Transferability, development of Single Sequence Repeat (SSR) markers and application to the analysis of genetic diversity and population structure of the African fan palm (*Borassus aethiopum* Mart.) in Benin

Mariano Joly Kpaténon[1,2,3], Valère Kolawolé Salako[2,4], Sylvain Santoni[5], Leila Zekraoui[3], Muriel Latreille[5], Christine Tollon-Cordet[5], Cédric Mariac[3], Estelle Jaligot [3,6] #, Thierry Beulé [3,6] #, Kifouli Adéoti [1,2] #*

Affiliations :

1 Laboratoire de Microbiologie et de Technologie Alimentaire (LAMITA), Faculté des Sciences et Techniques, Université d'Abomey-Calavi, Cotonou, Bénin

2 Biodiversité et Ecologie des Plantes (BDEP), Faculté des Sciences et Techniques, Université d'Abomey-Calavi, Cotonou, Bénin

3 DIADE, Univ Montpellier, IRD, Montpellier, France,

4 Laboratoire de Biomathématiques et d'Estimations Forestières (LABEF), Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, Cotonou, Bénin

5 AGAP, Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, Montpellier, France

6 CIRAD, UMR DIADE, Montpellier, France

# equal contribution as last authors

* corresponding author

Email: zoulade@yahoo.fr (KA)

Author contributions

Conceptualization: VKS, EJ, TB, KA

Funding acquisition: VKS, EJ, TB, KA

Methodology: all authors

Investigation: MJK, SS, ML, CT, KA, VKS

Data processing and analysis: MJK, KA, SS, CM, LZ, TB

Writing - Original Draft Preparation: all authors

Abstract:

In Sub-Saharan Africa, the fan palm *Borassus aethiopum* Mart. is an important non-timber forest product-providing palm that faces multiple anthropogenic threats to its genetic diversity. However, this species is so far under-studied, which prevents its sustainable development as a resource.

The present work is a first attempt at characterizing the genetic diversity of this palm species as well as its spatial structuration in Benin, West Africa. During a first phase we implemented a microsatellite markers-based approach relying on the reported transferability of primers developed in other palm species and found that, in disagreement with previously published results, only 22.5% of the 80 markers tested enabled amplification of African fan palm DNA and polymorphism detection was insufficient. During a second phase, we therefore generated a *B. aethiopum*-specific genomic dataset through high-throughput sequencing and used it for the *de novo* detection of potential microsatellite markers. Among these, 11 enabled polymorphism detection and were further used for analyzing genetic diversity in nine *B. aethiopum* populations.

Our results show that genetic diversity of Beninese fan palm populations is low, with an overall average expected heterozygosity (He) of 0.354. Moreover, the positive values of the fixation index (F) in populations from both the Central (Soudano-Guinean) and the Southern (Guinean) regions suggest limited gene flows. Our analysis show that sampled *B. aethiopum* populations are clustered into two groups, one spanning populations from both the Southern and most of the Central region, and the other including the Central population of Savè (which also has the highest He) and populations from the North.

In light of our results, we discuss the use of inter-species transfer vs. *de novo* development of microsatellite markers in genetic diversity analyses targeting under-studied species. We also suggest future applications for the molecular resources generated through the present study.

Introduction

Many plant species remain under-studied due to their low economic importance, complicated biology and/or the absence of available genome sequence information. Upon initiating a research project aimed at characterizing the genetic diversity of such a species, researchers may be confronted with the situation that some resources can be found in more or less distantly related taxa. In such cases, the first step is often to assess whether some of these resources, such as molecular markers, can be used to study the new species. Provided that the "source" species display enough genetic similarities to the "target" species and that marker transferability has been previously assessed, this first step may lead to quick progress in a cost-effective manner. In many instances, transferring markers between species is therefore seen as a smarter investment than developing and testing new markers, especially if the initial funding allocated to the project is scarce.

Over the last three decades, molecular markers have been widely used to study genetic variation among and within populations of various plant species [1–10]. Among the different types of markers that are available, microsatellites or Single Sequence Repeats (SSRs) are often selected because they are easy to use and their implementation has low resources (*i.e.* genomic, financial, lab equipment) requirements. As a result, they are markers of choice for the assessment of polymorphism among species, genetic structure within populations, phylogeny reconstruction, genetic mapping, evolutionary analysis, and molecular breeding [11–14]. However, the steps leading to the development of functional SSR markers, namely the initial identification of microsatellite loci, primer selection and assessment of amplification/polymorphism detection, require some prior knowledge of the genome of the target species and may prove to be expensive and time-consuming [13,15]. In order to overcome this difficulty, approaches relying on the transfer of SSR markers between species or genera have therefore been implemented. They have been successful in many instances, as documented across *Prunus* species and among members of the Rosaceae family [16,17]; between species of the *Hevea* genus and to other Euphorbiaceae [18]; among Lamiaceae [19]; among Legumes belonging to the *Vicia* genus [20] and from the *Phaseolus* genus to *Vigna* [21].

The African fan palm *Borassus aethiopum* Mart., also known as ron or toddy palm, is a dioecious species belonging to the Arecaceae family. It is widely distributed across West and Central Africa, where it is present as wild populations. The fan palm is classified as a non-timber forest products (NTFPs)-providing plant [22,23], since the different parts of the plant are used for various purposes by local populations: hypocotyls and fruits for food, fruit odor as shrew repellent, stipe for construction, roots and leaves for traditional medicine, leaves for crafts [24–28]

These multiple uses of products derived from *B. aethiopum* have put a strong anthropogenic pressure on the species, thus contributing to both fragmentations of its populations and their poor natural regeneration [24,29–32]. More specifically, the harvesting of *B. aethiopum* fruits for hypocotyl production and trade has become, over the last two decades, one of the most important household commercial activities associated with this species in Benin, West Africa [33]. Further fragmentation of the species' habitat has been observed as result of land clearing for agriculture or urban development [32,34,35]. As illustrated through similar examples in the literature [39,40], such phenomena may lead to restricted gene flow and ultimately, to loss of genetic diversity among *B. aethiopum* populations.

There is therefore an urgent need to define a sustainable management policy for *B. aethiopum* populations, in order to ensure its sustainable use. As a consequence, acquiring information on the genetic diversity of the species and on the spatial structuring of its populations is a major touchstone towards defining sustainable management actions. At the time of writing the present article, only a few chloroplastic sequences are publicly available for *B. aethiopum* through NCBI (https://www.ncbi.nlm.nih.gov/search/all/?term=borassus%20aethiopum). By contrast, abundant molecular resources, including genome assemblies or drafts, are available for model palm species such as the African oil palm *Elaeis guineensis* Jacq. [41], the date palm *Phoenix dactylifera* [42–44] and the coconut tree *Cocos nucifera* [42,43]. In each of these three palm species , large numbers of SSR markers have been identified and for a fraction of them, cross-species and cross-genera transferability tests among species belonging to the Palmaceae family have been performed [47–53]. In several instances [48–51,53], these tests included samples from the Asian relative of *B. aethiopum*, *B. flabellifer*.

In the present study, we describe how we first attempted to use SSR markers that had been identified in these other palm species for the analysis of genetic diversity in *B. aethiopum*. Then, in a second phase, we show how we performed a low-coverage sequencing of the fan palm genome in the aim of developing the first set of specific SSR markers targeting this species. We then used these to assess the genetic diversity of *B. aethiopum* populations in Benin, as a preliminary step towards more comprehensive studies.

Material and Methods

Plant material sampling and DNA extraction

Nine distinct populations of *B. aethiopum* separated by at least 50 km were selected from the three main climatic regions that are encountered in Benin (Fig 1): the Sudanian region in the North (four populations), the Sudano-Guinean region in the Centre (three populations) and the Guineo-Congolian region in the South (two populations). Additionally, among sampled populations, three were located in protected areas and six in farmlands. Within each population, we sampled young leaves from 20 male and female adult trees that were separated by at least 100 m, and stored them in plastic bags containing silica gel until further processing. The complete list of leaf samples and their characteristics is available in S1 Table.

Fig 1. Geo-climatic distribution of the Beninese *Borassus aethiopum* populations used in this study. Collection sites for each of the nine populations sampled are indicated; information for individual samples are available in S1 Table. The three main geo-climatic regions of Benin are (from the lighter- to the darker-colored): Guineo-Congolian, Sudano-Guinean and Sudanian, respectively.

Adapted from a map by the GingkoMaps project (http://www.ginkgomaps.com/), distributed under a Creative Commons Attribution (CC-BY) 3.0 license (https://creativecommons.org/licenses/by/3.0/).

Genomic DNA was extracted from 250 mg of leaves ground to powder under liquid nitrogen using the Chemagic DNA Plant Kit (Perkin Elmer, Germany), according to the manufacturer's instructions on a KingFisher Flex™ (Thermo Fisher Scientific, USA) automated DNA purification workstation. Final DNA concentration was assessed fluorometrically with the GENios Plus reader (TECAN) using bis-benzimide H 33258 (Sigma-Aldrich) as a fluorochrome.

Transferability of palms microsatellite markers: selection and amplification

We selected a total of 80 SSR markers from previous studies: 44 were developed on *Phoenix dactylifera*; 25 were identified in *Elaeis guineensis* and showed successful amplification on *Borassus flabellifer* and *P. dactylifera* DNA; and 11 from *Cocos nucifera*. The respective sequences and origins of these different primer sets are displayed in Table 1.

Table 1: Characteristics of the palm SSR markers tested for transferability on *B. aethiopum*.

Marker N°	Marker name	Primer	Sequence (5'-3' orientation)	Source palm species	Other palm cultivars or species tested for transferability	References
1	mEgCIR0230	F	CCCTGGCCCCGTTTTTC	*Elaeis guineensis*	*Elaeis oleifera* *Syragus sp.* *Cocos nucifera* *Phoenix roebelinii* *Phoenix canariensis* *Phoenix reclinata*	[51](
		R	AGCGCTATATGTGATTCTAA			
2	mEgCIR0326	F	GCTAACCACAGGCAAAAACA			
		R	AAGCCGCACTAACATACACATC			
3	mEgCIR0465	F	TCCCCCACGACCCATTC			
		R	GGCAGGAGAGGCAGCATTC			

4	mEgCIR0476	F	TTCCTCGGCCCCTTCTC			
		R	TCGCCGACCTTCCACTG			
5	EgCSSR-5781	F	TTCACGCTACTGATGGTTGG	*Elaeis guineensis*	*Borassus flabellifer*	[50]
		R	TCGATCCCTTCTCTGGAAAC			
6	EgCSSR-1461	F	GTCCTCTCCTACGCCTCCTC			
		R	ATGCGATCCGAGTTCAGAAG			
7	mEgCIR2332	F	GAAGAAGAGCAAAAGAGAAG	*Elaeis guineensis.*	*Borassus flabellifer*	[45,46]
		R	GCTAGGTGAAAAATAAAGTT			
	mEgCIR3295	F	TGCCTCCAGACAATCAC			
		R	GTAAGGCTTAACCAGATAAC			
9	mEgCIR3311	F	AATCCAAGTGGCCTACAG			
		R	CATGGCTTTGCTCAGTCA			
10	mEgCIR3413	F	AAAGCTATGGGGTGAAAGAT			
		R	TGGATAAGGGCGAGAAGAGA			
11	mEgCIR3477	F	CCTTCAAGCAAAGATACC			
		R	GGCACCAAACACAGTAA			
12	mEgCIR3592	F	GAGCCAAAACAGACTTCAA			
		R	ACCGTATATGACCCCTCTC			
13	mEgCIR3755	F	GCTCACCAAAAAGTGTTAAGTC			
		R	AGTTTCAACGGCAGGTATAT			
14	mEgCIR3788	F	TTGTATGACCAAAGACAGC			
		R	AGCGCAACATCAGACTA			
15	ESSR75	F	AGATGGTTGGAGATTTCATGGT	*Elaeis guineensis*	*Borassus flabellifer*	[45,46,48]
		R	AACTTGAGGGTGCCATTACAAG			
16	ESSR76	F	CCATACCAGCAGAAGAGGATGT			
		R	CTGAAGGTCATAGGGGTCTCTG			
17	ESSR82R	F	CCCTCGACACCCATAGTTATTT			
		R	CTCGATTTCTGGCCTCTCATAC			
18	ESSR332	F	AGTTAATGTGTCAGGGCCAGTT			
		R	CTTGGTTCACTTGGGTGTGTC			
19	ESSR553	F	ATAAATTGTGCGAGGGGAAAAC			
		R	AGATCCGCGACAGGTCTTAAC			
20	ESSR566	F	GTGTCATCAAATTCGGTCCTTT			
		R	CGGTTCTTCTGCTGCTCTACTT			
21	ESSR609	F	AGGCGGTGATGAAGATGAAG			
		R	CTCCTCTCAAACAGAGTGGGAT			
22	ESSR650	F	GCCTTTTCTGGTTAATGGACTG			
		R	GTTTGTCTATGGATGATTGTGAGG			
23	ESSR652	F	CATACCGTCACCACTCAGAAAC			
		R	GCCGTCATTCTACCAGTTGAG			
24	ESSR673	F	TTCTGGCTACGAGCATAAGGA			
		R	TCAATAACCCTGGCTAAACACA			
25	ESSR681	F	TCTGAATTGTCGGAGTGGC			
		R	CATCCTTGCGTAAACAAAAGAG			
26	CNZ34	F	CATGTCGATAATTATACCCAA	*Cocos nucifera*	*Borassus flabellifer* *Korthalsia laciniosa, Zalacca zalacca, Daemonorops kurzianus* *Calamus simplicifolia, C. mannan, C. thwaitesii, C. erectus, C. palustris*	[47,52]
		R	TGCAAATATGAATGCAAACAC			
27	CN2A5	F	AAGGTGAAATCTATGAACACA			
		R	GGCAGTAACACATTACACATG			
28	CNZ 12	F	TAGCTTCCTGAGATAAGATGC	*Cocos nucifera*	*B. flabellifer* *Phoenix dactylifera* *Elaeis guineensis* *Cocos nucifera*	[47,53]
		R	GATCATGGAACGAAAACATTA			
29	CNZ 24	F	TCCTAAGCTCAATACTCACCA			
		R	CGCATTGATAAATACAAGCTT			
30	CNZ 18	F	ATGGTTCAGCCCTTAATAAAC			
		R	GAACTTTGAAGCTCCCATCAT			
31	CNZ 42	F	TGATACTCCTCTGTGATGCTT			
		R	GTAGATTGTGGGAGAGGAATG			
32	CN2A4	F	CAGGATGGTTCAAGCCCTTAA			
		R	GGTGGAAGAGGGAGAGATTGA			
33	CAC 21	F	AATTGTGTGACACGTAGCC	*Cocos nucifera*	*Cocos nucifera* *B. flabellifer*	[54,55]
		R	GCATAACTCTTTCATAAGGGA			
34	CAC 71	F	ATAGCTCAAGTTGTTGCTAGG			
		R	ATATTGTCATGATTGAGCCTC			
35	CAC 84	F	TTGGTTTTTGTATGGAACTCT			
		R	AAATGCTAACATCTCAACAGC			
36	CN1H2	F	TTGATAGGAGAGCTTCATAAC	*Cocos nucifera*	*B. flabellifer* *Phoenix dactylifera*	[54]
		R	ATCTTCTTTAATGCTCGGAGT			
37	PdAG-SSR	F	TCTGATTTCGTTTACTTCTTAGGA	*Phoenix*	*Phoenix dactylifera*	[45]

		R	TTCATATTCAGTTGTCGGGTGTA	*dactylifera*		
38	mPdCIR015	F	AGCTGGCTCCTCCCTTCTTA			
		R	GCTCGGTTGGACTTGTTCT			
39	mPdCIR063	F	CTTTTATGTGGTCTGAGAGA			
		R	TCTCTGATCTTGGGTTCTGT			
40	mPdIRD1	F	CTCGGAAGGGTATGGACAAA	*Phoenix dactylifera*	*P. reclinata* *P. roebelenii* *P. rupicola T.* *P. theophrasti* *Hyphaene thebaica* *Livistona carinensis* *Chamaerops humilis*	[56]
		R	TTGCCTTCGACGTGGTAGTA			
41	mPdIRD3	F	CATTGATCCAACACCACCAC			
		R	GCCAAAACCAGCTCTGGTAAC			
42	mPdIRD4	F	TTGGTGGCCTTTCTCAGAGT			
		R	TGGGATCAAAGTAGGGTTGG			
43	mPdIRD5	F	CTATCAGGATGGGGGTGATG			
		R	ACCCATCTGCATAGCTCCAG			
44	mPdIRD7	F	TGCAATACGATGGCAGAGTC			
		R	CCTTGCAAGTTTTCCACACC			
45	mPdIRD8	F	CTATTGGGTCCCTTGGTGAG			
		R	TGACTGCTCGTCATCAGGTC			
46	mPdIRD10	F	ATGCGTTCATCTCCCTTGAG			
		R	GCTGCAAACATCATCCTCAC			
47	mPdIRD11	F	GAGTTGGAGGCAAAACCAGA			
		R	CCACAAAACCCTTGTCTTCC			
48	mPdIRD14	F	GAGGGGTTCACGTTTGTGTC			
		R	GCACCAAGCACAAGAGCAAT			
49	mPdIRD15	F	CCGAGTCTGGCGAAGTAAAC			
		R	CTCCCCTTCCTCATCCTCTC			
50	mPdIRD16	F	CTGTCCGATCGAATTCTGC			
		R	GGACATCTCTTTGCGGTCAT			
51	mPdIRD17	F	GTGGGAGAAACCCGAAGAAT			
		R	CTGCTGCCTCATCTGCATT			
52	mPdIRD20	F	TTGAATGGTCCCCTGTAGGT			
		R	GTCCCAGCATGATTGCAGTA			
53	mPdIRD22	F	GGCTGTATGGGAAAGACCTG			
		R	CCTGCTGCATATTCTTCGTG			
54	mPdIRD24	F	GCTCCTGCAGAACCTGAAAC			
		R	GGACATCACCGTCCAATTCT			
55	mPdIRD25	F	CACTGGAAATTCAGGGCCTA			
		R	CCCAATTTCTCAGCCAAGAC			
56	mPdIRD26	F	CCTCCAGTTCATGCTTCTCC			
		R	GAGCAGACCCGACAGACAAT			
57	mPdIRD28	F	GAAACGGTATCGGGATGATG			
		R	TTAACGACGCCGTTTCCT			
58	mPdIRD29	F	GGCTCCACCATCATTGACA			
		R	AACAGCATCGACTGCCTTCT			
59	mPdIRD30	F	GCAGATGGTTGAAAGCTCCT			
		R	CCCCATTAACAGGATCAACG			
60	mPdIRD31	F	GCAGGTGGACTGCAAAATCT			
		R	CTATTGGGGTGCTGATCCAT			
61	mPdIRD32	F	AAGAAGACATTCCGGCTGGT			
		R	GCGGGTGTGTGATATTGATG			
62	mPdIRD33	F	GGAGCATACAGTGGGTTTGC			
		R	CAGCCTGGGAATGAGGATAG			
63	mPdIRD35	F	CAGCCCCTTACTCAGACTGG			
		R	CCCATAAGCTGATTGTGCTG			
64	mPdIRD36	F	GACACGTTGACGATGTGGAA			
		R	CCATTGCTGTTGAGGAGGAG			
65	mPdIRD37	F	TTTCCTGCTCGAAAGACACC			
		R	CTTAGCCAGCCTCCACACTC			
66	mPdIRD40	F	GAGAGATGCGTCAGGGAATC			
		R	CCAGAATCTTCCAAGCAAGC			
67	mPdIRD42	F	GAGGCAAAACTATGGGAAGC			
		R	TTCACTGGAGCAAGGGTAGG			
68	mPdIRD43	F	GCAGCCATTGCTTACAGTGA	*Phoenix dactylifera*	*P. reclinata* *P. roebelenii* *P. rupicola* *P. theophrasti* *Hyphaene thebaica* *Livistona carinensis* *Chamaerops humilis*	[56]
		R	TAAACTGCTGCCTTCCTTGG			
69	mPdIRD44	F	CAGATCCGGGAGATGATGAA			
		R	AGCAGGAGCAGCTGCATAA			
70	mPdIRD45	F	TAGCCTGTGCATGTTCGTTG			
		R	AACAGCAGCTGATGGTGATG			
71	mPdIRD46	F	ATGGGTCCATTGGAGGAACT			
		R	GACGGAGACCTTGACTGCTC			
72	mPcCIR10	F	ACCCCGGACGTGAGGTG	*Phoenix dactylifera*	*Phoenix dactylifera*	Cherif, Castillo and Aberlenc-Bertossi, unpublished data.
		R	CGTCGATCTCCTCCTTTGTCTC			
73	mPcCIR20	F	GCACGAGAAGGCTTATAGT			

		R	CCCCTCATTAGGATTCTAC			
74	mPcCIR32	F	CAAATCTTTGCCGTGAG			
		R	GGTGTGGAGTAATCATGTAGTAG			
75	mPcCIR35	F	ACAAACGGCGATGGGATTAC			
		R	CCGCAGCTCACCTCTTCTAT			
76	mPcCIR50	F	CTGCCATTTCTTCTGAC			
		R	CACCATGCACAAAAATG			
77	mPcCIR57	F	AAGCAGCAGCCCTTCCGTAG			
		R	GTTCTCACTCGCCCAAAAATAC			
78	mPcCIR85	F	GAGAGAGGGTGGTGTTATT			
		R	TTCATCCAGAACCACAGTA			
79	mPdIRD41	F	ATCTTCCATGCAGCCTCAAG			
		R	CAGGTCGTCCCGTCTCTAAA			
80	mPdIRD47	F	GTTGGCATCACTTCAGAGCA			
		R	GCTCTTTCGGTGCTAGTTGC			

F, R: forward and reverse primers, respectively.

Transferability of the 80 palm SSR markers was assessed on a representative subset of 20 of *B. aethiopum* individuals sampled in the different populations, plus 4 positive controls from each source species for these markers (*P. dactylifera, C. nucifera, E. guineensis*). Microsatellite amplification was performed with a modification of the M13-tailed Primers protocol (Boutin-Ganache et al., 2001) adapted to the use of fluorescent labelling. The PCR reaction was performed on 20 ng of leaf DNA and a mix of 1X PCR buffer, 200 µM dNTP, 2 mM MgCl2, 0.4 pmol M13-tailed forward primer fluorescently labeled in 5' with FAM, HEX or TAMR, 4 pmol reverse primer, and 0.5 U of *Taq* polymerase (Sigma), in a final volume of 20 µl. The following program was used: 3 min of initial denaturation at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 50°C and 72°C for 1 min and a final extension at 72°C for 5 min. The resulting amplification products were then diluted to 1/10th mixed with 0.5 µL of an internal size standard (GeneScan 500 ROX, Thermo Fisher Scientific) and denatured for 5 minutes at 94°C prior to capillary electrophoresis (Applied Biosystems 3500 Genetic Analyzer, Thermo Fisher Scientific).

De novo identification of microsatellite loci in the *B. aethiopum* genome, marker selection and diversity analysis

One *B. aethiopum* leaf sample was randomly selected and genomic DNA purification was performed according to the protocol of [57]. This DNA extract was then used for the construction of an Illumina paired-end library, as described in [58], before high-throughput sequencing on a MiSeq v3 Illumina platform. Demultiplexing of the raw data output was performed using the Maillol script (https://github.com/maillol/demultadapt), with a 0-mistmatch threshold. Adapters were eliminated using Cutadapt v1.10 (Martin, 2011) (http://code.google.com/p/cutadapt/) with the following parameters: overlap length = 7, minimum length = 35 and quality = 20. High-quality reads (Q > 30) were filtered using the following script: https://github.com/SouthGreenPlatform/arcad-hts/blob/master/scripts/arcad_hts_2_Filter_Fastq_On_Mean_Quality.pl and the resulting filtered reads were deposited into GenBank SRA under accession number PRJNA576413. Paired-end reads were then merged using FLASH v1.2.11 (https://github.com/SouthGreenPlatform/arcad-hts/blob/master/scripts/arcad_hts_3_synchronized_paired_fastq.pl). Finally, microsatellite motif detection and specific primer design were carried out after elimination of redundant sequences using the QDD v3.1.2 software [59] with default settings.

Using selected primer pairs, test amplifications were performed with two randomly selected fan palm DNA samples, then primers showing successful amplification were further tested for polymorphism detection among seven randomly selected DNA samples. The M13 Tailed Primers protocol described previously was used, with the following program: 3 min of initial denaturation at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 55°C and 72°C for 1 min and a final extension at 72°C for 5 min. PCR products visualization was performed as previously indicated. Finally, the primer pairs enabling successful amplification of polymorphic, mono-locus bands were used for the analysis of genetic diversity among the complete set of 180 *B. aethiopum* individuals with the same conditions.

Data analysis

Amplification products were scored using the GeneMapper software V3.7 and only unambiguous amplification products were considered for data analysis. Genetic parameters such as total number of alleles, allelic frequency, expected heterozygosity (He), observed heterozygosity (Ho), were calculated for each locus and each population in the GenAIEx software Version 6.502 [60]. The F-statistics analysis assessing genetic differentiation and the Analysis of MOlecular VAriance (AMOVA) for estimation of genetic differentiation within and among populations were performed with the same software.

A Principal Coordinates Analysis (PCoA) was also performed using GenAIEx software to enable the visualization of genetic variation distribution across the individuals under study. We used the STRUCTURE software version 2.3.4 [62] for the determination of the most probable number of clusters for population structure (K value). Using the admixture model, eight simulations were performed for each inferred K value, with a running length composed of 300,000 burning periods and 50,000 Markov chain Monte Carlo (MCMC) iterations to allocate accessions to different populations. The output from this analysis was then used as input in the Structure HARVESTER online program to determine the exact [62]. Based on this value, a clustering analysis of the studied populations was performed and using the genetic distance matrix obtained from previous analysis, a dendrogram was constructed with the DendroUPGMA program accessible online at http://genomes.urv.cat/UPGMA/ [63]

Results

Assessment of palm SSR markers transferability to *B. aethiopum* and evaluation of their capacity for characterizing genetic diversity

Of the 80 microsatellite markers that have been selected from three palm species and tested for amplification on African fan palm DNA, 18 (22.5 %) generate amplification products (Table 2). No amplification is observed using the 11 *C. nucifera* markers, whereas 7 (15.9 %) and 11 (44%) of the *P. dactylifera* and *E. guineensis* markers, respectively, show a successful amplification. None of the amplification products generated with date palm primers display genetic polymorphism in our *B. aethiopum* test panel. Among oil palm-derived SSR markers however, two, namely ESSR566 and ESSR652, display polymorphism. However, it must be noted that the ESSR566 primer pair amplifies two distinct loci, ESSR566A and ESSR566B.

Table 2 : Summary of SSR markers transferability assessment

Species of origin	Number of SSR markers	Number of successful amplification (% of markers)	Number of polymorphic amplicons (% of amplifications)
Cocos nucifera	11	0 (0)	0 (0)
Phoenix dactylifera	44	7 (15.9)	0 (0)
Elaeis guineensis	25	11 (44.0)	2 (18.2)
Total	**80**	**18 (22.5)**	**2 (11.1)**

Overall, during this phase of the study we detect polymorphism in our *B. aethiopum* test panel with only 2 (11.1% of successfully amplified markers, 2.5% of total) of the palm SSR primer pairs that have been assayed. Only one of these markers proves to be both polymorphic and monolocus in the African fan palm, and might therefore be used for studying genetic diversity in this species.

De novo identification of microsatellite sequences in the *B. aethiopum* genome and assessment of potential SSR markers

In order to enable a more precise evaluation of genetic diversity among *B. aethiopum* populations, we developed specific *B. aethiopum* markers from *de novo* sequencing data. A total of 23,281,354 raw reads (average length 250 bp) have been generated from one MiSeq run. Raw sequence reads have been trimmed and generated 21,636,172 cleaned-up reads, yielding 493,636 high-quality reads after filtering ($Q > 30$) from which 216,475 contigs have been assembled.

From this latter output, the QDD software identifies a total of 1,630 microsatellite loci (see S2 Table), of which 81.41 % are perfect (*i.e.* repeat size 4 bp or smaller and repeat number 10- 20). Among these, 83.86 % of loci are composed of di-nucleotidic repeat units, 13.06 % of tri-nucleotidic units, 2.39 % of tetra-nucleotidic repeats and 0.67 % of repeats with five nucleotides and over. From these, we have selected SSR markers composed of di- (AG) or tri- nucleotide repeats, using the following criteria for specific amplification of easily scorable bands: primer lengths ranging from 18 to 22 bp, annealing temperatures 55–60°C and predicted amplicon sizes 90-200 bp.

The characteristics of the 57 selected primer pairs and the results of the test amplifications are presented in Table 3. Successful amplification of *B. aethiopum* DNA is obtained in most cases (94.7%). However, 34 of the putative markers tested (63.0% of amplifying ones) show no polymorphism. The remaining 20 putative markers are polymorphic and among them, nine correspond to multiple loci. As a result, 11 putative African fan palm SSR markers (representing 20.4% of successful amplifications and 55.0% of polymorphic markers in our study) are both polymorphic and mono-locus in our amplification test panel and may therefore be used for further analyses.

Table 3: List of selected primer pairs targeting putative *B. aethiopum* microsatellite loci and assessment of their polymorphism detection ability

Locus name	Repeat motif	Primers	**Sequence (5'-3' orientation)**	Expected amplicon size (bp)	Amplification product
MBo01	[AGG]7	F	CCTATCCTTCCATCCCGATCG	90	multiple loci, polymorphic
		R	TTGCCGTGAATCAGCCTCAA		
MBo02	[ATC]7	F	GGGAGAACAAGGATAACAGCAG	115	single locus, monomorphic
		R	TCCATTTCATCACTAGCTCGGT		
MBo03	[AGG]7	F	CTCCGAGCCCTAGCAACTTT	131	single locus, monomorphic
		R	TCTGGATGACGAAACCTTCACA		
MBo04	[ACC]7	F	GATGTGGCCGCTCTGATCTC	192	single locus, monomorphic
		R	ACATGCTGGCAAGGTATTCT		
MBo05	[AAG]7	F	GTCCTAGCACGCTGGCATTA	202	single locus, monomorphic
		R	TGGGTTGCCAATGAACCCTT		
MBo06	[ATC]7	F	TGGCCATTCAACTGCTTCAC	202	single locus, monomorphic
		R	GAATCTAGCACCAGCAAACCC		
MBo07	[AAG]7	F	GGCACTGGAGTCCACATCAA	239	single locus, monomorphic
		R	TCCTTCTGTACTGGCATCTCT		
MBo08	[AGG]8	F	TGATTGTTTCCTCTTCCCTCCT	90	single locus, monomorphic
		R	TTAATGAGCCGAAGAGGAGCC		
MBo09	[AGG]8	F	TCCCTCACTCCCATCCTCTC	163	single locus, monomorphic
		R	ACTCCACTCCTTCCCTCATACA		
MBo10	[AAC]8	F	GTTAAAGACGCAGGGCTGGA	166	single locus, monomorphic
		R	CCCACTTAGTGAGATAAGACTTGA		
MBo11	[ATC]8	F	GCATCACATGGTTTCAGGCT	219	single locus, monomorphic
		R	GCTCAACCATCGGCAGTGTA		
MBo12	[ATC]9	F	GGAGGAAAGGTTGCCCTAGAA	102	single locus, monomorphic
		R	TCTCAACCTGATGTCATTGCA		
MBo13	[AAG]9	F	CAGGTTGCATCGGCCCATT	103	multiple loci, polymorphic
		R	GGAGCCTAATGCACCCAGAG		
MBo14	[AAC]9	F	ATGGCCGATCCCACTTAGTG	117	single locus, monomorphic
		R	GAGAGAACGGCAATAATTTATGCA		
MBo15	[AAG]10	F	GCTGAAGAGGATGAAGAAGAAGC	92	multiple loci, monomorphic
		R	TCATCATCTCCCTCTCCTTCT		
MBo16	[AGG]10	F	CAGCACTGGCCTCACAGC	118	single locus, monomorphic

		R	CCGTCGATCAGTTGTTGGAGA		
MBo17	[ATC]10	F	ACACAATGACCTTTCGCTGA	124	single locus, monomorphic
		R	CCAAACAGGACCTTATGCCA		
MBo18	[AAG]10	F	ACATCCTCTCCTTCATCTCCTT	187	multiple loci, polymorphic
		R	GTTCCTACAATGCTTGGCGC		
MBo19	[AAG]10	F	TGCTATCACCCAATATCTAGGCT	202	single locus, monomorphic
		R	ACAGTCAACAACTACCATACTGC		
MBo20	[AAG]10	F	TGTGGTTAAAGCAATGGAAGCA	229	single locus, monomorphic
		R	GCCGAACTCCTACTCTCATACG		
MBo21	[AAG]11	F	ACAACAGAAGATCAGTATACGTTCT	171	single locus, monomorphic
		R	TTGAGGAATCATGCTTGTCAGT		
MBo22	[AAG]14	F	AGAAGAATTCGGTTAGGTCACAA	108	single locus, monomorphic
		R	AGATAACATGGGTAAGAATTGCCT		
MBo23	[AAT]5	F	TGAGTTCTTGTCTTGTCTTCGT	100	single locus, monomorphic
		R	GGTTTGGGACACCCTTCAGG		
MBo24	[AAT]9	F	AAAGTCATGTCTGGGTGATGAA	90	single locus, monomorphic
		R	ATGATGAGCACAGCTACAACTCT		
MBo25	[AAT]6	F	TCTTCAGGTGACAAGCAACA	96	single locus, monomorphic
		R	CCTGGGCATGGAGATAGCAT		
MBo26	[AAT]7	F	CCATAGGCCAGCCCACTATA	134	single locus, monomorphic
		R	ACCCTTTCTTCTTCCTCATTTGT		
MBo27	[AAT]7	F	TCTCTATTGCTTGGTGATCCC	103	single locus, monomorphic
		R	TCCAACAAGGGATGGTTATCATG		
MBo28	[AAT]8	F	GCCTTGAGAGTGGAAGAGGC	205	single locus, monomorphic
		R	TCTCTTCTTTGCGCCCTCAT		
MBo29	[AAT]16	F	AGACATGTAGAGGTGGGACT	211	single locus, monomorphic
		R	TCTGTATGAGAGACGTGTTACAGT		
MBo30	[AAT]8	F	TGACCATAACAAGCTACCAGGT	146	single locus, monomorphic
		R	GGTGGAAGCTATTGATATTGCATGT		
MBo31	[AAT]10	F	TGACAATGATGCATGCGATAACA	187	single locus, monomorphic
		R	GCATCACCCATGTCCTTTAGC		
MBo32	[AAT]10	F	TCCGAGGGCAGTATTTGTCG	117	single locus, monomorphic
		R	CACTATTTCGGAAACCTAAGCCC		
MBo33	[AAT]17	F	GCACACTTTGTATCCGACGC	147	single locus, monomorphic
		R	CAGGGATAGTAACCGTCAGGG		
MBo34*	[AG]28	F	GTGGCACCTCTGCGGTTT	192	single locus, polymorphic
		R	CGAGATGGAAGCACCTGGAG		

MBo35*	[AG]24	F	AGCATGCTTTCTGCTTCATGTG	137	single locus, polymorphic
		R	CCTTTCCCTGACTGCATTGC		
MBo36	[AG]23	F	TCGGAAGTCGAATGTGGCAG	180	no amplification
		R	TCGGAAGAGTGGTCAATCATGG		
MBo37	[AG]23	F	GCTCTACTCCCAGAGACGGA	142	multiple loci, polymorphic
		R	AACAGTCGACGGAATGCTCA		
MBo38*	[AG]20	F	AGTCCTCACTGCTGGTGGTA	130	single locus, polymorphic
		R	TCCTTGAATAGTCCATCTTGCA		
MBo39	[AG]19	F	AACGCAGGTTAAGAGGCTCC	168	multiple loci, monomorphic
		R	CCTCCTGGTGCAACCCTTAC		
MBo40	[AG]19	F	TGTGGAGTGTGAGTCGATGG	193	multiple loci, polymorphic
		R	GGCTGCATAATCTCATCACGC		
MBo41*	[AG]18	F	TTCTCCACCAGCCTCACAAC	184	single locus, polymorphic
		R	ATACGGCCCATCAACCCTTC		
MBo42	[AG]18	F	CCTGGTGGTACATGTGGTCA	136	multiple loci, polymorphic
		R	TGTGGCACATTCATTTCTGAAGG		
MBo43	[AG]18	F	AGTTTGTTCTGTGTGTTGTCAC	137	no amplification
		R	GCACACATCTTGCTTTGAAGAC		
MBo44	[AG]17	F	AACACACTTTAAATCGACTTCTTCA	193	multiple loci, polymorphic
		R	CACGGCTGCCATGTGAGG		
MBo45	[AG]17	F	TAGATCGGAAGTCAGGCCC	193	no amplification
		R	AGAGAAGTGGGAGGAGAGGTC		
MBo46	[AG]17	F	GCCGATATTAGCTTCTTCTTGGC	154	single locus, monomorphic
		R	GCCTTGTTGATCCCGTTTCAC		
MBo47	[AG]16	F	GGCACCTGACGCCTCTTT	188	single locus, monomorphic
		R	TCACTTCGACTCAATTGTATCCAT		
MBo48	[AG]16	F	AGGACAAAGAGATGAGAAGCCT	92	multiple loci, polymorphic
		R	ACCAATTCCCAGTTAGTTGACCA		
MBo49*	[AG]16	F	CATCACCCATTCTCTCTGCCT	141	single locus, polymorphic
		R	GAGAAACCATCCGCACCTCA		
MBo50*	[AG]15	F	AGAAGTCATCTTGAGGGCCC	150	single locus, polymorphic
		R	TTGCTAGAATGATACACAAATTGCT		
MBo51*	[AG]15	F	TGTGCTATTTGTTGGGAATGCA	191	single locus, polymorphic
		R	GCAAGCTCATGTTCTAGTTTCAAGT		
MBo52*	[AG]15	F	ACACATCCTACATGAATAGACCTCC	122	single locus, polymorphic
		R	TCTTGTCATAGCCTAGATTCCCT		
MBo53	[AG]15	F	AGGTTTAAGGGTTTGGGTTAGGG	131	single locus, monomorphic

		R	GGTGGAGTAAGTTTGAGGGTCA		
MBo54*	[AG]11NNN[AG]15	F	CATATGCTGATACAAGAGAGAGGG	124	single locus, polymorphic
		R	ACCTTATAAGCAGGATCCAGACA		
MBo55	[AG]15	F	TGGAATCAACCTTGGGTCTACA	198	multiple loci, polymorphic
		R	TCGTCGGTCTTCTAGCCACT		
MBo56*	[AG]15	F	ACCAAGATCAAGCACGAGGA	103	single locus, polymorphic
		R	AGGATCACCCTTTCTTTCTTTCT		
MBo57*	[AG]15	F	GGGTTCAATCCTGATGAGAGCA	136	single locus, polymorphic
		R	ACCGTTCGATCAACCATGGT		

Loci for which single-locus SSR polymorphism has been detected within our test panel of seven *B. aethiopum* individuals are signaled by an asterisk (*). Expected amplicon size is as predicted by QDD.

Microsatellite-based characterization of the genetic variation within *B. aethiopum* populations of Benin

The set of 11 *B. aethiopum*-specific SSR markers identified in the previous step has been used for the characterization of genetic diversity in our full panel of nine populations (180 individuals) distributed across Benin. Among our sample set, the number of alleles per microsatellite locus ranges from 2 for locus Mbo41 to 6 for loci Mbo34, Mbo35 and Mbo50, with an average value of 4.27, whereas expected heterozygosity (He) values range from 0.031 (locus Mbo56) to 0.571 (locus Mbo35; Table 4). Using these markers, the analysis of genetic diversity (Table 5) shows that the percentage of polymorphism detected at the microsatellite loci investigated ranges from 72.73% (populations of Togbin and Malanville) to 90.91% (populations of Savè, Agoua, Pendjari, Pingou and Trois Rivières), with a mean value of 84.85%. With the exception of the Savè, Hounviatouin and Malanville populations, 1 to 3 private alleles of the targeted microsatellite loci are observed in most populations. Regarding the genetic parameters, the number of effective alleles (Ne) ranges from 1.447 to 2.069 with an average number of 1.761. He values range from 0.263 (Hounviatouin) to 0.451 (Savè) with an average value of 0.354 whereas the observed heterozygosity (Ho) varied from 0.234 (Togbin) to 0.405 (Pingou) with an average value of 0.335. Negative values of Fixation index (F) are obtained for the populations of Pingou, Malanville and Trois rivières whereas positives F values are observed in all other populations investigated, indicating a deficit of heterozygosity in the latter.

Table 4: Characteristics of 11 polymorphic microsatellites markers used for genetic diversity analysis of *B. aethiopum*

Locus name	Number of alleles scored/locus	Expected Heterozygosity (He)
Mbo34	6	0.520
Mbo35	6	0.571
Mbo38	5	0.458
Mbo41	2	0.343
Mbo49	4	0.167
Mbo50	6	0.548
Mbo51	3	0.320
Mbo52	3	0.201
Mbo54	4	0.26
Mbo56	3	0.031
Mbo57	5	0.296

Table 5: Mean diversity parameters for each of the nine *B. aethiopum* populations

Geo-climatic region	**Populations**	**% polymorphic loci**	**Na**	**Ne**	**Nb of private alleles**	**Ho**	**He**	**F**
Guineo-Congolian (South)	*Togbin*	72.73%	2.273	1.584	3	0.234	0.288	0.145
	Hounviatouin	81.82%	2.182	1.447	0	0.272	0.263	0.007
Soudano-Guinean (Centre)	*Savè*	90.91%	2.909	2.069	0	0.384	0.451	0.134
	Biguina	81.82%	2.364	1.770	2	0.345	0.374	0.064
	Agoua	90.91%	2.273	1.722	1	0.329	0.358	0.059

Sudanian (North)	*Pendjari*	90.91%	2.818	1.900	3	0.368	0.396	0.055
	Pingou	90.91%	2.364	1.906	1	0.405	0.390	-0.063
	Malanville	72.73%	2.455	1.627	0	0.302	0.303	-0.020
	Trois rivières	90.91%	2.545	1.822	2	0.373	0.360	-0.055
	Overall mean	**84.85±2.62%**	**2.465±0.103**	**1.761±0.065**		**0.335±0.023**	**0.354±0.023**	**0.035±0.022**

Na : number of different alleles; Ne: Number of effective alleles; Ho= Observed Heterozygosity; He: Expected Heterozygosity; F: Fixation index

Genetic structure of the *B. aethiopum* populations under study

The calculation of Nei's genetic distance among populations (Table 6) shows values ranging from 0.073, as observed between Togbin and Hounviatouin (Guineo-Congolian region), to 0.577 between Togbin (Guineo-Congolian region) and Trois Rivières (Sudanian region). Overall, genetic distances between the fan palm populations under study are lowest within the same region, with the lowest genetic distances among populations of Savè, Pendjari, Pingou, and Trois Rivières which are all located in the Northern part of the country. One interesting exception is the Centre (Guineo-Sudanian) region of Benin, where we find that the most genetically distant population from Savè is the one collected within the Agoua forest reserve (0.339). Surprisingly, Savè displays its highest genetic identity value when compared to the other two populations sampled in protected areas, namely Pendjari (0.870) and Trois Rivières (0.882) which are both located in the Sudanian region. This is an unexpected finding considering the important geographic distances that are involved.

Table 6: Pairwise Population Matrix of Nei's genetic distance and genetic identity values

	Togbin	**Hounviatouin**	**Savè**	**Biguina**	**Agoua**	**Pendjari**	**Pingou**	**Malanville**	**Trois Rivières**
Togbin	-	0.073	0.477	0.253	0.337	0.517	0.494	0.487	0.577
Hounviatouin	0.929	-	0.419	0.110	0.215	0.435	0.317	0.375	0.535
Savè	0.621	0.658	-	0.270	0.339	0.140	0.265	0.238	0.126
Biguina	0.776	0.896	0.763	-	0.152	0.241	0.161	0.186	0.316
Agoua	0.714	0.806	0.713	0.859	-	0.408	0.304	0.359	0.490
Pendjari	0.596	0.647	0.870	0.786	0.665	-	0.167	0.108	0.103

Pingou	0.610	0.728	0.767	0.851	0.738	0.846	-	0.174	0.175
Malanville	0.614	0.688	0.788	0.831	0.699	0.898	0.841	-	0.145
Trois Rivières	0.561	0.585	0.882	0.729	0.613	0.902	0.840	0.865	-

Above the diagonal: Nei's genetic distance; below: genetic identity.

A similar structuration of genetic distances emerges from the analysis of pairwise population genetic differentiation (Fst) (Table 7), suggesting genetic differentiation according to geographic distances between populations, with the notable exception of the lower genetic differentiation between palms from Savè and those from either one of the forest reserves in the Northern region.

Table 7 : Pairwise populations Fst value

	Togbin	**Hounviatouin**	**Savè**	**Biguina**	**Agoua**	**Pendjari**	**Pingou**	**Malanville**	**Trois Rivières**
Togbin	0.000								
Hounviatouin	0.072	0.000							
Savè	0.233	0.221	0.000						
Biguina	0.168	0.086	0.145	0.000					
Agoua	0.215	0.153	0.157	0.105	0.000				
Pendjari	0.247	0.212	0.077	0.120	0.188	0.000			
Pingou	0.252	0.181	0.138	0.103	0.169	0.100	0.000		
Malanville	0.301	0.246	0.149	0.121	0.197	0.072	0.119	0.000	
Trois Rivières	0.285	0.279	0.076	0.178	0.224	0.073	0.104	0.107	0.000

Our analysis of molecular variance (AMOVA; Table 8) shows that within-population variation underlies the major part (53%) of total variance, whereas among-populations and among-regions variations explain variance to a similar extent (23 and 24%, respectively).

Table 8 : AMOVA results.

Source	**df**	**SS**	**MS**	**Est. Var.**	**% total variance**	**P value**

Among Regions	2	309.407	154.704	1.944	24%	<0.001
Among Populations	6	254.302	42.384	1.903	23%	<0.001
Within Populations	171	739.100	4.322	4.322	53%	<0.001
Total	179	1302.809		8.169	100%	

df=degree of freedom, SS=sum of squares, MS mean squares, Est. var.=estimated variance

In accordance with results from both the analysis of genetic distances and the AMOVA, the Principal Coordinates Analysis (PCoA) of our 180 individual *B. aethiopum* samples shows that the first axis (accounting for 24% of total variation) distinguishes roughly between two main groups of populations (Fig 2). Likewise, the Bayesian analysis of our data indicates an optimal value of K=2 for the clustering of the studied populations into two groups (Fig 3): one group that includes palms belonging to the populations of Togbin and Hounviatouin from the Southern part of the country, as well as most of the palms from Biguina and Agoua from the Western (Togolese) border of the Centre region; and one group composed of the majority of the palms collected in Savè (Eastern part of the Centre region) and palms from the Northern populations of Pendjari, Pingou, Malanville and Trois Rivières. The dendrogram derived from the UPGMA analysis of our data further shows that, within these two main groups, subgroups can be defined based on geo-climatic regions, Savè being the only exception to this general trend (Fig 4).

Fig 2. Principal Coordinates Analysis (PCoA) of individual *Borassus aethiopum* samples.

Fig 3. Bayesian cluster analysis

A: Determination of the optimal value of K from Structure Harvester.

B: Bayesian STRUCTURE bar plot analysis of Beninese *B. aethiopum* samples with K=2. Red: group 1; green: group 2. Populations are numbered as in S1 Table and displayed along the horizontal axis.

Fig 4. Genetic relationships among Beninese *B. aethiopum* populations.

Branch ends with identical colors represent populations from the same geo-climatic region.

Discussion

In flowering plant, the efficiency of cross-species transfer of SSR markers is highly variable among taxa, especially when important differences in genome complexity exist between the marker source and the target [64]. Nevertheless, this method has been used successfully for accelerating the analysis of genetic diversity in many plant species, including palms [13,68–70] In the present study, we find that the transferability rate of microsatellite markers developed in other palms genera to the African fan palm, *i.e.* their ability to successfully amplify genomic DNA from the latter species, is very low. Indeed, among the 80 primer pairs designed on either oil palm, date palm or coconut palm, we observe that only 22.5% produce amplicons from *B. aethiopum*. This percentage is very low when compared to both the inter-species and inter-genera transferability rates that have been found in similar studies targeting other palm species: from 17 to 93% in a panel of 32 palm species [50], 75% from *E. oleifera* to *E. guineensis* [67], 86% between the wooly jelly palm *Butia eriospatha* and related species *B. catarinensis* [68] and up to 100% in the licuri palm *Syagrus coronata* [70]. When considering other plant families, our transferability rate is also markedly lower than both the average rate of 50% found by [70] within the *Glycine* genus and among Legumes genera, and the overall rate of 35.2% calculated by [71] for within-family transferability among Gymnosperms and Angiosperms. The low transferability rate in our study might be explained in part by the fact that we used markers originating from genomic sequences. Indeed, as pointed out by [72], such markers have a lower transferability rate when compared to Expressed Sequence Tags (ESTs)-derived microsatellites due to the higher inter-species sequence variability within non-coding *vs.* coding sequences. Similarly, it is plausible that differences in genome size and complexity among palm species and genera account for our difficulty to identify palm SSR markers that successfully amplify in *B. aethiopum*. As a matter of fact, the size of the *B. aethiopum* genome, as determined by flow cytometry (1C = 7.73 Gb; Jaume Pellicer, unpublished data), is 3.2 to 11.5 times larger than those of the microsatellite source species used in the present study: the date palm genome is estimated to be 671 Mb [40] whereas the oil palm genome is 1.8-1.9 Gb [41,74] and the coconut genome is 2.42 Gb [46]. Most likely, these differences in genome sizes among related diploid plant species rely on differences in Transposable Element contents and associated structural variations such as copy number variants and homologous recombinations [75], which might eventually affect the cross-species amplification ability of SSR primers. The illustration of such a mechanism working at the intra-genus level has been provided by cultivated rice species *Oryza sativa* and its wild relative *O. australiensis* [76]. More generally, gaining a better understanding of genome structures within the *Borassus* genus could also help reconcile our results with previous published reports of successful transfer of SSR markers developed from other palm sources to *Borassus flabellifer* (see references cited in Table 1). Indeed, since the genome size of *B. flabellifer* is only marginally smaller than that of *B. aethiopum* (7.58 Gb; Jaume Pellicer, unpublished data), significant differences in genome composition may be underlying the lack of SSR transferability between both species.

In any case, from the low number of successfully transferred microsatellite markers we could only identify one displaying polymorphism in our fan palm test panel, making it impossible to rely on for analysis of genetic diversity. Still, the fact that so little microsatellite polymorphism (2 out of 18 amplifying primer pairs: 11.1%) could be detected in this subset of 20 palms sampled across different locations throughout Benin is somewhat surprising and its reasons remain to be elucidated. In addition to possibly being a symptom of habitat fragmentation and low gene flow between populations, this low diversity might also result from the extremely long juvenile phase that has been attributed to this palm species, for which authors have reported floral maturity occurring 30 to 50 years after germination [77,78].

Compared to other studies in which high-throughput sequencing techniques have been used for the development of new microsatellite markers in species with very little information available [78,79], our results are similar. We identified 57 potential SSR markers, of which 11 displayed polymorphism and were used to assess the genetic structure of *B. aethiopum* populations in Benin. We find a low genetic diversity, with an average He value (0.354) that is substantially below those reported for *Borassus flabellifer* [46] and for other non-timber forest products such as *Khaya senegalensis* (He = 0.53; [80] and *Phyllanthus sp* [81]. The positive F value that we observed in the majority (6 out of 9) of populations in the present study indicates an overall deficiency of heterozygotes across population. This deviation from the Hardy-Weinberg equilibrium (HWE) might reflect poor gene flows through pollen and seed dissemination, leading to crosses between related individuals. Accordingly, our data reveal limited genetic distances among populations, with values lower than those reported for others palm species. Indeed for *B. flabellifer*, genetic distance ranged from 0.716 to 0.957 [82] and among natural oil palm accessions an average of 0.769 was observed [84]. Both our Fst values and AMOVA analysis point to intra-population differentiation as being the main source of genetic variation. As illustrated by the agreement between our PCoA and Bayesian analyses, Beninese *B. aethiopum* populations cluster globally according to geographic distances between the collection sites. However, among the nine populations studied, the population from Savè appears to be the most diversified (He= 0.451) and constitutes an exception to this general distribution. This site located in the Sudano-Guinean transition zone of Benin is currently the most active for the production of fan palm hypocotyls, and it acts as a supplier for the whole national territory (VK Salako, personal communication), suggesting that it might be the largest population of *B. aethiopum* in the country. Moreover, our sampling of Savè individuals appear to be genetically distinct from palms belonging to other populations of the Central region and closer from those of the Northern region, despite the important geographical distances involved with the latter case. We postulate that seed dispersion by elephants might have played a major role in the observed pattern of genetic diversity and explain the singularity observed in Savè. As a matter of fact, [31,32] detected the presence of *B. aethiopum* seeds in elephant dungs and hypothesized that elephants may have played important role in the seed dissemination for this species through fruit consumption and long-distance herd migrations. In support to this assumption, Savè is part of a continuous forest corridor connecting with the Northern region that was used by elephants in their migrations. Up until 1982, the seasonal occurrence of the animal has been reported in the Wari-Maro forest of Central Benin [84].

The use of the specific microsatellite markers developed in this study from genomic sequencing of *B. aethiopum* appears to be efficient to assess the genetic diversity and population structure of this species. These microsatellite loci with respect to our results represent potential molecular marker set that can be used to elucidate the genetic diversity of *B. aethiopum* in other African countries. Additionally, and provided that genome divergence is not too extensive to allow marker transferability, our SSR markers may also been used in a palm species that belongs to the same genus and that is reported to share parts of its distribution area, namely *Borassus akeassii* B.O.G., which has long been confused with *B. aethiopum* due to its similar morphology [85]. High-throughput sequencing proves to be a good, fast and effective way to develop new microsatellite markers especially for plant species without published molecular data. The increasing availability and affordability of this technology makes it possible, both technically and financially, to overcome the difficulties arising in case studies such as ours, where marker transfer has proven to be limited or ineffective. To our knowledge, the data presented in the present article constitute the first sizeable molecular resource available for the African fan palm, which we have made available to the scientific community at large in order to facilitate the implementation of an increasing number of studies on this palm species. We have also performed the first analysis of the genetic diversity of *B. aethiopum* in an African country, which we see as a first step towards the elaboration of an evidence-based strategy for sustainable resource management and preservation in Benin. As a complement, the acquisition of agro-morphological data and the characterization of processes regulating the reproductive development of the species are currently under way. Beyond that, we also aim to extend our analysis of *B. aethiopum* diversity to the West African sub-region, and leverage the data acquired to improve knowledge of both other species within the *Borassus* genus, and of palms diversity as a whole.

Acknowledgements

The Authors wish to thank Emira Cherif, Carina Castillo and Frédérique Aberlenc (IRD, UMR DIADE, Montpellier, France) for sharing their unpublished microsatellite markers data for transferability tests, and Dr Jaume Pellicer (Comparative Plant & Fungal Biology, Jodrell Laboratory, Royal Botanical Gardens, Kew, UK), for allowing us to use the palm flow cytometry data.

Funding statement

The work described in this article was funded through travel grants to MJK and KA under the framework of the MooSciTIC project granted to EJ by Agropolis Fondation (ID 1501-011, "Investissements d’Avenir" Program - Labex Agro: ANR-10-LABX-0001-01). Additional funding was provided by the Sud Expert Plantes - Développement Durable (SEP2D) programme to KA (GenPhyB project, ID AAP3-64). MJK is the recipient of a PhD fellowship from the French Embassy in Benin.

Conflict of interest

The authors declare no conflict of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethics statement

In accordance with the Nagoya Protocol on Access and Benefit Sharing (ABS), a field permit allowing access and non-commercial use for research purposes of the plant material used in the present study has been submitted to the competent national authority (Direction Générale des Eaux, Forêts et Chasse/Ministère du Cadre de Vie et du Développement Durable, Benin).

References

1. Huang XQ, Cöster H, Ganal MW, Röder MS. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet. 2003;106(8):1379-89. doi.org/10.1007/s00122-002-1179-7
2. Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J. Genetic Structure and Diversity Among Maize Inbred Lines as Inferred From DNA Microsatellites. Genetics. 2003;165:2117-28.
3. Liu S, Cantrell RG, McCarty JC, Stewart JM. Simple Sequence Repeat–Based Assessment of Genetic Diversity in Cotton Race Stock Accessions. Crop Science. 2000;40(5):1459-69. doi.org/10.2135/cropsci2000.4051459x
4. Maughan PJ. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome. 1995; 38(4):715-23. doi.org/10.1139/g95-090
5. Milla SR, Isleib TG, Stalker HT. Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. Genome. 2005;48(1):1-11. doi.org/10.1139/g04-089
6. Mohamed A, García-Martínez S, Loumerem M, Carbonell P, Ruiz JJ, Boubaker M. Assessment of genetic diversity among local pea (*Pisum sativum* L.) accessions cultivated in the arid regions of Southern Tunisia using agro-morphological and SSR molecular markers. Genet Resour Crop Evol. 2019;66(6):1189-203. doi.org/10.1007/s10722-019-00784-8
7. Moon HS, Nicholson JS, Lewis RS. Use of transferable *Nicotiana tabacum* L. microsatellite markers for investigating genetic diversity in the genus *Nicotiana*. Belzile F, éditeur. Genome. 2008;51(8):547-59. doi.org/10.1139/g08-039

8. Muriira NG, Muchugi A, Yu A, Xu J, Liu A. Genetic Diversity Analysis Reveals Genetic Differentiation and Strong Population Structure in Calotropis Plants. Sci Rep. 2018;8(1):7832. doi.org/10.1038/s41598-018-26275-x

9. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, et al. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breeding. 1996;2(3):225-38. doi.org/10.1007/BF00564200

10. Yuan Q-J, Zhang Z-Y, Hu J, Guo L-P, Shao A-J, Huang L-Q. Impacts of recent cultivation on genetic diversity pattern of a medicinal plant, Scutellaria baicalensis (Lamiaceae). BMC Genet. 2010;11(1):29. doi.org/10.1186/1471-2156-11-29

11. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. Euphytica.2011;177(3):309-34. doi.org/10.1007/s10681-010-0286-9

12. Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, et al. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology & Biotechnological Equipment. 2018;32(2):261-85. doi.org/full/10.1080/13102818.2017.1400401

13. Satya P, Paswan PK, Ghosh S, Majumdar S, Ali N. Confamiliar transferability of simple sequence repeat (SSR) markers from cotton (*Gossypium hirsutum* L.) and jute (*Corchorus olitorius* L.) to twenty two Malvaceous species. 3 Biotech. 2016;6(1). doi.org/10.1007/s13205-016-0392-z

14. Wang C, Jia G, Zhi H, Niu Z, Chai Y, Li W, et al. Genetic Diversity and Population Structure of Chinese Foxtail Millet [*Setaria italica* (L.) Beauv.] Landraces. Genes|Genomes|Genetics . 2012;2(7):769-77. doi/10.1534/g3.112.002907

15. Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, et al. How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology. 2003;12(6):1339-48. doi/abs/10.1046/j.1365-294X.2003.01825.x

16. Mnejja M, Garcia-Mas J, Audergon J-M, Arús P. Prunus microsatellite marker transferability across rosaceous crops. Tree Genetics & Genomes. 2010;6(5):689-700. https://doi.org/10.1007/s11295-010-0284-z

17. Wünsch A. Cross-transferable polymorphic SSR loci in *Prunus* species. Scientia Horticulturae. 2009;120(3):348-52. doi.org/10.1016/j.scienta.2008.11.012

18. Yu F, Wang B-H, Feng S-P, Wang J-Y, Li W-G, Wu Y-T. Development, characterization, and cross-species/genera transferability of SSR markers for rubber tree (*Hevea brasiliensis*). Plant Cell Rep. 2011;30(3):335-44. https://doi.org/10.1007/s00299-010-0908-7

19. Karaca M, Ince AG, Aydin A, Ay ST. Cross-genera transferable e-microsatellite markers for 12 genera of the Lamiaceae family. J Sci Food Agric. 2013;93(8):1869-79.

20. Raveendar S, Lee G-A, Jeon Y-A, Lee YJ, Lee J-R, Cho G-T, et al. Cross-amplification of *Vicia sativa* subsp. *sativa* microsatellites across 22 other *Vicia* species. Molecules. 2015;20(1):1543-50. doi.org/10.3390/molecules20011543

21. Pratap A, Gupta S, Tomar R, Malviya N, Maurya R, Pandey VR, et al. Cross-genera amplification of informative microsatellite markers from common bean and scarlet runner bean for assessment of genetic diversity in mungbean (Vigna radiata). Plant Breeding. 2016;135(4):499-505. doi/abs/10.1111/pbr.12376

22. Blach-Overgaard A, Svenning J-C, Dransfield J, Greve M, Balslev H. Determinants of palm species distributions across Africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints. Ecography. 2010;33(2):380-91. doi/abs/10.1111/j.1600-0587.2010.06273.x

23. Gruca M, Yu W, Amoateng P, Nielsen MA, Poulsen TB, Balslev H. Ethnomedicinal survey and in vitro anti-plasmodial activity of the palm Borassus aethiopum Mart. Journal of Ethnopharmacology. 2015;175:356-69. doi.org/10.1016/j.jep.2015.09.010

24. Adomou CA, Agbani OP, Sinsin B. Plantes. In: Protection de la nature en Afrique de l'Ouest: Une liste rouge pour le Bénin / Nature Conservation in West Africa: Red List for Benin,. International Institute of Tropical Agriculture, Ibadan, Nigeria,: Neuenschwander P. S., Sinsin B., Goergen G.,; 2011. 365 p.

25. Giffard PL. Le palmier rônier *Borassus aethiopum* Mart. Revue bois et forêt des tropiques. 1967;(116):3-13.

26. Michon L, Adéoti K, Koudouvo K, Ewédjè E-E, Stauffer F. Notes on *Borassus aethiopum* Mart., a Multi-Purpose Palm in Togo. 2018;57-69.

27. Ohin BM, Adeoti K, Noumavo PA, Kouhounde SHS, Garba K, Tovide NS, et al. Technological Processes Itinerary and Socio-economic Impact Assessment of *Borassus aethiopum* Hypocotyles "Agonté" Produced in Centre of Benin Republic. Food and Public Health. 2018;7.

28. Sambou B, Lawesson EJ, Barfod A. Borassus aethiopum, a threatened multiple purpose palm in Senegal. Principes. 1992;36(3):148-55.

29. Gbesso F, Yedomonhan H, Tente B, Akoegninou A. Distribution géographique des populations de rôniers (*Borassus aethiopum* Mart, Arecaceae) et caractérisation phytoécologique de leurs habitats dans la zone soudano-guinéenne du Benin. Journal of Applied Biosciences. 2014;74(1):6099. doi.org/10.4314/jab.v74i1.14

30. Ouinsavi C, Gbémavo C, Sokpon N. Ecological Structure and Fruit Production of African Fan Palm (*Borassus aethiopum* Mart.) Populations. American Journal of Plant Sciences. 2011;02(06):733-43. doi.org/10.4236/ajps.2011.26088

31. Salako V, Assogbadjo A, Adomou A, Agbangla C, Glèlè Kakaï R. Latitudinal distribution, co-occurring tree species and structural diversity of the threatened palm *Borassus aethiopum* (Arecaceae) in Benin, West Africa. Plant Ecology and Evolution. 2015;148(3):335-49. doi.org/10.5091/plecevo.2015.1046

32. Salako , V, Houéhanou TH, Yessoufou K, Assogbadjo AE, Akoegninou A, Kakaï Glèlè RL. Patterns of elephant utilization of *Borassus aethiopum* Mart. and its stand structure in the Pendjari National Park, Benin, West Africa. Tropical Ecology. 2017;425-37.

33. Gbesso F, Akouehou G, Tente B. Aspects technico-économiques de la transformation de *Borassus aethiopum* Mart (Arecaceae) au Centre-Bénin. 2013;15.

34. Agyarko K, Dartey E, Amankwah Kuffour R, Abum Sarkodie P. Assessment of Trace Elements Levels in Sediment and Water in Some Artisanal and Small-Scale Mining (ASM) Localities in Ghana. Current World Environment. 2014;9(1):7-16.

35. Ouédraogo A, Boussim J, Zongo J-D, Guinko S. Caractéristiques morphologiques des rôniers (*Borassus* L.) du Burkina Faso. Etudes flo. Vég. Burkina Faso. 2002;(7):37-40.

36. Dixo M, Metzger JP, Morgante JS, Zamudio KR. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. Biological Conservation. 2009;142(8):1560-9. doi.org/10.1016/j.biocon.2008.11.016

37. Jump AS, Penuelas J. Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. Proceedings of the National Academy of Sciences. 2006;103(21):8096-100. doi/10.1073/pnas.0510127103

38. Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature. 2013;500(7462):335-9. doi.org/10.1038/nature12309

39. Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, et al. De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nat Biotechnol. 2011;29(6):521‑7. doi.org/10.1038/nbt.1860

40. Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, et al. Genome sequence of the date palm *Phoenix dactylifera* L. Nat Commun. 2013;4(1):2274. doi.org/10.1038/ncomms3274

41. He Z, Zhang C, Liu W, Lin Q, Wei T, Aljohi HA, et al. DRDB: An Online Date Palm Genomic Resource Database. Front Plant Sci. 2017;8. doi.org/10.3389/fpls.2017.01889

42. Huang Y-Y, Matzke AJM, Matzke M. Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (*Cocos nucifera*). Candela H, éditeur. PLoS ONE. 2013;8(8):e74736. doi.org/10.1371/journal.pone.0074736

43. Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, et al. The genome draft of coconut (*Cocos nucifera*). GigaScience.2017;6(11). doi/10.1093/gigascience/gix095/4345653

44. Bazzo BR, de Carvalho LM, Carazzolle MF, Pereira GAG, Colombo CA. Development of novel EST-SSR markers in the macaúba palm (*Acrocomia aculeata*) using transcriptome sequencing and cross-species transferability in Arecaceae species. BMC Plant Biology. 2018;18(1). doi/10.1186/s12870-018-1509-9

45. Billotte N, Marseillac N, Risterucci A-M, Adon B, Brottier P, Baurens F-C, et al. Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). Theoretical and Applied Genetics. 2005;110(4):754‑65. doi.org/10.1007/s00122-004-1901-8

46. Pipatchartlearnwong K, Swatdipong A, Vuttipongchaikij S, Apisitwanich S. Genetic evidence of multiple invasions and a small number of founders of Asian Palmyra palm (*Borassus flabellifer*) in Thailand. BMC Genetics. 2017;18(1). doi/10.1186/s12863-017-0554-y

47. Rivera R, Edwards KJ, Barker JHA, Arnold GM, Ayad G, Hodgkin T, et al. Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. 1999;42:8. doi.org/10.1139/gen-42-4-668

48. Ukoskit K, Chanroj V, Bhusudsawang G, Pipatchartlearnwong K, Tangphatsornruang S, Tragoonrung S. Oil palm (*Elaeis guineensis* Jacq.) linkage map, and quantitative trait locus analysis for sex ratio and related traits. Molecular Breeding. 2014;33(2):415-24. doi/10.1007/s11032-013-9959-0

49. Wu Y, Yang Y, Qadri R, Iqbal A, Li J, Fan H, et al. Development of SSR Markers for Coconut (*Cocos nucifera* L.) by Selectively Amplified Microsatellite (SAM) and Its Applications. Tropical Plant Biology. 2019;12(1):32-43. doi.org/10.1007/s12042-018-9215-1

50. Xiao Y, Xia W, Ma J, Mason AS, Fan H, Shi P, et al. Genome-Wide Identification and Transferability of Microsatellite Markers between Palmae Species. Frontiers in Plant Science. 2016;7. doi.org/10.3389/fpls.2016.01578

51. Billotte N, Risterucci AM, Barcelos E, Noyer JL, Amblard P, Baurens FC. Development, characterisation, and across-taxa utility of oil palm (*Elaeis guineensis* Jacq.) microsatellite markers. Genome. 2001;44(3):413-25. doi.org/10.1139/gen-44-3-413

52. Rao NM, Ramesha BT, Ravikanth G, Ganeshaiah KN, Shaanker RU. Cross-Species Amplification of Coconut Microsatellite Markers in Rattans. Silvae Genetica. 2007;56(1-6):282-6. doi.org/10.1515/sg-2007-0040

53. Anitha N, Jayaraj KL, Kumar EPA, George J, Rajesh MK. Assessment of cross-taxa utility of coconut microsatellite markers. Indian Journal of Horticulture. 2008;6.

54. Meerow AW, Wisser RJ, Brown SJ, Kuhn DN, Schnell RJ, Broschat TK. Analysis of genetic diversity and population structure within Florida coconut (*Cocos nucifera* L.) germplasm using microsatellite DNA, with special emphasis on the Fiji Dwarf cultivar.

Theoretical and Applied Genetics. 2003;106(4):715-26. doi.org/10.1007/s00122-002-1121-z

55. Perera L, Russell JR, Provan J, McNicol JW, Powell W. Evaluating genetic relationships between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. Theoretical and Applied Genetics. 1998;96(3-4):545-50. doi/10.1007/s001220050772

56. Aberlenc-Bertossi F, Castillo K, Tranchant-Dubreuil C, Chérif E, Ballardini M, Abdoulkader S, et al. In Silico Mining of Microsatellites in Coding Sequences of the Date Palm (Arecaceae) Genome, Characterization, and Transferability. Applications in Plant Sciences. 2014;2(1):1300058. Disponible sur: http://doi.wiley.com/10.3732/apps.1300058

57. Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, et al. Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers. Theoretical and Applied Genetics. 2006;114(1):49-58. doi/10.1007/s00122-006-0409-9

58. Mariac C, Scarcelli N, Pouzadou J, Barnaud A, Billot C, Faye A, et al. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. Molecular Ecology Resources. 2014;14(6):1103-13. http://doi.wiley.com/10.1111/1755-0998.12258

59. Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, et al. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics. 2010;26(3):403-4. doi.org/10.1093/bioinformatics/btp670

60. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics. 2012;28(19):2537-9. doi/10.1093/bioinformatics/bts460
61. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959. Genetics. 2000;155(2):945-59. doi.org/10.1534/genetics.116.195164
62. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology. 2005;14(8):2611-20. doi/10.1111/j.1365-294X.2005.02553.x
63. Garcia-Vallvé S, Palau J, Romeu A. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. Molecular Biology and Evolution. 1999;16(9):1125-34. doi.org/10.1093/oxfordjournals.molbev.a026203
64. Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C. Cross-species transfer of nuclear microsatellite markers: potential and limitations. Mol Ecol. 2007;16(18):3759-67. doi/10.1111/j.1365-294X.2007.03439.x
65. Manju KP, Manimekalai R, Naganeeswaran SA. Microsatellites mining in date palm (Phoenix dactylifera L.) and their cross transferability across Arecaceae family. Plant Omics. 2016;9(3):191-7. doi.org/10.21475/poj.16.09.03.p7793
66. Viana MVC, Miranda EA, de Francisco AK, Carvalho CAL, Waldschmidt AM. Transferability of microsatellite primers developed for stingless bees to four other species of the genus Melipona. Genet Mol Res. 2011;10(4):3942-7. doi.org/10.4238/2011.november.22.11
67. Zaki NM, Singh R, Rosli R, Ismail I. Elaeis oleifera Genomic-SSR Markers: Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae. IJMS. 2012;13(4):4069-88. doi.org/10.3390/ijms13044069

68. Nazareno AG, Zucchi MI, dos Reis MS. Microsatellite markers for *Butia eriospatha* (Arecaceae), a vulnerable palm species from the Atlantic Rainforest of Brazil. Am J Bot. juill 2011;98(7):e198-200.

69. Simplicio RR, Pereira DG, Waldschmidt AM. Transferability of microsatellite markers in *Syagrus coronata* (Mart.) Becc. (Arecaceae), an iconic palm tree from the Brazilian semiarid region. Genetics and Molecular Research. 2017;16(2). doi.org/10.4238/gmr16029699

70. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A. Cross-species amplification of soybean (Glycine max) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Molecular Biology and Evolution. 1998;15(10):1275‑87. doi.org/10.1093/oxfordjournals.molbev.a025856

71. Rossetto M. Sourcing of SSR markers from related plant species. In: Henry RJ, éditeur. Plant genotyping: the DNA fingerprinting of plants. Wallingford: CABI; 2001. p. 211‑24. doi.org/10.1079/9780851995151.0211

72. Fan L, Zhang M-Y, Liu Q-Z, Li L-T, Song Y, Wang L-F, et al. Transferability of Newly Developed Pear SSR Markers to Other Rosaceae Species. Plant Molecular Biology Reporter. 2013;31(6):1271‑82. doi/10.1007/s11105-013-0586-z

73. Rival A, Beule T, Barre P, Hamon S, Duval Y, Noirot M. Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq) tissue cultures and seed-derived plants. Plant Cell Reports. 1997;16(12):884‑7. doi.org/10.1007/s002990050339

74. Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene. 2012;509(1):7‑15. doi.org/10.1016/j.gene.2012.07.042

75. Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Saniyal A, et al. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. 2006;16(10):1262-9. doi.org/10.1101/gr.5290206

76. Cabannes Y, Chantry G, Willemin V. Le rônier et le palmier à sucre : production et mise en œuvre dans l'habitat. 1987 p. 44.

77. Salako KV. Population ecology, pattern of uses and conservation of the dioecious palm tree Borassus aethiopumMart. (Arecaceae) in Benin, West Africa. [Faculty of Sciences and Techniques (FAST)/ BENIN]: University of Abomey-Calavi (UAC); 2016.

78. Grubisha LC, Nelson BA, Dowie NJ, Miller SL, Klooster MR. Characterization of Microsatellite Markers for Pinedrops, *Pterospora andromedea* (Ericaceae), from Illumina MiSeq sequencing. Applications in Plant Sciences. 2014;2(11):1400072. doi.org/10.3732/apps.1400072

79. Yamamoto M, Handa Y, Aihara H, Setoguchi H. Development and characterization of 43 microsatellite markers for the critically endangered primrose *Primula reinii* using MiSeq sequencing. Plant Divers. 2018;40(1):41-4. doi.org/10.1016/j.pld.2017.09.003

80. Gaoue OG, Lemes MR, Ticktin T, Sinsin B, Eyog-Matig O. Non-timber Forest Product Harvest does not Affect the Genetic Diversity of a Tropical Tree Despite Negative Effects on Population Fitness. Biotropica. 2014;46(6):756-62. doi.org/10.1111/btp.12145

81. Geethika E, Triveni HN, Srirama R, Siva R, Setty S, Ravikanth G. Development and characterization of microsatellite markers for Phyllanthus emblica Linn., important nontimber forest product species. J Genet. 2018;97(4):1001-6. doi.org/10.1007/s12041-018-0979-8

82. Vinayagam P, Dhandapani J, Raman P, Alagarsamy RK, Muthusamy P, Balaraman SS. Molecular Characterization for Genetic Diversity of Palmyrah (*Borassus flabellifer*)

Accessions using Inter Simple Sequence Repeat (ISSR) Markers. The Asian and Australasian Journal of Plant Science and Biotechnology. 2009;3(1):11‑5.

83. Bakoumé C, Wickneswari R, Siju S, Rajanaidu N, Kushairi A, Billotte N. Genetic diversity of the world's largest oil palm (Elaeis guineensis Jacq.) field genebank accessions using microsatellite markers. Genet Resour Crop Evol. 2015;62(3):349‑60. doi.org/10.1007/s10722-014-0156-8

84. Roth H.H., Douglas-Hamilton I. Distribution and status of elephants in West Africa (1). Mamm. 1991;55(4):489‑527. doi.org/10.1515/mamm.1991.55.4.489

85. Bayton RP, Ouédraogo A, Guinko S. The genus Borassus (Arecaceae) in West Africa, with a description of a new species from Burkina Faso. Botanical Journal of the Linnean Society. 2006;150(4):419‑27. Disponible sur: doi.org/10.1111/j.1095-8339.2006.00485.x

Supporting information captions

S1 Table: List of sampled *B. aethiopum* individuals.

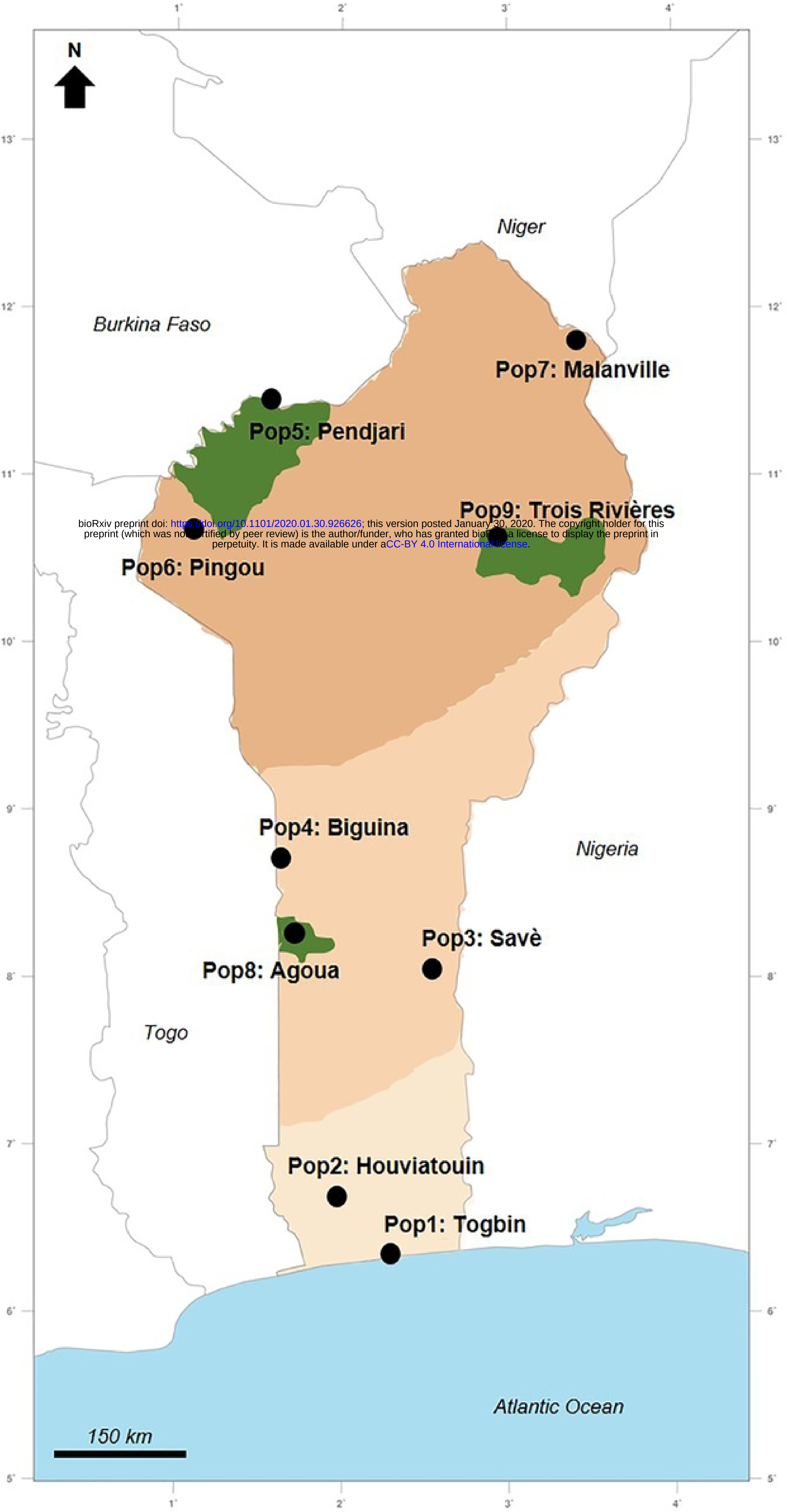
M, F: = male or female palm, respectively.

All geographic coordinates are provided as North from the Equator (latitude) and East from the Greenwich meridian (longitude), respectively.

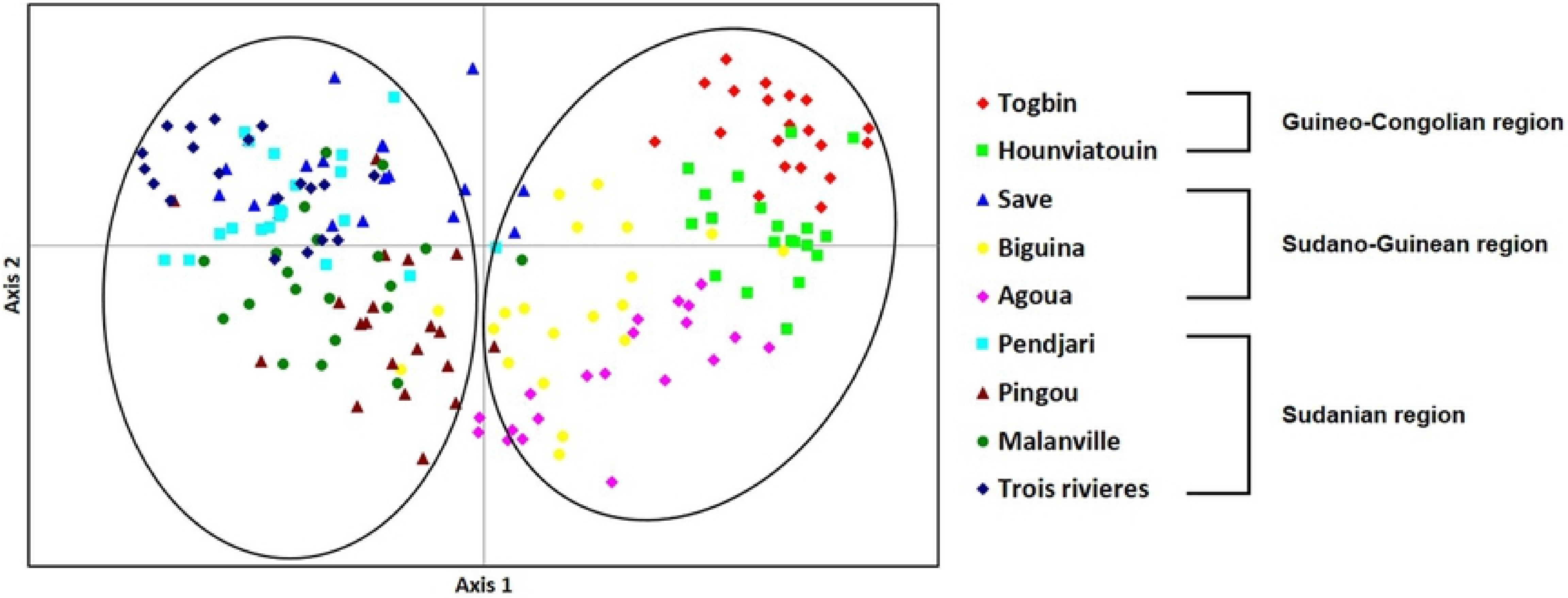
(DOCX)

S2 Table: List and characteristics of putative microsatellite loci identified in the genome of *B. aethiopum* through QDD analysis

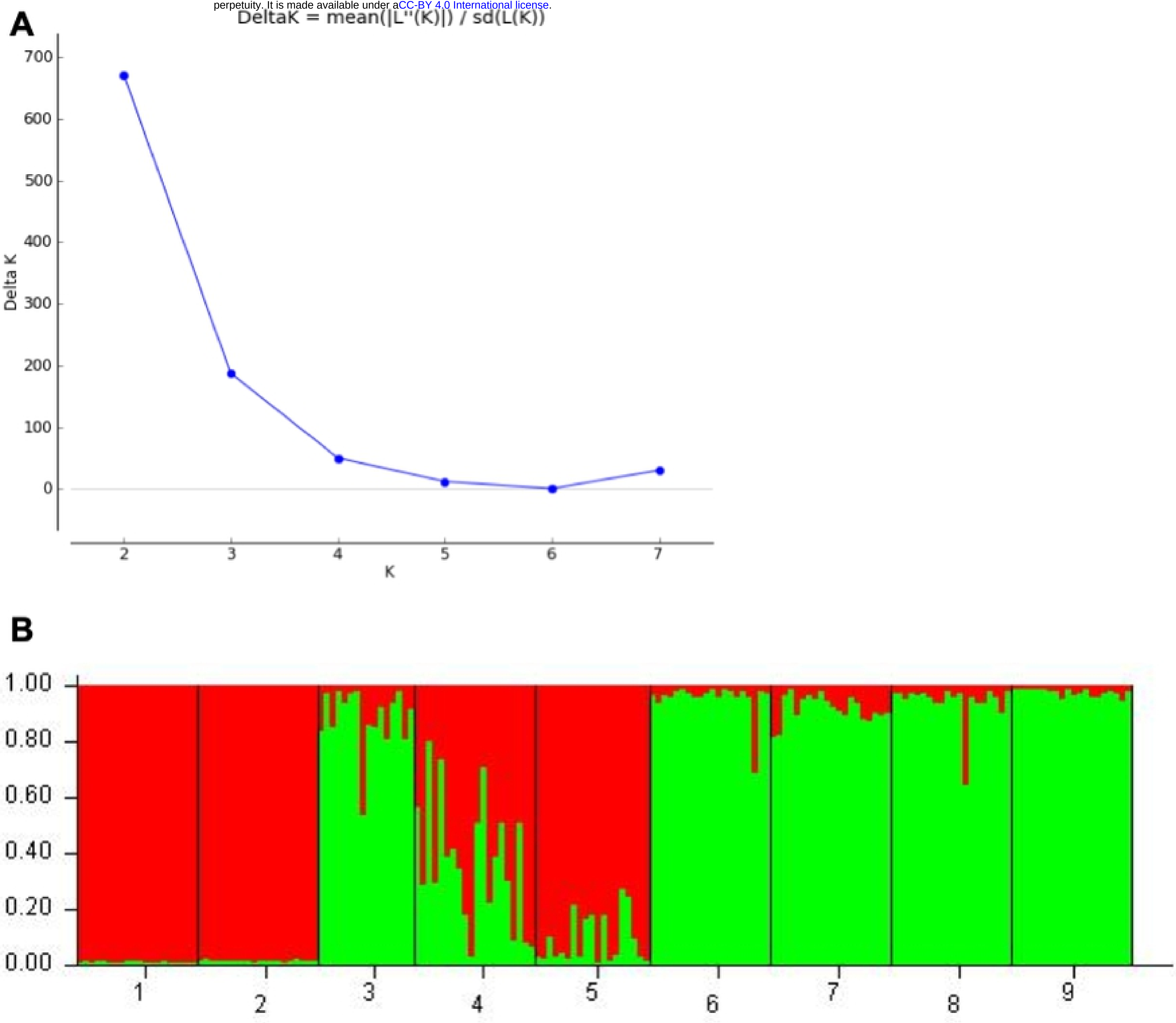
(XLSX)



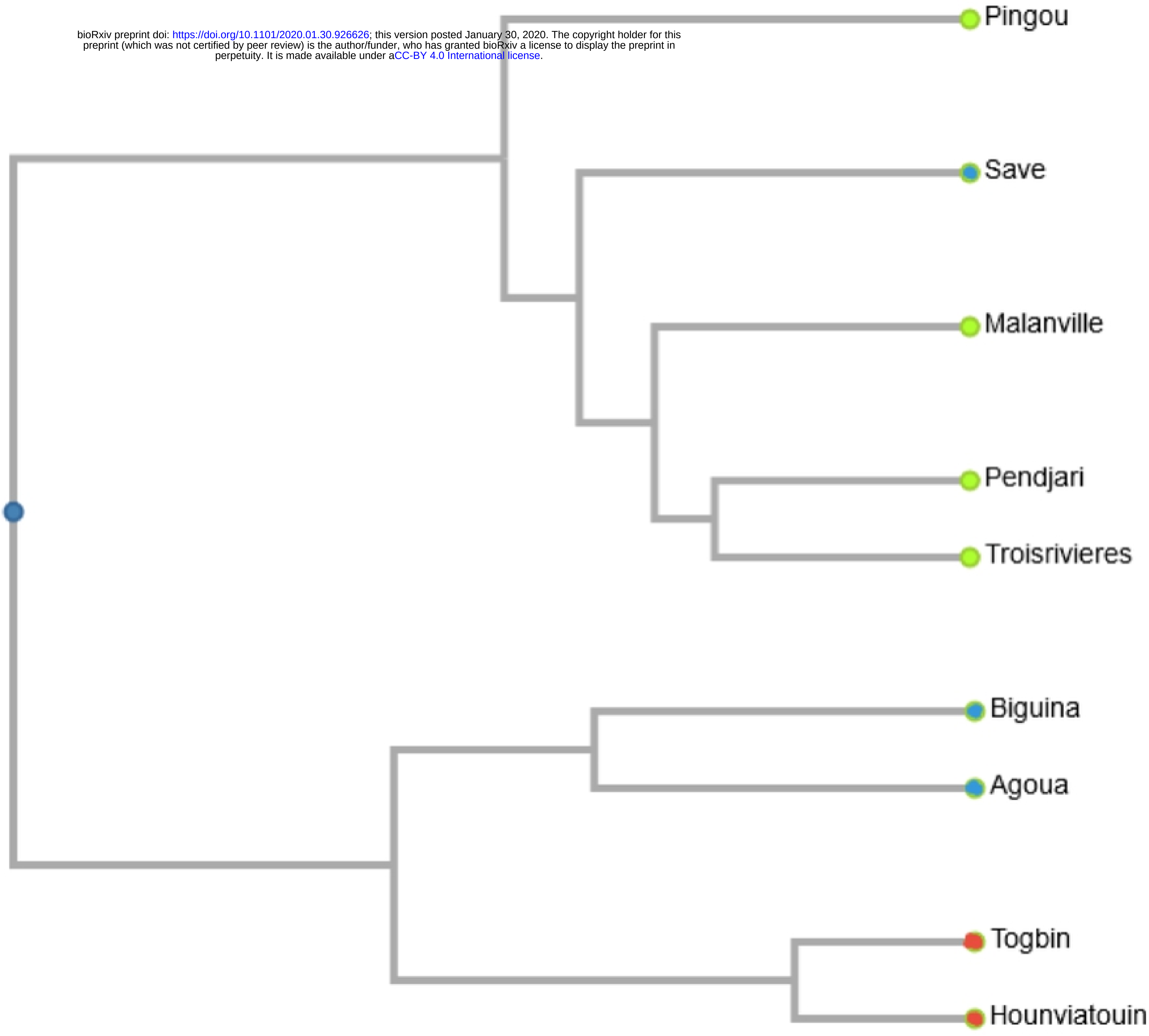
Geo-climatic distribution of beninese Borassu



Principal coordinates analysis



Bayesian cluster analysis



Genetic relationship among Beninese Borassus aethiopum popu