forests

MDPI

Article

Assessing Phenotypes, Genetic Diversity, and Population Structure of Shea Germplasm (*Vitellaria paradoxa* subsp. *paradoxa* C.F.Gaertn.) from Senegal and Burkina Faso

Adja Madjiguene Diallo [1,2,*], Sara Diallo [3], Robert Kariba [2], Samuel Muthemba [2], Jantor Ndalo [2], Djingdia Lompo [4,5], Tore Kiilerich Ravn [6], Mounirou Hachim Alyr [3,†] and Prasad Hendre [2]

1 Centre National de Recherches Forestières, Institut Sénégalais de Recherches Agricoles (ISRA/CNRF), Route des Pères Maristes, Dakar BP 2312, Senegal
2 AOCC Genomics Laboratory and Tree Genebank Research Unit, World Agroforestry, United Nations Avenue, Nairobi 00100, Kenya; r.kithure@cgiar.org (R.K.); s.muthemba@cifor-icraf.org (S.M.); j.ndalo@cifor-icraf.org (J.N.); p.hendre@cifor-icraf.org (P.H.)
3 Laboratoire Campus de Biotechnologies Végétales, Département de Biologie Végétale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop (UCAD), Dakar BP 5005, Senegal; sara1.diallo@ucad.edu.sn (S.D.); alyr.hachim@gmail.com (M.H.A.)
4 Centre National de Semences Forestières, Ouagadougou 01 BP 2682, Burkina Faso; d.lompo@cifor-icraf.org
5 Centre for International Forestry Research-World Agroforestry (CIFOR-ICRAF), Ouagadougou 06 BP 9478, Burkina Faso
6 Aarhus Karlshamn (AAK) Company, Slipvej 4, 8000 Aarhus, Denmark; tore.ravn@aak.com
* Correspondence: madjiguene.diallo@isra.sn or madjidiallo@yahoo.fr
† Current Address: Center for Applied Genetic and Technologies, Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA 30602, USA.

Abstract

Vitellaria paradoxa subsp. *paradoxa* C.F.Gaertn., is one of the most important components of sub-Saharan agroforestry systems, providing to rural communities, especially women, with socio- economic, environmental, and nutritional benefits. Despite its importance, the species is threatened and remains semi-domesticated. To better preserve and improve this resource, the genetic diversity and structure of 88 mother trees originated from Senegal and Burkina Faso were studied by analysing 17 phenotypic traits and 3196 SNP markers. The results revealed similar level of observed heterozygosity (Ho) between the Senegalese and Burkinabe populations (Ho = 0.16), whereas the average number of alleles per population (Na) and the expected heterozygosity (He) ranged from 0.33 to 0.34 and 0.38 to 0.39, respectively, indicating moderate to low genetic diversity. Furthermore, the polymorphic information content ranged from 0.15 for Senegal to 0.25 for Burkina Faso. Both ADMIXTURE and cluster analysis delineated our collection into two groups depending on the origin. The AMOVA showed that the highest fraction of variation was within individual, indicating a very low genetic differentiation (Fst = 0.0006) between population. At the phenotypic level, the G2 cluster representing the Senegalese genepool recorded the highest performance in terms of nut and kernel attributes, cariten and unsaponifiable matters contents, while higher crude fat, Diglyceride, Triglyceride, and Triacylglycerol Mono Stearoyl Olein Stearin contents were observed in the Burkina Faso collection (G1). The present findings on the species' genetic diversity and genetic structure constitute a good start to strengthen the species tree improvement and conservation programs.

Keywords: *Vitellaria paradoxa* subsp. *paradoxa*; single nucleotide polymorphisms (SNPs); population structure; genetic diversity; breeding

Check for updates

Academic Editor: John E. Carlson

Received: 23 October 2025
Revised: 25 December 2025
Accepted: 27 December 2025
Published: 31 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Vitellaria paradoxa subsp. *paradoxa* C.F.Gaertn., is one of the well-known multipurpose trees and represents an important component of the dryland parklands in the Sahel region of Africa.

Shea trees are deciduous, reaching heights of 10–15 m, and are found in a belt stretching from eastern Senegal to the Sudan/Ethiopia border [1]. The species is monoecious and mostly outcrosses, with a diploid genome of $2n = 24$ [2,3]. It is well known for its edible seed oil, which is rich in fatty acids and used in food, cosmetics, and pharmaceuticals. This provides significant economic and social benefits in many producing countries, especially for women. For example, it is estimated that about three million women in West Africa work in the shea sector, which generates between 90 and 200 million USD per year, supporting livelihoods and gender empowerment across Africa [4].

According to a new study, the worldwide market for shea butter is now expected to reach 3748 million USD by 2030 [5], presenting a significant opportunity for shea producing communities. Meeting such an increase will require invigorated efforts to manage and shape the future of the shea parklands, considering the entire supply chain, from growing/harvesting to procuring, processing, and marketing.

Despite their critical role in supporting livelihoods and gender empowerment, shea trees are under threat, with the species being classified as vulnerable by the International Union for Conservation of Nature [6]. These trees are facing an existential crisis due to excessive harvesting, foraging, overgrazing, charcoal production, and a lack of recruits, as well as the effects of climate change, which lead to a steady decline in their population [7,8].

To address the issues of genetic erosion and support the industry, urgent efforts must be made to significantly and measurably safeguard and improve shea landscapes with respect to shea tree populations and productivity. In this regard, the establishment of effective conservation and genetic improvement programs is crucial to secure a steadily improved and climate-stable shea production of high quality. Numerous projects and initiatives, focusing on germplasm collection, characterisation, evaluation, and conservation have been implemented in West Africa, such as the EU-funded "Innovative Tools and Techniques for Sustainable Use of the Shea Tree (INNOVKAR)" [9,10] have been implemented in West Africa and focus on germplasm collection, characterization, evaluation, and conservation. These initiatives have enabled the identification of superior trees (plus trees) with desirable phenotypes such as high nut yield and oil quality, through a participatory approach. Unfortunately, most of the identified "plus trees" are being conserved on farmers' land, making them more vulnerable to biotic and abiotic pressures [11]. Furthermore, only a few provenance/progeny trials have been established to conserve these germplasms, which limits the design and implementation of effective conservation and breeding programs, as well as the potential for elite germplasms deployment to farmers [8,11].

The study aims to investigate phenotypic and genotypic variation in nut and kernel attributes as well as the biochemical composition of kernels in two shea parklands. The parklands are located in Satiri, Burkina Faso, and in Kedougou, Senegal. This study is the first to be conducted in the field of shea genetic research in Senegal and Burkina Faso. Using a combination of single-nucleotide polymorphism (SNP) markers and biochemical techniques, it aims to ensure a sufficient supply of high-quality shea kernels to support further breeding and conservation efforts [12]. We hypothesized that (i) the distance would impact the genetic structure, the phenotypic and biochemical characteristics of kernels between stands within Senegal and between Senegalese and Burkinabe populations; and (ii) a positive phenotypic correlation is expected between nut/kernel traits and oil biochemical quality parameters.

2. Materials and Methods

2.1. Plant Material and Site Description

The plant material used in this study came from natural stands located in Senegal and Burkina Faso. In Senegal, shea trees are naturally found in the Kedougou region, in the south-east of the country. In contrast, shea parklands are very common in Burkina Faso. For this study, three populations representing the species' natural distribution of the species in Kedougou (Kenioto, Salemata, and Saraya), and Satiri parkland in Burkina Faso were selected (Table 1; Figure 1).

Table 1. Geographical and climate characteristics of the sites.

Site	Longitude	Latitude	Altitude	Annual Rainfall (mm)	Mean Temperature (°C)	
					Peak Hottest Months	Peak Coldest Months
Satiri	4° 03′ W	11° 34′ N	348	1055	35	25
Kenioto	12° 10′ W	12° 33′ N	167			
Salemata	12° 49′ W	12° 38′ N	171	1240	40	24
Saraya	11° 45′ W	12° 50′ N	151			

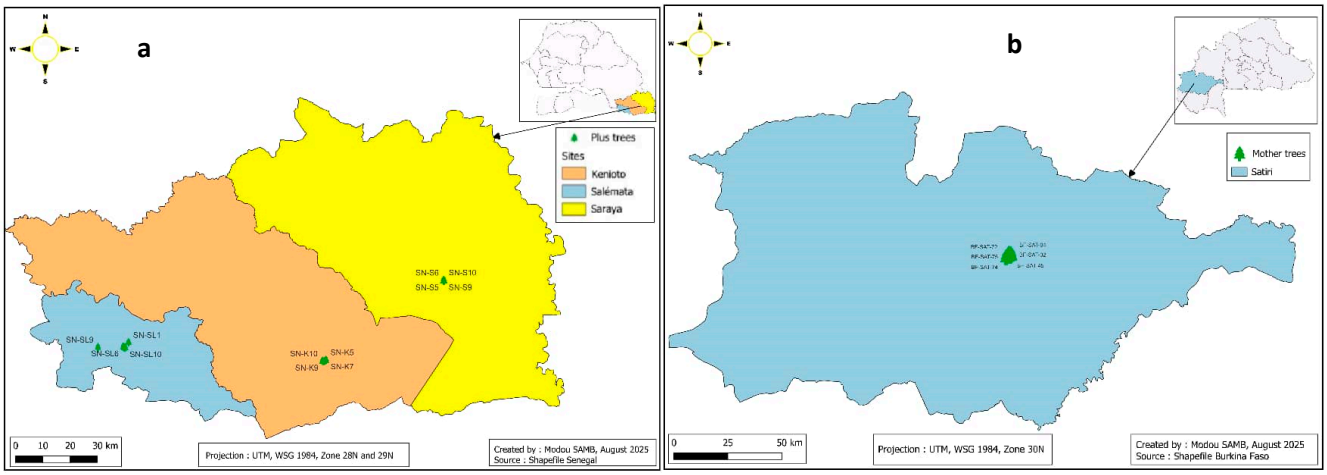


Figure 1. Maps showing the sample locations of shea trees collected from three sites in (**a**) Senegal and in (**b**) Burkina Faso.

2.2. Sampling

In the context of supporting the shea improvement program, "plus trees" were identified through a participatory approach in July 2021. In Senegal, discussions were organized with the President of the women's Federation at each site to identify and select the ten best-performing trees based on fruit and kernel attributes (nut/kernel yield and dimensions, oil yield and quality). Thus, a total of 30 trees were selected in Senegal.

In Burkina Faso, our sample comprised 80 mother trees that had already been identified in Satiri's parkland, near the "Centre de Recherche sur l'Arbre à Karité (CRAAK)" (Scheme 1). To capture as much diversity (allelic richness) as possible, both the selected "plus trees" and the mother trees in Senegal and Burkina were spaced 100 metres apart. The GPS coordinates of each tree were recorded.



Scheme 1. (**a**) Shea "plus tree" in Kénioto, Senegal; (**b**) Fresh nuts of the "plus tree" K9, originated from Kénioto, Senegal.

2.3. Genotyping Senegalese and Burkinabe Germplasm

2.3.1. DNA Extraction

Approximately 10–12 young, healthy leaf samples were collected from each of the selected trees, placed in zip-lock plastic bags containing a sufficient amount of silica gel, and sent to the genomics laboratory of the African Orphan Crops Consortium (AOCC) at the World Agroforestry Centre (CIFOR-ICRAF) in Nairobi, Kenya (worldagroforestry.org), for DNA extraction. As shea leaves are rich in mucilage, total genomic DNA was isolated using a modified CTAB procedure, developed at the African Orphan Crops Consortium (AOCC) Genomic Laboratory in Nairobi at World Agroforestry, Gigiri PO Box 30677, Nairobi, 00100, Kenya, and described in detail by [12,13]. This method includes two key steps to remove soluble mucilage: (i) adding 10% ethanol to the mixture of ground leaf sample and high-salt CTAB buffer to precipitate the mucilage; and (ii) using dichloromethane in addition to chloroform isoamyl alcohol for phasing out residual mucilage. This high-salt CTAB-based "selective mucilage exclusion protocol" has been used for whole genome DNA sequencing and re-sequencing of other orphan trees and crops like shea (*Vitellaria paradoxa*) [12].

After the DNA extraction, the integrity of the DNA was checked on a 0.8% agarose gel, and the optical density ratios were inspected using a Nanodrop ND-2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The DNA samples were then arranged in batches of 96. The arrayed DNA samples were genotyped using DArTSeq approach at Diversity Array Technology Inc., 02 6122 7300. Building 3, Level D, Kirinari Street University of Canberra, Bruce, ACT, Australia 2617, (https://www.diversityarrays.com).

2.3.2. Generation of DArT-SNP Markers

This analysis used the shea tree (*Vitellaria paradoxa*) reference genome developed by World Agroforestry (ICRAF) in Kenya in collaboration with the University of New Hampshire in the United States [12]. A low-density (250,000 reads, 72 bases on Illumina's HiSeq 2500) partial-genome profiling DArT sequencing service, which is one of the popular genotyping-by- sequencing (GBS) methods was used to discover and map the SNP markers as described by [14]. A total of 4821 phased bi-allelic DArT-SNP markers were generated from 88 DNA samples. To increase the power of the analysis, data with low heterozygosity (Minor Frequency Allele (MAF) of less than 5%,) and markers with more than 20% of

Missing data were filtered out. Using this criterion, 1625 SNPs were filtered out. Thus, a total of 3196 out of the original 4821 phased bi-allelic DArT SNP markers were used in the analysis. Additionally, individuals presenting loci with missing data (represented by "-9") in more than 20% of cases were also filtered out.

2.3.3. Population Structure and Diversity Analysis

Analyses were conducted using 3196 single-nucleotide polymorphisms (SNPs) to calculate the following genetic diversity parameters: the average number of allele per population (Na), polymorphism information content (PIC), observed and expected heterozygosity (Ho and He), and genetic differentiation (Fst). These parameters were computed using the snpReady package in R, version 4.3.0 [15]. Molecular variance analysis (AMOVA) was performed at different hierarchical levels (among and within populations and samples) using the adegenet package in R and was based on all four collecting sites (populations).

Analysis of population structure was performed using the ADMIXTURE version 1.3.0 in Unix [16]. This analysis required two steps. In step 1, we generated the necessary files for admixture with PLINK version 1.9.0-b.7.7 [17] using the following: script "plink --vcf genotypingData.vcf --make-bed --out genotypingData". Step 2 involved running ADMIXTURE version 1.3.0 using the following script "filename = "genotypingData.bed" for K in {1..5} do admixture --cv ${filename} -j16 $K | tee log_K${K}.out done". The optimal K value was determined by the lowest cross-validation error for each K value, thereby identifying the most suitable number of ancestry components.

The phylogenetic tree was constructed using the neighbour-joining method implemented in R. Fasta sequences were generated using the genotyping data and aligned using MUSCLE software version 3.8.31 [18]. These sequences were then used to construct an approximately maximum-likelihood phylogenetic tree, with 1000 times resampling without optimizing the branch lengths for the resampled alignments, using FastTree version 2.1.11 [19]. FastTree uses a heuristic variant of neighbor joining to get a rough topology. The resulting tree was plotted using FigTree software version 1.4.4 [20].

Discriminant Analysis of Principal Components (DACP) [21] was performed using the adegenet package in R, based on numerical data with genotype IDs to optimize the discrimination of individuals into groups, thus complementing the pattern of the genetic structure obtained with the ADMIXTURE method. In our study, the selection of the number of principal components was ad hoc, as reported in [22], and consisted in retaining the axis that explained more than 70% of the genetic variance.

2.4. Morphological and Biochemical Characteristics of the Selected "Mother, and Plus Trees"

2.4.1. Growth and Phenotypic Assessment of Nuts and Kernels

Quantitative growth traits were recorded in the field: height was measured using a graduate stick; trunk diameter was measured 30 cm above ground (for the sake of uniformity and consistency of measurement across trees, considering different growth habits of shea), and crown diameter (mean of two perpendicular crown diameter measures) was measured using a diameter tape. Fruit availability varied among the selected trees, leading to an unbalance. While fruits were collected from all Senegalese trees, in Burkina Faso, fruits were only available from 39 out of the 80 identified trees. A total of 69 trees were sampled for fruit. Around 1 kg of fresh fruit, corresponding to approximately 200 g of dried kernels, was collected from each tree in the field and taken to the laboratory for phenotypic characterisation. Twenty-five fruits were randomly chosen from each tree, numbered and used for morphological and biochemical characterisation. The pulp of the individually numbered fruits was then removed, and the nuts were mechanically extracted, cleaned and numbered accordingly. Furthermore, the size and weight of the fresh nuts

were measured, after which they were dried at 74 °C for 72 h in an oven (Salvis KVTS-22 Vacuum oven, Wetzikon, Switzerland). The dried nuts were then de-shelled using a stone to extract the kernels. The dimensions of the kernels were assessed using a digital calliper, and their weight was recorded. A total of 69 dried kernel samples were analysed for oil characteristics at the AAK biochemistry laboratory in Denmark.

2.4.2. Biochemical Analysis of Shea Oil Extracted from Kernels

The biochemical characterization of extracted shea oil obtained from harvested kernels was carried out using standard biochemical analysis methods developed by the Association of Official Analytical Chemists (AOAC) to evaluate its characteristics. A total of eight food quality parameters were measured: the crude fat content, fatty acids represented by: diglycerides, triglyceride, Triacylglycerol Mono Stearoyl Olein Stearin (TagMonoSOS); the fraction of unsaponifiable matter (1, 2-FFA, and 3) which is an indicator of the shea butter purity, consists of all compounds that are not esters of fatty acids and glycerol, such as phospholipids and phenolic compounds, sterols, squalene, and fat-soluble vitamins. The cariten content which refers to the rubber like substance found in shea butter. It is an isoprene polymer. The crude fat content, also known as total fat content, was measured using the Soxhlet extraction method (fatbySoxhlet), as described by [23]. The analysis of diglycerides, triglycerides and TAG, as well as the fraction of insaponifiables, was based on analytical high-performance liquid chromatography (HPLC), following the AOCS Official Method Cd 22-91 procedure [24], as described in [25]. HPLC analysis was conducted on Shimadzu LC-20AD (Shimadzu, Kyoto, Japan) equipped with CMB-20A controller, LC-20AD binary pump, SIL-20A autosampler, and CTO10AS column oven. Separation was achieved with an Agilent ZORBAX SB-C18 (4.6 mm × 250 mm; 5 µm, Agilent Technology, Santa Clara, CA, USA).

2.5. Statistical Analysis of Phenotypic Characters

Given the imbalance observed in our sampling for growth assessment (i.e., 30 trees in Senegal versus 80 trees in Burkina Faso) and in fruit sampling due to unexpected events in the field (fruit available from 39 out of 80 trees in Burkina Faso versus 30 samples from 30 trees in Senegal), an analysis of variance using the lm () function has been applied. Statistical analysis was performed in R version 4.4.3 Core Team 2017, following the model:

$$Y_{ij} = \mu + S_i + \text{error}_{ij}$$

where Y_{ij} is the value of the phenotypic trait (height, nut weight, Cariten. . .), µ is the grand mean, S_i is the effect of site i, error_{ij} is the residual error.

To visualize the relationships among the selected individuals with respect to the studied traits, a Principal Component Analysis (PCA) was performed in R, using the prcomp () function to group the genotypes into groups so that there is homogeneity within the clusters and heterogeneity between them. The data were scaled during the PCA analysis using the following script: "pcaresults<-prcomp(datascale,center=TRUE,scale=TRUE)". This was done to prevent variables with larger numerical ranges or units from dominating the analysis.

Standard Pearson correlation analysis using the rstatix package implemented in R version 4.4.3 have been performed, generating a matrix of correlation coefficients through the corr-mat () function and a corresponding matrix of *p*-value (cor_pmat ()) indicating the statistical significance of each relationship. Bonferroni corrections for multiple testing were not applied in any of the statistical tests. We find this justified since we only performed tests of biological meaningful hypotheses closely linked to research questions behind the

study. However it does mean that significant results based on tests statistics close to the 5% significance level should be interpreted with some care.

3. Results

3.1. Population Structure, Genetic Relationship, and Diversity Parameters

We investigated the population structure and genetic relationships within the studied population using the ADMIXTURE software. Given the expected close genetic relationship between samples collected in Senegal and Burkina Faso, the Bayesian clustering procedure was used to infer the population structure between these samples. The K value was determined basing on the cross-validation error rate; the K value corresponding to the lowest error rate was identified as the optimal K value (Figure 2a). When K = 2, the cross-validation error rate was the lowest, indicating that K = 2 was the best fit for our dataset (Figure 2a). This corresponds to two distinct clusters: shea trees from all sites in Senegal (G1) and individuals from Burkina Faso (G2) (Figure 2b).

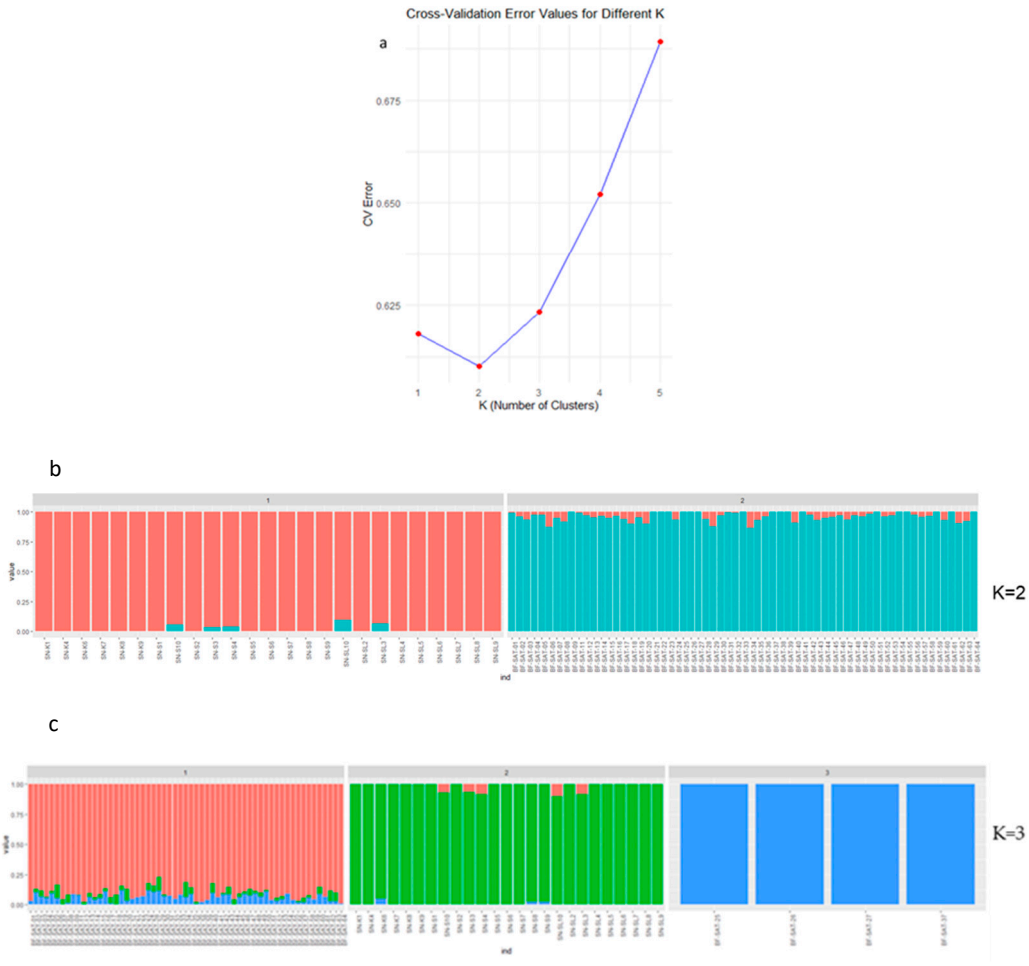


Figure 2. Genetic structure of 88 shea trees based on SNP markers. (**a**) Analysis of the cross validation error corresponding to different K values confirmed a higher likelihood at K = 2; (**b**) cluster of the 88 individuals at K = 2 showing 2 groups: Senegalese group in red and Burkinabe group in blue-green and (**c**) cluster of the 88 individuals at K = 3 showing 3 groups: Senegalese group in green; Burkinabe sub-groups in blue and red.

The neighbour-joining method allowed visualizing two clusters, following the geographic origin of our materials (Figure 3), this confirms the pattern observed in the Structure analysis.

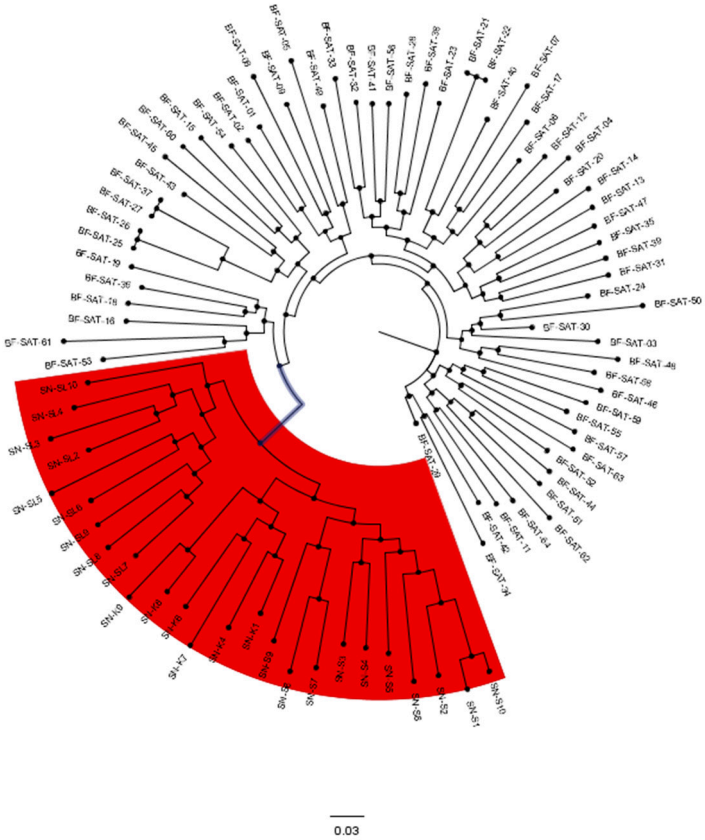


Figure 3. Neighbour-joining tree clustering of 88 shea trees. In white: Burkinabe group (G1) and in red: Senegalese group (G2).

In contrast, Discriminant Analysis of Principal Components (DAPC), which is based on the genetic distance between individuals (Figure 4), did not align with the ADMIXTURE analysis and the NJ tree results. Concerning the DA eigenvalue barplot in Figure 4, axes 1 and 2 were retained because they contained more than 80% of the information. The analysis revealed the presence of three major clusters. Here, all samples from Senegal fell into cluster 3 (G3), whereas samples from Satiri were categorised into two distinct clusters. The G1 subpopulation contained a relatively higher number of individuals (87%) than the G2 subpopulation, which consisted of eight genotypes.

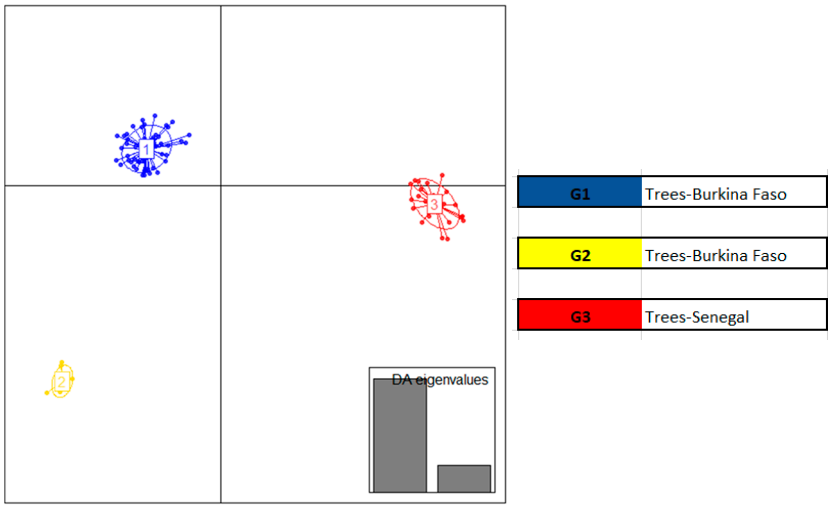


Figure 4. Discriminatory Analysis of Principal Component (DAPC). G1 and G2: samples collected in Burkina Faso (blue and yellow), and G3: samples from Senegal (red).

3.2. Allelic Pattern Across Populations and Molecular Analysis

The genetic diversity of the tested samples is summarized in Table 2. Overall, the genetic diversity was quite similar between the two groups, G1 and G2, with values of Na = 0.34, He = 0.39, and Ho = 0.16, and PIC = 0.25 for G1 versus Na = 0.33, He = 0.38, Ho = 0.16, and PIC = 0.15 for G2.

Table 2. SNP diversity parameters between the two groups of *V. paradoxa* subsp. *paradoxa*.

Populations	N	Na	He	Ho	PIC
G1 (Satiri-Burkina Faso)	66	0.34	0.39	0.16	0.25
G2 (Senegal)	25	0.33	0.38	0.16	0.15
Mean		0.335	0.385	0.16	0.20

N: sample size; Na: average number of alleles per population; Ho: observed heterozygosity; He: expected heterozygosity; PIC: Polymorphism Information content; G1: Group 1; G2: Group 2.

The AMOVA table has shown that the genetic variation between populations accounted for only 0.07% of the total variation, while the largest proportion of variation stemmed from differences within individuals (99.93%) (Table 3). Our study also showcased a very low and insignificant Fst value, suggesting low genetic differentiation among individuals (Fst = 0.0006).

Table 3. Analysis of Molecular Variance (AMOVA) showing the distribution of the variation within and among individuals.

Source	DF	SS	MS	Est. Variance	%
Among populations	1	57.15	28.57	0.01	0.07
Within samples	85	2404.73	28.29	28.29	99.93
Total					100

DF: Degree of Freedom; SS: Sum of Square; MS: Mean Square; Est. Variance: Estimated variance; %: percentage.

3.3. Morphological and Biochemical Characterization of Selected Germplasms

The analysis of the morphological characteristics revealed significant variations among sites for all studied traits (Table 4). At the country level, individuals from Senegal were significantly taller and larger than those from Burkina Faso (height: 14.18 m ± 2.24 versus 7.95 m ± 1.49; diameter: 49.11 cm ± 7.3 versus 26.14 cm ± 7.99; crown: 9.99 m ± 1.67 versus 6.3 m ± 2.68). In Senegal, a strong effect of sites was observed, with trees in the Saraya and Salemata sites presenting a lower growth rate than those in the Kenioto site (Figure 5).

Table 4. *F*-tests to test the significance of the difference in growth, nut, and biochemical attributes of kernel among sites in *V. paradoxa* subsp. *paradoxa*.

	Variation Among Populations		Means ± Standard Deviation	
Traits	***F* Value**	***p*-Value**	**Senegal**	**Burkina**
Growth traits				
Height (m)	44.61	<0.001	14.18 ± 2.24	7.95 ± 1.49
Diameter (cm)	25.73	<0.001	49.11 ± 7.3	26.14 ± 7.99
Crown (m)	12.68	<0.001	9.99 ± 1.67	6.3 + 2.68
Nut attributes				
Nut length (mm)	7.62	0.0002	29.05 ± 1.54	26.94 ± 2.27
Nut width (mm)	5.85	0.0013	23.63 ± 0.62	22.22 ± 1.41
Nut weight (g)	20.63	<0.001	6.72 ± 0.68	4.44 ± 0.77
Kernel length (mm)	3.70	0.0160	24.57 ± 0.93	23.02 ± 2.1
Kernel width (mm)	19.37	<0.001	19.08 ± 0.08	16.49 ± 1.21
Kernel weight (g)	19.56	<0.001	4.81 ± 0.17	3.02 ± 0.56

Table 4. *Cont.*

	Variation Among Populations		Means ± Standard Deviation	
Traits	***F* Value**	***p*-Value**	**Senegal**	**Burkina**
Biochemical characteristics				
Crude Fat content (wt%) *	12.61	<0.001	42.98 ± 2.13	47.62 ± 3.5
Diglyceride (wt%) **	13.33	<0.001	1.84 ± 0.27	3.06 ± 0.98
Triglyceride (wt%) **	7.66	0.0002	77.26 ± 6.28	77.72 ± 5.39
Cariten (wt%) **	5.22	0.0033	3.45 ± 0.63	3.08 ± 0.65
Unsapanifiable Matter 1 (wt%) **	4.57	<0.001	3.44 ± 0.61	2.59 ± 0.53
Unsaponifiable Matter 2-FFA (wt%) **	6.10	0.001	10.78 ± 3.99	10.83 ± 4.1
Unsaponifiable Matter 3 (wt%) **	14.59	<0.001	2.54 ± 0.56	1.99 ± 0.38
TagMonoSOS (wt)% ***	3.93	0.0122	33.22 ± 0.89	35.54 ± 3.01

* Relative to the mass of the crushed kernels; ** Relative to the mass of the extracted fatty matter; *** Relative to triglyceride content; TagMonoSOS: Triacylglycerol Mono Stearoyl Olein Stearin.

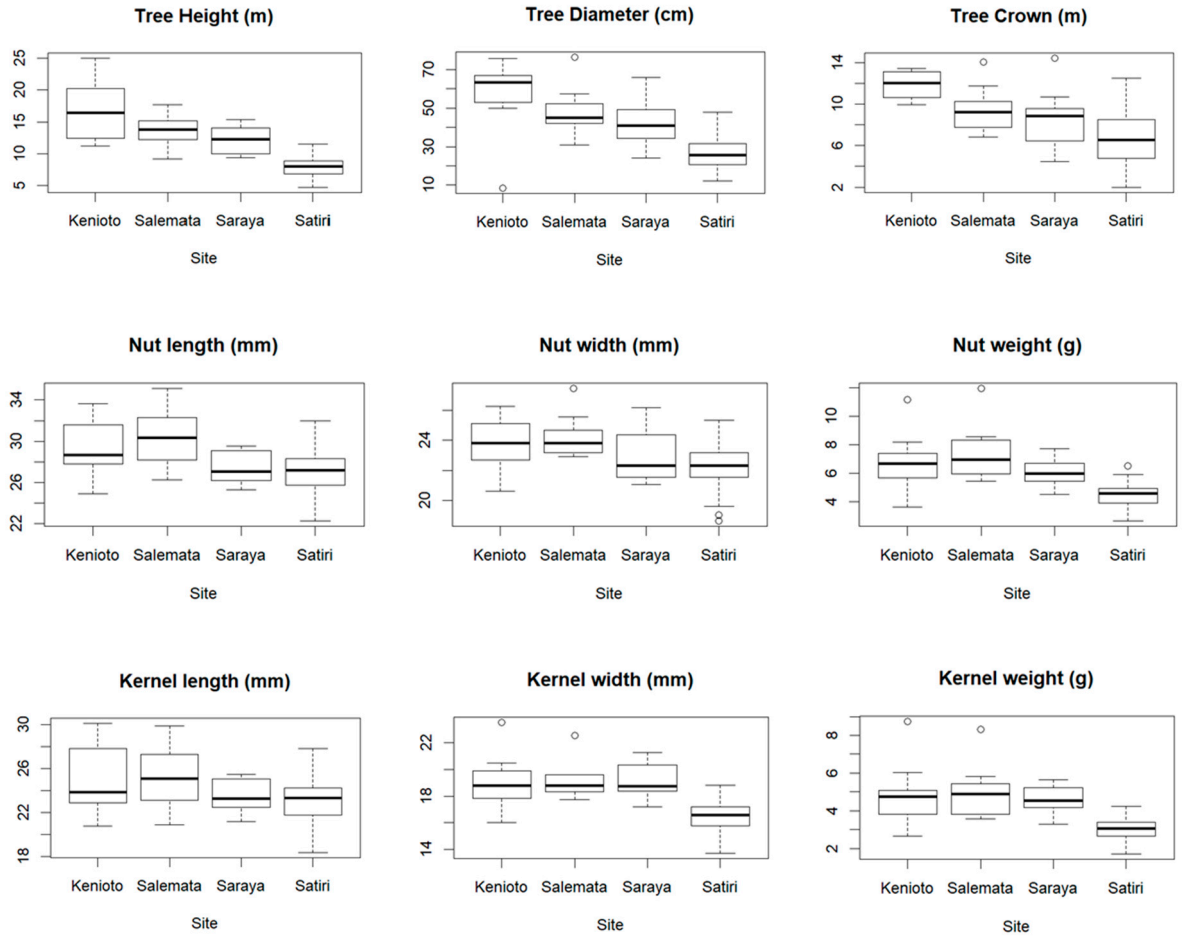


Figure 5. Comparison of average trait values across different sampled regions for selected plus trees.

Regarding nut characterization, no variation was found among Senegalese populations for nut attributes, except for nut length: the Salemata site differed significantly from the Saraya site (30.42 mm ± 2.76 versus 27.39 mm ± 1.49, respectively) (Figure 5). In contrast, a significant variation was observed among populations originating from Satiri in Burkina Faso and in Senegal, when all sites were considered together. The Salemata population presented the best performance in terms of nut size, with nuts that were 13% longer, 23% wider, and 54% heavier than those of the Satiri population (30.42 mm ± 2.76 versus 26.94 mm ± 2.27; 24.24 mm ± 1.41 versus 22.22 mm ± 1.41; 7.38 g ± 1.92 versus 4.44 g ± 0.77, respectively) (Table 4; Figure 5). A strong effect of sites

on kernel traits was also noted (Table 4; Figure 5): the Senegalese populations recorded better kernel attributes than the Burkina Faso population.

A strong effect of origin was observed in the biochemical composition of kernels. (Table 4). The Satiri population had the highest levels of crude fat, diglyceride, and TagMonoSOS contents (Table 4; Figure 6) compared to all the Senegalese sites combined. In contrast, kernels originating from Saraya contained 42% and 30% more unsaponifiable matter 2-FFA, and cariten than those belonging to the Satiri population, respectively (Table 4). A striking difference was spotlighted between the Senegalese populations: the kernels from the Salemata site contained 9% more crude fat content and 64% less amount of unsaponifiable matter 2-FFA than those from the Saraya site (Table 4).

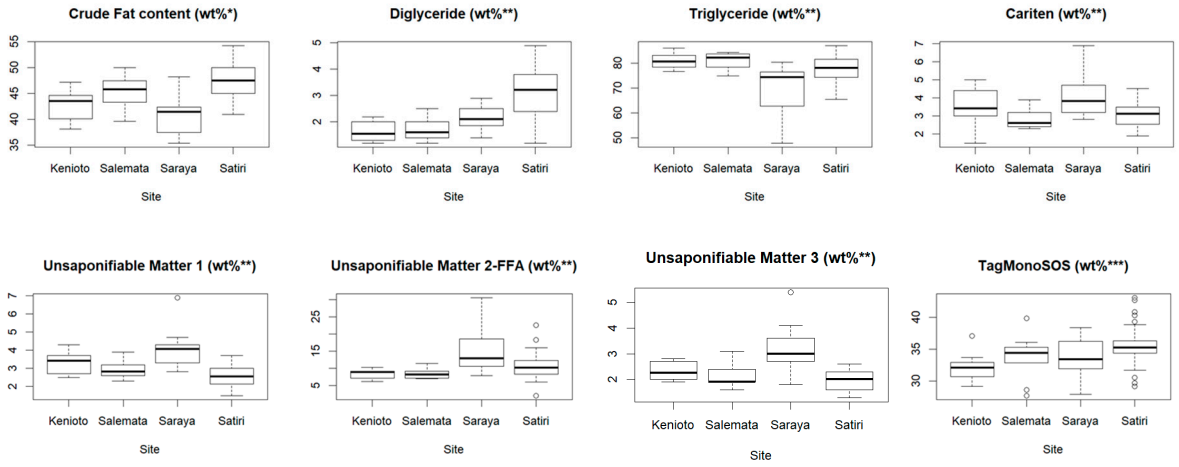


Figure 6. Biochemical composition of kernels among sites. * Relative to the mass of the crushed kernels; ** Relative to the mass of the extracted fatty matter; *** Relative to triglyceride content.

3.3.1. Multivariate Analysis and Correlation Among Traits

Principal component analysis (PCA) was performed separately on growth and nut/kernel characteristics (Figure 7) and on biochemical data (Figure 8). Regarding growth and nut/kernel attributes, the analysis identified two principal components (PCs) that both explained 86% of the total variance observed in the shea tree population. The PC1 (x-axis) accounted for 63% of the total variation, while PC2 (y-axis) accounted for 23%. As Figure 7 shows, the PC1 highlights the difference between the Satiri site, located on the positive side of the x-axis, and the other sites, on the negative side. Similarly, the growth traits (height, diameter, and crown) and the nut and kernel attributes were found to correlate negatively with PC1 (Figure 7). This indicates that shea trees originating from Senegal were characterized by superior nut and kernel attributes as well as growth traits, compared to those from the Satiri site.

Conversely, PCA analysis of the biochemical composition of the oil showed a clear differentiation of the variables among sites (Figure 8). Trees originating from the Satiri population were found to be positively correlated with the y-axis (24%), and they contained higher average levels of diglycerides, crude fat, and TagMonoSOS. However, the kernels from the Senegalese sites differed from those found in the Burkina Faso site, in terms of cariten and triglyceride content, as well as the amount of unsaponifiable matter 1 and 3.

In both PCAs, cluster analysis classified the trees into two distinct clusters depending on their origin (Figures 7 and 8). Cluster I was homogeneous and consisted mainly of individuals from the Satiri area, characterised by higher levels of diglycerides, TagMonoSOS, unsaponifiable matter 2-FFA, crude fat content, lower nut/kernel characteristics, and growth performance. Cluster II was heterogeneous and comprised trees from the three Senegalese sites (Salemata, Saraya, and Kenioto). The individuals in this cluster generally

presented better growth, nut and kernel attributes, higher levels of cariten, unsaponifiable matters 1 and 3, and triglycerides than those in cluster I.

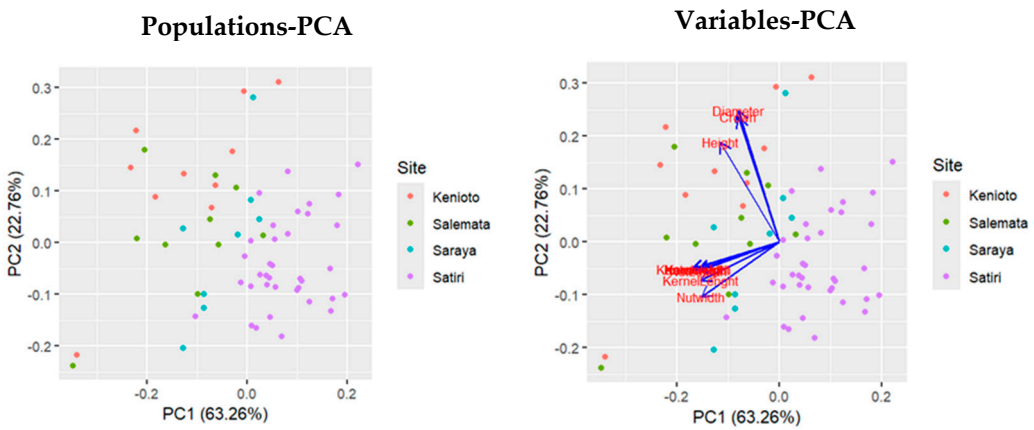


Figure 7. Principal Component analysis performed for growth and nut/kernel characters.

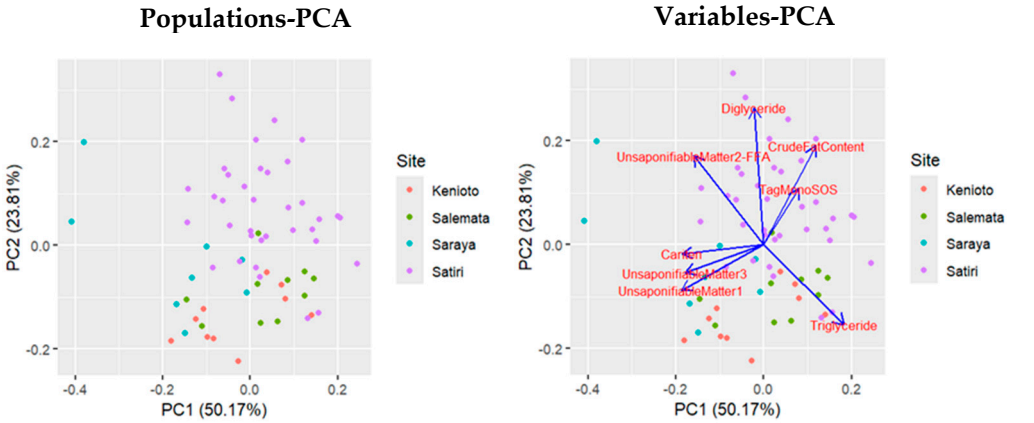


Figure 8. Principal Component analysis performed for biochemical data.

It is interesting to note that the Saraya population mainly comprised trees that outperformed others in terms of fractions of unsaponifiable matters and cariten content.

3.3.2. Phenotypic Correlation Among Traits

Phenotypic correlations between growth traits and nut/kernel dimensions and oil composition was in general moderate (Table 5). The results revealed highly significant positive correlation coefficients between growth traits (height, diameter and crown diameter) ranging from $0.69 < r_p < 0.81$. The same pattern was observed between nut and kernel dimensions (nut length, nut width, nut weight, kernel length, kernel width and kernel weight), with correlation coefficients ranging from 0.65 to 0.97, and between cariten content and the fraction of insaponifiable matters (1, 2-FFA and 3), with correlation coefficients ranging from 0.55 to 0.79. The same patterns were also observed between the amount of unsaponifiable matter 1 and 3 ($r_p = 0.82$), In contrast, there was a strong negative correlation between triglyceride content and cariten, as well as between triglyceride content and the fractions of unsaponifiable matters 1, 2-FFA and 3 ($r_p = -0.69$, -0.60, -0.96 and -0.61, respectively).

Table 5. Phenotypic correlation among traits Bold text indicates a significant correlation.

	Height	Diam	Crown	NutL	NutW	Nutwght	KernL	KernW	Kernwght	CrudeFat	Digly	Trigly	Cariten	USM1	USM2-FFA	USM3
Diam	**0.79** ***															
Crown	**0.69** ***	**0.81** ***														
NutL	0.43	0.23	0.29													
NutW	0.31	0.15	0.099	**0.78** ***												
Nutwght	0.51	0.29	0.29	**0.84** ***	**0.86** ***											
KernL	0.41	0.14	0.24	**0.95** ***	**0.75** ***	**0.82** ***										
KernW	0.47	0.27	0.24	**0.65** ***	**0.85** ***	**0.90** ***	**0.67** ***									
Kernwght	0.51	0.27	0.28	**0.79** ***	**0.83** ***	**0.97** ***	**0.80** ***	**0.94** ***								
CrudeFat	–0.35	–0.39	–0.29	–0.22	–0.18	–0.29	–0.17	–0.28	–0.33							
Digly	–0.44	–0.43	–0.25	–0.32	–0.33	–0.43	–0.24	–0.44	–0.4	0.3						
Trigly	0.14	0.12	0.11	0.19	0.17	0.14	0.18	0.11	0.086	0.18	–0.48					
Cariten	0.14	0.067	0.053	–0.049	–0.037	0.023	–0.096	0.025	0.055	–0.45	0.018	**–0.69** ***				
USM1	0.32	0.36	0.22	0.035	0.077	0.21	–0.013	0.25	0.26	–0.56	–0.21	**–0.60** ***	**0.79** ***			
USM2-FFA	–0.13	–0.12	–0.1	–0.16	–0.18	–0.13	–0.15	–0.11	–0.08	–0.11	0.46	**–0.96** ***	**0.55** ***	0.46		
USM3	0.26	0.27	0.11	0.049	0.18	0.26	0.029	0.32	0.31	–0.45	–0.13	**–0.61** ***	**0.72** ***	**0.83** ***	0.46	
TAGMS	–0.26	–0.24	–0.23	–0.24	–0.22	–0.33	–0.21	–0.26	–0.33	0.54	0.038	0.16	–0.24	–0.28	–0.17	–0.15

Diam: Diameter; NutL: Nut length; NutW: Nut width; Nutwght: Nut weight; KernL: Kernel length; KernW: Kernel width; Kernwght: Kernel weight; Digly: Diglyceride; Trigly: Triglyceride; USM1: Unsaponifiable Matter 1; USM2-FFA: Unsaponifiable Matter 2-FFA; USM3: Unsaponifiable Matter 3; TAGMS: triacylglycerol mono stearoyl olein stearin. ***: $p < 0.0001$.

4. Discussion

4.1. Genetic Diversity Analysis and Structuring

In the context of breeding and conservation of the shea resource in West Africa, investigating the genetic diversity and structure within *V. paradoxa* subsp. *paradoxa* could provide useful information to weave effective breeding and conservation strategies. To the best of our knowledge, this study is the first that uses single nucleotide polymorphisms (SNPs) to investigate the genetic diversity and population structure of *V. paradoxa* subsp. *paradoxa* in Senegal and Burkina Faso. The genetic structure of our germplasm was analyzed using various methodologies, including ADMIXTURE analysis, DAPC, and phylogenetic tree construction, thereby providing accurate and complementary information supporting the division of our shea tree resources into two major groups.

Overall, the results of the genetic diversity revealed a moderate level of gene diversity (mean He = 0.385) and polymorphism information content (PIC: 0.20). This is in line with previous results focused on shea germplasms originating from East Africa [26], the Ivory Coast [27], and Ghana [28]. The relatively moderate genetic diversity observed in this study could be attributed to the species' semi-domesticated nature [8,29], as compared to undomesticated (wild) species, for which higher heterozygosity is expected due to absence of human induced selection pressure which acts much faster than natural selection [30]. Alternatively, this could be due to genetic erosion resulting from anthropogenic pressures, such as logging for firewood and climate change. These factors lead to a reduction in tree density and to an increase of the vulnerability of the resource in its natural habitat [6]. The observed moderate genetic diversity may also be due to the bi-allelic of nature of SNPs and their low mutation rates of SNPs compared to multiallelic markers such as simple sequence repeats (SSRs,) which can have several alleles, even up to 10–15 as discussed by [27]. However, it is interesting to note that the average gene diversity in the Senegal and Burkina Faso genepools exceeded that in the Ghanaian and Ivory Coast gene pools (He = 0.26 and He = 0.24, respectively) [27,28], suggesting that these populations could be used for future hybridization and improvement work to expand the genetic diversity of the breeding panel.

In our study, the values of genetic diversity parameters Na, He, Ho were almost identical for the two groups G1 and G2. Additionally, there is no significant genetic differentiation between them regardless of their geographical location. The similarity in the extent of genetic diversity and the low genetic differentiation suggests that shea germplasm from Senegal and Burkina Faso was likely connected in the past by gene flow through the dispersal of pollen and seeds, facilitating constant gene exchange. As a result, the populations have a similar extent of genetic diversity. However, significant differences were spotted between the two groups at the phenotypic level, which could be explained by the influence of climate and environmental factors (e.g., altitude, rainfall, and soil characteristics).

Still, the expected heterozygosity was more than twice the observed heterozygosity, which may indicate local inbreeding [31,32]. This can lead to a decrease in genetic variation and adaptive potential [33,34]. Therefore, there is an urgent need for genetic management decisions to decrease the homozygosity levels in our populations and increase the adaptive potential of shea resources. This is crucial for the sustainability of this resource, which has already been classified as 'vulnerable' by the IUCN [6].

The ADMIXTURE analysis (K = 2), which is based on the Bayesian method and neighbour joining analysis, supported clustering based on geography. In fact, all the genotypes in G1 belonged to the Satiri site in Burkina Faso, while G2 consisted of individuals from all the sites in Senegal. Additionally, the two groups presented very low genetic differentiation (Fst = 0.0006), suggesting either gene flow or that these populations are essentially the same

population descended from a shared ancestor. However, the cluster pattern in DAPC was interesting in that it categorized the shea trees into three groups. Through this approach, all the samples from Senegal were clustered together (G3), while the Satiri shea materials were divided into two genetically distinct clusters (G1 and G2). Thus, this analysis further highlighted the grouping of cluster G1, dividing it into two additional groups. This confirms that the DAPC approach generally outperforms the ADMIXTURE analysis in characterising population subdivision [22]. As a non-model-based multivariate approach, DAPC is particularly sensitive to subtle genetic differentiation, a pattern commonly observed in crop and tree species with complex demographic histories [22]. Similarly, a divergence between DAPC and ADMIXTURE has been reported in cacao (*Theobroma cacao*), where DAPC resolved three to four genetic clusters associated with domestication and breeding history, while ADMIXTURE revealed extensive admixture among accessions [35]. In African rice (*Oryza glaberrima*), DAPC identified three major genetic clusters corresponding to agro-ecological zones, whereas ADMIXTURE primarily captured broad ancestry components with limited resolution of fine-scale population structure [36]. Comparable results were also observed in olive (*Olea europaea*), a perennial tree crop, where DAPC clearly discriminated regional genetic clusters that were only weakly supported by Bayesian clustering methods due to long-term gene flow and clonal propagation [37].

Examining the composition of the two smaller clusters revealed by ADMIXTURE (Figure 2, K = 3, blue group) and DAPC (Figure 4, yellow group) within the Burkina Faso gene pool showed that, the smaller cluster obtained from ADMIXTURE, consisted of only four individuals whereas the DAPC cluster contained four additional trees, totaling eight trees. Thus, this confirms that both clusters consist of more or less the same individuals, indicating that both methods captured the same underlying genetic signal. The presence of a second genetic cluster within such a relatively small sampling area in Burkina Faso is intriguing. While all of these individuals were sampled in the same location, the semi-domesticated nature of these shea parklands may indicate some degree of human intervention i.e., selection. However, it is important to emphasize that farmers never plant shea trees. Therefore, the observed pattern cannot be explained by planting or the introduction of new accessions by farmers. We are therefore speculating that spatial isolation at a local scale (i.e., topography or altitude) could have led to this pattern [38] However, the results of genetic diversity and differentiation obtained in this study should be interpreted with some care due to imbalance of the sampling. Therefore, further studies with a larger sample size are needed to confirm or elucidate the observed pattern.

4.2. Phenotypic Traits Structuring in Relation to Genetic Diversity Pattern

A high level of phenotypic variability of shea germplasm in terms of growth, nut and kernel attributes, and oil quality traits has been widely reported by various authors [11,39,40]. A similar pattern was observed in our study. In fact, the phenotypic differentiation followed the genotype clustering from ADMIXTURE and the NJ tree, suggesting that the two groups obtained with SNP markers reflected two morphological groups. Although some similarity in the extent of genetic diversity and low genetic differentiation among Senegalese and Burkinabe populations has been observed, significant differences were found between the Senegalese and the Burkinabe groups at the phenotypic level. These could be either explained by the influence of climate and environmental factors (e.g., altitude, rainfall and soil characteristics), which have been reported to significantly affect the expression of shea tree morphological traits in many countries [41], or by the DNA makeup of each population. Regarding nut and kernel attributes (dimensions), cariten content, and the amount of unsaponifiable matter, it is interesting to note that the G2 cluster, representing the Senegalese gene pool, recorded the highest performance. In

contrast, higher levels of fat, diglycerides, triglycerides, and TagMonoSOS were observed in the G1 group representing the Satiri collection. At the individual level, the trees with the best key food quality traits (unsaponifiable matter-free fatty acids (FFA) and cariten content) belonged to the Saraya site, which is located in the extreme east of the country. Trees from the Kenioto and Salemeta populations were specifically characterised by better growth and larger nuts and kernels. However, these findings did not entirely align with the cluster analysis based on molecular markers, which did not reveal any genetic differentiation among the Senegalese populations. Therefore, it would be interesting to investigate whether the observed differences between individuals and populations are under genetic or environmental control. The establishment of multi-location trials (progeny trials/Breeding Seed Orchards (BSOs)/Clonal Seed Orchards (CSOs)) of the best-performing genotypes in both countries, coupled with association mapping research to assess additive genetic variance, would help to determine the extent to which nut/kernel size and oil quality traits are under genetic control [42–45].

4.3. Traits Correlation and Breeding Implications

The correlation between different morphological traits in shea has been demonstrated by many authors [2,8,9,11]. In our study, we found a high phenotypic correlation between nut and kernel traits, as is usually reported in previous studies of shea and cashew [45]. The low and moderate phenotypic correlation between oil composition and growth/nut/kernel traits observed in our study implies that oil composition is only weakly dependent on the size of the tree, nut and kernel. The relatively high correlations between cariten content and the fractions of insaponifiable matters (1, 2-FFA and 3), and between the amounts of unsaponifiable matters 1 and 3, coupled with the strong negative correlations between triglyceride content and cariten, and between triglyceride content and the fractions of insaponifiable matters 1, 2-FFA and 3, suggest that selecting superior trees based on cariten content, for example, may result in the selection of individuals with higher fractions of insaponifiable matters and lower triglyceride content.

Growth traits are mainly influenced by environmental conditions (e.g., rainfall, topography), making them less heritable, whereas fruit size, oil composition are generally under strong genetic control. Therefore, further research is crucial to accurately estimate the correlation between shea oil components and nut/kernel sizes, to plan a more efficient breeding program.

5. Conclusions

This study foregrounds the importance of using high-density markers to evaluate genetic diversity and population structure, to support breeding and conservation activities in the shea industry of Senegal and Burkina Faso. Given the moderate to low level of genetic differentiation between the two populations, our findings suggest that the best individuals can be used as effective parents (i.e., in controlled crosses) for breeding to create a hybrid. Our study also revealed two genetic groups with specific phenotypic characteristics within the study population. Overall, we observed high phenotypic variation in nut and kernel attributes and oil composition, highlighting the potential gains from selection and breeding. Thus, the development of breeding zones, seed sources, and panel populations may lead to better utilisation and conservation of the species' genetic potential. Further research based on estimating narrow-sense heritability and additive genetic variance, as well as identifying genes of interest using genome-wide association studies (GWAS) and marker-assisted selection (MAS), is crucial to enhance genetic gain through i.e., bi-parental breeding programs, and ensure the success of breeding programs.

Author Contributions: Conceptualization, A.M.D. and P.H.; methodology, A.M.D., P.H., R.K., S.M., J.N., D.L. and T.K.R.; formal analysis, A.M.D., S.D., M.H.A., T.K.R. and P.H.; writing original draft preparation, A.M.D.; writing—review and editing: A.M.D., P.H., S.D., D.L. and T.K.R.; supervision: P.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the "TWAS-IsDB post-doctoral fellowship, letter of agreement Number 11/2020, Vendor Number 506988" and Theme Trees ICRAF, Nairobi. The APC was funded by Theme Trees, ICRAF, Nairobi.

Data Availability Statement: The data will be made available at ICRAF data Archive.

Acknowledgments: We are grateful to Ramni Jamnadass and Lars Graudal from CIFOR-ICRAF, who supported the conception and application of this postdoc project to the TWAS-IsDB fellowship. We wish to thank Marcel Badji for assistance with field work, Diatta Marone for facilitating the administrative procedure at ISRA/CNRF, Modou Samb and Mamadou Ousseynou Ly for map realization, AOCC genomics laboratory staff for assisting in the DNA extraction and genotyping. Financial support was provided by the TWAS-IsDB fellowship and by the Theme Trees at ICRAF, Nairobi.

Conflicts of Interest: The authors declare no conflicts of interest. Author Dr. Tore Kiilerich Ravn is employed by the company Aarhus Karlshamn (AAK) and undertook the biochemical analysis of kernels in the lab. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Hall, J.B.; Aebischer, D.P.; Tomlinson, H.F.; Osei-Amaning, E.; Hindle, H.R. *Vitellaria paradoxa: A Monograph*; School of Agriculture and Forest Sciences Publication No. 8; University of Wales, Bangor: Gwynedd, UK, 1996; 105p.
2. Sanou, H.; Lamien, N. *Vitellaria paradoxa*, Shea Butter Tree. In *Conservation and Sustainable Use of Genetic Resources of Priority Food Tree Species in Sub-Saharan Africa*; Bioversity International: Rome, Italy, 2011.
3. Choungo Nguekeng, P.B.; Hendre, P.; Tchoundjeu, Z.; Kalousová, M.; Tchanou Tchapda, A.V.; Kyereh, D.; Masters, E.; Lojka, B. The Current State of Knowledge of Shea Butter Tree (*Vitellaria paradoxa* C.F.Gaertner) for Nutritional Value and Tree Improvement in West and Central Africa. *Forests* **2021**, *12*, 1740. [CrossRef]
4. Chen, T. The Impact of the Shea Nut Industry on Women's Empowerment in Burkina Faso; 2017; FAO Report, ISBN 978-92-5-130005-3. Available online: https://openknowledge.fao.org/server/api/core/bitstreams/4c577507-9ba7-4316-aa81-9064a62d0b1b/content (accessed on 23 October 2025).
5. Available online: https://www.grandviewresearch.com/industry-analysis/shea-butter-market/request/rs1 (accessed on 24 November 2025).
6. IUCN Red List of Threatened Species: *Vitellaria paradoxa*. IUCN Red List Threat Species. 2023. Available online: https://www.iucnredlist.org/species/37083/10029534 (accessed on 19 December 2025).
7. Bayala, J.; Ky-Dembele, C.; Kalinganire, A.; Olivier, A.; Nantoumé, H. *A Review of Pasture and Fodder Production and Productivity for Small Ruminants in the Sahel*; ICRAF Occasional Paper No. 21; World Agroforestry Centre: Nairobi, Kenya, 2014.
8. Boffa, J.M. *Opportunities and Challenges in the Improvement of the Shea (Vitellaria paradoxa) Resource and Its Management*; Occasional Paper 24; World Agroforestry Centre: Nairobi, Kenya, 2015.
9. Diarrassouba, N.; Yao, S.D.M.; Traoré, B. *Identification Participative et Caractérisation des Arbres Elites de Karité Dans la Zone de Production en Côte d'Ivoire*; Côte d'Ivoire (projet FIRCA/Karité), Report No.: N° 069/2016; University Peleforo Gon Coulibaly: Korhogo, Côte d'Ivoire, 2017; 15p.
10. Sandwidi, A.; Diallo, B.O.; Lamien, N.; Vinceti, B.; Sanon, K.; Coulibaly, P.; Sawadogo, P.M. Participatory identification and characterisation of shea butter tree (*Vitellaria paradoxa* C.F. Gaertn.) ethnovarieties in Burkina Faso. *Fruits Int. J. Trop. Subtrop. Hortic.* **2018**, *73*, 141–152. [CrossRef]
11. Attikora, A.J.P.; Diarrassouba, N.; Yao, S.D.M.; Clerck, C.D.; Silue, S.; Alabi, T. Morphological traits and sustainability of plus shea trees (*Vitellaria paradoxa* C.F.Gaertn.) in Côte d'Ivoire. *Biotechnol Agron. Soc. Environ.* **2023**, *27*. [CrossRef]
12. Hale, I.; Ma, X.; Melo, A.T.O.; Padi, F.K.; Hendre, P.S.; Kingan, S.B.; Sullivan, S.T.; Chen, S.; Boffa, J.M.; Muchugi, A.; et al. Genomic Resources to Guide Improvement of the Shea Tree. *Front. Plant Sci.* **2021**, *12*, 720670. [CrossRef]
13. Bredeson, J.V.; Lyons, J.B.; Oniyinde, I.O.; Okereke, N.R.; Kolade, O.; Nnabue, I.; Nwadili, C.O.; Hřibová, E.; Parker, M.; Nwogha, J.; et al. Chromosome evolution and the genetic basis of agronomically important traits in greater yam. *Nat. Commun.* **2022**, *13*, 2001. [CrossRef]

14. Kilian, A.; Wenzl, P.; Huttner, E.; Carling, J.; Xia, L.; Blois, H.; Caig, V.; Heller-Uszynska, K.; Jaccoud, D.; Hopper, C.; et al. Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods Mol. Biol.* **2012**, *888*, 67–89.
15. Granato, I.S.C.; Galli, G.; Couto, E.G.O.; e Souzza, M.B.; Mendonca, L.F.; Fritsche-Neto, R. *snpReady*: A tool to assist breeders in genomic analysis. *Mol. Breed.* **2018**, *38*, 102. [CrossRef]
16. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [CrossRef]
17. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
18. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2024**, *32*, 1792–1797. [CrossRef]
19. Price, M.N.; Paramvir, S.D.; Adam, P.A. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef]
20. Rambaut, A. FigTree v1.4.4: Tree Figure Drawing Tool. 2018. Available online: https://tree.bio.ed.ac.uk/software/figtree/ (accessed on 17 November 2025).
21. Jombart, T.; Caitlin, C. *A Tutorial for Discriminant Analysis of Principal Components (DAPC) Using Adegenet 2.1.0*; Imperial College: London, UK, 2017.
22. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 94. [CrossRef]
23. Peiris, T.L.V. Determination of Crude Fat Content. 2009. GS/MSc/Food/3630/08. Available online: https://fr.scribd.com/document/480055415/AOAC-920-39 (accessed on 25 November 2025).
24. AOCS. *AOCS Official Method Cd 22-91: Determination of Polymerized Triglycerides by Gel-Permeation HPLC*; AOCS Press: Champaign, IL, USA, 2009.
25. Wang, J.; Long, Q.; Zhong, H. Influence of Temperature on Triacylglycerol Degradation in Camellia Seed Oil during Accelerated Thermal Oxidation. *J. Food Nutr. Res.* **2018**, *6*, 320–328. [CrossRef]
26. Odoi, J.B.; Adjei, E.A.; Hendre, P.; Nantongo, J.S.; Ozimati, A.A.; Badji, A.; Nakabonge, G.; Edema, R.; Gwali, S.; Pouyan, T.L.O.P.; et al. Genetic diversity and population structure among Ugandan shea tree (*Vitellaria paradoxa* subsp. *nilotica*) accessions based on DarTSeq markers. *Crop Sci.* **2023**, *63*, 2297–2309. [CrossRef]
27. Attikora, A.J.P.; Yao, S.D.M.; Dago, D.N.; Silué, S.; De Clerck, C.; Kwibuka, Y.; Diarrassouba, N.; Alabi, T.; Achigan-Dako, E.G.; Lassois, L. Genetic diversity and population structure of superior shea trees (*Vitellaria paradoxa* subsp. *paradoxa*) using SNP markers for the establishment of a core collection in Côte d'Ivoire. *BMC Plant Biol.* **2024**, *24*, 913. [CrossRef]
28. Anyomi, W.E.; Barnor, M.T.; Danquah, A.; Ofori, K.; Padi, F.K.; Avicor, S.W.; Hale, I.; Danquah, E.Y. Heritability and Genetic Advance Estimates of Key Shea Fruit Traits. *Agronomy* **2023**, *13*, 640. [CrossRef]
29. Odoi, J.B.; Muchugi, A.; Okia, C.A.; Gwali, S.; Odong, T.L. Local knowledge, identification and selection of shea tree (*Vitellaria paradoxa*) ethnovarieties for pre-breeding in Uganda. *J. Agric. Nat. Resour. Sci.* **2020**, *7*, 22–33.
30. Benlioğlu, B.; Adak, M.A. Importance of Crop Wild Relatives and Landraces. Genetic Resources in Plant Breeding Programmes. *J. Exp. Agric. Int.* **2019**, *37*, 1–8. [CrossRef]
31. Hoffmann, A.A.; White, V.; Jasper, M.; Yagui, H.; Sinclair, S.; Kearney, M. An endangered flightless grasshopper with strong genetic structure maintains population genetic variation despite extensive habitat loss. *Ecol. Evol.* **2021**, *11*, 5364–5380. [CrossRef]
32. Schmidt, T.L.; Jasper, M.E.; Weeks, A.R.; Hoffmann, A.A. Unbiased population heterozygosity estimates from genome wide sequence data. *Methods Ecol. Evol.* **2020**, *12*, 1888–1898. [CrossRef]
33. Ndiaye, L.; Diallo, A.M.; Vu, G.; Mueller, M.; Ngom, D.; Mbaye, T.; Gailing, O. Genetic diversity of populations of *Dalbergia melanoxylon* Guill. & Perr. in the Ferlo zone (Senegal) using chloroplast and nuclear microsatellite markers. *Genet. Resour. Crop Evol.* **2024**, *72*, 4901–4913. [CrossRef]
34. Ralls, K.; Ballou, J.D.; Dudash, M.R.; Eldridge, M.D.B.; Fenster, C.B.; Lacy, R.C.; Sunnucks, P.; Frankham, R. Call for a paradigm shift in the genetic management of fragmented populations. *Conserv. Lett.* **2018**, *11*, e12412. [CrossRef]
35. Lombardi, R.; Caccamo, M.; Materazzi, A. Genetic structure and domestication history of cacao revealed by multivariate and Bayesian approaches. *Tree Genet. Genomes* **2018**, *14*, 22.
36. Philippe Cubry, P.; Tranchant-Dubreuil, C.; Thuillet, A.C.; Monat, C.; Ndjiondjop, M.N.; Labadie, K.; Cruaud, C.; Engelen, S.; Scarcelli, N.; Rhoné, B.; et al. The rise and fall of African rice cultivation revealed by genomic analyses. *Curr. Biol.* **2018**, *28*, 2274–2282. [CrossRef]
37. Belaj, A.; Dominguez-García, M.D.C.; Atienza, S.G.; Urdíroz, N.M.; De la Rosa, R.; Satovic, Z.; Martín, A.; Kilian, A.; Trujillo, I.; Valpuesta, V.; et al. Developing a core collection of olive cultivars using SSR markers and multivariate analysis. *Tree Genet. Genomes* **2011**, *8*, 365–379. [CrossRef]

38. Liu, H.; Wang, Z.; Zhang, Y.; Li, M.; Wang, T.; Su, Y. Geographic isolation and environmental heterogeneity contribute to genetic differentiation in *Cephalotaxus oliveri*. *Ecol. Evol.* **2023**, *13*, e9869. [CrossRef] [PubMed]
39. Sanou, H.; Lovett, P.N.; Bouvet, J.M. Comparison of quantitative and molecular variation in agroforestry populations of the shea tree in (*Vitellaria paradoxa* C.F. Gaertn) Mali. *Mol. Ecol.* **2005**, *14*, 2601–2610. [CrossRef]
40. Luo, Z.; Brock, J.; Dyer, J.M.; Kutchan, T.; Schachtman, D.; Augustin, M.; Ge, Y.; Fahlgren, N.; Abdel-Haleem, H. Genetic Diversity and Population Structure of a *Camelina sativa* Spring Panel. *Front. Plant Sci.* **2019**, *10*, 184. [CrossRef]
41. Yao, S.D.M.; Diarrassouba, N.; Attikora, A.; Fofana, I.J.; Dago, D.N.; Silue, S. Morphological diversity patterns among selected elite Shea trees (*Vitellaria paradoxa* C.F. Gaertn.) from Tchologo and Bagoué districts in Northern Côte d'Ivoire. *Int. J. Genet. Mol. Biol.* **2020**, *12*, 1–10. [CrossRef]
42. Dhakal, L.P.; Lillesø, J.P.B.; Kjær, E.D.; Jha, P.K.; Aryal, H.L. *Seed Sources of Agroforestry Trees in a Farmland Context—A Guide to Tree Seed Source Establishment in Nepal; Forest and Landscape Development and Environment Series 1*; Forest & Landscape Denmark Hørsholm Kongevej 11 DK-2970 Hørsholm Denmark: Fredensborg, Denmark, 2005; Development and Environment Series no. 1-2005, ISBN 87-7903-251-6. Available online: https://www.cifor-icraf.org/publications/downloads/Publications/PDFS/b13782.pdf (accessed on 25 November 2025).
43. Diallo, A.M.; Nielsen, L.R.; Hansen, J.K.; Ræbild, A.; Kjær, E.D. Study of quantitative genetics of gum Arabic production complicated by variability in ploidy level of *Acacia senegal* (L.) Willd. *Tree Genet. Genomes* **2015**, *11*, 80–92. [CrossRef]
44. Coşkun, O.F.; Gulsen, O. Determination of markers associated with important agronomic traits of watermelon (*Citrullus lanatus* L.). *J. Agric. Sci. Technol.* **2024**, *26*, 1359–1371. [CrossRef]
45. Sankharé, M.; Diallo, A.M.; Ba, H.S.; Diatta, S.; Samb, C.O.; Touré, M.A.; Badiane, S. Phenotypic diversity of growth, leaf and yield-related traits in cashew (*Anacardium occidentale* L.): Implications for the development of a cashew breeding program in Senegal. *Genet. Resour. Crop Evol.* **2025**, *72*, 6771–6781. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.