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 Kpatènon, Kolawolé Valère Salako, Sylvain Santoni, Leila Zekraoui, Muriel Latreille, Christine Tollon-Cordet, Cédric Mariac, Estelle Jaligot, Thierry Beulé, Kifouli Adéoti

To cite this version:
Mariano Joly Kpatènon, Kolawolé Valère Salako, Sylvain Santoni, Leila Zekraoui, Muriel Latreille, et al.. 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. BMC Genetics, BioMed Central, 2020, 21 (1), 10.1186/s12863-020-00955-y. hal-03105540

HAL Id: hal-03105540
https://hal.inrae.fr/hal-03105540
Submitted on 11 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
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 Kpatènon, Kolawolé Valère Salako, Sylvain Santoni, Leila Zekraoui, Muriel Latreille, Christine Tolland-Cordet, Cédric Maria, Estelle Jaligot, Thierry Beulé and Kifouli Adéoti

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.

In a second phase, we generated a *B. aethiopum*-specific genomic dataset through high-throughput sequencing and used it for the de novo detection of microsatellite loci. Among the primer pairs targeting these, 11 detected polymorphisms and were further used for analyzing genetic diversity. Across the nine sites, expected heterozygosity (He) ranges from 0.263 to 0.451 with an overall average 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. Accordingly, the low number of migrants and positive values of the fixation index (F) in sites from both the Central (Sudano-Guinean) and (Continued on next page)
the Southern (Guinean) climatic regions suggest limited gene flow between sites. The global correlation between genetic and geographic distances is weak; however, our clustering analyses indicate that *B. aethiopum* palms from Savè (Center) are genetically more similar to those from the North than to samples from 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, and suggest future applications for our molecular resources. We propose that, while prominent short-range pollen and seed dispersal in Benin explain most of our results, gene flux between the Central and Northern regions, as a result of animal and/or human migrations, might underlie the Savè discrepancy.

**Keywords:** *Borassus aethiopum*, Genetic diversity, Microsatellite, Marker transferability, High-throughput sequencing, Simple sequence repeat, Under-studied species

**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 makes 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 Arecales 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 \textit{B. aethiopum} have put a strong anthropogenic pressure on the species, thus contributing to both fragmentation of its populations and their poor natural regeneration \cite{27, 29–32}. Further fragmentation of the species’ habitat has been observed as a result of land clearing for agriculture or urban development \cite{32–34}. As illustrated through similar examples in the literature \cite{35, 36}, such phenomena may lead to restricted gene flow and ultimately, to loss of genetic diversity among \textit{B. aethiopum} populations. A sustainable management policy for \textit{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 \textit{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 \textit{Elaeis guineensis} Jacq \cite{37}, \textit{Phoenix dactylifera} L. \cite{38–40} and \textit{Cocos nucifera} L. \cite{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 Palmaeae family have been performed \cite{43–49}. In several instances \cite{44–47, 49} these tests included samples from \textit{Borassus flabellifer}, the Asian relative of \textit{B. aethiopum}. The primary objective of the present study is to generate the first set of genetic data on \textit{Borassus aethiopum}, as a first step towards improving the management of this species through a better knowledge of its diversity. In order to achieve this, we first describe attempts to use SSR markers identified in these other palm species. Then, we describe the low-coverage sequencing of the \textit{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 \textit{B. aethiopum} samples collected across the three different climatic regions of Benin, a country that was most readily accessible to us for sample collection, as an important first step towards more comprehensive studies spanning the West African sub-region.

**Results**

**Assessment of palm SSR marker transferability to \textit{Borassus aethiopum} and evaluation of their capacity for characterizing genetic diversity**

Of the 80 microsatellite markers selected from the three model palm species \textit{Elaeis guineensis}, \textit{Phoenix dactylifera} and \textit{Cocos nucifera} and tested for amplification on \textit{B. aethiopum} DNA, 18 (22.5\%) generate amplification products (Table 1). No amplification is observed using the 11 \textit{C. nucifera} markers, whereas 7 (15.9\%) and 11 (44\%) of the \textit{P. dactylifera} and \textit{E. guineensis} markers, respectively, show a successful amplification. None of the amplification products generated with \textit{P. dactylifera} primers display genetic polymorphism in our \textit{B. aethiopum} test panel. Among \textit{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.

Overall, during this phase of the study we detect polymorphism in our \textit{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 \textit{B. aethiopum}, and might therefore be used for studying genetic diversity in this species.

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

In order to enable a more precise evaluation of genetic diversity in \textit{B. aethiopum}, we developed specific \textit{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,

| Species of origin      | Number of SSR markers tested | Number of successful amplifications (% of markers) | Number of polymorphic amplicons (% of amplifications) |
|------------------------|------------------------------|----------------------------------------------------|-----------------------------------------------------|
| \textit{Cocos nucifera} | 11                           | 0 (0)                                              | 0 (0)                                               |
| \textit{Phoenix dactylifera} | 44                           | 7 (15.9)                                           | 0 (0)                                               |
| \textit{Elaeis guineensis} | 25                           | 11 (44.0)                                          | 2 (18.2)                                            |
| Total                  | 80                           | 18 (22.5)                                          | 2 (11.1)                                            |
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 1618 microsatellite loci (Additional file 1), of which 1327 (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-nucleotide repeat units, 13.06% of tri-nucleotide units, 2.39% of tetra-nucleotide 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 20.4% of primer pairs associated with successful amplification and 55.0% of those detecting polymorphic products, however nine of them yield 20.4% of primer pairs associated with successful amplification and 55.0% of those detecting polymorphic products (60.0% of amplifying couples) show no polymorphism. The remaining 20 primer pairs enable the amplification of polymorphic products, however nine of them yield 20.4% of primer pairs associated with successful amplification and 55.0% of those detecting polymorphic products (60.0% of amplifying couples) show no polymorphism.

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

The newly identified 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 (Fig. 1).

Map generated from publicly available resources of the Institut Géographique National du Bénin (IGN; https://geobenin.bj/) and the “Major Rivers of the World” dataset from the World Bank Data Catalog (https://datacatalog.worldbank.org/dataset/major-rivers-world; Creative Commons Attribution 4.0 International license), using the ArcGIS software by ESRI (www.esri.com).

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.

**Population structure of *B. aethiopum* in Benin**

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.

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.

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. 2, the positive correlation between both variables is weak, but significant ($R^2 = 0.1139, P = 0.040$).

The results of the non-hierarchical AMOVA (Table 7) show that within-site variation underlies the major part (53%) of total variance, whereas among-sites and among-regions variations explain genetic variance to a
## 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       | CCTATCTTCCATCCCGATCG TTGGCCTGAATCGCTGCAA | 90                          | complex, polymorphic  |
| MBo02      | [ATC]₇       | GGGGAAACAGAGCGTACGGACGAG TCCATTTCATCATACGGT   | 115                         | single locus, monomorphic |
| MBo03      | [AGG]₇       | CTCGAGCCTAGCAGCTT TCTGGATGAGGAACCTCAC     | 131                         | single locus, monomorphic |
| MBo04      | [ACC]₇       | GATGGGCCGCTTGCTGTCT ACATGCTGCGAAGAGCTT | 192                         | single locus, monomorphic |
| MBo05      | [AAG]₇       | GTCCATTAGCCGCTCGGATTA TGGGTGCAAATGACCCCTT | 202                         | single locus, monomorphic |
| MBo06      | [ATC]₇       | TGCCATTCAACTGCTTAC GAACTAGCACCAGCAAACCC  | 202                         | single locus, monomorphic |
| MBo07      | [AAG]₇       | GCCACCTGGAGTCCACATCAA TCCCTCTGACTGCGATCCTC | 239                         | single locus, monomorphic |
| MBo08      | [AGG]₈       | TGATGTGTTTCCTTCCCTCT TTAATGACCGAGAGGAGGCC | 90                          | single locus, monomorphic |
| MBo09      | [AGG]₈       | TCCCTACCTCCTCATTCCTC ACTCCACTCTCTCTCATAACA | 163                         | single locus, monomorphic |
| MBo10      | [AAC]₈       | GTATAAGACCGAGGGCTGGA CCCACTTAGTGAAGATAGACTG | 166                         | single locus, monomorphic |
| MBo11      | [ATC]₈       | GCATACATGGTCCGAGCT GCTCAACATCGAGACGTTGA | 219                         | single locus, monomorphic |
| MBo12      | [ATC]₉       | GGAGGAAGGGTTGCTCAGAA TCTCAACGTGCTATGGCAGA | 102                         | single locus, monomorphic |
| MBo13      | [AAG]₉       | CAGTGTTGACCCCGAGCCATT GGAGCCCTATGCAACCCAGAG | 103                         | complex, polymorphic  |
| MBo14      | [AAC]₁₀      | ATGGCCEGACTCCCACTTAGT GAGAGAAAGGCTAAATTTATGCA | 117                         | single locus, monomorphic |
| MBo15      | [AAG]₁₀      | GCTGGAAGAGGATGGAAGAGGAGGC TCATCACTCTCCCTCCCT | 92                          | complex, monomorphic  |
| MBo16      | [AGG]₁₀      | CAGGACTGTCCTCACAGC CTCGACATAGTTGTGAGA | 118                         | single locus, monomorphic |
| MBo17      | [ATC]₁₀      | ACACAATGACCTTTGCCTGGA CCAACAGGACCTTATGCCA | 124                         | single locus, monomorphic |
| MBo18      | [AAG]₁₀      | ACATCCCTCCCTCATCTCCTT GTTCCTAATGCTGGCGC | 187                         | complex, polymorphic  |
| MBo19      | [AAG]₁₀      | TGCTATACCCAAATATCTAGGCT ACAGTCACATACTACCATCGC | 202                         | single locus, monomorphic |
| MBo20      | [AAG]₁₀      | TGCTGTTAAGGCAATTGGAAGAGA GCCGAATCTTCATCTCATAACG | 229                         | single locus, monomorphic |
| MBo21      | [AAG]₁₁      | ACAACAGAAGATGATAGTTGTTTCTTAC TGGAGGAATGCTGCTTAC | 171                         | single locus, monomorphic |
| MBo22      | [AAG]₁₄      | AGAAGAATCGGTTAGGTCACAA | 108                         | single locus, monomorphic |
Table 2  List of selected primer pairs targeting putative *B. aethiopum* microsatellite loci and assessment of their polymorphism detection ability (Continued)

| Locus name | Repeat motif | Primer sequences (5′-3′ orientation) | Expected amplicon size (bp) | Amplification product |
|------------|--------------|-------------------------------------|-----------------------------|-----------------------|
| MBo23      | [AAT]_{5}    | AGATAAACATGGTGTAAGRATTTGCT          | 100                         | single locus, monomorphic |
| MBo24      | [AAT]_{6}    | AAAGCTATGGCTGGGTGATGAA              | 90                          | single locus, monomorphic |
| MBo25      | [AAT]_{6}    | TCTTCAGTGACACCTCTCAGG              | 96                          | single locus, monomorphic |
| MBo26      | [AAT]_{7}    | CATAGGCGAGGCCCACTATA               | 134                         | single locus, monomorphic |
| MBo27      | [AAT]_{7}    | TCTCTATTTGGTTGGATCCC               | 103                         | single locus, monomorphic |
| MBo28      | [AAT]_{8}    | GCCCTGAGGTGAGAGAGGSC               | 205                         | single locus, monomorphic |
| MBo29      | [AAT]_{16}   | AGACATGTAGAGTGGGGACT               | 211                         | single locus, monomorphic |
| MBo30      | [AAT]_{8}    | TGTGACTACACTCTTCTGTCAGT            | 146                         | single locus, monomorphic |
| MBo31      | [AAT]_{10}   | TGACATGATGGACATGGGATAA             | 187                         | single locus, monomorphic |
| MBo32      | [AAT]_{10}   | TCCGAGGGCATATTTTGTCG               | 117                         | single locus, monomorphic |
| MBo33      | [AAT]_{17}   | GCACACTTTGTATCGCCAGGC              | 147                         | single locus, monomorphic |
| MBo34*     | [AG]_{28}    | GTGACCCTTCTGGCGCTTT               | 192                         | single locus, polymorphic |
| MBo35*     | [AG]_{24}    | AGCATGCTTTCTGCTCTAGT             | 137                         | single locus, polymorphic |
| MBo36      | [AG]_{23}    | TCGGAAGTCAATGTCGCCG               | 180                         | no amplification         |
| MBo37      | [AG]_{23}    | GCTCTACTCCAGAGACGGA              | 142                         | complex, polymorphic     |
| MBo38*     | [AG]_{20}    | AGTCCTCAGTGGGTGGAATGA            | 130                         | single locus, polymorphic |
| MBo39      | [AG]_{10}    | AACAGTGAGCCAAAGGCTCA             | 168                         | complex, monomorphic     |
| MBo40      | [AG]_{10}    | TGCTGGATGAGTGAAGTGGAC            | 193                         | complex, polymorphic     |
| MBo41*     | [AG]_{18}    | TCTCTCCACCCAGCTCAAC             | 184                         | single locus, polymorphic |
| MBo42      | [AG]_{18}    | CGAGTGACCTTGACCTGTCG             | 136                         | complex, polymorphic     |
| MBo43      | [AG]_{18}    | AGTTGTGGTGTGTGTGTGTCCAC          | 137                         | no amplification         |
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.

Hierarchical analyses performed with K = 2 and K = 3, respectively, yield an identical proportion of genetic variation at the within-individual level (62% of total; Table 7). Analysis using K = 3 allows for a balanced representation of variation between the among-regions and among-sites scales (16% of total variance for each), whereas among-regions variation is not as well accounted for under K = 2 (7% of total variance, vs. 24% for among-sites variation).

The Principal Coordinates Analysis (PCoA) of 180 B. aethiopum samples (Fig. 3a) 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. 3b) 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 sub-groups within the first group, corresponding to sites belonging to the Southern region and to those from the Central one, respectively.

Table 2 List of selected primer pairs targeting putative B. aethiopum microsatellite loci and assessment of their polymorphism detection ability (Continued)

| Locus name | Repeat motif | Primer sequences (5′-3′ orientation) | Expected amplicon size (bp) | Amplification product |
|------------|--------------|------------------------------------|-----------------------------|----------------------|
| MBo44      | [AG]17       | AACACACTTTAAATCGACTCTTCA           | 193                         | complex, polymorphic |
|            |              | CACGGCTGCCCATGTGAGG                |                             |                      |
| MBo45      | [AG]17       | TAGATCGGAAGTCAGGCCC                | 193                         | no amplification     |
|            |              | AGAGAAGTGGAAGGAGGCCTGTC           |                             |                      |
| MBo46      | [AG]17       | GCCGATATTAGCTTCCTCGTGCG           | 154                         | single locus, monomorphic |
|            |              | GCCCTGTTGATCCCGTTTTAC             |                             |                      |
| MBo47      | [AG]16       | GGCACCTGACGGCTCTTT                | 188                         | single locus, monomorphic |
|            |              | TCACCTGGACTCAATTGTATCCAT          |                             |                      |
| MBo48      | [AG]16       | AGGACAAGAAGATGAGAAAGCCT           | 92                          | complex, polymorphic |
|            |              | ACCAATTCCAGTTAGTTGACCA            |                             |                      |
| MBo49*     | [AG]16       | CATCACCATTCTTCTCGCTT              | 141                         | single locus, polymorphic |
|            |              | GAGAACCACATCAGGGCTGCT            |                             |                      |
| MBo50*     | [AG]15       | AGAAGTCATCTTGAGGGGCC              | 150                         | single locus, polymorphic |
|            |              | TTGCTAGGAATGATACACAAATTGCT       |                             |                      |
| MBo51*     | [AG]15       | TGTGCTATTTTGTGGGAATGCA            | 191                         | single locus, polymorphic |
|            |              | GCAAGCTCATGTTCTAGTTCAAGG        |                             |                      |
| MBo52*     | [AG]15       | ACACATCTCATGTAAGAGACTCC          | 122                         | single locus, polymorphic |
|            |              | TCTTGCTCATAGCTGCGATCTC            |                             |                      |
| MBo53      | [AG]15       | AGGTATAGGATTTGAGGGTTAGG       | 131                         | single locus, monomorphic |
|            |              | GGTGAGTAAGTTTGAGGGCTA           |                             |                      |
| MBo54*     | [AG]11NNN[AG]15 | CATATGCTGATACAAGAGAGAGGAGG  | 124                         | single locus, polymorphic |
|            |              | ACCCTTAAAGCAGGATCAGACAA         |                             |                      |
| MBo55      | [AG]15       | TGGATCAACCTTGGGTCTACA           | 198                         | complex, polymorphic |
|            |              | TCTGTGCTTCTCTAGGCACT            |                             |                      |
| MBo56*     | [AG]15       | ACACAGATCAAGCAGCAGGGA           | 103                         | single locus, polymorphic |
|            |              | AGGATACCCCTTCTTCTTCTCTCT      |                             |                      |
| MBo57*     | [AG]15       | GGGTTCATCCCTGGATGAGAGCA        | 136                         | single locus, polymorphic |
|            |              | ACCCTGCTGATCACAACTCATGT        |                             |                      |

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 [N1N2]n or [N1N2N3]n for dinucleotide and trinucleotide loci, respectively, where N1, N2, and N3 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.
Likewise, the Bayesian analysis of our data indicates an optimal value of $K = 2$ for the clustering of the samples into two groups (Fig. 4a and b): 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 Center region; and one group composed of the majority of samples collected in Savè (Eastern part of the Center region) and from the Northern locations of Pendjari, Pingou, Malanville, and Trois Rivières. Since

---

### Table 3 Characteristics of 11 newly identified 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                       |
there is a possibility that the ΔK method used for estimating K leads to over- or under-estimated values [50], clustering with higher values of K have also been tested. As is apparent in Fig. 4b, for values of K = 4 and above standard deviations increase considerably, therefore we present results for both K = 2 and K = 3 (Fig. 4c; see also Additional Figure 4 for the summary of the complete analyses with K = 1 to K = 10). As previously observed with the location-based PCoA, under K = 3 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. 5) 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 [51]. Nevertheless, this method has been used successfully for accelerating the analysis of genetic diversity in many plant species, including palms [11, 52–54]. In the present study, we find that the transferability rate of microsatellite markers developed in other palms genera to *Borassus 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 *Elaeis guineensis*, *Phoenix dactylifera* or *Cocos nucifera*, we observe that only 18 (22.5%) produce amplicons from *B. aethiopum*. This percentage is very low when compared to both the inter-species and inter-genera transferability

---

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

| Geo-climatic region | Site    | Nb of polymorphic markers | Na    | Ne    | Nb of private alleles | Allelic richness | Ho    | He    | F       |
|---------------------|---------|---------------------------|-------|-------|-----------------------|------------------|-------|-------|---------|
| Guineo-Congolian (South) | Togbin  | 8                         | 2.273 | 1.584 | 3                     | 2.08             | 0.234 ± 0.063 | 0.288 ± 0.073 | 0.145 ± 0.066 |
|                     | Hounviatouin | 9                        | 2.182 | 1.447 | 0                     | 2.1              | 0.272 ± 0.066 | 0.263 ± 0.054 | 0.007 ± 0.094 |
| Sudano-Guinean (Center) | Savè    | 10                        | 2.909 | 2.069 | 0                     | 2.72             | 0.384 ± 0.075 | 0.451 ± 0.066 | 0.134 ± 0.088 |
|                     | Biguina  | 9                         | 2.364 | 1.770 | 2                     | 2.21             | 0.345 ± 0.069 | 0.374 ± 0.070 | 0.064 ± 0.062 |
|                     | Agoua    | 10                        | 2.273 | 1.722 | 1                     | 2.16             | 0.329 ± 0.063 | 0.358 ± 0.066 | 0.059 ± 0.071 |
| Sudanian (North)     | Pendjari | 10                        | 2.818 | 1.900 | 3                     | 2.49             | 0.368 ± 0.070 | 0.396 ± 0.071 | 0.055 ± 0.064 |
|                     | Pingou   | 10                        | 2.364 | 1.906 | 1                     | 2.29             | 0.405 ± 0.072 | 0.390 ± 0.073 | 0.063 ± 0.049 |
|                     | Malanville | 8                        | 2.455 | 1.627 | 0                     | 2.2              | 0.302 ± 0.072 | 0.303 ± 0.074 | 0.020 ± 0.051 |
|                     | Trois Rivières | 10                      | 2.545 | 1.822 | 2                     | 2.43             | 0.373 ± 0.074 | 0.360 ± 0.073 | 0.055 ± 0.039 |
| 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 |

Values are provided ± standard deviation wherever applicable.

*Na* average number of different alleles, *Ne* effective number of alleles, *Ho* Observed Heterozygosity, *He* Expected Heterozygosity, *F* Fixation index.

---

**Table 5** Pairwise location matrix of Nei’s genetic distance and genetic identity values

| Site        | 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| 0.029      | –            | 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.
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* [54], 86% between the wooly jelly palm (*Butia eriospatha* Mart.) and related species *Butia catarinensis* [55] and up to 100% in the licuri palm (*Syagrus coronate* Mart.) [56]. When considering other plant families, our transferability rate is also markedly lower than both the average rate of 50% found by Peakall et al. [57] within the *Glycine* genus and among Legumes genera, and the overall rate of 35.2% calculated by Rossetto [58] 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, 59] 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

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

**Table 6** Pairwise sampling locations Fst value

![Fig. 2](image-url)  
**Fig. 2** Correlation between pairwise Fst vs. pairwise geographical distance
differences in transposable element (TE) content, which in turn might have induced structural alterations throughout the genome through indels, copy number variations and recombinations [60, 61]. 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 Oryza australiensis [60]. 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 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 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.

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 [68]. The manner of seed and pollen dispersal, which have so far not been studied extensively 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 [69, 70].

Table 7 AMOVA results

| Analysis                          | Scale       | df  | SS     | MS     | Est. var. | % Total variance | P value |
|-----------------------------------|-------------|-----|--------|--------|-----------|------------------|---------|
| Non-hierarchical AMOVA           |             |     |        |        |           |                  |         |
| Among regions                     | 2           | 2   | 154.704| 76.338 | 1.944     | 24%              | < 0.001 |
| Among sites                       | 6           | 6   | 2.220  | 0.195  | 1.831     | 7%               | < 0.001 |
| Within sites                      | 171         | 171 | 1.831  | 1.831  | 4.322     | 53%              | < 0.001 |
| Total                             | 179         | 179 | 4.322  | 1.831  | 8.169     | 100%             |         |
| Hierarchical AMOVA, K = 2         |             |     |        |        |           |                  |         |
| Among regions                     | 1           | 1   | 76.338 | 76.338 | 0.205     | 7%               |         |
| Among sites                       | 7           | 7   | 0.703  | 0.703  | 0.574     | 24%              |         |
| Among individuals                 | 171         | 171 | 1.831  | 1.831  | 4.322     | 53%              |         |
| Total                             | 359         | 359 | 2.933  | 2.933  | 8.169     | 100%             |         |

| F-statistics                      |             |     |        |        |           |                  |         |
|-----------------------------------|-------------|-----|--------|--------|-----------|------------------|---------|
| Fst                               | 0.310       | 0.010|
| Fis                               | 0.096       | 0.010|
| Fit                               | 0.376       | 0.010|

| Hierarchical AMOVA, K = 3         |             |     |        |        |           |                  |         |
| Among regions                     | 2           | 2   | 152.676| 76.338 | 0.478     | 16%              |         |
| Among sites                       | 6           | 6   | 21.085 | 21.085 | 0.472     | 16%              |         |
| Among individuals                 | 171         | 171 | 2.220  | 2.220  | 0.195     | 7%               |         |
| Within individuals                | 180         | 180 | 1.831  | 1.831  | 4.322     | 53%              |         |
| Total                             | 359         | 359 | 2.975  | 2.975  | 8.169     | 100%             |         |

| F-statistics                      |             |     |        |        |           |                  |         |
|-----------------------------------|-------------|-----|--------|--------|-----------|------------------|---------|
| Fst                               | 0.319       | 0.010|
| Fis                               | 0.096       | 0.010|
| Fit                               | 0.385       | 0.010|

df degree of freedom, SS sum of squares, MS mean squares, Est. var. estimated variance, Fst inter-sites genetic differentiation, Fis genetic differentiation of individuals within sites, Fit differentiation of individuals from the total.
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, 71]. 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) [72] and *Phyllanthus* sp. (0.607 and 0.582 for *Phyllanthus emblica* and *Phyllanthus indofischeri* respectively [73]. 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 that are lower than those reported for others palm species. Indeed for *B. flabellifer*, genetic distances ranged from 0.716 to 0.957 [74] and among natural *E. guineensis* accessions an average of 0.769 was observed [75]. 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 ([76]; 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. A part of the explanation for the genetic distance observed between the Western (i.e. Biguina and Agoua) and the Eastern (Savè) collection sites within the Central region may reside in their physical separation by the Ouémé river, which further forms a natural corridor between Savè and the sites of Trois Rivières and Malanville in the North-East (see Fig. 1) [77]. We postulate that seed dispersal by humans and/or animals along this corridor 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, members of the Bariba ethnic group, who live in the Eastern part of the country up to Malanville, share strong historical ties with the Shabè people from Savè, and exchanges between both groups are frequent [78]. The same corridor is also used annually for transhumance by the Fulani people [79], for whom *B. aethiopum* is an important plant: the role of their mobility in the dispersal of the plant, similar to what has been proposed for *Caesalpina bonduc* [80], is therefore plausible. Regarding the impact of animal migrations, 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

---

**Fig. 4** Bayesian cluster analysis. **a** Determination of the optimal value of K from Structure Harvester. **b** Evanno plot. **c** Bar plot representations of Bayesian STRUCTURE analysis of Beninese *B. aethiopum* samples with K = 2 (top) or with K = 3 (bottom) generated with CLUMPAK. Sampling sites are displayed along the horizontal axis.
elephants in their migrations. Up until 1982, the seasonal occurrence of the animal has been reported in the Wari-Maro forest of Central Benin [81].

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 be used in a palm species that belongs to the same genus and 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 [82]. 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 *Borassus 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. Using 11 newly identified SSR markers, 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. Our results support the hypothesis that pollen and seed dispersal mainly occur within sites, leading to crosses among related individuals. The exception to this general rule in the region of Savè (Center) seems to indicate long-range transfer of genes as a result to animal and/or human movements towards and from forest reserves of the North. Further research into the characteristics of these migrations and their impact on gene flow among *B. aethiopum* populations is required in order to confirm this assumption. As a complement to the present work, the acquisition of agro-morphological data is currently under way, in a bid to elucidate the reproductive development and breeding system of the species. As a longer-term perspective, we also plan 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 *Borassus aethiopum* were collected from wild populations 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.
| Marker name | Primer sequences (5'-3' orientation) | $T_a$ (°C) | Source palm species | Successful transfer to other palm species | References |
|------------|--------------------------------------|------------|---------------------|------------------------------------------|------------|
| mEgCIR0230 | CCCTGGGCCCGTTTTTC AGGCCTATATGTGATTTTAA | 57.0       | E. guineensis       | E. oleifera Syringus sp. C. nucifera P. roebellinii P. canariensis P. reclinata | [62]       |
| mEgCIR0326 | GCTAACCACAGGCAAAAAACA AAGCGCAGACTAAATACA CTAC | 59.0       |                     |                                           |            |
| mEgCIR0465 | TCCCCCACGACCAATTC GCAGAGGAGGCAGCATTC | 63.1       |                     |                                           |            |
| mEgCIR0476 | TTCTTCGCCCGCTTCCTCTC TCGGCCACCTCCACTG | 61.6       |                     |                                           |            |
| EgCSSR-5781 | TACAGCTACTAGTGGTGG TGATCCCTTCCTGGAAAAC | 59.4       | E. guineensis       | B. flabellifer                            | [49]       |
| EgCSSR-1461 | GCTCTCTCTACGCCTCCTC ATGCGATCCGAGTTCAGAAG | 60.3       |                     |                                           |            |
| mEgCIR2332 | GAGAAAGAGGCAAAGAGAGGAAG | 55.0       | E. guineensis       | B. flabellifer                            | [44, 45]   |
| mEgCIR3295 | GCTACGTCGAGACAAATCAC GAAAGGATTAACAGATAAC | 55.0       |                     |                                           |            |
| mEgCIR3311 | AATCCAATGCGCTTACAG CATGCTTTGCTTCATGCA | 55.0       |                     |                                           |            |
| mEgCIR3413 | AAAAGCTATGGGGGTAAAGAT TGGATAAGGCGAGAAGAAGA | 55.0       |                     |                                           |            |
| mEgCIR3477 | CTTCAAGCAAGATACCC GCACCAACACAGTAA | 55.0       |                     |                                           |            |
| mEgCIR3592 | GACCAAAAAACGACTCAGA CCCTGATATAGCCCTTCTC | 55.0       |                     |                                           |            |
| mEgCIR3755 | GCTCACGAAAAATGGTTAG CCCATGTTGTCAGTATAT | 55.0       |                     |                                           |            |
| mEgCIR3788 | TGATGACGCAAAAGACAGC GAGCGCAATCAGAAGC | 55.0       |                     |                                           |            |
| ESSR75    | AGATGGTTTTGAGATTTCAT TGGT AACTTGGGAGGTTGCCATTA CAAG | 60.0       | E. guineensis       | B. flabellifer                            | [44, 45, 47] |
| ESSR76    | CCATACCGCAAGAAGGATG CTAAGGATCGTATAGGGGTC TCTG | 60.0       |                     |                                           |            |
| ESSR82R   | CCCTCGACACCCCATAGT TATT | 60.0       |                     |                                           |            |
| ESSR332   | AGTTAATTGTCTCGACGGGAC AGTT CTTGGATCTCGGTTGTC | 60.0       |                     |                                           |            |
| ESSR553   | ATAATTTGTCGAGGGGA AAAC | 60.0       |                     |                                           |            |
### Table 8 Characteristics of the palm SSR markers tested for transferability to *B. aethiopum* (Continued)

| Marker name | Primer sequences (5′-3′ orientation) | T<sub>s</sub> (°C) | Source palm species | Successful transfer to other palm species | References |
|-------------|--------------------------------------|-------------------|---------------------|-------------------------------------------|------------|
| ESSR566    | AGATCCGCAGCAGTGCTTTATCCCTATG       | 60.0              |                    |                                            |            |
|             | GGTCCCTCAATACTCGGTGCCCTT           |                   |                    |                                            |            |
| ESSR609    | AGGCCGGTATAGGAAGATGAAGC             | 59.0              |                    |                                            |            |
|             | CTCCTCTCAACAGAGTGCGAGT             |                   |                    |                                            |            |
| ESSR650    | GCCCTTCTGGTTAATPGACCTG              | 59.0              |                    |                                            |            |
|             | GTTGTCTATGGATGATTGTGAGG             |                   |                    |                                            |            |
| ESSR652    | CATACGTCACACCTCAGAAACC             | 60.0              |                    |                                            |            |
| ESSR673    | GCGGTCTCTCTCAGATGTTGAG             | 59.0              |                    |                                            |            |
|             | TCAATAACCTGGCTTAACACACA             |                   |                    |                                            |            |
| ESSR681    | TCTGAATTGTCGGAGTGGCC               | 59.0              |                    |                                            |            |
|             | CATCTTGGTAACACAAAAGAG              |                   |                    |                                            |            |
| CNZ34      | CATCGATGATAATTACCCAA               | 55.0              | C. nucifera        | B. flabellifer                             | [46, 63]  |
|             | TGCATATAATGAATGCAAAAC              |                   |                    | Klaciniosa                                 |            |
| CNZ2A5     | AAGTTGAAATCTATGAAACAA              | 53.2              |                    | D. kurzianus                               |            |
|             | GGCAGTAACACATTACATAG               |                   |                    | C. simplicifolia                            |            |
|             |                                    |                   |                    | C. mannex                                  |            |
|             |                                    |                   |                    | C. thrutesii                               |            |
|             |                                    |                   |                    | C. erectus                                 |            |
|             |                                    |                   |                    | C. palustris                               |            |
| CNZ 12     | TAGCTCTCTGAGATAAGATGCT             | 54.6              | C. nucifera        | B. flabellifer                             | [46, 64]  |
|             | GATCATGAAACCTGAACAA                |                   |                    | P. dactyliera                              |            |
| CNZ 24     | TCTTAAGCTCAATACCTACCC              | 55.0              |                    | E. guineensis                              |            |
|             | CGCATTGAATGCAAAAC                 |                   |                    |                                            |            |
| CNZ 18     | ATGGTGCACTGAGCCTATTAAAAAC          | 60.3              |                    |                                            |            |
|             | GAACCTTGGACCTCCCTACCAT             |                   |                    |                                            |            |
| CNZ 42     | TGATACCCCTCTGATGCTTT              | 55.5              |                    |                                            |            |
|             | GTAGATTTGCGGAGGAGGA                |                   |                    |                                            |            |
|             | ATG                                |                   |                    |                                            |            |
| CNZ2A4     | CAGGATGGTGCAAGCCCTTTA             | 61.0              |                    |                                            |            |
|             | GGTGGAGAGGGGAGGAGGA                |                   |                    |                                            |            |
|             | TGA                                |                   |                    |                                            |            |
| CAC 21     | AATTGTGTGACACGTGAC                 | 54.1              | C. nucifera        | B. flabellifer                             | [65, 66]  |
|             | GCATAAATCTTCTTCAATAGGGA            |                   |                    |                                            |            |
| CAC 71     | ATAGCTCAAGTGTGGTGCAATGG            | 54.2              |                    |                                            |            |
|             | ATATTGTCTAGATGAGGGCC              |                   |                    |                                            |            |
| CAC 84     | TTGGTTTTTTGATGGAACCTCT             | 54.4              |                    |                                            |            |
|             | AAAGCTACACATCTCAACAGC             |                   |                    |                                            |            |
| CN1H2      | TTGATAGGAGAGCTCATCATAAC            | 53.2              | C. nucifera        | B. flabellifer                             | [65]      |
|             | ATTCCTTTTTATGGTCCGAG               |                   |                    | P. dactyliera                              |            |
| PdAG-SSR   | TCCTGTCTCGTTTCTCTTCCTAGG          | 58.0              | P. dactyliera      |                                            | [44]      |

---

Kpaténon *et al.* *BMC Genetics* (2020) 21:145 Page 16 of 23
Table 8 Characteristics of the palm SSR markers tested for transferability to *B. aethiopum* (Continued)

| Marker name | Primer sequences (5’-3’ orientation) | T<sub>s</sub> (°C) | Source palm species | Successful transfer to other palm species | References |
|-------------|--------------------------------------|------------------|---------------------|------------------------------------------|------------|
| mPdCIR015   | TTCATATTCAGTTGTCGG GTGTA             | 59.1             | P. dactylifera      | P. reclinata, P. roebelenii, P. rupicola, P. theophrasti, H. thebaica, L. carinensis, C. humilis | [67]       |
| mPdCIR063   | CTTTTATGTGTCCTGAGAGA TCTCTGATCTGGTCTG | 52.5             |                     |                                          |            |
| mPdIRD1     | AGCTGGCTCCTCCCTCTTTA TGCCTCTGAGCTGTA | 59.6             | P. dactylifera      |                                          |            |
| mPdIRD3     | CATTGATCAACACCACCAC GCCAAAACCACCCTCTGTA | 60.3             |                     |                                          |            |
| mPdIRD4     | TGGGTGCGCTTCTCAGAGT TGGGATCAAGTGGTGG | 59.8             |                     |                                          |            |
| mPdIRD5     | TCTACGAGGATGGGGTGTAGT ACCCATCTGCAAGCCTCAG | 60.2             |                     |                                          |            |
| mPdIRD7     | TCGATACTGAGGCGAGATCT CTTTGCAAGTTCCTCCACC | 60.2             |                     |                                          |            |
| mPdIRD8     | CTATTGGGCTTCTGGT GAGTTGGAGGCAAAACCACA | 59.7             |                     |                                          |            |
| mPdIRD10    | ATGGTTTGCATCTCCTTGGTAG TGAGTCTGCTGTCATCAC | 59.7             |                     |                                          |            |
| mPdIRD11    | GAGTTGGAGGCAAAACCACA CCCAAAACCCTCTCCGAGT | 59.8             |                     |                                          |            |
| mPdIRD14    | GAGGGTGCTAGGTGTCAG GCACCAAGCACAAGAGCAAT | 60.9             |                     |                                          |            |
| mPdIRD15    | CCGAGTCTGCCAGAAGTAAACC TCCTCTCTCCTCTC | 60.0             |                     |                                          |            |
| mPdIRD16    | CTGGCTCGATAGAATTCTTCGC GGACATCTCTGGTGCAT | 50.7             |                     |                                          |            |
| mPdIRD17    | GTGGGAGAAACCCCAAGAATGCTGCTGTCAT CGCTCTCTGCTGGAAGTCT | 60.2             |                     |                                          |            |
| mPdIRD20    | TTGAATGGGCTCCTCTGAGT GTCCAGCTATGATTCCAGTA | 59.5             |                     |                                          |            |
| mPdIRD22    | GGCTGTATGGGAGAAGACCTG ACTGCTGATATTCCTCTGGAAG | 59.5             |                     |                                          |            |
| mPdIRD24    | GCTCTGCAAGACCTGAAAGGAACGGATGCTGGAAGCTCAAGGG | 59.9             |                     |                                          |            |
| mPdIRD25    | CACTGGAAATTCCAGGCTGTTCC ACCAAATTTTCGCTGG | 59.9             |                     |                                          |            |
| mPdIRD26    | CTCCTGATGCTGCTCGTTCTCC GAGCGAGACCCGAGACGAC | 60.0             |                     |                                          |            |
| mPdIRD28    | GAAACGATCGGGGATGCTG TTAACGCAGCCGGTTTCTT | 59.7             |                     |                                          |            |
| mPdIRD29    | GGCTCCACCATCATATTGACA | 60.3             |                     |                                          |            |
Table 8 Characteristics of the palm SSR markers tested for transferability to *B. aethiopum* (Continued)

| Marker name | Primer sequences (5′-3′ orientation) | Ta (°C) | Source palm species | Successful transfer to other palm species | References |
|-------------|--------------------------------------|---------|---------------------|------------------------------------------|------------|
| mPdIRD30    | AACAGCATCGACTGCCTTCT               | 59.8    |                     |                                           |            |
|             | GCAGATGGTGAAAGCTCCT                |         |                     |                                           |            |
|             | CCCATTAACAGGATCAACG                 |         |                     |                                           |            |
| mPdIRD31    | GCAGGTGACTGCAAAACTCT               | 60.0    |                     |                                           |            |
|             | CTATTTGACGTGACATCCAT                |         |                     |                                           |            |
| mPdIRD32    | AAGAAGCACATCCGGCTGTG               | 59.9    |                     |                                           |            |
|             | GCGGGTGTGATTTAGTG                  |         |                     |                                           |            |
| mPdIRD33    | GGAGCATACAGTTGGTTG                 | 60.1    |                     |                                           |            |
|             | CAGCGTGGAAATGGAATAG                |         |                     |                                           |            |
| mPdIRD35    | AAGCCCTTACTCAAGCTGG                | 59.6    |                     |                                           |            |
|             | CCCATAAGCTGATTGCTG                 |         |                     |                                           |            |
| mPdIRD36    | GACACGTGACGTGATGGGA                | 60.7    |                     |                                           |            |
|             | CCATTGCTGAGGAGGAGG                 |         |                     |                                           |            |
| mPdIRD37    | TTTCTCTGCTGAAAGACACC               | 60.2    |                     |                                           |            |
|             | CTTAGCCAGCCTCCACACTC               |         |                     |                                           |            |
| mPdIRD38    | GAGAGATCGGTCAGGGAGTAG              | 59.2    |                     |                                           |            |
|             | CCAGAAATCCAAAGCAGGCAGGCAGG        |         |                     |                                           |            |
| mPdIRD39    | GAGGAAAATATTGGAGAGG                | 59.5    |                     |                                           |            |
|             | TTCACGTGAGCAGAGGATGG               |         |                     |                                           |            |
| mPdIRD40    | GACGCATTGCTACATTGGA                | 60.2    |                     |                                           |            |
|             | TAAACTGCTGCTCCTCCTG                |         |                     |                                           |            |
| mPdIRD41    | CAGATCCGGAGATGATGGA                | 60.4    |                     |                                           |            |
|             | AACGAGGACAGACGTGCTA                |         |                     |                                           |            |
| mPcCIR10    | ACCCCGACCTGAGGGTGT                | 62.8    | *P. dactylifera*     |                                           | Cherif, Castillo and Aberlenc-Bertossi, unpublished data. |
|             | CTCGCATCTCCTCCTCCTC                |         |                     |                                           |            |
| mPcCIR20    | GCACGAGAGGCTTATAGT                 | 51.7    |                     |                                           |            |
|             | CCCCTCATAGGATCCTAC                 |         |                     |                                           |            |
| mPcCIR32    | CAAATCTTCGCCTGAG                  | 53.3    |                     |                                           |            |
|             | GGTGGAGATTAATCATGT                 |         |                     |                                           |            |
| mPcCIR35    | ACAAACGGGCATGGATCTA                | 60.8    |                     |                                           |            |
|             | CCGCAGTACCTCCTCCTAT                |         |                     |                                           |            |
| mPcCIR50    | CTGCCATTTTCTGAC                   | 50.6    |                     |                                           |            |
|             | CACCATGCACAAAAATG                  |         |                     |                                           |            |
| mPcCIR57    | AAGCACGCACCTCCCTGAG                | 62.0    |                     |                                           |            |
|             | GTCTCACTCGCCACAAAATAC              |         |                     |                                           |            |
| mPcCIR85    | GAGAGGCTGGTGGTTAT                 | 51.8    |                     |                                           |            |
|             | TCTATCCAGAACACAGTA                 |         |                     |                                           |            |
| mPdIRD41    | ATCTTCCATGCAGCCCTAAG              | 60.3    |                     |                                           |            |
(Fig. 1). According to White [83], Benin covers three contrasted climatic regions which are the Sudanian region in the North, the Sudan-Guinean region in the Center 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 Che-magic 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) assessed fluorometrically with the GENios Plus reader purification workstation. Final DNA concentration was flexing 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.

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

| Marker name | Primer sequences (5′-3′ orientation) | T<sub>a</sub> (°C) | Source palm species | Successful transfer to other palm species | References |
|-------------|--------------------------------------|-------------------|---------------------|------------------------------------------|------------|
| CAGGTGTCGCCGCTCTAAA | mPdIRD47 GTTGGCATACCTGACAGCA mCAGGCCTGGCTAGTGCCTG | 60.1 | *Hyphaene thebaica* | | |

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

T<sub>a</sub>: average annealing temperature for each primer pair

Species names are abbreviated as follows: *P. roebelini*: Phoenix roebelini; *P. canariensis*: Phoenix canariensis; *Phoenix reclinata*; *H. thebaica*: Hyphaene thebaica; *L. carinensis*: Livistona carinensis; *C. humiltis*: Chamaerops humiltis; *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. theophra*: Phoenix theophra*

Transferability of the 80 palm 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 *Phoenix dactylifera* [67]; 25 developed for *Elaeis guineensis* [44, 62]; and 11 developed for *Cocos nucifera* [65]. 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 [63] adapted to the use of fluorescent labelling [64]. 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 MgCl<sub>2</sub>, 0.4 pmol M13-tailed forward primer, 4 pmol M13 primer, (5′-CACGAGCTGTGAACACGCAC-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 min 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. [84]. The DNA was then used for the construction of an Illumina paired-end library, as described in Mariac et al. [85], 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-mismatch threshold. Adapters were eliminated using Cutadapt version 1.10 [86]. (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 PRJN576413. 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 [87] with default settings (detailed in Additional file 3).

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 visualization 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 [88]. 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 \( H_0 \) is observed heterozygosity and He is expected heterozygosity [89].

F-statistics analysis assessing genetic differentiation (Fst), genetic identity, number of migrants (Nm) [90] and non-hierarchical analysis of molecular variance (AMOVA) for estimating genetic differentiation within and among locations were performed with the same software. Allelic richness was calculated using the SPA-GeDi software version 1.5 [http://ebeulb.ac.be/ebe/SPA-GeDi.html [91]). Consecutively to K determination (see below), successive hierarchical AMOVA analyses were carried out with \( K = 2 \) and \( K = 3 \). The Mantel permutation test was used for assessing the correlation between genetic and geographic distances between sampling sites [92, 93]. 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 [94] 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 HARVEST online program version 0.6.94 (http://taylor0biology.ucla.edu/structureHarvester/) to determine the optimal value of K using the ΔK method of Evanno et al. [95] and allowing for different estimates of K in accordance with Janes et al [50]. Based on the resulting values of K, a clustering analysis of the studied sampling sites was performed and graphical output was generated using CLUMPAK’s main pipeline (http://clumpak.tau.ac.il [96]; ). In order to further assess genetic clustering, a UPGMA tree based on Fst values using 1000 bootstrap replications was constructed using the POPTREE2 software [97].

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12863-020-00955-y.

Additional file 1. List and characteristics of putative microsatellite loci identified in the genome of Borassus aethiopum through QDD analysis. Characteristics (basic motif, length) of the microsatellite loci and of the primer pairs (length, position, Tm, stability, amplicon size and sequence) designed for their targeted amplification.

Additional file 2. List of sampled Borassus aethiopum individuals. M: male or female palm, respectively. All geographic coordinates are provided as North from the Equator (latitude) and East from the Greenwich meridian (longitude), respectively.

Additional file 3. Default QDD software settings. Parameters used in the QDD detection of microsatellite loci.

Additional file 4. Results of the Bayesian cluster analysis with variable values of K. Graphical summary generated from STRUCTURE results by CLUMPAK’s main pipeline with values of K ranging from 1 to 10.

Abbreviations
AMOVA: Analysis of molecular variance; F: Fixation index; Fst: 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; Nm: Number of migrants; PCR: Principal coordinate analysis; SSR: Simple sequence repeat; UPGMA: Unweighted pair-group method with arithmetic mean.

Acknowledgements
The Authors wish to thank Emira Cherif, Karina Castillo and Frédérique Aberlerc (IRD, UMR DIADÉ, 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.

Authors’ contributions
KVS, 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, KVS performed the experiments.
JMK, KA, SS, CM, LZ and TB processed and analyzed data. All authors contributed to writing and revising the manuscript. The author(s) read and approved the final manuscript.

Funding
The work described in this article was funded through travel grants to MJK and KA under the framework of the MooSoCITIC project granted to EJ by Agropolis Foundation (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 (GenPhyF 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.

Availability of data and materials
Data generated from genome sequencing (filtered reads) were deposited into GenBank under BioProject ID PRJNAS76413. 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).

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.

Competing interests
The authors declare that they have no competing interests.

Author details
1Laboratoire de Microbiologie et de Technologie Alimentaire (LAMITA), Faculté des Sciences et Techniques, Université d’Abomey-Calavi, Cotonou, Bénin. 2Biodiversité et Écologie des Plantes (BDEP), Faculté des Sciences et Techniques, Université d’Abomey-Calavi, Cotonou, Bénin. 3DIADE, Univ Montpellier, IRD, Montpellier, France. 4Laboratoire de Biomathématiques et d’Estimations Forestières (LABEF), Faculté des Sciences Agronomiques, Université d’Abomey-Calavi, Cotonou, Bénin. 5AGAP, Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, Montpellier, France. 6CIRAD, UMR DIADE, Montpellier, France.

Received: 18 June 2020 Accepted: 17 November 2020

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 Mol Biol. 2013;83:1271–82.
2. Oliveira EJ, Padua JG, Zucchi MI, Vencovsky R, Vilela MLC. Origin, evolution and genome distribution of microsatellites. Genet Mol Biol. 2006;29:294–307.
3. Huang XQ, Coster 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 HSM, Nicholson JSNS, RJSLS L. Use of transferable Nicotiana tabacum L. microsatellite markers for investigating genetic diversity in the genus Nicotiana. Genome. 2008. https://doi.org/10.1139/G08-039.
5. Murita 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, Garcia-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;11:29.
8. Elleegro H. Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. 2000;16:551–8.
9. Kalla RK, Rai MK, Kalla S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. Euphytica. 2011;173:309–34.
10. Nadeem MA, Nawaz MA, Shahid MQ, Dogar Y, Comertpay G, Yildiz M, et al. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing, Biotechnol Biotechnol Equip. 2018;32:261–85.
11. Satya P, Paswan PK, Ghosh S, Majumdar S, Ali N. Conflamibility transferability of simple sequence repeat (SSR) markers from cotton (Gossypium hirsutum L) and jute (Corchorus olitorius L.) to twenty two Malvaceae species. 3. Biotech. 2016;665.
12. Wang C, Ji G, Zhi H, Niu Z, Chai Y, Li W, et al. Genetic diversity and population structure of chinese foxtail millet (Sesamum indicum (L.) Beauv.) landraces. G3: genes, genomes, Genetics. 2012;2:769–77.
13. Squirrel J, Hollingsworth PM, Woodhead M, Russell Lowe J, Gibby M, et al. How much effort is required to isolate novel miosatellites from plants? Mol Ecol. 2003;12:1339–48.
14. Mnejja M, Garcia-Mas J, Audergon J-M, Arús P. Development, characterization, and cross-species/genus transferability of SSR markers for rubber (Hevea brasiliensis). Plant Cell Rep. 2011;30:335–44.
15. Karaca M, Ince AG, Aydin A, Ay ST. Cross-genera transferable e-microsatellite markers for 12 genera of the Lamiaceae family: cross-genus transferable e-microsatellite markers for Lamiaceae. J Sci Food Agric. 2013;93:1869–79.
16. Raveendar S, Lee G-A, Jeon Y-A, Lee Y, Lee J-R, Cho G-T, et al. Cross- amplification of Vicia sativa subsp. sativa microsatellites across 22 other Vicia species. Molecules. 2015;20:543–50.
17. Pratap A, Gupta S, Tomar R, Malviya N, Pandey VR, et al. Cross- amplification of informative microsatellite markers from common bean and scarlet runner bean for assessment of genetic diversity in mungbean (Vigna radiata). Plant Breed. 2016;135:499–505.
18. Taheri S, Lee Abdulbolah T, Yousuf M, Hanafi M, Sahebi M, Azpi Z, et al. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Molecules. 2018;23:399.
19. Purvokol D, Carthey K, Tajuddin T, Dinanti D, Sudansom S. SSR identification and marker development for sago palm based on NGS genome data. Breed Sci. 2019;69:1–10.
20. Grubisha LC, Nelson BA, Dowle NJ, Miller SL, Klooster MR. Characterization of microsatellite markers for pinedrops, Pterospora andromedea (Ericaceae), from illumina MSeq sequencing. Appl Plant Sci. 2014:2:1400072.
21. Sun Z-Y, Ye L-J, Zhang F, Hu W, Fan D-M, Zhang Z-Y. Development of microsatellite markers for Sarcodonsosia conueto (Lindisbaliaceae) using next-generation sequencing technology. Appl Plant Sci. 2016;4:1600003.
22. Grupa M, Yu W, Arnaoeteng P, Nielsen TB, Balishev B. Ethnomedicinal survey and in vitro anti-plasmodial activity of the palm Borassus aethiopum Mart. J Ethnopharmacol. 2015;175:356–69.
23. Blach-Overgaard A, Svenning J-C, Dransfield J, Greve M, Balslev H. Cross-transferable polymorphic SSR loci in Borassus species. Plant Biotechnol Equip. 2018;32:261–85.
24. Ouinsavi C, Gbémavo C, Sokpon N. Ecological structure and fruit production of Borassus aethiopum populations. Int J Bio Chem Sci. 2017;11:1512.
25. Purwoko D, Cartealy IC, Tajuddin T, Dinarti D, Sudarmanto S. SSR identification and marker development for sago palm based on NGS genome data. Breed Sci. 2019;69:1–10.
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 non-climatic environmental factors, and spatial constraints. Ecography. 2010;33:380–95.
27. Ouinsavi C, Gbémavo C, Sokpon N. Ecological structure and fruit production of Borassus aethiopum populations. Int J Bio Chem Sci. 2017;11:1512.
28. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. Euphytica. 2011;173:309–34.
29. Adomou AC, Agbani OP, Sinsin B. Plantes. In: Neuenschwander P, Sinsin B, editors. Protection de la nature en Afrique de l’Ouest. Ibadan: International Institute of Tropical Agriculture (IITA); 2011. p. 21–46.
30. Gbeso F, Yedomonhan H, Tente B, Akoegninou A. Distribution géographique des populations de rôniers (Borassus aethiopum Mart,
31. Salako VK, Asogbadjo AE, Adomou AC, Agbangla C, Glélé Kakai RL. Lattitudinal distribution, co-occurring tree species and structural diversity of the threatened palm Borassus aethiopum (Arecaceae) in Benin, West Africa. Plant Ecol Evol. 2015;148:335–49.

32. Salako VK, Houehanou TH, Yessoufou K, Asogbadjo AE, Akoegninou A, Glélé Kakai RL. Patterns of elephant utilization of Borassus aethiopum Mart. And its stand structure in the Pendjari National Park, Benin, West Africa. Trop Ecol. 2017;58:425–37.

33. Montoya D. Habitat loss, dispersal, and the probability of extinction of tree species. Commun Integr Biol. 2008;1:146–7.

34. Ouattara A, Boussim I, Zongo I-D, Guindo S. Caractéristiques morphologiques des roniers (Borassus L.) du Burkina Faso. Etudes flor vég Burkina Faso. 2002;7:37–40.

35. Dixo M, Metzger JP, Morgante JS, Zarudni KR. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic coastal Forest. Biol Conserv. 2009;142:1560–9.

36. Jump AS, Perelsnas J. Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. Proc Natl Acad Sci U S A. 2006;103:8066–100.

37. Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, et al. Development of SSR markers for Syagrus coronata (Mart.) Becc. (Arecaceae), an iconic palm tree from the Atlantic rainforest of Brazil. Am J Bot. 2011;98:e198–200.

38. 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. https://doi.org/10.4238/gmr16029699.

39. Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, et al. Genome sequence of Phoenix dactylifera. Nat Commun. 2018;9:1–7.

40. Billotte N, Marseillac N, Risterucci A-M, Adon B, Brottier P, Baurens F-C, et al. Development, characterisation, and across-taxa utility of oil palm (Elaeis guineensis Jacq.) microsatellite markers. Front Plant Sci. 2016;7. https://doi.org/10.3389/fpls.2016.01578.

41. 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. 2017;1:251–7.

42. Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, et al. The genome draft of Phoenix dactylifera. Nature. 2013;500:335–40.

43. Bazzo BR, de Carvalho LM, Carazzolle MF, Pereira GAG, Colombo CA. Genetic evidence of multiple invasions and a small number of founders of the endangered macaúba palm (Acrocomia aculeata) using transcriptome sequencing and cross-species transferability. Plant Ecol Evol. 2015;148:335–49.

44. Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, et al. The genome draft of Phoenix dactylifera. Nat Commun. 2018;9:1–7.

45. He Z, Zeng C, Liu W, Lin Q, Wei T, Aljohi HA, et al. DRDDB: an online date palm genomic resource database. Front Plant Sci. 2017;8. https://doi.org/10.3389/fpls.2017.01889.

46. Huang Y-X, Matze AJM, Matze M. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (Cocos nucifera). PLoS One. 2013;8:e74736.

47. Xiao Y, Xu P, Fan H, Baudouin L, Xia W, Bocs S, et al. The genome draft of Phoenix nucifera. Gigascience. 2017;6. https://doi.org/10.1093/ gigascience/gox095.

48. 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;18:276.

49. Sablone T, Fiszman E, Góra K, Sawicki J, Dantas SC, et al. Development of microsatellite markers for the critically endangered plant species (Strobotila Jacq.). Theor Appl Genet. 2005;110:754–65.

50. Anitha N, Jayaraj KL, Kumar EPA, George J, Rajesh MK. Assessment of cross-taxa utility of coconut microsatellite markers in rattans. Silvae Genet. 2005;56:282–6.

51. Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C. Cross-species transferability of nuclear microsatellite markers: potential and limitations. Mol Ecol. 2017;26:3594–402.

52. Viana MVC, Miranda EA, de Francisco AK, Carvalho CAL, Waldschmidt AM. Transferability of microsatellite primers developed for single-leaf bees to four other species of the genus Melipona. Genet Mol Res. 2011;10:3942–7.

53. Nazarenos AG, Zucchi MI, dos Reis MS. Microsatellite markers for Butia enopsea (Arecaceae), a vulnerable palm species from the Atlantic rainforest of Brazil. Am J Bot. 2011;98:e198–200.

54. Viana MVC, Miranda EA, de Francisco AK, Carvalho CAL, Waldschmidt AM. Transferability of microsatellite primers developed for single-leaf bees to four other species of the genus Melipona. Genet Mol Res. 2011;10:3942–7.

55. Nazarenos AG, Zucchi MI, dos Reis MS. Microsatellite markers for Butia enopsea (Arecaceae), a vulnerable palm species from the Atlantic rainforest of Brazil. Am J Bot. 2011;98:e198–200.

56. Viana MVC, Miranda EA, de Francisco AK, Carvalho CAL, Waldschmidt AM. Transferability of microsatellite primers developed for single-leaf bees to four other species of the genus Melipona. Genet Mol Res. 2011;10:3942–7.

57. Nazarenos AG, Zucchi MI, dos Reis MS. Microsatellite markers for Butia enopsea (Arecaceae), a vulnerable palm species from the Atlantic rainforest of Brazil. Am J Bot. 2011;98:e198–200.

58. Viana MVC, Miranda EA, de Francisco AK, Carvalho CAL, Waldschmidt AM. Transferability of microsatellite primers developed for single-leaf bees to four other species of the genus Melipona. Genet Mol Res. 2011;10:3942–7.

59. Nazarenos AG, Zucchi MI, dos Reis MS. Microsatellite markers for Butia enopsea (Arecaceae), a vulnerable palm species from the Atlantic rainforest of Brazil. Am J Bot. 2011;98:e198–200.
75. Bakoumé C, Wickneswari R, Siju S, Rajanandu 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. 2014;62:349–60.

76. Michon L, Adeati K, Koffi K, Ewedje E, Stauffer FW. Notes on Borassus aethiopum Mart., a multi-purpose palm in Togo and Benin. Palms. 2018;62:57–69.

77. Moniod F. Régime hydrologique de l’Ouémé (Dahomey). Cahiers ORSTOM Série Hydrologie. 1973;10:171–83.

78. Affouda EO. Le royaume Shabé et ses rapports socio-ethniques de voisinage jusqu’à la création de la colonie du Dahomey. In: Mama VI, editor. Symposium culturel de Yaoui – Première édition 2018 - Peuplement du pays Shabé et occupation du site historique Etoo de Vacou. Bénin Yaoui; 2018. p. 20–8.

79. Houessou SO, Dossa LH, Diogo RVC, Houinato M, Buerkert A, Schlecht E. Change and continuity in traditional cattle farming systems of west African coast countries: a case study from Benin. Agric Syst. 2019;168:112–22.

80. Assogbadjo AE, Fandohan B, Kakaï RG, Kyndt T, Hardy OJ, Gheysen G, et al. Genetic evidence of the contribution of ethnic migrations to the propagation and persistence of the rare and declining scrambling shrub Gesnolphia bonduc L. Hum Ecol. 2012;40:117–28.

81. Roth HH, Douglas-Hamilton I. Distribution and status of elephants in West Africa. Mammalia. 1991;55:489–527.

82. 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.

83. White F. The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNESCO vegetation map of Africa. UNESCO. Paris Natural resources research; 1983.

84. Maric 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.

85. Mariac C, Scarcelli N, Pouzado J, Baraud 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. Mol Ecol Resour. 2014;14:1103–13.

86. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.J. 2011;17:10–2.

87. Meglécz E, Costedoat C, Dubut V, Gilles A, Malusa 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.

88. Peakall R, Smouse PE. GenAlEx 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol. 2006;6:288–95.

89. Hartl DL, Clark AG. Principles of population genetics. 3rd ed. Sunderland: Sinauer Associates Incorporated; 1997. https://www.cabdirect.org/cabdirect/abstract/19980108046.

90. Wright S. Evolution and the genetics of populations. Chicago: University of Chicago Press; 1969. https://press.uchicago.edu/ucp/books/book/chicago/E/bo5961634.html. Accessed 5 Nov 2020.

91. Hardy OJ, Vekemans X. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes. 2002;2:618–20.

92. Slatkin M. Isolation by distance in equilibrium and non-equilibrium populations. Evolution. 1993;47:264–79.

93. Slatkin M. Gene flow and the geographic structure of natural populations. Science. 1987;236:787–92.

94. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.

95. 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.

96. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 2015;15:1179–91.

97. 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.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:
• fast, convenient online submission
• thorough peer review by experienced researchers in your field
• rapid publication on acceptance
• support for research data, including large and complex data types
• gold Open Access which fosters wider collaboration and increased citations
• maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.
Learn more biomedcentral.com/submissions