la Côte d'Ivoire*. Vol.I. 2^e éd. révisée. Publication n°15.
484 Nogent-sur- Marne, France : Centre Technique Forestier Tropical

485 Bationo B.A., Ouédraogo S.J. & Guinko S. (2001) Longévité des graines et contraintes à la
486 survie des plantules d'*Afzelia africana* Sm. ex Pers. dans une savane boisée du Burkina
487 Faso. *Ann. For. Sci.*, **58**(1), 69-75

488 Bizoux J.P., Daïnou K., Bourland N., Hardy O.J., Heuertz M., Mahy G. et al. (2009). Spatial
489 genetic structure in *Milicia excelsa* (Moraceae) indicates extensive gene dispersal in a low-
490 density wind-pollinated tropical tree. *Molecular Ecology*, **18**, 4398–4408

491 Bolger A.M., Lohse M. & Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence
492 data. *Bioinformatics*, **30** (15), 2114–2120

493 Born C., Hardy O.J., Chevallier M.H. et al. (2008) Small scale spatial genetic structure in the
494 central African rainforest tree species *Aucoumea klaineana*: a stepwise approach to infer
495 the impact of limited gene dispersal, population history and habitat fragmentation.
496 *Molecular Ecology*, **17**, 2041–2050

497 Brummitt R.K., Chikuni A.C., Lock J.M. & Polhill R.M. (2007) Leguminosae, subfamily
498 Caesalpinoideae. In: Timberlake J.R., Pope G.V., Polhill R.M. & Martins E.S., eds. *Flora*
499 *Zambesiaca*. Volume 3, part 2. Kew, UK: Royal Botanic Gardens, 1-228

500 Bryja J., Granjon L., Dobigny G. et al. (2010) Plio-Pleistocene history of West African Sudanian
501 savanna and the phylogeography of the *Praomys daltoni* complex (Rodentia): the
502 environment / geography / genetic interplay. *Molecular Ecology*, **19**, 4783–4799.

503 Caye K., Deist T.M., Martins H., et al (2016) TESS3: fast inference of spatial population structure
504 and genome scans for selection. *Mol. Ecol. Resour.*, **16**:540–548

505 Chybicki I.J., Burczyk J. (2009) Simultaneous estimation of null alleles and inbreeding
506 coefficients. *J. Hered.*, **100**, 106–113

507 Coates Palgrave M. and Coates Palgrave K. (2002) Trees of Southern Africa (3rd ed), Struik
508 Publishers (Pty) Ltd., Cape Town

509 Cornuet J.M. and Luikart G. (1996) Description and power analysis of two tests for detecting
510 recent population bottlenecks from allele frequency data. *Genetics*, **144** (4), 2001-2014

511 Daïnou K., Mahy G., Duminil J., Dick C.W., Doucet J.-L., Donkpegan A.S.L., Pluijgers M., Sinsin
512 B., Lejeune P. and Hardy, O.J. (2014) Speciation slowing down in widespread and long-
513 living tree taxa: insights from the tropical timber tree genus *Milicia* (Moraceae). *Heredity*,
514 **113**, 74–85

515 Daïnou K., Blanc-Jolivet C., Degen B., Kimani P., Ndiade-Bourobou D., Donkpegan A.S.L.,
516 Tosso F., Kaymak E., Bourland N., Doucet, J.-L., (2016) Revealing hidden species diversity
517 in closely related species using nuclear SNPs, SSRs and DNA sequences—a case study in
518 the tree genus *Milicia*. *BMC Evol. Biol.*, **16**, 259

519 David P., Pujol B., Viard F. et al. (2007) Reliable selfing rate estimates from imperfect population
520 genetic data. *Molecular Ecology*, **16**, 2474–2487

521 Debout G.D., Doucet J.-L. & Hardy O.J. (2011) Population history and gene dispersal inferred
522 from spatial genetic structure of a Central African timber tree, *Distemonanthus*
523 *benthamianus* (Caesalpinoideae). *Heredity*, **106**, 88–99

524 Demenou B.B., Doucet J-L. & Hardy O.J. (2018) History of the fragmentation of the African rain
525 forest in the Dahomey Gap: insight from the demographic history of *Terminalia superba*.
526 *Heredity*, **120**, 547–561

527 Depristo M.A. et al. (2011) A framework for variation discovery and genotyping using next-
528 generation DNA sequencing data. *Nat. Genet.*, **43**, 491–498.

529 Dick C.W., Hardy O.J., Jones F.A. and Petit R.J. (2008) Spatial scales of pollen and seed-
530 mediated gene flow in Tropical Rain Forest Trees. *Tropical Plant Biology*, **1**, 20–33

531 Donkpegan A.S.L., Hardy O.J., Lejeune P., Oumorou M., Daïnou K. & Doucet J.-L. (2014) Un
532 complexe d'espèces d'*Afzelia* des forêts africaines d'intérêt économique et écologique
533 (synthèse bibliographique). *Biotechnol. Agron. Soc. Environ.* **18** (2), 233-246

534 Donkpegan A.S.L., Doucet J.-L., Daïnou K., & Hardy O.J. (2015) Microsatellite Development and
535 Flow Cytometry in the African Tree Genus *Afzelia* (Fabaceae, Caesalpinoideae) Reveal a
536 Polyploid Complex. *Applications in Plant Sciences*, **3**, 1400097

537 Donkpegan A.S.L., Doucet JL., Migliore J., Duminil J., Daïnou K., Rosalia P., Wieringa J.,
538 Champluvier D. & Hardy O.J. (2017). Evolution in African tropical trees displaying ploidy-
539 habitat association: the genus *Afzelia* (Leguminosae). *Mol. Phyl. Evol.*, **107**, 270-281.

540 541 Doyle J.J. & Doyle J. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf
tissue. *Phytochemical Bulletin*, **19**, 11–15

542 543 544 545 Duminil J., Mona S., Mardulyn P., Doumenga C., Walmacq F., Doucet J.-L., & Hardy O.J. (2015) Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis. *Journal of Biogeography*, **42** (8), 1443–1454

546 547 548 Earl D.A., Vonholdt B.M. (2012) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

549 550 Ekblom R. & Galindo J. (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15

551 552 553 Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**:e19379

554 555 Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620

556 557 Evrard Q. (2015) Ecologie de reproduction du doussié, *Afzelia bipindensis* Harms, en forêt dense humide tropicale gabonaise. Mémoire: ULiège/Gembloux Agro-Bio Tech, 22 p.

558 559 560 Ewédjè E.B.K., Ahanchédé A. & Hardy O.J. (2017) Breeding system, gene dispersal and small-scale spatial genetic structure of a threatened food tree species, *Pentadesma butyracea* (Clusiaceae) in Benin. *Conservation Genetics*, **18** (4), 799–811

561 562 563 Falush D., Stephens M. & Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587

564 565 566 567 Fischer M. C., Rellstab C., Leuzinger M., Roumet M., Gugerli F., Shimizu K. K., Holderegger R. and Widmer A. (2017) Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics*, 18:69

568 569 Fritchot E., Mathieu, F., Trouillon T., Bouchard G. & François O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, **196**, 973–983

570 571 Frankham, R., Ballou, J.D., Ralls, K., et al. (2017). Genetic management of fragmented animal and plant populations. Oxford University Press, Oxford, U.K.

572 573 Gathua M. (2000) The effects of primates and squirrels on seed survival of a canopy tree, *Afzelia quanzensis*, in Arabuko-Sokoke Forest, Kenya. *Biotropica* **32**, 127–132

574 575 576 Gérard J. & Louppe D., (2011) *Afzelia africana* Sm. ex Pers. In: Lemmens R.H.M.J., Louppe D. & Oteng- Amoako A.A., eds. Record from PROTA4U. Wageningen, The Netherlands: PROTA (Plant Resources of Tropical Africa), <http://www.prota4u.org/search.asp>, (09.04.14)

577 578 Gerhardt K. & Todd C. (2009) Natural regeneration and population dynamics of the tree *Afzelia*
quanzensis in woodlands in Southern Africa. *Afr. J. Ecol.*, **47**, 583–591

579 580 581 Guichoux E., Lagache L., Wagner S., Chaumeil P., Leger P., Lepais O., Lepoittevin C., Malausa
T., Revardel E., Salin F., et al. (2011) Current trends in microsatellite genotyping. *Mol. Ecol. Resour.*, **11** (4):591–611

582 583 Hardy O.J., Vekemans X. (2002) Spagedi: a versatile computer program to analyse spatial
genetic structure at the individual or population levels. *Mol. Ecol.*, **2**, 618–620

584 585 586 587 Hardy O.J., Born C., Budde K., Daïnou K., Dauby G., Duminil J., Ewédjé E-EB., Gomez C.,
Heuertz M., Koffi G.K. (2013) Comparative phylogeography of African rain forest trees: a
review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes
Rendus Geoscience*, **345**, 284–296

588 589 Hardy O.J. (2015) Population genetics of autopolyploids under a mixed mating model and the
estimation of selfing rate. *Molecular Ecology Resources*, **16** (1), 103–117

590 591 592 Houehanou T.D., Prinz K. & Hellwig F. (2019) Characterization of 15 nuclear microsatellite
markers for *Afzelia africana* (Fabaceae) and related species. *Applications in Plant Sciences*
7(5): e1249

593 594 595 596 Ikabanga D.U.D.U., Stévert T., Koffi K.G.G., Monthé F.K., Doubindou E.C.N., Dauby G., Souza
A., M'batchi B., Hardy O.J. (2017) Combining morphology and population genetic analysis
uncover species delimitation in the widespread African tree genus *Santiria* (Burseraceae).
Phytotaxa **321**:166–180

597 Jacana, J. (1997) Sappi Tree Spotting Lowveld. Jacana Education, Johannesburg, South Africa.

598 599 600 Jinga P., Palagi J., Ashley V.A. (2016) Development of microsatellite loci of pod mahogany,
Afzelia quanzensis (Fabaceae), by Illumina shotgun sequencing, and cross-amplification in
A. africana. *Appl. Plant. Sci.*, **4**(6):16000010

601 602 603 Jinga P. & Ashley M.V. (2018) A mountain range is a strong genetic barrier between populations
of *Afzelia quanzensis* (pod mahogany) with low genetic diversity. *Tree Genetics &
Genomes*, **14**:4

604 605 Kato M., Kosaka Y., Kawatika A. et al (2008) Plant-pollinator interactions in tropical monsoon
forests in southeast Asia. *Am. J. Bot.*, **95**(11): 1375–1394.

606 607 608 Kyndt T., Assogbadjo A.E., Hardy O.J., Glele Kakaï R., Sinsin B., Van Damme P. and Gheysen
G. (2009) Spatial genetic structuring of baobab (*Adansonia digitata*, Malvaceae) in the
traditional agroforestry systems of West Africa. *Am. J. Bot.*, **96** (5), 950–957

609 610 Lézine A.M. (1989) Late Quaternary vegetation and climate of the Sahel. *Quaternary Research*,
32, 317–334

611 612 Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows–Wheeler
transform. *Bioinformatics*, **25**, 1754–1760. doi:10.1093/bioinformatics/btp324

613 Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. & Durbin
614 R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25** (16),
615 2078–2079

616 Librado P. & Rozas J., (2009) DnaSP v.5: a software for comprehensive analysis of DNA
617 polymorphism data, *Bioinformatics*, **25**, 1451–1452.

618 Linder H. P., de Klerk H. M., Born J., Burgess N. D., Fjeldsa J. & Rahbek C. (2012) The
619 partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa.
620 *Journal of Biogeography*, **39**, 1189–1205

621 Liu N., Chen L., Wang S., Oh C. and Zhao H. (2005) Comparison of single-nucleotide
622 polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, **6**
623 (Suppl 1):S26

624 Logossa Z.A., Camus - Kulandaivelu L., Allal F., Vaillant A., Sanou H., Kokou K. and Bouvet J.
625 (2011) Molecular data reveal isolation by distance and past population expansion for the
626 shea tree (*Vitellaria paradoxa* C.F. Gaertn) in West Africa. *Molecular Ecology*, **20**: 4009–
627 4027. doi:10.1111/j.1365-294X.2011.05249.

628 Loiselle B.A., Sork V.L., Nason J. and Graham C. (1995) Spatial genetic-structure of a tropical
629 understory shrub, *Psychotria Officinalis* (Rubiaceae). *Am. J. Bot.*, **82**, 1420–1425

630 Lombo D., Vinceti B., Konrad H., Gaisberger H., and Geburek T. (2018) Phylogeography of
631 African Locust Bean (*Parkia biglobosa*) Reveals Genetic Divergence and Spatially
632 Structured Populations in West and Central Africa. *Journal of Heredity*,
633 doi:10.1093/jhered/esy047

634 Lovett P.N. and N. Haq. (2000) Diversity of the Sheanut tree (*Vitellaria paradoxa* CF Gaertn.) in
635 Ghana. *Genetic Resources and Crop Evolution*, **47**(3) : 293-304

636 Lowe W.H. & Allendorf F.W. (2010) What can genetics tell us about population connectivity?
637 *Mol. Ecol.* **19**: 3038–3051.

638 Lyam, Paul Terwase, et al. (2018) Genetic diversity and distribution of *Senegalia senegal* (L.)
639 Britton under climate change scenarios in West Africa. *PLoS one* 13.4 : e0194726.

640 Monthe F.K., Duminil J., Kasongo Yakusu E., Beeckman H., Bourland N., Doucet J-L., Sosef
641 M.S.M. & Hardy O.J. (2018) The African timber tree *Entandrophragma congoense* (Pierre
642 ex De Wild.) A.Chev. is morphologically and genetically distinct from *Entandrophragma*
643 *angolense* (Welw.) C.DC. *Tree Gen. & Genomes* **14**(5), doi:10.1007/s11295-018-1277-6

644 Ndiade-Bourobou D., Hardy O.J., Favreau B., Moussavou H., Nzengue E., Mignot A. and Bouvet
645 J.-M. (2010) Long-distance seed and pollen dispersal inferred from spatial genetic structure
646 in the very low-density rainforest tree, *Baillonella toxisperma* Pierre, in Central Africa.
647 *Molecular Ecology*, **19** (22), 4949–4962

648 Nielsen R., Tarpy D.R., Reeve H.K. (2003) Estimating effective paternity number in social
649 insects and the effective number of alleles in a population. *Mol. Ecol.*, **12**, 3157–3164

650 Odee D.W., Telford A., Wilson J., Gaye A., Cavers S. (2012) Plio-Pleistocene history and
651 phylogeography of *Acacia senegal* in dry woodlands and savannahs of sub-Saharan
652 tropical Africa: evidence of early colonisation and recent range expansion. *Heredity*, **109**
653 (6), 372–382

654 Omondi S.F., Kireger E., Dangasuk O.G., Chikamai B., Odee D.W., Cavers S. et al. (2010)
655 Genetic diversity and population structure of *Acacia senegal* (L.) Willd. in Kenya. *Trop.
656 Plant. Biol.*, **3**, 59–70.

657 Ouédraogo-Koné S., Kaboré-Zoungrana C.Y. & Ledin I. (2008) Important characteristics of some
658 browse species in an agrosilvopastoral system in West Africa. *Agrofor. Syst.*, **74**, 213-221

659 Pakkad G., Ueno S. & Yochimaru H. (2009) Isolation and characterization of microsatellite loci in
660 an endangered tree species, *Afzelia xylocarpa* (Kurz) Craib (Caesalpinoideae). *Molecular
661 Ecology Resources*, 9 (3): 880-882

662 Pakkad G., Kanetani S. & Elliott S. (2014) Genetic diversity and differentiation of an endangered
663 tree species, *Afzelia xylocarpa* (Kurz) craib in Thailand revealed by nuclear microsatellite
664 markers. *Afr. J. Biotechnol.*, **13**:366–377

665 Patterson N., Price A.L. & Reich D. (2006) Population structure and eigenanalysis. *PLoS Genet.*
666 **2**:e190. doi: 10.1371/journal.pgen.0020190

667 Piñeiro R., Dauby G., Kaymak E. and Hardy O. J. (2017) Pleistocene population expansions of
668 shade-tolerant trees indicate fragmentation of the African rainforest during the Ice Ages.
669 *Proc. R. Soc. B*, **284**, <http://doi.org/10.1098/rspb.2017.1800>

670 Piry S., Luikart G. & Cornuet J.M. (1999) BOTTLENECK: a computer program for detecting
671 recent reductions in the effective population size using allele frequency data. *Journal of
672 Heredity*, 90, 502 – 503

673 Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. & Reich D. (2006).
674 Principal components analysis corrects for stratification in genome-wide association
675 studies. *Nat. Genet.* **38**, 904–909. doi: 10.1038/ng1847

676 Quézel P. (1965) La végétation au Sahara. Du Tchad à la Mauritanie. *Gustav Fischer Verlag*,
677 Stuttgart, 333 p

678 R Development Core Team. R (2017) A language and environment for statistical computing.
679 Available: <http://www.R-project.org>

680 Salzmann U., Hoelzmann P. and Morczinek I. (2002) Late Quaternary climate and vegetation of
681 the Sudanian zone of Northeast Nigeria. *Quaternary Research*, **58**, 73–83

682 Satabié B. (1994) Biosystématique et vicariance dans la flore camerounaise. *Bull. Jardin Bot.
683 Nat. Belg.*, **63**, 125- 170.

684 Sexton G.J., Frere C.H., Kalinganire A., Uwamariya A., Lowe A.J. , Godwin I.D., Prentis P.J.
685 and Dieters M. J. (2015) Influence of putative forest refugia and biogeographic barriers on

686 the level and distribution of genetic variation in an African savannah tree, *Khaya*
687 *senegalensis* (Desr.) A. Juss. *Tree Genetics & Genomes*, **11**:103

688 Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA
689 polymorphism. *Genetics* **123**: 585–595

690 Tsy J.L., Lumaret R., Mayne D., Mohamed V.A., Abutaba Y.I.M., Sagna M., Raoseta S., Danthu
691 P. (2009) Chloroplast DNA phylogeography suggest a West African centre of origin for the
692 baobab, *Adansonia digitata* L (Bombacoideae, Malvaceae). *Mol. Ecol.*, **18**, 1707–1715

693 Van Wyk B. & Van Wyk P. (1997) Field Guide to Trees of Southern Africa. Struik Publishers,
694 Cape Town, South Africa.

695 Vekemans X. and Hardy O.J. (2004) New insights from fine-scale spatial genetic structure
696 analyses in plant populations. *Mol. Ecol.*, **13**, 921–935

697 Vincens A., Bremond L., Brewer S., Buchet G. and Dussouillez P. (2006) Modern pollen-based
698 biome reconstructions in East Africa expanded to southern Tanzania. *Review of*
699 *Palaeobotany and Palynology*, **140**, 187–212

700 Wardill T.J., Graham G.C., Zalucki M., Palmer W.A., Playford J. and Scott K.D. (2005) The
701 importance of species identity in the biocontrol process: identifying the subspecies of
702 *Acacia nilotica* (Leguminosae: Mimosoideae) by genetic distance and the implications for
703 biological control. *Journal of Biogeography*, **32** (12), 2145–2159

704 Yang X., Xu Y., Shah T., Li H., Han Z., Li J., et al. (2011) Comparison of SSRs and SNPs in
705 assessment of genetic relatedness in maize. *Genetica*, **139**, 1045–1054. doi:
706 10.1007/s10709-011-9606-9

707 Zheng X., Levine D., Shen J., Gogarten S.M., Laurie C. & Weir B.S. (2012) A high-performance
708 computing toolset for relatedness and principal component analysis of SNP data.
709 *Bioinformatics*, **28**, 3326–3328. doi: 10.1093/bioinformatics/bts60

710 Zigelski P., Rudolph B., Oldeland J., Lages F., Jürgens N., & Finckh M. (2019). The Tough, the
711 Wet and the Hidden: Evolutionary Strategies of a Polyploid Tropical Tree in a Changing
712 Environment. *Perspectives in Plant Ecology, Evolution and Systematics*, **38**, 1-12,

713 Waller M. & Salzmann U. (1999) Holocene vegetation changes in the Sahelian zone of NE
714 Nigeria: The detection of anthropogenic activity. *Palaeoecol. Africa*, **26**, 85–102

715 Watrin J., Lézine A-M. & Hély C. (2009) Plant migration and plant communities at the time of the
716 “green Sahara”. *Comptes Rendus Geoscience*, **341** (8-9), 656 – 670.

717 Watterson G.A. (1975) On the number of segregating sites in genetical models without
718 recombination. *Theor. Popul. Biol.*, **7**: 256–276.

719

720 **ACKNOWLEDGMENTS**

721
722 This work received financial support from the “Fonds pour la Formation à la Recherche dans
723 l’Industrie et l’Agriculture (FRIA-FNRS, Belgium)” through a research grant to A.D., from the
724 Marie Curie FP7-PEOPLE-2012-IEF program (project AGORA) awarded to R.P., from the Fonds
725 de la Recherche Scientifique (F.R.S.-FNRS) through project J.0292.17F, the Belgian Science
726 Policy (project AFRIFORD) and the CGIAR Research Program on Forests, Trees and
727 Agroforestry. This work has benefited from support of a grant from Investissement d’Avenir
728 grants of the ANR (CEBA:ANR-10-LABX-25-01). A.D. acknowledges a Labex COTE Mobility
729 grant to INRA. The authors are grateful to Nils Bourland who helped us during field expeditions,
730 through the project PD 620/11 Rev.1 (M): “Development and implementation of species
731 identification and timber tracking in Africa with DNA fingerprints and stable isotopes” by the
732 International Tropical Timber Organization (ITTO). We also thank the Botanic Garden of Meise
733 (BR-Herbarium, Belgium), ULB (BRLU-Herbarium) and Naturalis (WAG-Herbarium,
734 Netherlands) who provided us with material from their herbarium collections; and Esra Kaymak
735 and Tom Gilbert for their assistance in generating GBS data.

736

737 **DATA ACCESSIBILITY**

738 Microsatellite and GBS data are being submitted in DRYAD and GenBank’s Sequence Read
739 Archive respectively.

740

741 **AUTHOR CONTRIBUTIONS**

742 A.D., J-L.D. and O.H. conceived the study. A.D. collected the data and performed the analyses.
743 R.P. generated the GBS data sequencing. A.D., R.P., M.H., J.D., K.D., J-L.D., O.H. interpreted
744 the results, contributed to drafting and writing the article.

745

746 **CONFLICT OF INTEREST STATEMENT**

747 The authors declare no conflict of interest.

748

749

750 FIGURE LEGENDS

751

752 **Figure 1.** Genetic structure of African diploid *Afzelia* species using GBS-based SNPs (N = 30 *A. africana* with 2800 SNPs; N = 12 *A. quanzensis* with 3841 SNPs). **A)** Geographic origin of
753 samples and population genetic structure of *A. africana* at K=3 (Western Africa) and of *A. quanzensis* at K=2 (East and Austral Africa), where pie charts represent individual ancestry
754 proportions in the assumed populations, as estimated using sNMF. **B)** Histograms of individual
755 ancestry proportions for each species, as estimated using sNMF for K=2 to K=5 assumed
756 ancestral populations. **C - D)** PCA ordinations along the first two PCA axes of **(C)** *A. africana*
757 and **(D)** *A. quanzensis*, where symbols distinguish sNMF clusters (k0 represent samples not
758 assigned to a cluster at q>0.7).

761

762 **Figure 2.** Spatial genetic structures (kinship-distance curves) of *A. africana* (square) and *A. quanzensis* (triangle) based on SSRs (stippled lines) and SNPs (plain lines).

764

765

766 TABLE LEGENDS

767

768 **Table 1.** Genetic diversity parameters and selfing rate estimates in populations of two *Afzelia*
769 species. Number of genotyped trees (N), number of alleles per locus (N_a), effective number of
770 alleles (N_{ae}), expected (H_E) and observed (H_O) heterozygosity, inbreeding coefficient estimated
771 from heterozygote deficit ($F = 1 - H_O/H_E$), inbreeding coefficient estimated while accounting for null
772 alleles following the method implemented in INEst ($F_{(null)}$). *p < 0.05 indicates significant
773 deviation from HWE. NC indicates that no estimation was computed by INEst.

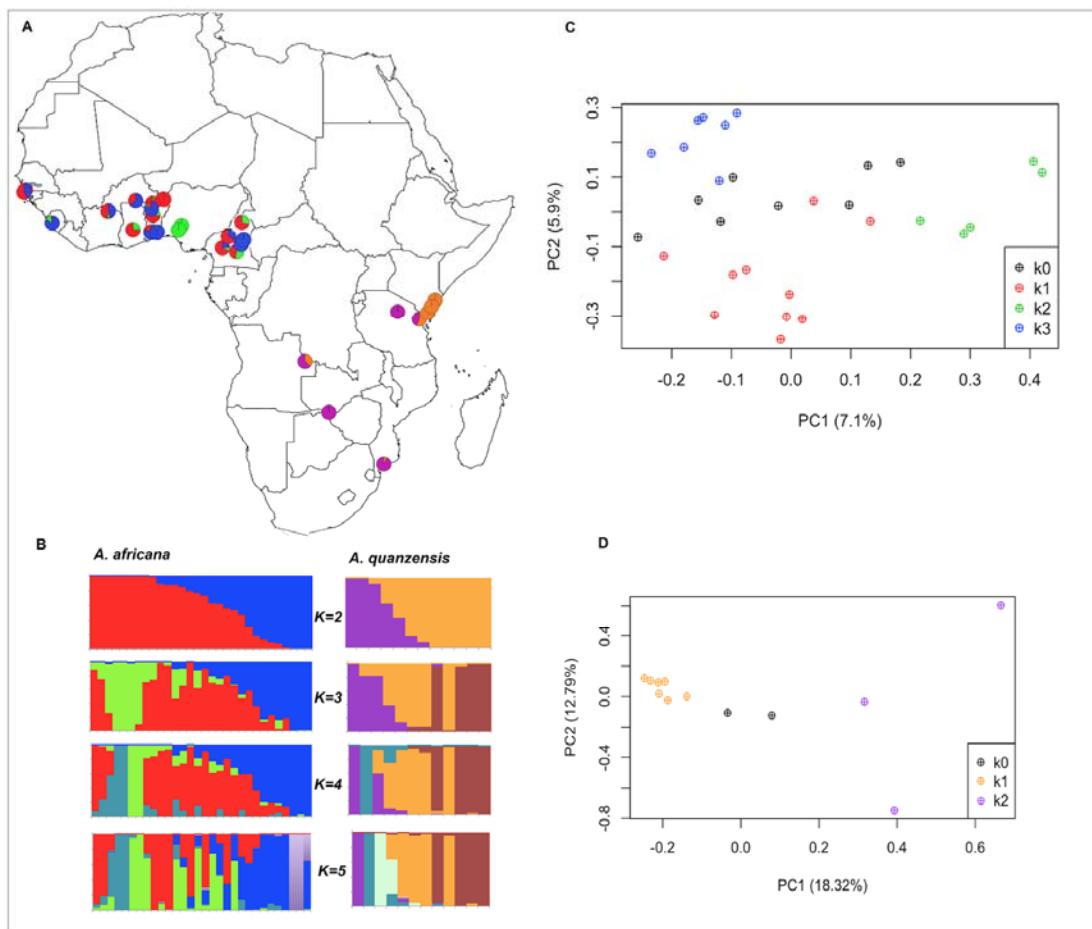
774

775 **Table 2.** An evaluation of alternative demographic models for the total population of both
776 species with SSR and SNP. T2 is the bottleneck statistic of different models; IAM, infinite allele
777 model; TPM, two-phase model; SMM, stepwise mutation model; SNM, standard neutral model;
778 n, number of individuals; π, nucleotide diversity and K1, K2 and K3 represent the genetic groups
779 defined for each species (see Figure 1). nc, not computed; ns: not significative; * P<0.05; **
780 P<0.01; *** P<0.001

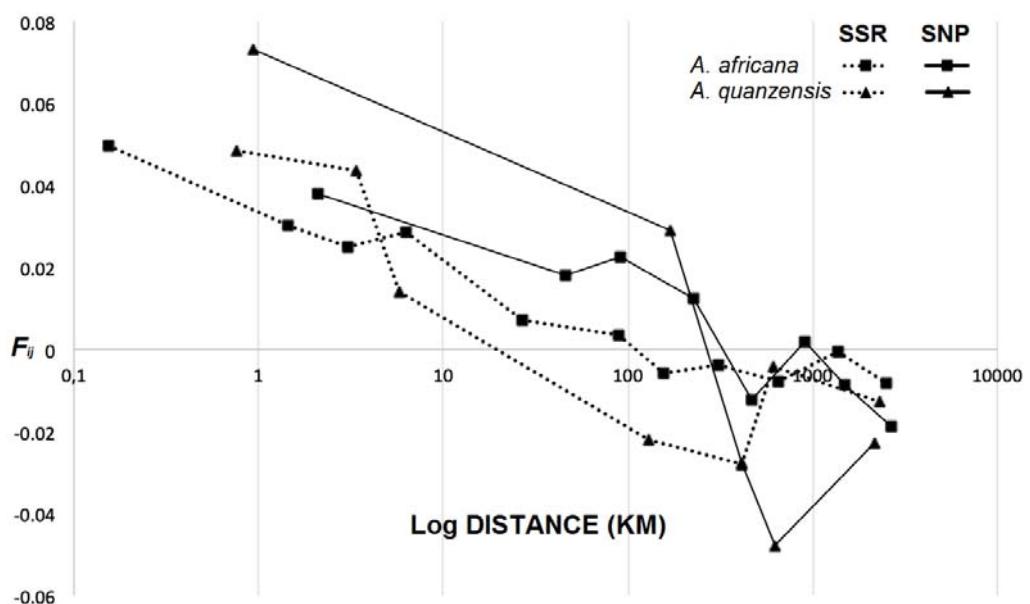
781

782

783 **Figure 1.**



784
785 **Figure 2.**



786

787

Table 1.

	Country	Populations	Latitude	Longitude	N	N_a	N_{ae}	H_E	H_o	F	$F_{(null)}$	Selfing (S)
<i>A africana</i>	Benin	BassilaS	2.4227	9.0037	7	3.00	2.53	0.459	0.595	-0.327*	0	☒
	Benin	BassilaN1	2.2886	9.0122	9	4.83	4.71	0.588	0.556	0.059	0	☒
	Benin	BassilaN2	2.2737	8.8221	11	4.33	4.28	0.516	0.470	0.094	0	☒
	Benin	BassilaN3	1.5556	9.2671	5	3.67	3.81	0.594	0.558	0.067	0	☒
	Benin	Lama	2.1141	6.97694	34	5.00	2.45	0.497	0.434	0.129	0	0.33±0.16
	Benin	Natitingou	1.3808	10.2785	9	4.50	3.78	0.561	0.444	0.218*	0	☒
	Benin	ParcW1	11.2962	5.5864	18	5.17	3.69	0.589	0.574	0.026	0	☒
	Benin	ParcW2	1.2886	6.9592	6	3.17	3.17	0.483	0.528	-0.105	0	☒
	Benin	Pendjari	2.9909	11.5101	25	7.17	4.35	0.641	0.693	-0.083	0	0
	Benin	Penessoulou1	3.0543	11.4788	32	5.67	3.30	0.483	0.458	0.052	0	0±0.1
	Benin	Penessoulou2	1.5242	10.9432	8	3.00	2.65	0.442	0.479	-0.092	0	☒
	Togo	Notse	1.5071	9.2856	12	4.17	2.99	0.525	0.528	-0.005	0	☒
	Cameroon	Ngambetica	1.6583	8.9988	7	3.83	3.08	0.566	0.500	0.125	0	☒
	Cameroon	Yoko	12.0073	5.4309	7	3.33	2.47	0.500	0.524	-0.052	0	☒
<i>A quanzensis</i>	Kenya	Gede	-3.3014	39.9965	31	5.50	3.45	0.473	0.409	0.139*	0	0±0.07
	Kenya	Witu	-2.3837	40.5212	48	6.00	3.23	0.521	0.411	0.212*	0	0±0
	DRC	Lubembe	-10.9166	22.5346	9	1.83	1.68	0.457	0.278	0.643*	NC	☒

Table 2.

	SSR			SNP		
	T2 model			SNM model		
	IAM	TPM	SMM	n	π	Tajima's D
<i>A. africana</i>	-4.5 ***	-10.8 ***	-21.1 ***	K1	10	0.007
				K2	5	0.006
				K3	7	0.022
				All	30	0.004
<i>A. quanzensis</i>	-4.4 ***	-8.2 ***	-14.1 ***	K1	2	nc
				K2	7	0.012
				All	12	0.033