BMC Plant Biology

RESEARCH Open Access

Genetic resources of African mahogany in Brazil: genomic diversity and structure of forest plantations

Júlio Cézar Tannure Faria[1,9*], Enéas Ricardo Konzen[2], Marcos Vinicius Winckler Caldeira[1], Tiago de Oliveira Godinho[3], Luciana Prata Maluf[4], Sarah Ola Moreira[5], Carolina da Silva Carvalho[6], Bárbara Simões Santos Leal[6], Cássia dos Santos Azevedo[1], Dione Richer Momolli[1], Gracielle Teodora da Costa Pinto Coelho[4], Conceição de Maria Batista de Oliveira[7] and Taís Cristina Bastos Soares[8]

Abstract

Background African mahogany species (*Khaya* sp.) have been introduced to Brazil gaining increasing economic interest over the last years, as they produce high quality wood for industrial applications. To this date, however, the knowledge available on the genetic basis of African mahogany plantations in Brazil is limited, which has driven this study to examine the extent of genetic diversity and structure of three cultivated species (*Khaya grandifoliola*, *Khaya senegalensis* and *Khaya ivorensis*) and their prospects for forest breeding.

Results In total, 115 individuals were genotyped (48 of *K. grandifoliola*, 34 of *K. senegalensis* and 33 of *K. ivorensis*) for 3,330 filtered neutral loci obtained from genotyping-by-sequencing for the three species. The number of SNPs varied from 2,951 in *K. ivorensis* to 4,754 in *K. senegalensis*. Multiloci clustering, principal component analysis, Bayesian structure and network analyses showed a clear genetic separation among the three species. Structure analysis also showed internal structure within each species, highlighting genetic subgroups that could be sampled for selecting distinct genotypes for further breeding, although the genetic distances are moderate to low.

Conclusion In our study, SNP markers efficiently assessed the genomic diversity of African mahogany forest plantations in Brazil. Our genetic data clearly separated the three *Khaya* species. Moreover, pairwise estimates of genetic distances among individuals within each species showed considerable genetic divergence among individuals. By genotyping 115 pre-selected individuals with desirable growth traits, allowed us not only to recommend superior genotypes but also to identify genetically distinct individuals for use in breeding crosses.

Keywords Genotyping-by-sequencing, Population genomics, SNP, Forest breeding

*Correspondence:
Júlio Cézar Tannure Faria
jc.tannure@gmail.com

Full list of author information is available at the end of the article

Background

Khaya is a genus of woody trees generally known as African mahogany, comprising economically important species that provide noble wood for a variety of uses, enabling higher profitability when compared to traditional forest species already in the market. In general, African mahoganies reach large dimensions in diameter and height, with straight and cylindrical trunks and with no branches, desirable traits for timber use. In addition, they show excellent growth and management traits in both pure and intercropped plantations [1, 2]. According to the International Tropical Timber Organization, from 2009 to 2022 there was an increase of 108.24% in the price of African mahogany wood (air-dried) in the international market, rising from €595.00 to €1239.00 per m^3 [3].

Currently, the world market for African mahogany wood is concentrated in native African forests [3], but Brazilian forest plantations have been gaining ground, particularly due to their faster production cycle, taking around 20 years to obtain sawn wood [4–6]. The cultivation of African mahogany in Brazil began in the 1970s, when Italo Claudio Falesi, then a researcher at Embrapa Eastern Amazon, received seeds from the government of the Ivory Coast. The seeds were grown into trees within Embrapa headquarters in the northern city of Belém, state of Pará, Brazil. The trees had excellent growth performance and served as matrices for producing seeds in the 1990s. Those seeds might have been used to founding new plantations of African mahogany in other regions of Brazil, along with other materials imported from Africa at the time. Despite the latter, it is believed that the genetic basis of main plantations in Brazil are descendant of the matrices located in northern Brazil [2].

Currently, seedlings from seeds are the main means for propagating African mahogany in Brazil, where there are no officially registered cultivars or clones. It challenges the development and dissemination of genetic materials of high quality and productivity. Therefore, more studies dedicated to clonal forestry and selection of matrices of commercial interest are needed to ensure better marketability for producers and investors [7, 8]. Currently, breeding of African mahogany in Brazil requires the selection of genotypes with desirable growth traits as well as the production of cellulose, charcoal, and other potential uses [9, 10]. Genetic improvement of African mahogany also needs to address goals such as improved mechanical properties of the wood, greater pest and disease resistance and better performance under climate fluctuations [11, 12]. Therefore, as for any other species, breeding requires genetic variation to an extent that can be effectively selected for desirable genetic gains from selection.

After the introduction of PCR in the late 1980's, several molecular methods have been introduced to routine breeding programs of plants. However, no methods have qualified so well as those involving next generation sequencing technologies (NGS), which have enabled genomic analyses to unprecedent levels, with reduced costs and the delivery of big datasets for analyses [13, 14]. In summary, NGS methods implicate a few basic and consecutive steps: genomic DNA extraction, DNA fragmentation and adapter ligation, construction of genomic libraries and sequencing. In general, the data are then aligned to a reference genome, which allows the identification of genetic variants, including SNP, INDEL and structural variants. The big advances of NGS technologies have also pulsed genomic studies with non-model plants, bringing significant contributions toward breeding and genomic selection of those species [15].

Genotyping-by-sequencing, usually known as GBS [16] is an important application of NGS, which allows whole genome resequencing (WGR) or reduced representation sequencing (RRS) of genomes. It has been adapted for non-model species, enabling numerous population genetic studies [17]. Being model or non-model plants, GBS has facilitated the identification of SNP markers, the most abundant and widely used molecular markers due to their broad genomic coverage, access to neutral loci as well as those markers under selection. RRS technologies offer rapid and high-quality analysis, with low error rates and ability to SNP identification without the need for reference genomes [18–21]. GBS has enabled studies on association mapping, QTL identification, high density linkage maps, genomic selection, and germplasm characterization [22].

Genetic variation of African mahogany has been addressed by a few studies in natural populations through microsatellite and SNP markers. Out of 20 novel microsatellite loci developed for the big-leaf mahogany (*Swietenia macrophylla* King, Meliaceae), ten were transferable to the African mahogany *Khaya senegalensis* (Desr.) A. Juss [23]. Another study provided 11 microsatellites markers obtained through next-generation sequencing that were used for accessing the genetic variation of 73 accessions of *K. senegalensis* collected through the natural range of distribution of the species. The authors detected high genetic diversity from the materials of western Africa [24]. Gaoue et al. [25] also used microsatellite markers to characterize natural populations of *K. senegalensis* subjected to long-term bark and foliage extraction in Benin, finding moderate levels of genetic variation. A larger study, encompassing more species of *Khaya*, used a set of 101 single nucleotide polymorphisms (SNP) developed by Pakull et al. [26]. More than 2,000 individuals were sampled, belonging to natural populations of *K. ivorensis*, *K. anthotheca*, *K. nyasica*, *K.*

grandifoliola, *K. senegalensis*, and *K. madagascariensis*. In general, the SNP markers were able to distinguish the species studied, except for *K. nyasica* and *K. madagascariensi*, that were not clearly separated from each other [27]. Just recently, a novel study has released the genomes of two important mahoganies, *Swietenia macrophylla* (274,49 Mb) and *Khaya senegalensis* (406,50 Mb), which brings novel possibilities for addressing novel breeding endeavors of such species [28].

In Brazil, as *Khaya* materials have been introduced from a few seeds and other few events of importation, researchers believed that the genetic diversity of natural populations from Africa was not well represented [2]. However, Soares et al. [29] found considerably high genetic diversity from introduced materials of *K. grandifoliola* located in plantations in the northern state of Pará, using microsatellites originally developed for *K. senegalensis* [23].

So far, the study of Soares et al. [29] is the only that addressed the genetic diversity of introduced populations of *Khaya* in Brazil. Moreover, to date, no study employing a large set of SNP markers has been conducted in natural populations of the genus. Here we present the first large SNP dataset obtained through genotyping-by-sequencing for three species of *Khaya* (*K. ivorensis*, *K. senegalensis* and *K. grandifoliola*) from forest plantations in southeastern Brazil. We were aimed at investigating the genetic diversity and structure from sampled individuals of the three species to select genetically contrasting individuals for further breeding of superior and desirable genotypes. We hypothesized that limited but significant genetic diversity is present within species in these plantations, that could enable the selection of contrasting materials and with superior performance.

Results

Criteria of choice for samples genotyped

We conducted genotyping-by-sequencing using Illumina technology with high-quality DNA samples obtained from originally 120 individuals belonging to two experimental forest plantations located in the Reserva Natural Vale (Linhares, Espírito Santo state, Brazil) and Viveiro Origem (Felixlândia, Minas Gerais state, Brazil) (Fig. 1a). The samples belonged to three species: *K. grandifoliola* (50 individuals) (Fig. 1b and c), *K. senegalensis* (35 individuals) (Fig. 1d and e) and *K. ivorensis* (35 individuals) (Fig. 1f and g). The individuals originated from seeds from Pará states or were imported from Africa (further details shown in methods). The experimental areas have been phenotyped for several growth, trunk shape and health status variables to select desirable trees within the objectives of a breeding program for African mahogany, in Brazil, that is, we selected the individuals with the highest values or the best scores among all evaluations of the forest inventory. The following variables were evaluated for selecting those trees: diameter at breast height (*DBH*), total height (*H*), merchantable height (*Mh*),

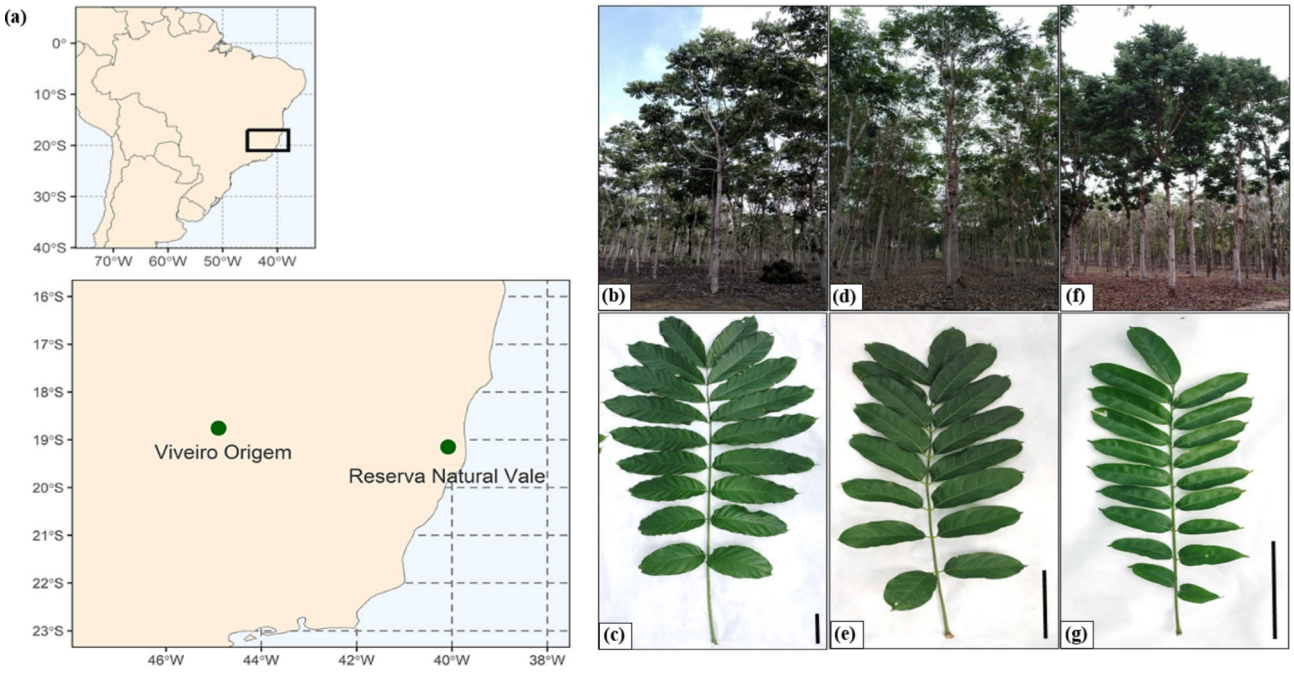

Fig. 1 **(a)** Location of experimental fields of *Khaya* spp. sampled in this study: Reserva Natural Vale (Linhares, state of Espírito Santos, Brazil) and Viveiro Origem (Felixlândia, state of Minas Gerais, Brazil); **(b)** adult individual and **(c)** individual leaf sample of *K. grandifoliola*; **(d)** adult individual and **(e)** individual leaf of *K. senegalensis*; **(f)** adult individual and **(g)** individual leaf of *K. ivorensis*. Bar = 14 cm

quality of the trunk and overall health status of each individual.

Genomic diversity of *Khaya* spp.

After sequencing, the obtained files underwent all filtering process for overall quality parameters, including missing data. In total, 533,852,208 reads were generated from all samples. After demultiplexing, 348,152,512 reads remained. The total number of reads per species were 140,923,785 for *K. grandifoliola*, 89,302,281 for *K. ivorensis*, and 117,926,446 for *K. senegalensis*. Overall, the mean sequence depth per loci was of 172.13 for all individuals. Following the filtering parameters (please check our Methods), a relatively high number of SNP markers was obtained for each species and for all species combined after data filtering (Table 1). In total, 115 individuals (48 samples of *K. grandifoliola*, 33 of *K. senegalensis* and 24 of *K. ivorensis)* and 3,330 neutral loci were retained after all filtering procedures, including pruning for linkage disequilibrium. The datasets for each species separately resulted in 3,366, 4,754, and 2,951 loci, respectively, for *K. grandifoliola*, *K. senegalensis* and *K. ivorensis* (Table 1). These datasets were used for further genetic diversity and structure analyses.

Based on the dataset with all species combined, the mean observed heterozygosity (H_O), the expected heterozygosity (H_E) and the total heterozygosity (H_T) were estimated at 0.121, 0.119, and 0.161, respectively. The coefficients of genetic differentiation among the species (species set as populations) were estimated with G_{ST} (0.259) and Wright's F_{ST} (0.258), indicating considerable genetic differentiation among the three species (Table 1).

By analyzing the three species of African mahoganies separately, the observed heterozygosity (H_O) varied from 0.192 to 0.252, while the expected heterozygosity (H_E) ranged from 0.219 to 0.271 (Table 1). *K. senegalensis* showed the lowest estimates of genetic diversity within the germplasm evaluated, while *K. ivorensis* had the highest estimates. When all species were combined, the fixation index (F) was slightly negative, while their values were positive for each species separately, indicating some degree of endogamy within each species for the germplasm that was sampled (Table 1).

Genetic structure

In order to process the genetic structure data of the species, a Bayesian analysis was conducted using statistics based on the distribution of the evaluated parameters. Therefore, using a prior set of SNP calls, we were able to attribute the most probable group to which each individual of the three species belonged. The analyses with SNP markers for 115 individuals of *Khaya*, using computations from Structure, suggested the occurrence of three genetic groups Fig. 2a) based on ΔK. Structure-based analyses showed that all samples had ancestry coefficients > 0.95, which coincides with the observation that all individuals were assigned to their groups according to the species they belonged (*K. grandifoliola*, *K. senegalensis* and *K. ivorensis*). That is, each cluster was composed exactly by all the individuals belonging to a single species. Cluster 1 was composed by 48 samples of *K. grandifoliola*, while Cluster 2 was composed by 34 individuals of *K. senegalensis*. Finally, cluster 3 encompassed 33 samples of *K. ivorensis*, independently from the location (and origin) of the materials.

A phylogenetic tree obtained through neighbor-joining inference (Fig. 2b), a principal component analysis (Fig. 2c) and a haplotype network (Fig. 2d) showed similar and complementary results to the Bayesian inference of genetic structure. The principal component analysis demonstrated a clear separation among the species based on the genomic data and the first two principal components (Fig. 2c). The first two components of the PCA explained 11.3% (PC1) and 10.5% (PC2) of the variation, while the next components explained much lower variation. A detailed examination of the phylogenetic tree and the haplotype network enabled the verification of more genetically similar or dissimilar individuals. The most genetically similar individuals within each species are A75 and A72 (*K. grandifoliola*), A111 and A102 (*K. grandifoliola*) and A22 and A24 (*K. senegalensis*) (Fig. 2b and d).

Considering that each species was assigned to a single group, we further derived pairwise estimates of F_{ST} (Fig. 3) to compare them. The F_{ST} estimate varied from 0.317 (*K. senegalensis* vs. *K. ivorensis*) to 0.346 (*K. grandifoliola* vs. *K. ivorensis*), which shows a moderate

Table 1 Number of samples, number of SNP markers, number of retained loci after filtering (RAD tags) and estimates of genetic diversity parameters *K. Grandifoliola*, *K. senegalensis* and *K. Ivorensis*

Assembly	Number of samples	Number of SNPs	Number of loci (RAD tags)	H_O	H_E	F	H_T	G_{ST}	F_{ST}
All species combined	115	7648	3330	0.121	0.119	0.017	0.161	0.259	0.258
K. grandifoliola	48	6624	3366	0.233	0.260	0.104	0.260		
K. senegalensis	33	10,131	4754	0.192	0.219	0.123	0.219		
K. ivorensis	34	5504	2951	0.252	0.271	0.070	0.271		

(H_O) Observed heterozigosity; (H_E) Expected heterozigosity; (H_T) Total heterozigosity; (F): fixation index calculated as the relative deviation between the expected heterozygosity and the observed heterozygosity; (G_{ST}) Proportion of genetic variation among speies; (F_{ST}) Wright's genetic differentiation coefficient

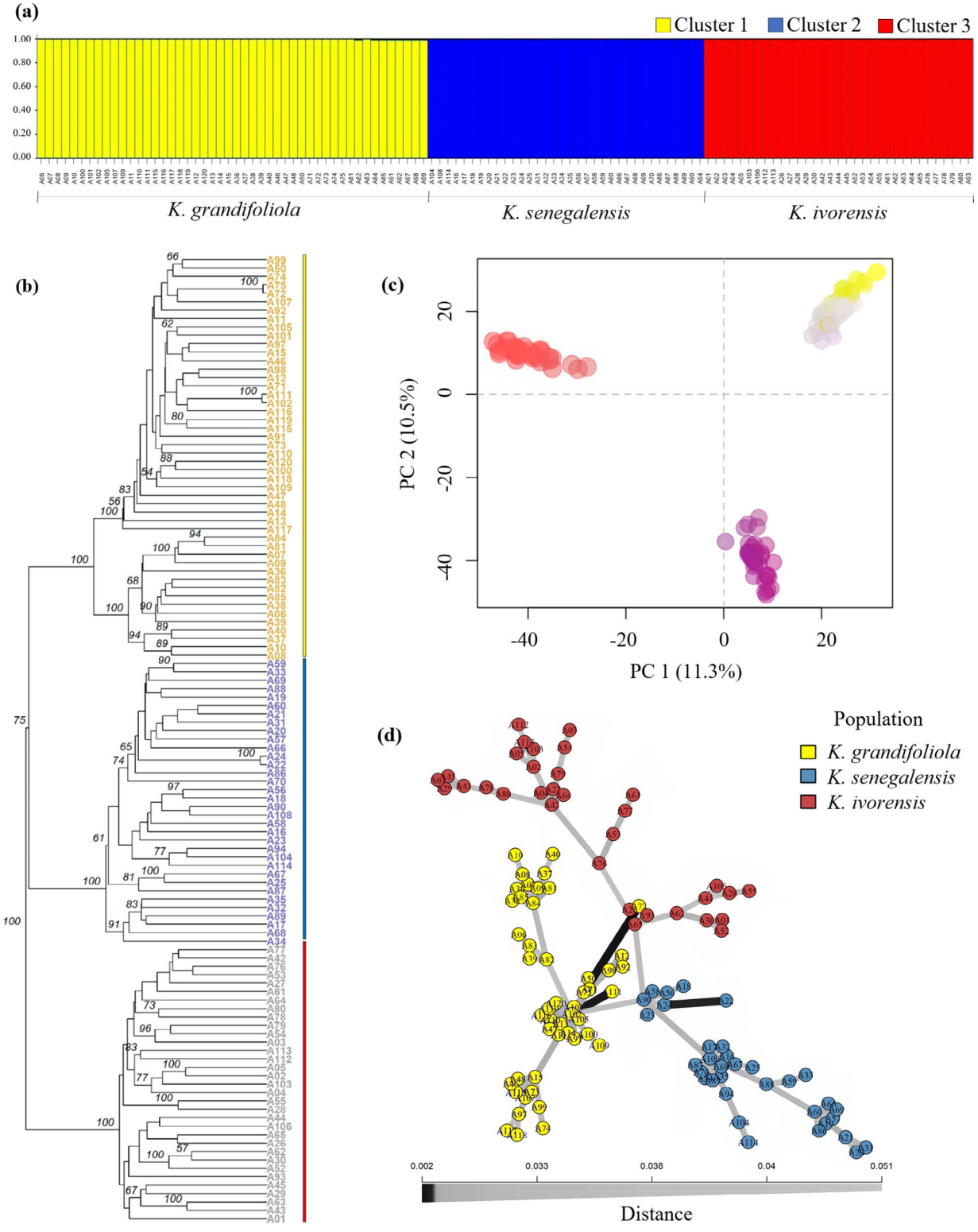

Fig. 2 Genomic structure of 115 individuals of *Khaya* spp. (*K. grandifoliola*, *K. senegalensis* and *K. ivorensis*) based on 3330 neutral SNP loci. **(a)** Genomic structure based on Bayesian analyses (K = 3). Ancestry proportions are represented in the y axis. All individuals are represented in the x axis. Individuals were assigned to groups matching exactly to the species they belonged. (**b)** Neighbor-joining phylogenetic tree among individuals and species, based on Nei's genetic distances. **(c)** Discriminant analysis of principal components (dAPC) showing the two first components of the analysis. **(d)** Median haplotype network with all samples

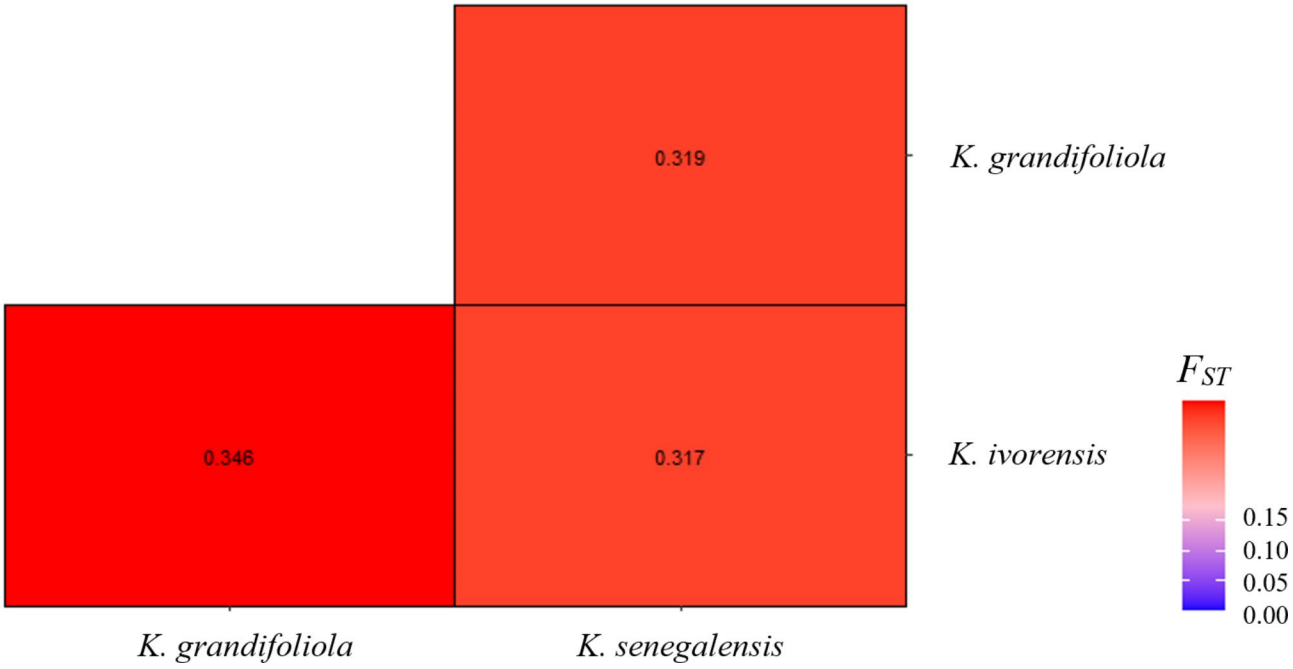

Fig. 3 Pairwise Wright's coefficients of genetic differentiation (F_{ST}) among three species of *Khaya* spp., based on 3,330 SNP markers obtained through genotyping-by-sequencing

differentiation among the three species, especially between *K. grandifoliola* and *K. ivorensis*.

Genetic structure and phenotypic variation within each species of Khaya

Structure analyses were also performed separately for each species. The Bayesian analysis revealed two main genetic groups within the 48 individuals of *K. grandifoliola*. The 34 individuals of *K. senegalensis* were divided into three genetic groups. *K. ivorensis*, with 33 individuals, was divided into two genetic groups (Fig. 4). Individuals with ancestry coefficients > 0.70 were designated as pure groups, while samples with < 0.70 were considered admixed.

K. grandifoliola was subdivided into two subgroups, with all individuals with coancestry higher than 0.70, therefore, belonging to a major group. Cluster 1 was composed by 19 individual samples from Viveiro Origem in Minas Gerais state, and 14 individuals from Reserva Natural Vale, Espírito Santo. The cluster 2 retained 15 individuals from Reserva Natural Vale (Fig. 4a). Among all samples, the individuals of cluster 2 showed more uniform phenotypes, with the highest mean diameter at breast height and tree heights (individuals A85, A37, and A38 at Reserva Natural Vale, ES). Nonetheless, three individuals from cluster 1 (A100, A109, and A120, at Viveiro Origem identified by M12, M7, and M4, respectively) had excellent values for growth variables, therefore, being potential matrices for further endeavors at clonal propagation and future cultivation (Table 2).

K. senegalensis individuals were divided into three subgroups. While cluster 1 encompassed three individuals form Viveiro Origem and eight from Reserva Natural Vale, clusters 2 and 3 showed greater admixture in the remaining individuals (Fig. 4b). Individuals A24 and A22 showed high admixture between clusters 2 and 3, as well as the individual A34 between clusters 1 and 2 (Fig. 4b). Little variation was observed for the phenotypic variables evaluated for *K. senegalensis*. In each group, a few individuals with superior phenotypes could be recommended: in cluster 1, individuals A17 and A35; in cluster 2, individuals A31 and A66; and in cluster 3, individuals A18 and A16, all from Reserva Natural Vale, in Espírito Santo. Compared to the other species, *K. senegalensis* showed the best performance for trunk quality and tree health, with cylindrical trunks, visual absence of diseases and no predation from insects (Table 2).

Finally, the individuals of *K. ivorensis* were assigned to two genetic groups. The individuals sampled from Reserva Natural Vale (ES) were assigned to both clusters, while the four individuals from Viveiro Origem were allocated to cluster 2, but with moderate admixture with cluster 1 (Fig. 4c). Three individuals from Reserva Natural Vale showed the most prominent phenotypes. In cluster 1, A63 had the highest values of diameter at breast height (26.5 cm), while in cluster 2, A61 had a DBH of 27.7 cm (Table 2).

Table 2 shows the individuals that presented the most promising phenotypic characteristics in each genetic group, among all samples. These individuals have

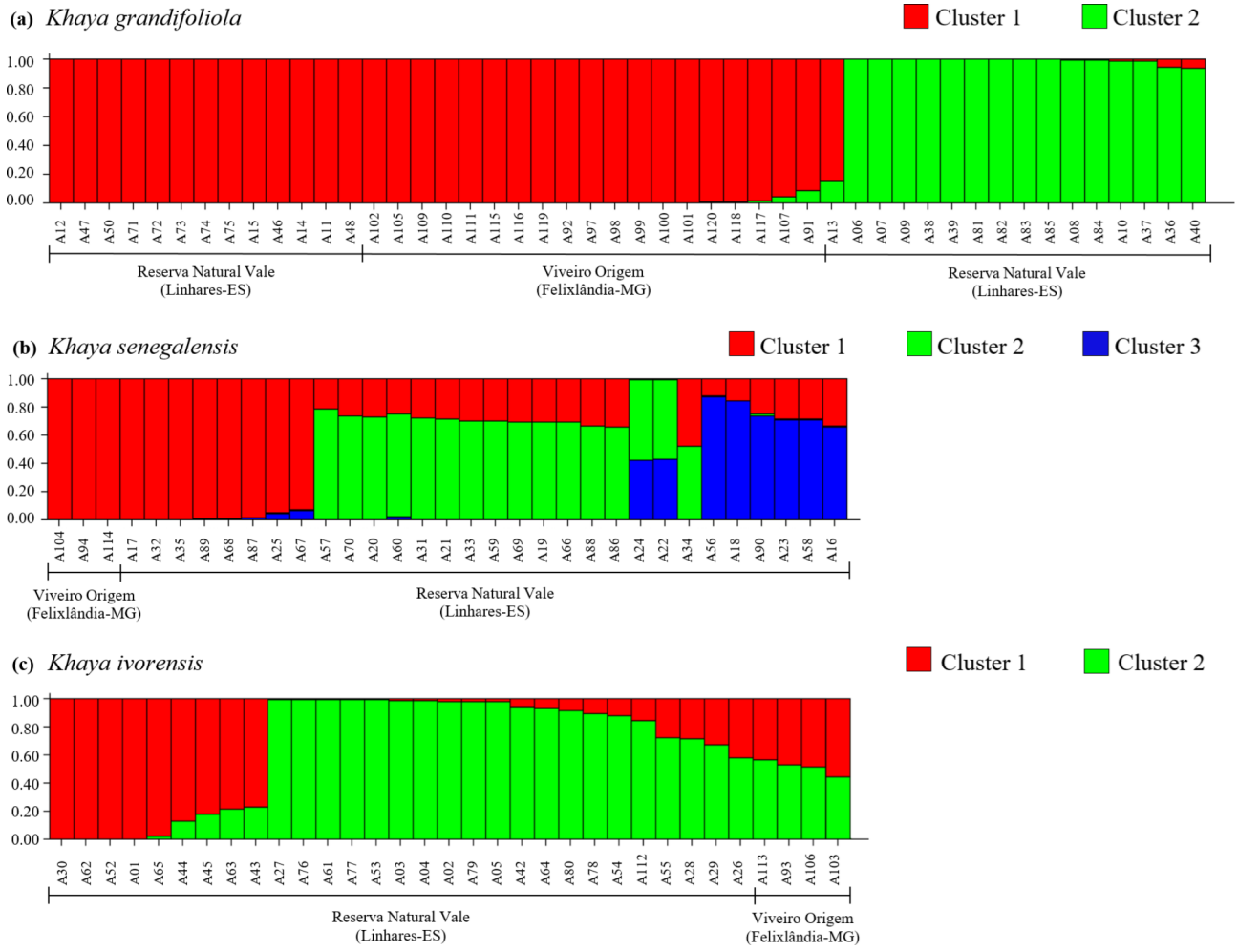

Fig. 4 Population genomic structure within each species of *Khaya* spp. **(a)** Genomic structure from Bayesian analyzes (K = 2) of the species *K. grandifoliola* based on 3366 SNPs from neutral loci; **(b)** Genomic structure from Bayesian analyses (K = 3) of the species *K. senegalensis* based on 4754 SNPs from neutral loci; **(c)** Genomic structure from Bayesian analyzes (K = 2) of the species *K. ivorensis* based on 2951 SNPs from neutral loci. The y-axis is the population membership, and the x-axis is the sample. Each bar represents an individual and each color is an inferred association within each group

desirable phenotypic traits than can be used in clonal propagation for retaining their characteristics, as well as further steps of breeding aimed at crossings between individuals with contrasting genotypes. Therefore, we also calculated the genetic distances among all pairs within each species (Fig. 5).

Discussion

This is the first study using GBS to identify SNPs and their application to infer the genetic diversity and structure of genetic resources of the three main African mahogany species introduced in Brazil. A prior study with SNP markers was published by Pakull et al. [27], that characterized natural populations of *Khaya* species. However, their study involved a set of 101 SNP using a MassARRAY®iPLEX™ genotyping, developed earlier by Pakull et al. [26]. Our study enabled the detection of ~3,000 or more SNP markers.

So far, population genetic studies on African mahoganies are scarcely available from the literature. In natural areas, most species of the genus *Khaya* are classified as vulnerable, due to the intense exploitation of their wood, and a few molecular studies were aimed at diagnosing their conservation status. In a study with *K. senegalensis* natural populations showed moderate to high levels of genetic diversity based on microsatellite markers and a genetic structure associated with the geographic distribution of populations [24]. Moderate genetic variation was also found in populations of *K. senegalensis* undergoing extractivism, also using microsatellites [25]. The recent publication of chromosome-scale genomes of *Swietenia macrophylla* and *K. senegalensis* has enabled the assembly of 274.49 Mb and 406.50 Mb, respectively, assigned to 28 pseudo-chromosomes. In total, 34,129 and 31,908 protein coding genes were predicted, respectively, for *S. macrophylla* and *K. senegalensis* [28].

Table 2 Selected individuals of *Khaya* spp. through the clusters formed and the phenotypic characteristics, diameter, breast height, total height, marketable height, stem quality and health status of the trees

Species	Genetic cluster	Individual	DBH (cm)	H (m)	TQ	TH
K. grandifoliola	1	A100	30.30	8.60	2	1
	1	A109	28.70	16.50	1	2
	1	A120	28.70	16.50	2	1
	2	A85	26.25	16.00	1	2
	2	A37	25.55	17.00	1	1
	2	A38	25.30	17.25	1	1
K. senegalensis	1	A17	23.10	12.75	1	1
	1	A35	22.60	12.50	1	1
	2	A31	23.30	12.75	1	1
	2	A66	24.50	12.25	1	1
	3	A18	24.90	12.00	2	1
	3	A16	23.20	13.00	1	1
K. ivorensis	1	A63	26.50	14.00	1	2
	1	A44	25.80	15.75	2	2
	1	A29	25.80	14.50	1	2
	2	A61	27.70	14.75	1	1
	2	A42	25.60	15.70	1	1
	2	A04	25.40	14.00	1	2

(*DBH*) Diameter at breast height; (*H*) Total height; (*TQ*) Trunk quality; (*TH*) Tree health status. Trunk quality and tree health were visually scored using the scale: 1 for excellent, 2 for regular, 3 for low. Selected individuals of *Khaya* spp. (9-years old)

The genetic resources of *Khaya* available to Brazil resulted from a few introductions only. At the headquarters of Embrapa Eastern Amazon, in Pará state, five trees were established in 1976 and their seeds have been distributed to producers. Entrepreneurs have also imported seeds for securing larger cultivation areas [30]. Despite this, significant genetic variation and structure was detected within each species from our study, revealing potential sources for selection and breeding from artificial populations of *K. grandifoliola*, *K. senegalensis* and *K. ivorensis*. Therefore, our study highlights previous findings that, despite the limited sources of genetic materials, considerable genetic variation is present in the available germplasm. And another study in Brazil also concludes that for genetic materials of *K. grandifoliola*, to that microsatellites were screened in 53 individuals in Pará and 24 individuals in Goiás, using seeds of the main provenances in Brazil, generally from Pará [29]. The authors found moderate genetic variation measured from the expected heterozygosity ($H_E = 0.56$), that was lower than the observed heterozygosity ($H_O = 0.74$), indicating that prior selection may have favored heterozygous plants and promoted heterosis [29].

Taking the results of Soares et al. [29] into further consideration, the difference between the expected and observed heterozygosity suggested that prior selection promoted outbreeding of the germplasm that they evaluated. In our case, however, some extent of inbreeding was revealed from the *F* estimates, although it was low, and the type of marker and genomic representation levels were also different. Anyway, the genetic diversity found in our study demonstrated an opportunity for selecting contrasting genotypes.

Moreover, pairwise estimates of genetic distances among individuals within each species showed considerable genetic divergence among individuals. As we genotyped 115 individuals priorly selected for desirable phenotypes for growth (diameter at breast height and total tree height), trunk quality and tree health, we could not only recommend superior genotypes but also identify individuals genetically distinct to be employed in crosses for breeding. So far, the studies applied for breeding were based on phenotypic variation only, such as a description of the phenotypic variation for growth traits in two provenances of *K. ivorensis* in Minas Gerais state, Brazil [31]. Aggregating genotypic and phenotypic data may accelerate breeding strategies for these important woody species.

Our research also becomes a pioneer in the exploration of genetic resources in two main research sites available in Brazil, which indicated limited genetic diversity, but clearly supported a phylogenetic differentiation among the three species. Moreover, the identification of population structure within each species shows the importance of collecting seeds representing the gene pool and not only maintain, but also promote recombination for amplifying the variability available. Despite the variation detected, we also recommend the importation of novel germplasm to increase the genetic diversity and promote novel crosses for genetic improvement. An interchange of genetic resources for the goal of conservation and breeding is also advised.

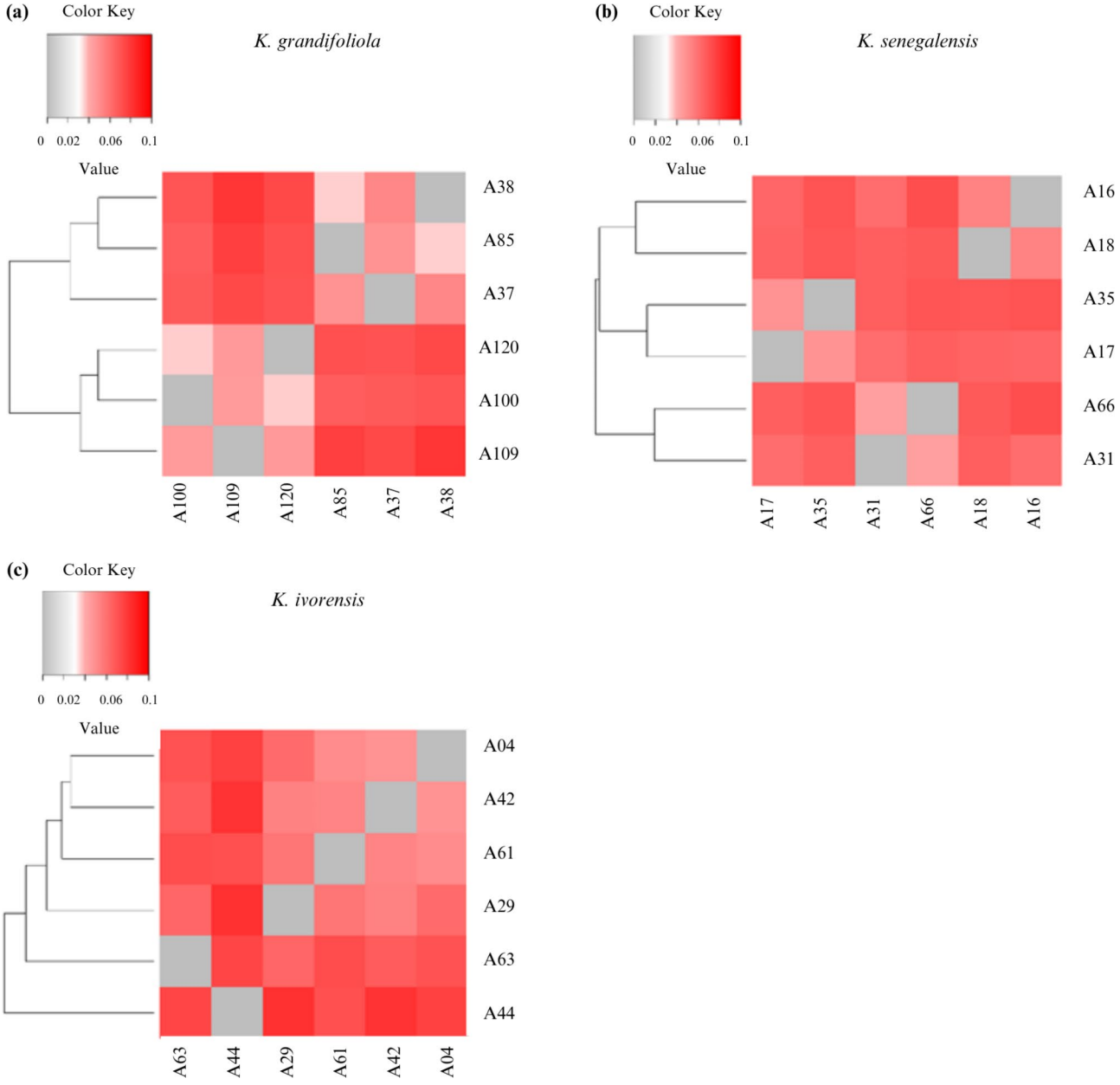

Fig. 5 Heatmap based on genetic distances calculated from SNP markers for selected individuals of *Khaya* spp., where stronger reddish tons indicate greater genetic distance between the individuals sampled. **(a)***K. grandifoliola*; **(b)***K. senegalensis*; **(c)***K. ivorensis*

Although the three species of African mahoganies here studied are morphologically similar, the GBS analyses showed genetic differences among them. Therefore, our data also suggested an important taxonomic discussion. From paired estimates of the proportion of genetic diversity among species (F_{ST}), moderate divergence was found among the three species and the phylogenetic inference resulted in each species allocated in distinct clades. In turn, *K. grandifoliola* and *K. ivorensis* were the most genetically divergent among all comparisons (Fig. 3). This is an important observation since a common misclassification between *K. grandifoliola* and *K. ivorensis* has occurred. In 2019, when the professor and researcher Dr. Ulrich Gaël Bouka Diplet from Africa visited the main plantations in Brazil, he clarified morphological differences of individuals that were wrongly classified as *K. grandifoliola*, instead of *K. ivorensis*, as they should be [32, 33]. Studies available from the literature have perpetuated this taxonomic mistake, making it necessary to carefully analyze scientific publications prior to 2019 [33]. Our genetic analyses supported the separation among the three species, which is also important for identifying genotypes that belong with a single

and specific species, enabling the selection of traits that might species-specific.

The discussion on taxonomy of *Khaya* was also provoked by a molecular and phenotypic study combined [34]. The work was conducted with SNP markers developed by Pakull et al. [26], added by four other markers, highlighting uncertainties on the taxonomical delimitation of *Khaya* species. After genotyping 498 individuals of *K. anthotheca* sampled across several countries in Africa, five consistent and distinct genetic groups were identified. In fact, the fifth group was further divided in two subgroups based on their analyses. The authors also verified that the genetic groups were consistent with morphogroups, based on morphological traits that were screened in the same individuals. Altogether, the results led the authors to infer that more species than just *K. anthotheca* were sampled [34]. In our study, the three species of *Khaya* were clearly separated by genotypic data, consistent with the morphological classification already established. Our genetic data provided a clear separation of the three species, therefore, a proper phylogenetic inference for identification is the most recommended.

The molecular data here presented alongside with phenotypic selection can be moved toward next steps of breeding for African mahoganies in Brazil. The selection of superior genotypes is needed toward establishing commercial plantations that ensure high product quality, resistance against pests and diseases, adaptation to soil and climate conditions, in addition to increasing productivity, as well as a reduction in the rotation interval [12]. Moreover, the cultivation of African mahogany species in Brazil is primarily based on seedlings of seminal origin, which has its role in maintaining genetic diversity, however, limits large-scale production of superior wood. Consequently, the application of asexual propagation techniques, together with the use of molecular markers as a tool to investigate genetic diversity, assumes vital importance in carrying out studies aimed at improving the forestry production of these species.

Thus, through the genotypic and phenotypic survey of the materials in the present study, in which three species were evaluated under the same soil and climate conditions, there is great potential for inference on the initial planning in research around genetic improvement with the selected individuals. These materials can become the main materials available in African mahogany in Brazil with potential in the global hardwood market. However, studies regarding the feasibility of implementation in the country must still be developed, with research that allows the appropriate management of the species to guarantee the desirable economic return.

Conclusions

In our study, SNPs markers were efficient in investigating the genomic diversity and structure of forest plantations of African mahogany in Brazil. Our genetic data provided a clear separation of the three species of *Khaya*, consistent with the morphological classification already established. Moreover, pairwise estimates of genetic distances among individuals within each species showed considerable genetic divergence among individuals. As we genotyped 115 individuals priorly selected for desirable phenotypes for growth, we could not only recommend superior genotypes but also identify individuals genetically distinct to be employed in crosses for breeding. Furthermore, to increase the potential for selection gains, the genetic variability of the population can be enhanced by introducing genetic material from their native environments.

Methods

Plant materials

This study was conducted from DNA samples extracted from adult individuals of *Khaya* spp. from two forest plantations located in the Reserva Natural Vale (Linhares, state of Espírito Santos, Brazil) and Origem Nursery (Felixlândia, state of Minas Gerais, Brazil) (Fig. 1a). Three species of African mahogany are cultivated in these areas: *Khaya grandifoliola*, *Khaya senegalensis* and *Khaya ivorensis*. The forest plantations resulted from seedlings acquired from distinct geographic regions (Table 3), with seed lots composed from at least 20 selected tree matrices.

The individuals selected for the genetic analyses were chosen based on desirable values of growth, shape of the

Table 3 Location, species, origin of seeds and number of adult individuals sampled for genetic diversity assessment in experimental plantations of *Khaya* spp. in southeastern Brazil

Location*	Species	Origin	Number of sampled individuals
Reserva Natural Vale (Linhares-ES)	*Khaya grandifoliola*	Pará (Brazil)	30
Reserva Natural Vale (Linhares-ES)	*Khaya senegalensis*	Senegal (Africa)	30
Reserva Natural Vale (Linhares-ES)	*Khaya ivorensis*	Espírito Santo (Brazil)	30
Viveiro Origem (Felixlândia-MG)	*Khaya grandifoliola*	Minas Gerais (Brazil)	20
Viveiro Origem (Felixlândia-MG)	*Khaya senegalensis*	Minas Gerais (Brazil)	5
Viveiro Origem (Felixlândia-MG)	*Khaya ivorensis*	Minas Gerais (Brazil)	5

* ES and MG refer, respectively, to the Brazilian states of Espírito Santo and Minas Gerais

trunk and disease resistance, in a forest plantation with over a thousand individuals. Superior individuals of the three species were selected from data of a forest inventory aimed at detailing their growth variation: diameter at breast height (*DBH*), total height (*H*), merchantable height (*Mh*), quality of the trunk and overall health status of each individual. Total and merchantable height were measured with a hypsometer, while *DBH* was obtained with a tree caliper 130 cm from the soil level. Trunk quality and tree health were visually scored using the following scale: 1 for excellent, 2 for regular, 3 for low. The trunk quality evaluation was based on the level of tortuosity of each trunk. The overall health evaluation of each individual was based on leaf appearance, presence of trunk injuries and visual detection of diseases and insect predation.

Species studied

Distinguishing the species that were accessed in this study is a quite complex task at the morphological level. *K. grandifoliola* (Fig. 1b) is a medium size tree, reaching up to 40 m in height and a diameter between 120 and 200 cm. This mahogany has rapid growth, natural pruning, straighter shaft and considerably big leaves. Overall, leaves are elliptical, varying from elliptical to oblong-elliptical and a slightly pointed apex (Fig. 1c) [1, 35].

K. senegalensis (Fig. 1d) has lower size and is adapted to dry climates, tolerating longer drought episodes. Individuals can reach between 30 and 35 m in height and diameter between 100 and 250 cm. Leaves have elliptical folioles with slightly pointed apex (Fig. 1e). Contrary to the other species, sapopemas (tabular expansions in stems) are not prominent [1, 36].

K. ivorensis (Fig. 1f) can reach up to 60 m in height and the diameter varies from 160 to 210 cm. Trees have compost leaves with three to seven pairs of folioles disposed in opposite directions. The leaflets are oblong and/or elliptical in shape with a markedly acuminate apex (Fig. 1g) [1, 32].

Selection of individuals and DNA extraction

The 120 most superior individuals, based on the previous phenotypic evaluation, were selected for DNA analyses: 50 individuals of *K. grandifoliola*, 35 individuals of *K. senegalensis* and 35 individuals of *K. ivorensis*. Leaf samples of each individual were harvested and stored in plastic zip lock bags containing silica gel for dehydration. The samples were taken to the laboratory and stored in freezer until DNA extractions were performed.

The initial steps of extraction, quality control, quantification, and lyophilization of DNA samples were carried out in the Department of Biochemistry and Molecular Biology at the Federal University of Espírito Santo, Alegre-ES. The genomic DNA was extracted using Qiagen DNeasy Plant Mini kit, using grinded tissue of each individual. The protocol for DNA extractions followed the instructions of the manufacturer. After that, the DNA samples were qualitatively evaluated using Nanodrop 2000 (ratio between 1.7 and 1.8 after measuring absorbances at 260/280 nm) as well an agarose gel 1.5%. DNA quantification was performed using Qubit fluorimeter. Samples were then liofilized and sent to Eco-Mol Consultoria e Projetos facility (Piracicaba, Sao Paulo, Brazil) for library construction and sequencing.

GBS library and sequencing

For GBS (genotyping-by-sequencing) library construction, the lyophilized samples were initially resuspended in 30 µl of ultrapure water, resulting in an approximate concentration of $10ng.\mu l^{-1}$ per sample. We employed the GBS method developed by Elshire et al. [37]. After digestion with *PstI* restriction enzyme, each DNA sample was ligated with adaptors containing indexing sequences (barcodes) that enabled their identification after sequencing. Indexes and adaptors were ligated to the ends of each restriction site using T4 DNA ligase. After adaptor ligation, all samples were pooled together and purified with magnetic beads (Agencourt AMPure XP – Beckman Coulter). Restriction fragments were then amplified and once more purified with magnetic beads. The GBS library resulted in fragments varying from 200 to 450 bp.

Library quality was evaluated with BioAnalyzer Agilent 2100 using the High Sensitivity DNA kit. No primer dimers and adaptor excess were observed (peak between 100 and 150 bp were absent), and the majority of fragments was between the expected size. Finally, the library was quantified using qPCR with KAPA Biosystems Quantification kit (Illumina), diluted to 10nM, and once more quantified with qPCR. The pooled library was sent to Centro de Genômica Funcional at ESALQ/USP (University of Sao Paulo) to be sequenced in a flowcell with NextSeq2000 Illumina using P2V2 kit (100 cyles with single read). Sequencing data were deposited to NCBI as BioProject PRJNA1136886.

De novo assembly and SNP calling

The *de novo* asembly of raw sequecing data were conducted with ipyrad [38] for all the three species together, as well as separately for each species. To do so, we set al.most all parameters according to the recommended settings of the software. The clustering threshold (parameter 14), though, was adjusted from 0.85 to 0.95; and the max indels locus (parameter 23) was adjusted from default 8 to 4. Each VCF file obtained from *de novo* assembly was then filtered using vcftools [39] to exclude all SNP with more than 30% of missing data and all individuals with more than 50% of missing data. The filtering procedures also consisted on removing non-biallelic SNP,

with depth lower than 10 ou higher than 400 and that did not fit to Hardy-Weinberg expectations ($P<0.0001$). After that, VCF files were filtered for a single SNP per locus. Finally, VCF files were also filtered to exclude outlier SNPs using the method implemented in OutFLANK package [40], that is based on a trimmed distribution of neutral F_{ST}. Only putative non-linked and neutral loci were maintained for the genetic analyses.

Genomic diversity

Estimates of genomic diversity were conducted using neutral loci only, after outlier removal. The data were used to estimate the observed heterozigosity (H_O), the expected heterozygosity under Hardy-Weinberg equilibrium (H_E), the total heterozygosity (H_T), the proportions of the genetic diversity among species (G_{ST}) and Wright's coefficient of population differentiation (F_{ST}) using 'diveRsity' [41], 'poppr' [42] and PopGenKit [43] in R (version 4.3.1, R Core Team).

Population structure, phylogenetic inference and haplotype networks

To infer the genetic structure and most probable number of genetic groups within and among the three species, we used Structure v. 2.3.4 [44], using neutral loci (RAD tags) only. Structure analyses were conducted from VCF files for each species separately, as well as the file with the SNP markers for the three species. Each analysis was conducted with 100,000 burn-in iterations, followed by 500,000 Mont Carlo Markov chain (MCMC) replicated in 10 independent simulations and with no prior information to define clusters. The K number of clusters was determined using mean likelihood values implemented with ΔK method using Structure Harvester software [45]. Ancestry coefficients of each sample were determined based on the alignment of five replicates of the best K number using CLUMPP method though the CLUMPAK software [46].

Population structure was also addressed through principal component analysis (PCA) using 'ade4' package [47] and graphically represented using 'ggplot2' [48]. To further evaluate the genetic relationships among individuals and species, Nei's pairwise genetic distances were calculated, and a Neighbor-joining tree was generated, using 2000 bootstrap replicates with package 'poppr'. We further visualized population structure with a minimum spanning haplotype network, considering genetic distance calculations as implemented in 'poppr'.

Acknowledgements

We thank: Fundação de Amparo à Pesquisa e Inovação do Espírito Santo and Conselho Nacional de Desenvolvimento Científico e Tecnológico (Fapes/CNPq N.º 11/2019 – 531/2020); Reserva Natural Vale (RNV – Linhares-ES); Instituto Ambiental Vale (IAV); Vale S.A.; Viveiro Origem (Felixlândia-MG); Universidade Federal do Espírito Santo (Ufes – Jerônimo Monteiro-ES and Alegre-ES); Centro de Estudos Costeiros, Limnológicos e Marinhos (UFRGS/CECLIMAR-RS); Instituto Tecnológico Vale (ITV – Belém-PA); Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper – Linhares-ES).

Author contributions

JCTF, ERK, MVWC and TOG designed the research, performed the experiments, and analyzed data. LPM, SOM and TCBS, analyzed data and provided critical feedback. JCTF, ERK, CSC and BSSL performed statistical analysis. JCTF, ERK, CSA, DRM, GTCPC and CMBO revised and edited the final version of the manuscript. All authors read and approved the final manuscript.

Funding

Fundação de Amparo à Pesquisa e Inovação do Espírito Santo and Conselho Nacional de Desenvolvimento Científico e Tecnológico (Fapes/CNPq N.º 11/2019 – 531/2020); Instituto Ambiental Vale (IAV); Vale S.A.

Data availability

Sequencing data were deposited to NCBI as BioProject PRJNA1136886. Any other datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All methods were performed in compliance with institutional, national, and international guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Departamento de Ciências Florestais e da Madeira, Universidade Federal do Espírito do Santo, Jerônimo Monteiro, Espírito Santo 29550-000, Brasil
[2]Departamento Interdisciplinar, Universidade Federal do Rio Grande Do Sul, Imbé, Rio Grande Do Sul 95625-000, Brasil
[3]Recursos Naturais e Áreas Protegidas, Vale, BR 101N, Linhares, Espírito Santo 29911-080, Brasil
[4]Viveiro Origem, Fazenda Riachão, Felixlândia, Minas Gerais 39237-000, Brasil
[5]Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), BR 101N, Linhares, Espírito Santo 29915-140, Brasil
[6]Instituto Tecnológico Vale - Desenvolvimento Sustentável (ITV-DS), Belém, Pará 66055-090, Brasil
[7]Departamento de Agronomia, Universidade Federal do Espírito Santo, Alegre, Espírito Santo 29500-000, Brasil
[8]Departamento de Farmácia e Nutrição, Universidade Federal do Espírito Santo, Alegre, Espírito Santo 29500-000, Brasil
[9]Departamento de Ciências Ambientais, Universidade Federal de São Carlos, Sorocaba, São Paulo 18052-780, Brasil

Received: 7 October 2023 / Accepted: 2 September 2024
Published online: 13 September 2024

References

1. Pinheiro AL, Couto L, Pinheiro DT, Brunetta JMF. Ecologia, silvicultura e tecnologia de utilização Dos Mognos Africanos (*Khaya* spp). Viçosa: Sociedade Brasileira de Agrossilvicultura; 2011.
2. Ribeiro A, Ferraz AC, Scolforo JRS. O cultivo do mogno africano (*Khaya* spp.) e o crescimento da atividade no Brasil. Florest Ambient. 2017;24. https://doi.org/10.1590/2179-8087.076814.
3. ITTO. International Tropical Timber Organization. Biennial review and assessment of the world timber situation 2019–2020. 2021. https://www.itto.int/direct/topics/topics_pdf_download/topics_id=6783&no=1. Accessed 30 Aug 2023.
4. Ribeiro A, Silva CSJ, Ferraz-Filho AC, Scolforo JRS. Financial and risk analysis of African mahogany plantations in Brazil. Cienc Agrotec. 2018;42(2):148–58. https://doi.org/10.1590/1413-70542018422026717.

5. Oliveira RS, Franca TM. Climate zoning for the cultivation of African mahogany species in Brazil. Cerne. 2020;26(3):369–80. https://doi.org/10.1590/01047760202026032748.
6. Souza CO, Arantes MDC, Pinto JÁ, Silva JGM, Carneiro MF, Lima ACB, Passos RR. Qualidade dos resíduos madeireiros de mogno-africano e eucalipto para briquetagem. Cienc Florest. 2022;32:637–52. https://doi.org/10.5902/1980509843299.
7. Silva RAN, David AMSDS, Figueiredo JC, Pereira KKG, Fogaça CA, Alves FRP, Soares LM. Germinação e vigor de sementes de mogno africano sob diferentes temperaturas. Cienc Florest. 2020;30(4):1245–54. https://doi.org/10.5902/1980509837337.
8. Teixeira G, Rodrigues GSSC. Trajetória geográfica Da silvicultura em Minas Gerais. Mercator. 2021;20:e20004. https://doi.org/10.4215/rm2021.e20004.
9. Paiva HN, Gomes JM. Propagação vegetativa de espécies florestais. 1rd ed. Viçosa: UFV; 2011.
10. Ramalho MAP, Abreu AFB, Santos JB, Nunes JAR. Aplicações Da genética Quantitativa no melhoramento de plantas autógamas. Lavras: UFLA; 2012.
11. Assis TF, Mafia RG. Hibridação E clonagem. In: BORÉM A, editor. Biotecnologia Florestal. Viçosa: Ed. Suprema; 2007. pp. 93–121. Chapter 5.
12. Xavier A, Wendling L, Silva RL. Silvicultura clonal: princípios e técnicas. 2rd ed. Viçosa, MG: Ed. UFV; 2013.
13. Zhang A, Wang H, Beyene Y, Semagn K, Liu Y, Cao S, Zhang X. Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. Front Plant Sci. 2017;8:1916. https://doi.org/10.3389/fpls.2017.01916.
14. Resende RT, Brondani C. Melhoramento De Precisão: aplicações e perspectivas na genética de plantas. 1rd ed. Brasilia, DF: Ed. Embrapa; 2023.
15. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: an overview. Hum Immunol. 2021;82(11):801–11. https://doi.org/10.1016/j.humimm.2021.02.012.
16. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Bucker ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6:e19379. https://doi.org/10.1371/journal.pone.0019379.
17. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE. 2012;7(2):e32253. https://doi.org/10.1371/journal.pone.0032253.
18. Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome. 2012;5(3). https://doi.org/10.3835/plantgenome2012.05.0005.
19. Manimekalai R, Suresh G, Govinda-Kurup H, Athiappan S, Kandalam M. Role of NGS and SNP genotyping methods in sugarcane improvement programs. Crit Rev Biotechnol. 2020;40(6):865–80. https://doi.org/10.1080/07388551.2020.1765730.
20. Younessi-Hamzekhanlu M, Gailing O. Genome-wide SNP markers accelerate perennial forest tree breeding rate for disease resistance through marker-assisted and genome-wide selection. Int J Mol Sci. 2022;23(20):12315. https://doi.org/10.3390/ijms232012315.
21. Pálsson S, Wasowicz P, Heiðmarsson S, Magnússon KP. Population structure and genetic variation of fragmented mountain birch forests in Iceland. J Heredity. 2023;114(2):165–74. https://doi.org/10.1093/jhered/esac062.
22. Rayaprolu L, Deshpande SP, Gupta R. Genotyping-by-sequencing (GBS) method for accelerating marker-assisted selection (MAS) program. In: Wani SH, Kumar A, editors. Genomics of cereal crops. New York: Humana; 2022. pp. 245–57. https://doi.org/10.1007/978-1-0716-2533-0_12.
23. Lemes MR, Esashika T, Gaoue OG. Microsatellites for mahoganies: twelve new loci for *Swietenia macrophylla* and its high transferability to *Khaya senegalensis*. Am J Bot. 2011;98:e207–9. https://doi.org/10.3732/ajb.1100074.
24. Karan M, Evans DS, Reilly D, Schulte K, Wright C, Innes D, Holton TA, Nikles DG, Dickinson GR. Rapid microsatellite marker development for African mahogany (*Khaya senegalensis*, Meliaceae) using next-generation sequencing and assessment of its intra-specific genetic diversity. Mol Ecol Resour. 2012;12(2):344–53. https://doi.org/10.1111/j.1755-0998.2011.03080.x.
25. Gaoue OG, Lemes MR, Ticktin T, Sinsin B, Eyog-Matig O. Non-timber forest product harvest does not affect the genetic diversity of a tropical tree despite negative effects on population fitness. Biotropica. 2014;46(6):756–62. https://doi.org/10.1111/btp.12145.
26. Pakull B, Mader M, Kersten B, Ekué MR, Bouka-Dipelet UG, Paulini M, Degen B. Development of nuclear, chloroplast and mitochondrial SNP markers for *Khaya* Sp. Conserv Genet Resour. 2016;8(3):283–97. https://doi.org/10.1007/s12686-016-0557-4.
27. Pakull B, Ekué MR, Bouka-Dipelet UG, Doumenge C, McKey DB, Loumeto JJ, Opuni-Frimpong E, Yorou SN, Nacoulma BMY, Guelly KA, Ramamonjisoa L, Thomas D, Guichoux E, Loo J, Degen B. Genetic diversity and differentiation among the species of African mahogany (*Khaya* spp.) based on a large SNP array. Conserv Genet. 2019;20:1035–44. https://doi.org/10.1007/s10592-019-01191-3.
28. Sahu SK, Liu M, Wang G, Chen Y, Li R, Fang D, He C. Chromosome-scale genomes of commercially important mahoganies, *Swietenia macrophylla* and *Khaya senegalensis*. Sci Data. 2023;10(1):832. https://doi.org/10.1038/s41597-023-02707-w.
29. Soares SD, Bandeira LF, Ribeiro SB, Telles MPC, Silva JA, Borges CT, Coelho ASG, Novaes E. Genetic diversity in populations of African mahogany (Khaya Grandioliola C. DC.) Introduced in Brazil. Genet Mol Biol. 2020;43(2):e20180162. https://doi.org/10.1590/1678-4685-GMB-2018-0162. 2020.
30. Falesi IC, Baena ARC. Mogno Africano (*Khaya Ivorensis* A. Chev. Em sistema silvipastoril com leguminosa e revestimento natural do solo. Belém: Embrapa Amazônia Oriental, ;; 1999. (Document 4).
31. Oliveira LGM, Dias PC, Gonçalves EJ, Soares JR, Oliveira LS. Genetic variability of two provenances of African mahogany (*Khaya Ivorensis* A. Chev) in the cerrado. Sci for. 2019;47(124):624–31. https://doi.org/10.18671/scifor.v47n124.04.
32. Reis CAF, Oliveira EB, Santos AM. Mogno-africano (*Khaya* spp.): atualidades e perspectivas do cultivo no Brasil; 2019. https://ainfo.cnptia.embrapa.br/digital/bitstream/item/202696/1/Mogno-Africano-08-10-2019.pdf. Accessed 20 Sept 2023.
33. Ferraz-Filho AC, Ribeiro A, Bouka GU, Frank-Júnior M, Terra G. African mahogany plantation highlights in Brazil. Florest Ambient. 2021;28(3):e20200081. https://doi.org/10.1590/2179-8087-FLORAM-2020-0081.
34. Bouka GU, Doumenge C, Ekué MR, Daïnou K, Florence J, Degen B, Hardy OJ. *Khaya* revisited: genetic markers and morphological analysis reveal six species in the widespread taxon *K. Anthotheca*. Taxon. 2022;71(4):814–32. https://doi.org/10.1002/tax.12720.
35. Praciak A, Pasiecznik N, Sheil D, Van Heist M, Sassen M, Correia CS, Dixon C, Fyson G, Rushford K, Teeling C, editors. The CABI encyclopedia of forest trees. Oxfordshire: CABI; 2013.
36. Opuni-Frimpong E, Tekpetey SL, Owusu AS, Obiri BD, Appiah-Kubl E, Opoku S, Nyarko-Duah NY, Essien C, Opoku EM, Storer AJ. Managing mahogany plantation in the tropics: field guide for farmers. Kumasi/Ghana: Forest Institute of Ghana; 2016. p. 95.
37. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) Approach for High Diversity species. PLoS ONE. 2011;6(5):e19379. https://doi.org/10.1371/journal.pone.0019379.
38. Eaton DA, Overcast I. Ipyrad: interactive assembly and analysis of RADseq datasets. Bioinformat. 2020;36(8):2592–4. https://doi.org/10.1093/bioinformatics/btz966.
39. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformat. 2011;27(15):2156–8. https://doi.org/10.1093/bioinformatics/btr330.
40. Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F ST. Am Nat. 2015;186(S1):S24–36. https://doi.org/10.1086/682949.
41. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA, diveRsity. An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol. 2013;4(8):782–8. https://doi.org/10.1111/2041-210X.12067.
42. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014;2:e281. https://doi.org/10.7717/peerj.281.
43. Paquette SR, Paquette MSR. Package 'PopGenKit'. 2011.
44. Pritchard JK, Wen X, Falush D, Documentationfor STRUCTURE. software. 2010. [Documentation file]. Available with the program at http://pritch.bsd.uchicago.edu/structure.html. Accessed 30 Aug 2023.
45. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359–61. https://doi.org/10.1007/s12686-011-9548-7.
46. 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(5):1179–91. https://doi.org/10.1111/1755-0998.12387.

47. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. J Stat Softw. 2007;22:1–20. https://doi.org/10.18637/jss.v022.i04.
48. Wickham H. ggplot2. Wiley Interdisciplinary Reviews: Comput Stat. 2011;3(2):180–5. https://doi.org/10.1002/wics.147.

Publisher's note