Transferability, development of simple 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
Université d'Abomey-Calavi
Valère Kolawolé Salako
Universite d'Abomey-Calavi Faculte des Sciences Agronomiques
Sylvain Santoni
INRA Centre de Montpellier
Leila Zekraoui
Institut de recherche pour le developpement, Montpellier
Muriel Latreille
INRA Centre de Montpellier
Christine Tollon-Cordet
INRA Centre de Montpellier
Cédric Mariac
Institut de Recherche pour le Développement, Centre de Montpellier
Estelle Jaligot
CIRAD, Centre IRD, Montpellier
Thierry Beulé
CIRAD, Centre IRD, Montpellier
Kifouli ADEOTI (zoulade@yahoo.fr)
Universite d'Abomey-Calavi Faculte des Sciences et Techniques https://orcid.org/0000-0003-2276-6042

Research article

Keywords: Borassus aethiopum, genetic diversity, microsatellite, marker transferability, high-throughput sequencing, simple sequence repeat, under-studied species

Posted Date: July 14th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-37023/v1

Version of Record: A version of this preprint was published at BMC Genetics on December 3rd, 2020. See the published version at https://doi.org/10.1186/s12863-020-00955-y.

Abstract

Background: In Sub-Saharan Africa, *Borassus aethiopum* Mart. (African fan palm) 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 and population structure of *B. aethiopum* across nine collection sites spanning the three climatic regions of Benin, West Africa, through the use of microsatellite markers.

Results: During a first phase we relied on the reported transferability of primers developed in other palm species. We find that, in disagreement with previously published results, only 22.5% of the markers tested enable amplification of *B. aethiopum* DNA and polymorphism detection is very low.

We thus generated a *B. aethiopum*-specific genomic dataset through high-throughput sequencing and used it in a second phase for the *de novo* detection of microsatellite loci. Among the primer pairs designed to target these, 11 enabled polymorphism detection and were further used for analyzing genetic diversity. Across the nine collection sites, expected heterozygosity (He) ranges from 0.263 to 0.451 with an overall average value of 0.354, showing a low genetic diversity. Analysis of molecular variance (AMOVA) shows that within-site variation accounts for 53% of the genetic variation, and accordingly the low number of migrants and the positive values of the fixation index (F) in sites from both the Central (Sudano-Guinean) and the Southern (Guinean) climatic regions suggest limited gene flow between sites. While we globally observe a weak correlation between genetic and geographic distances, our clustering analyses indicate that *B. aethiopum* palms from Savè (Center) are genetically more similar to those from the Northern sites than to samples from the other Central sites.

Conclusions: In the 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.

Background

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 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. Often, transferring markers between species is seen as a smarter investment than developing and testing new markers, especially if limited funding is available [1, 2].

Over the last three decades, molecular markers have been widely used to study genetic variation among and within populations of various plant species [3–7]. Among the different types of markers that are available, microsatellites or simple sequence repeats (SSRs) are often selected due to their high mutation frequency, which ranges from 10^{-2} to 10^{-6} nucleotides per locus per generation [2, 8] and generates multiple allelic forms, and their co-dominant nature. The combination of both characteristics make them sensitive tools for the assessment of genetic diversity among species, determination of population structure, phylogenetic reconstruction, genetic mapping, evolutionary analyses, and molecular breeding [9–12]. From a practical perspective, the popularity of SSRs is also related to their low resource requirements (*i.e.* technical skills, laboratory equipments and consumables) that enable their easy implementation and the reproducibility of results in most research environments [2, 8]. 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 [11, 13]. 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 [14, 15]; between species of the *Hevea* genus and to other Euphorbiaceae [16]; among Lamiaceae [17]; among Legumes belonging to the *Vicia* genus [18] and from the *Phaseolus* genus to *Vigna* [19]. In other cases, the ever-increasing affordability of high-throughput sequencing technologies and the development of dedicated bioinformatics data mining tools have enabled the identification of microsatellite loci and the development of SSR markers, including in non-model plant species with limited or no background genetic information [20–23].

Borassus aethiopum Mart., also known as ron palm, toddy palm or African fan 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 [24]. The species is classified as a non-timber forest products (NTFPs)-providing plant, since different parts of the plant are used for various purposes by local populations [24, 25]. In Benin (West Africa) for instance, 121 different uses distributed in seven categories (medicinal, handicrafts, food, construction, firewood, ceremonies and rituals) have been reported for the species [26]. Among these, the consumption of ripe fruits (fresh or roasted) and hypocotyls as food, the use of the weather- and pest-resistant stipe as construction wood and that of leaves and petioles in handicrafts, are the most widespread in local populations [26–28]. These different products are also sold in markets, mostly by women, to whom they provide additional income: it is indeed estimated that in Benin, sales of hypocotyls alone may represent 50% to nearly three times the minimum wage of 40,000 CFA Francs (*ca.* 61 euros) a month [27].

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 [27, 29–32]. Further fragmentation of the species' habitat has been observed as a result of land clearing for agriculture or urban development [32–34]. As illustrated through similar examples in the literature [35, 36], such phenomena may lead to restricted gene flow and ultimately, to loss of genetic diversity among *B. aethiopum* populations. A sustainable management policy for *B. aethiopum* populations is therefore urgently needed and acquiring information on the genetic diversity of the species and population structure is a major step towards defining sustainable management actions. At the time of writing the present article, only a few chloroplast 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 *Elaeis guineensis* Jacq. [37], *Phoenix dactylifera* L. [38–40] and *Cocos nucifera* L. [41, 42]. 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 [43–49]. In several instances [44–47, 49] these tests included samples from *B. flabellifer*, the Asian relative of *B. aethiopum*.

Here we first describe attempts to use SSR markers identified in these other palm species, for the analysis of genetic diversity in *B. aethiopum*. Then, we describe the low-coverage sequencing of the *B. aethiopum* genome with the aim of developing the first set of specific SSR markers targeting this species. Finally, we used the novel SSR markers to assess the genetic diversity and population structure of *B. aethiopum* samples collected across the three different climatic regions of Benin, as an important first step towards more comprehensive studies.

Results

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

Of the 80 microsatellite markers selected from the three model palm species *E. guineensis*, *P. dactylifera* and *C. nucifera* and tested for amplification on *B. aethiopum* DNA, 18 (22.5%) generate amplification products (Table 1). 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 *P. dactylifera* primers display genetic polymorphism in our *B. aethiopum* test panel. Among *E. guineensis*-derived SSR markers however, two, namely ESSR566 and ESSR652, display polymorphism. However, it must be noted that depending on the DNA sample the ESSR566 primer pair generates a variable number of amplicons with distinct sizes, which may be an indication that more than one locus is targeted.

Table 1
Summary of SSR markers transferability assessment

Species of origin	Number of SSR markers tested	Number of successful amplifications (% 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 assayed. Only one of these markers, namely ESSR652, enables unambiguous detection of microsatellite locus polymorphism in *B. aethiopum*, 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 in *B. aethiopum*, we developed specific *B. aethiopum* markers from *de novo* sequencing data. A total of 23,281,354 raw reads with an average length of 250 bp have been generated from one MiSeq run. Raw sequence reads have been trimmed resulting in 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 the contigs, the QDD software identifies a total of 1,618 microsatellite loci (Additional file 1), of which 1,327 (82.01%) are perfect (*i.e.* repeat size 4 bp or smaller and repeat number 10–20). Among the perfect microsatellite loci, 83.86% 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 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 2. Successful amplification of *B. aethiopum* DNA is obtained for 54 (94.7%) primer pairs and of these, 34 (60.0% of amplifying couples) show no polymorphism. The remaining 20 primer pairs enable the amplification of polymorphic products, however nine of them yield complex, ambiguous amplification profiles that prevent their use for reliable detection of genetic variation. As a result, 11 putative *B. aethiopum* SSR markers (representing 20.4% of primer pairs associated with successful amplification and 55.0% of those detecting polymorphic products in our study) are both

polymorphic and unambiguously mono-locus in our amplification test panel and may therefore be used for further analyses.

Table 2

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

Locus name	Repeat motif	Primer sequences (5'-3' orientation)	Expected amplicon size (bp)	Amplification product
MBo01	$[AGG]_7$	CCTATCCTTCCATCCCGATCG	90	complex, polymorphic
		TTGCCGTGAATCAGCCTCAA		
MBo02	$[ATC]_7$	GGGAGAACAAGGATAACAGCAG	115	single locus, monomorphic
		TCCATTTCATCACTAGCTCGGT		
MBo03	$[AGG]_7$	CTCCGAGCCCTAGCAACTTT	131	single locus, monomorphic
		TCTGGATGACGAAACCTTCACA		
MBo04	$[ACC]_7$	GATGTGGCCGCTCTGATCTC	192	single locus, monomorphic
		ACATGCTGGCAAGGTATTCT		
MBo05	$[AAG]_7$	GTCCTAGCACGCTGGCATTA	202	single locus, monomorphic
		TGGGTTGCCAATGAACCCTT		
MBo06	$[ATC]_7$	TGGCCATTCAACTGCTTCAC	202	single locus, monomorphic
		GAATCTAGCACCAGCAAACCC		
MBo07	$[AAG]_7$	GGCACTGGAGTCCACATCAA	239	single locus, monomorphic
		TCCTTCTGTACTGGCATCTCT		
MBo08	$[AGG]_8$	TGATTGTTTCCTCTTCCCTCCT	90	single locus, monomorphic
		TTAATGAGCCGAAGAGGAGCC		
MBo09	$[AGG]_8$	TCCCTCACTCCCATCCTCTC	163	single locus, monomorphic
		ACTCCACTCCTTCCCTCATACA		
MBo10	$[AAC]_8$	GTTAAAGACGCAGGGCTGGA	166	single locus, monomorphic
		CCCACTTAGTGAGATAAGACTTGA		
MBo11	$[ATC]_8$	GCATCACATGGTTTCAGGCT	219	single locus, monomorphic
		GCTCAACCATCGGCAGTGTA		
MBo12	$[ATC]_9$	GGAGGAAAGGTTGCCCTAGAA	102	single locus, monomorphic
		TCTCAACCTGATGTCATTGCA		
MBo13	$[AAG]_9$	CAGGTTGCATCGGCCCATT	103	complex, polymorphic
		GGAGCCTAATGCACCCAGAG		

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

Conventionally, microsatellite motifs are displayed under the form $[N_1N_2]_x$ or $[N_1N_2N_3]_x$ for dinucleotide and trinucleotide loci, respectively, where N_1, N_2 and N_3 represent nucleotides included in the elementary unit of the motif

Loading [MathJax]/jax/output/CommonHTML/jax.js

pected amplicon size is as predicted by QDD.

Locus name	Repeat motif	Primer sequences (5'-3' orientation)	Expected amplicon size (bp)	Amplification product
MBo14	$[AAC]_9$	ATGGCCGATCCCACTTAGTG	117	single locus, monomorphic
		GAGAGAACGGCAATAATTTATGCA		
MBo15	$[AAG]_{10}$	GCTGAAGAGGATGAAGAAGAAGC	92	complex, monomorphic
		TCATCATCTCCCTCTCCTTCT		
MBo16	$[AGG]_{10}$	CAGCACTGGCCTCACAGC	118	single locus, monomorphic
		CCGTCGATCAGTTGTTGGAGA		
MBo17	$[ATC]_{10}$	ACACAATGACCTTTCGCTGA	124	single locus, monomorphic
		CCAAACAGGACCTTATGCCA		
MBo18	$[AAG]_{10}$	ACATCCTCTCCTTCATCTCCTT	187	complex, polymorphic
		GTTCCTACAATGCTTGGCGC		
MBo19	$[AAG]_{10}$	TGCTATCACCCAATATCTAGGCT	202	single locus, monomorphic
		ACAGTCAACAACTACCATACTGC		
MBo20	$[AAG]_{10}$	TGTGGTTAAAGCAATGGAAGCA	229	single locus, monomorphic
		GCCGAACTCCTACTCTCATACG		
MBo21	$[AAG]_{11}$	ACAACAGAAGATCAGTATACGTTCT	171	single locus, monomorphic
		TTGAGGAATCATGCTTGTCAGT		
MBo22	$[AAG]_{14}$	AGAAGAATTCGGTTAGGTCACAA	108	single locus, monomorphic
		AGATAACATGGGTAAGAATTGCCT		
MBo23	$[AAT]_5$	TGAGTTCTTGTCTTGTCTTCGT	100	single locus, monomorphic
		GGTTTGGGACACCCTTCAGG		
MBo24	$[AAT]_9$	AAAGTCATGTCTGGGTGATGAA	90	single locus, monomorphic
		ATGATGAGCACAGCTACAACTCT		
MBo25	$[AAT]_6$	TCTTCAGGTGACAAGCAACA	96	single locus, monomorphic
		CCTGGGCATGGAGATAGCAT		
MBo26	$[AAT]_7$	CCATAGGCCAGCCCACTATA	134	single locus, monomorphic
		ACCCTTTCTTCTTCCTCATTTGT		
MBo27	$[AAT]_7$	TCTCTATTGCTTGGTGATCCC	103	single locus, monomorphic
		TCCAACAAGGGATGGTTATCATG		

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

Conventionally, microsatellite motifs are displayed under the form $[N_1N_2]_x$ or $[N_1N_2N_3]_x$ for dinucleotide and trinucleotide loci, respectively, where N_1, N_2 and N_3 represent nucleotides included in the elementary unit of the motif

pected amplicon size is as predicted by QDD.

Locus name	Repeat motif	Primer sequences (5'-3' orientation)	Expected amplicon size (bp)	Amplification product
MBo28	$[AAT]_8$	GCCTTGAGAGTGGAAGAGGC	205	single locus, monomorphic
		TCTCTTCTTTGCGCCCTCAT		
MBo29	$[AAT]_{16}$	AGACATGTAGAGGTGGGACT	211	single locus, monomorphic
		TCTGTATGAGAGACGTGTTACAGT		
MBo30	$[AAT]_8$	TGACCATAACAAGCTACCAGGT	146	single locus, monomorphic
		GGTGGAAGCTATTGATATTGCATGT		
MBo31	$[AAT]_{10}$	TGACAATGATGCATGCGATAACA	187	single locus, monomorphic
		GCATCACCCATGTCCTTTAGC		
MBo32	$[AAT]_{10}$	TCCGAGGGCAGTATTTGTCG	117	single locus, monomorphic
		CACTATTTCGGAAACCTAAGCCC		
MBo33	$[AAT]_{17}$	GCACACTTTGTATCCGACGC	147	single locus, monomorphic
		CAGGGATAGTAACCGTCAGGG		
MBo34*	$[AG]_{28}$	GTGGCACCTCTGCGGTTT	192	single locus, polymorphic
		CGAGATGGAAGCACCTGGAG		
MBo35*	$[AG]_{24}$	AGCATGCTTTCTGCTTCATGTG	137	single locus, polymorphic
		CCTTTCCCTGACTGCATTGC		
MBo36	$[AG]_{23}$	TCGGAAGTCGAATGTGGCAG	180	no amplification
		TCGGAAGAGTGGTCAATCATGG		
MBo37	$[AG]_{23}$	GCTCTACTCCCAGAGACGGA	142	complex, polymorphic
		AACAGTCGACGGAATGCTCA		
MBo38*	$[AG]_{20}$	AGTCCTCACTGCTGGTGGTA	130	single locus, polymorphic
		TCCTTGAATAGTCCATCTTGCA		
MBo39	$[AG]_{19}$	AACGCAGGTTAAGAGGCTCC	168	complex, monomorphic
		CCTCCTGGTGCAACCCTTAC		
MBo40	$[AG]_{19}$	TGTGGAGTGTGAGTCGATGG	193	complex, polymorphic
		GGCTGCATAATCTCATCACGC		
MBo41*	$[AG]_{18}$	TTCTCCACCAGCCTCACAAC	184	single locus, polymorphic
		ATACGGCCCATCAACCCTTC		

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

Conventionally, microsatellite motifs are displayed under the form $[N_1N_2]_x$ or $[N_1N_2N_3]_x$ for dinucleotide and trinucleotide loci, respectively, where N_1, N_2 and N_3 represent nucleotides included in the elementary unit of the motif

Loading [MathJax]/jax/output/CommonHTML/jax.js

pected amplicon size is as predicted by QDD.

Locus name	Repeat motif	Primer sequences (5'-3' orientation)	Expected amplicon size (bp)	Amplification product
MBo42	$[AG]_{18}$	CCTGGTGGTACATGTGGTCA	136	complex, polymorphic
		TGTGGCACATTCATTTCTGAAGG		
MBo43	$[AG]_{18}$	AGTTTGTTCTGTGTGTTGTCAC	137	no amplification
		GCACACATCTTGCTTTGAAGAC		
MBo44	$[AG]_{17}$	AACACACTTTAAATCGACTTCTTCA	193	complex, polymorphic
		CACGGCTGCCATGTGAGG		
MBo45	$[AG]_{17}$	TAGATCGGAAGTCAGGCCC	193	no amplification
		AGAGAAGTGGGAGGAGAGGTC		
MBo46	$[AG]_{17}$	GCCGATATTAGCTTCTTCTTGGC	154	single locus, monomorphic
		GCCTTGTTGATCCCGTTTCAC		
MBo47	$[AG]_{16}$	GGCACCTGACGCCTCTTT	188	single locus, monomorphic
		TCACTTCGACTCAATTGTATCCAT		
MBo48	$[AG]_{16}$	AGGACAAAGAGATGAGAAGCCT	92	complex, polymorphic
		ACCAATTCCCAGTTAGTTGACCA		
MBo49*	$[AG]_{16}$	CATCACCCATTCTCTCTGCCT	141	single locus, polymorphic
		GAGAAACCATCCGCACCTCA		
MBo50*	$[AG]_{15}$	AGAAGTCATCTTGAGGGCCC	150	single locus, polymorphic
		TTGCTAGAATGATACACAAATTGCT		
MBo51*	$[AG]_{15}$	TGTGCTATTTGTTGGGAATGCA	191	single locus, polymorphic
		GCAAGCTCATGTTCTAGTTTCAAGT		
MBo52*	$[AG]_{15}$	ACACATCCTACATGAATAGACCTCC	122	single locus, polymorphic
		TCTTGTCATAGCCTAGATTCCCT		
MBo53	$[AG]_{15}$	AGGTTTAAGGGTTTGGGTTAGGG	131	single locus, monomorphic
		GGTGGAGTAAGTTTGAGGGTCA		
MBo54*	$[AG]_{11}NNN[AG]_{15}$	CATATGCTGATACAAGAGAGAGGG	124	single locus, polymorphic
		ACCTTATAAGCAGGATCCAGACA		
MBo55	$[AG]_{15}$	TGGAATCAACCTTGGGTCTACA	198	complex, polymorphic
		TCGTCGGTCTTCTAGCCACT		

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

Conventionally, microsatellite motifs are displayed under the form $[N_1N_2]_x$ or $[N_1N_2N_3]_x$ for dinucleotide and trinucleotide loci, respectively, where N_1, N_2 and N_3 represent nucleotides included in the elementary unit of the motif

pected amplicon size is as predicted by QDD.

Locus name	Repeat motif	Primer sequences (5'-3' orientation)	Expected amplicon size (bp)	Amplification product
MBo56*	$[AG]_{15}$	ACCAAGATCAAGCACGAGGA	103	single locus, polymorphic
		AGGATCACCCTTTCTTTCTTTCT		
MBo57*	$[AG]_{15}$	GGGTTCAATCCTGATGAGAGCA	136	single locus, polymorphic
		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 (*).				
Conventionally, microsatellite motifs are displayed under the form $[N_1N_2]_x$ or $[N_1N_2N_3]_x$ for dinucleotide and trinucleotide loci, respectively, where N_1, N_2 and N_3 represent nucleotides included in the elementary unit of the motif and x is the number of unit repetitions. Expected amplicon size is as predicted by QDD.				

Microsatellite-based characterization of genetic variation of **B. aethiopum** in Benin

The previously defined set of 11 *B. aethiopum*-specific SSR markers has been used for the characterization of genetic diversity in our full panel of 180 individual samples from nine locations distributed across Benin. As shown in Table 3, among our sample set the number of alleles per microsatellite locus ranges from 2 for marker Mbo41 to 6 for markers Mbo34, Mbo35, and Mbo50, with an average value of 4.27, whereas expected heterozygosity (He) values range from 0.031 (marker Mbo56) to 0.571 (marker Mbo35). Using these markers, the analysis of genetic diversity (Table 4) shows that the number of polymorphic markers detected at the microsatellite loci investigated ranges from 8 (sites of Togbin and Malanville) to 10 (Savè, Agoua, Pendjari, Pingou and Trois Rivières), with a mean value of 9 ± 0.865. With the exception of Savè, Hounviatouin and Malanville, 1 to 3 private alleles of the targeted microsatellite loci are observed in most sampling locations. 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 Pingou, Malanville and Trois rivières whereas positive F values, indicating a deficit of heterozygosity, are observed in all other sites investigated.

Table 3
Characteristics of 11 polymorphic microsatellites markers used for genetic diversity analysis of *B. aethiopum*

Locus name	Number of alleles scored/locus	Expected Heterozygosity (He)	Observed Heterozygosity (Ho)
Mbo34	6	0.520	0.383
Mbo35	6	0.571	0.522
Mbo38	5	0.458	0.513
Mbo41	2	0.343	0.356
Mbo49	4	0.167	0.146
Mbo50	6	0.548	0.542
Mbo51	3	0.320	0.304
Mbo52	3	0.201	0.232
Mbo54	4	0.26	0.435
Mbo56	3	0.031	0.034
Mbo57	5	0.296	0.263

Table 4
Mean diversity parameters for each of the nine *B. aethiopum* sampling sites.

Geo-climatic region	Site	Number of polymorphic markers	Na	Ne	Number of private alleles	Ho	He	F
Guineo-Congolian (South)	*Togbin*	8	2.273	1.584	3	0.234	0.288	0.145
	Hounviatouin	9	2.182	1.447	0	0.272	0.263	0.007
Sudano-Guinean (Center)	*Savè*	10	2.909	2.069	0	0.384	0.451	0.134
	Biguina	9	2.364	1.770	2	0.345	0.374	0.064
	Agoua	10	2.273	1.722	1	0.329	0.358	0.059
Sudanian (North)	*Pendjari*	10	2.818	1.900	3	0.368	0.396	0.055
	Pingou	10	2.364	1.906	1	0.405	0.390	-0.063
	Malanville	8	2.455	1.627	0	0.302	0.303	-0.020
	Trois rivières	10	2.545	1.822	2	0.373	0.360	-0.055
	Overall mean	**9 ± 0.865**	**2.465 ± 0.103**	**1.761 ± 0.065**		**0.335 ± 0.023**	**0.354 ± 0.023**	**0.035 ± 0.022**

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

Nei's genetic distance among locations (Table 5) ranges 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 *B. aethiopum* sampling locations are lowest within the same region, with the lowest genetic distances among the sites of Pendjari, Pingou, and Trois Rivières which are all located in the Northern part of the country. One interesting exception is the Central (Guineo-Sudanian) region of Benin, where we find that the most genetically distant location from Savè is the one from the Agoua forest reserve (0.339). Surprisingly, Savè displays its highest genetic identity value when compared to the other two collection sites located within 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 geographic distances involved.

Table 5
Pairwise location 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 structure of genetic distances emerges from the analysis of pairwise location genetic differentiation (Fst) (Table 6), suggesting genetic differentiation according to geographic distances between collection sites, with the notable exception of the lower genetic differentiation between samples from Savè and those from either one of the forest reserves in the Northern region, namely Pendjari and Trois Rivières.

Table 6
Pairwise sampling locations 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

In order to assess the strength of the relationship between genetic and geographic distances, we plotted them as a linear regression and performed the Mantel permutation test. As shown in Fig. 1, the positive correlation between both variables is weak, but significant (R^2 = 0.1139, P = 0.040).

AMOVA (Table 7) shows that within-site variation underlies the major part (53%) of total variance, whereas among-site and among-regions variations explain genetic variance to a similar extent (23 and 24%, respectively). Accordingly, the average Number of migrants between collection sites (Nm = 1.019) is low, indicating very limited gene flow.

Table 7
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 Locations	6	254.302	42.384	1.903	23%	< 0.001
Within Locations	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

The Principal Coordinates Analysis (PCoA) of 180 *B. aethiopum* samples (Fig. 2A) shows that the first axis (accounting for 24% of total variation out of a sum of 33.90 for axes 1 and 2) roughly separates individual samples in two main groups, a result that is in agreement with the analysis of genetic distances. The sampling locations-based PCoA (Fig. 2B) confirms the genetic separation along the first axis (accounting for 44.08% of total variation over a total of 61.06% for the sum for axes 1 and 2) between sites from the Guineo-Congolian (Southern) region, plus the sites of Agoua and Biguina (Center) *vs.* sites from the Sudanian (Northern) region, plus the site of Savè (Center). Although the distinction is not as clearly marked, the second axis (accounting for 16.98% of total variation) further allows to distinguish two subgroups within the first group, corresponding to sites belonging to the Southern region and to those from the Central one, respectively.

Likewise, the Bayesian analysis of our data indicates an optimal value of K = 2 for the clustering of the samples into two groups (Fig. 3A and Fig. 3B): one group that includes samples from Togbin and Hounviatouin in the Southern part of the country, as well as most samples from Biguina and Agoua at the Western (Togolese) border of the Centre region; and one group composed of the majority of samples collected in Savè (Eastern part of the Centre region) and from the Northern locations of Pendjari, Pingou, Malanville, and Trois Rivières. Since there is a possibility that the ΔK method used for estimating K leads to over- or under-estimated values, clustering with a value of K = 3 has also been tested (Fig. 3C). As previously observed with the location-based PCoA, under this hypothesis further clustering emerges within the first group, involving samples from Togbin and Hounviatouin (South) and those from Biguina and Agoua (Center), respectively.

The Unweighted pair-group method with arithmetic mean (UPGMA) tree constructed from our data (Fig. 4) distinguishes two main groups matching the ones defined through the Bayesian analysis with K = 2, and which are supported by bootstrap values above 50. Within each of these groups, subgroups corresponding to those observed with K = 3 clustering and that globally match geo-climatic regions (Savè excepted) can further be defined. However, in this case most bootstrap values attached to these secondary branches are not significant.

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 [50]. Nevertheless, this method has been used successfully for accelerating the analysis of genetic diversity in many plant species, including palms [11, 51–53]. In the present study, we find that the transferability rate of microsatellite markers developed in other palms genera to *B. aethiopum*, *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 *E. guineensis*, *P. dactylifera* or *C. nucifera*, 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 [49], 75% from *E. oleifera* to *E. guineensis* [53], 86% between the wooly jelly palm (*Butia eriospatha* Mart.) and related species *B. catarinensis* [54] and up to 100% in the licuri palm (*Syagrus coronate* Mart) [55]. When considering other plant families, our transferability rate is also markedly lower than both the average rate of 50% found by Peakall *et al.* [56] within the *Glycine* genus and among Legumes genera, and the overall rate of 35.2% calculated by Rossetto [57] 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 Fan *et al.* [1], 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: *P. dactylifera* genome is estimated to be 671 Mb [39] whereas the *E. guineensis* genome is 1.8–1.9 Gb [37, 58] and *C. nucifera* genome is 2.42 Gb [42]. It is possible that these differences in genome sizes among related diploid plant species rely on differences in transposable element (TE) content, which in turn might have induced structural alterations throughout the genome through indels, copy number variations and recombinations [59, 60]. The illustration of such a mechanism working at the intra-genus level has been provided by cultivated rice species *Oryza sativa* L. and its wild relative *O. australiensis* [59]. Ultimately, TE-induced structural variations may have a negative effect on the cross-species amplification ability of some of the SSR primers. Indeed, in a recent study Xiao *et al.* [49] showed that over 70% of the conserved microsatellite loci between *E. guineensis* and *P. dactylifera* are located within genic regions of the genome with low TE content, and which are therefore less likely to be submitted to TE-dependent structural variations. More generally, gaining a better understanding of genome

also help reconcile our results with previous published reports of successful

transfer of SSR markers developed from other palm sources to *B. flabellifer* (see references cited in Table 8, Methods section). Indeed, since the genome size of *B. flabellifer* (7.58 Gb; Jaume Pellicer, unpublished data) is only marginally smaller than that of *B. aethiopum*, significant differences in genome composition may be underlying the lack of SSR transferability between both species.

Table 8
Characteristics of the palm SSR markers tested for transferability to *B. aethiopum*.

Marker name	Primer sequences (5'-3' orientation)	T_a (°C)	Source palm species	Successful transfer to other palm species	References
mEgCIR0230	CCCTGGCCCCGTTTTTC	57.0	*E. guineensis*	*E. oleifera*	[73]
	AGCGCTATATGTGATTCTAA			*Syragus* sp.	
mEgCIR0326	GCTAACCACAGGCAAAAACA	59.0		*C. nucifera*	
	AAGCCGCACTAACATACACATC			*P. roebelinii*	
mEgCIR0465	TCCCCCACGACCCATTC	63.1		*P. canariensis*	
	GGCAGGAGAGGCAGCATTC			*P. reclinata*	
mEgCIR0476	TTCCTCGGCCCCTTCTC	61.6			
	TCGCCGACCTTCCACTG				
EgCSSR-5781	TTCACGCTACTGATGGTTGG	59.4	*E. guineensis*	*B. flabellifer*	[49]
	TCGATCCCTTCTCTGGAAAC				
EgCSSR-1461	GTCCTCTCCTACGCCTCCTC	60.3			
	ATGCGATCCGAGTTCAGAAG				
mEgCIR2332	GAAGAAGAGCAAAAGAGAAG	55.0	*E. guineensis.*	*B. flabellifer*	[44, 45]
	GCTAGGTGAAAAATAAAGTT				
mEgCIR3295	TGCCTCCAGACAATCAC	55.0			
	GTAAGGCTTAACCAGATAAC				
mEgCIR3311	AATCCAAGTGGCCTACAG	55.0			
	CATGGCTTTGCTCAGTCA				
mEgCIR3413	AAAGCTATGGGGTGAAAGAT	55.0			
	TGGATAAGGGCGAGAAGAGA				
mEgCIR3477	CCTTCAAGCAAAGATACC	55.0			
	GGCACCAAACACAGTAA				
mEgCIR3592	GAGCCAAAACAGACTTCAA	55.0			
	ACCGTATATGACCCCTCTC				
mEgCIR3755	GCTCACCAAAAAGTGTTAAGTC	55.0			

For each marker, forward (top) and reverse primers (bottom) are provided.

T_a: average annealing temperature for each primer pair.

Species names are abbreviated as follows: *P. roebelinii: Phoenix roebelinii; P. canariensis: Phoenix canariensis; Phoenix reclinata; H. thebaica : Hyphaene thebaica*; *L. carinensis : Livistona carinensis; C. humilis : Chamaerops humilis ; K. laciniosa : Korthalsia laciniosa; Z. zalacca : Zalacca zalacca ; D. kurzianus : Daemonorops kurzianus ;C. simplicifolia : Calamus simplicifolia ;C. mannan : Calamus mannan ; C. thwaitesii : Calamus thwaitesii ; C. erectus : alustris; P. rupicola : Phoenix rupicola; P. theophrasti: Phoenix theophrasti.*

Marker name	Primer sequences (5'-3' orientation)	T_a (°C)	Source palm species	Successful transfer to other palm species	References
	AGTTTCAACGGCAGGTATAT				
mEgCIR3788	TTGTATGACCAAAGACAGC	55.0			
	AGCGCAACATCAGACTA				
ESSR75	AGATGGTTGGAGATTTCATGGT	60.0	*E. guineensis*	*B. flabellifer*	[44, 45, 47]
	AACTTGAGGGTGCCATTACAAG				
ESSR76	CCATACCAGCAGAAGAGGATGT	60.0			
	CTGAAGGTCATAGGGGTCTCTG				
ESSR82R	CCCTCGACACCCATAGTTATTT	60.0			
	CTCGATTTCTGGCCTCTCATAC				
ESSR332	AGTTAATGTGTCAGGGCCAGTT	60.0			
	CTTGGTTCACTTGGGTGTGTC				
ESSR553	ATAAATTGTGCGAGGGGAAAAC	60.0			
	AGATCCGCGACAGGTCTTAAC				
ESSR566	GTGTCATCAAATTCGGTCCTTT	60.0			
	CGGTTCTTCTGCTGCTCTACTT				
ESSR609	AGGCGGTGATGAAGATGAAG	59.0			
	CTCCTCTCAAACAGAGTGGGAT				
ESSR650	GCCTTTTCTGGTTAATGGACTG	59.0			
	GTTTGTCTATGGATGATTGTGAGG				
ESSR652	CATACCGTCACCACTCAGAAAC	60.0			
	GCCGTCATTCTACCAGTTGAG				
ESSR673	TTCTGGCTACGAGCATAAGGA	59.0			
	TCAATAACCCTGGCTAAACACA				
ESSR681	TCTGAATTGTCGGAGTGGC	59.0			
	CATCCTTGCGTAAACAAAAGAG				
CNZ34	CATGTCGATAATTATACCCAA	55.0	*C. nucifera*	*B. flabellifer*	[46, 75]

For each marker, forward (top) and reverse primers (bottom) are provided.

T_a: average annealing temperature for each primer pair.

Species names are abbreviated as follows: *P. roebelinii: Phoenix roebelinii; P. canariensis: Phoenix canariensis; Phoenix reclinata; H. thebaica : Hyphaene thebaica*; *L. carinensis : Livistona carinensis; C. humilis : Chamaerops humilis ; K. laciniosa : Korthalsia laciniosa; Z. zalacca : Zalacca zalacca ; D. kurzianus : Daemonorops kurzianus ;C. simplicifolia : Calamus simplicifolia ;C. mannan : Calamus mannan ; C. thwaitesii : Calamus thwaitesii ; C. erectus : [illegible]alustris; P. rupicola : Phoenix rupicola; P. theophrasti: Phoenix theophrasti.*

Marker name	Primer sequences (5'-3' orientation)	T_a (°C)	Source palm species	Successful transfer to other palm species	References
	TGCAAATATGAATGCAAACAC			*K.laciniosa*	
CN2A5	AAGGTGAAATCTATGAACACA	53.2		*Z. zalacca*	
	GGCAGTAACACATTACACATG			*D.kurzianus*	
				C.simplicifolia	
				C. mannan	
				C. thwaitesii	
				C. erectus	
				C. palustris	
CNZ 12	TAGCTTCCTGAGATAAGATGC	54.6	*C. nucifera*	*B. flabellifer*	[46, 76]
	GATCATGGAACGAAAACATTA			*P. dactylifera*	
CNZ 24	TCCTAAGCTCAATACTCACCA	55.0		*E. guineensis*	
	CGCATTGATAAATACAAGCTT				
CNZ 18	ATGGTTCAGCCCTTAATAAAC	60.3			
	GAACTTTGAAGCTCCCATCAT				
CNZ 42	TGATACTCCTCTGTGATGCTT	55.5			
	GTAGATTGTGGGAGAGGAATG				
CN2A4	CAGGATGGTTCAAGCCCTTAA	61.0			
	GGTGGAAGAGGGAGAGATTGA				
CAC 21	AATTGTGTGACACGTAGCC	54.1	*C. nucifera*	*B. flabellifer*	[74, 77]
	GCATAACTCTTTCATAAGGGA				
CAC 71	ATAGCTCAAGTTGTTGCTAGG	54.2			
	ATATTGTCATGATTGAGCCTC				
CAC 84	TTGGTTTTTGTATGGAACTCT	54.4			
	AAATGCTAACATCTCAACAGC				
CN1H2	TTGATAGGAGAGCTTCATAAC	53.2	*C. nucifera*	*B. flabellifer*	[74]
	ATCTTCTTTAATGCTCGGAGT			*P. dactylifera*	

For each marker, forward (top) and reverse primers (bottom) are provided.

T_a: average annealing temperature for each primer pair.

Species names are abbreviated as follows: *P. roebelinii: Phoenix roebelinii; P. canariensis: Phoenix canariensis; Phoenix reclinata; H. thebaica : Hyphaene thebaica*; *L. carinensis : Livistona carinensis; C. humilis : Chamaerops humilis ; K. laciniosa : Korthalsia laciniosa; Z. zalacca : Zalacca zalacca ; D. kurzianus : Daemonorops kurzianus ;C. simplicifolia : Calamus simplicifolia ;C. mannan : Calamus mannan ; C. thwaitesii : Calamus thwaitesii ; C. erectus : C. palustris; P. rupicola : Phoenix rupicola; P. theophrasti: Phoenix theophrasti.*

Marker name	Primer sequences (5'-3' orientation)	T_a (°C)	Source palm species	Successful transfer to other palm species	References
PdAG-SSR	TCTGATTTCGTTTACTTCTTAGGA	58.0	*P. dactylifera*		[44]
	TTCATATTCAGTTGTCGGGTGTA				
mPdClR015	AGCTGGCTCCTCCCTTCTTA	59.1			
	GCTCGGTTGGACTTGTTCT				
mPdClR063	CTTTTATGTGGTCTGAGAGA	52.5			
	TCTCTGATCTTGGGTTCTGT				
mPdIRD1	CTCGGAAGGGTATGGACAAA	59.6	*P. dactylifera*	*P. reclinata* *P. roebelenii* *P. rupicola* *P. theophrasti* *H. thebaica* *L. carinensis* *C. humilis*	[72]
	TTGCCTTCGACGTGGTAGTA				
mPdIRD3	CATTGATCCAACACCACCAC	60.3			
	GCCAAAACCAGCTCTGGTAAC				
mPdIRD4	TTGGTGGCCTTTCTCAGAGT	59.8			
	TGGGATCAAAGTAGGGTTGG				
mPdIRD5	CTATCAGGATGGGGGTGATG	60.2			
	ACCCATCTGCATAGCTCCAG				
mPdIRD7	TGCAATACGATGGCAGAGTC	60.2			
	CCTTGCAAGTTTTCCACACC				
mPdIRD8	CTATTGGGTCCCTTGGTGAG	59.7			
	TGACTGCTCGTCATCAGGTC				
mPdIRD10	ATGCGTTCATCTCCCTTGAG	59.7			
	GCTGCAAACATCATCCTCAC				
mPdIRD11	GAGTTGGAGGCAAAACCAGA	59.8			
	CCACAAAACCCTTGTCTTCC				
mPdIRD14	GAGGGGTTCACGTTTGTGTC	60.9			
	GCACCAAGCACAAGAGCAAT				
mPdIRD15	CCGAGTCTGGCGAAGTAAAC	60.0			
	CTCCCCTTCCTCATCCTCTC				

For each marker, forward (top) and reverse primers (bottom) are provided.

T_a: average annealing temperature for each primer pair.

Species names are abbreviated as follows: *P. roebelinii: Phoenix roebelinii; P. canariensis: Phoenix canariensis; Phoenix reclinata; H. thebaica : Hyphaene thebaica*; *L. carinensis : Livistona carinensis; C. humilis : Chamaerops humilis ; K. laciniosa : Korthalsia laciniosa; Z. zalacca : Zalacca zalacca ; D. kurzianus : Daemonorops kurzianus ;C. simplicifolia : Calamus simplicifolia ;C. mannan : Calamus mannan ; C. thwaitesii : Calamus thwaitesii ; C. erectus : alustris; P. rupicola : Phoenix rupicola; P. theophrasti: Phoenix theophrasti.*

Marker name	Primer sequences (5'-3' orientation)	T_a (°C)	Source palm species	Successful transfer to other palm species	References
mPdIRD16	CTGTCCGATCGAATTCTGC	50.7			
	GGACATCTCTTTGCGGTCAT				
mPdIRD17	GTGGGAGAAACCCGAAGAAT	60.2			
	CTGCTGCCTCATCTGCATT				
mPdIRD20	TTGAATGGTCCCCTGTAGGT	59.5			
	GTCCCAGCATGATTGCAGTA				
mPdIRD22	GGCTGTATGGGAAAGACCTG	59.5			
	CCTGCTGCATATTCTTCGTG				
mPdIRD24	GCTCCTGCAGAACCTGAAAC	59.9			
	GGACATCACCGTCCAATTCT				
mPdIRD25	CACTGGAAATTCAGGGCCTA	59.9			
	CCCAATTTCTCAGCCAAGAC				
mPdIRD26	CCTCCAGTTCATGCTTCTCC	60.0			
	GAGCAGACCCGACAGACAAT				
mPdIRD28	GAAACGGTATCGGGATGATG	59.7			
	TTAACGACGCCGTTTCCT				
mPdIRD29	GGCTCCACCATCATTGACA	60.3			
	AACAGCATCGACTGCCTTCT				
mPdIRD30	GCAGATGGTTGAAAGCTCCT	59.8			
	CCCCATTAACAGGATCAACG				
mPdIRD31	GCAGGTGGACTGCAAAATCT	60.0			
	CTATTGGGGTGCTGATCCAT				
mPdIRD32	AAGAAGACATTCCGGCTGGT	59.9			
	GCGGGTGTGTGATATTGATG				
mPdIRD33	GGAGCATACAGTGGGTTTGC	60.1			
	CAGCCTGGGAATGAGGATAG				

For each marker, forward (top) and reverse primers (bottom) are provided.

T_a: average annealing temperature for each primer pair.

Species names are abbreviated as follows: *P. roebelinii: Phoenix roebelinii; P. canariensis: Phoenix canariensis; Phoenix reclinata; H. thebaica : Hyphaene thebaica*; *L. carinensis : Livistona carinensis; C. humilis : Chamaerops humilis ; K. laciniosa : Korthalsia laciniosa; Z. zalacca : Zalacca zalacca ; D. kurzianus : Daemonorops kurzianus ;C. simplicifolia : Calamus simplicifolia ;C. mannan : Calamus mannan ; C. thwaitesii : Calamus thwaitesii ; C. erectus : alustris; P. rupicola : Phoenix rupicola; P. theophrasti: Phoenix theophrasti.*

Marker name	Primer sequences (5'-3' orientation)	T_a (°C)	Source palm species	Successful transfer to other palm species	References
mPdIRD35	CAGCCCCTTACTCAGACTGG	59.6			
	CCCATAAGCTGATTGTGCTG				
mPdIRD36	GACACGTTGACGATGTGGAA	60.7			
	CCATTGCTGTTGAGGAGGAG				
mPdIRD37	TTTCCTGCTCGAAAGACACC	60.2			
	CTTAGCCAGCCTCCACACTC				
mPdIRD40	GAGAGATGCGTCAGGGAATC	59.2			
	CCAGAATCTTCCAAGCAAGC				
mPdIRD42	GAGGCAAAACTATGGGAAGC	59.5			
	TTCACTGGAGCAAGGGTAGG				
mPdIRD43	GCAGCCATTGCTTACAGTGA	60.2			
	TAAACTGCTGCCTTCCTTGG				
mPdIRD44	CAGATCCGGGAGATGATGAA	60.4			
	AGCAGGAGCAGCTGCATAA				
mPdIRD45	TAGCCTGTGCATGTTCGTTG	60.4			
	AACAGCAGCTGATGGTGATG				
mPdIRD46	ATGGGTCCATTGGAGGAACT	60.2			
	GACGGAGACCTTGACTGCTC				
mPcClR10	ACCCCGGACGTGAGGTG	62.8	*P. dactylifera*		Cherif, Castillo and Aberlenc-Bertossi, unpublished data.
	CGTCGATCTCCTCCTTTGTCTC				
mPcClR20	GCACGAGAAGGCTTATAGT	51.7			
	CCCCTCATTAGGATTCTAC				
mPcClR32	CAAATCTTTGCCGTGAG	53.3			
	GGTGTGGAGTAATCATGTAGTAG				
mPcClR35	ACAAACGGCGATGGGATTAC	60.8			
	CCGCAGCTCACCTCTTCTAT				

For each marker, forward (top) and reverse primers (bottom) are provided.

T_a: average annealing temperature for each primer pair.

Species names are abbreviated as follows: *P. roebelinii: Phoenix roebelinii; P. canariensis: Phoenix canariensis; Phoenix reclinata; H. thebaica : Hyphaene thebaica*; *L. carinensis : Livistona carinensis; C. humilis : Chamaerops humilis ; K. laciniosa : Korthalsia laciniosa; Z. zalacca : Zalacca zalacca ; D. kurzianus : Daemonorops kurzianus ;C. simplicifolia : Calamus simplicifolia ;C. mannan : Calamus mannan ; C. thwaitesii : Calamus thwaitesii ; C. erectus : alustris; P. rupicola : Phoenix rupicola; P. theophrasti: Phoenix theophrasti.*

Marker name	Primer sequences (5'-3' orientation)	T_a (°C)	Source palm species	Successful transfer to other palm species	References
mPcCIR50	CTGCCATTTCTTCTGAC	50.6			
	CACCATGCACAAAAATG				
mPcCIR57	AAGCAGCAGCCCTTCCGTAG	62.0			
	GTTCTCACTCGCCCAAAAATAC				
mPcCIR85	GAGAGAGGGTGGTGTTATT	51.8			
	TTCATCCAGAACCACAGTA				
mPdIRD41	ATCTTCCATGCAGCCTCAAG	60.3			
	CAGGTCGTCCCGTCTCTAAA				
mPdIRD47	GTTGGCATCACTTCAGAGCA	60.1			
	GCTCTTTCGGTGCTAGTTGC				
For each marker, forward (top) and reverse primers (bottom) are provided.					
T_a: average annealing temperature for each primer pair.					
Species names are abbreviated as follows: *P. roebelinii: Phoenix roebelinii; P. canariensis: Phoenix canariensis; Phoenix reclinata; H. thebaica : Hyphaene thebaica*; *L. carinensis : Livistona carinensis; C. humilis : Chamaerops humilis ; K. laciniosa : Korthalsia laciniosa; Z. zalacca : Zalacca zalacca ; D. kurzianus : Daemonorops kurzianus ;C. simplicifolia : Calamus simplicifolia ;C. mannan : Calamus mannan ; C. thwaitesii : Calamus thwaitesii ; C. erectus : Calamus erectus ; C. palustris : Calamus palustris; P. rupicola : Phoenix rupicola; P. theophrasti: Phoenix theophrasti.*					

In any case, from the low number of successfully transferred microsatellite markers we could only identify one displaying polymorphism in our *B. aethiopum* 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, this low diversity might also result from the extremely long juvenile phase that has been attributed to this palm species. Indeed, floral maturity has been reported to occur 30 to 50 years after germination [61]. The manner of seed and pollen dispersal, which have so far not been studied in *B. aethiopum*, might also play a role. Indeed, in pollen-mediated gene flow species, the distance the pollen travel is of importance in the occurrence of crossing between populations [62, 63].

Regarding the development of novel SSR markers, our results are similar to other studies based on the use of high-throughput sequencing techniques in species where very little information is available [22, 64]. We identified 57 microsatellite loci, from which we selected 11 markers displaying polymorphism that were used to assess the genetic structure of *B. aethiopum* sampled from different sites in Benin. We find low genetic diversity, with an average He value (0.354) that is substantially below those reported for *B. flabellifer* (0.417) [45] and for other non-timber forest products such as *Khaya senegalensis* (0.53) [65] and *Phyllanthus sp.* (0.607 and 0.582 for *P. emblica* and *P. indofischeri* respectively [66]. The positive F value that we observed in the majority (6 out of 9) of locations in the present study indicates an overall deficiency of heterozygotes across sites. This deviation from the Hardy-Weinberg equilibrium (HWE) might reflect low gene flow through pollen and seed dissemination, leading to crosses between related individuals, as supported by the low average number of migrants between sites. Accordingly, our data reveal limited genetic distances among collection sites, with values lower than those reported for others palm species. Indeed for *B. flabellifer*, genetic

nd among natural *E. guineensis* accessions an average of 0.769 was

observed [68]. Both our Fst values and AMOVA analysis point to intra-site differentiation as being the main source of genetic variation.

As illustrated by the global agreement between our PCoA and Bayesian analyses, Beninese *B. aethiopum* samples cluster into two main groups that are mostly dependent on geo-climatic regions and geographic distances between collection sites, although the correlation between genetic and geographic distance is poorly significant. There might be further genetic separation between Southern *B. aethiopum* samples and those from the Central sites of Agoua and Biguina, resulting in the splitting of one group into two subgroups. However, with our current dataset it is not possible to achieve this level of discrimination in our analyses. Additional sampling campaigns from intermediate locations in the Central and Northern regions will be necessary in order to make progress on the subject.

Among the nine locations studied in Benin, samples from Savè appear to be the most diversified (He = 0.451) and constitute the exception to the general distribution according to geographical distances. This site located in the Sudano-Guinean transition zone of Benin is currently the most active for the production of *B. aethiopum* hypocotyls, and it acts as a supplier for the whole national territory (V.K. Salako, personal communication), suggesting that it might be the largest population of *B. aethiopum* in the country. Moreover, individuals sampled in Savè appear to be genetically distinct from those sampled in other locations of the Central region and closer to those originating from the Northern region, despite the considerable geographical distances involved in 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, Salako et al. [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 likely 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 [69].

The specific microsatellite markers developed in this study from the partial genomic sequencing of *B. aethiopum* appear to be efficient to assess the genetic diversity and population structure of this species. 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 [70]. High-throughput sequencing techniques are an effective way of developing new microsatellite markers in plant species without significant molecular data. The increasing technical performances and financial affordability of these technologies make it feasible to overcome the difficulties arising in case studies such as ours, where marker transfer was proved to be limited or ineffective.

Conclusions

To our knowledge, the data presented in the present article constitute the first sizeable molecular resource available for *B. aethiopum*, 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 other species within the *Borassus* genus, and of palms diversity as a whole.

Methods

Plant material sampling and DNA extraction

Samples of *B. aethiopum* were collected in nine distinct sites (three located in protected forest areas, six in farmlands) that were distant from each other by at least 50 km and which spanned the three main climatic regions encountered in Benin (Fig. 5). According to White [71], Benin covers three contrasted climatic regions which are the Sudanian region in the North, the Sudano-Guinean region in the Centre and the Guineo-Congolian region in the South. Along a South-North gradient, the rainfall regime switches from bimodal to unimodal, the climate becomes globally drier [29] and the density of *B. aethiopum* distribution increases [31]. At each location, young leaves from 10 male and 10 female adult trees separated by at least 100 m were collected and stored in plastic bags containing silica gel until further processing. The complete list of samples and their characteristics is available in Additional file 2.

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

A total of 80 SSR markers from previous studies were selected for assessment of their transferability to *B. aethiopum*: 44 developed for *P. dactylifera* [72]; 25 developed for *E. guineensis* [44, 73]; and 11 developed for *Cocos nucifera* [74]. The respective sequences and origins of these primer sets are displayed in Table 8.

Transferability of the 80 palm SSR markers was assessed on a representative subset of 20 *B. aethiopum* individuals sampled at the different locations, plus four positive controls from each

source species for these markers (*i.e. P. dactylifera, C. nucifera*, and *E. guineensis*). Microsatellite amplification was performed with a modification of the M13-tailed Primers protocol [75] adapted to the use of fluorescent labelling [76]. The PCR reaction was performed on 20 ng of leaf DNA in volume of 20 µL with the following final concentrations or amounts: 1X PCR buffer, 200 µM dNTP, 2 mM MgCl2, 0.4 pmol M13-tailed forward primer, 4 pmol M13 primer, (5'-CACGACGTTGTAAAACGAC-3') fluorescently labeled at the 5' end with FAM, HEX or TAMR, 4 pmol reverse primer, and 0.5 U of KAPA *Taq* polymerase (Sigma-Aldrich). 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 separation through capillary electrophoresis on an Applied Biosystems 3500 Genetic Analyzer (Thermo Fisher Scientific). Amplification products visualization was performed using the GeneMapper software version 3.7 (Applied Biosystems).

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

One *B. aethiopum* leaf sample (originating from the Togbin site) was randomly selected and used for genomic DNA purification according to the protocol of Mariac *et al.* [77]. The DNA was then used for the construction of an Illumina paired-end library, as described in Mariac *et al.* [81], before high-throughput sequencing on a MiSeq v3 platform (Illumina; average read size 250 bp). 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 version 1.10 [82] (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 under BioProject ID PRJNA576413. Paired-end reads were then merged using FLASH version 1.2.11 (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 software version 3.1.2 [83] with default settings (Additional file 3: File S1).

Using selected primer pairs, test amplifications were performed with two randomly selected *B. aethiopum* 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. Separation and vizualization of amplification products were performed as described previously. Finally, the primer pairs enabling successful and unambiguous amplification of polymorphic bands were used for the analysis of genetic diversity among the complete set of 180 *B. aethiopum* individuals under the same PCR conditions.

Data analysis

Amplification products were scored using the GeneMapper software version 3.7 (Applied Biosystems) and only unambiguous amplification products were considered for data analysis. Genetic diversity parameters were calculated for each locus and each sampling location using the GenAlEx software version 6.502[84]. Expected heterozygosity (He) was calculated using the formula:

$$He = 1 - \sum p_i^2$$

where p_i is the frequency of each allele. The fixation index (F) was calculated as:

$$F = 1 - \frac{H_0}{He}$$

where Ho is observed heterozygosity and He is expected heterozygosity [85].

F-statistics analysis assessing genetic differentiation (Fst), genetic identity, number of migrants (Nm) [86] and analysis of molecular variance (AMOVA) for estimating genetic differentiation within and among locations were performed with the same software. The Mantel permutation test was used for assessing the correlation between genetic and geographic distances between sampling sites [87, 88]. Two Principal Coordinates Analyses (PCoA) enabling the visualization of genetic variation distribution across individuals and sampling sites, respectively, were performed using GenAlEx. The STRUCTURE software version 2.3.4[89] was used 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 burn-in periods and 50,000 Markov chain Monte Carlo (MCMC) replicates. The output from this analysis was then used as input in the Structure HARVESTER online program version 0.6.94 (http://taylor0.biology.ucla.edu/structureHarvester/)to determine the optimal value of K using the ΔK method of Evanno *et al.* [90] and allowing for different estimates of K in accordance with Janes *et al* [91]. Based on the resulting values of K, a clustering analysis of the studied sampling sites was performed. In order to further assess genetic clustering, a UPGMA tree based on Fst values using 1,000 bootstrap replications was constructed using the POPTREE2 software [92].

Abbreviations

AMOVA: Analysis of molecular variance; F: Fixation index; F_{st}: inter-population genetic differentiation coefficient; He: Expected Heterozygosity; Ho: Observed Heterozygosity; HWE: Hardy-Weinberg equilibrium; Na: average number of different alleles; Ne: effective number of alleles; N_m : Number of migrants; PCoA: Principal coordinate analysis; SSR: Simple sequence repeat; UPGMA: Unweighted pair-group method with arithmetic mean

Declarations

Ethics approval and consent to participate

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).

Consent for publication

Not applicable

Availability of data and materials

Data generated from genome sequencing (filtered reads) were deposited into GenBank under BioProject ID PRJNA576413. Capillary electrophoresis profiles are available upon reasonable request to the Corresponding Author. All other data generated or analyzed during this study are included in this published article (and its supplementary information files).

Competing interests

The authors declare that they have no competing interests

Funding

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

Authors' contributions

VKS, EJ, TB and KA conceived and designed the experiments and were responsible for funding acquisition. All authors were involved in defining the experimental strategy. MJK, SS, ML, CTC, KA, VKS performed the experiments. MJK, KA, SS, CM, LZ and TB processed and analyzed data. All authors contributed to writing the manuscript.

Acknowledgements

The Authors wish to thank Emira Cherif, Karina 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.

References

1. 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;31:1271–82.
2. Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira MLC. Origin, evolution and genome distribution of microsatellites. Genetics Molecular Biology. 2006;29:294–307.
3. 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;106:1379–89.
4. Moon HS, Nicholson JS, Lewis RS. Use of transferable Nicotiana tabacum L. microsatellite markers for investigating genetic diversity in the genus Nicotiana. Genome. 2008;51:547–59.
5. 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;8:7832.
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;66:1189–203.
7. 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;11:1–13.
8. Ellegren H. Microsatellite mutations in the germline:: implications for evolutionary inference. Trends Genet. 2000;16:551–8.
9. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. Euphytica. 2011;177:309–34.
10. 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;32:261–85.
11. 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;6:1–7.
12. 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. G3: Genes|Genomes|Genetics. 2012;2:769–77.
13. 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? Mol Ecol. 2003;12:1339–48.
14. Mnejja M, Garcia-Mas J, Audergon J-M, Arús P. Prunus microsatellite marker transferability across rosaceous crops. Tree Genetics Genomes. 2010;6:689–700.
15. Wünsch A. Cross-transferable polymorphic SSR loci in Prunus species. Sci Hortic. 2009;120:348–52.
16. 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;30:335–44.
17. 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;93:1869–79.
18. 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;20:1543–50.
19. 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;135:499–505.

20. Taheri S, Lee Abdullah T, Yusop M, Hanafi M, Sahebi M, Azizi P, et al. Mining and Development of Novel SSR Markers Using Next Generation Sequencing (NGS) Data in Plants. Molecules. 2018;23:2–20.
21. Purwoko D, Cartealy IC, Tajuddin T, Dinarti D, Sudarsono S. SSR identification and marker development for sago palm based on NGS genome data. Breed Sci. 2019;69:1–10.
22. 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;2:1–4.
23. Sun Z-X, Ye L-J, Zhang F, Hu W, Fan D-M, Zhang Z-Y. Development of Microsatellite Markers for Sargentodoxa cuneata (Lardizabalaceae) Using Next-Generation Sequencing Technology. Applications in Plant Sciences. 2016;4:1–4.
24. 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. J Ethnopharmacol. 2015;175:356–69.
25. 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;33:380–91.
26. Salako KV, Moreira F, Gbedomon RC, Tovissodé F, Assogbadjo AE, Glèlè Kakaï RL. Traditional knowledge and cultural importance of Borassus aethiopum Mart. in Benin: interacting effects of socio-demographic attributes and multi-scale abundance. J Ethnobiol Ethnomed. 2018;14:1–16.
27. Ouinsavi C, Gbémavo C, Sokpon N. Ecological Structure and Fruit Production of African Fan Palm (Borassus aethiopum) Populations. American Journal of Plant Sciences. 2011;02:733–43.
28. Gbesso F, Nassi KM, Gbesso GHF. Utilisation sociale de Borassus aethiopum Mart et de ses habitats dans les Communes de Savè et de Glazoué au Bénin. International Journal of Biological Chemical Sciences. 2017;11:1512–22.
29. Adomou CA, Agbani OP, Sinsin B. ,., : Neuenschwander PS. 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. Nigeria: International Institute of Tropical Agriculture, Ibadan; 2011. Sinsin B, Goergen G ,.
30. 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;74:6099–111.
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 Evolution. 2015;148:335–49.
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 strcture in the Pendjari National Park, Benin, West Africa. Tropical Ecology. 2017;:425–37.
33. 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;9:7–16.
34. Ouédraogo A, Boussim J, Zongo J-D, Guinko S. Caracteristiques morphologiques des roniers (Borassus L.) du Burkina Faso. Etudes flo vég Burkina Faso. 2002;:37–40.
35. Dixo M, Metzger JP, Morgante JS, Zamudio KR. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. Biol Cons. 2009;142:1560–9.
36. Jump AS, Penuelas J. Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. Proceedings of the National Academy of Sciences. 2006;103:8096–100.

37. 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;500:335–9.
38. 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;29:521–7.
39. 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;4:1–9.
40. 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;8:1–9.
41. Huang Y-Y, Matzke AJM, Matzke M. Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (Cocos nucifera). PLoS ONE. 2013;8:1–12.
42. Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, et al. The genome draft of coconut (Cocos nucifera). GigaScience. 2017;6:1–11.
43. 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 Biol. 2018;18. doi:10.1186/s12870-018-1509-9.
44. 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.). Theor Appl Genet. 2005;110:754–65.
45. 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 Genet. 2017;18:1–8.
46. Rivera R, Edwards KJ, Barker JHA, Arnold GM, Ayad G, Hodgkin T, et al. Isolation and characterization of polymorphic microsatellites in Cocos nucifera L. Genome. 1999;42:668–75.
47. 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. Mol Breeding. 2014;33:415–24.
48. 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;12:32–43.
49. 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. Front Plant Sci. 2016;7:1–10.
50. 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;16:3759–67.
51. Manju KP, Manimekalai R, Naganeeswaran SA, Biotechnology Section SB, Institute, Coimbatore, Tamil nadu, India, Naganeeswaran SA, Biotechnology Section, Central Plantation Crops Research Institute (ICAR), Kasaragod, Kerala, India, et al. Microsatellites mining in date palm (Phoenix dactylifera L.) and their cross transferability across Arecaceae family. Plant Omics. 2016;9:191–7.
52. 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;10:3942–7.
53. 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;13:4069–88.
54. 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. 2011;98:198–200.
55. 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. Genet Mol Res. 2017;16:1–6.

56. 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. Mol Biol Evol. 1998;15:1275–87.
57. Rossetto M. Sourcing of SSR markers from related plant species. In: Henry RJ, editor. Plant genotyping: the DNA fingerprinting of plants. Wallingford: CABI; 2001. pp. 211–24. doi:10.1079/9780851995151.0211.
58. 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 Rep. 1997;16:884–7.
59. 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;16:1262–9.
60. Hanin M, Paszkowski J. Plant genome modification by homologous recombination. Curr Opin Plant Biol. 2003;6:157–62.
61. Cabannes Y, Chantry G, Willemin V. Le rônier et le palmier à sucre: production et mise en oeuvre dans l'habitat. Montpellier, France; 1987.
62. Rasmussen IR, Brødsgaard B. Gene flow inferred from seed dispersal and pollinator behaviour compared to DNA analysis of restriction site variation in a patchy population of Lotus corniculatus L. Oecologia. 1992;89:277–83.
63. Szczecińska M, Sramko G, Wołosz K, Sawicki J. Genetic Diversity and Population Structure of the Rare and Endangered Plant Species Pulsatilla patens (L.) Mill in East Central Europe. PLOS ONE. 2016;11:1–24.
64. 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;40:41–4.
65. 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;46:756–62.
66. 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;97:1001–6.
67. 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 Biotechnology. 2009;3:11–5.
68. 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;62:349–60.
69. Roth HH, Douglas-Hamilton I. Distribution and status of elephants in West Africa (1). mamm. 1991;55:489–527.
70. Bayton RP, Ouédraogo A, Guinko S. The genus Borassus (Arecaceae) in West Africa, with a description of a new species from Burkina Faso. Bot J Linn Soc. 2006;150:419–27.
71. White. The vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO UNESCO. Nat Resour Res. 1983;:1–356.
72. 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;2:1–5.
73. 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;44:413–25.
74. 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. Theor Appl Genet. 2003;106:715–26.

75. Rao NM, Ramesha BT, Ravikanth G, Ganeshaiah KN, Shaanker RU. Cross-Species Amplification of Coconut Microsatellite Markers in Rattans. Silvae Genetica. 2007;56:282–6.
76. Anitha N, Jayaraj KL, Kumar EPA, George J, Rajesh MK. Assessment of cross-taxa utility of coconut microsatellite markers. Indian Journal of Horticulture. 2008;65:317–21.
77. 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. Theor Appl Genet. 1998;96:545–50.
78. Boutin-Ganache I, Raposo M, Raymond M, Deschepper FC. M13-Tailed Primers Improve the Readability and Usability of Microsatellite Analyses Performed with Two Different Allele- Sizing Methods. Biotechniques. 2001;31:25–8.
79. Boutin P, Hani EH, Vasseur F, Roche C, Bailleul B, Hager J, et al. Automated fluorescence-based screening for mutation by SSCP: use of universal M13 dye primers for labeling and detection. Biotechniques. 1997;23:358–62.
80. 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. Theor Appl Genet. 2006;114:49–58.
81. 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;14:1103–13.
82. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17:10–2.
83. 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;26:403–4.
84. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics. 2012;28:2537–9.
85. Hartl D, Clark A. Principles of Population Genetics. 1997.
86. Wright S Evolution and the Genetics of Populations. Theory of gene frequencies. Chicago, USA: University of Chicago Press: University of Chicago Press; New edition edition; 1969. https://www.press.uchicago.edu/ucp/books/book/chicago/E/bo5961634.html. Accessed 19 Mar 2020.
87. Slatkin M. Isolation by distance in equilibrium and non-equilibrium populations. Evolution. 1993;:264–279.
88. Slatkin M. Gene flow and the geographic structure of natural populations. Science. 1987;:787–792.
89. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics, 155(2), 945–959. Genetics. 2000;155:945–59.
90. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005;14:2611–20.
91. Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, et al. The K = 2 conundrum. Mol Ecol. 2017;3594–602.
92. Takezaki N, Nei M, Tamura K. POPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface. Mol Biol Evol. 2010;27:747–52.

Figures

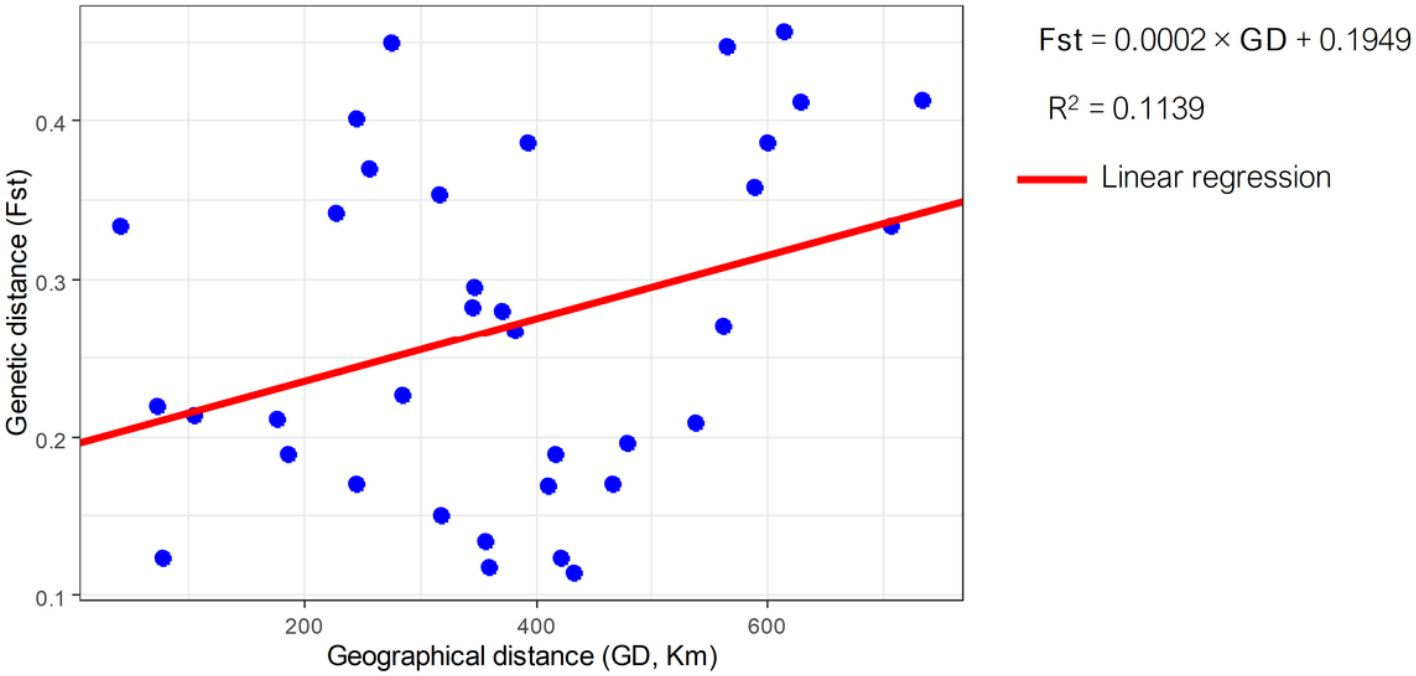

Figure 1

Correlation between pairwise Fst vs. pairwise geographical distance

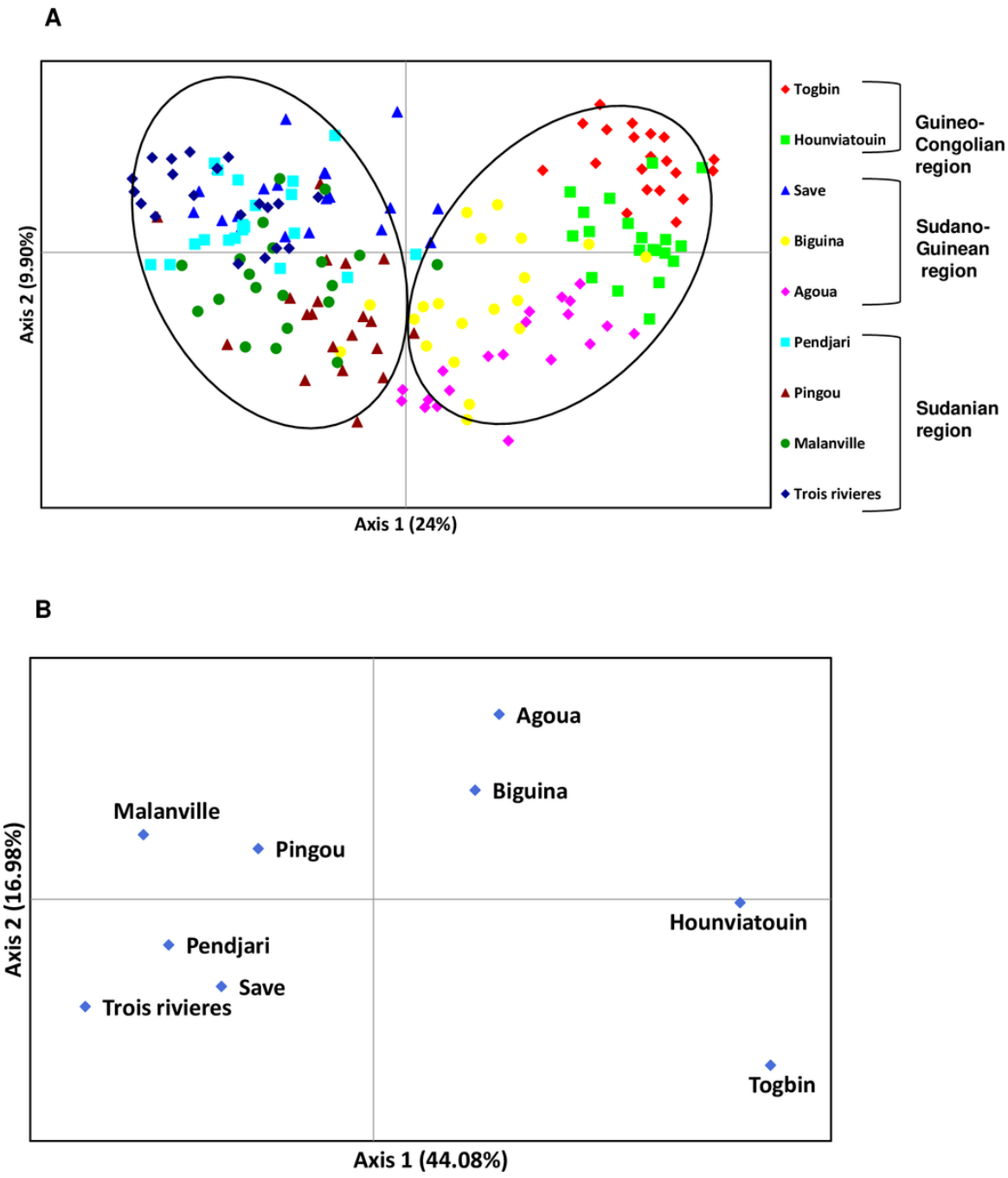

Figure 2

Principal Coordinates Analysis (PCoA) A: PCoA analysis of individual B. aethiopum samples. B: Sampling locations-based PCoA

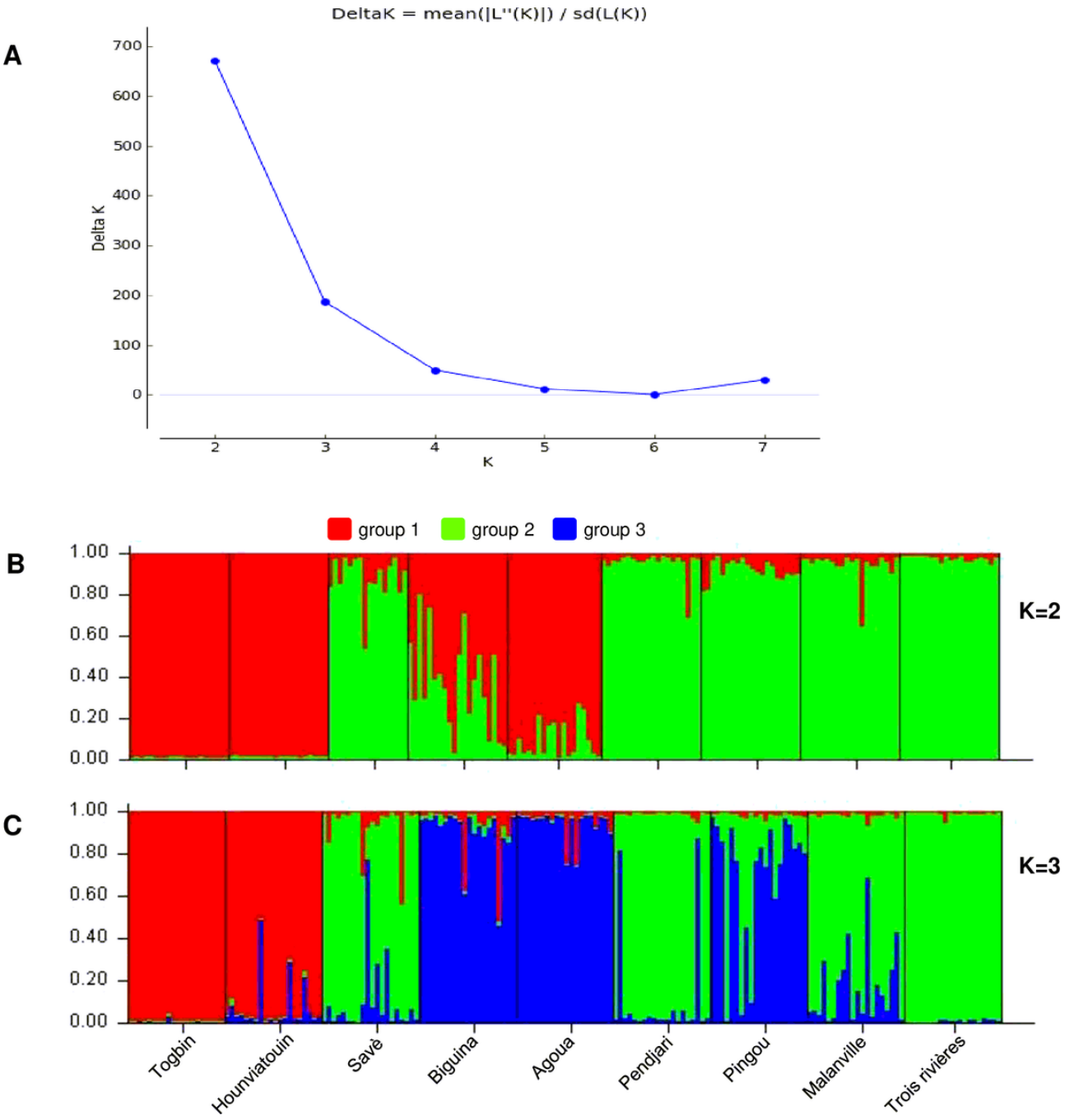

Figure 3

Bayesian cluster analysis A: Determination of the optimal value of K from Structure Harvester. Bar plot representations of Bayesian STRUCTURE analysis of Beninese B. aethiopum samples with K=2 (B) or with K =3 (C). Sampling sites are displayed along the horizontal axis.

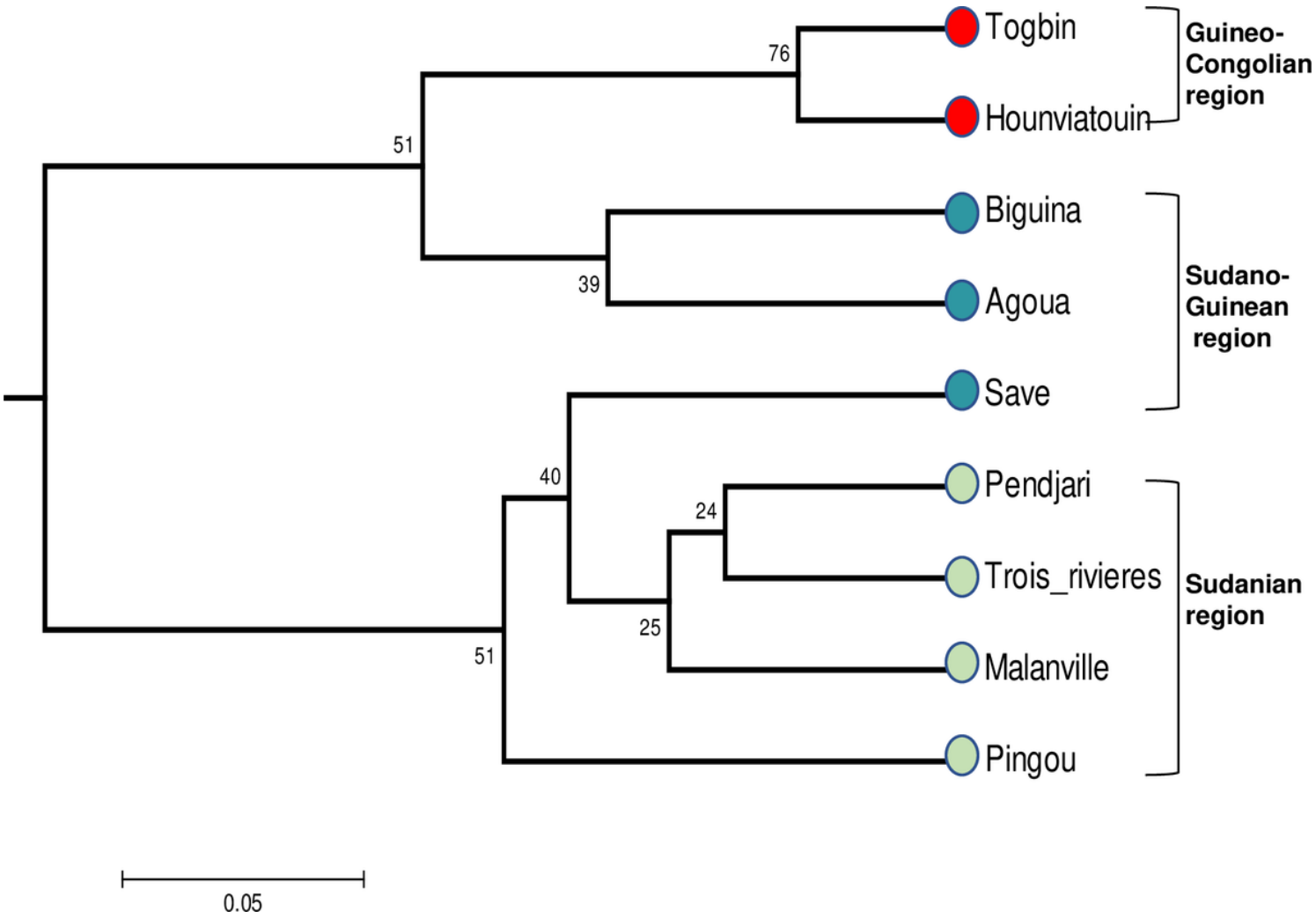

Figure 4

UPGMA dendrogram based on genetic distances between Beninese B. aethiopum sampling locations. Bootstrap values supporting each branch are indicated on the nodes.

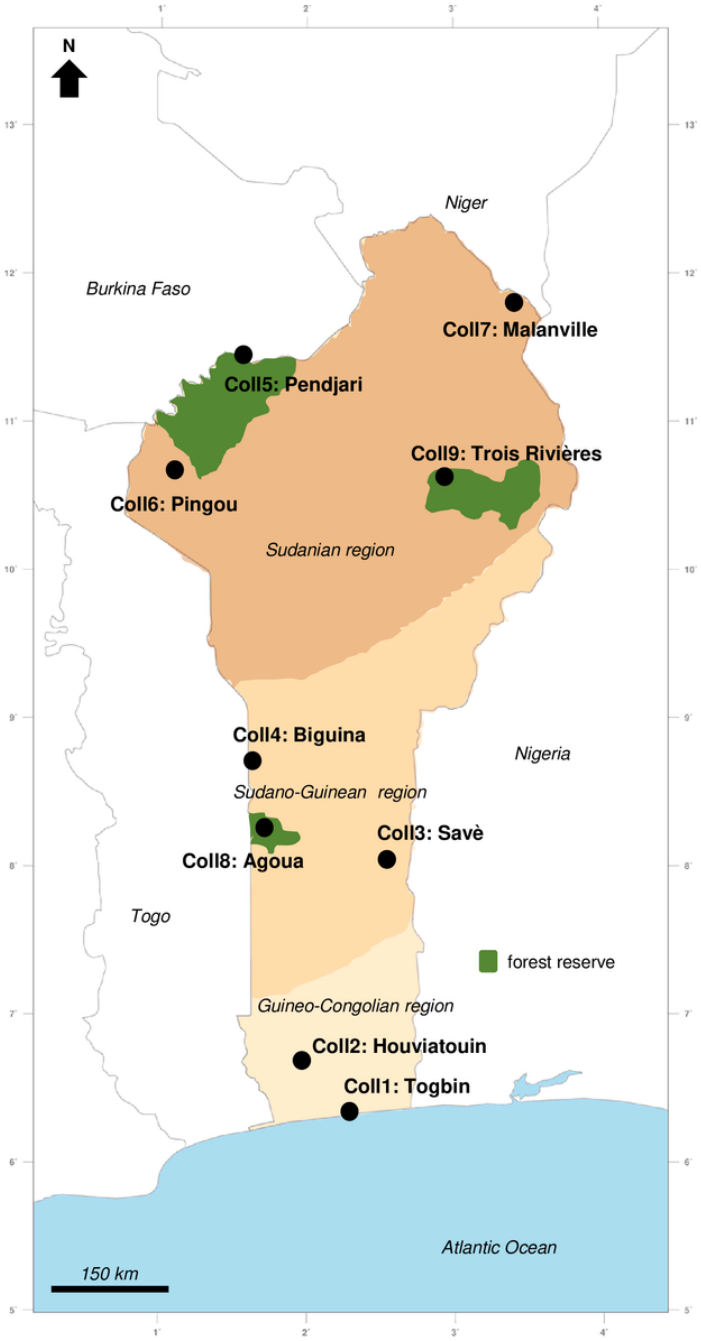

Figure 5

Sampling locations of the Beninese B. aethiopum used in this study. Information for individual samples from the nine collection sites (Coll1-9) are available in Additional file 2. 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/).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile3.txt
- Additionalfile2.docx

- Additionalfile1.xlsx