LIÈGE université
Gembloux
Agro-Bio Tech

Conservation and Genetic Improvement of Shea Tree (*Vitellaria paradoxa* C.F. Gaertn) in Côte d'Ivoire

Affi Jean Paul ATTIKORA

FRENCH COMMUNAUTY OF BELGIUM
UNIVERSITY OF LIEGE – GEMBLOUX AGRO-BIO TECH

Conservation and Genetic Improvement of Shea Tree (*Vitellaria paradoxa* C.F. Gaertn) in Côte d'Ivoire

Affi Jean Paul ATTIKORA

Original dissertation presented for the degree of Doctor of Philosophy (Ph.D.) in Agronomic Sciences and Bioengineering.

Promoter: Dr Ludivine LASSOIS

Civil Year: 2025

Abstract

Shea tree (*Vitellaria paradoxa*) is a vital indigenous African species with significant economic, social and environmental importance. Despite the economic importance of shea butter as a multipurpose product, research efforts toward the genetic improvement of the shea tree have remained limited. In addition, the species faces multiple threats from biotic and abiotic pressures, which compromise both its natural populations and the sustainable supply of shea butter. Addressing these challenges requires dedicated conservation and breeding efforts to unlock the full economic potential of the species.

The general objective of this study was to provide scientific solutions for the sustainable management and efficient utilization of shea tree genetic resources in Côte d'Ivoire. Specifically, the study aimed to: (1) establish an *ex situ* core collection for conservation and breeding, (2) identify genetic markers associated with shea butter production, and (3) develop vegetative propagation techniques to support domestication.

To achieve these objectives, research was conducted on superior shea trees (SSTs) from the Bagoué, Hambol, Poro, and Tchologo regions, selected through participatory surveys to form an *in situ* collection. The study revealed significant conservation challenges: 8% of the identified superior trees were dead three years after the identification and 84% were heavily infested with parasitic plants of the Loranthaceae family, threatening the long-term viability of the collection. These findings underscore the need for a dual conservation strategy, integrating both *in situ* and *ex situ* approaches, to safeguard shea genetic resources. The establishment of an *ex situ* collection that captures high genetic diversity is critical for long-term conservation and breeding programs.

Prior to genetic analysis, a novel protocol was developed to extract high-quality, high-yield genomic DNA from both roots and leaves. DNA yield varied by tissue type, with higher concentrations obtained from leaves (5.17 µg) compared to roots (3.96 µg) per 100 mg of tissue. The extracted DNA was successfully used for downstream applications, including SSR-PCR, ISSR-PCR, RAPD-PCR, restriction enzyme digestion, and Sanger sequencing.

Genetic diversity was assessed using 12 morphological traits and 7,559 SNP markers on 220 and 333 SSTs, respectively. The study revealed substantial morphological variation (Shannon diversity index: 0.55–0.98) and moderate genetic diversity (HE = 0.26; PIC = 0.24). Both morphological and molecular analyses clustered the population into three distinct genetic groups. AMOVA based on SNP markers indicated that 100% of the variation occurred within individuals, with a low fixation index (Fst = 0.004), suggesting minimal genetic differentiation. This genetic diversity facilitated the selection of a core collection of 100 trees (30% of the original population) while maintaining full allelic and morphological variation, ensuring its utility for long-term conservation and breeding initiatives.

Furthermore, six multi-locus genome-wide association study (GWAS) methods identified 25 quantitative trait nucleotides (QTNs) associated with shea butter traits and 24 candidate genes involved in fatty acid biosynthesis. Oleic and stearic acids were the predominant fatty acids (85–90%), with regional variations influencing shea butter quality.

In the context of vegetative propagation, in vitro culture successfully induced axillary shoot regeneration using MS/2 medium supplemented with 3/1.2/1 mg/L BAP/Kin/NAA, while 100% of the regenerated shoots developed roots on MS1B/2 medium containing 3/0.1/40 mg/L IBA/mT/putrescine. Genetic fidelity of regenerated plants was confirmed using ISSR markers, ensuring clonal integrity.

This study significantly enhances the understanding of shea tree genetic diversity and conservation in Côte d'Ivoire. It provides practical strategies for sustainable management and breeding, supporting local livelihoods while contributing to the global shea butter industry.

Keywords: Shea tree, Genetic resources conservation, Morphological variability, DNA extraction, Single-Nucleotide Polymorphism, Genetic diversity, Population structure, Core collection, Shea butter, Fatty acid composition, GWAS, QTNs, Candidate gene, *In vitro* regeneration.

Résumé

Le karité (*Vitellaria paradoxa*) est une espèce indigène africaine essentielle, présentant une importance économique, sociale et environnementale significative. Malgré la valeur économique du beurre de karité en tant que produit à multiples usages, les efforts de recherche consacrés à l'amélioration génétique du karité restent limités. De plus, l'espèce est confrontée à de multiples menaces dues aux pressions biotiques et abiotiques, compromettant à la fois ses populations naturelles et l'approvisionnement durable en beurre de karité. Faire face à ces défis nécessite des efforts ciblés de conservation et d'amélioration génétique afin d'exploiter pleinement le potentiel économique de l'espèce.

L'objectif général de cette étude était de fournir des solutions scientifiques pour la gestion durable et l'utilisation rationnelle des ressources génétiques du karité en Côte d'Ivoire. Plus précisément, l'étude visait à : (1) établir une collection de référence *ex situ* pour la conservation et l'amélioration génétique, (2) identifier des marqueurs génétiques associés à la production de beurre de karité, et (3) développer des techniques de propagation végétative pour soutenir la domestication.

Pour atteindre ces objectifs, des recherches ont été menées sur des arbres de karité supérieurs (*Superior Shea Trees*, SSTs) issus des régions de Bagoué, Hambol, Poro et Tchologo, sélectionnés à travers des enquêtes participatives pour constituer une collection *in situ*. L'étude a révélé des défis majeurs en matière de conservation : 8 % des arbres supérieurs identifiés étaient morts trois ans après l'identification et 84 % étaient fortement infestés par des plantes parasites de la famille des Loranthaceae, mettant en péril la viabilité à long terme de la collection. Ces résultats soulignent la nécessité d'une stratégie de double conservation, intégrant à la fois des approches *in situ* et *ex situ*, pour préserver les ressources génétiques du karité. L'établissement d'une collection *ex situ* capturant une diversité génétique élevée est crucial pour la conservation et les programmes d'amélioration à long terme.

Avant l'analyse génétique, un protocole innovant a été développé pour l'extraction d'ADN génomique de haute qualité et à haut rendement à partir de racines et de feuilles. Le rendement en ADN variait selon le type de tissu, avec des concentrations plus élevées dans les feuilles (5,17 µg) que dans les racines (3,96 µg) pour 100 mg de tissu. L'ADN extrait a été utilisé avec succès pour des applications en aval, notamment le SSR-PCR, ISSR-PCR, RAPD-PCR, la digestion enzymatique et le séquençage Sanger.

L'analyse de la diversité génétique, basée sur 12 caractères morphologiques et 7 559 marqueurs SNP chez respectivement 220 et 333 SSTs, a révélé une variation morphologique importante (indice de diversité de Shannon : 0,55–0,98) ainsi qu'une diversité génétique modérée (HE = 0,26 ; PIC = 0,24). Les analyses morphologiques et moléculaires ont permis de regrouper la population en trois groupes génétiques distincts. L'AMOVA, basée sur les marqueurs SNP, a montré que 100 % de la variation était intra-individuelle, avec un indice de fixation faible (Fst = 0,004), indiquant une différenciation génétique minimale. Cette diversité génétique a facilité la sélection d'une collection de base de 100 arbres (30 % de la population initiale),

capturant l'ensemble de la variation allélique et morphologique, assurant ainsi son utilité pour la conservation et l'amélioration à long terme.

Par ailleurs, six méthodes d'association pangénomique multi-locus (*GWAS*) ont permis d'identifier 25 QTNs (quantitative trait nucleotides) associés aux caractéristiques du beurre de karité et 24 gènes candidats impliqués dans la biosynthèse des acides gras. L'acide oléique et l'acide stéarique étaient les acides gras dominants (85–90 %), avec des variations régionales influençant la qualité du beurre de karité.

Dans le cadre de la propagation végétative, la culture in vitro a permis l'induction de pousses axillaires sur un milieu MS/2 supplémenté avec 3/1,2/1 mg/L de BAP/Kin/NAA, tandis que 100 % des pousses régénérées ont développé des racines sur un milieu d'enracinement MS1B/2 contenant 3/0,1/40 mg/L d'IBA/mT/putrescine. La fidélité génétique des plants régénérés a été confirmée à l'aide de marqueurs ISSR, garantissant leur intégrité clonale.

Cette étude apporte une contribution significative à la compréhension de la diversité génétique et de la conservation du karité en Côte d'Ivoire. Elle propose des stratégies pratiques pour une gestion durable et une amélioration génétique, soutenant ainsi les moyens de subsistance locaux tout en contribuant à l'industrie mondiale du beurre de karité.

Mots clés : Karité, Conservation des ressources génétiques, Variabilité morphologique, Extraction d'ADN, Polymorphisme nucléotidique simple (SNP), Diversité génétique, Structure de la population, Collection de base, Beurre de karité, Composition en acides gras, GWAS, QTNs, Gène candidat, Culture *in vitro*

Acknolwedgements

My sincere thanks to the Direction des bourses hors Côte d'Ivoire for granting me the subsidy to complete this thesis. I would also like to thank the Universities of Liège and Péléforo Gon Coulibaly of Korhogo for the cooperation that enabled us to pursue our studies at the Faculty of Gembloux Agro Bio-Tech, and for the scientific mission grants of the University of Liège that enabled us to complete this thesis.

At the end of this research project, we would like to express our sincere thanks to all those who helped us in any way to carry out this study. In particular, we would like to thank.

I would like to express my heartfelt gratitude to my promoter, **Dr. Ludivine Lassois**, for your invaluable support throughout this scientific journey. Your intellectual rigor, availability, and insightful guidance have been indispensable to the success of this work.

I also want to highlight your dedication in securing funding, which enabled me to carry out the various activities essential to the progress of this thesis. Thanks to your efforts and determination, I was able to access the necessary resources to conduct my research, participate in training programs, and enhance my academic skills.

Your patience, kindness, and ability to guide me while valuing my ideas have made this experience particularly enriching, both scientifically and personally. You encouraged me during moments of doubt and celebrated every achievement, motivating me to consistently give my best.

Through your expertise, unwavering support, and tireless commitment, I was able to complete this thesis under the best possible conditions. During the years of my thesis, you proved that you embody the true meaning of the word “Promoter”. To me, you were like a good mother who always takes care of her children. For all of this, I am deeply grateful.

Prof. **Hervé Vandeschuren**, for welcoming me in the Plant Genetics laboratory that you head. Thank you for the lab meetings you organized where each researcher had to present his monthly results. This allowed me to structure my thesis objectives very quickly. The individual meetings we had with my promoter, particularly concerning *in vitro* culture, allowed me to work towards this aspect of my thesis.

Prof. Mongomaké Koné and Dr. Nakpalo Silué for assistance with the in vitro culture objective of this work.

I sincerely thank Prof. **Adama Coulibaly**, first President of the University Peleforo Gon Coulibaly (UPGC), for believing in us and facilitating the cooperation between the University of Liege and the UPGC. I also thank Prof. Diarrassouba Nafan, Dr. Silué Souleymane and **Dr. Alabi Taofic** for their valuable efforts for this cooperation.

I would like to thank ULiege members **Prof. Frédéric Francis** and **Prof. Eric Haubruge** for their support and efforts in this cooperation. Thank you so much!

I would like to extend my sincere gratitude to **Professors Sanbine Danthine** and **Maurie-Laure Fauconnier** for generously making their laboratories available to me.

Their support has been invaluable in facilitating my study of shea butter content and fatty acid composition.

To all my thesis committee and Jury members' : **Prof. Jean-Louis Doucet**, **Prof. Marie-Laure Fauconnier**, **Prof. Martine Schroyen**, and **Dr. Yordan Muhovski**, thank you for your collaboration and acceptance to review this thesis.

I sincerely thank **Prof. Enoch Achigan-Dako**, **Dr. Saraka Didier Martial Yao** and **Dr. Dougba Noel Dago** for their scientific assistance during this journey.

I am grateful to **Dr. Ekra Jean-Yves**, Mr. Blé Kpagni Antoine, Mr. Soro Nangalourou Adama, Diallo Rokia and all the ANADER workers and farmers who helped me during the field data collection and prospecting.

My thanks also goes to my classmates: Alfred Kouakou Kouassi, Bintou Traoré and Madinatou Zié, your support in everything motivated me to continue. You acted as mothers and fathers when I was ill, and as collaborators in sharing scientific ideas throughout this dissertation. Thank you very much guys.

A special thanks to the Team of Plant Genetics, particularly **Léonard, Sok Lay, Ariadna, Yvan, Kumar, Jimmy, Martin** and, **Quentin.** Besides being a united professional team, you also played the role of mentoring. Thank you so much for your patience and everyday guidance.

Finally, and most importantly, I would like to express my deepest gratitude to my family, led by our father, Ettien ATTIKORA, and our mother, Anouhoua AHISSAN, who have always supported us in the academic world. My brothers: Assomoi Marcel ATTIKORA and Ahissan Jean Constant ATTIKORA and my sisters: Tanoh Patricia ATTIKORA and Yah Rose ATTIKORA, with this thesis I honor your immeasurable sacrifices. I am immensely proud not only to have accomplished this great feat, but also of your unwavering faith in me. To my fiancée: Ama Estelle BEDIA and my children: Ettien David Eliam ATTIKORA and twins Affo Marie-Lyse ATTIKORA and Akoua Marie-Lyne ATTIKORA, thank you for supporting me and my long absence during this thesis. To my nephews: Laeticia, Lynda-Flore and Aldresse, and my nephews: Orthiniel, Trésor and Briyan, thank you for spicing up my Ph.D. journey in your unique ways, and as you grow up, remember that "the only impossible dreams are the ones you never pursue."

Table of contents

List of figures

List of tables

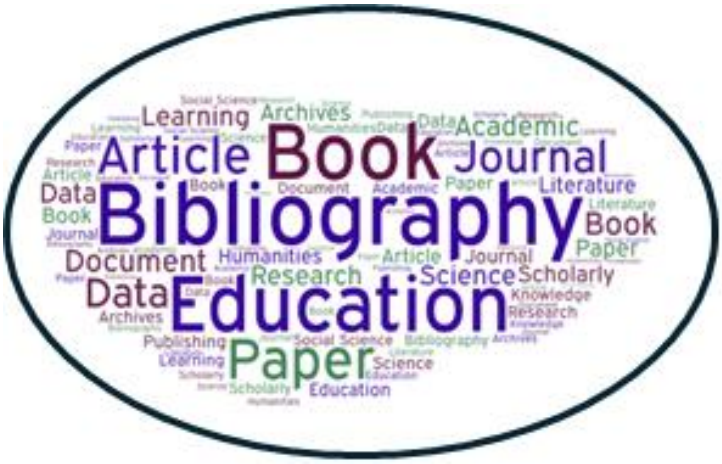
List of acronyms

AFLP	:	Amplified Fragment Length Polymorphism
ALC	:	Adult Leaf Color
AMOVA	:	Analysis Of Molecular Variance
BAP	:	6-Benzylaminopurine
C16:0	:	Palmitic Acid
C18:0	:	Stearic Acid
C18:1	:	Oleic Acid
C18:2	:	Linoleic Acid
DAPC	:	Discriminant Analysis of Principal Components
DNA	:	Deoxyribonucleic Acid
FA	:	Fatty Acid
FC	:	Fat Content
F_{ST}	:	Genetic Differentiation
GBS	:	Genotyping By-Sequencing
gDNA	:	genomic DNA
GWAS	:	Genome-Wide Association Study
HE	:	Expected Heterozygosity
HO	:	Observed Heterozygosity
I:	:	Shannon's Information Index of Diversity
IAA	:	3-Indoleacetic acid
IBA	:	Indole-3-butyric acid
ISSR	:	Inter-Simple Sequence Repeat
Kin	:	Kinetin (6-furfuryl aminopurine)
LAS	:	Leaf Apex Shape
LL	:	Limb Length
LW	:	Limb Width
MAS	:	Marker-Assisted Selection
mg/L	:	Milligram per Liter
MS	:	Murashige and Skoog
MS3B	:	Modified Murashige and Skoog
***m*T**	:	*meta*-Topolin
NAA	:	1-Napthaleneacetic acid
NL	:	Nut Length
NWD	:	Nut Width
NWG	:	Nut Weight
PCA	:	Principal Component Analysis
PCoA	:	Principal Coordinate Analysis
PCR	:	Polymerase Chain Reaction
PGR	:	Plant Growth Regulators
PIC	:	Polymorphism Information Content
PL	:	Petiole Length

PSA	:	Proportion Of Superior Allele
QTN	:	Quantitative Trait Nucleotide
RAPD	:	Random Amplified Polymorphic DNA
RFLP	:	Restriction Fragment Length Polymorphism
SB	:	Shea Butter
SEES	:	Seed Coat Color
SNP	:	Single Nucleotide Polymorphism
SSR	:	Simple Sequence Repeats
SST	:	Superior Shea Tree
TC	:	Trunk Circumference
TGH	:	Tree Growth Habit
HSD test	:	Tukey Honest Significant Difference test.

Chapter 1

General Introduction



Chapter 1. General Introduction

1. Context of the Study

The shea tree (*Vitellaria paradoxa*) is a plant species of significant economic, social, and ecological value, particularly in Côte d'Ivoire, where it serves as a key resource for local livelihoods, agricultural development, and environmental sustainability by supporting biodiversity, preventing soil erosion, and contributing to carbon sequestration.

Shea butter, extracted from the seeds of the shea tree, is a highly desirable product in the food, cosmetic, and pharmaceutical industries, making the species a cornerstone of rural economies in West Africa.

In Côte d'Ivoire, the University of Peleforo Gon Coulibaly (UPGC) has recognized the strategic importance of shea and other agricultural resources, such as mango, cashew, and poultry, in driving regional development. Through its "Center for the Valorization of Bioresources", the university, is working to enhance the utilization of these resources to boost the local economy.

In parallel, since 2017, the "African Center for Shea Research and Application" (CRAK) at UPGC has been actively involved in an extensive program dedicated to the conservation and genetic improvement of the shea tree. This program aims to ensure the sustainable production of shea butter while maintaining the genetic integrity of *Vitellaria paradoxa* in Côte d'Ivoire.

As part of this initiative, 1250 Superior Shea Trees (SSTs) have been identified through a participatory survey on farmers' lands, based on the production capacities of trees recognized by farmers. This survey took place during 2017 in the Bagoué and Tchologo regions, and then during 2020 in the regions of Poro and Hambol. These SSTs constitute an *in situ* collection of genetically superior shea trees across the country and represent an invaluable genetic resource with potential for both conservation and future genetic improvement programs.

Although significant progress has been made, further efforts and research are needed to improve shea tree conservation and genetic enhancement for high-quality butter production. This Ph.D. research contributes to these efforts and results obtained will play a pivotal role in shaping the future of shea tree conservation and genetic improvement in Côte d'Ivoire and beyond.

2. Background

Plants provide humans with a range of essential resources, including oxygen, food, medicines, and materials for the production of shelter and clothing. Indeed, the conservation and genetic improvement of plant species are central to ensure global food security, biodiversity preservation, and sustainable development. The loss of genetic diversity due to habitat destruction, climate change, and overexploitation has increased the need for effective conservation strategies. The accelerated decline in plant biodiversity has emerged as a major global concern. According to the International Union for Conservation of Nature (IUCN), approximately 40% of the world's plant species are currently threatened with extinction. Globally, numerous programs aim to safeguard genetic resources through *in situ* and *ex situ* methods, such as seed banks, field gene banks, biotechnology tools, and participatory management practices (1). These approaches are essential for maintaining genetic variation, which is the basis for improving traits such as yield, stress tolerance, and quality in both wild and domesticated plant species (2). Traditional tree breeding methods rely on selective breeding, hybridization, and provenance trials, while modern approaches integrate molecular markers, genome-wide association studies (GWAS), and marker-assisted selection (MAS) to accelerate genetic gains (3,4).

Tree species, particularly those in tropical and subtropical regions, face unique challenges in conservation and genetic improvement (5). Unlike annual crops, many tree species exhibit long juvenile phases, high levels of heterozygosity, and dependence on natural regeneration, making large-scale breeding programs more complex (3). Agroforestry systems have emerged as a key solution for conserving tree species in multifunctional landscapes, integrating economic production with ecological benefits such as soil fertility, carbon sequestration, and biodiversity preservation (6).

One prominent example of a tree species with high economic and ecological value is the shea tree (*Vitellaria paradoxa*). Native to the Sudano-Sahelian zone, it extends across 21 countries from Senegal to Uganda (7–10) and is an important economic and ecological component of agroforestry systems. *Vitellaria paradoxa* belongs to the Sapotaceae family and has two subspecies: *paradoxa* in West Africa and *nilotica* in East Africa (11,12). It reproduces predominantly through insect-mediated cross-pollination but can also self-pollinate (13). The fruit is primarily dispersed by gravity (barochory) with secondary dispersal by animals, including elephants, birds, rodents, and humans (14,15).

Shea production is a major economic activity in West Africa, with an estimated annual nut yield of 1,760,600 metric tons. Nigeria and Mali account for 70% of total production, while Côte d'Ivoire ranks fifth with about 250,000 metric tons annually (16–18). The industry is predominantly run by women, with about 3 million in Africa and 152,000 in Côte d'Ivoire involved in collection and processing (18). Shea butter, extracted from shea kernels, is a major export commodity used extensively in food,

cosmetics, and pharmaceuticals (19). Its high concentration of stearic, oleic, palmitic, and linoleic acids makes it a valuable raw material in the food industry. In addition, its significant unsaponifiable content contributes to its use in luxury cosmetics and pharmaceuticals, making shea a multi-million dollar export product (20–22).

In Côte d'Ivoire, shea trees are mainly found in the northern savannahs, often coexisting with other large tree species (23,24). Despite their economic and ecological importance, shea tree populations face severe threats, including deforestation, charcoal production, uncontrolled bushfires, and agricultural expansion (25). Human activities, such as shorter fallow periods and systematic fruit collection, further limit natural regeneration (11). The increasing conversion to more profitable crops, such as cashew exacerbates genetic erosion in shea populations. The species is classified as vulnerable by the International Union for Conservation of Nature (26), highlighting the urgent need for conservation and genetic improvement strategies.

Unlike many commercial crops, *V. paradoxa* is not traditionally cultivated due to its long juvenile phase (10–25 years) (27). Instead, its presence in agroforestry systems results from natural propagation and systematic farmer management. Through selective protection of high-yielding trees, farmers have contributed to a form of partial domestication (12,28). To transition from wild harvesting to organized plantations, research has focused on shea domestication through vegetative propagation techniques, such as grafting (29–32), cuttings (33–37), air-layering (34), and in vitro culture (38–44). Other studies have investigated shea tree parasitism (45–49), genetic diversity (14,24,50–54), fruiting patterns (55–58), biosynthesis pathways of shea butter (25,59), and superior genotype selection (60–62). Advances in genomics, including reference genome sequencing (25) and genome-wide association studies (GWAS) (63), have paved the way for molecular breeding.

In situ conservation of shea trees has traditionally been driven by local knowledge, where farmers select and protect superior phenotypes based on traits such as fruit size, oil yield, and quality. These practices have preserved genetic resources while supporting sustainable agroforestry systems. The initiatives of researchers to identify and preserve superior phenotypes of shea resources have been done (60–62). However, the introduction of modern scientific tools, such as molecular markers (e.g. single nucleotide polymorphism), and genomic approaches, provides new opportunities for designing effective conservation strategies and accelerating genetic improvement (25). Recent advancements in genome-wide association studies (GWAS) and molecular breeding have enabled the identification of genetic loci associated with economically important traits, paving the way for marker-assisted selection (25,54,59,63).

Thus, the conservation and genetic improvement of the shea tree represent a critical intersection of traditional knowledge and modern science. By integrating participatory approaches with advanced genomic tools, it is possible to enhance the sustainability of shea-based agroforestry systems, address the challenges of genetic erosion, and unlock the economic potential of this vital species.

3. Importance of Shea Tree

3.1. Economic Potential of Shea Tree in Côte d'Ivoire

Shea butter is a multi-million dollar commodity valued for its wide range of applications in the food, cosmetics, and pharmaceutical industries. It is the second most important oil crop in Africa after oil palm (19). In Côte d'Ivoire, shea contributes significantly to the economy through both local use and export. Global demand for shea butter continues to grow, with West Africa accounting for approximately 90% of the world's production. Côte d'Ivoire plays a crucial role in this market, contributing both raw kernels and processed butter to meet export demand (18,64,65). The country has an additional advantage due to its geographical and ecological conditions, supporting high-quality shea production.

Shea collection and processing in Côte d'Ivoire, as in many other African countries, are predominantly managed by women (65). Women are the main actors in the value chain, from collecting the nuts to processing and trading the butter (65). This provides an important source of income and increases their economic autonomy (9,66,67).

The shea industry also promotes social and economic empowerment by creating opportunities for women to form cooperatives. These groups enable women to pool resources, negotiate better prices, and access training or funding for processing technologies. However, gender disparities remain significant. Women often face challenges in accessing land, as traditional land tenure systems in Côte d'Ivoire frequently limit their ownership or control of shea parklands (65,68–72). As a result, women are highly dependent on communal lands for shea nut collection, which can lead to conflicts over access to resources.

Woman's right to harvest shea nuts varies considerably across West Africa (72). In Ghana, a woman collects nuts from her husband's plots, while elsewhere women collect shea from trees in fallow fields (66). In Mali, rights to shea nuts in cultivated fields are granted to any woman, regardless of land claims (73).Throughout the region, access to shea trees in uncultivated fields is usually open to all women (66,74). Nut collectors who do not have access rights to trees on cultivated fields use open access uncultivated areas for collection, further accentuating the differences in nut and oil quality. In addition to tenure arrangements that regulate access to fruit, shea nut collection is mediated by socio-cultural conventions and local institutions (66,69).

Land ownership plays a critical role in the sustainability and productivity of shea trees. In Côte d'Ivoire, customary land tenure systems dominate, and men often hold primary ownership rights. This limits women’s ability to make decisions about tree management and replanting, even though they are the primary beneficiaries of the shea economy. Furthermore, the lack of secure land tenure discourages long-term conservation efforts, as tree planting and maintaining trees require a significant investment of time (67).

The expansion of agricultural land for cash crops, such as cashew and cotton, poses additional challenges for shea conservation. Farmers often prioritize short-term financial gains from cash crops over the long-term benefits of maintaining shea trees,

contributing to deforestation and declining shea tree populations in some regions (25,67,75).

To fully realize the economic potential of shea trees in Côte d'Ivoire, there is a need for policies that address gender and land ownership challenges. Land reforms that promote women's ownership or secure access to shea parklands can significantly increase their economic contribution and provide incentives for sustainable management practices. In addition, investments in processing infrastructure, training, and access to markets are crucial for increasing the value-added potential of shea butter, especially for women-led cooperatives. Moreover, the processing of shea fruit pulp into sirop, confiture and alcohol has great local economic potential to alleviate poverty in shea-producing areas.

Additionally, promoting agroforestry systems that integrate shea trees with other crops can balance economic productivity with environmental conservation. Public-private partnerships and international cooperation can play a key role in advancing these initiatives, while ensuring that the benefits are shared equitably.

The shea tree is not only an economic asset, but also a vehicle for promoting gender equality and sustainable development in Côte d'Ivoire. By addressing land ownership challenges and supporting women in the shea value chain, the country can unlock the full potential of this resource. Strengthening the shea industry has the potential to transform rural livelihoods and contribute significantly to both local and national economies.

3.2. Importance of Shea tree in the Context of Agroforestry

Shea is considered to be one of the most important tree species in agroforestry parks in sub-Saharan Africa, where the species occurs naturally (57). It has a high percentage of standing biomass and contributes to the reduction of soil degradation. It has a high carbon sequestration capacity and can be used in climate change mitigation strategies (76–79).

The microclimate is positively affected by the presence of shea trees in agroforestry plantations. The shade provided by the trees has a significant effect on soil moisture content. This increases crop yields (80). Higher shea fruit yields have been observed in the field stand compare to the fallow stand (57,58). In Burkina Faso, for example, 51% of farmland has been identified as suitable for growing shea tree and is an important species for soil restoration (81). The cultivation of crops such as maize and soya has been used to generate income and improve food security in shea parks (82). According to the authors, different yield responses were observed for soya and maize. Because of competition for light and nutrients, combined crops of mature shea gave very poor maize and soybean yields compared with combined crops of young shea (82). Another study was conducted on the influence of shea and *Parkia biglobosa* on sorghum production in Burkina Faso (83). The authors found that sorghum yield was negatively affected by *P. biglobosa* by 50-70%. Some studies have shown that as shea trees age, the yield of adjacent crops decreases. Yield variability can be explained by shading and competition for water and nutrients. However, soybean can be

recommended for cultivation in shea parks, as the yield reduction was less significant than for cereals (75,82,84).

4. Botanical description of Shea tree

The taxonomic classification of the shea tree (*Vitellaria paradoxa*) has been subject to debate. Hepper in 1962, proposed that Butyrospermum be retained as the genus name instead of Vitellaria, but this proposal was repeatedly rejected by votes (85). The nomenclature was later clarified and the original binomial name "*Vitellaria paradoxa*", first described by Carl Friedrich Gaertner in 1807, was retained (8).

The shea tree is currently classified in the plant kingdom as follows:

Domain: ……..Eukaryota
Kingdom: …….Plantae
Phylum: …….Spermatophyta
Subphylum: …….Angiospermae
Class: …….Dicotyledonae
Order: ……..Ebenales
Family: …….Sapotaceae
Genus: …….Vitellaria
Species: ……..*Vitellaria paradoxa*

In 1943, Chevalier described eight varieties based on leaf and fruit characteristics, but these classifications are rarely used (86). The shea tree is known by numerous indigenous names that vary among different ethnographic groups (8).

Vitellaria paradoxa is belong to the Sapotaceae family, which includes 53 genera of trees, shrubs and woody lianas (87). Most of them are from the tropical and subtropical regions of South America and Asia (8,11). Many species in this family have a sticky white latex found in the bark, branches, leaves and fruits. And like slow-growing species, they are found in arid areas (*Vitellaria paradoxa*) and rainforest (*Pouteria spp*) (88,89).

The genus *Vitellaria* has only one species, *V. paradoxa*. Two subspecies were identified: *V. paradoxa* subsp. *paradoxa* and *V. paradoxa* subsp. *nilotica*. Differences between the two subspecies are related to flower and fruit characteristics (90) and shea butter fatty acid composition (10,91). Subspecies *paradoxa* is found in 14 countries from Senegal to the Central African Republic, while subspecies *nilotica* is found in 7 countries from Tchad to Ethiopia (10). The ranges of *vitellaria* and *nilotica* do not overlap, but are within 175 km radius of the boundary between Lake Chad and River Congo to the west, and the Nile River to the east and northeast (14). According to the International Union for Conservation of Nature, this species is classified as vulnerable on the Red List of Threatened Species. The species has been overexploited for timber, firewood and charcoal production. Its habitat is also suffering from agricultural encroachment and increasing population pressure (26).

Shea trees vary in height from 7 to 25 m, and in diameter from 0.3 to 1 m (**Figure 1a**). Their bark has thick, longitudinal and deep striations (92–94). The leaves are in dense clusters and spirally arranged at the tips of the sturdy branches. Petiole length varies from 5 to 15 cm. Young leaves are rusty red and hairy, then become leathery, hairless, dark green, brilliant, 12 to 25 cm long and 4 to 7 cm wide (8). The tree bears multiple flowers in the axils of the terminal leaves or on the scars of leafless twigs. The number of flowers per inflorescence is highly variable and may exceed 100 (**Figure 1b**). The flower is hermaphrodite and actinomorphic, with a very small bract at the base of the stem (95,96). The flowers are borne on long pedicels (22-25 cm) and are creamy white, very fragrant and melliferous (90).

Shea flowers are generally heterostyly polygamous: the male flowers, characterized by short styles containing in the flower bud, fall shortly after flowering; the hermaphrodites, with styles protruding from the bud, are few in number (2 to 4 per umbel) and leave a young, initially pubescent fruit after the corolla falls (86). Flowers in the same umbel flower successively, with the males flowering first. Although cross-fertilization predominates, self-fertilization is still possible because anther opening occurs late in relation to ovary maturity, in accordance with the anatomy of the flower. Fertilization is generally cross-fertilization, but it is also possible for the male flowers of an umbel to fertilize the long-stemmed flowers of the same umbel at sexual maturity (86). Experiments showed a fruit set rate by self-fertilization of 9.6% compared with 23% for the control flowers which were not bagged (86,97,98).

Insects, mainly bees, are the main pollinators for shea. Although wind may play a role, it has been observed that few insects visit shea flowers during the day, probably because of the strong wind, particularly the north-easterly wind, which blows at flowering time (86). It is likely that most insects visit the shea flowers at night and early in the morning. In the Sudanese zone, each shea tree generally carries one or more beehives on its branches.

The shea fruit can be described as a spherical to elliptical berry measuring 4-5 cm × 2.5-5 cm and weighing 20-30g (99) **(Figure 1c**). The fruit is a yellow, green or yellowish berry with thick flesh containing an oval or round reddish-brown seed. The fruit is initially green as it develops, but takes on a yellowish-green to brown color as it ripens. The shell of the seed is relatively thin, shiny, with large scars; the seed consists of two thick, fleshy, tightly fused cotyledons, with no apparent radicle (90).

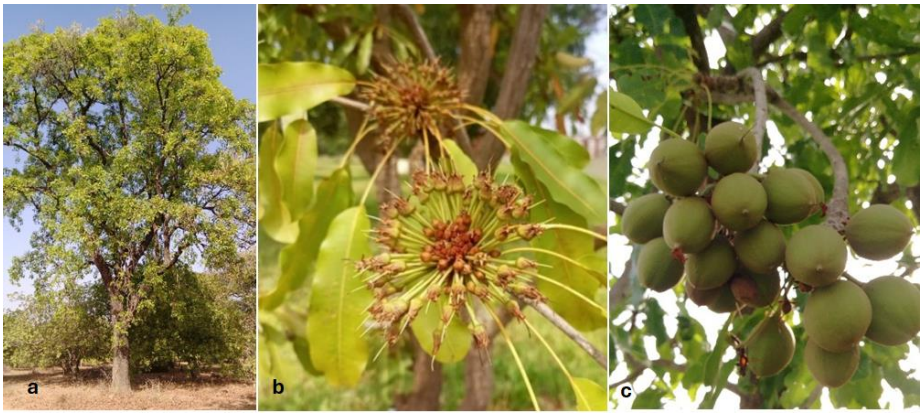


Figure 1. Presentation of shea tree. (a) a shea tree plant in cashew plantation; (b) shea inflorescence; (c) shea fruits

5. Ecology of Shea tree

Vitellaria paradoxa is native to the West African savannah, but is also found in the southern Sahel. It grows in a wide belt of approximately 5,000 kilometers from west to east and 500 kilometers from north to south, across 21 different countries (7–10) as shown in **Figure 2**. *V. paradoxa* occurs from Senegal in the west to the foothills of the Ethiopian highlands (13,100). The genus *Vitellaria* consists of a single species belonging to the Sapotaceae family, and has two described subspecies: subspecies *nilotica*, found mainly in East Africa, subspecies *paradoxa*, found in West Africa from Senegal to the Central African Republic (11,12). Shea is generally an inland tree, except in Ghana and Nigeria where it grows within 50 km of the coast (8). The ecology of shea is influenced by several environmental factors. Shea is commonly found on flat land. It is rarely found on slopes and is almost absent from valleys. Distribution, density and trunk circumference vary depending on where the shea trees grow. It grows at annual temperatures of 25-29°C, and at altitudes of 100 to 600 m (95). The presence of shea trees was noted in the mountainous region of western Cameroon at 1300 m (101). Shea also grows in areas with annual rainfall of 600 to 1400 mm and a dry season of 5 to 8 months (95). Edaphically, shea grows on clay, sandy-clay, stony and lateritic soils. The tree is adapted to iron-rich sandy soils with a good humus content for optimal growth and development. It prefers soils with a neutral pH, but can also tolerate acidic soils (92). Shea trees require good drainage and do not tolerate high soil water content, although they adapt fairly well to site hydrology. The roots of *V. paradoxa* are extensive and can tolerate prolonged periods of drought (8,102). Phytosociologically, shea is abundant in wooded savannah and agroforestry parklands (12,103). Shea may be sensitive to interspecific competition. It is a species that prefers light, open and airy sites (104). However, it is frequently found in association with several other woody species (27,105).

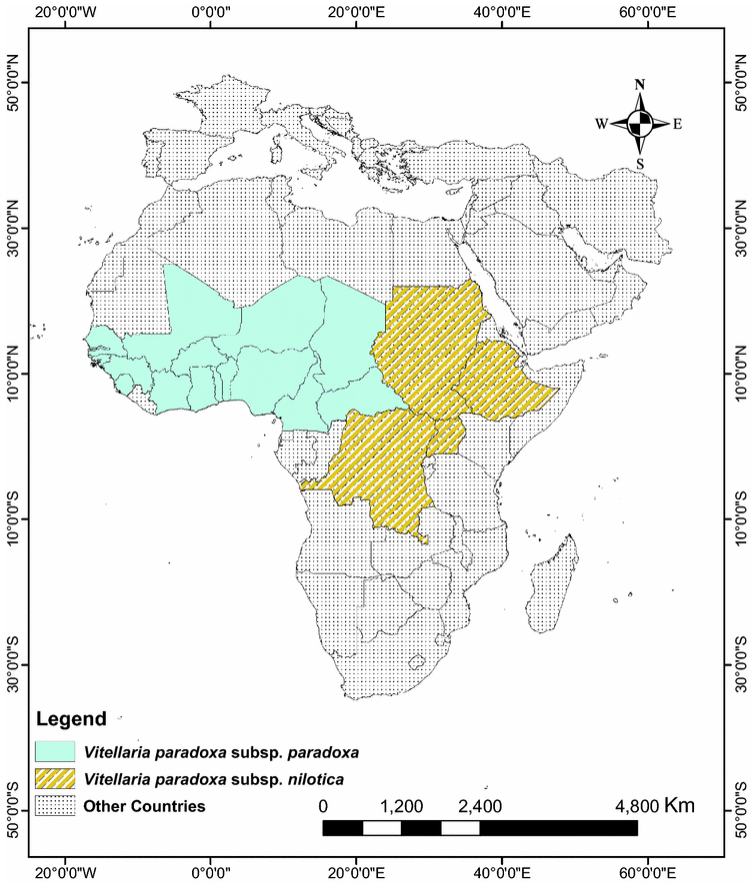


Figure 2. Shea distribution map showing the occurrence of the two subspecies (100).

6. Phenology of Shea

The shea tree has a particularly long development cycle, with the reproductive phase typically beginning at 15–20 years of age under optimal conditions, while economically viable production is usually achieved around 40 years of age (106).

Shea trees are deciduous, losing their leaves during the dry season between November and March. Defoliated trees flower from December to April, depending on the region (55), with new leaves appearing immediately after the flowering period. It should be noted that fertility is correlated with the rate of flowering. The seasonal periodicity of phenology varies considerably, with successive stages beginning in the northern and western regions and extending eastwards. The fruit begins to ripen between March and August and is harvested mainly between April and September. In Côte d'Ivoire, flowering begins in November and ends in May. The fruiting period begins in December and ends in July, and the harvest takes place from June to August, depending on the region (27).

7. Morphological and Genetic Diversity of Shea Tree

Genetic diversity plays an important role in the domestication and conservation of plant species. Morphological, biochemical and molecular markers are used to study the genetic diversity of plants. Morphological analysis reveals diversity as perceived and selected by local farmers, who are the main actors in the management of varietal diversity (107). Molecular markers, especially microsatellite markers, are used to understand the spatial structuring of diversity and to construct hypotheses about the different events of plant spread. Genetic markers are used as tools to characterize the diversity of plant genetic resources and the functional variants of relevant genes. Genetic diversity is thus a selective resource for improving tree species and contributes to genetic gain for many traits (108–110).

The current diversity and distribution of shea trees reflect a long history of evolution and responses to climate change and other environmental factors that determine the morphology of its organs such as the stem, leaf, flower, fruit and seed (14).

In Côte d'Ivoire, as in many producing countries, shea tree genetic resources are conserved and maintained in the field by farmers. The role of genetic variation studies in supporting crop conservation approaches and variety improvement has long been recognized (111). High genetic diversity is essential for the potential to adapt to current and future biotic and abiotic constraints in the context of global change (112).

7.1. Morphological markers

The expression of morphological traits in shea trees (*Vitellaria paradoxa*) is influenced by environmental factors, yet it remains a crucial tool for plant evaluation (113). Indeed, discriminative morphological descriptors have been applied in the regeneration of plant collections (114).

Over time, studies have identified various descriptors, including fruit, seed, leaf, stem, and root traits, to capture the species' morphological diversity (94,104). Notably, the IPGRI/INIA developed a comprehensive list of descriptors in 2006, including fruit characteristics and specific leaf features such as shape, apex, margin, and veins, to differentiate individuals within shea populations (27).

Local populations classified shea tree varieties based on morphological traits such as fruit, nut variants. For instance, 25 shea ethno-varieties were identified based on 11 primary fruit and nut traits in Burkina Faso (115), while 28 shea ethno-varieties were identified in Uganda based on fruit/nut organoleptic and morphological attributes (116). In Chad, 6 shea morphotypes have been described based on morphological characteristics related to fruit shape, and these have been found to be the most discriminating (117). Morphological diversity has been mentioned with the agro-ecological locations of shea tree. Indeed, a study carried out in Cameroon revealed that lower-altitude trees had larger trunks and smaller fruits compared to high-altitude trees, showing a correlation between morphological traits and agro-ecological regions (118). In Ghana, trees on farmland displayed greater height than those in forests, likely

due to protection from environmental pressures (53). Similarly, in Mali shea trees in farmland (shea parks) showed higher shea flower and fruit densities than in forest.

In Côte d'Ivoire, studies indicated significant diversity in quantitative traits such as trunk circumference and fruit, nut and leaf traits. Qualitative trait analysis further identified variability in traits like crown shape, fruit shape, nut color, and branch orientation (27,94,99,104,119).

Studies have consistently shown that morphological diversity in shea trees, often categorized into distinct morphotypes, is influenced by vegetative form and environmental context.

7.2. Molecular diversity of *V. paradoxa*

The molecular diversity of trees is studied using molecular markers located on the DNA. A molecular marker is a polymorphic locus that provides information about the genotype of the individual carrying it (120). Molecular markers facilitate direct analysis of the polymorphism of DNA sequences in order to construct a genetic map. Several molecular markers are used to identify the genotype in plants. The most commonly used are RAPD, AFLP, SSR, RFLP, SNP markers (121). Research on the molecular diversity of the shea tree has focused on four main types of molecular markers: isoenzyme, RAPD, SSRs and SNPs (122).

Microsatellite DNA has been widely used in population genetics because of its ability to detect differences between closely related species and within the same species (100). In shea, these markers have been used to distinguish the two subspecies of the species. Ten microsatellite loci were used to investigate the effect of anthropogenic activity on the spatial and temporal genetic structure of the shea populations. As results, low genetic differentiation was observed, possibly due to the buffering effect of high gene flow between managed and unmanaged populations (123). Microsatellites have also been used to illustrate diversity within shea tree populations (24,124); to assess the effects of land use and agro-ecozone types (53); and to differentiate ethno-varieties (52). Random amplified polymorphic DNA (RAPD) and chloroplast marker analysis of several populations distributed in West Africa and Central and East Africa suggested a genetic distinction between West and Central African and East African shea populations (14,50,125). The recognition of the existence of distinct gene pools between the two subspecies present in different regions implies that they may adapt differently to future climate change. The diversity of the shea gene pool within and between countries, and even regional populations, calls for organized and concerted collection and evaluation of superior germplasm based on previously separate independent research across the region (67).

Genome-wide analysis of the genetic diversity of *V. paradoxa* plays a crucial role in the species conservation and improving accelerated breeding methods, especially marker-assisted selection (MAS). Previous research efforts have led to the sequencing of the complete reference genome of *V. paradoxa*. This reference genome, consisting of 12 chromosomes with a total size of approximately 656.7 megabases (Mb), has been annotated to contain 38,505 coding genes (25).

Analysis of the evolutionary history using this reference genome revealed clear evidence of two ancient genome-wide duplication events in the shea tree evolutionary lineage (25). These events occurred at different times, one prior to the Astrid-Rosid divergence, estimated to be between 116 and 126 million years ago (Mya), and the other at the origin of the Ericales order, estimated to be between 65 and 90 Mya (25).

The availability of the reference genome has led to new studies based on new markers such as SNPs. For example, studies on the genetic diversity and population structure of the shea tree using SNPs have been reported (54,122,126), and the genome-wide association study on fat content was studied in Uganda (63). Molecular marker based on SNP has been used to study the genetic diversity of the superior shea trees in Côte d'Ivoire in order to establish a core collection for ex-situ conservation of shea genetic resources (122).

8. Shea Butter Characteristics

8.1. Chemical composition of shea butter

The oil or fat content of shea kernels is generally high with good post-harvest processing. This content averages between 45-55%, although oil contents of less than 25% to more than 60% can be obtained (127) **(Figure 3**).

There are different techniques for extracting shea butter: traditional method (churning), semi-mechanized method (press, motorized grinder, centrifugation system), solvent method, and other modern techniques (microwave assisted, enzymatic, supercritical CO_2). Afterwards, a comparison between the different extraction methods showed that microwave-assisted extraction gives a higher extraction yield (88%), followed by screw press extraction (82.20%), solvent extraction yield (66.9%), enzymatic (43%), supercritical CO_2 (39.5%), the mechanical method (37%), and finally the traditional process with 28% as yield (128). The different shea butters evaluated by Fourier Transform Infrared Spectroscopy (FTIR) did not show any significant differences according to the extraction methods (128).

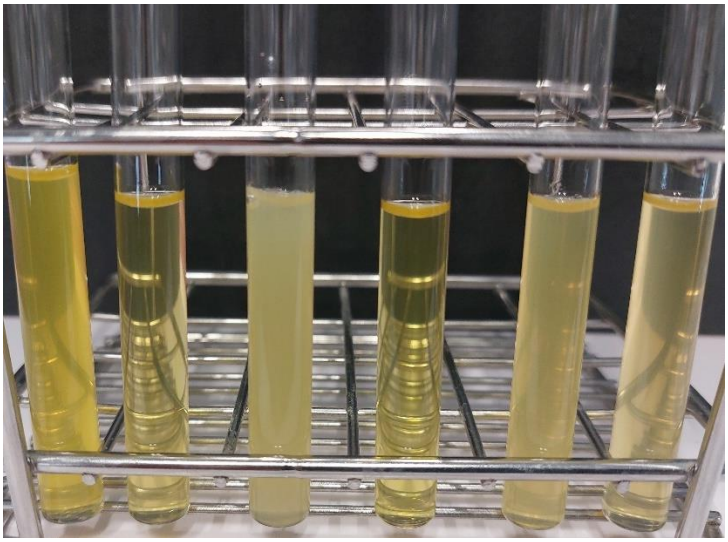


Figure 3. Shea oil newly extracted with maceration method. Each tube represents a different genotypes with different fatty acid composition.

The four main fatty acids in shea are oleic acid (34-62%), stearic acid (25-56%), palmitic acid (2-10%) and linoleic acid (1-10%) (129–131). As for arachidic acid, it represents a small percentage (1-2%) and traces of other fatty acids such as gadoleic, behenic and lignoceric acids are present in shea butter. The composition of stearic and oleic acids represents 85-90% of the total fatty acids (132). The ratio of oleic acid to stearic acid in shea butter extracted from the kernels of subspecies *paradoxa* is generally between 1 and 0.8-1.1 and between 1.5-1.9 for subspecies nilotica. Higher percentages of oleic acid have been recorded at the extremes of the range of *paradoxa* subspecies (Gambia and eastern Nigeria) and at the extreme south-eastern end of the range of *nilotica*, suggesting recent colonization and a role for the fatty acid profile in the viability of recalcitrant seeds (127).

Shea butter is composed of three groups of triglycerides. These are polyunsaturated, di-saturated, and monounsaturated. The major polyunsaturated triglyceride in shea butter is oleic-oleic-oleic (O-O-O), while the major di- and monounsaturated triglycerides are stearic-oleic-oleic (St-O-O), stearic-oleic-stearic (St-O-St), respectively.

The high content of symmetrical di-saturated TAGs in West African shea butter has a significant impact on the melting profile of shea butter from both subspecies (9). While it is possible to modify the composition and melt profile through fractionation or hydrogenation techniques, this natural variation has many implications for the range of applications and processing economics most appropriate for shea butter from different geographical areas. Another unique feature of shea butter is its higher unsaponifiable content than most other vegetable oils (128). Its average unsaponifiable content is 5-7%, although extreme values of over 19% from individual tree seeds have been recorded (127). Analysis of these unsaponifiables has revealed a wide range of interesting and useful bioactive compounds with antioxidant, anti-

inflammatory, anti-tumor, UV-protective and protease-inhibiting properties, including phytosterols, triterpenes, polyisoprenic hydrocarbons (karitene), tocopherols and catechins, with dried almonds containing up to 0.4% of the latter (12,132,133).

Shea butter is the most important product of the species. Selecting genotypes with high-fat and fatty acid profile is crucial for enhancing shea butter quantity and production. Several methods for genome-wide association study (GWAS) based on oil content have been used to identify candidate genes involved in oil production. These methods can be used to identify shea genes associated with shea butter production within a population.

8.2. Physicochemical Quality and Classification of Unrefined Shea Butter

Shea butter, derived from the kernels of *Vitellaria paradoxa*, is a valuable vegetable fat widely utilized in food, cosmetics, and pharmaceutical industries. Unlike many other plant-based oils, shea butter remains solid at room temperature, primarily due to its high stearic acid content, which contributes to its unique texture and thermal stability (134). The physicochemical properties of unrefined shea butter are critical determinants of its quality, influencing both its nutritional value and industrial applications (135,136).

The classification of unrefined shea butter into different quality grades is based on key chemical parameters that reflect its freshness, purity, and suitability for consumption or processing (135). These parameters include moisture content, free fatty acid (FFA) levels, peroxide value, and the concentration of insoluble impurities. Collectively, they provide insights into the butter's stability, susceptibility to rancidity, and the efficacy of post-harvest processing methods (137,138).

Based on these criteria, unrefined shea butter is categorized into two primary grades as presented in **Table 1**.

Table 1. Quality criteria of unrefined shea butter according to Codex Alimentarius

Characteristics	**Unrefined shea butter**	
	Grade 1[a]	Grade 1[b]
Water content (%)	≤ 0.05	0.06 – 0.2
Free fatty acids (%)	≤ 1	1.1 – 3
Peroxide value (mEqO2/kg)	≤ 10	11 – 15
Insoluble impurities (% m/m)	≤ 0.09	0.1 – 0.2

These quality indicators are primarily influenced by factors such as extraction techniques, packaging materials, and storage conditions rather than the genetic variability of the shea tree. Poor handling practices during kernel processing, exposure to moisture, and prolonged storage under suboptimal conditions can significantly increase FFA and peroxide values, leading to reduced shelf life and rancidity (127,128,130,135,139).

Beyond chemical composition, the physicochemical properties of shea butter such as color, texture, and odor, play a pivotal role in consumer preference and industrial applications. These attributes are influenced by both genetic and environmental factors, as well as post-harvest processing techniques.

- **Color:** Unrefined shea butter exhibits a range of colors, including black, yellow, and beige. The color variation is closely linked to the treatment of the nuts prior to extraction and, to a lesser extent, genetic differences among shea tree populations. For example, the roasting intensity of kernels during traditional processing can lead to darker butter, while minimal heat application results in lighter shades (140).

- **Texture:** The texture of shea butter can range from fondant (smooth and creamy) to compact (firm and crumbly). This variation is often correlated with its color and fatty acid composition, particularly the ratio of stearic to oleic acids. Higher stearic acid content tends to produce a firmer butter, while increased oleic acid levels contribute to a softer, more pliable texture (132,135,139).

- **Odor:** The intensity of the characteristic shea butter odor, ranging from strong to mild, is influenced by storage conditions and processing methods. Improper drying of kernels, inadequate storage, and prolonged exposure to high humidity can result in strong, sometimes rancid odors due to lipid oxidation. Conversely, well-processed and properly stored butter retains a mild, pleasant nutty aroma (139).

The quality of unrefined shea butter is determined by a combination of chemical and physicochemical parameters that reflect both intrinsic properties and external handling conditions. While genetic factors influence certain attributes like color and texture, the primary quality indicators such as moisture content, FFA, peroxide value, and impurities are largely shaped by post-harvest practices. Enhancing the quality of shea butter requires not only the selection of superior shea tree genotypes but also improvements in processing techniques, storage infrastructure, and quality control protocols.

9. Threats, Preservation and Restoration of Shea Parklands

Shea trees dominate woody vegetation in agroforestry parks across several countries where they grow (56,105). The presence of other woody species could protect old *V. paradoxa* trees from bushfires (56). In these ecosystems, other woody species, such as *Acacia senegal* (L.) Wild, *Annona senegalensis* Pers., and *Parkia biglobosa* (Jacq.), often share ecological niches with *V. paradoxa*, providing protection against bushfires for older trees (141).

However, the sustainability of *V. paradoxa* agroforestry systems faces numerous challenges, including poor regeneration due to the inappropriate agricultural practices, and excessive harvesting for energy needs (142). In Côte d'Ivoire, natural shea populations are severely affected by bushfires, overexploitation, climate change, the introduction of new cash crops and parasitic infestations (113).

Shea trees are highly susceptible to hemiparasitic plants infestations, especially from the Loranthaceae family. Major hemiparasites in West Africa include *Tapinanthus pentagonia*, *T. dodoneifolius*, *T. globiferus* and *T. ophiodes* (143), with infestation rates reaching 95% in Burkina Faso (144), 87% in Benin (48), and 81% in Nigeria (145). In addition, two hemiparasitic species such as *Tapinanthus bangwensis* and *Agelanthus dodoneifolius* have been reported in Côte d'Ivoire, affecting 59.7% to 65.5% of shea populations (45,49). These parasitic plant species are strongly associated with high tree mortality and pose a serious threat to the sustainability of shea parklands (143).

In addition to parasitic plants, fungal diseases and insect pests also affect shea trees (90). The main fungal pathogens, *Fusicladium butyrospermi* and *Pestalozzia heterospora*, cause dark gray leaf spots (7,67). Among insect pests, *Salebria spp.* (Lepidoptera: Pyralidae) are particularly damaging, with infestation rates of 49%-80% on trees and 4%-15% on fruits in Burkina Faso (46). Without effective control strategies, these bioaggressors could further threaten shea tree populations and production.

Several studies showed that the species is vulnerable to severe threats related to human activities and climate change (67,146–148) and that natural stands are aging (105,149) (**Figure 4a and 3b**).

The degradation of natural shea stands has been well-documented, with agricultural systems significantly influencing regeneration dynamics (17,99,142). Aging shea populations dominate cultivated fields, whereas fallow lands show higher regeneration potential (105). However, this regeneration is short-lived, often failing due to mesological factors such as inappropriate farming practices and the shortened fallow periods, which are typically less than 15 years (149).

The degradation of shea parks is further accelerated by intensified agriculture, which requires environmentally polluting inputs such as herbicides. These pressures have led to a significant decline in shea tree number per hectare over time. For example, in 1946, the number of shea tree in the Sudanese savannah zone was 230 trees/ha, but by 2008, it had dropped to just 11–5 trees/ha (150),(151). From 2005 to 2017, shea tree number per hectare in parklands declined by 75% in Benin (148). In addition, a recent study in Côte d'Ivoire showed an average number of 14 shea trees per hectare, with 6 shea trees per hectare in cashew orchards (152). Major factors contributing to this decline include the cutting of adult trees for charcoal production, slow tree growth, and inadequate maintenance of young seedlings.

Efforts to restore shea populations have focused on various regeneration techniques, including planting, transplanting, direct seeding (29,142), assisted natural regeneration (ANR) (142,153), and suckering or adventive shoot induction (142). Among these, ANR has emerged as the most effective method for restoring degraded shea parklands. This technique involves protecting and managing natural seedlings until they mature and integrate into the agricultural landscape (154) as shown in **Figure 4c**.

The survival rate of ANR exceeds 70% after two years, significantly outperforming planting and direct seeding, which show survival rates of 13.33% and 6.67%, respectively, in the sub-Sahelian sector, and similarly low rates in the northern Sudan and southern Sudan regions (142). In Côte d'Ivoire, ANR has been successfully applied to restore 56.6 ha of shea parklands for 20 producers in the Bagoué and Tchologo regions (153).

Despite its effectiveness, ANR and other restoration techniques face challenges such as the slow juvenile growth of shea trees and a lack of knowledge on their silviculture (155). Young shea trees require at least a year in nurseries before transplantation and only begin production after 20 years, provided they are adequately protected from grazing and other threats (143).

The restoration of shea populations is essential for sustaining agroforestry systems, particularly in areas with low tree density. In Côte d'Ivoire, ANR offers a promising agricultural innovation for rehabilitating degraded parklands and improving their productivity while awaiting the establishment of plantations using improved planting materials.

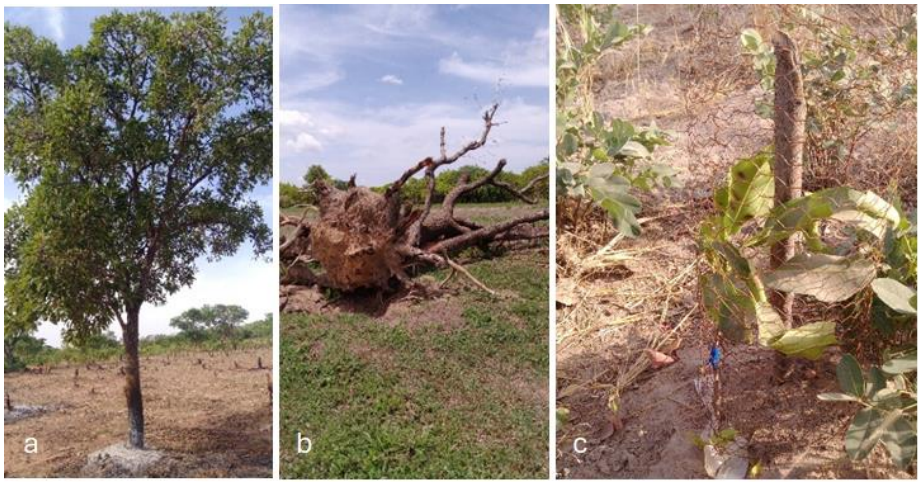


Figure 4. Threats and restoration of shea tree populations. (a) shea tree fired in a new farm installation; (b) shea tree uprooted by wind; (c) ANR: grafted and protected shea plantlet.

10. Traditional Management and Conservation of Shea Trees

Shea tree is largely non-domesticated, traditional knowledge and practices play a critical role in its management and conservation. The main traditional management practices include weeding, on-farm sparing, natural coppicing, pruning and seasonal burning.

Local communities have developed various strategies to conserve and manage shea trees in agroforestry systems (75,156). Farmers often integrate shea trees into parklands, which are characterized by scattered tree cover combined with crops such as millet, sorghum, and maize (67) as shown in **Figure 5**. These systems provide multiple benefits, including soil fertility improvement, microclimate regulation, and diversified sources of income (157).

Figure 5. Shea parkland in combination with a crop of maize

Traditional pruning practices are often used to optimize tree health and fruit yield. Farmers selectively prune branches to improve light penetration and reduce competition between branches, which promotes flowering and fruiting (75). Controlled burning is another widely used practice in shea parklands to manage undergrowth, reduce pest infestations, and create favorable conditions for regeneration (67,75).

Farmers play an active role in the conservation of shea tree populations by selecting and protecting superior phenotypes based on traits such as fruit size, pulp taste, and

oil yield (12,75,158). These selected trees are often left untouched during land clearing for agriculture, ensuring their survival and genetic diversity within the landscape.

Participatory selection has been shown to contribute to the preservation of locally adapted genetic material and the maintenance of trait diversity (69). This method not only supports conservation but also provides a basis for future domestication and breeding efforts.

Traditional management practices, such as agroforestry systems, pruning, and participatory selection, play an important role in the conservation of shea trees. However, these efforts need to be supported by modern scientific methods and supportive policies to address ongoing threats. A collaborative approach that values both local knowledge and scientific innovation is key to ensure the long-term sustainability of shea tree resources.

11. Vegetative propagation of shea tree

Vegetative propagation methods that have been used for shea tree regeneration are cuttings, grafting, layering, and in vitro propagation (90). These are valuable techniques for reducing time to maturity, increasing yields, and producing planting materials at all time. Stem cuttings, when rooted in substrates such as rice hulls and treated with indole-3-butyric acid (IBA), show high rooting rates depending on maturity of the cutting source and hormone concentration (33–37). Grafting methods, such as slit grafting, achieve high success rates, especially with young seedlings, and shorten the juvenile period, making them cost-effective for maintaining superior shea traits (29–32). However, the layering method has lower success rates, limiting its scalability (34,67).

On the other hand, *in vitro* propagation allows regeneration of shea shoots from axillary nodes and apical buds on Murashige and Skoog (MS) media containing phytohormones, although success rates vary (41,43,44). Techniques such as callus formation using immature plant tissue show promise for genetic improvement via somatic embryogenesis, although the production of whole plants from callus remains a challenge (38,40,42). Overall, vegetative propagation supports genetic fidelity, rapid growth, and the possibility of genetic modification, potentially increasing the agricultural value of the shea tree.

12. Use of shea

All shea tree organs are used (90). Roots, leaves, bark and flowers are used in traditional medicine to treat headaches, stomachaches, diarrhea, malaria and rheumatism (159). Certain organs are also used to treat gynaecological problems (latex), eye problems (leaves and bark) and to facilitate childbirth and breastfeeding (bark) (160). The trunk and branches are used to make charcoal or directly as firewood. The sweet pulp that surrounds the seed is consumed by local populations during the lean season, or used to make local drinks (161). It is also used to feed livestock such as pigs, cattle and sheep (**Figure 6**). Butter extracted from almonds is

used in local cooking as cooking oil (8). The residues left after extracting the butter are also used as pesticides (162).

Figure 6. Pigs eating shea pulp

Shea butter composition varies with climate, species, and processing and is valuable in the food, cosmetic, and pharmaceutical industries due to its unique fatty acid and bioactive profiles (128). In the food industry, the high content of saturated - unsaturated - saturated triacylglycerols (SUS TAG) and fatty acids, mainly oleic acid and stearic acid, makes shea butter a viable alternative to cocoa butter for the production of chocolate (128,163,164). Its higher melting point and crystallization properties also position shea butter as a potential palm oil substitute (128).

In cosmetics, shea butter is noted for its unusually high unsaponifiable content (5-7%), which exceeds other oils and contributes to skin-nourishing properties (127). In comparison, palm oil, soybean oil and cocoa butter have significantly lower levels of unsaponifiable, while avocado oil approaches shea butter levels (128).

In pharmaceuticals, shea butter is valued for its anti-inflammatory, antioxidant, anti-tumor and UV-protective effects (165). These are due to its bioactive unsaponifiable components, including triterpenes, catechins, tocopherols and phytosterols such as α-spinasterol and β-sitosterol. Triterpenes such as α- and β-amyrin make up the majority of the unsaponifiable portion, while tocopherols provide antioxidant benefits (166).

Chapter 2

Objectives

Chapter 2: Objectives

In the Sudanian savannas of Côte d'Ivoire, shea tree parks play a central role in the agricultural landscape. Despite its economic and social importance, the shea sector lacks a structured breeding and improvement program to meet the needs of stakeholders. Fruit production varies significantly from year to year, with over half of the trees producing little to no fruit, and only 10% being consistently good producers (166). In addition, genetic erosion is occurring within existing shea tree populations, further threatening productivity and biodiversity.

Given these challenges, conservation and genetic enhancement initiatives have been launched to identify and preserve high-yielding shea trees. In 2017, through extensive field surveys, researchers compiled a list of superior shea trees based on farmers' knowledge, leading to the establishment of an *in situ* collection of elite trees.

Although significant progress has been made, further research and development are required to enhance conservation strategies and genetic improvement to ensure high-quality shea butter production.

This Ph.D. research aligns with these ongoing efforts, aiming to provide scientific solutions for the sustainable management and efficient utilization of shea tree genetic resources in Côte d'Ivoire. The ultimate goal is to strengthen its economic and social impact by improving both production quality and market value.

This study specifically aims to:

1. Establish an *ex situ* core collection to serve as a reference for shea tree conservation and improvement. To achieve this, several key questions must be addressed:

    - Is *in situ* conservation a reliable method for preserving genetic material, and to what extent is this system vulnerable? (Chapter 3)

    - What is the level of morphological diversity (leaves, canopy, fruits, etc.) within the shea tree collection selected based on productivity criteria? (Chapter 3)

    - What is the molecular genetic diversity of this collection, and can a structure be identified within it? Which genotypes, when grouped into an *ex situ* core collection, will allow the maximum conservation of genetic diversity for the country? (Chapter 4)

2. Identify key genes for future breeding programs, specifically those involved in high-quality shea butter production, including traits related to butter content and fatty acid composition. (Chapter 5)

3. Develop a rapid propagation method for promising genotypes to ensure their conservation, facilitate research applications, and enable their large-scale distribution to support the production of high-quality shea butter in Côte d'Ivoire. (Chapter 6)

This research is expected to provide critical insights into the conservation, genetic improvement, and sustainable development of the shea sector, contributing to its long-term economic and environmental viability in Côte d'Ivoire and beyond.

Chapter 3

Morphological Traits diversity and Sustainability of the *in situ* collection of superior shea trees (*Vitellaria paradoxa* C.F. Gaertn.) in Côte d'Ivoire

Chapter 3. Morphological Traits diversity and Sustainability of the *in situ* collection of superior shea trees (*Vitellaria paradoxa* C.F. Gaertn.) in Côte d'Ivoire

As part of the shea tree improvement program in Côte d'Ivoire, an *in situ* collection of high-yielding trees was identified and georeferenced. The quantitative morphological diversity of these genotypes, which have been conserved on farmers' fields, was previously characterized, and the results were published before this thesis:

"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.
By Yao Saraka Didier Martial, DIARRASSOUBA Nafan, ATTIKORA Affi Jean Paul, FOFANA Inza Jesus, DAGO Dougba Noel and SILUE Souleymane

This study revealed a high level of morphological diversity in the quantitative traits analyzed, particularly in fruit productivity (number and weight of fruits).

However, to obtain a comprehensive dataset, it was also essential to characterize the diversity of qualitative traits. This chapter, therefore, aims to evaluate the variability of these qualitative traits to provide critical insights into the phenotypic diversity of these valuable genetic resources. Thus, a detailed morphological characterization was conducted on 220 SSTs from the Bagoué and Tchologo regions. These regions were chosen as pilot sites for SST selection due to their historical significance in shea-growing and the practical constraints posed by financial and technical limitations.

Additionally, since this evaluation was conducted three years after the georeferencing of the trees, it was possible to assess the mortality rate of these trees, providing an overview of the reliability of this conservation method. The results presented help address key research objectives:

1. The selected trees exhibit significant qualitative morphological diversity, which is essential for the future establishment of a core collection beneficial to all improvement programs.

2. *In situ* conservation is not sufficiently reliable for long-term genetic preservation, as trees are exposed to various human-induced constraints and the impact of parasitic plants. We recommend establishing an *ex situ* collection of shea trees, managed by a research organization, to ensure the preservation of this vital genetic heritage.

The finding of this chapter has been published in :

Attikora AJP, Diarrassouba N, Yao SMD, De Clerck C, Silué S, Alabi T, Lassois L,. *Morphological traits and sustainability of plus shea trees (Vitellaria paradoxa C.F. Gaertn.) in Côte d'Ivoire. Biotechnol Agron Soc. Environ.* 25 sept 2023 DOI: 10.25518/1780-4507.20462

Morphological traits and sustainability of the *in situ* Collection of superior shea trees (*Vitellaria paradoxa* C.F. Gaertn.) in Côte d'Ivoire

This chapter is an original research article published in *Biotechnology, Agronomy, Society and Environment* Volume 27; September, 25th 2023 DOI: 10.25518/1780-450. The expression "Superior Shea Tree" has been used instead of "Plus Shea Tree" for uniformity and text formatting.

Affi Jean Paul ATTIKORA[1,*], Nafan DIARRASSOUBA[2], Saraka Didier Martial YAO[2], Caroline DE CLERCK[4], Souleymane SILUÉ [2], Taofic ALABI [2,3], Ludivine LASSOIS [1]

[1]University of Liege, Gembloux Agro Bio-Tech, Terra Research Center, Plant Genetics and Rhizosphere Processes lab. Passage des déportés 2, 5030 Gembloux, Belgium. E-mail: ajpattikora@uliege.be

[2]University of Peleforo Gon Coulibaly (UPGC), Faculty of Biological Sciences, Department of Biochemistry-Genetics, , Educational and Research Unit of Genetic, BP 1328 Korhogo, Côte d'Ivoire.

[3]University of Liege, Gembloux Agro Bio-Tech, Functional and Evolutive Entomology. Passage des déportés 2, 5030 Gembloux, Belgium.

[4]University of Liege, Gembloux Agro Bio-Tech, AgricultureIsLife. Passage des déportés 2, 5030 Gembloux, Belgium

3.1. Abstract

Description of the project: Shea tree **(***Vitellaria paradoxa*, C.F Gaertn) is an essential component of natural/semi-natural savannas and agroforestry systems of Africa. It contributes to local household incomes and represents an important cash crop for export. It is important to improve our knowledge about the viability and genetic diversity of "Superior Shea Tree" (SST: shea trees whose genetic superiority has not yet been proven by an appropriate progeny test) genotypes to conserve the genetic diversity of the species and the potential for producing improved species. **Objectives**: We characterized the diversity of an *in situ* SST collection in Côte d'Ivoire and evaluated its long-term sustainability. **Methods:** Twelve qualitative morphological traits were studied on 220 SSTs randomly sampled among 405 SSTs also randomly selected from *in situ* collections in the Bagoué and Tchologo regions of northern Côte d'Ivoire to assess *in situ* conservation sustainability. **Results:** Most qualitative morphological traits were highly diverse based on the Shannon diversity index (0.55-0.98 range), with significant differences between regions. Hierarchical ascending classification gathered the SSTs into three groups. The lack of sustainability of the *in situ* collection for long-term conservation of the species was confirmed because it is endangered by a high mortality rate (8.15% after three years) and high levels of infestation by *Loranthaceae* (83.68%). **Conclusion:** We recommend establishing an *ex situ* collection of shea trees, managed by a research organization, to ensure the preservation of this vital genetic heritage. To maximize the genetic diversity of the future *ex situ* collection, we suggest establishing a core collection based on molecular diversity characterized from molecular markers, such as single nucleotide polymorphism (SNP).

Keywords: Morphology, conservation of genetic resources, plant collections, Superior trees, *Vitellaria paradoxa*, parasitic plants, resource management

3.2. *Introduction*

Shea butter tree or shea tree (*Vitellaria paradoxa* C. F. Gaertn.) is an indigenous African tree species belonging to the *Sapotaceae* family. It grows naturally within a wide belt of more than 3.4 million km^2 across 21 countries between western Senegal and eastern Uganda, and it supports an estimated 16.2 million of shea nut collectors (7,8,10,167). The species is taxonomically subdivided into two subspecies, namely (1) *V. paradoxa* subsp. *paradoxa* found in West and Central Africa, and (2) *V. paradoxa* subsp. *nilotica*, located in East Africa. The tree grows up to 20 m height and 1 m diameter at breast height, and has a lifespan of more than 200 years. In Côte d'Ivoire, shea tree populations grow in the semi-arid zone of the savanna region (extreme north of the country), similar to other sub-Saharan African countries; however, some populations also occupy the pre-forest transition zone thanks to their acquired high level of adaptation (113). In theory, these adaptation processes should generate higher genetic richness in Côte d'Ivoire compared to other countries where this plant is restricted to the semi-arid zone of the savanna regions.

The locations from which shea tree germplasm was introduced to other countries in Africa are not well known because of changes in vegetation. The species probably spread out from refugia. Furthermore, human management indicates ongoing semi-domestication (14). The current diversity and distribution of shea trees reflects a long history of evolution and responses to climate change and other environmental factors that determine the morphology of its organs (stem, leaf, flower, fruit, and seed) (14). The large distribution of shea was doubled by practicing allogamy as a mode of reproduction, but this practice might have led to high intra-specific diversity (113).

Shea tree is a socio-economically important species for many traditional communities in Africa. It provides substantial income to rural households in Western and Central Africa. Because women predominantly collect shea fruit for transformation, it is considered as a “female crop” in all production areas. Several tree parts are used in rituals, pharmaceutics, cosmetics, construction, and food products (99). The fruit and associated main derived product – shea butter (“beurre de karité” in French) – are used for local consumption and for trade on local and international markets. The concept of exporting shea kernels to Europe on a large scale was first introduced by French colonial administrations in the early 20th century. However, at the end of the 1950s, low shea production led shea collectors and processors to use the kernels and butter for their own domestic consumption and/or for sale on local periodic and cross-border markets (168). It was only after independence in 1960 that the new international demand for edible cocoa butter equivalents (CBEs) led to the introduction of several efforts to regulate and control the shea trade through stabilization funds and parastatal marketing boards (169).

These products provide enormous benefits and contribute to national economies. The export of raw shea kernel and shea butter to international markets in Europe, Asia, and the USA has risen in recent years (by 25% between 1994 and 2004), and is expected to keep rising (22,69,170–172). Shea kernels are processed to manufacture a wide range of food products (including chocolate) and cosmetics (173). However,

shea trees face various threats, including climate change, emerging bioagressors, natural regeneration issues, demographic pressure, bushfires, overexploitation, changing land uses, and cutting for firewood and charcoal (72). Consequently, shea tree was listed as a "vulnerable species" by the International Union for Conservation of Nature (IUCN) in 1998, and is expected to be upgraded to "endangered species" in the future. The function of genetic variation studies in assisting plant conservation approaches and improving cultivars has long been recognized (111). High genetic diversity is key to the adaptation potential to current and future biotic and abiotic constraints in the context of global change (112).

Despite the socio-economic importance of shea butter as a multi-purpose product and the risk of shea tree becoming an endangered species, key knowledge gaps remain regarding its diversity, optimal conservation approaches, and ways of improving its resilience. In Côte d'Ivoire, as in many producing countries, the genetic resources of shea trees are conserved and maintained in the field by farmers. This diversity is an important underexploited heritage, but it is increasingly under threat. Therefore, it is essential to protect and conserve shea trees in a sustainable way. Genetic and morphological trait studies are required to manage the genetic resources of shea trees effectively. In particular, the development of national and regional plans to conserve their genetic resources should be promoted (67). Following this author's recommendations and based on a participatory survey, the shea breeding program of Côte d'Ivoire identified 639 "superior shea trees" (SSTs) (shea trees whose genetic superiority has not yet been proven by an appropriate progeny test) in the Bagoué and Tchologo regions (60). These SSTs constitute part of the *in situ* collection of shea trees in Côte d'Ivoire, in place since 2017.

We characterized the diversity of this collection based on the variation of qualitative morphological traits and evaluated its durability. Our results are expected to provide a basis for relevantly involving the identified SSTs in subsequent hybridization programs, generating high-performance plant material and restoring the shea tree sector in Côte d'Ivoire. In particular, our results are expected to help reinforce and adapt the conservation process of this endangered species.

3.3. Materials and Methods

3.3.1. Study area

Côte d'Ivoire is divided into 31 regions with varying cropping patterns, climates, livelihoods, soil types, and natural resources. The country covers 322,462 km² of terrestrial area, and is located in West Africa (4° 30' to 10° 30' N latitude and 2° 30' to 8° 30' W longitude). It belongs to the tropical zone. Based on climatic factors and variation in vegetation, the savannas of the northern part of Côte d'Ivoire – where shea trees grow – are subdivided in two main zones called the Sudanian and sub-Sudanian savannas. The Sudanian savanna corresponds to the main shea tree production zone, whereas the sub-Sudanian savanna corresponds to a transition production zone. Our study was carried out in the Sudanian savannas of the Tchologo and Bagoué regions in the northern region of Côte d'Ivoire (**Figure 7**), where shea trees are a significant component of the farming system. Tchologo region covers

17,382 km², with 467,958 inhabitants (density 26.92 inhabitants/km²). This region includes three departments: Ferkessédougou, Kong, and Ouangolodougou (**Figure 7**). In contrast, Bagoué region covers 10,678 km², with 375,687 inhabitants (35.2 inhabitants/km²). Bagoué region includes three departments: Boundiali, Kouto, and Tengréla (**Figure 7**).

The climate in these two regions is Sudanese, with two main seasons (dry and rainy). The dry season extends from November to April, and overlaps with the Harmatan (dry and dusty) season extending between December and February, with temperatures peaking in March and April. The rainy season extends from May to October, and peaks in August and September. Annual rainfall is around 1,200 mm (174). The annual average temperature is 27 °C. The vegetation is Sudano-Guinean, dominated by wooded and grassy savannas (175).

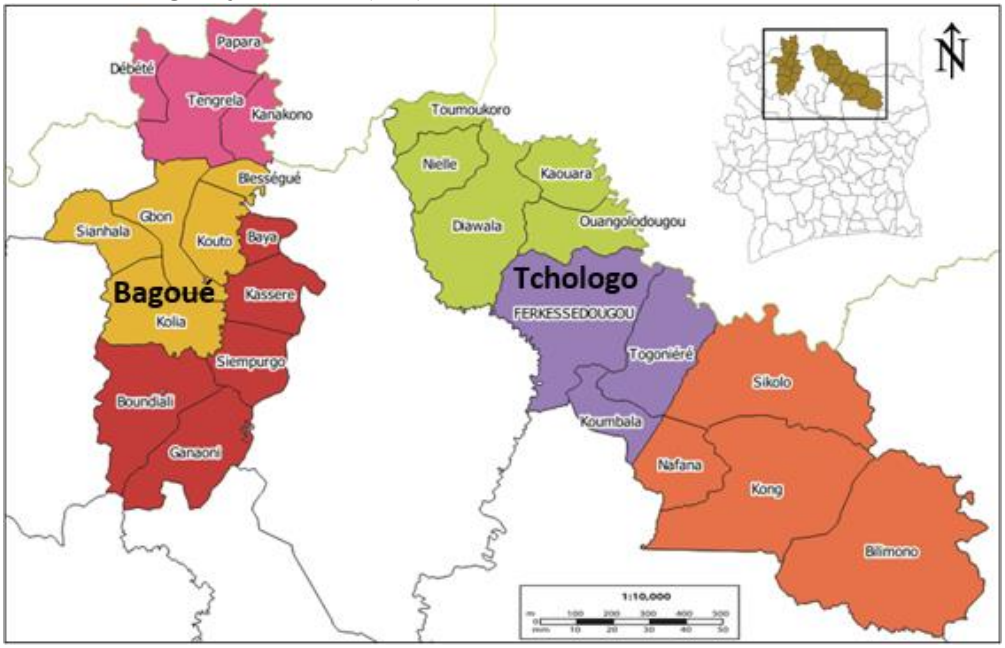


Figure 7. Localization of the two study areas in Côte d'Ivoire. Enlarged image: Bagoué region (left) including Tengrela (pink), Kouto (yellow) and Boundiali (red) departments; Tchologo region (right) including Ouangolodougou (green) Ferkéssédougou (purple) and Kong (orange) departments.

3.3.2. Plant material

The shea trees considered in this study were mature trees regenerated/protected by farmers, probably for centuries (168), and belonged to the subspecies *paradoxa*. They had been previously designated as "superior shea trees" (SSTs), i.e., identified by the Shea breeding program of the University of Peleforo Gon Coulibaly (Côte d'Ivoire) in collaboration with the Agence Nationale d'Appui au Développement Rural (ANADER, Côte d'Ivoire), following an extensive survey between February and June 2017. This program consisted of a participatory survey with farmers allowed to select SSTs based on specific criteria, including high fruit yield, sweet taste of the fruit pulp,

large fruit size, early flowering every year, and periodicity of fruit production. Six hundred and thirty-nine shea trees – considered as "superior" trees – were selected, geo-referenced for future identification and conserved *in situ* on the land plots of 400 farmers in the Bagoué and Tchologo regions. Four hundred and five SSTs were randomly sampled in the two regions to evaluate their capacity for *in situ* survival. Then, 200 SSTs were randomly sampled among these 405 SSTs, and 20 more SSTs from Tengréla Department were selected for diversity analyses of qualitative traits (**Table 2**).

Table 2. Number of shea trees observed *per* region and department

Region	**Department**	**Qualitative trait diversity study**		***In situ* survival study**	
Bagoué	Boundiali	23	146	98	210
	Kouto	50		57	
	Tengréla	73		53	
Tchologo	Ouangolodougou	12	74	38	197
	Ferkéssédougou	21		82	
	Kong	41		77	
Total		220		405	

3.3.3. Analysis of qualitative trait diversity

Twelve qualitative morphological variables were selected based on the minimum descriptors of shea trees as defined by the International Plant Genetic Resources Institute (176). These traits were evaluated on 220 randomly selected SSTs (**Table 2**). The descriptors were observed on trees, branches, healthy adult and young fresh leaves, and seeds, as listed in **Table 3**. Five descriptors were observed from 10 leaves randomly selected from the crown. The two seed-related descriptors were observed from 10 seeds of each individual tree. Inflorescence density was determined by looking at the presence and density of flowers on the terminal branches. The remaining 4 descriptors were based on the crown shape, the branching pattern, the tree growth habit and the branch density.

Table 3. Shea tree descriptors, qualitative traits, and modalities (IPGRI, 2006)

Descriptors	Qualitative variables	Modalities
Growth descriptors	Crown shape (CRS)	Pyramidal, broadly pyramidal, spherical, oblong, semicircular, elliptical
	Tree growth habit (TGH)	Erect, semi-erect, spreading
	Branching density (BRD)	Sparse, medium, dense
	Branching pattern (BRP)	Erect, opposite, verticillate, horizontal, irregular, plagiotropic
Leaf Descriptors	Leaf blade shape (LBLS)	Obovate, elliptic, broadly elliptic, narrowly elliptic, oblong, obovate-oblong, ovate-oblong
	Leaf base shape (LBS)	Oblique, rounded, cuneate, shortly attenuate
	Leaf apex shape (LAS)	Acute, acuminate, retuse, obtuse
	Young leaf color (YLC)	Red, pink, light green, green
	Adult leaf color (ALC)	Light green, green, dark green, pinkish green
Inflorescence descriptor	Inflorescence density (IND)	Sparse, intermediate, dense
Seed descriptors	Seed coat color (SCC)	Creamish, dull brown, brown, pale brown, dark brown
	Seed shape (SES)	Spheroid, ellipsoid, oval, ovoid

The data were analyzed using Rstudio version 4.2.2. The normalized Shannon-Weaver diversity index (H') (range: 0 to 1) was used to assess the diversity of each qualitative trait (114). It was calculated as follows:

$$H' = -\sum_{i}^{n} \left(\frac{Ni}{N}\right) \times \ln\left(\frac{Ni}{N}\right) / \ln(n)$$

where Ni is the number of individuals observed for a given modality and a given descriptor, N is the total number of individuals, and n is the number of modalities for a given characteristic. When H' is close to 1, the diversity of a given qualitative trait is high in all populations (114). The equal probability of the frequencies (pi = **Ni**/**N**) of appearance of the modalities of a given qualitative trait was tested using the $\chi 2$ test

at the 5% threshold. It was also used to test the homogeneity of the proportion of descriptor modalities between the Bagoué and Tchologo regions. A significant result indicated that at least two of the observed frequencies were different. Multiple correspondence analysis (MCA) and hierarchical ascending classification (HAC) structured the SSTs based on the qualitative descriptors. MCA is a multivariate descriptive method that groups variables according to the variability that they explain in a given population. This method allowed us to generate a table of the contribution to the inertia showing the degree of linkage between variables and components. The choice of axes was made on the basis of the elbow method. HAC was carried out using the unweighted paired groups method of analysis (UPGMA) to cluster shea trees based on their morphological diversity.

3.3.4. Evaluation of the sustainability of the *in situ* SST collection

The sustainability of the *in situ* collection was evaluated based on two parameters of the randomly selected 405 SSTs (**Table 2**). The mortality rate was monitored over the 3 years following initial SST identification and selection by the Shea breeding program of the University of Peleforo Gon Coulibaly (Côte d'Ivoire) in collaboration with the Agence Nationale d'Appui au Développement Rural (ANADER, Côte d'Ivoire) in 2017. The parasitic plant *Loranthaceae* was the main biotic threat to the survival of the *in situ* collection. Its occurrence was quantified for each SST.

3.4. *Results*

3.4.1. Diversity of qualitative traits

3.4.1.1. Diversity of growth descriptors

Crown shape (CRS)

The six crown shape phenotypes were distributed unevenly. The broadly pyramidal phenotype was most common (26.36%; **Figure 8**). The other phenotypes (elliptical, oblong, spherical, semicircular, and pyramidal) represented 18.18%, 17.27%, 15.45%, 13.18%, and 9.55% of total phenotypes, respectively. The $\chi 2$ test was highly significant ($\chi 2 = 21.26$; $P = 0.001$). The Shannon-Weaver index was high ($H' = 0.97$; **Table 4**). The proportions of CRS modalities differed significantly in the Bagoué and Tchologo regions ($\chi^2 = 12.05$; $p = 0.034$).

Figure 8. Crown shape diversity of superior shea trees from the Bagoué and Tchologo regions in northern Côte d'Ivoire. Left to right: broadly pyramidal (26.36%), elliptical (18.18%), oblong (17.27%), spherical (15.45%), semicircular (13.18%) and pyramidal (9.55%) phenotypes.

Tree growth habit (TGH)

The $\chi 2$ test was significant (χ^2 = 11.16; p = 0.004) for the three phenotypes describing tree bearing. Extended and erect phenotypes were most common (38.18% and 39.09% of the population, respectively). In comparison, the semi-erect phenotype represented 22.73% (**Table 4**) of the total trees. The Shannon-Weaver index was H'= 0.98. The proportions of TGH modalities differed significantly in the Bagoué and Tchologo regions (χ^2 = 18.7; p < 0.001).

Branching density (BRD)

Three types of branch density (sparse, medium, dense) were identified, and occurred at unequal frequencies (χ^2 = 111.45; p < 0.001). The medium-density phenotype was prevalent (69.97%; **Table 4**). The Shannon-Weaver index was H'= 0.78. The proportions of all BRP modalities were statistically similar in the Bagoué and Tchologo regions (χ^2 = 2.38; p = 0.3).

Branching pattern (BRP)

The irregular phenotype was most common (29.55%) among the branching pattern phenotypes. The verticillate, plagiotropic, erect, opposite and horizontal phenotypes represented 19.55%, 16.82%, 15.54%, 10.91% and 7.73% of the total, respectively. The Shannon-Weaver index was H'= 0.95. The proportions of the BRP modalities differed significantly in the Bagoué and Tchologo regions (χ^2 = 14.62; p = 0.01). Plagiotropism of the BRP was higher in Tchologo region (29.7%) than in Bagoué region (10.3%). The others modalities were statistically similar.

3.4.1.2. Diversity of leaf descriptors

Leaf blade shape (LBLS)

Five leaf blade shapes were observed (obovate, elliptic, broadly elliptic, narrowly elliptic, and oblong), while the obovate-oblong shape was not observed. Oblong and narrowly elliptic leaves were most frequent (51.36% and 34.55%, respectively; **Table 4**). The Shannon-Weaver index was H'= 0.59. The LBLS modalities were statistically similar in the Bagoué and Tchologo regions (χ^2 = 6.45; p = 0.26).

Leaf base shape (LBS)

Three out of four types of leaf base shape were observed (oblique, rounded, and cuneate) in varying proportions (χ^2 = 26.21; P < 0.001). The short attenuated type was not observed. The wedge-shaped phenotype was most common (45.91%). The other two phenotypes (rounded and oblique) represented 25% and 29.09% of total phenotypes, respectively (**Table 4**). The Shannon-Weaver index was H' = 0.77. The proportions of all LBS modalities were statistically similar in the Bagoué and Tchologo regions (χ^2 = 0.09; p = 0.96).

Leaf apex shape (LAS)

The four phenotypes of apex shape (obtuse, acute, rounded, acuminate) were present at varying frequencies ($\chi^2 = 48.25$; $p < 0.001$), with a Shannon-Weaver index H' = 0.89. The acute, rounded, and obtuse shapes were most common (31.36%, 29.09%, and 34.55% of the total, respectively). The acuminate phenotype was least common (5%). The proportions of all LAS modalities were statistically similar in the Bagoué and Tchologo regions ($\chi^2 = 5.33$; $p = 0.17$).

Young leaf color (YLC)

Four colors (green, light green, pink, red) of young leaves were observed at varying proportions ($\chi^2 = 213.74$; $p < 0.001$), with a Shannon-Weaver index H' = 0.70. The predominantly light green phenotype was most prevalent (67.27%), followed by the green phenotype (16.36%), and the red (9.09%) and pink (7.27%) phenotypes (**Table 4**). The proportions of all YLC modalities were statistically similar in the Bagoué and Tchologo regions ($\chi^2 = 4.6$; $p = 0.2$).

Table 4. Qualitative traits diversity of 220 superior shea trees from the Bagoué and Tchologo regions in northern Côte d'Ivoire.

Traits	Modalities	Frequency (%)	Bagoué (%)	Tchologo (%)	χ 2	p	H'
CRS	Pyramidal	9.55	12.3	4.1	21.3	0.001**	0.97
	Broadly pyramidal	26.36	25.3	28.4			
	Spherical	15.45	18.5	9.5			
	Oblong	17.27	13.0	25.7			
	Semicircular	13.18	14.4	10.8			
	Elliptical	18.18	16.4	21.6			
TGH	Erect	38.18	39.7	35.1	11.2	0.004**	0.98
	Semi-erect	22.73	14.4	39.2			
	Spreading	39.09	45.9	25.7			
BRD	Sparse	14.55	13.0	17.6	111.4	P< 0.001***	0.78
	Medium	66.36	69.9	59.5			
	Dense	19.09	17.1	23.0			
BRP	Erect	15.45	15.8	14.9	38.1	P< 0.001***	0.95
	Opposite	10.91	12.3	8.1			
	Verticillate	19.55	20.5	17.6			
	Horizontal	7.73	9.6	4.1			
	Irregular	29.55	31.5	25.7			
	Plagiotropic	16.82	10.3	29.7			
LBLS	Obovate	1.36	0.7	2.7	295.9	p< 0.001***	0.59
	Elliptic	7.73	9.6	4.1			
	Broadly elliptic	1.36	1.4	1.4			

	Narrowly elliptic	34.55	34.2	35.1			
	Oblong	51.36	52.1	50.0			
	Obovate-oblong	3.64	2.1	6.8			
	Ovate-oblong	0	0	0			
LBS	Oblique	29.09	29.5	28.4	26.2	p< 0.001***	0.77
	Rounded	25	25.3	24.3			
	Cuneate	45.91	45.2	47.3			
	Shortly attenuate	0	0	0			
LAS	Acute	31.36	31.5	31.1	48.2	p< 0.001***	0.89
	Acuminate	5	4.1	6.8			
	Retuse	29.09	25.3	36.5			
	Obtuse	34.55	39.0	25.7			
YLC	Red	9.09	11.6	4.1	213.7	P< 0.001***	0.70
	Pink	7.27	6.8	8.1			
	Light green	67.27	63.7	74.3			
	Green	16.36	17.8	13.5			
ALC	Light green	14.55	20.5	2.7	143.2	P< 0.001***	0.58
	Green	71.36	74.7	64.9			
	Dark green	14.09	4.8	32.4			
	Pinkish green	0	0	0			
IND	Sparse	33.68	41.1	25.6	20.5	p< 0.001***	0.97
	Intermadiate	23.68	25.0	22.1			
	Dense	42.64	33.9	52.3			
SCC	Creamish	0	0	0	88.5	P< 0.001***	0.55
	Dull brown	0	0	0			
	Brown	60.91	60.3	62.2			
	Pale brown	9.55	8.9	10.8			
	Dark brown	29.55	30.8	27.0			
SES	Spheroid	5	5.5	4.1	79.7	p< 0.001***	0.85
	Ellipsoid	18.18	15.1	24.3			
	Oval	31.36	28.1	37.8			
	Ovoid	45.45	51.4	33.8			

*CRS, crown shape; TGH, tree growth habit; BRD, branching density; BRP, branching pattern; LBLS, leaf blade shape; LBS, Leaf base shape; LAS, leaf apex shape; YLC, young leaf color; ALC, adult leaf color; IND, inflorescence density; SCC, seed coat color; SES, seed shape; H', normalized Shannon-Weaver diversity index; ** highly significant difference; *** very highly significant difference.*

Adult leaf color (ALC)

Three colors (light green, green, dark green) were observed at varying proportions on adult leaves (χ^2 = 143.19; p < 0.001), with a Shannon-Weaver index H' = 0.58. Pinkish green leaves were not observed. The green, light green and dark green phenotypes represented 71.36%, 14.55%, and 14.09% of total phenotypes,

respectively (**Table 4**). The light green and dark green modalities differed significantly in Bagoué region *vs*. Tchologo region. Light green was higher in Bagoué (20.5%) than in Tchologo (2.7%), whereas dark green was higher in Tchologo (32.4%) than in Bagoué (4.8%).

3.4.1.3. Diversity of inflorescence descriptors

Inflorescence density (IND)

Dense, intermediate, and sparse inflorescence densities represented 42.64%, 23.68%, and 33.68% of the phenotypes, respectively ($\chi^2 = 13.71$; $p = 0.0001$) and the Shannon-Weaver index was H'= 0.97 (**Table 4**). The proportion of IND modalities differed significantly in the Bagoué and Tchologo regions ($\chi^2 = 19.59$; $p < 0.001$).

Dense and sparse flowering types dominated in Tchologo (55.55%) and Tengréla (56.25%) departments, respectively (**Figure 9**).

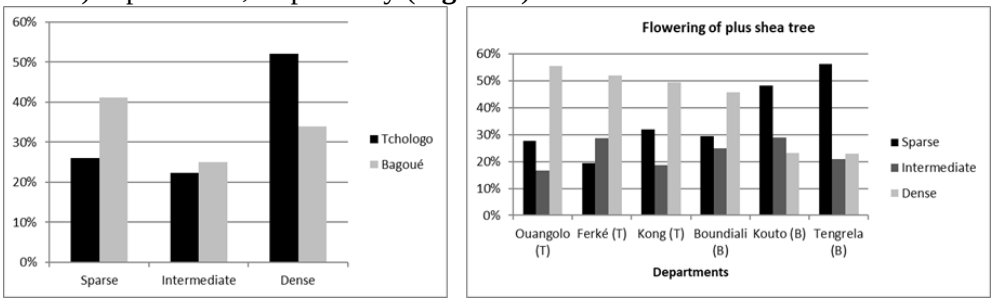


Figure 9. Diversity of the flowering characteristics in superior shea trees. Ouangolo, Ouangolodougou; Ferké, Ferkéssédougou; T, Tchologo; B, Bagoué.

3.4.1.4. Diversity of seed descriptors

Seed coat color (SCC)

Three of the five possible phenotypes (brown, pale brown, dark brown) were present in both regions ($\chi^2 = 88.48$; $p < 0.001$). Creamish and dull brown seed coat colors were not observed. The brown phenotype was most common (60.91%), followed by dark brown (29.54%) and pale brown (9.55%) (**Figure 10**; **Table 4**). The Shannon-Weaver index was H' = 0.55. The proportions of all SCC modalities were statistically similar in the Bagoué and Tchologo regions ($\chi^2 = 0.46$; $p = 0.8$).

Figure 10. Seed color. Left, brown; middle, dark brown; right, pale brown.

Seed shape (SES)

Three out of four seed shape phenotypes were observed at varying proportions ($\chi^2 = 79.67$; $p < 0.001$). The ovoid and oval forms were most common (49.54% and 31.82%, respectively), followed by the ellipsoid phenotype (18.64%). The spheroid

shape was not observed. The Shannon-Weaver index was H' = 0.85. The proportions of all SES modalities were statistically similar in the Bagoué and Tchologo regions ($\chi^2 = 7.34$; $p = 0.06$).

3.4.2. Structuring and clustering based on SST qualitative traits

The multiple correspondence analysis (MCA) of the qualitative descriptors showed that the total inertia was 3.1, with 44 active modalities. The elbow method was applied to the 34 eigenvalues obtained in the MCA to select the first three factorial axes for the analysis. A significant jump was detected just after the third eigenvalue; therefore, the three factorial planes formed by the first three factorial axes of the MCA were used. These three axes represented 18.4% of total inertia in the plot, while axes 1, 2 and 3 represented 6.5%, 6.4%, and 5.5% of the total, respectively. The representation of variables in the factorial planes showed that axis 1 crossed with axis 2 and axis 3, and axis 2 crossed with axis 3, representing 12.35%, 11.7%, and 11.33% of reality, respectively. For each axis, the average theoretical percentage of inertia explained by each modality was 2.27% (100%/44). However, the relative contribution of each modality to each axis varied considerably (axis 1: 0-17%; axis 2: 0-11.25%; axis 3: 0-13.25%). Axis 1 was mostly explained by CRS and certain leaf variables (LBLS, LAS, YLC, and ALC), axis 2 by LAS, LBLS, and certain growth descriptors (CRS, TGH, BRD, and BRP), and axis 3 by certain growth (BRP, TGH, and CRS) and leaf (ALC and YLC) descriptors. The variables contributing to the formation of the different axes are showed in **Figure 11**. Following this analysis, we used the nine descriptors listed in **Table 5** to perform a hierarchical ascending classification (HAC).

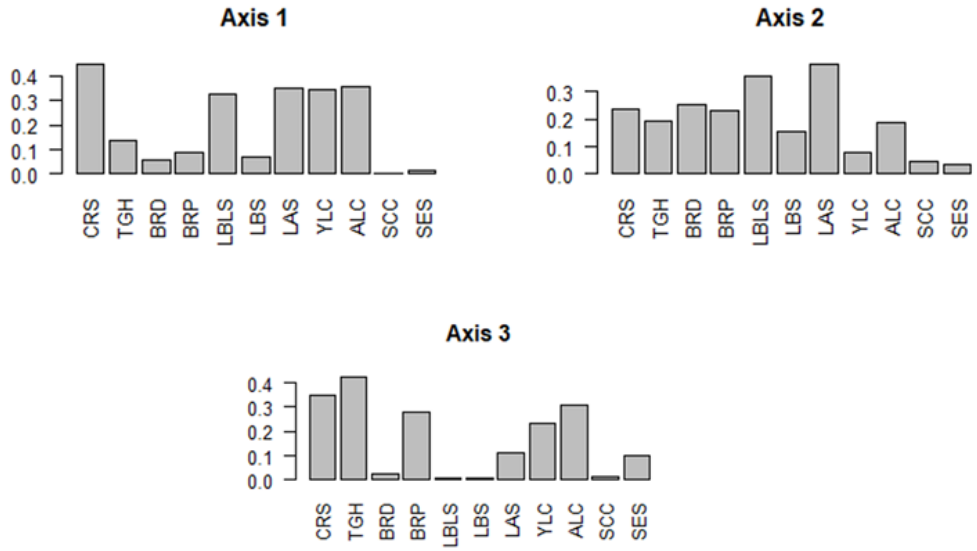


Figure 11. Contribution of the qualitative traits to axes 1, 2 and 3 of the multiple correspondence analysis (MCA)**.** *CRS, crown shape; TGH, tree growth habit; BRD, branching density; BRP, branching pattern; LBLS, leaf blade shape; LBS, leaf base shape; LAS, leaf apex shape; YLC, young leaf color; ALC, adult leaf color; SCC, seed coat color; SES, seed shape.*

HAC of 220 SSTs separated them into different classes according to their degree of similarity based on morphological qualitative traits. The dendrogram obtained from Gower's distance and the unweighted paired groups method of analysis (UPGMA) delineated three clusters (**Figure 12**). These three phenotypic groups represented the morphological characteristics of the SST population.

Group 1 was characterized by 'Light Green Adult Leaf', 'Red Young Leaf', 'Oblique Leaf Blade', 'Irregular Branching Patterns', 'Spherical Crown' and 'Spreading Tree Growth Habit';

Group 2 was characterized by 'Green Adult Leaf', 'Acute Leaf Apex', 'Cuneate Leaf Blade' and 'Narrow Leaf Blade Shape';

Group 3 included 'Retuse or Obtuse Leaf Apex', 'Oblong Leaf Blade', 'Medium Branch Density' and 'Erect Tree Growth Habit'.

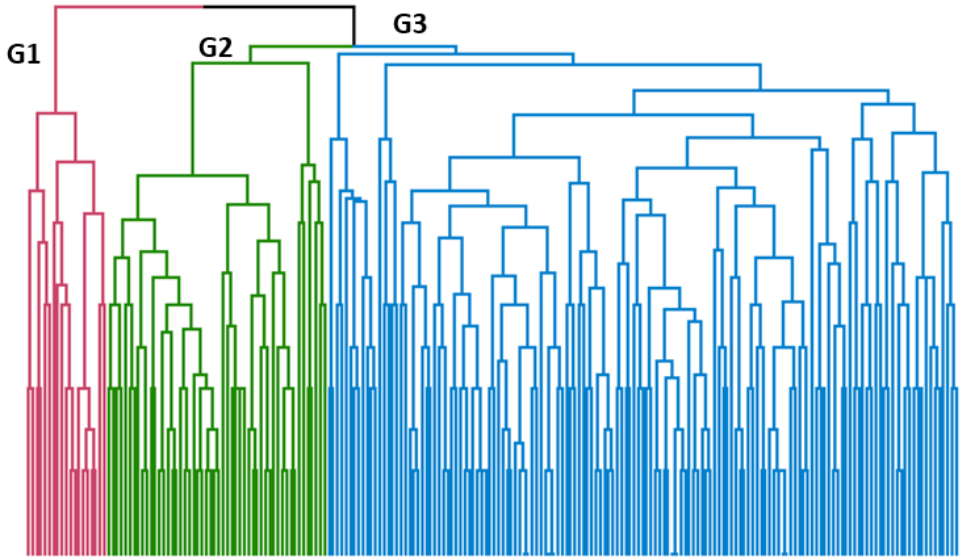


Figure 12. Clustering of the superior shea tree (SST) sample based on morphological traits using the UPGMA method.

The Wilks test was significant (alpha = 0.01, $p < 0.0001$), rejecting the null hypothesis of equality of the three vector groups, whose characteristics are presented in **Table 5**.

Table 5. Clusters and their characteristics.

Clusters	Characteristics	Number of Superior shea trees (SSTs)
G1	Light green ALC, red YLC, oblique LBS, irregular BRP, spherical CRS and spreading TGH	19
G2	Green ALC, acute LAS, cuneate LBS and narrow LBLS	52
G3	Retuse or obtuse LAS, oblong LBLS, medium BRD and erect TGH	149

*CRS, **c**rown shape; TGH, **t**ree growth habit; BRD, **b**ranching density; BRP, **b**ranching pattern; LBLS, **l**eaf blade shape; LBS, **l**eaf base shape; LAS, leaf apex shape; YLC, **y**oung leaf color; ALC, **a**dult leaf color.*

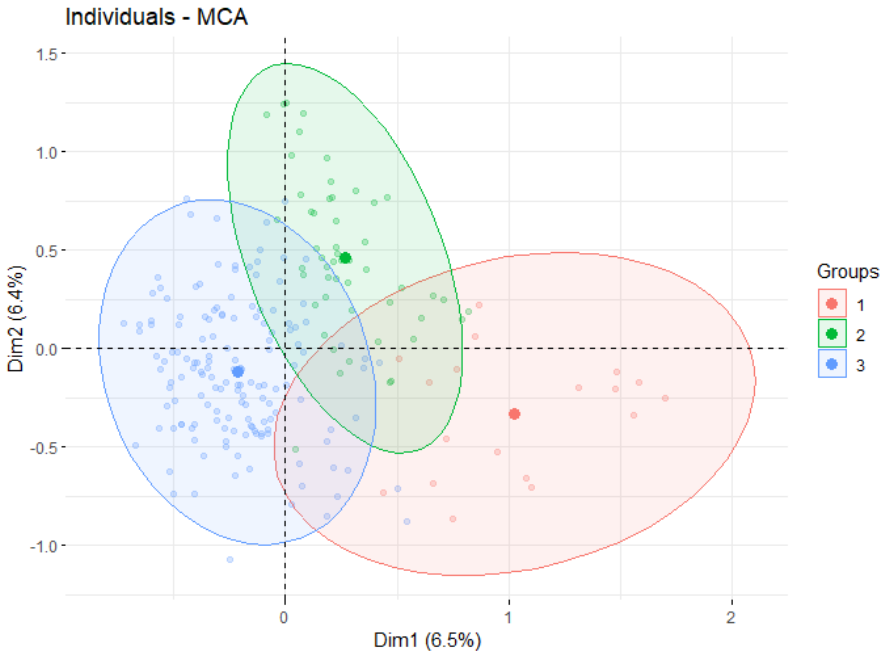


Figure 13. Distribution of individuals from the three groups of SSTs from the Bagoué and Tchologo regions.

Figure 13 shows the dispersion of individuals in the three groups in factorial plans 1 and 2. Distinct groups were not obtained when the Bagoué and Tchologo regions were used as grouping variables, based on the dispersion of individuals in the first two factorial planes (**Figure 14**).

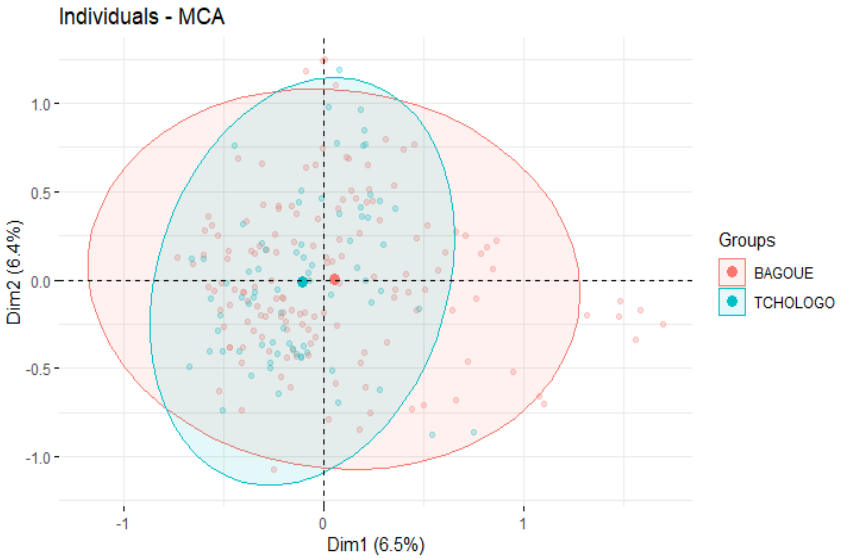


Figure 14. Distribution of individuals of SSTs using Bagoué and Tchologo regions as groups.

3.4.3. Evaluation of the sustainability of the *in situ* SST collection

3.4.3.1. Trees surviving three years after the setting up of *in situ* conservation

Three years after SSTs were identified and *in situ* conservation was established, 33 out of 405 trees had died (8.15% of the collection). A higher mortality rate was recorded in Bagoué region (11.54%) compared to Tchologo region (4.57%). Kouto department (12.07%; Bagoué) had the highest mortality rate, and Kong department (2.60%; Tchologo) had the lowest (**Figure 15**).

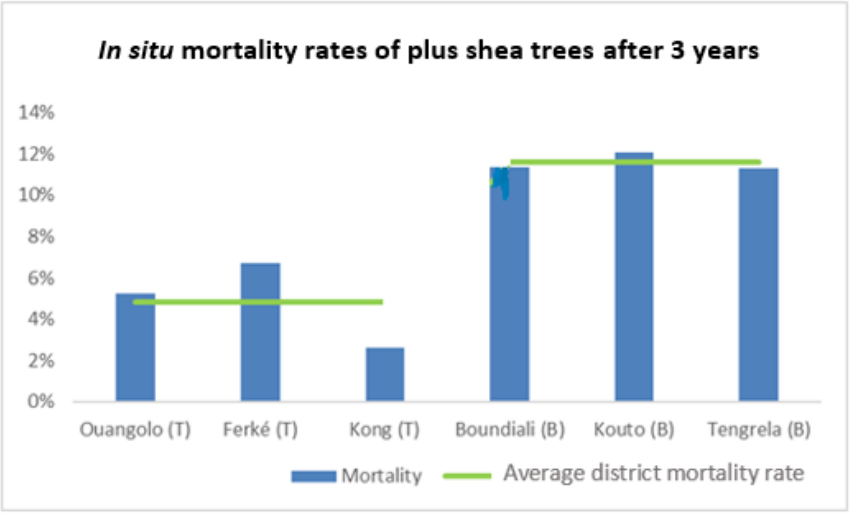


Figure 15. Mortality rate of superior shea trees after three years of an *in situ* conservation (Ouangolo, Ouangolodougo; Ferké, Ferkéssédougo; T, Tchologo; B, Bagoué).

3.4.3.2. Presence of the parasitic plant Loranthaceae

Overall, 83.64% of the SSTs had at least one clump of mistletoe. The SSTs of Tengréla Department were most impacted, with at least one clump *per* tree. Kong was the least affected department (62.67% infestation rate), followed by Boundiali (73.91%). Overall, the trees of Bagoué region were more infested than those of Tchologo region, but no significant difference was evidenced (85.34% *versus* 81.91% infestation rates, respectively; **Figure 16**).

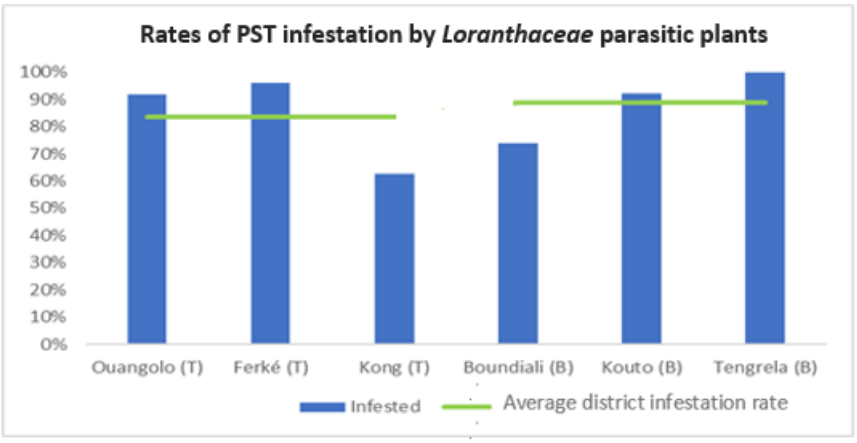


Figure 16. Rates of Superior-shea tree infestation by *Loranthaceae* (Ouangolo, Ouangolodougo; Ferké, Ferkéssédougo; T, Tchologo; B, Bagoué).

3.5. *Discussion*

This study demonstrates that the morphological diversity of the qualitative traits of the *in situ* SST collection of Côte d'Ivoire is high. Most shea tree descriptors (176) displayed high diversity (Shannon diversity index – a reliable indicator of trait diversity – range: 0.55-0.98). However, in view of the descriptors studied in the present work, the Shannon index is a mixture of the number of modalities, the frequencies in each of them and especially the easiness to distinguish the different modalities of each descriptor. For example, distinguishing the 'dark green', 'green' and 'light green' ALCs was difficult because of borderline situations with a qualitative modality as opposed to a quantitative one. Conversely, scoring the 4 YLC modalities (green, light green, pink, and red) was much easier, especially when differentiating between red or pink from (light) green.

The whole range of diversity of the growth descriptors and inflorescence density was assessed, even though frequencies were variable and unequal. The SST collection was primarily characterized by medium branch density (67%) and intermediate to dense inflorescence density (66%). However, inflorescence density was higher in Tchologo region than in Bagoué region (52.1% *versus* 33.9% dense flowering, respectively). Lower percentages in Bagoué region might be explained by the low rate of dense flowering in Kouto and Tengréla departments (23% each). The high percentage of medium branch density might be explained by farmers preferentially keeping trees with medium shade to limit potential negative effects like reduced

speculative yields in these agro-forestry systems. Shading by shea trees negatively impacted maize grain and straw yields (46.64% and 32.8% reductions, respectively) compared to exposed areas (177). Four types of crown shape have been identified in shea parkland located at Tengréla (99,119). In contrast, we identified six types of crown shape, likely due to our survey area being larger. Crown shape can be modified by agricultural practices, but competition with other trees may also have an effect; however, this explanation is less likely because our study was conducted on isolated trees.

Certain leaf characteristics were not observed in our study, e.g., ovate-oblong LBLS, short-attenuate LBS, and pinkish-green ALC. Furthermore, even though young leaves exhibited a variety of colors, green dominated (83%) over red/pink (17%). Mature leaves were entirely green, without any pinkish-green hue. LBLS was primarily characterized by narrow elliptic and oblong shapes (86%), while the remaining five types represented only 14%. This study highlights higher trait diversity compared to another study (27), possibly due to differences in scale, as our study evaluated SST distributions across two major regions as opposed to one park. Only three of the five potential seed colors delineated in the IPGRI were found, with brown seeds dominating (61%). In contrast, all the four seed shape categories were found, with ovoid and oval-shaped seeds dominating (45.45% and 31.36%, respectively). Anyomi et al. (2023) reported variations in shea nut shape and recorded two ellipsoid seed shapes in 86.32% of the samples (178). This might be explained by farmers preferentially selecting certain characteristics, hence a negative selection pressure on the others. For instance, local communities can define shea varieties according to certain criteria such as fruit size, pulp taste, fruiting period, and seed color (27).

Seven of the twelve qualitative descriptors were statistically similar the Bagoué and Tchologo regions. This could mean that these descriptors are common to the Bagoué and Tchologo regions and could be excluded from a potential future similar study comparing shea trees in these regions. Others descriptors (CRS, TGH, BRP, ALC and IND) were significantly different in the Bagoué and Tchologo regions. These differences were mainly linked to tree growth descriptors and could reveal that shea trees from the two regions differed in their vegetative growth. This type of differentiation is often linked to the farmers' choice to keep the trees on their land, in a crop-dependent manner. These differences could also be explained by the difference in the number of individuals studied in each region (74 individuals in Tchologo; 146 in Bagoué): the small number of modalities increased the differences in their proportions (**Table 4**).

Hierarchical ascending classification based on descriptive morphological characteristics was used to structure the trees and generated three clusters. The low proportion of variance explained by the first three axes is a characteristic of the MCA, which generally gives pessimistic measures of extracted information. The results of this analysis should be treated with caution; however, it identified a profile of shea trees according to their characteristics that could be used to identify and classify SSTs using qualitative descriptors. Diarrassouba (2008) also concluded that the qualitative morphological traits of shea trees could be used to discern classes (27). Our results

also showed no genetic difference in the qualitative traits of Bagoué region trees *versus* Tchologo region trees.

Our study also confirmed that the *in situ* collection is endangered because of a high mortality rate and high *Loranthaceae* infestation (83.64% overall; 85.34% and 81.91% in the Bagoué and Tchologo regions, respectively). Our results corroborate those of (48)), who recorded 87.25% of shea tree infestation by *Loranthaceae* under field conditions. The SSTs of the Kong and Boundiali departments had the lowest infestation rates (63% and 74%, respectively); however, infestation rates above 90% were measured in other departments. Different levels of shea tree infestation by *Loranthaceae* have been found among departments (49). Furthermore, the shea parklands in northern Côte d'Ivoire are infested by two genera of *Loranthaceae*, namely, *Tapinanthus bangwensis* (Engl. And Krause) Danser and *Agelanthus dodoneifolius* (DC. Polh And Wiens) (45,49). *Loranthaceae* seeds are mainly spread by birds that consume their fruit and excrete them in their feces (49). Because shea trees are the major tree species of the West African savanna, they serve as resting refuges for many birds; this might explain their high rate of infestation by *Loranthaceae.* Therefore, *Loranthaceae* represent a major threat to SSTs; yet, the spread of this plant species cannot be controlled, although it is causing significant damage to cultivated and wild trees and shrubs (47,179). To reduce damage, annual control actions must be implemented, during which *Loranthaceae* are manually removed from the *in situ* SST collection. However, the long distances separating individual trees in the collection would require significant human and financial resources to control this biotic stressor. Another solution would be to group individual trees of interest in an *ex situ* collection that could be managed and controlled near a research center in Côte d'Ivoire. This approach would also ensure the sustainability of the collection by protecting it from other constraints that negatively affect survival. Thirty-three SSTs were lost in Bagoué and Tchologo over a 3-year period. Consequently, keeping plant genetic resources *in situ* on farming land poses clear management issues. In five shea parklands in Benin, the number of trees *per* hectare decreased by 0, 9, 12, 16 and 37 trees in Bohicon, Savé, Parakou, Bembéréké and Kandi, respectively, from 2005 to 2017 (Gnanglè, C. P (2017) cited in Lovett and Phillips, 2018) (148). This represents mortality rates of 0%, 34,62%, 44.44%, 39.02% and 75.51%, respectively.

Mortality differed between the Bagoué and Tchologo regions (12% and 5%, respectively) over the 3-year period. Farmers stated different reasons for tree mortality in the two regions. In Tchologo region, mortality was attributed to the uprooting of trees by strong winds. In Bagoué region, multiple reasons were given, including uprooting by wind, uprooting by machines during the installation of the new electrification network in the region, and cutting down of trees by some farmers who considered them to hinder crop growth. Therefore, human actions clearly have a strong impact on the loss of shea tree diversity in Côte d'Ivoire. Cubero (1997) previously highlighted that *in situ* conservation *sensu stricto* is vulnerable. Because farmers have the right to grow what they want, this preference will always be driven by profitable agriculture (180). Consequently, the collection in its *in situ* form is

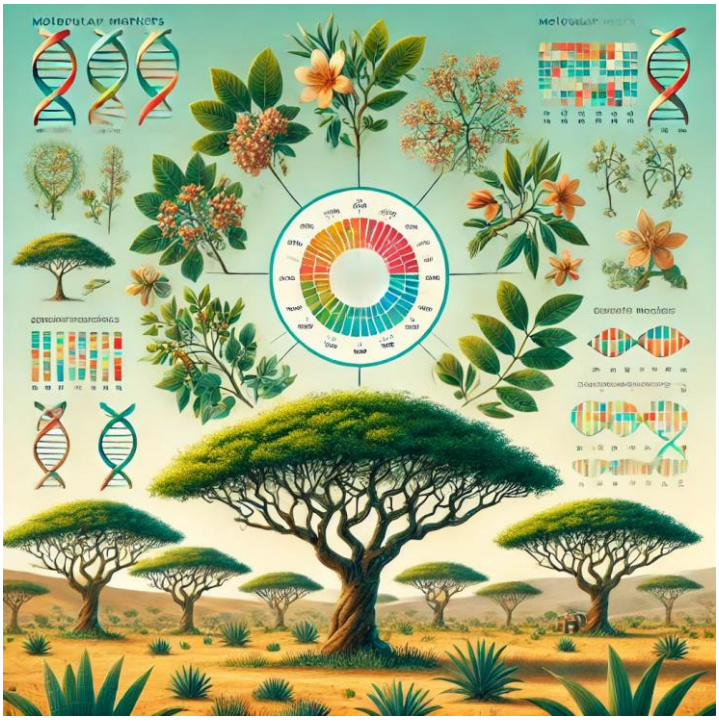
threatened. Therefore, alternative, complementary forms of safeguarding must be implemented that will preserve diversity over the long term. An ideal solution would be to create a core collection in a protected and monitored environment. To accomplish this, an analysis of the genetic diversity of the *in situ* collection is required using single nucleotide polymorphism (SNP) markers to optimize its genetic diversity (181). This genetic diversity analysis will be important for long-term improvement of shea trees and future studies using genomic selection like marker-assisted selection (MAS) and genome-wide association study (GWAS) (25).

3.6. *Conclusions*

This study demonstrates that the qualitative morphological traits of SSTs display high genetic diversity across two regions in Côte d'Ivoire. This important diversity constitutes a heritage that must be preserved for future generations. However, current *in situ* conservation efforts are not sufficient to ensure sustainable conservation because farmers are under pressure to produce crops. Biotic constraints (i.e., *Loranthaceae* infestation on most trees) and abiotic constraints (i.e., wind and human actions) endanger the SST population both in the short and medium term. Monitoring is required to ensure long-term conservation. Based on these results, we recommend the creation of an *ex situ* collection that could be managed by a research center to guarantee sustainability. Furthermore, we recommend selecting a core collection based on the molecular diversity of SSTs to maximize the genetic diversity of this collection.

Chapter 4

Molecular diversity analyses and core collection constitution



Chapter 4. Molecular diversity analyses and core collection constitution

Chapter 3 highlighted the limitations of *in situ* conservation for shea trees. To ensure the preservation of this genetic heritage and address future needs, establishing an *ex situ* collection is essential.

Managing and maintaining such a collection requires significant human and financial resources, which are often limited. Therefore, it is crucial to reduce the number of individuals to facilitate its management.

However, this reduction in the number of genotypes must not come at the expense of genetic diversity, which serves as the foundation for any conservation and breeding program.

Thus, an optimal core collection must meet two fundamental criteria:

1. A reduced size to ensure efficient management.
2. Maximum genetic diversity to preserve its potential for improvement.

In recent years, molecular markers have revolutionized species identification and provided powerful tools for creating a reduced yet genetically representative core collection. Among these markers, Single Nucleotide Polymorphisms (SNPs) stand out for their high precision in measuring genetic diversity.

A key prerequisite for any molecular study is the extraction of high-quality genomic DNA. A literature review and preliminary DNA extraction tests revealed a lack of effective protocols for extracting high-quality shea tree DNA, highlighting a major methodological gap in shea molecular research.

This chapter is divided into two parts:

1. **Development of an optimized DNA extraction protocol for shea tree leaves and roots.** While an efficient protocol was successfully developed, the leaf samples for SNP analysis were ultimately sent directly to the sequencing company, which opted to perform the DNA extraction themselves to ensure full control over the analytical process. Nonetheless, this protocol remains a valuable tool for future shea tree research.
2. **Analysis of genetic structure, diversity and core collection constitution.** Analyses were performed in a population of 333 superior shea trees (SSTs) from the Bagoué, Hambol, Poro, and Tchologo regions, using SNP markers. These regions were selected due to their importance in shea cultivation and their representation of diverse ecological conditions. The core collection were established based on both morphological traits and allelic diversity, ensuring the representation of both genetic and phenotype variations.

Through the integration of molecular tools and conservation strategies, this chapter contributes to a deeper understanding of the genetic landscape of superior shea trees in Côte d'Ivoire. The findings presented here provide a solid foundation for the development of effective breeding programs and the sustainable management of shea

genetic resources, supporting both biodiversity conservation and the economic empowerment of communities reliant on this vital species.

The results of this chapter have been published in:

Attikora AJP, Silué S, Yao SMD, De Clerck C, Shumbe L, Diarrassouba N, Fofana IJ, Alabi T, Lassois L. *An innovative optimized protocol for high-quality genomic DNA extraction from recalcitrant Shea tree (Vitellaria paradoxa, C.F. Gaertn) plant and its suitability for downstream applications*. *Mol Biol Rep* 51, 171 (2024). https://doi.org/10.1007/s11033-023-09098-6

Attikora AJP, Yao SMD, Dago DN, Silué S, De Clerck C, Kwibuka Y, Diarrassouba N, Alabi T, Achigan-Dako EG, 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* 24, 913 (2024). https://doi.org/10.1186/s12870-024-05617-0

4.1. An innovative optimized protocol for high-quality genomic DNA extraction from recalcitrant Shea tree (*Vitellaria paradoxa* C.F. Gaertn) plant and its suitability for downstream applications

This part is an original research article published in Molecular Biology Reports Volume 51; January, 25th 2024 https://doi.org/10.1007/s11033-023-09098-6.

Affi Jean Paul ATTIKORA[1,*], Souleymane SILUÉ[2], Saraka Didier Martial YAO[2], Caroline DE CLERCK[4], Leonard SHUMBE[1], Nafan DIARRASSOUBA[2], Inza Jésus FOFANA[2], Taofic ALABI[2,3], Ludivine LASSOIS[1]

[1]University of Liege, Gembloux Agro Bio-Tech, Terra Research Center, Plant Genetics and Rhizosphere Processes lab. Passage des déportés 2, 5030 Gembloux, Belgium. E-mail: ajpattikora@uliege.be

[2]University of Peleforo Gon Coulibaly (UPGC), Faculty of Biological Sciences, Department of Biochemistry-Genetics, Educational and Research Unit of Genetic, BP 1328 Korhogo, Côte d'Ivoire.

[3]University of Liege, Gembloux Agro Bio-Tech, Functional and Evolutive Entomology. Passage des déportés 2, 5030 Gembloux, Belgium.

[4]University of Liege, Gembloux Agro Bio-Tech, AgricultureIsLife. Passage des déportés 2, 5030 Gembloux, Belgium.

**Corresponding author*: Affi Jean Paul ATTIKORA, E-mail: ajpattikora@uliege.be

4.1.1. Abstract

Background: It is not always easy to find a universal protocol for the extraction of genomic DNA (gDNA) from plants. Extraction of gDNA from plants such as shea with a lot of polysaccharides in their leaves is done in two steps: a first step to remove the polysaccharides and a second step for the extraction of the gDNA. In this work, we designed a protocol for extracting high-quality gDNA from shea tree and demonstrate its suitability for downstream molecular applications. **Methods:** Fifty milligrams of leaf and root tissues were used to test the efficiency of our protocol. The quantity of gDNA was measured with the NanoDrop spectrometer and the quality was checked on agarose gel. Its suitability for use in downstream applications was tested with restriction enzymes, SSRs and RAPD polymerase chain reactions and Sanger sequencing. **Results:** The average yield of gDNA was 5.17; 3.96; 2.71 and 2.41 µg for dry leaves, dry roots, fresh leaves and fresh roots respectively per 100 mg of tissue. Variance analysis of the yield showed significant difference between all tissue types. Leaf gDNA quality was better compared to root gDNA at the absorbance ratio $A_{260/280}$ and $A_{260/230}$. The minimum amplifiable concentration of leaf gDNA was 1 pg/µl while root gDNA remained amplifiable at 10 pg/µl. Genomic DNA obtained was also suitable for sequencing. **Conclusion**: This protocol provides an efficient, convenient and cost effective DNA extraction method suitable for use in various *vitellaria paradoxa* genomic studies.

Keywords: DNA purity; Downstream applications; genomic DNA extraction protocol; Polymerase chain reaction; Sanger sequencing

4.1.2. Introduction

Shea, *vitellaria paradoxa*, is a tropical plant with a thick and resistant leaf. It belongs to the *Sapotaceae* family and grows naturally between eastern Senegal and western Ethiopia across 21 countries (10,167). The shea sector is growing rapidly and represents a major source of income for all stakeholders. Several parts of the shea tree are used, but the most coveted product is shea butter (99). The high global demand for this product in recent decades has generated a lot of interest among stakeholders who are trying to organize shea tree sector in order to make it sustainable and productive. The exported butter is destined for the cosmetic, chocolate and pharmaceutical industries (21) due to its high unsaponifiable matter content (5-7%) and physico-chemical properties similar to those of cocoa butter (129). However, because of its wild character, many shea trees in shea parklands are not good producers. In addition, abiotic and biotic threats have led to the shea tree being declared a vulnerable species (26). Several identifications and core collections creation programs have been initiated in the countries where the plant grows (60). However, in order to create core collections, the identified individuals must be characterized at the molecular level to retain the real diversity.

Obtaining a high quality DNA is a prerequisite for all molecular downstream analyses. Although high-quality DNA extraction from most plants can be carried out according to simple protocols know to all or by the use of universal commercial kits, for some recalcitrant plants the task becomes much more laborious. Several recalcitrant species-specific optimized protocols have been developed to access high-quality DNA of plant species such *Taxus baccata*, *Butela pendula*, *Quercus brantii*, *Eugenia dysenterica*, *Coffea sp*, and *Annona reticulata* (182–187). However, none of the published methods have been shown to be universally applicable to all plant varieties (184,188). Therefore, researchers must modify or combine different protocols to obtain the desired DNA quality depending on the manipulation being considered and the plant material. Varma et al. published a review summarizing the factors influencing extraction efficiency in terms of quantity and quality of obtained DNA from recalcitrant plants (188). These factors are: source material, processing (age and type of tissue, collection and storage and homogenization) and the presence of contaminants (polysaccharides, polyphenols, proteins and RNA and non-nuclear DNA). For almost two decades, shea DNA obtained using different DNA extraction protocols, has been used in molecular studies to assess its genetic diversity. Most of those studies have revealed difficulties with extraction and problems with DNA quality (14,23–25,50,52,123–125,189,190) probably due to the presence of contaminants such as polysaccharides in the DNA. Recurrent problems with the quality of DNA extracted from Shea trees are hampering the development of knowledge about this species using modern analytical techniques such as high-throughput sequencing. A high-quality DNA extraction protocol is thus essential for characterizing and analyzing the genetic diversity of this endangered species.

In this study, we designed a) a cost effective and affordable method for Shea tree high-quality gDNA extraction, based on one-step protocol for root tissues and on a

two-step protocol for leaf tissues: polysaccharides elimination and DNA extraction; b) demonstrated the suitability for use of the extracted DNA in downstream applications such as PCR, molecular marker assays (RAPD, SSR) and sanger sequencing.

This protocol has taken into account root tissues because roots are available all the year compared to leaves and could also be used for mycorrhizal studies.

4.1.3. Materials and methods

4.1.3.1. Plant material

The biological material consisted of:

- Dry root and mature leaf: leaves were collected from mature shea producing trees (>15 years old) in Côte d'Ivoire between June and July 2021 and roots were collected on 24-months-old shea plants grown in laboratory greenhouse and dried with Silica gel;
- Fresh root and mature leaf: leaves and roots were collected on 24-months-old shea plants grown in laboratory greenhouse.

4.1.3.2. DNA Extraction protocols

Three extraction methods were applied to extract DNA from shea leaf and root tissues: the Doyle and Doyle 1987 CTAB protocol (191); the DNA extraction protocol published by, Youssef et al. for recalcitrant plant (192); and a novel two-step innovative protocol for improving gDNA extraction from Shea, hereafter referred to as the "innovative shea DNA extraction protocol". Fifteen samples were used for each protocol.

For DNA extraction from root, the polysaccharides elimination step is not necessary to obtain high quality DNA. Only step 2 of the protocol is required.

4.1.3.3. The Innovative shea DNA extraction protocol

Reagents:

Polysaccharide elimination buffer: 15mM Tris-HCl, 2mM EDTA, 80mM KCl, 20mM NaCl, 0.5% x 100-Tritone, bring the solution to pH = 9, and add 0.1% β – mercaptoethanol

DNA extraction buffer: 0.1M Tris-HCl, 0.05M EDTA, 2M NaCl, 1%(w/v) Polyethylene Glycol 8000 (PEG), 0.5%(w/v) Sodium metabisulfite, 2% (w/v) Polyvinyl pyrrolidone (PVP) and 0.2% (w/v) Polyvinylpolypyrrolidone (PVPP), bring the solution to pH = 8.

Step1: Polysaccharides elimination

Fifty milligrams of mature shea leaves were ground using a mortar and pestle chilled with liquid nitrogen. The ground sample was transferred into a 2 ml Eppendorf tube and 1.8 ml of polysaccharide elimination buffer, adapted from Ky et al. (187) was added and vortexed to mix. The extract was then incubated for 5 minutes at room temperature in an Eppendorf rotator with shaking .The extract was subsequently centrifuged for 10 minutes at room temperature at 5000 rpm and the supernatant was discarded. The process was repeated twice, however, without the incubation step. The

samples were used directly in step 2: "Nuclei Lysis and DNA Extraction" hereafter. Alternatively, they could be stored at -20°C for later or continue directly with step2.

Step2: Nuclei lysis and DNA extraction

The protocol for nuclei lysis and DNA extraction was adapted from the protocol of Youssef et al. (192), after major modifications. To each polysaccharide eliminated sample, 900 µl of the extraction buffer and 100 µl of SDS (20%) was added .The tubes were vortexed vigorously and placed on ice for 5 minutes, after which they were incubate in a water bath at 65°C for 30 minutes and then placed back on ice for 5 minutes. To each sample tube, 400 µl of 5 M potassium acetate (pH 8) was added, mixed gently and placed on ice for another 5 min. The samples were centrifuged at 14000 rpm for 15 min at 10°C and the supernatant was transferred to a new 2 ml tube. 5 µl of RNAse A (10 mg/ml) was added to each tube and incubated at 37°C for 30 minutes. To precipitate the DNA, an equal volume of cold isopropanol was added to each tube, mixed by gently inverting the tubes, and incubate at -20°C for 15 minutes. The samples were then centrifuge at 14000 rpm for 15 min at 4°C. The supernatant was discarded and the pellet was washed twice with 500 µl of cold 70% ethanol at 14000 rpm for 3 minutes. The pellet was allowed to completely dry and dissolved in 600 µl of TE buffer. The suspension was again centrifuged at 14000 rpm for 10 min at 10°C and the supernatant was transferred to a new 1.5 ml tube. Equal volume of cold isopropanol and volume of 30 µl of sodium acetate (3 M, pH 5.2) were added to the supernatant and the precipitation steps were repeated. After centrifugation, the pellet was allowed to completely dry, then dissolved in 100 µl of ddH_2O.

4.1.3.4. DNA quantity and quality evaluation

Electrophoresis

Genomic DNA presence and quality were evaluated for the three methods by electrophoresis on 1% agarose gel using 1X TAE (Tris-base, acetic acid, EDTA) buffer. Gelred was used as the intercalating agent at a final concentration of 0.1 µl/mL.

Spectrophotometer evaluation for the innovative Shea DNA extraction protocol

The quantity and quality of DNA extracted from leaf and root samples using the "innovative shea DNA extraction protocol" were evaluated at 230, 260 and 280 nm wavelengths using a spectrophotometer (NanoDrop, Thermo Scientific Type UV1, England) (193). The yield of genomic DNA from 100 mg of tissue was estimated according to the following equation:

$$y = 2x * v,$$

where y is the estimated yield, x represents the quantity of gDNA converted into micrograms per microliter (µg/µl) and v represents the volume of ddH_2O used to dilute the gDNA.

Optimized dilution test

Twenty samples of 10 ng/µl DNA obtained from leaf and root tissues using our extraction protocol, were used for the dilution test. This test allowed to determine the smallest amplifiable DNA concentrations from the respective tissues, for molecular studies. Five 10-folds serial dilutions (1; 0.1; 0.01; 0.001; 0.0001ng/µl) dilutions were

made from the initial dilution (10 ng/µl). Quality and the intensity of the PCR product from the diluted DNA were checked on agarose gel.

Downstream applications

To determine the suitability of gDNA for downstream applications, several tests such SSR and RAPD PCR amplification, restriction enzymatic and Sanger sequencing were performed.

SSR markers amplification

Five SSR markers (*mCIRvp8, mCIRvp28, mCIRvp70, mCIRvp134, mCIRvp159*) of *Vitellaria paradoxa* identified by Allal et al. (189) were used to verify the quality of the extracted DNA for use in PCR reaction. The Sequences of the PCR primers used and their characteristics are shown in **Table 6**. PCR amplification was performed in 10 µl total volume: 5 µl of Go Taq Green Master Mix (2X), 0.5 µl (10 pmol/µl) of each primers F and R, 1 µl of DNA template (10 ng/µl) and 3 µl of sterile distilled water. The PCR run program was carried out under the following conditions: Denaturation of DNA at 94 °C for 4 min, followed by 30 cycles of amplification 94 °C for 30 s, 45 s at the primer's optimized annealing temperature (**Table 6**), and 72 °C for 1 min and a final extension at 72 °C for 5 min. The presence or absence of the amplified marker was checked by electrophoresis on 1% agarose gel ran at 180 V for 20 min in 1X TAE (Tris-base, acetic acid, EDTA) buffer. The gel was stained with 0.1µl/mL gelred for DNA visualization.

Table 6. List of the SSR and RAPD primers and characteristics

SSR				
Locus	**Primer sequence**	**Tm (°C)**	**Repeat motif**	**Allele size range (bp)**
mCIRvp8	F : 5'AATTCATTGGAGGACAGCA 3'	56.1	(TG)12	310-316
	R : 5'ACACCAATCGCAACACAG 3'			
mCIRvp28	F : 5'ATTGTTAGTTATGGTTTTGG 3'	50.0	(CA)6	162-170
	R : 5'TGATTTGCTATTTTGCTTAC 3'			
mCIRvp134	F : 5'CTCTTCTCCTCCCCTTCAAC 3'	50.0	(TG)10	292-298
	R : 5'ACCATAATCCCTCAGCAATC 3'			
mCIRvp159	F : 5'CACGAAGAAATATGCTG 3'	56.1	(TG)5,(GT)6,(GA)10, AA(GA)3	240-268
	R : 5'ATGGATTGCTTTAGGTG 3'			
mCIRvp70	F : 5' TGCCCAACAAGAGAGTC 3'	50.0	(AG)13	254-264
	R : 5' ACCTCATAAAATCCCCAC 3'			
RAPD				
Locus	**Primer sequence**	**Tm (°C)**		
OPB7	5'GGT GAC GCA G 3'	34		
OPB10	5'CTG CTG GGA C 3'	34		
OPB17	5'AGG GAA CGA G 3'	32		
OPC8	5'TGG ACC GGT G 3'	34		
OPF12	5'ACG GTA CCA G 3'	32		

Tm(°C): Annealing temperature, F: forward, R: Reverse

Shea tree gDNA restriction digestion analysis

Two hundred nanograms of gDNA from fresh and dry tissues were digested using 5U each of two enzymes, *Hin*dIII (NEB #R3104) and *Eco*RI (NEB #R3101), at 37°C for 4 hours, according to the manufacturer's protocol (NEB, USA) and results was visualized on 1% agarose gel.

RAPD markers amplification

Five RAPD markers were used to test the amplification efficiency of our DNA. PCR amplification was carried out in 15 µl total volume: 7.5 µl of Go Taq Green Master Mix (2X), 1 µl of the RAPD single primers (Table 5) (10 pmol/µl), 1µl of DNA template and 5.5 µl of sterile ddH2O. The PCR program was performed under the following conditions: Denaturation: 4 min at 94°C followed by 45 cycles [Denaturation 1 min at 94°C, Fixation 1 min at 32 or 34°C depending on the primer, Elongation 2 min at 72°C], terminal elongation for 10 min at 72°C and holding time at 4°C. The amplified fragments were separated on a 1.5% agarose gel (Gel Doc XR+ system; BIO RAD model Universal Hood II).

Sanger sequencing

Five genomic DNA samples from leaf tissue with an average concentration of 10 ng/µl were Sanger sequenced using 4 SSRs markers; *mCIRvp8, mCIRvp70, mCIRvp134 and mCIRvp159* (**Table 6**), as sequencing primers.

4.1.3.5. Data analysis

The data were analyzed by R software version 4.2.2. Analysis of variance was used to test the difference between the mean concentrations and the ratio $A_{260/280}$ and $A_{260/230}$ of the different tissues. Post-hoc pairwise multiple comparisons were performed to determine treatments that were significantly different at the 95% using Tukey HSD test.

The sequences obtained by Sanger sequencing were used for sequence alignment in NCBI with the program BLASTn.

4.1.4. Results

4.1.4.1. DNA extraction protocols comparison

The results obtained on the agarose gels show that shea genomic DNA was extracted using all protocols. DNA bands of sizes above 10 kb were observed on the agarose gel for samples from all DNA extraction methods (**Figure 17a, b and c**). However, the genomic DNA obtained from the protocols of CTAB and Youssef et al. (**Figure 17a** and **Figure 17b** respectively) were degraded gDNA (smears) compare to our protocol showing high molecular weight gDNA, intense bands and clear gel wells (**Figure 17c**). By testing our "innovative shea DNA extraction protocol" on young shea tree leaves, we came with the same results as mature leaves without the step 1 (elimination of polysaccharides) of the protocol (data not shown here).

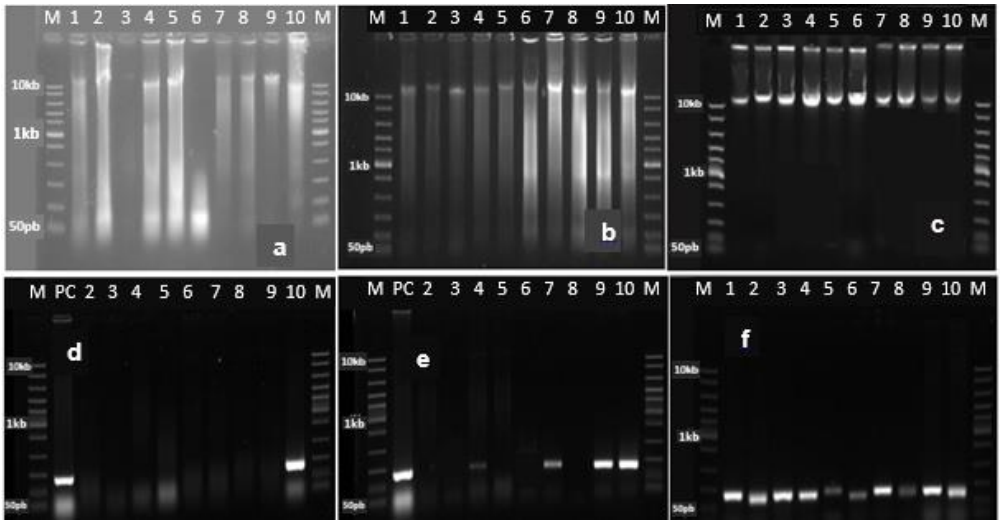


Figure 17 Electrophoresis analysis of gDNA obtained from different protocols. fresh root (wells 1 to 3), fresh leaf (wells 4 to 6), dry leaf (wells 7 and 8) and dry root (wells 9 and 10) tissues using the (a) the CTAB protocol, (c) Youssef et al. protocol and (c) our "Shea DNA extraction protocol". M = Molecular weight markers (*Fast DNA ladder*) and SSR markers PCR amplification of DNA extract with (d): CTAB protocol; (e): Youssef et al., (2015) protocol; (f): "Shea DNA extraction protocol"

Furthermore, Polymerase chain reaction (PCR) using SSR markers was performed on the gDNA from the different protocols. No amplification was observed in the PCR reactions using DNA obtained from the CTAB protocol (**Figure 17d**), only three out of eight gDNA samples obtained using the protocol of Youssef et al. were amplified in the PCR reaction (**Figure 17e**), whereas all gDNA obtained using our "innovative shea DNA extraction protocol" were amplified (**Figure 17f**).

4.1.4.2. DNA quality and yield performances of the "innovative shea DNA extract protocol"

***Quality**:*

The purity of shea gDNA isolated from root and leaf tissues was checked with the NanoDrop spectrophotometer at different wavelengths (230, 260 and 280). The absorbance ratios at $A_{260/280}$ and $A_{260/230}$ were then estimated (**Table 7**).

Results showed that $A_{260/280}$ ratios of DNA from the root tissues were low (fresh root: 1.35 and dry root: 1.46) whereas those from the leaf tissues were high (fresh leaf: 1.75 and dry leaf: 1.63). Analysis of variance showed a significant difference in $A_{260/280}$ ratio between tissue types (F: 65.2; $p < 0.001$). Tukey's HSD post hoc test revealed that all means are different (**Table 7**).

The ratio $A_{260/230}$ was 1.97; 1.79; 1.8 and 1.89 for fresh leaf, fresh root, dry leaf and dry root respectively. This ratio $A_{260/230}$ showed no significant differences (F:0.03; $p = 0.87$) between tissue types (**Table 7**).

Quantity:

The results showed a higher average yield of gDNA in the dry leaf samples with 5.17µg/100mg of tissue. Dry root and fresh leaf average gDNA yields was 3.96µg/100mg and 2.71µg/100mg tissue, respectively. The smallest average yield of gDNA was in fresh root with 2.42 µg/100mg of tissue (**Table 7**). The ANOVA of the average yields showed a significant difference between tissue types (F: 153.9; p<0.001).

Table 7. Shea tree tissue types, gDNA concentration and A260/280 and A260/230 ratios

Quantity							
Tissue Type	**Number of samples**	**Min (ng/µl)**	**Max (ng/µl)**	**Mean (ng/µl)**	**Yield (µg/100mg)**	***F***	***P***
FL	45	8.33	18.07	13.53±2.65^{a}	2.71±0.41^{a}	153.9	*P*<0.001***
DL	54	20.28	35.62	25.84±2.86^{b}	5.17±0.9^{b}		
FR	45	9.76	14.91	12.08±1.65^{c}	2.42±0.35^{a}		
DR	52	14.27	24.72	19.8±2.66^{d}	3.96±0.86^{c}		
Quality							
Tissue type	**Number of samples**	**$A_{260/280}$ ± SE**	***F***	***P***	**$A_{260/230}$ ± SE**	***F***	***P***
FL	45	1.75±0.17^{d}	65.2	*P*> 0.001***	1.97±0.4^{a}	0.03	0.87
DL	54	1.63±0.26^{c}			1.8±0.72^{a}		
FR	45	1.35±0.14^{b}			1.79±0.51^{a}		
DR	52	1.46±0.14^{a}			1.89±0.22^{a}		

*FL: Fresh Leaf, DL: Dry Leaf, FR: Fresh Root and DR: Dry Root, Min: minimum, Max: maximum. Means that have the same letter are not statistically different, ***: high significant difference, SE = standard deviation.*

Optimized dilution of the gDNA

The dilution of gDNA from the root showed optimal amplification at the concentration 1ng/µl. Indeed, at this concentration all samples were amplified with intense and well visible bands. Concentrations 0.1ng/µl and 0.01ng/µl were amplified but with sometimes less intense bands (**Figure 18c**). However, at the concentration 1pg/µl, there was no amplification.

For leaf tissue DNA, from the initial dilution after extraction to the 10^4 folds dilution, the gDNA was amplified. The best bands were found at concentrations 10; 1; 0.1 and 0.001ng/µl (between initial dilution and 10^3 dilution). However, the 10^4 dilution (1pg/µl) showed a clear band but the intensity of the band was low. The concentration 0.1pg/µl did not show a band (**Figure 18b**).

Efficiency for downstream applications

Analysis of gDNA quality by enzymatic digestion

The gDNA of dry and fresh leaf and root from shea tree was checked by a restriction digestion using *Hind*III and *Eco*RI. Extensive smears were obtained in wells containing gDNA and restriction enzymes whereas wells containing only gDNA did not show degradation (**Figure 18d**).

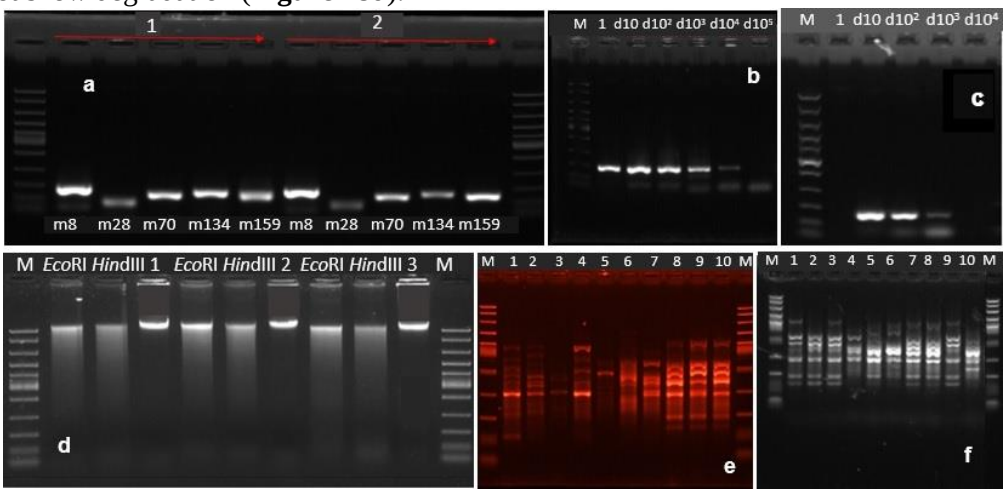


Figure 18. Electrophoresis gel of amplified SSR, RAPD makers and enzymatic digestion. a) PCR amplification of SSR markers of the gDNA of Shea tree. b and c) Efficiency for PCR amplification of shea tree gDNA after dilution (b: leaf sample and c: root sample), d) Digestion of gDNA by the restriction enzymes EcoRI and HindIII. 1; 2 and 3 are undigested gDNA, **e** and **f)** RAPD amplification of the gDNA of Shea tree. e: root samples and f: leaf samples; M: DNA fast Ladder.

Analysis of gDNA quality by PCR

Polymerase chain reaction (PCR) was used to target different microsatellites loci from *Vitellaria paradoxa* gDNA. All samples from leaf tissue (Fresh and Dry) were successfully amplified. PCR showed microsatellites genes fragment with a clear and sharp band (**Figure 18a**). However, PCR performed on root gDNA gave only 10% amplification with initial dilution.

RAPD amplification test

Optimized gDNA concentration was used to amplify RAPD markers on root and leaf DNA. The results showed clear and intense bands for all samples (**Figure 18e** and **Figure 18f**).

Sanger sequencing

The sequences obtained after sequencing and blasted in NCBI confirmed that they belonged to the *vitellaria paradoxa* species. Indeed, Sequences only produced significant alignments to shea tree SSRs markers described by Allal et al (2008) in NCBI GenBank (**Table 8**). Identity percentage varied from 83.47 to 99.21% with an E-value from 3e-55 to 3e-121 and a query cover from 94 to 99% for the four SSR markers used in this study (**Table 8**). These results confirm clearly that gDNA obtained from the "innovative Shea DNA extraction protocol" is suitable for sequencing.

Table 8. Sequence alignment characteristics of our samples with *vitellaria paradoxa* species in NCBI

Samples	Characteristics	*mCIRVP8*	*mCIRVP70*	*mCIRVP134*	*mCIRVP159*
A	Query Cover	97%	98%	**99%**	95%
	E-Value	3e-64	1e-67	1e-87	1e-61
	Identity	**83.47%**	92.97%	94.64	92%
B	Query Cover	97%	97%	95%	95%
	E-Value	**3e-121**	1e-86	1e-93	2e-57
	Identity	**99.21%**	96.08%	94.35%	87.05%
C	Query Cover	98%	98%	97%	**94%**
	E-Value	3e-115	6e-84	5e-79	**3e-55**
	Identity	97.62%	95.63%	91.77%	87.08%
D	Query Cover	97%	95%	98%	*NA*
	E-Value	7e-110	1e-66	4e-93	*NA*
	Identity	97.18%	89.45%	95.58%	*NA*
E	Query Cover	98%	98%	96%	96%
	E-Value	1e-112	3e-87	3e-95	2e-58
	Identity	97.59%	97%	95.54%	86.15%

Values in bold in the table represent the highest and lowest values of the blast results in NCBI. *NA: Not available.*

4.1.5. Discussion

Due to the high polysaccharide content of the shea leaf, it is recalcitrant to direct DNA extraction with CTAB (191) or MATAB. Direct extraction with these buffers resulted in a gelatinous pellet with DNA signifying the presence of polysaccharides. Downstream applications (PCR and enzymatic digestion) using the genomic DNA extracted with those methods appeared limited.

In order to obtain high gDNA quality from recalcitrant shea tree, three extraction methods were compared. The results displayed on agarose gel showed that the quality of the gDNA obtained from the protocols of CTAB and Youssef et al. were inferior to the quality of the gDNA from our protocol, as shown by the degraded DNA (smears) in **Figure 17a, Figure 17b** and **Figure 17c** respectively. The presence of smear is a sign of degradation of the extracted DNA which could easily affects the quality of the subsequent molecular applications (184). These results indicate that our protocol provides high gDNA quality without degradation.

Furthermore, all gDNA from our protocol were amplified by PCR whereas the others were not. This trend was observed in PCR on *Allium* and *Helianthus* species gDNA obtained using different DNA extraction methods (194). Polymerase chain reaction (PCR) with SSR markers performed on the genomic DNA from the different protocols indicated that the extraction method had an influence on the PCR amplification success (195). So, our protocol allows to obtain better quality of gDNA from shea compared to the protocol of Youssef et al. (192) and the CTAB protocol (191). Based on these results, our protocol "innovative shea DNA extraction protocol" was retained for the further parts of this study.

The purity of the gDNA from leaf and root obtained by our protocol was assessed. The absorbance ratio $A_{260/280}$ of leaf's gDNA was high (dry leaf: 1.63 ± 0.26 and fresh leaf: 1.75 ± 0.17), while that of root was low (fresh root: 1.35 ± 0.14 and dry root: 1.46 ± 0.14). Rezadoost et al. obtained equal $A_{260/280}$ ratio for root and leaf DNA in Betula and Grape species (185). In contrast, Ahmadi et al. obtained high $A_{260/280}$ ratio of root genomic DNA compare to leaf's genomic DNA in Persian oak species (186). The low ratio $A_{260/280}$ of the root in our study could indicate the likely presence of proteins in the DNA extracted. This could be solve by adding protease to degrade DNA-associated proteins and other cellular proteins (196). However, although the ratio $A_{260/280}$ value of 1.8 indicates high purity of the DNA, other studies have demonstrated that DNA with $A_{260/280}$ ratio between 1.3 and 1.8 can be amplified (188,197).

The ratio $A_{260/230}$ was 1.97; 1.79; 1.8 and 1.89 for fresh leaf, fresh root, dry leaf and dry root respectively and did not show significant effects on this purity parameter of the gDNA between tissue types. Stulnig et al. (193) reported that the absorbance at 230 nm should be half the value at 260 nm to give a ratio $A_{260/230}$ of nearly 2 and a significant absorbance at 230 nm indicates contamination by phenolate ions, thiocynates and other organic compounds. Fresh shea leaf tissue in this study gave similar values to the standards DNA quality at both $A_{260/280}$ and $A_{260/230}$ ratios (1.75 and 1.97 respectively). Our protocol was effective in yielding higher quality DNA in shea. Similar range of quantity and quality were observed with DNA extracted from a panel of Cerrado plant species and Bryophytes species (198,199). The high quantity of genomic DNA obtained in the dried tissues compared to the fresh tissues can be explained by the higher quantity of dry matter in the dried tissues.

By using restriction enzymes such as *Hin*dIII and *Eco*RI on the gDNA, extensive smears were observed, showing that the gDNA was digested, and could be used for Southern Blot analysis (**Figure 18d**). Therefore, the quality and purity of the gDNA is sufficient for use in downstream restriction enzymatic reactions, such as restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (ALFP) (192,197,200,201).

Successful PCR amplification of the leaf samples indicated that the gDNA was of high quality at the initial concentration (10ng/µl). However, we obtained 100% of amplification of root gDNA when diluted 10 folds (1ng/µl) and could indicated that the concentration of inhibitors decreases when the gDNA is diluted, which allows the primers to reach their binding site easily. Failure of amplification at initial dilution

could be due to the presence of proteins, as indicated by the low $A_{260/280}$ ratio (1.34) . The sample type had a high effects on the PCR amplification success. Our results were consistent with results obtained by others authors who observed that the sample type had a higher influence on the PCR amplification success than the choice of the extraction method (194,195). At the 1pg/µl concentration, there was amplification for leaf gDNA, while no amplification was observed for root gDNA at the same concentration. This indicates that the concentration of DNA used in PCR depends on the type of tissue, the quantity and quality of DNA as well as the amount of secondary compounds remaining in the DNA extract (195). It should be noted that the electrophoresis conditions applied in this work were not suitable for separating SSR fragments. However, this has no negative impact on the value of the results presented in this study, since the objective of this part was not to genotype the samples.

All gDNA optimized concentration from leaf and root samples were well amplified with RAPD markers, showing polymorphic, sharp and scorable bands. Similar results were shown in other efficient protocols for gDNA extraction on recalcitrant plant species (185,202). Our protocol can therefore be used to isolate high-quality DNA from shea tissues, suitable for hybridization studies using assisted selection molecular markers such as RAPD markers (202).

The results obtained from Sanger sequencing confirm clearly that gDNA from the "innovative Shea DNA extraction protocol" is of high quality and suitable for sequencing, as shown by sequencing data highly matching the genomic sequence of shea in NCBI. High concentration of NaCl (2M) combined with the used of PVP (2%) directly in the extraction buffer which efficiently remove polysaccharides and polyphenol and DNA precipitating with isopropanol and sodium acetate help to obtain gDNA in an intact form (182,202).

4.1.6. Conclusion

With this study we provide a protocol for the extraction of high quality of gDNA from recalcitrant shea to be employed in characterizing genetic diversity in shea tree population. The leaf gDNA was of higher quantity and quality than the root gDNA. Sanger sequencing of the SSRs markers confirm the high quality of the genomic DNA obtained with our "Innovative DNA extract protocol" through the high identity to the sequence of *Vitellaria paradoxa* species in NCBI. This protocol could be extended to other plant species where high polysaccharide content hinders good quality DNA extraction.

4.2. 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

This part is an original research article published in BMC Plant Biology Volume 24; October, 1st 2024 https://doi.org/10.1186/s12870-024-05617-0.

Affi Jean Paul ATTIKORA*[1], Saraka Didier Martial YAO[2,3], Dougba Noel DAGO[2], Souleymane SILUE[2], Caroline DE CLERCK[4], Yves KWIBUKA[5], Nafan DIARRASSOUBA[2,3], Taofic ALABI[2,6], Enoch G. ACHIGAN-DAKO[7], Ludivine LASSOIS[1]

1.University of Liege, Gembloux Agro Bio-Tech, Terra Research Center, Plant Genetics and Rhizosphere Processes lab. Passage des déportés 2, 5030 Gembloux, Belgium. E-mail: ajpattikora@uliege.be

2.University of Peleforo Gon Coulibaly (UPGC), Faculty of Biological Sciences, Department of Biochemistry-Genetics, Educational and Research Unit of Genetic, BP 1328 Korhogo, Côte d'Ivoire.

3. African Center for Shea Research and Application (CRAK), Korhogo, Côte d'Ivoire.

4. University of Liege, Gembloux Agro Bio-Tech, AgricultureIsLife. Passage des déportés 2, 5030 Gembloux, Belgium.

5. Université Catholique de Bukavu (UCB), Faculté des sciences agronomiques, département de production végétale. Bukavu, D.R. Congo.

6. University of Liege, Gembloux Agro Bio-Tech, Functional and Evolutive Entomology. Passage des déportés 2, 5030 Gembloux, Belgium.

7. Genetics, Biotechnology, and Seed Science Unit (GBioS), Department of Plant Sciences, Faculty of Agronomic Sciences, University of Abomey-Calavi, 01 BP 526, Abomey-Calavi, Benin.

* Corresponding author: Affi Jean Paul ATTIKORA; affijean0121@gmail.com, ajpattikora@uliege.be

4.2.1. Abstract

Background: The shea tree is a well-known carbon sink in Africa that requires a sustainable conservation of its gene pool. However, the genetic structure of its population is not well studied, especially in Côte d'Ivoire. In this study, 333 superior shea tree genotypes conserved *in situ* in Côte d'Ivoire were collected and genotyped with the aim of investigating its genetic diversity and population structure to facilitate suitable conservation and support future breeding efforts to adapt to climate change effects. **Results**: A total of 7,559 filtered high-quality single nucleotide polymorphisms (SNPs) were identified using the genotyping by sequencing technology. The gene diversity (HE) ranged between 0.1 to 0.5 with an average of 0.26, while the polymorphism information content (PIC) value ranged between 0.1 to 0.5 with an average of 0.24, indicating a moderate genetic diversity among the studied genotypes. The population structure model classified the 333 genotypes into three genetic groups (GP1, GP2, and GP3). GP1 contained shea trees that mainly originated from the Poro, Tchologo, and Hambol regions, while GP2 and GP3 contained shea trees collected from the Bagoué region. Analysis of molecular variance (AMOVA) identified 55% variance within populations and 45% variance within individuals, indicating a very low genetic differentiation (or very high gene exchange) between these three groups (FST = 0.004, gene flow Nm = 59.02). Morphologically, GP1 displayed spreading tree growth habit, oval nut shape, higher mean nut weight (10.62 g), wide leaf (limb width = 4.63 cm), and small trunk size (trunk circumference = 133.4 cm). Meanwhile, GP2 and GP3 showed similar morphological characteristics: erect and spreading tree growth habit, ovoid nut shape, lower mean nut weight (GP2: 8.89 g; GP3: 8.36 g), thin leaf (limb width = 4.45 cm), and large trunk size (GP2: 160.5 cm, GP3: 149.1 cm). A core set of 100 superior shea trees, representing 30% of the original population size and including individuals from all four study regions, was proposed based on allelic diversity using the "maximum length sub-tree function" in DARwin v. 6.0.21. **Conclusion**: These findings provide new knowledge of the genetic diversity and population structure of Ivorian shea tree genetic resources for the design of effective collection and conservation strategies for the efficient use of inbreeding.

Keywords: *Vitellaria paradoxa*, single nucleotide polymorphisms (SNPs), population structure, genetic diversity, DArTseq, analysis of molecular variance (AMOVA), core collection.

4.2.2. Introduction

The shea tree (*Vitellaria paradoxa*, C. F. Gaertn) is an African native plant species. It has been recorded in 21 countries across semi-arid Sub-Saharan Africa within a wide belt of more than 3.4 million km (10). It supports an estimated 16.2 million of shea nut collectors (10). *V. paradoxa* is distributed from Senegal in the West to Uganda in the East (13). The genus *Vitellaria* comprises a single species belonging to the Sapotaceae family, and it includes two subspecies: subsp. *nilotica*, primarily found in East Africa, and subsp. *paradoxa*, which occurs in West Africa from Senegal to the Central African Republic (72,91).

In Côte d'Ivoire, shea trees are predominantly found in the northern region. This species exhibits monoecious characteristics and possesses a bisexual reproduction pattern that makes it rely primarily on insect-mediated cross-pollination. However, the plant has also hermaphroditic flowers capable of self-pollination (13). The dissemination of the shea fruits is primarily achieved through barochory, with secondary dispersal by various animals such as birds, monkeys, rodents, and even humans (14).

Shea tree fruits have been considered as a significant source of economic benefit in a number of countries within the semi-arid savanna regions of West Africa. In fact, dating back to medieval times, their kernels have been used in the production of an important primary derivative: shea butter (203,204). Additionally, due to the presence of edible fatty acids that are used in the food, cosmetic, and pharmaceutical industries, shea butter holds a prominent position as a multimillion-dollar export commodity (20,22,167).

Despite its importance, the International Union for Conservation of Nature (26) has classified the shea tree as one of the plant species facing significant threats and experiencing vulnerability. It is confronted with extensive degradation, which is primarily driven by activities such as charcoal production and frequent, uncontrolled bushfires. Furthermore, factors like population growth, which reduces the duration of fallow periods, and the systematic collection of fruit from beneath shea trees by local communities, pose significant obstacles to the natural regeneration of the species (72). In addition, the overexploitation of shea trees and the expansion of new, more profitable agricultural crops such as cashew plantations in northern Côte d'Ivoire are contributing to the decline in shea tree densities.

To safeguard the genetic resources of *Vitellaria paradoxa* in West Africa, national and regional strategies (67) for the identification and preservation of local shea tree varieties have been promoted (14). Several additional initiatives aimed at identifying and conserving shea tree resources have also been launched to address these concerns (60,61).

In Côte d'Ivoire, superior shea trees have been identified and rereferred in several regions based on a participatory survey. These trees constitute the *in situ* collection of shea trees in the country (60). The genetic diversity and structure of this population have been not studied. In addition, a recent study on the morphological traits and sustainability of part of these trees demonstrated that the superior shea trees conserved

on farmers' lands are threatened because of biotic and abiotic pressures (205). Consequently, establishing a core collection based on the genetic diversity of superior shea trees will be helpful for suitable conservation and management (205).

Studying the genetic diversity and population structure is important for designing effective conservation and breeding programs (206), as well as for characterizing the natural selection history and genetic relationships of *V. paradoxa* (25).

Several authors have mentioned molecular markers such as random amplified polymorphic DNA (RAPDs) (125,50,207) and single sequence repeats (SSRs) (123,23,189,14,124,52,53,190,207) in studies of shea tree species. Most of these studies have used these molecular markers to access the genetic diversity and population structure of shea tree species. Recently, single nucleotide polymorphism (SNP) markers were applied to study the genetic diversity and population structure of the Ugandan *nilotica* subspecies of the shea tree (54).

The approach of establishing core collections has emerged to increase the efficiency of the conservation and use of plant genetic resources while preserving the genetic diversity of the entire collection as much as possible (208,209). For plant species with recalcitrant seeds, the establishment of a core collection is the most suitable and cost-effective alternative method for *in situ* conservation of their genetic resources (210).

In the present study, the genotyping by sequency technology was used to genotype a panel of 333 superior *V. paradoxa* trees from the four key shea production regions in Côte d'Ivoire. The objectives were: (1) to characterize the genetic diversity and population structure of the shea tree using SNP data; (2) to characterize the genetic differentiation among and within Ivoirian shea populations; and (3) to establish a core collection for suitable conservation and management. This study is the first to use SNP markers to assess the genetic diversity and population structure in a *V. paradoxa* subspecies of Côte d'Ivoire. It lays a foundation for the effective conservation of shea trees and future genome wide association studies (GWAS) in shea tree breeding programs.

4.2.3. Methods

4.2.3.1. Plant Material

From an initial collection of 1,200 superior shea trees previously identified by a shea breeding program from Côte d'Ivoire now known as the Centre Africain de Recherches et d'Applications sur le Karité (CRAK, the African Center for Shea Research and Application), 333 were randomly selected based on geographical distribution and population density. Superior shea trees were identified as described in a previous study (205). A participatory survey with farmers allowed for the selection of superior trees based on criteria such as high fruit yield, sweet taste of the fruit pulp, large fruit size, early flowering every year, and periodicity of fruit production (205). These genotypes are being conserved *in situ* in four northern regions of Côte d'Ivoire: Bagoué, Hambol, Poro, and Tchologo (**Figure 19**). To obtain samples, two mature leaves from the lower part of the tree were collected from each individual shea tree between May and July 2020 and 2021. The collected leaves were immediately dried using silica gel and then stored at 4°C, awaiting DNA extraction.

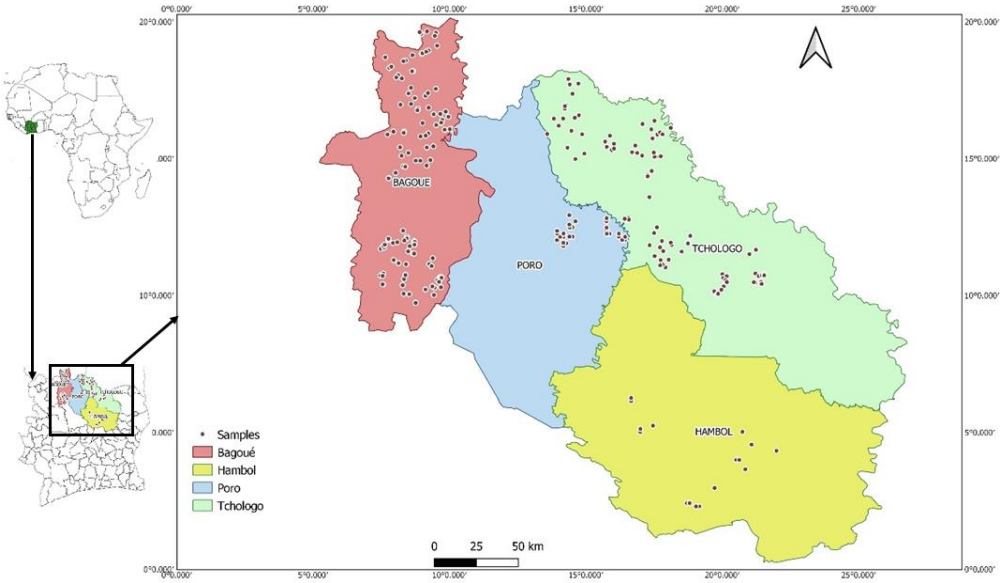


Figure 19. Spatial distribution of the 333 superior shea trees from northern Côte d'Ivoire (163 in Bagoué, 53 in Poro, 97 in Tchologo and 20 in Hambol); the colors represent the different regions (red for Bagoué, blue for Poro, green for Tchologo, and yellow for Hambol); the dots represent the samples

The savannas of northern Côte d'Ivoire, where shea trees grow, are divided into two main zones: the Sudanese and sub-Sudanese savannas, based on climatic factors and differences in vegetation (205). The Sudanese savanna (Bagoué, Poro, and Tchologo regions) corresponds to the main production zone of the shea tree with a monomodal rainfall pattern (1,200 mm/year) (174,211); the sub-Sudanese savanna (Hambol region) corresponds to a transitional production zone with a bimodal rainfall pattern (1,050 mm/year) (211,212). The average annual temperature is around 27°C. The vegetation is Sudano-Guinean and consists of wooded savanna and grassy savanna with scattered gallery forests, particularly along waterways (175). The pedology of this zone is characterized by three subclasses of ferralitic soils: soils on basic rocks, tropical ferruginous soils and hydromorphic soils (211). The main crops grown in these regions are cotton, cashew nuts and mangoes. The tree and shrub species commonly found in the study area are *Vitellaria paradoxa, Paria biglobosa, Pilliostigma thonningii*, *Parinari curratellifolia*, *Terminalia avicennioides* and *Ficus sciarophylla* (213).

4.2.3.2. DNA Extraction and Sequencing

The collected samples were sent to SEQART AFRICA, located at the International Livestock Research Institute (ILRI) in Nairobi, for genotyping. DNA extraction was conducted using the NucleoMag plant DNA extraction kit (Takara Bio USA), following the manufacturer recommendations. The concentration of extracted genomic DNA varied between 50 and 100 ng/µl. DNA integrity was checked on 0.8% agarose gel loaded. Libraries were constructed according to the DArTseq complexity

reduction method through digestion of the genomic DNA using a combination of *PstI* and *MseI* restriction enzymes and ligation of barcoded adapters and the common adapter, followed by PCR amplification of adapter-ligated fragments (214). Libraries were sequenced using single read sequencing runs for 77 bases. The sequencing was carried out using the Illumina HiSeq 2500 system.

SEQART AFRICA uses genotyping by sequencing DArTseq technology, which provides rapid, high-quality, and affordable genome profiling, even from the most complex polyploid genomes. DArTseq markers scoring was achieved using DArTsoft14, an in-house marker scoring pipeline based on algorithms. DArTseq SNP markers were scored as binary for the presence or absence (1 and 0, respectively) of the restriction fragment with the marker sequence in the genomic region of the corresponding sample.

SNP markers were aligned to the Vitpa_HiCP0_Assembly reference genome freely accessible online at https://bioinformatics.psb.ugent.be/orcae/overview/Vitpa to locate their corresponding chromosome positions.

4.2.3.3. SNP Marker Filtering

Two criteria were used to discard low-quality SNP markers and ensure data integrity. First, regarding the proportion of missing data (>20%), SNP markers with missing data exceeding 20% were excluded from the dataset. Second, regarding the minor allele frequency (MAF), SNPs with a minor allele frequency of less than 5% were considered rare and were therefore discarded.

The next step consisted of selecting the SNPs with substantial information content for further analysis. Specifically, a threshold based on the polymorphism information content (PIC) value equal to or greater than 0.1 was established. In addition, only biallelic SNP markers were kept for this study.

After this rigorous quality control process, a dataset consisting of 7,559 SNP markers and the 333 superior shea tree genotypes were considered for further analyses. This stringent filtering ensured that our dataset was of high quality and suitable for robust genetic analysis.

4.2.3.4. Genetic Properties of Markers

A custom Perl script was used to compute the allele counts and allele frequencies from the selected SNPs. Furthermore, to estimate markers-associated statistics such as observed heterozygosity (HO), expected heterozygosity (HE) and minor allele frequency (MAF), the GenAlEx version 6.503 software was used (38). The PIC values were computed using the formula proposed by Botstein et al. (1980) using the Excel software (39). The web-based SNiPlay software was used to determine transversion and transition mutations (40). The Plink software “recordeA” function (41) was used for the generation of the SNP dosage format 0, 1, and 2, respectively representing the homozygote, the homozygote alternative and the heterozygote.

4.2.3.5. Population Structure Analysis

To assess the population structure of superior shea trees, three complementary methods were employed: Bayesian model-based clustering using Structure version 2.3.4 software (42), principal coordinates analysis (PCoA) was performed using GenAlex version 6.503 software, and discriminant analysis of principal components (DAPC) was performed using the R software version 4.3.0.

For the Bayesian model-based clustering analysis, the Markov chain Monte Carlo (MCMC) method with the admixture model excluding the LOCPRIOR option was used. This analysis was iteratively run 10 times for each K value ranging from 1 to 10. A burn-in period of 50,000 iterations followed by 100,000 MCMC iterations was used. Additionally, we assumed an admixture model with correlated allele frequencies. The most probable K value for each test was determined using the delta K (ΔK) (43) method based on the rate of change in [LnP(D)] between successive K-values. Genotypes with membership probabilities greater than 0.7 were considered as belonging to the same group.

DAPC was used to complement the model-based population structure results obtained from Structure. DAPC is a multivariate method designed to identify and describe clusters of genetically related individuals (44). It was performed in R software version 4.3.0 using the package "adegenet" with the function "find.clusters". In the absence of a predefined grouping pattern, DAPC employs sequential K-means and model selection to establish genetic clusters based on genetic data. The bayesian information criterion (BIC) guided the determination of the optimal number of genetic clusters (K) to best describe the data. The calculation of the α-score was instrumental in retaining the optimal number of principal components. DAPC also furnished membership probabilities for each individual with respect to each identified group, which is comparable to the admixture proportions obtained from Structure.

A neighbor-joining (NJ) phylogenetic tree was reconstructed using R software (version 4.3.0) based on Nei's genetic distance with 1,000 bootstrap replicates. The tree was customized using the online tree annotation platform iTOL (Interactive Tree of Life) (45).

The number of clusters determined by Bayesian model-based clustering was subsequently used in the analysis of molecular variance (AMOVA) to assess the genetic differentiation of the genotypes. This comprehensive AMOVA was performed using the GenAlEx software version 6.503, and it allowed for the estimation of the fixation index (FST) and the gene flow per haploid number of migrants (Nm). FST values range from 0 (no differentiation between groups) to 1 (complete differentiation). Moreover, genetic diversity indices such as the number of different alleles (Na), the number of effective alleles (Ne), the number of loci with private alleles, Shannon's information index (I), observed heterozygosity (HO), and expected heterozygosity (HE) were also computed using the above-mentioned software (38).

4.2.3.6. Morphological Characteristics of the Groups Obtained with Bayesian Model-based Clustering from Structure

To evaluate the morphological characteristics of 160 genotyped shea trees (availability of their morphological traits), 11 quantitative and qualitative morphological traits (**Table S 1**), described in previous studies (94,205) were used. A comparison of the morphological traits between the groups obtained using SNP markers with Structure was also accessed. Principal component analysis (PCA) and a heat map of the correlation matrix were performed to structure the quantitative traits. Finally, a Mantel test with 10,000 permutations and Spearman correlation method was performed for matrices comparison between morphological traits and SNP markers to access the relationship between both markers.

4.2.3.7. Design of the Superior Shea Tree Core Collection

The design of the core collection followed the methodology proposed in a previous study (215). DARwin software version 6.0.21 was used for the reconstruction of the diversity of trees using SNP dataset (216). Dissimilarities were computed and transformed into Euclidean distances. The un-weighted neighbor-joining method was applied to the Euclidean distances to build a tree with all genotypes. The "maximum length sub tree" function was then used to identify the individuals of the core collection: The "maximum length sub-tree" is a step-by-step process that successively eliminates redundant individuals. We then selected the last 100 individuals that retain the largest diversity. The size of the core collection was fixed a priori (112), and the efficiency of the strategy was assessed by comparing and keeping the total number of alleles captured for each run using the same software. The size of the core collection was expressed as the ratio of individuals kept in the core collection to the number of individuals in the entire collection. Principal component analysis (PCA) was plotted using R software (version 4.3.0) to see the distribution of the core sample relative to the entire sample.

4.2.4. Results

4.2.4.1. Distribution of SNPs, Genetic Diversity, and Polymorphism Information Content in the *Vitellaria paradoxa* Genome

After the filtering process, 7,559 SNPs , representing 17.7% of the 42,705 SNPs initially yielded, were retained. These 7,559 SNPs were unequally distributed across the 12 chromosomes with, an average marker density of 1 marker per 87.38 kb.

A genome-wide SNP marker analysis revealed that chromosome 2 had the highest number of SNPs, with 12.9% (978 SNPs) of the filtered SNPs, while chromosome 12 had the lowest number, with 6.2% (469 SNPs). In terms of marker density, chromosome 6 had the highest marker density with 1 marker per 74.67 kb, while chromosome 4 had the lowest density with 1 marker per 105.77 kb (**Table 9**).

Table 9. Genomic distributions of the 7,559 filtered SNPs physically mapped on the 12 chromosomes of *Vitellaria paradoxa* and the corresponding SNPs densities

Chromosomes	No. of SNPs	% SNPs	Length (Mpb)	Density (kb)
Chr01	791	10.46	80,731,948	102.06
Chr02	978	12.94	74,439,616	76.11
Chr03	662	8.76	57,704,473	86.17
Chr04	564	7.46	59,651,551	105.77
Chr05	654	8.65	59,580,608	91.10
Chr06	659	8.72	49,210,429	74.67
Chr07	606	8.02	55,380,075	91.39
Chr08	664	8.78	52,298,408	78.76
Chr09	488	6.46	47,443,901	97.22
Chr010	522	6.91	46,597,090	89.27
Chr011	502	6.64	38,413,107	76.52
Chr012	469	6.2	37,276,254	79.48
Mean	630	8.33	-	87.38
Total	7559	100	-	-

Within the collected shea genome, transition-type SNPs (4,647, or 61.48% of SNP markers) were more common than transversion-type SNPs (2,992, or 38.52% of SNP markers). This resulted in a ratio of transitions to conversion SNPs of 1.6 (4,647/2,912). Specifically, transition A/G and T/C types were more common than G/A and C/T types. For transversion, the types T/G, A/T, A/C, and G/C were more common than G/T, T/A, C/A, and C/G (**Table 10**).

Table 10. Percentage of transition and transversion SNPs across the *Vitellaria paradoxa* genome

SNP type	Transition SNPs		Transversion SNPs			
	A/G	T/C	A/T	A/C	G/T	G/C
Number of SNPs	2316	2331	851	685	673	703
Frequency %	30.6%	30.8%	11.3%	9.1%	8.9%	9.3%
Total (Percent of total)	4647 (61.5%)		2912 (38.5%)			

The average observed heterozygosity (*HO*) in this study was 0.17, while the expected heterozygosity (*HE*) values varied from 0.1 (for 152 SNPs) to 0.5 (for 1,438 SNPs), with an average of 0.26. In parallel, the PIC values ranged from 0.1 (for 237 SNPs) to 0.5 (for 879 SNPs), with an average of 0.24 (**Figure S 1.a** and **Figure S 1.b**). A consistent number of SNPs (5,104 or 67.52% of the filtered SNPs) had a MAF value less than 0.2 (**Figure S 1.c**).

4.2.4.2. Population Structure and Genetic Relationships

The population structure and the genetic relationships among the studied population were investigated using the Structure version 2.3.4 software. The *K* value was used to estimate the number of clusters in the shea tree population based on genotypic data across the whole genome. The optimal *K* value was determined by plotting the number

of clusters (*K*) against Δ*K*. It showed a sharp peak at K = 3 (**Figure 20a**), suggesting that the studied population can be clustered into three groups with different genetic backgrounds (GP1, GP2, GP3). The number of genotypes in each group was 168 in GP1, 93 in GP2, and 72 in GP3. Among the population, 154 genotypes were considered as admixed **Table S 2**. Consequently, these results were considered in the subsequent population genetics analyses.

The net nucleotide distance between the groups is shown in **Table S 2**, and the maximum distance was recorded between GP1 and GP2 (0.003). The genetic distance between GP1 and GP3 (0.0026) was closely related to the distance between GP2 and GP3 (0.0025). The results from Structure estimated the fixation index (F_{ST}) for each group and suggested a significant divergence within the three groups (**Table S 2**). GP3 displayed the highest F_{ST} value (0.0182), while GP1 had the lowest F_{ST} value (0.0177). In addition, the heterozygosity values were 0.259, 0.261 and 0.262 for GP1, GP2, and GP3, respectively (**Table S 2**).

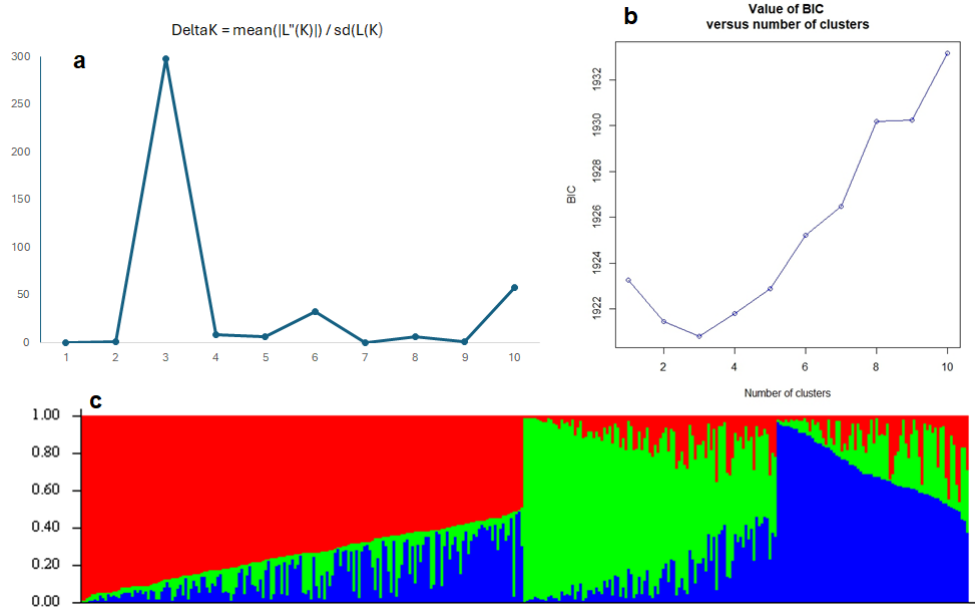


Figure 20 . Genetic structure of the 333 superior shea trees (a) Delta K for various number of clusters (K); (b) Values of the BIC (1,920.82) versus the number of clusters; (c) Bayesian model-based analysis (K = 3) of 333 *Vitellaria paradoxa* individuals; accessions in red are clustered into GP1, accessions in green are clustered in GP2, and accessions in blue are clustered into GP3

Consistent with the findings from Bayesian model-based clustering in Structure, the discriminant analysis of principal components also suggested three distinct clusters based on the value of BIC (1,920.82) (**Figure S 2.a**). In the DAPC results, group1, group2, and group3 have, respectively, 93, 72, and 168 individuals. The probability of each individual belonging to a single cluster was 100%. Therefore, no individual was considered mixed (**Figure S 2.a** and **Figure S 3**). The principal coordinates

analysis (PCoA) showed that GP2 is an intermediate group between GP1 and GP3 (**Figure S 4**). In fact, the Structure software triangular plot showed that most of the shea trees from the Hambol, Poro, and Tchologo regions were clustered in GP1, whereas genotypes included in GP2 and GP3 are mainly from the Bagoué region (**Figure S 2.b**). The same results were observed with the PCoA plot (**Figure S 5**).
In addition to the three methods above, a neighbor-joining phylogenetic tree also clustered the shea trees into three groups (**Figure 21**). Based on the true clades confirmed with the bootstrap values, the samples should be clustered by origin (**Figure 21**). This confirm the true structure of our shea tree samples using Structure.

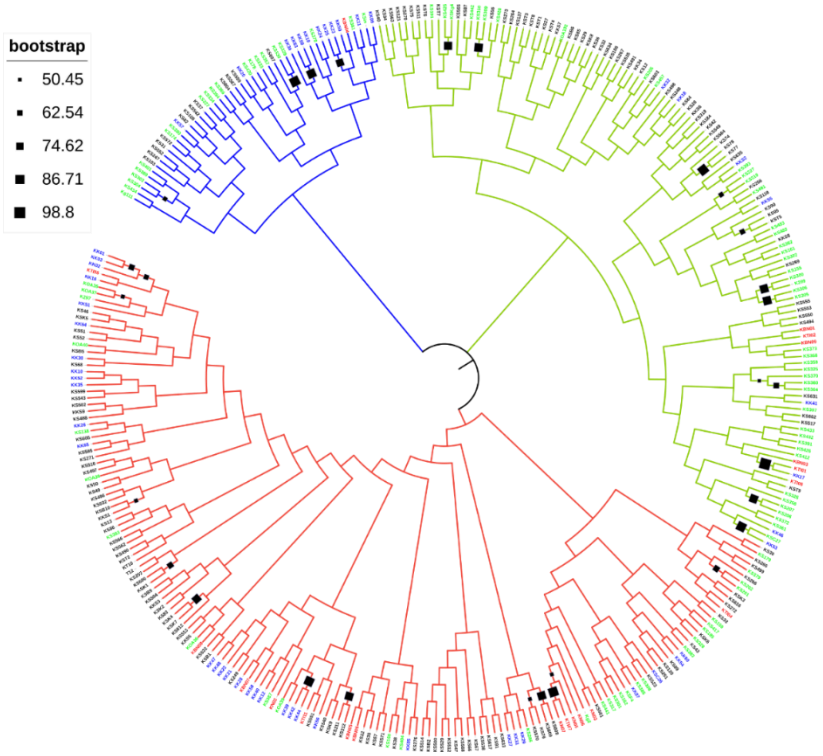


Figure 21. Neighbor-joining phylogenetic tree with bootstrap values (black squares represent the bootstrap values; 1,000 replicates) showing relationships between the 333 superior shea trees based on Nei's genetic distance matrix using 7,559 SNP markers; branch colors represent genetic groups (red for GP1, light green for GP2, and blue for GP3; label colors represent the origin of the genotypes (red for Hambol, black for Bagoué, green for Tchologo, and blue for Poro)

4.2.4.3. Genetic Differentiation of Populations

The three groups from the Bayesian model-based clustering in Structure were then used to calculate various genetic parameters such as AMOVA, Nei's genetic distance, and genetic diversity indices using GenAlEx 6.503 software. The results of these analyses are shown in **Table 11**. Overall, the results showed a low level of genetic differentiation between groups but a high level of genetic differentiation within

groups. In addition, Nei's genetic distance analysis revealed a very low fixation index value (F_{ST}: 0.004) and a significant number of migrants (*Nm*: 59.02), confirming the low level of genetic differentiation between the three genetic groups. The pairwise F_{ST} values were found to be 0.005, 0.004, and 0.003 for GP1-GP2, GP1-GP3, and GP2-GP3, respectively (**Table 11**).

Table 11. Analysis of molecular variance (AMOVA) among the 333 superior shea trees based on genetic variation among and within the identified groups.

Source	**Df**	**SS**	**MS**	**Estimated Variance**	**%**
Among Pops	2	6607.628	3303.814	5.763	0%
Within Pops	330	696593.908	2110.891	750.481	55%
Within Indiv.	333	203106.500	609.929	609.929	45%
Total	665	906308.036		1366.173	100%
Fixation index (F_{ST})	0.004				
Nm (Haploid)	59.02				
	GP1-GP2	GP1-GP3	GP2-GP3		
Pairwise F_{ST} values	0.005	0.004	0.003		

4.2.4.4. Allelic Pattern Across Populations

The three groups showed a grand mean of 2 for the number of different alleles (Na) and 1.395 for the number of effective alleles (Ne) (**Table 12**). Within the whole population, the means for Shannon's index (I), gene diversity (*HE*), and *unbiased gene diversity (uHE)* were 0.414, 0.258, and 0.260, respectively. The three genetic groups were closely related in terms of diversity metrics (**Table 12**). The percentage of polymorphic loci per population (PPL) was 100% for GP1 and GP2, while GP3 had a PPL of 99.93%.

Table 12. The means of different genetic parameters in each of the three groups

Pop	**N**	**Na**	**Ne**	***I***	***H0***	***HE***	***uHE***	***F***	**PPL**
GP1	154.418	2	1.395	0.415	0.173	0.259	0.259	0.327	100%
GP2	87.645	2	1.395	0.414	0.172	0.258	0.260	0.328	100%
GP3	65.345	1.999	1.396	0.413	0.173	0.258	0.260	0.320	9.93
Mean	102.469	2	1.395	0.414	0.173	0.258	0.260	0.325	99.98%

N: number of samples, Na: number of different alleles, Ne: number of effective alleles, I: Shannon's index, HO: observed heterozygosity, HE: diversity index, uHE: unbiased diversity index, PPL: percentage of polymorphic loci

4.2.4.5. Comparison of Our Results with Previous Studies

Several studies assessed the genetic diversity in shea tree during the last two decades (**Table S 3**). SSR markers were widely used to assess the genetic diversity of shea tree, representing 54.55% of the studies. SNP markers were the most recent markers used, representing 27.27% including our study. Only two studies mentioned RAPD markers in shea genetic diversity study (18.18%). Moderate genetic diversity was observed in our study and a study conducted in Uganda using SNP markers. However, the HE obtained in the present study (HE: 0.26) was higher than the HE obtained in Uganda (HE: 21). Using the same type of marker, a study conducted in Ghana showed low genetic diversity, with an HE value of 0.041. However, the population genetic

differentiation parameter such as the fixation index (F_{ST}) was higher in the Ugandan population than in our study and in the population of Ghana (**Table S 3**).

For other studies using SSR markers, low HE value (0.32) to moderate HE value (0.73) was observed. The HE values of SSR markers were higher than HE in our study because SSRs are co-dominant and the HE value varies from 0 to 1, whereas for SNP and RAPD markers, HE is calculated as dominant and the value ranges from 0 to 0.5. Overall, studies conducted in a single country or subspecies showed low genetic differentiation. In contrast, studies conducted in the natural range of *V. paradoxa* showed high genetic differentiation, and the authors attributed this differentiation to the subspecies *nilotica* and *paradoxa* (**Table S 3**).

4.2.4.6. Design of the Core Germplasm Collection

A core germplasm collection is a critical step in creating a manageable and representative sample that can reflect the diversity within the larger germplasm collection. This process is also relevant to modern plant breeding efforts. Maintaining genetic diversity within the core collection is essential, and the primary criterion for selecting its members is the average genetic distance in the population. Using the "maximum length sub-tree" function in DARwin 6.0.21, we successfully designed a core germplasm set of 100 individuals, representing 30% of the entire population (**Figure S 6**). Of the core germplasm, 49% of the shea trees are from the region of Bagoué. These individuals belong to GP2 (30 individuals) and GP3 (19 individuals). The remaining 51% of the core germplasm, are from the other three regions (Hambol, Poro, and Tchologo) and belong mainly to GP1 (**Table 13**). The grand mean diversity metrics of this core collection set were similar to those of the entire population (**Table S 4**).

Table 13. Percentage of individuals from each region and genetic group in the core set

Region	**N population**				**N core set**			
	GP1 (%)	GP2 (%)	GP3 (%)	Total	GP1 (%)	GP2(%)	GP3 (%)	Total
Bagoué	-	95 (28.5)	68 (20.4)	163	-	30 (30)	19 (19)	49
Hambol	20(6.01)	-	-	20	8 (8)	-	-	8
Poro	51 (15.3)	-	2 (0.6)	53	28 (30)	-	2 (2)	30
Tchologo	96 (28.8)	-	1 (0.3)	97	12 (12)	-	1 (1)	13
Total	167 (50.2)	95 (28.5)	71 (21.3)	333	48 (48)	30 (30)	22 (22)	100

A phylogenetic tree of the 100 individuals in the core set was reconstructed with the neighbor joining method based on Nei's genetic distance. The obtained dendrogram is similar to that obtained with the whole population (**Figure S 6**).

A good spatial representation of the core set within the entire sample was observed in the PCA plot of the entire panel of 333 shea trees (**Figure 22**).

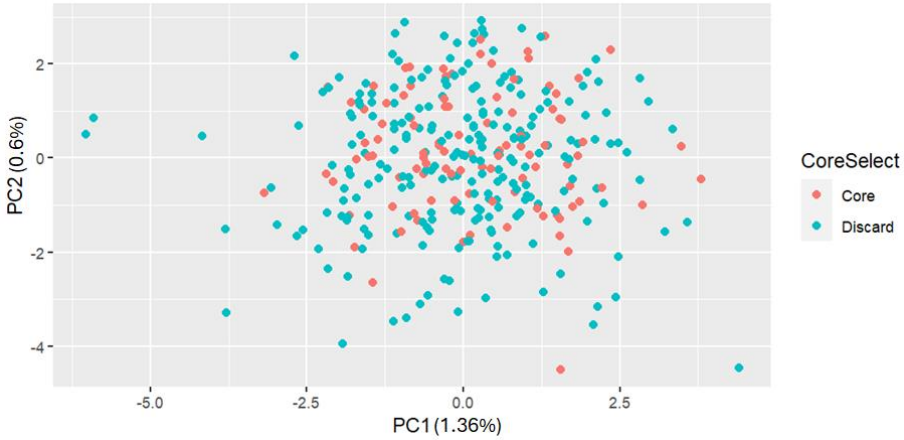


Figure 22. Principal component analysis plot showing the distribution of the core sample within the full panel of 333 shea trees. Blue dot represents the discarded genotypes and red dot represents the core sample.

4.2.4.7. Morphological Characteristics of the Genotyped Shea Trees

The analysis of the morphological characteristics of 160 genotyped shea trees revealed significant variations (**Table S 5**). Principal component analysis (PCA) identified three principal components (PCs) that explained 73.03% of the total variance observed among shea trees population (**Table S 6**). The PC1 (x-axis) explained 34.55% of the total variance. The leaf traits (PL, LL, and LW) and nut traits (NL, NWD and NWG) are correlated positively towards PC1 (**Figure 23**). However, trunk circumference (TC) correlated negatively towards PC1. Similarly, PC2 (y-axis) accounted for 26.03% of the total variation. The nut traits (NL, NWD, and NWG) and TC positively correlated towards PC2 while the leaf traits (PL, LL, and LW) were negatively correlated (**Figure 23**). PC3 captured 12.45% of the total variance, and TC and NWG were positively correlated. In contrast, NWD and NGW were negatively correlated towards PC3 (**Table S 6**).

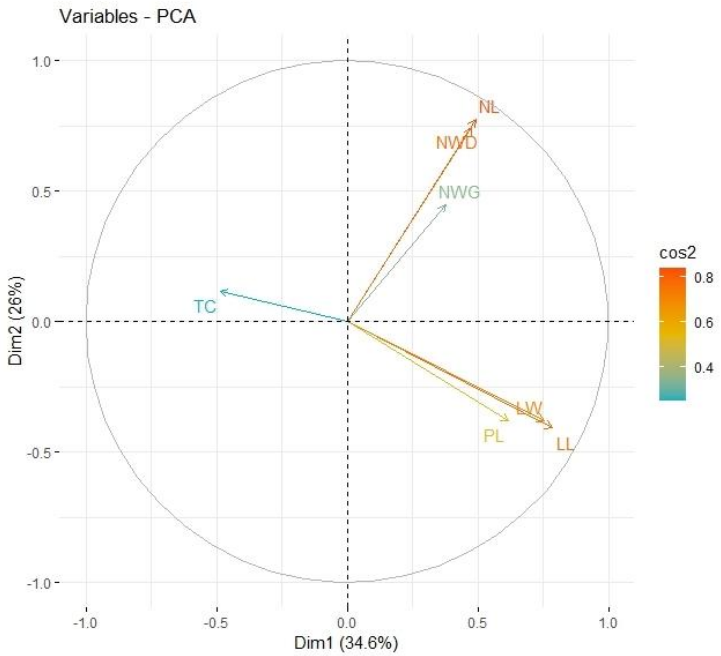


Figure 23. Estimated PC1 and PC2 for quantitative morphological traits of 160 genotyped shea trees. TC: trunk circumference, PL: petiole length, LL: limb length, LW: limb width, NL: nut length, NWD: nut width, NWG: nut weight

The correlation heatmap analysis revealed three classes of studied traits: (i) trunk circumference, (ii) nut size (nut length, nut width, and nut weight) and (iii) leaf size (limb width, limb length, and petiole length). Positive correlations were observed within the descriptors of each identified class (see color key in **Figure S 7**).

Using clustering from Structure as a priori groups, significant differences were found between the three groups in nut weight (F: 14.11; $p < 0.001$), trunk circumference (F: 5.04; $p < 0.01$), and limb width (F: 4.16; $p < 0.05$). GP1 showed a weak mean of trunk circumference (133.4 cm), large leaf size (limb width = 4.63 cm), and a high mean of nut weight (10.62 g), whereas GP2 showed a high mean of trunk circumference (160.5 cm), a thin leaf (limb width = 4.46 cm), and a weak mean of nut weight (8.89 g). GP3 had the lowest value for significant traits such as nut weight (8.36 g), and limb width (4.44 cm) and, it was intermediate for trunk circumference (149.1 cm). A detailed examination of the results shows that GP1 differs from GP2 and GP3, while no significant differences are observed between GP2 and GP3 (**Table S 7**). There were no statistical differences between the groups for the other quantitative traits. The means of these quantitative traits are presented in **Table S 7**.

Meanwhile, all the qualitative traits used to characterize the three groups showed statistically significant differences (**Table S 7**).

While the seed coat colors of "creamy" and "dull brown" were present only in GP1, the spheroid seed shape was found only in GP2 and GP3 (**Table S 7**). In terms of group differentiation, the same trend observed for quantitative traits was also observed for qualitative traits. Therefore, the three groups obtained with SNP data can only be clustered into two groups based on morphological characteristics: (I): GP1 and (II): GP2 and GP3 as a single group.

The Mantel test results revealed that there is no relationship between morphological traits and SNP markers (Mantel r = 0.0176, p = 0.3324).

4.2.4.8. Morphological Characteristics of the Core Collection

Of the 160 genotyped shea trees, 52 were captured in the established core collection. The analysis of the morphological characteristics of these core individuals showed important variations for the quantitative traits (**Table S 8**). For example, trunk circumference varied from 65.5 to 242 cm with a mean of 19.7 cm. Concerning qualitative traits, observations demonstrated that almost all the modalities of traits were found in the core collection. Only "spheroid shape", a modality of seed shape (SEES) was not captured in the core collection (**Table 14**). However, that modality had a very low proportion (2.5%) in the entire collection. In addition, multivariate analyses showed that there is no significant difference between the core collection morphological characteristics compared to those of the entire collection (**Table 14**).

Table 14. Quantitative and qualitative traits associated with the core and entire collection of *Vitellaria paradoxa*

Quantitative Traits		**Means ± standard deviation**		**F-value**	**p-value**
		Initial collection (N=160)	Core set (N=52)		
Trunk Circumference (cm)		144.6 ± 45.93	149.7± 44.3	0.48	0.49
Petiole length (cm)		8.29 ± 1.49	8.12 ± 1.61	0.51	0.475
Limb length (cm)		14.77 ± 2.31	14.33± 2.13	1.44	0.232
Limb width (cm)		4.61 ± 0.79	4.44 ± 0.78	1.82	0.18
Nut length (cm)		3.11 ± 0.38	3.66 ± 0.42	0.19	0.67
Nut width (cm)		2.9 ± 0.3	2.26 ± 0.41	0.78	0.38
Nut weight (g)		9.55 ± 2.61	10.09 ± 2.73	1.6	0.21
Qualitative traits	**Modalities**	**Proportions N(%)**		**χ^2**	**p-value**
		Initial collection (N=160)	Core set (N=52)		
Tree Growth Habit (TGH)	Erect	55 (34.38)	13 (25)	2.73	0.26
	Semi-erect	60 (37.5)	26 (50)		
	Spreading	45 (28.12)	13 (25)		
Leaf Apex Shape (LAS)	Acute	47 (29.38)	18 (34.62)	4.4	0.22
	Acuminate	24 (15)	13 (25)		
	Retuse	49 (30.63)	11 (21.15)		
	Obtuse	40 (25)	10 (19.23)		
Adult Leaf Color (ALC)	Light green	13 (8.13)	3 (5.77)	0.86	0.65
	Green	133 (83.13)	46 (88.46)		
	Dark green	14 (8.75)	3 (5.77)		
Seed Coat Color (SCC)	Creamish	13 (8.13)	8 (17.33)	3.08	0.54
	Dull brown	3 (1.88)	1 (1.92)		
	Brown	78 (48.75)	20 (38.46)		
	Pale brown	15 (9.38)	5 (9.62)		
	Dark brown	51 (31.88)	18 (34.62)		
Seed shape (SEES)	Spheroid	4 (2.5)	0 (0)	2.81	0.42
	Ellipsoid	37 (23.13)	16 (30.77)		
	Oval	65 (40.63)	24 (46.15)		
	Ovoid	54 (33.75)	14(26.92)		

4.2.5. Discussion

In order to assess the genetic diversity within *Vitellaria paradoxa*, a collection of 333 genotypes was gathered from the *in situ* superior shea collection in Côte d'Ivoire. The genotypic information from these individuals was used to investigate genetic diversity and population genetics, which could provide valuable insights for future breeding efforts, such as genome-wide association studies (GWAS). To the best of our knowledge, this study is the first one that uses single nucleotide polymorphisms (SNPs) to investigate the genetic diversity and population structure of *Vitellaria paradoxa* subspecies in Côte d'Ivoire.

4.2.5.1. Single Nucleotide Polymorphism Markers and Mutation Types

In this study, 7,559 high-quality SNP markers were retained after data filtering. The analysis of SNPs distribution and mutation types in the genome was conducted. A higher frequency of transitions than transversions was observed (transitions/transversions = 1.6). A ratio of 1.3 in transition SNPs and transversion SNPs was obtained in shea trees in Uganda (54). Similar transition/transversion results have been reported in other plant species, such as such as *Hevea brasiliensis* (217), *Camellia sativa* (206), *Vigna unguiculata* (215), *Colocasia esculenta* (218), and *Oryza sativa* (219), and in living organisms in general (220). This suggests that during the natural selection of *V. paradoxa*, transition mutations tend to be more tolerated than transversion mutations.

This may be due to the presence of higher frequencies of synonymous mutations in the protein-coding sequences (220). However, the non-synonymous SNPs are of interest for this study because they generate new variants that are important in breeding programs (54). In natural selection, non-synonymous SNPs are important for the conservation of shea tree species because the variants become more adapted to environmental changes (54). Transition mutation types are also important in evaluating the distribution, extent, and amount of genetic variation among and within shea tree populations (54). The frequencies of A/G and C/T transitions were similar (A/G: 30.6% and T/C: 30.8%). This result is consistent with those obtained for shea tree and rice species (54,219).

4.2.5.2. Gene Diversity

The genetic diversity and the population structure of a species are good indicators of its management and conservation status (54).

The findings in this study revealed moderate level of genetic diversity (HE: 0.26) and moderate polymorphism information content (PIC: 0.24) in *V. paradoxa*. This is consistent with what has been reported for shea trees in East Africa using SNP (54) and SSR markers (52), in West Africa with SSR markers (53,124) and in the natural range of the species using RAPD markers (50,125). This suggests that the SNP markers used in our study were reasonably informative markers. In contrast, a *V. paradoxa* diversity study in Ghana using SNP markers revealed very low genetic diversity (126). Other studies showed low genetic diversity in *V. paradoxa* using SSR markers (24,123). These results suggest that the genetic diversity in shea tree species is low to moderate in its natural range.

The moderate genetic diversity may be the result of selection based on farmer-preferred traits such as large fruit size, tasty pulp, and high oil content (28,62). Furthermore, this moderate genetic diversity provides an opportunity to generate shea varieties that can be successfully grown in shea belts in different geographical regions (54). Studies involving other plant species such as *Camelina sativa* (206), *Ziziphus jujuba* (221), and winter wheat (222) observed the same trend and attributed these findings to the bi-allelic nature and low mutation rates of SNPs. Our marker density is sufficient to perform genome-wide association studies, as a genotype panel with

minor allele frequency (MAF) > 0.1 is desirable for genome-wide association mapping (25,219).

4.2.5.3. Population Structure and Relationships

Population structure analyses are essential for understanding genetic diversity and facilitating subsequent association mapping studies (222). In our study, the cluster pattern in Structure grouped the shea trees into three groups, which concurs with the DAPC, PCoA, and the neighbor-joining tree reports. These results were also consistent with a previous study using SNP markers in *V. paradoxa* (126) in Ghana. However, our results were different from those obtained in Uganda using SNP markers. The authors clustered their accessions into two groups. These results can confirm that West African shea trees exhibit higher genetic diversity than East African shea trees using SSR markers (14). The structure of the Ivorian shea population collection suggests a geographical effect. In fact, the genotypes clustered in GP1 are from the regions of Poro, Tchologo, and Hambol, while the genotypes collected in GP2 and GP3 come from the region of Bagoué. There are thus two geographical groups of shea trees in our collection: one group in the western part of the Poro region, which can also be divided into two subgroups representing GP2 and GP3, and the second geographical group located in the eastern art of the Poro region (see **Figure 19**). Similar results were observed in the Ugandan shea tree population structure study (54). These findings suggest that these geographical groups can serve as genetic resources for hybridization programs to create improved varieties and align with the objectives of the annotated genome (25).

However, a certain number of individuals showed a mixed genetic profile across the three groups (**Figure 20c**). This observation is consistent with findings on *V. paradoxa* using DArTseq SNP markers (54,126). Gene flow between individuals in the neighborhood or individuals overlapping the study areas may be responsible for the presence of these admixed individuals. As a result, a limited number of accessions may show clear membership to one group, while the majority may show some degree of membership to the three groups. This indicates that common allelic/gene combinations continue among the collection of shea trees.

4.2.5.4. Genetic Differentiation of Populations

The F_{ST} value is the most relevant F-statistic used to study the degree of genetic differentiation between and within populations (219). In this study, a fixation index (F_{ST}) value of 0.004 was found for the whole population and the low pairwise F_{ST} values found between the three groups (**Table 11**) indicate low genetic differentiation between these groups (219,223). Similar results were found in the genetic differentiation in *V. paradoxa* populations implying SNP markers (54,126) and SSR markers (123). The neighbor-joining tree based on genetic distances confirmed the low differentiation between groups, with few bootstrap values higher than 50.45, with the main branches having bootstrap values lower than 50.45 (**Figure 21**). This trend has been reported in shea tree populations using nuclear SSR markers (124). These results suggest the presence of extensive and anthropogenic gene flow, outcrossing and admixture in the study area. Comparison of our findings with previous studies

revealed that low genetic differentiation is observed in shea tree populations when the study is carried out in a small area, limited to one or two countries (24,52,123,126,224) or within a subspecies (124). This trend is expected in the shea tree due to its status as a long-lived woody perennial, insect-pollinated outcross, and widespread in a continuous range (124). However, high genetic differentiation is observed when the study is conducted in the natural range of the subspecies (14,50,125). This high differentiation can be explained by the fact that these studies included both subspecies of *V. paradoxa*, and differentiation is observed between the *nilotica* and *paradoxa* subspecies (14,50,125). These findings suggest that many individual shea trees should be considered for efficient sampling of genetic diversity within a population. This is consistent with the proposal to develop a breeding population of *Vitellaria paradoxa* (125).

The findings from Structure are consistent with the outcomes from the AMOVA, where the total variation was primarily attributed to variations within the groups. Furthermore, it has been suggested that a high Nm value, indicative of substantial gene flow, can lead to low differentiation between populations (206). Our study aligned with this observation, as a very high Nm value of 59.02 was obtained. This suggests that in terms of gene flow, the Nm value obtained in our study was higher than that reported in shea trees using SSR markers (24). Therefore, our results suggest that population genetic differentiation using SNP markers is more informative than other markers.

4.2.5.5. Allelic Pattern and Genetic Diversity Indices

The allelic patterns and genetic diversity indices offered valuable insights into the genetic diversity present within each of the three genetic groups. While the three groups exhibited relatively close levels of expected heterozygosity (HE), GP1 showed a slightly higher HE compared to GP2 and GP3. This indicates that GP1 has a slightly higher level of diversity compared to the other groups, as HE considers both the number of alleles (referred to as richness) and the distribution (or evenness) of those alleles within a population. This could be explained by the origin of the genotypes clustered in GP1. In fact, the genotypes in GP1 are from three different regions (Hambol, Poro, and Tchologo), while the genotypes in GP2 and GP3 are mainly from the Bagoué region.

Genetic diversity ranged from low to high. This result is consistent with recent studies using SNP markers (54) and previous population genetic studies using SSR markers (14,24,53,123,124). Understanding the genetic diversity within *V. paradoxa* populations is essential, this is fundamental element for robust shea tree improvement programs and for future studies using genomic screening methods such as marker-assisted screening (MAS) and genome-wide association studies (GWAS) (25).

4.2.5.6. Core Collection

Sustaining living collections, a common practice for perennial tree crops, can be a costly and labor-intensive endeavor. Creating core collections is an efficient strategy for managing germplasm, effectively reducing costs while retaining the highest possible genetic diversity within the germplasm pool while minimizing redundancy

(208,209). Developing core collections has been the focus of various approaches (225–227), and the choice of the most appropriate evaluation methods depends on the specific objectives associated with these core collections (228).

In this study, we used the "maximum length sub-tree function" of DARwin version 6.0.21 to carefully curate a core set of 100 superior shea trees (30% of the initial population). The analysis indicates that the core collection successfully captures the full genetic diversity of the entire population. This is evidenced by the PCA plot, which showed good coverage of the core sample across the whole panel (**Figure 22**), and the representation of each genetic group within the core, which includes 48% from GP1, 30% from GP2 and 22% from GP3. These proportions reflect the overall genetic structure of the entire population, ensuring that no significant genetic diversity is lost in the core. In addition, several genetic parameters in the core closely mirror those of the total population: heterozygosity, allele frequency, and polymorphism rates. This similarity underscores the effectiveness of the core collection in maintaining the genetic variation present in the larger population, making it a robust subset for further genetic studies and conservation efforts. Thus, the core collection can safely be used for genetic research, breeding programs, and conservation strategies without compromising the genetic integrity of the original population (229).

The proportion of this core set is similar to the proportion established for other crops, such as palm oil (31.2%) (230). However, it differs from the core set proportions observed for *E. oleifera* (6.4%) (231), *S. superba* (19.87%) (232), and *P. massoniana* (19.46%) (229). A core collection sample size of 5-10% of the original germplasm resources can be sufficient to represent over 70% of the genetic variation present in the entire germplasm (208). For effective conservation of the genetic diversity of the entire population, a size of 20-30% of the population is required for the core set (233).

However, there is no universal approach to selecting a core size. It depends on factors such as the extent of variability and redundancy within the collection, the resources available for core set management, and the frequency of species regeneration (233,234).

Field genebanks provide convenient and immediate access to germplasm resources. However, relying solely on *in situ* conservation is not the most reliable long-term conservation strategy. Shea trees can be uprooted by adverse weather conditions, such as tropical cyclones (205). To address this, a shift in focus for germplasm conservation of shea trees can be directed towards the identified core set. A viable approach to germplasm conservation is to clone the identified core set through grafting for *ex situ* conservation. This method ensures the precise conservation of maximum genetic diversity while reducing redundancy and avoiding genetic erosion. Furthermore, the number of accessions to be planted in the field can be reduced, resulting in a more manageable and cost-effective approach compared to progeny trials.

The use of molecular markers to establish a core collection offers distinct advantages because they can accurately capture genetic diversity regardless of plant growth status, developmental stage, and environmental conditions (235). Importantly,

genetic diversity is often positively associated with population persistence and resilience to environmental changes (236).

4.2.5.7. Morphological Characteristics of the Genotyped Shea Trees

Variability in superior shea trees has been widely demonstrated using morphological traits (94,205). A similar trend was observed in this study. Using genotype clustering from the Structure software showed significant differences between the groups for trunk circumference, nut weight, and leaf width. These differences are essentially between GP1 and GP2 or GP3. No significant differences were observed between GP2 and GP3. These results suggest that the three groups obtained with molecular markers reflected two morphological groups. This can be explained by the significant influence of the environment on the expression of morphological traits. A savanna gradient has been reported in the study area (94). The effect of climate on the expression of shea tree morphological traits has been reported in several countries (94). In addition, we suggest that the SNP markers that structured Bagoué superior shea trees into two groups are synonymous SNPs, while those that separated GP1 and GP2/GP3 could be designed as nonsynonymous SNPs. SNPs may change the encoded amino acids (nonsynonymous SNP) and change the amino acid sequence or be silent (synonymous SNP), thereby maintaining the amino acid sequence or simply occurring in the noncoding regions (54). A Mantel test realized between morphological traits and SNP markers did not show significant differences (Mantel $r = 0.0176$; $p = 0.3324$). This suggests that the morphological characteristics observed in the groups are not the direct effects of the SNP markers. In addition to environmental conditions, the varying mineral composition of the soil from one ecological zone to another could also have an effect on the morphological trait expression. Studies have reported soil mineral composition effects on shea tree morphology in Mali (51), Eastern Ghana (160), and West Africa (237).

The morphological characteristics of the core collection showed a similar trend to those of the entire collection. A similar trend was found in a core establishment in *Synsepalum dulcificum* (210). These results suggest that the core collection captured the full morphological characteristics of the entire collection. It is also important to ensure that the core collection is well representative of the whole collection in all genetic aspects, confirming the quality of the established core based on morphological traits as well as molecular markers (SNPs). Therefore, morphological traits could be confidently used in the establishment of core collections of *V. paradoxa* species without loss of genetic diversity.

4.2.6. Conclusion

In this study, we used high-throughput DArTseq technology to investigate the genetic diversity and population structure of *Vitellaria paradoxa* in Côte d'Ivoire. The primary objective was to explore the potential utility of SNP markers for various genomic analyses in the context of genetic improvement efforts. Our data revealed that the collection under investigation exhibited a moderate degree of genetic diversity.

This rich genetic diversity serves as a promising foundation upon which to develop novel *Vitellaria* cultivars boasting desirable traits, including high yield potential, high oil production, and resilience to biotic and abiotic stresses, all while being well-suited for adaptation to diverse environmental conditions.

Furthermore, our research unveiled the presence of three genetic groups within the study population. The differentiation between these groups can be attributed to a combination of factors and natural selection pressures. Notably, the three groups demonstrated very close diversity across multiple parameters, including Shannon's information index (I), expected heterozygosity (HE), and unbiased expected heterozygosity (uHE).

A core collection of 100 superior shea trees, representing 30% of the entire population was captured. This study marks the inaugural endeavor to molecularly characterize and validate the creation of a core set for shea tree germplasm resources. The core collection successfully captured all the variation in morphological traits and the alleles present within the accessions, while preserving the genetic diversity and structure of the original population.

These findings provide important information for suitable conservation and future allelic/gene identification using genome-wide association studies (GWAS) and marker-assisted selection (MAS) to enhance genetic gain in *V. paradoxa* breeding programs.

4.2.7. Data availability

The datasets generated and/or analyzed during the current study are available in the Science Data Bank (ScienceDB) repository,
https://doi.org/10.57760/sciencedb.17559

4.2.8. Supplementary materials

Table S 1: Measured seven quantitative and four qualitative morphological traits

Quantitative traits	***Code***	**Unit (SI)**
Trunk Circumference	TC	cm
Petiole length	PL	cm
Limb length	LL	cm
Limb width	LW	cm
Nut length	NL	cm
Nut width	NWD	cm
Nut weight	NWG	g
Qualitative traits	**Code**	**Modalities**
Tree Growth Habit	TGH	Erect, semi-erect, spreading
Leaf Apex Shape	LAS	Acute, acuminate, retuse
Adult Leaf Color	ALC	Light green, green, dark green
Seed Coat Color	SEES	Spheroid, ellipsoid, oval, ovoid

Table S 2. Results of Bayesian model-based clustering implemented in Structure of 333 *Vitellaria paradoxa* accessions

Cluster	Net nucleotide distance		*HE*	*Mean F_{ST}*	Number of genotypes	Admixed genotypes
	GP1	*GP2*				
GP1	0		0.2588	0.0177	167	69
GP2	0.003	0	0.2606	0.0178	95	47
GP3	0.0026	0.0025	0,2615	0.0182	71	38

F_{ST} : fixation index, HE: expected heterozygosity

Table S 3. Studies using molecular markers to assess the genetic diversity of *V. paradoxa* in the last two decades

Subspecies	Marker	HE	F_{ST}	N	Study area	Gene. diff	Reference
ssp. paradoxa	SNP	0.26	0.004	333	Côte d'Ivoire	Low	Our study
ssp. nilotica	SNP	0.21	0.01	623	Uganda	Low	(54)
ssp. paradoxa	SNP	0.04	0.002	282	Ghana	Low	(126)
ssp. paradoxa	SSR	0.67	0.113	200	Ghana	Moderate	(53)
ssp. nilotica	SSR	0.73		118	Uganda	Low	(52)
ssp. paradoxa	SSR	0.63	0.085	673	West Africa	Low	(124)
ssp. paradoxa ssp. nilotica	SSR	0.51	0.21	374	Nat. range	High	(14)
ssp. paradoxa	SSR	0.32	0.047	169	Mali and Côte d'Ivoire	Low	(24)
ssp. paradoxa ssp. nilotica	RAPD	0.26	0.23	179	Nat. range	High	(50)
ssp. Paradoxa Ssp. nilotica	RAPD	0.26	0.15	118	Nat. range	High	(125)
ssp. paradoxa	SSR	0.41	0.001	441	Mali	Low	(123)

N: number of samples, Na: number of different alleles, Ne: number of effective alleles, I: Shannon's index, HO: observed heterozygosity, HE: diversity index, uHE: unbiased diversity index, PPL: percentage of polymorphic loci

Table S 4. Mean of different genetic parameters in the core set

Pop	N	Na	Ne	*I*	*H0*	*HE*	*uHE*	*F*	PPL
GP1	40.282	1.993	1.393	0.408	0.128	0.255	0.259	0.483	99.29%
GP2	25.483	1.969	1.392	0.402	0.132	0.252	0.258	0.443	96.89%
GP3	18.372	1.938	1.392	0.397	0.128	0.250	0.258	0.448	93.81%
Mean	28.046	1.967	1.392	0.402	0.129	0.253	0.258	0.459	96.66%

N: number of samples, Na: number of different alleles, Ne: number of effective alleles, I: Shannon's index, HO: observed heterozygosity, HE: diversity index, uHE: unbiased diversity index, PPL: percentage of polymorphic loci

Table S 5. Variation in quantitative morphological traits of the entire collection

Traits	Minimum	Mean ± Sd	Maximum
TC	29.7	144.6 ± 46.07	266
PL	5.32	8.29 ± 1.5	13.42
LL	9.67	14.77 ± 2.31	21.48
LW	2.94	4.61 ± 0.79	7.16
NL	1.78	3.11 ± 0.38	4.1
NWD	0.9	2.31 ± 0.3	2.9
NWG	4.33	9.55 ± 2.62	19.57

Sd: Standard deviation

Table S 6. Factor loadings in the first three factor components.

Factor components	PC1	PC2	PC3
Eigenvalues	2.42	1.82	0.87
Variance (%)	34.55	26.03	12.45
Cumulative variance (%)	34.55	60.58	73.03
Trunk circumference	-0.49	0.12	0.78
Petiole length	0.62	-0.38	0.2
Limb length	0.79	-0.41	0.15
Limb width	0.76	-0.38	0.07
Nut length	0.49	0.77	-0.03
Nut width	0.48	0.74	-0.12
Nut weight	0.38	0.45	0.42

PC1, PC2, PC3: principal components 1; 2; 3

Table S 7. Quantitative and qualitative traits associated to the identified *Vitellaria paradoxa* groups.

Quantitative Traits		**Means ± standard deviation**			**F-value**	**p-value**
		GP1 (N=75)	GP2 (N=40)	GP3 (N=45)		
Trunk Circumference (cm)		133.4 ± 49.79a	160.5± 38.81b	149.1 ± 41.29ab	5.04	< 0.01
Petiole length (cm)		8.23 ± 1.47	8.45 ± 1.46	8.26 ± 1.59	0.28	0.755
Limb length (cm)		15.01 ± 2.26	14.54 ± 2.41	14.58 ± 2.32	0.76	0.468
Limb width (cm)		4.8 ± 0.78a	4.46 ± 0.86ab	4.44 ± 0.67b	4.16	< 0.05
Nut length (cm)		3.09 ± 0.45	3.15 ± 0.29	3.08 ± 0.3	0.44	0.647
Nut width (cm)		2.38 ± 0.38	2.26 ± 0.2	2.34 ± 0.2	2.27	0.107
Nut weight (g)		10.62 ± 2.7a	8.89 ± 2.14b	8.36 ± 2.17b	14.11	< 0.001
Qualitative traits	**Modalities**	**Proportions N(%)**			**χ^2**	**p-value**
		GP1 (N=75)	GP2 (N=40)	GP3 (N=45)		
Tree Growth Habit (TGH)	Erect	19 (25.33)	16 (40)	20 (44.44)	38.78	< 0.001
	Semi-erect	46 (61.33)	8 (20)	6 (13.33)		
	Spreading	10 (13.33)	16 (40)	19 (42.22)		
Leaf Apex Shape (LAS)	Acute	20 (26.67)	9 (22.5)	18 (40)	31.93	< 0.001
	Acuminate	20 (26.67)	1 (2.5)	3 (6.67)		
	Retuse	26 (34.67)	10 (25)	13 (28.89)		
	Obtuse	9 (12)	20 (50)	11 (24.44)		
Adult Leaf Color (ALC)	Light green	1 (1.33)	5 (12.5)	7 (15.56)	11.9	< 0.05
	Green	65 (86.67)	33 (82.5)	35 (77.78)		
	Dark green	9 (12)	2 (5)	3 (6.67)		
Seed Coat Color (SCC)	Creamish	13 (17.33)	0 (0)	0 (0)	25.37	< 0.01
	Dull brown	3 (4)	0 (0)	0 (0)		
	Brown	31 (41.33)	18 (45)	29 (64.44)		
	Pale brown	4 (5.33)	7 (17.5)	4 (8.89)		
	Dark brown	24 (32)	15 (37.5)	12 (26.67)		
Seed shape (SEES)	Spheroid	0 (0)	1 (2.5)	3 (6.67)	33.39	< 0.001
	Ellipsoid	24 (32)	4 (10)	9 (20)		
	Oval	40 (53.33)	14 (35)	11 (24.44)		
	Ovoid	11 (14.67)	21 (52.5)	22 (48.89)		

Values in brackets represent the percentage of modalities for each qualitative trait.

Table S 8. Variation in quantitative morphological traits of the core set

Traits	**Minimum**	**Mean ± Sd**	**Maximum**
TC	65.5	149.7 ± 44.34	242
PL	5.32	8.12 ± 1.61	13.42
LL	10.96	14.33 ± 2.13	19.62
LW	2.94	4.44 ± 0.78	6.2
NL	1.78	3.08 ± 0.42	3.66
NWD	0.9	2.26 ± 0.41	2.76
NWG	4.33	10.09 ± 2.73	17.93

Sd: Standard deviation

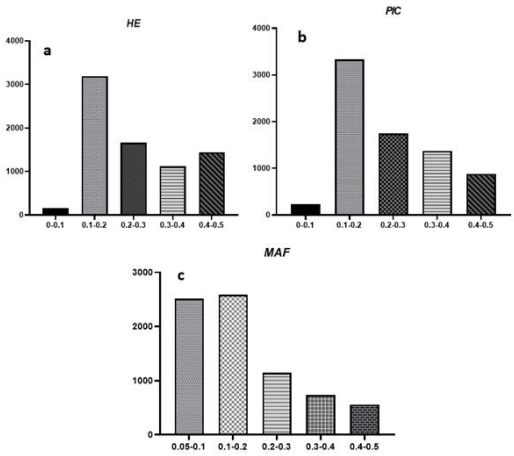


Figure S 1. Distribution of genetic diversity for the 7,559 filtered SNP markers from the 333 *Vitellaria paradoxa* genotypes. a: gene diversity (*HE*); b: Polymorphism information content (PIC); c: Minor allele frequency (MAF). X-axis indicates HE, PIC and MAF value classes; Y-axis indicates the number of SNP markers.

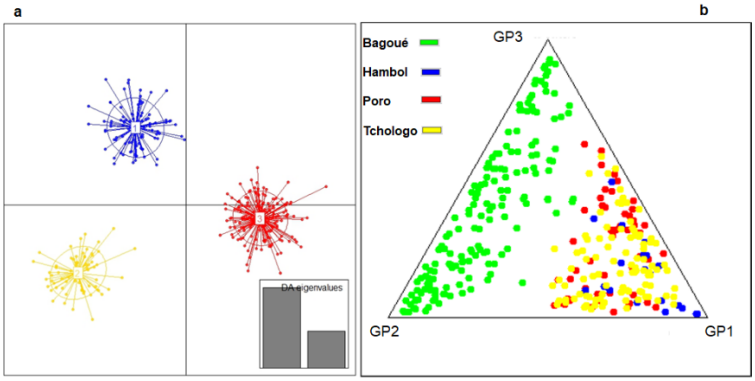


Figure S 2. (a) DAPC showing the genetic networks for the three genetic groups, with node sizes indicating genetic relationships between different accessions; (b) Structure software triangular plot showing the distribution of the four populations in the three genetic groups; dot colors represent the origin of the genotypes (green for Bagoué, blue for Hambol, red for Poro, and yellow for Tchologo).

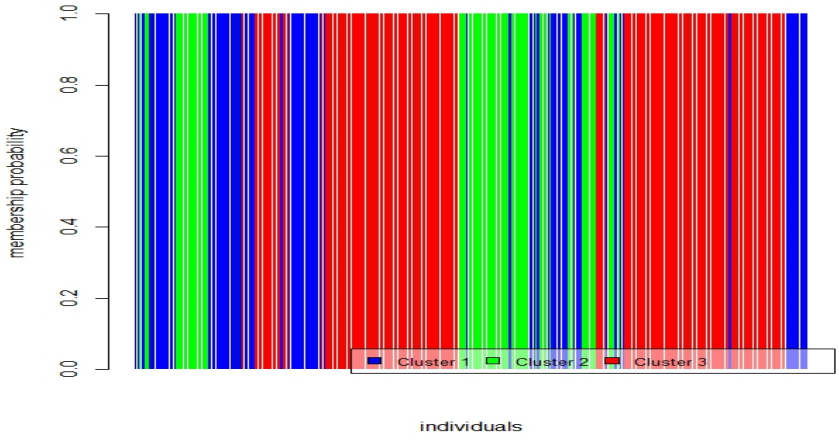


Figure S 3. DAPC population structure plot showing no admixture individuals of 333 shea tree accessions, K =3; each color represents one cluster. X-axis represents individuals and Y-axis represents the membership probability.

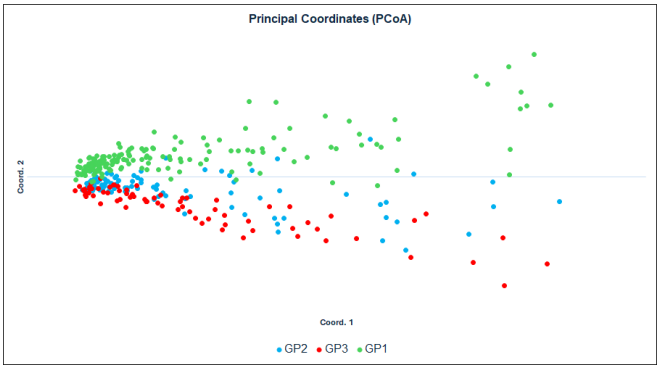


Figure S 4. Scatter plot of the 333 shea trees based on the genetic distance along the first two principal coordinates PCoA 1 and PCoA 2. The colors of the dots along the principal coordinate analysis (PCoA) coordinates 1 and 2 represent the different groups.

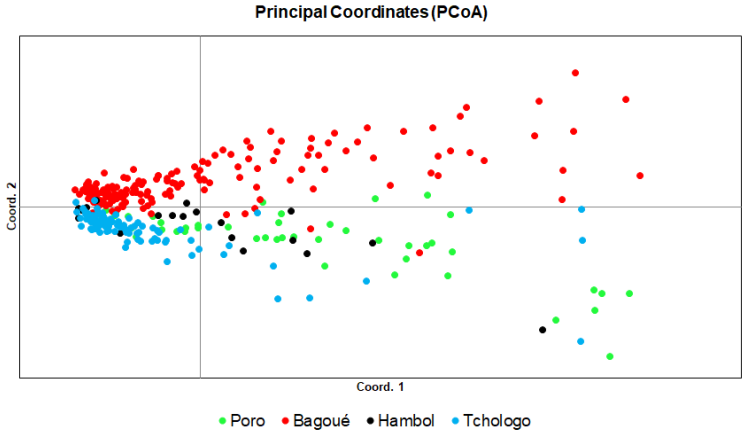


Figure S 5. Scatter plot of the 333 shea trees based on the genetic distance along the first two principal coordinates PCoA 1 and PCoA 2. The colors of the dots along the principal coordinate analysis (PCoA) coordinates 1 and 2 represent shea trees from different locations.

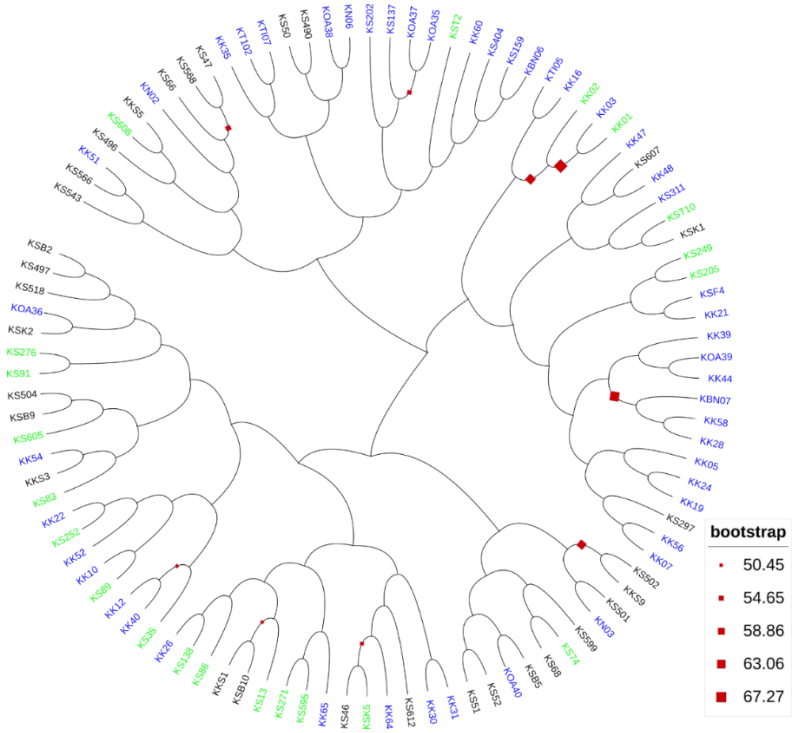


Figure S 6. Neighbor-joining phylogenetic tree with bootstrap values (red squares represent the bootstrap values; 1,000 replicates) showing relationships between the 100 core set of individual superior shea trees based on Nei's genetic distance matrix using 7,559 SNP markers; the color of the labels indicates the genetic group to which a sample belongs (blue for GP1 samples, black for GP2 samples, and green for GP3 samples); analyses were carried out in R, and the generated tree was customized using iTOL

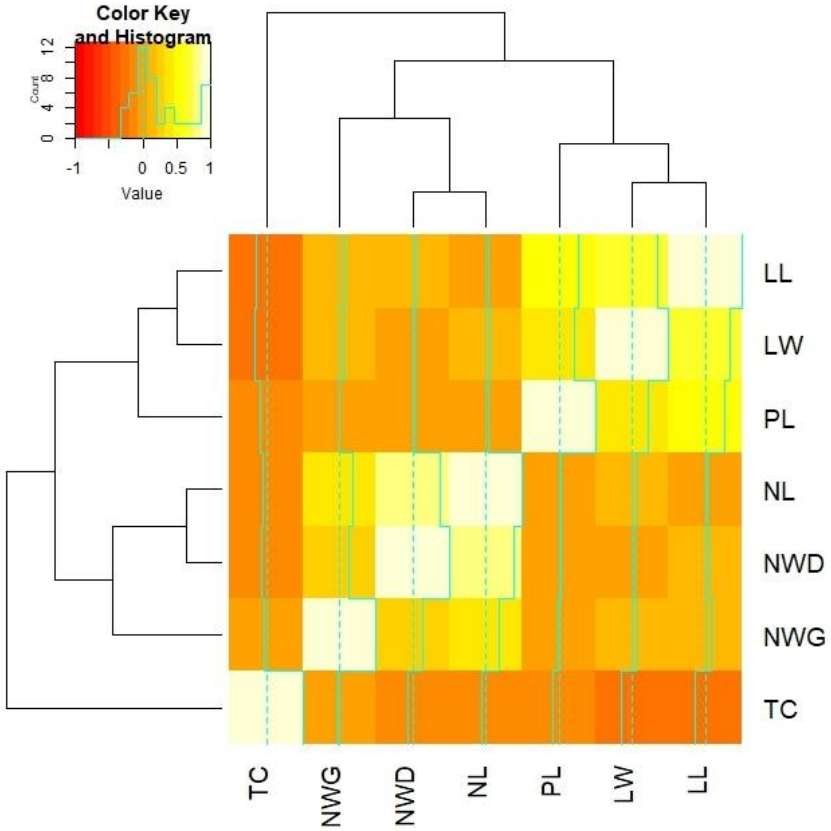


Figure S 7. Pearson correlation heat map showing relationships between morphological traits measured on elite shea trees in Bagoué, Hambol, Poro and Tchologo regions in Northern Côte d'Ivoire. TC: Circumference of trunk at 130 cm above the soil; NWG: Nut weight; NWD: Nut width; NL: Nut length; PL: Petiole length; LW: Limb width; LL: Limb length.

4.3. Conclusions of Chapter 4

In this chapter 4, we present a cost-effective, optimized protocol for high-quality genomic DNA (gDNA) extraction from the recalcitrant shea tree, designed for compatibility with downstream applications such as PCR amplification, enzymatic digestion and Sanger sequencing. This protocol successfully addresses the challenges posed by the high polysaccharide content in shea leaves, which often hinders traditional DNA extraction methods. Amplification tests confirmed the reliability of the protocol, yielding successful results even at low DNA concentrations (as low as 1 pg/µl for leaves). Due to its simplicity and reproducibility, this protocol is highly suited for routine laboratory use in molecular studies of the shea tree and can be readily adapted for other recalcitrant plant species.

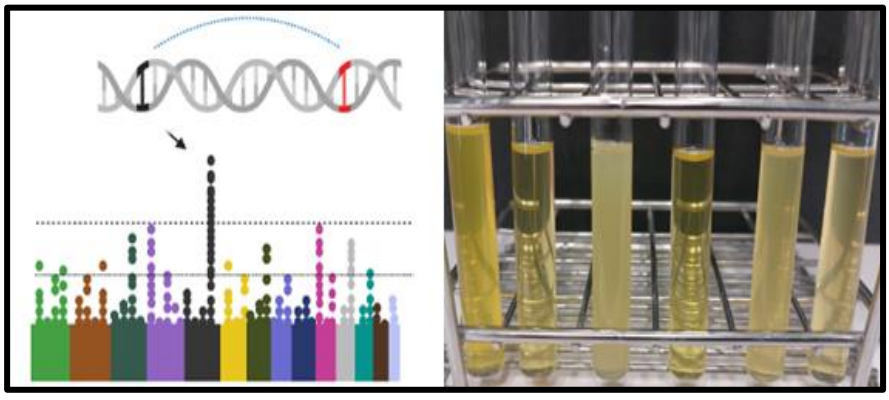
In addition to this methodological advancement, we assessed the genetic diversity and population structure of superior shea tree genotypes from an *in situ* collection using SNP markers. The analysis revealed moderate genetic diversity (HE = 0.26, PIC = 0.24) within the collection. The genotypes were grouped into three distinct genetic clusters, but low genetic differentiation was observed among these groups, attributed to extensive gene flow and the out-crossing nature of the species.

To further enhance conservation efforts, a core collection comprising 30% of the initial population was developed. This subset captured the full genetic diversity and morphological traits of the entire collection, ensuring its representativeness. The core collection provides a robust framework for the sustainable conservation and management of shea tree genetic resources in Côte d'Ivoire.

Overall, these findings contribute significantly to the development of effective conservation strategies and the establishment of a manageable, genetically diverse germplasm. Moreover, the use of SNP markers lays a solid foundation for future genome-wide association studies (GWAS) and marker-assisted selection (MAS) aimed at improving shea tree traits, such as yield and resilience to environmental stresses.

Chapter 5

Genome-wide Association Study of Fat Content and Fatty Acid Composition of Shea Tree (*Vitellaria paradoxa* subsp. *paradoxa*)



Chapter 5. Genome-wide Association Study of Fat Content and Fatty Acid Composition of Shea Tree (*Vitellaria paradoxa* C.F. Gaertn subsp. *paradoxa*)

Shea tree (*Vitellaria paradoxa* subsp. *paradoxa*) is a vital socio-economic and ecological resource across sub-Saharan Africa, with shea butter being its most commercially valuable product due to its extensive use in the food, cosmetic, and pharmaceutical industries. While significant efforts have been made to assess the morphological traits and genetic diversity of superior shea trees (SSTs), there remains a critical gap in understanding the genetic basis of economically important traits, particularly shea butter (SB) content and its fatty acid (FA) composition. These traits are key determinants of shea butter quality, influencing its functional properties and market value, and thus are of paramount importance for targeted breeding and domestication programs.

In this chapter, we present a genome-wide association study (GWAS) aimed at uncovering the genetic architecture underlying SB content and FA composition in SSTs. Understanding the variation in these traits is crucial for advancing shea tree improvement efforts, as it enables the identification of genetic markers that can be harnessed in marker-assisted selection (MAS) to accelerate the development of superior cultivars. To achieve this, shea nuts were collected from 122 out of the 170 previously genotyped SSTs, including 51 trees from the core collection. These samples were sourced from the Hambol, Poro, and Tchologo regions, which represent key shea-growing areas with diverse agro-ecological conditions. Unfortunately, due to the seasonality of shea fruit production, combined with financial and logistical constraints, samples from the Bagoué region were not available for this study.

Through comprehensive genotypic and phenotypic analyses, we identified quantitative trait nucleotides (QTNs) significantly associated with SB content and FA composition. Moreover, the study revealed potential candidate genes involved in the biosynthetic pathways regulating these traits. The identification of these QTNs and candidate genes represents a pivotal step toward precision breeding in shea, providing valuable molecular tools that can enhance the efficiency and accuracy of selection processes.

The findings from this GWAS have broad applications in both scientific research and practical breeding programs. By facilitating the early selection of genotypes with desirable SB and FA profiles, these genetic insights support faster, more precise breeding strategies. This will not only improve the productivity and quality of shea butter but also contribute to the sustainable domestication and commercialization of the shea tree, ultimately benefiting local communities and the broader shea value chain.

Genome-wide Association Study of Fat Content and Fatty Acid Composition of Shea Tree (*Vitellaria paradoxa* C.F. Gaertn subsp. *paradoxa*)

This part is an original research article published in *BMC Genomics* **26**, 164 (2025). https://doi.org/10.1186/s12864-025-11344-z.

Affi Jean Paul ATTIKORA*[1], Kouakou Alfred KOUASSI[2], Saraka Didier Martial YAO[3,4], Dougba Noel DAGO[3], Souleymane SILUE[3], Caroline DE CLERCK[5], Nafan DIARRASSOUBA[3,4], Taofic ALABI[6], Enoch G. ACHIGAN-DAKO[7], Maurie-Laure FAUCONNIER[8], Sabine DANTHINE[2], Ludivine LASSOIS[1]

1. Plant Genetics and Rhizosphere Processes lab, University of Liege, Gembloux Agro Bio-Tech, 5030 Gembloux, Belgium; ludivine.lassois@uliege.be.
2. Food Science and Formulation, University of Liège, Gembloux Agro-Bio Tech, 5030 Gembloux, Belgium.
3. Faculty of Biological Sciences, Department of Biochemistry-Genetics, Educational and Research Unit of Genetic, University of Peleforo Gon Coulibaly (UPGC), BP 1328 Korhogo, Côte d'Ivoire.
4. African Center for Shea Research and Application (CRAK), BP 39 Korhogo, Côte d'Ivoire.
5. AgricultureIsLife, University of Liege, Gembloux Agro Bio-Tech, 5030 Gembloux, Belgium.
6. Functional and Evolutive Entomology, University of Liege, Gembloux Agro Bio-Tech, 5030 Gembloux, Belgium.
7. Genetics, Biotechnology, and Seed Science Unit (GBioS), Department of Plant Sciences, Faculty of Agronomic Sciences, University of Abomey-Calavi, 01 BP 526, Abomey-Calavi, Republic of Benin.
8. Laboratory of Chemistry of Natural Molecules, University of Liège, Gembloux Agro-Bio Tech, 5030 Gembloux, Belgium

* Corresponding author: Affi Jean Paul ATTIKORA; ajpattikora@uliege.be , https://orcid.org/0009-0005-5752-8337

5.1. Abstract

Background: Fat content (FC) and fatty acids (FA) are the most important traits in shea tree breeding, controlled by several genes with relatively small effects. Therefore, determining the genes involved in the biosynthesis of such traits is crucial for improving oil quantity and quality and for the domestication process of the species. To identify the quantitative trait nucleotides (QTNs) controlling FC and FA, we conducted a multi-locus genome-wide association study (GWAS) using six multi-locus GWAS methods for FC and FA in 122 superior shea trees (SSTs). SSTs were genotyped using DArTseq, resulting in 7,559 non-redundant single nucleotide polymorphism markers.

Results: Fat content varied from 36% to 58% with a mean of 50%. Fatty acid composition was 51.26 ± 4.21, 38.76 ± 4.67, 6.45 ± 0.76 and 3.53 ± 0.52 % for oleic, stearic, linoleic and palmitic acids, respectively. A very high negative correlation coefficient (-0.98) was found between stearic and oleic acids. A total of 47 significant QTNs associated with fat-related traits were detected by the GWAS methods. Among these QTNs, 25 were identified as common QTNs based on their detection by multiple GWAS methods. Using superior allele information of the 4 common QTNs associated with fat content in 17 high-fat and 21 low-fat SSTs, we found a higher percentage of superior alleles in SSTs with high FC (47.1%) than in SSTs with low FC (14.3%). Pathway analysis of the common QTNs identified 24 potential candidate genes likely involved in the biosynthesis of FC and FA composition in shea tree seeds.

Conclusions: These findings will contribute to the discovery of the polygenic networks controlling FC in shea tree, improve our understanding of the genetic basis and regulation of FC, and be useful for molecular breeding of high-fat shea tree cultivars.

Keywords: Fatty acid components, multi-locus GWAS, *Vitellaria paradoxa*, Quantitative Trait Nucleotide, Candidate gene, Superior Allele proportion, Single Nucleotide Polymorphism

5.2. Introduction

Shea tree (*Vitellaria paradoxa*), a member of the Sapotaceae family, is an economically important species in the Sudano-Sahelian zone (67). It consists of two subspecies: *V. paradoxa subsp. paradoxa* from West and Central Africa and *V. paradoxa subsp. nilotica* found in East Africa (10). Shea butter, extracted from shea's kernels, is a key resource in local economies and international markets, serving as an essential ingredient in the food, cosmetic, and pharmaceutical industries (59). The extraction process for shea butter is still not standardized and is typically performed by manual or semi-mechanized methods (135). Regardless of factors such as origin, genetic variation and climatic conditions, the qualitative and quantitative composition of shea butter is mainly related to the extraction process (131,135). Therefore, the Codex Alimentarius specifies quality parameters for unrefined shea butter, allowing its classification into two categories based on water content, free fatty acids, peroxide value and insoluble impurities: butter for direct consumption (grade 1a) and butter for use in the food industry (grade 1b) (135).

Shea butter predominantly consists of four main fatty acids: approximately 48% oleic acid (C18:1), 40% stearic acid (C18:0), 5% linoleic acid (C18:2), and 3% palmitic acid (C16:0) (129). Its high stearic acid and oleic acid content makes it particularly suitable for applications in the chocolate and confectionery industries. Furthermore, shea butter is the most widely utilized commercial source of Sat-O-Sat (Sat: saturated fatty acid and O: oleic acid), a key component in the production of confectionery fats (238).

Despite its economic and industrial importance, traditional shea cultivation faces significant challenges. These include the species' long juvenile phase of 15-20 years (158) and the high heterogeneity of natural populations due to its outcrossing nature (28). This variability often leads to inconsistencies in oil quality and yield, hindering its full market potential (25).

V. paradoxa is a diploid species (2n = 24) with a genome estimated at 658.7 Mbp containing over 38,000 coding genes (25). This genomic information provides a strong foundation for the study of genetic traits associated with agronomic value. Through participatory surveys, potential superior shea trees have been identified based on traits such as tree yield, fruit size, pulp taste, and early flowering (60). Previous studies have characterized these superior trees at both morphological and molecular levels (94,122,205). However, their fat content and fatty acid composition, key traits that influence shea butter quality, have not been thoroughly investigated.

Determining these traits in superior shea trees and using advanced genomic tools, such as genome-wide association studies (GWAS), will provide new opportunities to identify quantitative trait nucleotides (QTNs) associated with fat content and fatty acid profile. Multi-locus GWAS methods have proven to be highly efficient and accurate in identifying genetic markers associated with complex traits. Unlike single-marker models, these approaches reduce false positives and provide a comprehensive understanding of genetic influences.

In this study, multi-locus GWAS approaches were applied to dissect the genetic basis of fat content and fatty acid composition in a genetically diverse population of superior shea trees. The specific objectives were: (I) to determine the fat content and fatty acid composition; (II) to identify significant QTNs associated with fat content and fatty acids; and (III) to discover candidate genes controlling fat-related traits in shea trees.

Ultimately, the results of this research, combined with the recently developed affordable and efficient DNA extraction protocol described Attikora et al (239), are expected to increase scientific knowledge of superior shea trees and advance shea tree breeding programs. This will increase the economic potential of the species by addressing market challenges such as variability in oil quality, in terms of fatty acid profiles, and yield. To our knowledge, this is the first genome-wide association study (GWAS) focused on fat content and fatty acid composition in Vitellaria paradoxa. These findings will serve as a foundation for future genetic improvement efforts, benefiting local farmers and global industries that rely on shea butter.

5.3. Materials and Methods

5.3.1. Plant Materials and Leaf Sampling for DNA Extraction

An initial population of 170 genotyped mature superior shea trees (SSTs) from the Hambol, Poro and Tchologo regions of Côte d'Ivoire were considered for shea fruit sampling. The savannas of northern Côte d'Ivoire, where shea trees grow, are divided into the Sudanese savanna (Poro, Tchologo regions) with monomodal rainfall (1,200 mm/year) and the sub-Sudanese savanna (Hambol region), a transitional zone with bimodal rainfall (1,050 mm/year). Average annual temperatures are approximately 27°C, and vegetation includes wooded and grassy savannas with gallery forests along waterways. Soils are predominantly ferralitic, with subclasses including soils on basic rocks, tropical ferruginous soils, and hydromorphic soils. Key crops include cotton, cashew nuts, and mangoes, while common tree and shrub species include *Vitellaria paradoxa*, *Paria biglobosa*, *Pilliostigma thonningii*, and others (122).

The number of initial SSTs in each region was based on the density of the shea population. In addition, the selection of superior shea trees consisted of a participatory survey in which farmers were allowed to select SSTs based on specific criteria, including high fruit yield, large fruit size, early flowering each year, and periodicity of fruit production (205).

A total of 122 mature SSTs from the initial population (170 SSTs) were selected for fruit sampling based on the initial number of SSTs in each region. In fact, in Hambol region, 100% of the SSTs were sampled (25/25 SSTs) and in Poro region, 48 out of 53 SSTs were sampled since any fruit was found under the five remaining SSTs. Regarding the region of Tchologo, only 50% of the initial SSTs, representing 48 out of 97, were randomly selected for fruit sampling for this study. The sequences of the genotypes include in this study were uploaded in NCBI as Sequence Read Archive (SRA) under the BioProject "PRJNA1167878". Leaves sampling and DNA extraction were described in our previous study (122). One hundred fruits (100) were randomly collected from each genotype for shea butter extraction. The pulp was removed, and

the nuts were boiled in water at 98 ± 2°C under atmospheric pressure for 15 minutes. They were then sun-dried for two weeks and the coat was removed from the kernels. The kernels were oven-dried at 40°C for 24 hours, then ground into a paste using a high-speed laboratory grinder (FRITSCH, 19.1020/00426, ROHS, Oberstein, Germany). The obtained pastes were stored at 4°C in sealed plastic containers under vacuum until fat extraction.

The fat extraction process was carried out in the Food Science and Formulation Laboratory at TERRA, ULiège GxABT, Belgium. The extraction of the fat was carried out according to the maceration method described by Kaoussi et al (240) to preserve the physicochemical properties of the fat while also enhancing the yield. One hundred grams (100 g) of the sample were mixed with 200 mL of hexane in a 500 mL capacity Duran flask and heated to 40°C while stirring for 90 minutes using a temperature-controlled heating agitator system. The extracts were then centrifuged at 7000 rpm for 15 minutes at 30°C using a Jouan C312 centrifuge (France). Finally, the clarified supernatant was filtered using a vacuum filtration setup consisting of a Buchner funnel with Whatman No. 1 filter paper (Ø125 mm) placed on an Erlenmeyer flask and separated by a gasket. A vacuum pump is attached to the setup to generate the vacuum. The extraction process was repeated three times on the same matrix to deplete it of fats, with the filtrates collected and combined in a 1000 mL flask. The solvent was then removed using a rotary evaporator (Büchi Labortechnik AG, Flawil, Switzerland), and any remaining solvent traces were eliminated by nitrogen flushing. The extracted shea butters were stored in the dark at −20°C until analysis. All extractions were conducted in triplicate for each sample.

5.3.2. Fatty Acid Composition

The fatty acid composition of the extracted shea butter was determined following transesterification with BF3, according to the AOCS Ce 2-66 method. Fatty acid methyl esters (FAME) were analyzed using a GC ULTRA gas chromatograph (Thermo Scientific Interscience) equipped with a flame ionization detector (FID) and an HP-Innowax column (Agilent Technology) of 30 m x 0.5 µm x 0.25 µm (length x thickness x diameter). The injection was carried out in splitless mode (splitless time: 2 min) at 250°C. Helium served as the carrier gas, with a constant flow rate of 1 mL/min. The temperature program was set as follows: starting at 50°C with a 1-minute hold, then increasing to 150°C at a rate of 30°C/min, followed by a rise to 240°C at 5°C/min with a 25-minute hold. The FID was set to 250°C. Fatty acid methyl esters were identified by comparing their retention times with those of pure reference standards. Analyses were performed in triplicate and the mean values for each sample were considered.

5.3.3. Phenotypic Data Analysis

The fat related-traits data were analyzed using R software, version 4.3.3. An analysis of variance (ANOVA) was performed to determine the variations within and among the genotypes. The correlation coefficients between the studied traits were calculated and presented in graphical form. Additionally, a principal component analysis (PCA) was performed to structure the studied traits.

5.3.4. SNP Genotyping Data Analysis

The methods used for SNP genotyping and mapping were described in previous study (122). Sequencing was performed using genotyping by sequencing DArTseq technology. DArTseq SNP markers were aligned to the Vitpa_HiCP0_Assembly reference genome (https://bioinformatics.psb.ugent.be/orcae/overview/Vitpa) to locate the corresponding chromosomal positions. A total of 42,705 SNP markers were mapped. To discard low-quality SNP and ensure data integrity, markers with more than 20% missing data were removed. In addition, minor allele frequency (MAF) SNPs with less than 5% were considered rare and were therefore excluded. A final dataset consisting of 7,559 SNP markers was used for further analysis.

5.3.5. Analysis of Population Structure and Linkage Disequilibrium

Bayesian clustering approach was performed using STRUCTURE 2.3.4 software to investigate the structure of the shea panel based on an admixture model excluding the LOCPRIOR option was used (241). The algorithm of the model-based clustering is to identify genetic groups in terms of K values. The analysis was performed in 10 runs, with successive values of K ranging from 1 to 10 and burn-in period of 50,000 and 100,000 Markov-chain Monte Carlo (MCMC) replicates. The optimal K value was determined based on the delta-K [Δ(K)] method using the rate of change in [Ln(P(D)] between successive K values. An unweighted neighbor-joining (NJ) tree was constructed based on a dissimilarity matrix (DM) estimated from the 7,559 SNPs using TASSEL 5.2.80 (242). In addition, discriminant analysis of principal component (DAPC) was performed using the "find.clusters" function of the "adegenet" package in R software version 3.4. 4 to access the structure of the Shea panel. Genome-wide linkage disequilibrium (LD) was generated by plotting average r^2 (correlation frequency among SNPs) values as a function of genetic distance in base pairs (bp) against the twelve chromosomes across the shea tree genome using the TASSEL 5.2.80. The LD decay plot was calculated in R.

5.3.6. Genome-wide Association Study

The R platform mrMLM 4.0.2 (https://cran.r-project.org/web/packages/mrMLM.GUI/index.html) for ML-GWAS was used to map candidate QTNs. Six multi-locus GWAS methods within the mrMLM R package were used to identify significant QTNs, including mrMLM (243), FASTmrMLM (244), FASTmrEMMA (245), pLARmEB (246), pKWmEB (247), and ISIS EM-BLASSO (248). All parameters were set to default values, and the critical LOD score was set to 3 for robust QTNs in the final step. In this study, the six multi-locus GWAS methods were applied because they have demonstrated their advantages over single-locus GWAS methods. In addition, the combination of multi-locus methods is also highly recommended to improve the power and robustness of GWAS. To control false positives, the Q + K model, in which are included the population structure matrix (Q) and the kinship matrix (K), were used in the analysis. The kinship matrix was calculated using the R package mrMLM 4.0.2.

5.3.7. Superior Allele Analysis

We considered the QTNs detected by at least two ML-GWAS methods as common QTNs. Based on the effect values of each common QTN and the genotype for code 1, we could determine the superior alleles of each QTN. If the QTN effect value is positive, the genotype for code 1 is the superior allele; if the effect value is negative, the alternative genotype is the superior allele. For each QTN, the proportion of superior alleles in 38 SSTs, consisting of 17 with high fat content and 21 with low fat content, was equal to the number of genotypes containing the superior allele divided by the total number of genotypes. For each genotype, the proportion of superior alleles in these QTNs was calculated as the number of superior alleles divided by the total number of QTNs. These 38 SSTs were selected based on the average fat content.

5.3.7. Candidate Genes Annotation

QTNs detected by ML-GWAS methods were selected as candidate genes. To account for putative genes associated with traits, a window range of 10 kb (upstream and downstream) was defined; and genes were searched from the *V. paradoxa* Whole Genome v2.0 Assembly and Annotation in the ORCAE database (https://bioinformatics.psb. ugent.be/orcae, accessed on July 26, 2024), with a search for candidate genes associated with fat content traits. The gene name, description, and AGPv4 coordinates with its protein were then searched in the *Vitellaria paradoxa* reference genome database. The putative functional candidate genes associated with the corresponding SNPs were then annotated according to any initially annotated genes from other species.

5.4. Results

5.4.1. Fat content and fatty acid composition of superior shea trees

A significant degree of variability was observed in the fat content of the superior shea trees (**Figure 24 and Table S 9**). The mean fat content of the 122 superior shea trees was found to be 49.7%, ranging from 36.2-58.1% (**Figure 24a**). A summary of the fatty acid composition of the shea genotypes is presented in **Table S 9**. This study examined the four main fatty acids found in shea butter: palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), and linoleic acid (C18:2). Oleic acid (51.3%) and stearic acid (38.8%) were the most abundant, with proportion ranging from 40.3-65.7% and 22.3-50.5%, respectively (**Figure 24c and Figure 24d)**. Linoleic acid followed with an average of 6.4%, ranging from 4.8 to 9% (**Figure 24e**).

Furthermore, the coefficient of variation (CV) of fat content (7.44%) and oleic acid (8.21%) observed in superior shea trees was found to be low, while CV of palmitic acid (14.73%), stearic acid (12.05%) and linoleic acid (11.78%) was medium (**Table S 9**). This indicates that the panel of superior shea trees exhibited low to moderate variation.

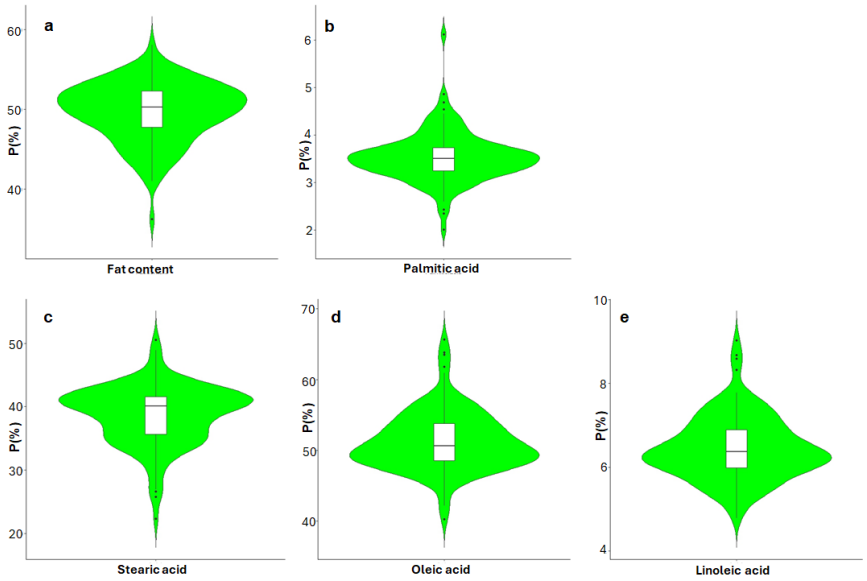


Figure 24. Box plots of the distributions of the proportion relative-fat traits: a) Fat content, b) palmitic acid, c) stearic acid, d) oleic acid, and e) linoleic acid, P (%): proportion

A positive correlation was observed between fat content and stearic acid (r = 0.31). In contrast, negative correlation was observed between fat content and C16:0 (r = -0.13), C18:1 (r = -0.25), and C18:2 (r = -0.43). Moreover, a strong negative correlation was found between stearic acid and oleic acid (r = -0.98) (**Figure 25**).

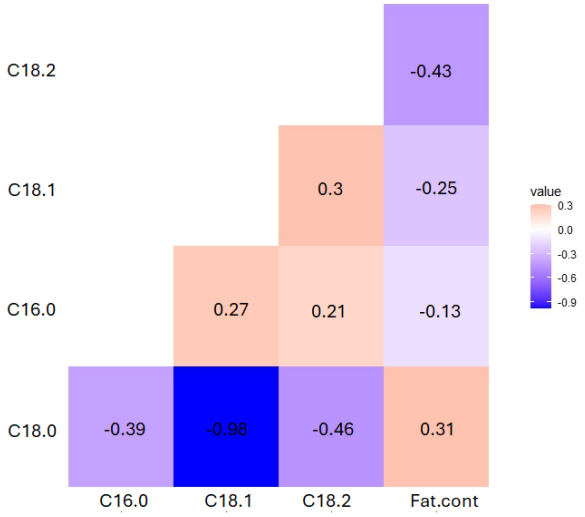


Figure 25. Correlation between five fat-related traits (C18.0: Stearic acid, C18.1: Oleic acid, C18.2: Linoleic acid and Fat.cont: Kernel Fat content) of the 122 superior shea trees. Color in the boxes indicates the value of the correlation relationship.

The fat content and fatty acid composition of superior shea trees in different regions are shown in **Table S 10**. Slight variations in fat content, palmitic acid, and linoleic acid were observed among superior shea trees across different regions (**Table S 10, Figure 26**). The mean fat content was 49% in Hambol and 49.9% in both the Poro and in Tchologo regions (**Figure 26a**). Similarly, the proportion of palmitic acid was 3.6% in Hambol and 3.5% in both Poro and Tchologo regions (**Figure 26b**). Mean linoleic acid content was 6.6% in Hambol, 6.3% in Poro, and 6.6% in Tchologo (**Figure 26e**). However, significant quantitative variation was noted for stearic acid and oleic acid. The mean proportion of stearic acid was 36.9% in Hambol, 39.8% in Poro, and 38.6% in Tchologo (**Figure 26c**). Oleic acid proportions were 52.9% in Hambol, 50.4% in Poro, and 51.3% in Tchologo (**Figure 26d**).

An analysis of variance (ANOVA) was performed to evaluate the effect of geographical regions on the fat content and fatty acid composition of superior shea trees. The results indicated statistically significant differences in the levels of stearic acid ($p = 0.03$) and oleic acid ($p = 0.05$) across regions (**Table S 10**). However, no significant variations were found in fat content, palmitic acid, or linoleic acid between the regions.

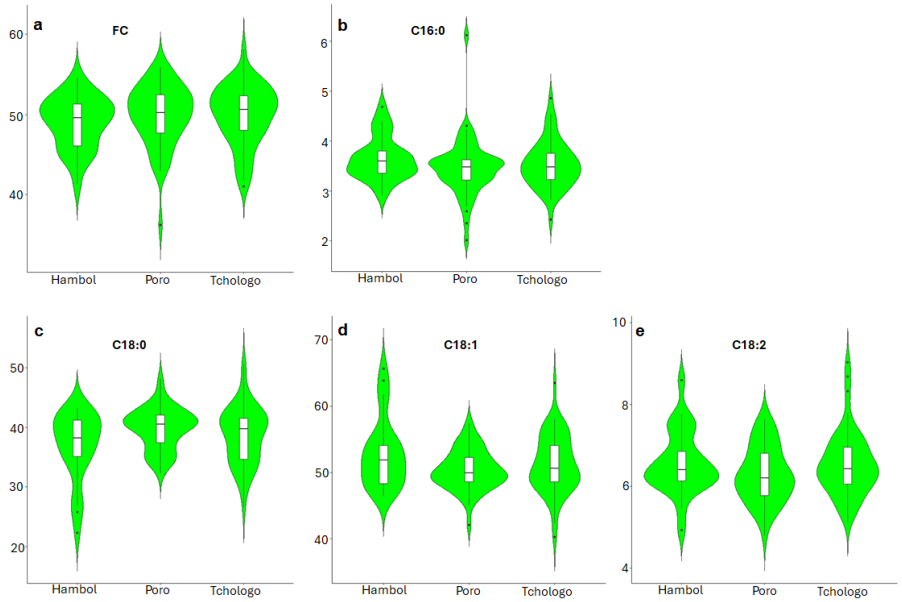


Figure 26. Box plots of the distributions of Fat content and Fatty Acid Composition of Superior Shea Trees by region; a) Fat content, b) palmitic acid, c) stearic acid, d) oleic acid, and e) linoleic acid

5.4.2. Principal component Analysis (PCA)

PCA was conducted to identify the variables that significantly influence the principal components (PCs), thereby explaining the variability in the data set. The PCA generated five principal components (PCs) in total, with the first two main PCs (eigenvalues > 1) accounting for 72.4% of the total variation (**Figure S 8**). PC1 explained 52% of the total variance, mainly driven by C18:0 (positive loading) and

opposed by C18:1 variable (negative loading). The residual variance in PC2 (20.46%) was mostly explained by C18:2 (negative loading) and fat content (positive loading).

5.4.3. Population Structure and Linkage Disequilibrium

The population structure of 122 superior shea trees was analyzed using the 7559 high-quality genome-wide SNP markers. The density and distribution of SNPs on each chromosome across the shea tree genome are presented in **Figure 27**.

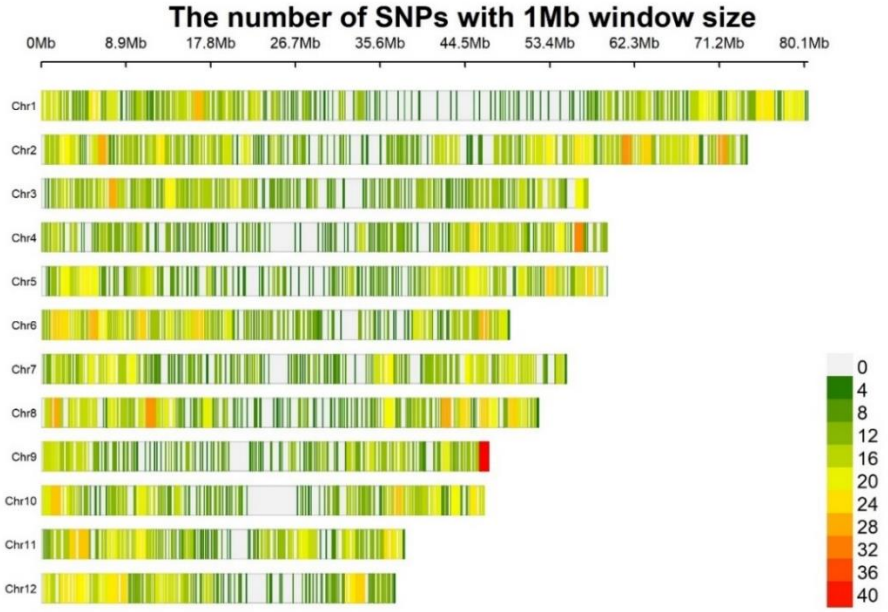


Figure 27. The number and size of SNPs within 1 Mb window size of *V. paradoxa* genome

Model-based simulation of population structure showed the highest peak at K = 3 as the number of genetic groups plotted against delta K (ΔK) by Structure Harvester (**Figure 28a**). This indicates the presence of three genetic groups (GG1, GG2 and GG3) in the reference set (**Figure 28b**). GG1 was the largest with 52 superior shea trees, including 33 pure types and 19 admixture types that constituted 42.62% of the shea panel. GG2 with 30 SST included 20 pure types and 10 admixtures constituting 24.59% of the total accessions and GG3 with 40 SST included 21 pure types and 19 admixtures, 32.79% of the entire population. The estimated fixation index (F_{ST}) was 0.012, 0.018 and 0.021 for GG1, GG2 and GG3 respectively. The highest allelic frequency of divergence was found between GG2 and GG3 (**Table S 11**).

Consistent with the findings from Bayesian model-based simulation of population structure, the discriminant analysis of principal components (DAPC) also suggested three distinct clusters based on the value of BIC (822.91) (**Figure 28c**). Cluster I had 24 accessions, cluster II had 66 accessions and cluster III had 32 accessions. Similarly, the unweight neighbor-joining (NJ) tree method clustered the accessions into three groups. Cluster I had 89 accessions, cluster II had 24 and cluster III had 9 accessions. Individuals from each region were found in the three clusters (**Figure 28d**).

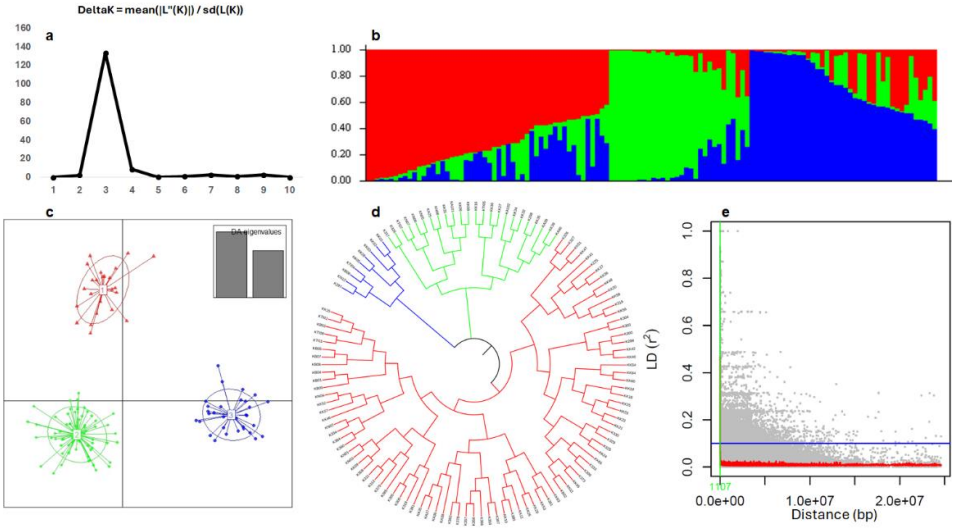


Figure 28. Population structure, phylogenetic analysis and linkage disequilibrium (LD). **a** Delta K for various numbers of clusters (K); **b** Population structure inferred into three subgroups (K = 3) based on delta K values; **c** Scatter plot of DAPC showing the genetic networks for the three groups; **d** Phylogenetic analysis using the neighbor-joining method grouped into three clusters; **e** LD decay plot from all population. The x-axis represents the physical distance, and the y-axis represents the average pairwise correlation coefficient (r2) of SNPs.

The average distance of LD decay (r^2) based on 7,559 SNP markers in the whole genome were calculated. The value of r^2 declined rapidly to reach a plateau at 0.01. The corresponding distance was considered as the average distance of LD decay in this population. The overall LD decay was very low ($r^2 > 0.2$) at a physical distance of 1107 pb in shea tree germplasm (**Figure 28e**).

5.4.4. Multi-Locus Genome-Wide Association Study Analysis of fat-related traits in *V. paradoxa*

All five fat-related traits were analyzed using six ML-GWAS methods (mrMLM, FASTmrMLM, FASTmrEMMA, pKWmEB, pLARmEB and, ISIS EM-BLASSO) to identify QTNs. A total of 47 significant QTNs (LOD≥3) were identified on all twelve chromosomes (**Table S 12 and Figure 29**). Of these, 21, 16, 5, 29 and 26 were detected with mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, and ISIS EM-BLASSO respectively. In contrast, pKWmEB ML-GWAS method did not identified any QTN. Of the detected QTNs, 8, 9, 18, 12, and 9 were defined to be associated with FC, C16:0, C18:0, C18:1, and C18:2 respectively. It should be noticed that 9 of the detected QTNs were associated with both C18:0 and C18:1 (**Table S 12).**

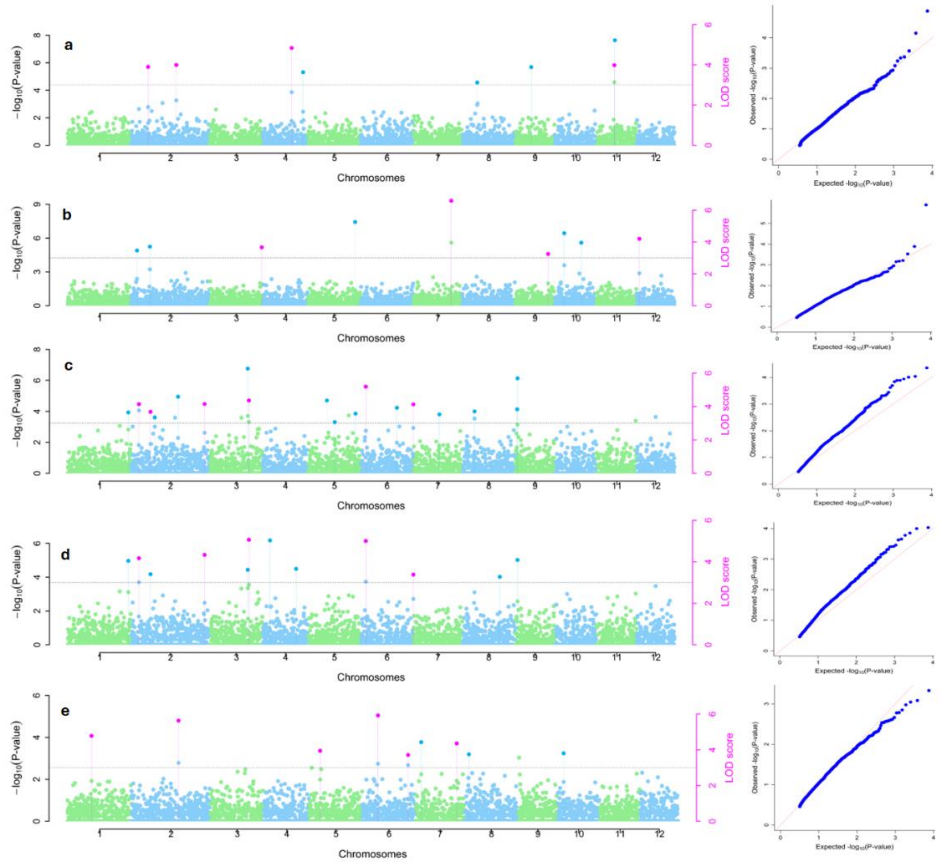


Figure 29. Manhattan and QQ plots for five fat-related traits in GWAS using mrMLM v4.0.2. with 7559 SNP markers. Left is Manhattan plot, while right is QQ plot. Loci discovered by multiple methods together are marked with pink dots in the Manhattan diagram, those discovered by a single method are marked in blue above the horizontal line that indicates a critical LOD score of 3.0. **a** Fat content; **b** Palmitic acid; **c** Stearic acid; **d** Oleic acid; **e** Linoleic acid.

A total of 25 identified QTNs were detected by at least two ML-GWAS methods (20 QTNs) or/and co-associated with two traits (9) or/and were flanking near a putative coding region (10 QTNs) that have a crucial role in *V. paradoxa* lipid biosynthesis. These QTNs were then chosen as common QTNs for the five fat-related traits.

For the 20 QTNs detected by at least two ML-GWAS methods, 6 were found to be tightly associated with both C18:0 and C18:1 (**Table S 12**). For the remaining QTNS, 4, 4, and 6 were found to be associated with FC, C16:0 and C18:2, respectively. The six common QTNs detected for both C18:0 and C18:1 were located on chromosomes 2, 3, and 6. For FC, the 4 common QTNs were distributed on chromosomes 2, 4 and, 11. The 4 common QTNs of C16:0 were distributed on chromosomes 3, 7, 9 and 12.

A total of 6 QTNs which commonly associated with C18:2, spread over 1, 2, 5, 6, and 7 chromosomes. Of these, fourteen QTNs were co-detected by at least 3 three ML-GWAS methods while 10 QTNS were co-identified by at least 4 ML-GWAS methods. Notably, *q6_3096993* and *q4_46157457* were determined across all five ML-GWAS approaches.

5.4.6. Distribution of Superior Alleles in Superior Shea Trees

As fatty acids composition depends on the fat content, the 4 common QTNs associated with fat content were used to explore the proportion of superior alleles in 38 superior shea trees (SSTs). Hence, 17 genotypes with higher fat content were considered as SSTs with higher phenotypic values while 21 genotypes with low fat content were considered as SSTs with lower phenotypic values. Genotypes with fat content above the addition of the average fat content and the standard deviation of the shea tree panel were considered as SSTs with higher phenotypic values. In contrast, genotypes with fat content below the subtraction of standard deviation in the average fat content were considered as SSTs with lower phenotypic values. Therefore, 17 SSTs had higher phenotypic values (53.51-58.09%) and 21 had lower phenotypic values (36.24-46.03%). For each of the 21 SSTs with lower fat content, the proportion of superior alleles ranged from 0 to 25%, while the proportion of superior alleles in the 17 SSTs with higher fat content ranged from 25 to 100%. Thus, the superior shea trees with high fat content have more superior alleles than the shea trees with low fat content (**Figure 30**).

Figure 30. Heat map of the superior allele distribution for the 4 common QTNs associated to fat content in 17 high-fat (lower part of the figure) and 21 low-fat (upper part of the figure) superior shea trees. Green and white colors represent superior and inferior alleles, respectively.

However, it is observed that there is no proportional relationship between the fat content and the percentage of superior alleles among SSTs with high phenotypic value. For example, sample K359 exhibited a fat content of 58.09% with 25% of superior alleles, while sample K360 displayed a fat content of 54.42% with 100% of superior alleles (**Table 15**).

Table 15. Phenotypic averages of kernel fat content and proportion of superior alleles in 38 genotypes across 4 common QTNs.

Individual	FC(%)	PSA(%)	Individual	FC(%)	PSA(%)	Individual	FC(%)	PSA(%)
K359	58.09	25	K297	53.86	25	**KK11**	44.42	25
KK46	55.86	75	K328	53.81	50	**KK03**	44.18	0
KK60	55.54	25	KK19	53.52	75	**K300**	43.79	25
KK08	55.54	25	K387	53.51	25	**KK16**	43.78	25
K385	55.11	50	**KK54**	46.03	0	**KK22**	43.55	25
KK41	55.1	25	**KBN4**	45.97	25	**K379**	43.3	25
KK48	54.56	50	**K358**	45.78	0	**KK24**	42.92	0
KBN5	54.49	50	**KBN3**	45.43	25	**K308**	42.62	0
K360	54.42	100	**KA35**	45.36	25	**KA38**	41.72	0
KK29	54.33	50	**KTI07**	45.27	0	**KBN9**	41.28	25
KA40	54.28	50	**K319**	44.71	25	**K334**	41.01	25
K304	54.24	50	**KTI05**	44.65	0	**KK02**	36.24	25
K380	54.15	50	**KN06**	44.42	0	**-**	-	-

FC: Fat content, PSA: proportion of superior allele, characters in bolt represent genotypes with low fat content.

Based on the superior allele information of these 4 important QTNs within the 38 superior genotypes, the PSAs for QTNs ranged from 0 to 82.35 %. Among them, 2 QTNs showed a PSA above 50%, while the remaining 2 QTNs showed PSA lower than 20%.

Within the 17 SSTs with high fat content, the PSA for QTNs ranged from 11.76 to 82.35%. Two QTNs had PSA values higher than 60% while the other two QTNs had PSA value lower than 50%. The range of PSAs of QTNs was 0–33.33% in the 21 SSTs with lower phenotypic values. All the QTNs had PSAs lower than 50% (**Table 16** and **Figure 30**). The number of QTNs with superior alleles proportion, higher than 50%, was more in the 17 SSTs than in the 21 SSTs with lower phenotypic values.

We further explored the superior alleles within the 17 SSTs with high fat content for stearic and oleic acids as they are the major fatty acids in *V. paradoxa*. As results, 44.12% of superior alleles were found for C18:0 while 27.45% of superior alleles were found for C18:1 (**Table 16**).

Table 16. Superior alleles and their proportions of 25 common QTNs and the five fat-related traits in 17 high FC SSTs and 21 low FC SSTs.

Trait	QTN	Superior genotypes	Chr	QTN position	LPV (%)		HPV (%)	
FC	*q4_46157457*	CC	4	46157457	33.33	14.29	82.35	47.06
	q11_13733557	GG	11	13733557	0		29.41	
	q2_13023583	TT	2	13023583	23.81		64.71	
	q2_49928619	TT	2	49928619	0		11.76	
C16 :0	*q3_57006090*	CC	3	57006090	61.9	41.61	64.71	39.71
	q7_45956782	GG	7	45956782	0		0	
	q9_42857113	TT	9	42857113	9.52		17.65	
	q12_1933151	CC	12	1932377	95.24		76.47	
C18 :0	*q2_70518009*	AA	2	70518009	80.95	35.42	76.47	44.12
	q2_5991246	AA	2	5991246	0		23.53	
	q6_3096993	GG	6	3096993	74.43		82.35	
	q3_45009201	CC	3	45009201	23.81		35.29	
	q6_48934498	CC	6	48934498	4.76		17.67	
	q2_15731361	GG	2	15731361	28.57		29.41	
	q9_1241251	TT	9	1241251	4.76		0	
	q3_43202766	TT	3	43202766	0		0	
	q1_78304163	TT	1	78304163	38.1		35.29	
	q6_28898438	TT	6	28898438	28.57		5.88	
	q9_1135448	CC	9	1135448	85.71		88.24	
C18 :1	*q2_5991246*	GG	2	5991246	80.95	43.65	52.94	27.45
	q6_3096993	CC	6	3096993	4.76		5.88	
	q2_70518009	GG	2	70518009	9.52		0	
	q3_45009201	TT	3	45009201	38.1		23.53	
	q6_48934498	AA	6	48934498	80.95		58.82	

	q2_15731361	AA	2	15731361	47.62		23.53	
	q9_1241251	CC	9	1241251	95.24		82.35	
	q3_43202766	CC	3	43202766	76.19		64.71	
	q1_78304163	GG	1	78304163	23.81		23.53	
C18 :2	*q1_20786407*	TT	1	20786407	42.86	42.18	17.65	28.57
	q2_51742656	GG	2	51742656	33.33		5.88	
	q5_ 9317929	TT	5	9317929	71.43		58.82	
	q6_43871960	CC	6	43871960	19.05		17.65	
	q6_10189604	TT	6	10189604	4.76		0	
	q7_48678186	TT	7	48678186	80.95		64.71	
	q10_3420702	CC	10	3420702	42.86		35.29	

Chr: chromosome, LPV: low fat content, HPV: high fat content

5.4.7. Potential Candidates Genes and annotations

To predict candidate genes for loci significantly associated with fat content and fatty acid composition, the detected QTNs were used to confirm the genomic regions in *V. paradoxa* reference genome. We identified 24 putative genes that possibly influence fat content and fatty acid composition (**Table 17**). These putative genes were associated with eight gene/protein families involved in the fatty acid biosynthesis of shea nuts. For fat content, four putative genes *Vitpa04g20550*, *Vitpa11g11370*, *Vitpa02g10450*, *Vitpa02g28870*, corresponding to 3 gene/protein families were discovered: Long Chain Acyl-CoA Synthetase (*LACS*) on chromosome 4 at locus *q4_46157457*, SNF1-related protein kinase regulatory subunit beta-2 (KINB2) on chromosome 11 at locus *q11_13733557*, 3-Ketoacyl-ACP synthase (*KAS*) on chromosome 2 at loci *q2_13023583* and *q2_49928619* (**Table 17**). These genes are all involved in the fatty acid biosynthesis pathway.

For the Fatty acids (palmitic, stearic, oleic and linoleic), 20 putative genes corresponding to 7 gene/protein families were discovered in shea butter: Fatty acid desaturases (*FADs*) on chromosomes 1, 3, 5 and, 6; Acyl-CoA-binding protein (*ACBP*) on chromosomes 1, 2, 7 and 10; Long Chain Acyl-CoA Synthetase (*LACS*) on chromosomes 3, 6, 9 and 12; Acetyl-CoA Carboxylase (*CAC*) on chromosomes 2 and 6 ; Fatty Acid Export (*FAX*) on chromosome 9; Biotin carboxyl carrier protein of acetyl-CoA carboxylase (*BCCP*) on chromosome 6; and 3-Ketoacyl-ACP synthase (*KAS*) on chromosome 2 **(Table 17**).

For palmitic acid, three family of genes were found; ACBP1 on chromosome 7; LACS on chromosomes 9 and 12 and; FAD2 on chromosome 3. Stearic and oleic acid identified the same putative genes. Four gene families were discovered: CAC, KAS, LACS, BCCP, ACBP, FAX and FADs. Finaly, FADs, LACS, CAC and ACBP gene families were identified for linoleic acid (**Table 17**). The genes identified from fatty acid are involved in Fatty acid biosynthesis and Fatty acid transmembrane transport pathways (**Table 17**).

Table 17. Gene annotation for the common QTNs for the fat content and fatty acid composition of *V. paradoxa.*

Trait	QTN	Chr	pos	Gene ID	G.O	Function
FC	*q4_46157457*	4	46157457	Vitpa04g20550 Long chain Acyl-CoA Synthetase 5: LACS5	GO:0001676	Long-chain fatty acid metabolic process
	q11_13733557	11	13733557	Vitpa11g11370 SNF1-related protein kinase regulatory subunit beta-2: KINB2	GO:0006633	Regulation of fatty acid synthesis by phosphorylation of acetyl-CoA carboxylase.
	q2_13023583	2	13023583	Vitpa02g10450 3-ketoacyl-CoA synthase 1: KASI	GO:0016021	Fatty acid biosynthesis
	q2_49928619	2	49928619	Vitpa02g28870 KASII: Ketoacyl-ACP synthase 2	GO:0006633	Fatty acid biosynthetic process
C16 :0	*q3_57006090*	3	57006090	Vitpa01g27780 FAD2: Fatty acid desaturase 2	GO:0006629	Lipid metabolic process
	q7_45956782	7	45956782	Vitpa07g24400 ACBP1: Acyl-CoA binding protein 1	GO:0005515	Protein binding
	q9_42857113	9	42857113	Vitpa09g21310 Long chain Acyl-CoA Synthetase 6: LACS6	GO:0022857	Transmembrane transporter activity
	q12_1933151	12	1933151	Vitpa12g00850 Long chain Acyl-CoA Synthetase 1: LACS1	GO:003677	DNA binding
C18 :0	*q2_70518009*	2	70518009	Vitpa02g441500 Acetyl-CoA Carboxylase alpha-CT subunit: CAC3	GO:0003989	Acetyl-CoA carboxylase activity
	q2_5991246	2	5991246	Vitpa02g04990	IPR029058	Lipid metabolism; fatty acid biosynthesis

				KASIII: 3-ketoacyl-ACP synthase 3		
	q6_3096993	6	3096993	Vitpa06g02800 Acetyl-CoA Carboxylase 1: CAC1	GO:0016874	Ligase activity
	q3_45009201	3	45009201	Vitpa03g23170 Long chain Acyl-CoA Synthetase 2: LACS2	GO:0005524	ATP binding
	q6_48934498	6	48934498	Vitpa06g31150 Biotin carboxyl carrier protein of acetyl-CoA carboxylase 1: BCCP1	PTHR45667	Component of the acetyl coenzyme A carboxylase complex
	q2_15731361	2	15731361	Vitpa02g12030 Acyl-CoA binding Protein 5: ACBP5	GO:0008080	N-acetyltransferase activity
	q6_28898438	6	28898438	Vitpa06g20320 Long-chain Acyl-CoA Synthase 2: LACS2	PTHR43272:SF4	Long-chain fatty acid-CoA ligase activity
	q9_1135448	9	1135448	Vitpa09g00560 Fatty Acid Exporter 2: FAX2	GO:0035338	Long-chain fatty-acyl-CoA biosynthetic process
	q3_43202766	3	43202766	Vitpa03g22110 Long chain Acyl-CoA Synthetase 2: LACS 2	GO:0006636	Unsaturated fatty acid biosynthetic process
	q1_78304163	1	78304163	Vitpa01g39960 Acyl-CoA binding protein 1: ACBP1	GO:0005515	Protein binding
	q9_1241251	9	1241251	Vitpa09g00560 Fatty Acid Exporter 2: FAX2	GO:0035338	Long-chain fatty-acyl-CoA biosynthetic process
C18 :1	*q2_5991246*	2	5991246	Vitpa02g04990 3-ketoacyl-ACP synthase 3: KASIII	IPR029058	Lipid metabolism; fatty acid biosynthesis

	q3_43202766	3	43202766	Vitpa03g22110 Long chain Acyl-CoA Synthetase 2: LACS 2	GO:0006636	Long-chain fatty acid-CoA ligase activity
	q1_78304163	1	78304163	Vitpa01g39960 Acyl-CoA binding protein 1: ACBP1	GO:0005515	Protein binding
	q9_1241251	9	1241251	Vitpa09g00560 Fatty Acid Exporter 2: FAX2	GO:0035338	Long-chain fatty-acyl-CoA biosynthetic process
	q6_3096993	6	3096993	Vitpa06g02800 Acetyl-CoA Carboxylase 1: CAC1	GO:0016874	Ligase activity
	q2_70518009	2	70518009	Vitpa02g44080 CAC3: Acetyl-CoA Carboxylase alpha-CT subunit	GO:0008324	Cation transmembrane transporter activity
	q3_45009201	3	45009201	Vitpa03g23170 Long chain Acyl-CoA Synthetase: LACS	GO:0005524	ATP binding
	q6_48934498	6	48934498	Vitpa06g31150 Biotin carboxyl carrier protein of acetyl-CoA carboxylase 1: BCCP1	PTHR45667	Component of the acetyl coenzyme A carboxylase complex
C18 :2	*q1_20786407*	1	20786407	Vitpa01g16230 Fatty acid desaturases: FADs	GO:0005506	Fatty acid metabolism
	q2_51742656	2	51742656	Vitpa02g30850 Acetyl-CoA Carboxylase: CAC	GO:0004672	Acetyl-CoA carboxylase activity
	q5_ 9317929	5	9317929	Vitpa05g05820 Fatty acid Desaturase 1: FAD1	GO:0016702	Oxidoreductase activity
	q6_43871960	6	43871960	Vitpa06g27010 Long-chain acyl-CoA synthetase 5: LACS 5	GO:0005682	Long-chain fatty acid-CoA ligase activity

q6_10189604	6	10189604	Vitpa06g09060 Fatty acid Desaturase 2: FAD2	GO:0006636	Unsaturated fatty acid biosynthetic process
q7_48678186	7	48678186	Vitpa07g26220 Acyl-CoA binding protein 2: ACBP2	PTHR43840:SF2	Metal ion binding
q10_3420702	10	3420702	Acyl-CoA binding protein 6: ACBP6	GO:0000062	Fatty-acyl-CoA binding

Chr: chromosome, pos: QTN position, G.O: Gene ontology

5.5. Discussion

5.5.1. Phenotypic characteristics of shea fat-related traits

The shea industry prioritizes shortening the juvenile maturity period, increasing oil yield per hectare, and improving oil quality (63).

The analysis revealed significant variation within populations but non-significant variation among populations for fat content. The lack of variation among populations suggests that newly bred varieties could adapt well across in the Côte d'Ivoire shea parklands (122), supported by the high proportion of admixture genotypes observed (**Figure 28b**). Reliable heritability estimates indicate that the selection of marker-associated traits for high-fat yield can significantly enhance genetic progress in shea breeding.

The fat content observed in this study (49.7%) exceeded that reported for *V. paradoxa* subsp. *paradoxa* in Nigeria (45.5%) (129), likely due to the participatory selection of high oil-yielding genotypes (54-58%). Similar findings have been reported in Uganda, although their fat content exceeded ours, reflecting genetic differences between *V. paradoxa* subsp. *paradoxa* in West Africa and *V. paradoxa* subsp. *nilotica* in East Africa (249). Methodological differences in fat extraction are likely to influence the reported variation, highlighting the need for standardization when comparing studies (129,135). Variation in fat content may also be due to the influence of environmental conditions. Authors have reported that environmental cues such as high light intensity increase seed oil content, while high temperature, drought and salinity decrease seed oil content in plant species (250,251).

As expected, the fatty acid composition of shea butter exhibited its characteristic profile, with oleic acid (C18:1) and stearic acid (C18:0) as the dominant components. Oleic acid was more prevalent (50-53%) compared to stearic acid (36.8-40%), a trend consistent with findings from Burkina Faso (135). West and Central African shea butter typically contains higher stearic acid and lower oleic acid levels, while the *nilotica* subspecies in East Africa is characterized by higher oleic acid and lower stearic acid levels (166,252,253). Interestingly, "soft shea butter," high in oleic acid, has also been observed in some *paradoxa* regions, including Côte d'Ivoire (91). As with fat content, environmental conditions influence fatty acid composition (250,251,254). In addition, developmental cues such as gibberellins, auxin and jasmonates can alter seed oil content and modify fatty acid composition (250).

These phenotypic findings are critical for evaluating genetic and environmental interactions influencing stearic and oleic acid traits in Côte d'Ivoire. They also highlight the importance of identifying and using superior shea trees to meet industry demands for quality and consistency in shea butter production.

5.5.2. Detected QTNs by ML-GWAS in Shea Tree

In this study, six ML-GWAS methods were used to analyze five fat-related traits in 122 superior shea tree germplasm. A total of 47 significant QTNs were identified across the methods, with pLARmEB detecting the highest number (29 QTNs), suggesting its relative efficiency (255). Similar findings have highlighted the complementarity of multi-locus GWAS methods in the analysis of complex traits, as

each method captures some distinct QTNs (256) . The identified QTNs were distributed across all 12 chromosomes, underscoring the robustness of ML-GWAS in detecting small-effect loci (257) for fat-related traits in *V. paradoxa*.

Among the QTNs, 25 were commonly detected by at least two methods, with many located in coding regions associated with fatty acid biosynthesis genes. This highlights their importance in the regulation of fat-related trait in shea tree. Nine QTNs were associated with stearic acid (C18:0) and oleic acid (C18:1), demonstrating opposite effect values for these fatty acids. This aligns with the strong negative correlation (-0.98) observed between stearic and oleic acids, suggesting that these fatty acids are controlled, regulated or influenced by the same factors with opposite effects. Similar trends have been reported in other species, including mango, where fatty acid proportions are influenced by genetic and environmental factors (258).

High-fat-content genotypes exhibited a greater proportion of superior alleles, with notable QTNs such as q4_46157457 and q2_13023583 showing strong associations with fat biosynthesis. These findings highlight the genetic complexity of fat content, which is influenced by developmental and environmental cues (250). In addition, superior alleles for C18:0 were more abundant than those for C18:1, consistent with their observed correlations with fat content.

Advances in omics technologies have facilitated the identification of candidate genes involved in lipid biosynthesis in shea tree (25,59). This study identified 24 candidate genes associated with fatty acid biosynthesis pathways. These genes are part of different protein family, including acetyl-CoA carboxylase (ACCase), the key enzyme that catalyzes the first committed step of *de novo* fatty acid synthesis in *V. paradoxa* (25). The genes of the fatty acid complex synthase consisting of Ketoacyl-ACP synthase (*KAS*) that combines acetyl-CoA with malonyl-ACP to produce C16-C18 fatty acids (25,250). It has been shown that a higher number of lipid biosynthesis genes, such as ketoacyl ACP synthase genes in the shea tree might be responsible of the high lipid content in shea fruits (59). Further desaturation and elongation produce longer and more unsaturated fatty acids in the endoplasmic reticulum (25,59,250). This is assured by fatty acid desaturase (*FADs*) genes. Finally, several genes involved in the transmembrane transport of fatty acids from the plastid to the endoplasmic reticulum, including Fatty Acid Export (*FAX*), acyl-CoA binding proteins (*ACBPs*) and Long Chain Acyl-CoA Synthetases (*LACS*) were identified. These genes regulate key steps in fatty acid synthesis, elongation, and desaturation, contributing to the high levels of C18 fatty acids characteristic of shea butter (25,59). Notably, *V. paradoxa* encodes more lipid biosynthesis genes than species like *Arabidopsis thaliana* or *Theobroma cacao*, consistent with its superior fat yield (25).

The identified QTNs and candidate genes provide a foundation for breeding programs aimed at improving shea butter yield and quality. Three strategies can be used to incorporate these findings into breeding programs. First, the identified QTN-allele matrix can be used to predict optimal crosses, such as selecting the top 10 crosses based on their fat content and the proportion of superior allele frequency. Second, SSR markers can be developed near the identified QTNs and applied in marker-assisted selection to enhance crop improvement. Third, the significant SNPs

associated with the traits of interest can be integrated into genomic selection models to improve breeding accuracy and efficiency. However, the identified genes need to be validated through functional analysis to strengthen the biological relevance of the findings.

This research underscores the potential of ML-GWAS in addressing key challenges in shea cultivation, advancing genetic improvement efforts, and supporting the economic and industrial value of *V. paradoxa*.

5.6. Conclusion

In this study, six multi-locus GWAS approaches were used to identify quantitative trait nucleotide (QTN) associated with fat content and fatty acid composition of *V. paradoxa* based on 7559 SNP markers. A total of 47 significant QTNs corresponding to 9, 18, 12, 9 and 8 were associated with fat content, palmitic acid, stearic acid, oleic acid and linoleic acid respectively, with 9 QTNs associated with both stearic and oleic acids. Among these QTNs, 25 were commonly detected by at least two GWAS methods. In total, 24 candidate genes were obtained based on the common QTNs, with 10 previously reported to be involved in the shea tree seed oil and fatty acid biosynthesis and transmembrane transport pathway. Based on 38 SSTs corresponding to 17 SSTs with high fat content and 21 SSTs with low fat content, the proportion of superior alleles of FC common QTNs ranged from 0 to 82.35%. In addition, the proportion of superior alleles within the genotypes with high fat content was higher than that with the genotypes with low fat content. This suggests that these superior alleles exhibit an additive effect on the shea tree seed oil accumulation. These findings suggest that an improvement of the shea tree seed oil yield can be achieved by integrating more superior alleles into shea genotypes by marker-assisted selection (MAS). The 17 high-fat SSTs can be directly propagated by grafting and *in vitro* culture to provide farmers with high-performing plant material or can be included in breeding programs for the development of new cultivars.

5.7. Data availability

The raw sequencing data reported here has been deposited at Genome Sequence Archive of National Genomics Data Center under accession number PRJNA1167878

5.8. Supplementary materials

Table S 9. Summary Statistics and Coefficient of Variation for Fat Content and Fatty Acids in Superior Shea Trees

	Minimum	**Maximum**	**Mean ± sd**	**Coefficient of variation(%)**
Fat content	36.24	58.09	49.73 ± 3.7	7.44
Palmitic acid	2.01	6.12	3.53 ± 0.52	14.73
Stearic acid	22.33	50.54	38.76 ± 4.67	12.05
Oleic acid	40.27	65.66	51.26 ± 4.21	8.21
Linoleic acid	4.78	9.03	6.45 ± 0.76	11.78

Table S 10. Statistical Summary of Fat Content and Fatty Acid Composition of Superior Shea Trees Across Different Regions.

	Region	Min.	Max.	Mean ± sd	CV (%)	Skewness	Kurtosis	F-value	p-value
Fat c.	Hamb.	41.28	54.49	49 ± 3.2	6.57	-0.5	-0.56	0.57	0.57
	Poro	36.24	55.86	49.9 ± 3.9	7.82	-0.95	1.33		
	Tchol.	41.01	58.09	49.9 ± 3.8	7.55	-0.49	-0.28		
C16:0	Hamb.	2.91	4.69	3.7 ± 0.4	11.78	0.66	-0.27	1.06	0.35
	Poro	2.01	4.11	3.5 ± 0.6	17.29	1.35	6.39		
	Tchol.	2.43	4.86	3.6 ±0.5	13.52	0.49	0.31		
C18:0	Hamb.	22.33	43.24	36.9^{a} ± 5.8	15.65	-0.99	-0.02	3.51	0.03*
	Poro	32.25	48.16	39.8^{b} ± 3.8	9.54	-0.18	-0.4		
	Tchol.	26.59	50.54	38.6^{ab} ± 4.9	12.63	-0.09	-0.16		
C18:1	Hamb.	46.4	65.66	52.9^{a} ± 5.3	10.04	0.93	-0.11	3.04	0.05
	Poro	42.13	57.51	50.4^{b}± 3	6.03	0.08	0.09		
	Tchol.	40.27	63.51	51.3^{ab} ± 4.4	8.6	0.15	0.17		
C18:2	Hamb.	4.93	8.59	6.6 ± 0.8	12.02	0.48	0.2	1.77	0.18
	Poro	4.78	7.64	6.3 ± 0.7	10.97	0.19	-0.72		
	Tchol.	5.12	9.03	6.6 ± 0.8	12.37	0.95	1.16		

Min: minimum, Max: maximum, CV: coefficient of variation, Fat c: Fat content, Hambol, Tchol.: Tchologo

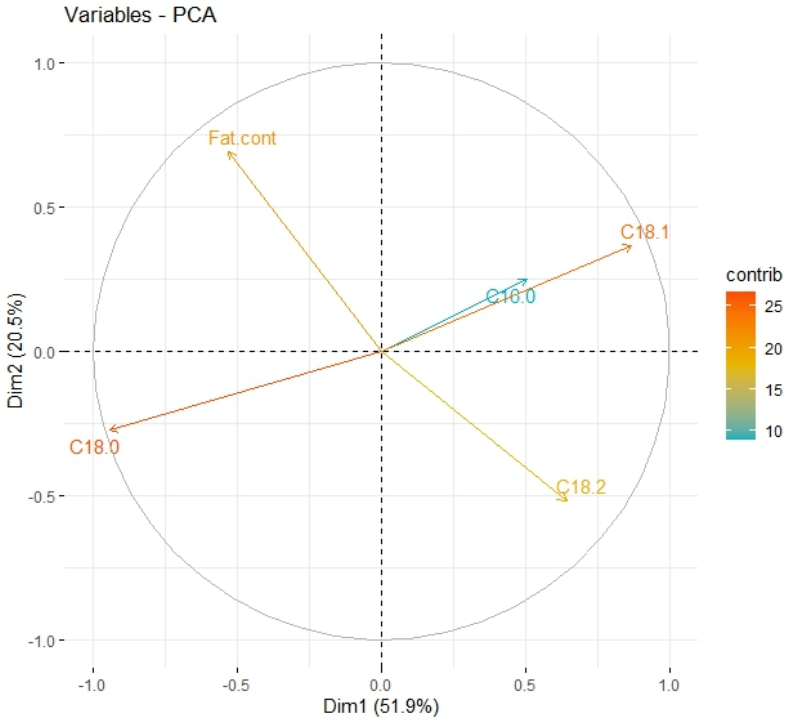


Figure S 8. Loading and score plots of PC1-PC2, "Variables-PCA".

Table S 11. Results of Bayesian model-based clustering implemented in Structure of 122 SSTs

Cluster No	**GG1**	***GG2***	***GG3***	***HE***	F_{ST}	**Number of genotypes**	**Admixed genotypes**
GG1	0			0.2599	0.012	52	19
GG2	0.0016	0		0.2604	0.018	30	10
GG3	0.0018	0.0022	0	0,2597	0,021	40	19
Total	-	-	-	0.26	-	122	48

HE: Expected heterozygosity; F_{ST}: Fixation index; GG1, GG2, GG3: genetic groups 1, 2 and 3 respectively.

Table S 12. Quantitative trait nucleotides (QTNs) associated with fat content and fatty acid content in *V. paradoxa* accessions.

Traits	Marker	Gwas Method	Chr	Pos	QTN effect	LOD Score	-log10	r^2	MAF
C16:0	*q3_57006090*[b]	1, 5	3	57006090	0.2264, 0.1699	36.933, 36.351	44.291, 43.678	97.826, 66.324	0.238
	q7_45956782	1, 2, 4, 5	7	45956782	-0.537, -0.397, -0.376, -0.454	62.718, 42.023, 69.118, 76.237	71.139, 49.637, 77.737, 85.058	336.766, 24.47, 197.545, 289.897	0.092
	q9_42857113[b]	1, 2	9	42857113	-0.193, -0.135	30.231, 34.732	37.198, 41.969	73.341, 47.868	0.27
	q12_1933151[b]	1, 2, 4, 5	12	1933151	0.3395, 0.2786, 0.2322, 0.2222	41.414, 46.872, 42.555, 31.404	48.999, 54.704, 50.194, 38.445	162.921, 14.562, 91.167, 83.967	0.12
	q2_4344942	4	2	4344942	0.1786	34.622	41.853	68.075	0.17
	q2_14738783	4	2	14738783	-0.162	37.044	44.408	57.715	0.2
	q5_55626422	4	5	55626422	-0.163	52.575	60.638	56.337	0.2
	q10_35612747	4	10	35612747	0.1377	39.578	47.073	44.096	0.23
	q10_8515795	5	10	8515795	-0.207	45.518	53.291	86.167	0.17
C18:0	***q2_70518009***[a, b]	1, 2, 4	2	70518009	25.630, 2.206, 19.779	41.425, 41.509, 42.318	49.01, 49.098, 49.945	88.886, 112.583, 86.867	0.11
	q2_5991246[a, b]	1, 2, 4	2	5991246	-27.37, -17.75, -17.61	59.194, 36.946, 41.381	67.497, 44.305, 48.964	145.993, 10.503, 99.159	0.19
	q2_19257675	1	2	19257675	-21.96	33.436	40.599	82.528	0.15
	q3_43202766[a]	1	3	43202766	-27.76	62.683	71.102	163.817	0.21

	q6_3096993[a]	1, 2, 4	6	3096993	25.763, 17.085, 19.205	73.867, 37.316, 51.863	82.623, 44.694, 59.898	130.982, 98.484, 119.445	0.2
	q8_9172814	1	8	9172814	-22.46	37.054	44.419	84.533	0.15
	q9_1135448[b]	1	9	1135448	25.697	38.323	45.754	97.945	0.12
	q3_45009201[a]	2, 4	3	45009201	14.201, 14.498	37.756, 49.525	45.158, 57.467	71.634, 71.662	0.43
	q6_48934498[a]	2, 5	6	48934498	-11.78, -17.6	30.555, 52.002	37.543, 60.042	51.283, 116.913	0.28
	q1_78304163[a]	4	1	78304163	11.464	36.487	43.821	46.408	0.39
	q2_15731361[a]	4, 5	2	15731361	-0.739, -11.79	34.676, 38.989	4.191, 46.454	17.213, 46.655	0.5
	q5_36247653	4	5	36247653	15.118	30.657	37.651	58.664	0.14
	q7_34884350	4	7	34884350	10.663	35.243	42.509	39.662	0.39
	q9_1241251[a, b]	4	9	1241251	-24.05	56.843	65.063	11.968	0.1
	q2_51742656	5	2	51742656	15.183	45.893	53.682	88.938	0.33
	q5_19261711	5	5	19261711	14.429	43.614	51.302	79.295	0.38
	q5_55907613	5	5	55907613	14.107	35.712	43.004	75.486	0.25
	q6_28898438[b]	5	6	28898438	-12.94	39.253	46.731	64.615	0.33
C18:1	*q1_78304163*[a]	1	1	78304163	-15.72	40.431	47.968	76.142	0.39
	q2_5991246[a, b]	1, 2, 4, 5	2	5991246	20.125, 17.516, 16.691, 2.054	38.52, 40.963, 42.476, 66.346	45.961, 48.526, 50.111, 74.881	112.983, 11.87, 105.835, 15.845	0.19
	q3_43202766[a]	1	3	43202766	19.558	36.107	43.421	116.374	0.21
	q6_3096993[a]	1, 2, 3, 4, 5	6	3096993	-31.50, -2.172, -36.43, -19.47, -19.2	95.051, 57.854, 34.979, 49.996, 45.514	104.33, 6.611, 4.223, 57.957, 53.287	280.241, 184.793, 66.249, 145.848, 140.201	0.2
	q9_1241251[a, b]	1	9	1241855	26.178	40.873	48.431	123.613	0.1

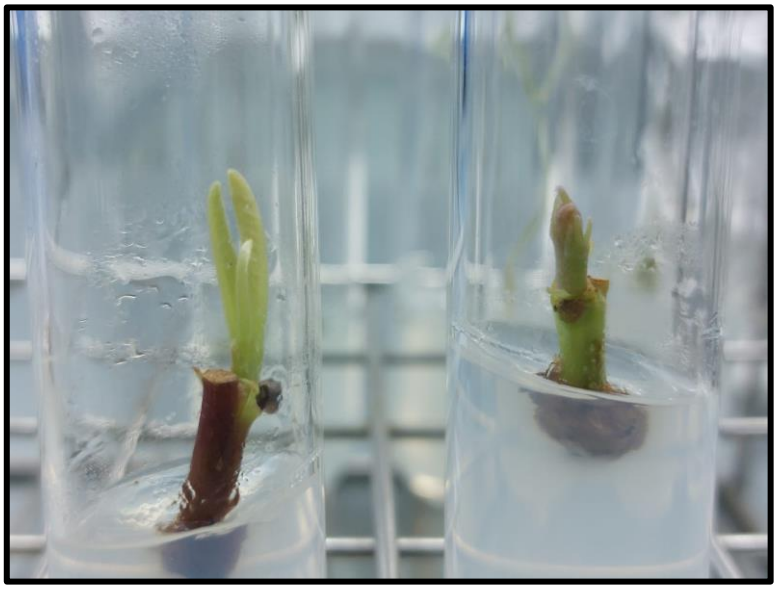
	***q2_70518009*[a, b]**	2, 4, 5	2	70518009	-20.44, -20.77, -16.87	43.335, 48.946, 36.178	5.101, 56.864, 43.495	112.221, 113.747, 74.243	0.11
	***q3_45009201*[a]**	2, 4, 5	3	45009201	-12.42, -14.15, -13.79	35.626, 50.611, 56.447	42.913, 58.597, 64.653	63.643, 84.045, 76.098	0.43
	q4_49530029	4	4	49530029	-0.921	36.584	43.923	36.494	0.32
	***q6_48934498*[a]**	4, 5	6	48934498	10.153, 10.259	31.444, 36.152	38.488, 43.468	43.455, 43.871	0.28
	q2_15731361[a]	5	2	15731361	0.6843	34.004	4.12	17.355	0.5
	q4_7790254	5	4	7790254	-12.24	5.027	58.242	62.963	0.27
	q8_41435892	5	8	41435892	-0.902	32.746	39.868	25.686	0.14
C18:2	*q1_20786407*	1, 2, 4, 5	1	20786407	0.3172, 0.2385, 0.275, 0.2518	42.493, 35.764, 53.124, 53.254	50.129, 43.059, 61.207, 61.342	98.526, 67.616, 96.032, 72.536	0.47
	q2_51742656	1, 2, 3, 5	2	51742656	-0.354, -0.307, -0.457, -0.337	53.915, 58.718, 31.075, 88.003	62.028, 67.004, 38.095, 9.712	142.896, 130.252, 48.076, 151.484	0.33
	q5_ 9317929	1, 2, 5	5	9317929	0.2994, 0.2262, 0.246	39.459, 34.047, 45.983	46.948, 41.246, 53.776	102.34, 70.849, 80.684	0.33
	q6_43871960	1, 2, 5	6	43871960	-0.316, -0.255, -0.218	37.103, 42.238, 31.335	44.47, 49.862, 38.372	109.756, 86.752, 60.655	0.36
	q6_10189604	1, 2, 4, 5	6	10189604	0.429, 0.3224, 0.4396, 0.3415	51.462, 47.971, 85.302, 66.999	59.481, 55.849, 94.355, 75.554	45.464, 31.151, 61.879, 33.654	0.22

	q7_48678186	1, 2, 4, 5	7	48678186	0.4206, 0.3462, 0.3936, 0.3498	40.838, 41.728, 59.801, 45.359	48.395, 49.328, 68.125, 53.125	175.724, 144.471, 199.426, 142.037	0.18
	q7_5074043	4	7	5074043	0.2227	44.244	5.196	72.687	0.31
	q8_2105500	4	8	2105500	-0.215	37.457	44.843	58.519	0.19
	q10_3420702[b]	5	10	3420702	0.1426	38.003	45.417	25.685	0.42
FC	*q4_46157457*	1, 2, 3, 4, 5	4	46157457	26.41, 22.561, 42.502, 19.889, 21.292	47.868, 52.273, 49.296, 48.381, 47.381	55.742, 60.324, 57.229, 56.276, 55.234	289.169, 239.309, 93.099, 174.944, 215.294	0.16
	q11_13733557	1, 2, 3, 4	11	13733557	-20.44, -16.83, -34.31, -14.77	40.872, 36.559, 38.836, 46.511	48.43, 43.897, 46.293, 54.327	204.689, 15.735, 85.125, 114.002	0.24
	q8_11065258	3	8	11065258	-20.33	31.227	38.257	41.936	0.49
	q2_13023583	4, 5	2	13023583	-1.22, -0.967	42.896, 35.059	5.055, 42.315	34.697, 23.412	0.29
	q2_49928619	4, 5	2	49928619	-18.57, -15.52	45.411, 34.502	53.179, 41.726	135.098, 101.353	0.13
	q4_56145911	4	4	56145911	-0.879	36.305	43.629	41.206	0.36
	q9_23061170	4	9	23061170	-12.27	38.858	46.316	58.967	0.14
	q11_14542132	4	11	14542132	10.248	52.239	60.289	56.696	0.32

*GWAS method: (1) mrMLM, (2) FASTmrMLM, (3) FASTmrEMMA, (4) pLARmEB and (5) ISIS EM-BLASSO; Chr: chromosome; Pos: QTN position; MAF: Minor allele frequency; Markers with letter **a** were associated with both stearic acid (C18:0) and oleic acid (C18:1); Markers with letter **b** were detected near a putative coding region that are involved in shea tree lipid biosynthesis; markers in **bolt** were tightly associated with both C18:0 and C18:1.*

Chapter 6

Efficient *in vitro* Regeneration and Genetic Fidelity Analysis of Shea Tree (*Vitellaria paradoxa* Gaertn) Using ISSR Marker



Chapter 6. Efficient *in vitro* Regeneration and Genetic Fidelity Analysis of Shea Tree (*Vitellaria paradoxa* Gaertn) Using ISSR Markers

Shea (*Vitellaria paradoxa*) is a species that reproduces predominantly through cross-pollination, resulting in heterogeneous populations. This heterogeneity often includes a prevalence of genotypes with suboptimal productivity and inferior shea butter quality. To address this challenge, vegetative propagation methods present a promising solution for achieving uniformity. Among these methods, cuttings have demonstrated limited success in shea tree propagation. Grafting, on the other hand, has yielded more promising outcomes but is constrained by several factors, including the influence of rootstock on scion development, the use of non-sanitized plant material, the restricted availability of scions per plant, and the dependence on the species' phenological cycle.

While traditional propagation methods, such as seed germination and vegetative cutting, have shown some promise, particularly with optimized cutting techniques that are easily adoptable by farmers, there remains a critical need for advanced propagation approaches to support large-scale domestication and breeding programs. *In vitro* culture, a technique that offers the potential for producing disease-free plants in large quantities, rapid and efficient multiplication of elite genotypes, has been highly anticipated by the scientific community working on shea tree improvement. Despite its potential, limited progress has been made in establishing reliable *in vitro* protocols for shea, largely due to technical challenges and restricted access to plant material in certain research environments.

This study was conducted in Belgium, where the lack of access to fresh vegetative material for traditional cutting techniques necessitated the exploration of *in vitro* culture methods. To address this, shea seeds were sourced from Côte d'Ivoire, the center of origin for the superior shea trees (SSTs) characterized in earlier chapters. These seeds were germinated and nurtured in a controlled nursery environment within the tropical chamber of the Plant Genetics Laboratory at Gembloux Agro-Bio Tech. This controlled setting provided optimal conditions for the growth and development of seedlings suitable for *in vitro* experimentation.

For the *in vitro* culture process, nodal segments from the shoots of these nursery-grown seedlings were selected as explants. Nodal explants are commonly used in micropropagation due to their high meristematic activity, which promotes shoot proliferation and plant regeneration. The development of this *in vitro* protocol represents a pivotal step toward overcoming the propagation bottlenecks associated with shea tree cultivation. By enabling the rapid multiplication of genetically superior individuals, this approach holds significant promise for supporting both conservation and breeding initiatives, ultimately contributing to the sustainable cultivation of shea across its native range. However, this approach needs more technical tools and knowledge and is expensive. Therefore, it cannot be used by farmers.

Efficient *in vitro* Regeneration and Genetic Fidelity Analysis of Shea Tree (*Vitellaria paradoxa* Gaertn) Using ISSR Markers

Affi Jean Paul ATTIKORA[1*], Souleymane SILUÉ[2], Mongomaké KONE[3], Napkalo SILUE[4], Yves KWIBUKA[5], Saraka Didier Martial YAO[2,6], Caroline DE CLERCK[7], Sok Lay HIM[1], Nafan DIARRASSOUBA[2,6], Taofic ALABI[8], Ludivine LASSOIS[1]

1.University of Liege, Gembloux Agro Bio-Tech, Terra Research Center, Plant Genetics and Rhizosphere Processes lab. Passage des déportés 2, 5030 Gembloux, Belgium.

2.University of Péléforo Gon Coulibaly (UPGC), Faculty of Biological Sciences, Department of Biochemistry-Genetics, Educational and Research Unit of Genetic, BP 1328 Korhogo, Côte d'Ivoire.

3.Université Nangui Abrogoua, UFR des Sciences de la nature, Laboratoire de Biologie et d'Amélioration des Productions Végétales, Abidjan, 02 BP 801 Abidjan 02, Côte d'Ivoire.

4.Université de San Pédro, UFR Agriculture, Ressources Halieutiques et Agro-Industrie, Spécialité Agrophysiologie et Phytopathologie, Biotechnologie Végétale, BP V1800 San Pedro, Côte d'Ivoire.

5.Université Catholique de Bukavu (UCB), Faculté des Sciences Agronomiques, Département de Production Végétale, Bukavu, D.R. Congo.

6. African Center for Shea Research and Application (CRAK), Korhogo, Côte d'Ivoire.

7.University of Liege, Gembloux Agro Bio-Tech, AgricultureIsLife. Passage des déportés 2, 5030 Gembloux, Belgium.

8.University of Liege, Gembloux Agro Bio-Tech, Functional and Evolutive Entomology. Passage des déportés 2, 5030 Gembloux, Belgium.

*Corresponding author: Affi Jean Paul ATTIKORA, E-mail: ajpattikora@uliege.be

6.1. Abstract

Shea tree is an economically valuable tree crop in the food, cosmetic, and pharmaceutical industries due to its seed oil, known as shea butter. Rapid propagation of superior shea trees through *in vitro* culture is essential to support domestication efforts, poverty alleviation, and climate change mitigation. This study aimed to establish an efficient and reliable *in vitro* culture protocol for the regeneration of shea true-to-type plantlets. Nodal explants were incubated on Murashige and Skoog (MS) medium supplemented with 6-benzylaminopurine (BAP) and/or kinetin (Kin) in combination with 1-naphthaleneacetic acid (NAA) for shoot regeneration. Four- to eight-week-old micro-shoots were grown on MS and Modified Murashige and Skoog (MS1B) media at different strengths for rooting. These media were enriched with indole-3-butyric acid (IBA) alone or in combination with NAA, IAA, *meta*-topolin (*m*T), and putrescine. To assess the genetic fidelity of the regenerated plants, five polymorphic inter-simple sequence repeat (ISSR) primers were used.

The results showed that MS/2 medium containing 3:1.2:1 mg/L BAP:Kin:NAA gave the best regeneration of axillary shoots. Four-week-old axillary shoots were 100% rooted on MS1B/2 medium containing 3:0.1:40 mg/mL IBA:*m*T:putrescine. Rooted plantlets were successfully acclimated in vivo. The polymorphism of the ISSR Primers ranged from 50 to 87.5%, with an average of 65%, and the polymorphism information content was 0.22. For genetic fidelity assessment, 34 scorable and reproducible markers were obtained. All markers were monomorphic and identical to the mother plant. The micropropagation protocol proposed in this study is suitable for large-scale *in vitro* regeneration of shea without genetic alteration. However, further studies are needed for the induction of multiple micros-hoots.

Keywords: Axillary shoot induction, Genetic homogeneity, ISSR markers, Micropropagation, Root induction, *Vitellaria paradoxa*

6.2. Introduction

Shea tree (*Vitellaria paradoxa*) grows across 21 countries that forms a large belt in the semi-arid zone of sub-Saharan Africa (67). In Côte d'Ivoire, shea trees are located in the northern part. It is an important economic plant species due to its multipurpose use (90). The shea butter produced from its kernel is relevant in the food, cosmetic and pharmaceutical industries (67). Income from shea has been shown to make a significant contribution to rural livelihoods throughout the Sahel-Savannah (170). Shea tree is recognized as the second largest oil-producing plant after oil palm in Africa (25). The global shea market is estimated at $30 billion in 2020 (63). This high demand is due to its use in the confectionery and cosmetics industries. In the European market, the demand for these natural and organic cosmetics reached a value of EUR 3.90 billion in 2019 (136). Cosmetics alone exceeded $530 million in 2020 and is expected to reach $1025 million by 2027 (63).

Shea trees are the major constituents of the existing plant biomass in their growing area. In that way, they not only form a substantial carbon reserve with a huge potential for carbon sequestration (necessary for climate change mitigation), but also contribute to protect against the environmental degradation prevalent in their growing area (79). However, the over-exploitation of its timber for firewood and charcoal production (259) has led to the collapse of the tree's population at the extent that currently, it is indexed as a vulnerable plant species by the International Union for Conservation of Nature (26).

Indeed, the current population of shea trees are the result of natural regeneration, protection, and wild management during cultivation cycles (205). Due to cross-pollination, these wild-managed stands resulting in heterogeneous populations. These heterogeneity often includes a prevalence of genotypes with poor genetic background and therefore produce an unreliable crop in terms of shea nut quantity and shea butter quality (41).

The necessity to produce superior true-to-type varieties that conserve the desired genetic traits (51), has prompted research into the clonal propagation strategy (260). However, some vegetative propagation methods, such as the use of stem cuttings and the air-layering, were reported to have limited success (34,36,37,261,262). However, studies have reported a successful grafting of shea tree plantlets on fields (33) and on one-year-old plantlets in nursery (32). Grafting was recommended in a recent study to clone the individuals within the Design Core Collection of Ivorian shea trees for *ex situ* conservation (122). Unfortunately, the grafting method can only be used on already established plantlets. Therefore, there is a need to develop a reliable asexual propagation technique that can be used to both produce superior true-to-type varieties and conserve shea tree desired traits to aid in the domestication of the species.

On the other hand, the in-vitro culture is an alternative approach to improve the propagation and enhance a large-scale cultivation of the shea tree (44). Explant culture methods have shown success in propagating various recalcitrant tree crops and woody plants (263) including various members of the Sapotaceae, such as *Argania spinosa* (264), *Madhuca latifolia* (265), *Pouteria lucuma* (266) and *Achras sapota* (267). A

variant of the in-vitro approach, the micropropagation of woody species, has become a widely used technique for rapidly regenerating elite germplasm and reducing the juvenile post-propagation period for conservation purposes (268).

Previous shea micropropagation attempts have used nodal and shoot tip explants (41,43,44), young leaf explants (38,42,269), and immature cotyledon and pulp techniques (39,40). Most of the studies using nodal and shoot tip explants reported low rates of axillary shoot induction and did not attempt root induction (43,44). However, in the study conducted by Lovett and Haq, the authors reported very low rates (8%) of rooted plantlets, but did not mention acclimatization, suggesting that they failed at this important stage (41). Also, none of the shea micropropagation studies mentioned multiple shoot regeneration. With other explants mentioned above, the authors reported callus formation but failed with shoot regeneration. All authors of these previous studies recommended further research to standardize micropropagation techniques for *Vitellaria paradoxa*. The main causes of these failures or low rates of micro-shoot induction were associated with the combination and/or concentration of growth regulators used.

It is essential to determine the genetic uniformity of *in vitro* plants, as the goal of micropropagated plants is to produce a uniform genetic material for commercial use (270). *In vitro* plantlets often have genetic variants due to exposure to artificial conditions (271–273). An effective method for studying somaclonal changes in *in vitro* plantlets is the use of molecular markers (271,273,274). The inter-simple sequence repeat (ISSR) markers are the most commonly used PCR-based molecular markers in genetic stability studies (273,275,276). ISSR markers are dominant, reproducible, and capable of amplifying genome-wide regions without prior sequence information (277). These characteristics make ISSR markers valuable tools for various applications in plant genetics, such as genetic diversity analysis and marker-assisted genetic fidelity (273,275,278).

The aim of this study was to establish an efficient *in vitro* culture protocol for the regeneration of true-to-type shea plantlets.

6.3. Materials and Methods

6.3.1.1. Preparation of explants and surface sterilization

Shea seeds were sourced from Côte d'Ivoire and were germinated and nurtured in a controlled nursery environment within the tropical green-house of the Plant Genetics Laboratory at Gembloux Agro-Bio Tech.

Shoots (stems and buds) from a well-managed stock of *V. paradoxa* seedlings maintained in the tropical green-house, were used for the preparation of explants (**Figure 31a**). Excised shoots (**Figure 31b**) were either pink or green in color. The collected shoots were rinsed under running tap water for 5 min to remove any surface dirt. The explants were prepared by carefully removing older leaves, leaving 1 to 2 new leaves per shoot (**Figure 31c**). They were agitated in a disinfectant solution of 3% liquid soap for 10 min and left under running tap water for 10 min. The shoots were then dipped in a 2% (w/v) fungicide solution (Benlate WP) for 5 min and left under running tap water for 5 min. The shoots were then transferred to a sterile

environment under a horizontal laminar air flow cabinet (Clean Air Technology; Filtest, Benelux, Belgium) for aseptic sterilization.

The shoots were surface sterilized with 70% (v/v) ethanol for 2 min and further sterilized with 10% (v/v) sodium hypochlorite solution (14% Cl_2; VWR, Belgium) for 10 min. After sterilization, the shoots were rinsed five times with sterile distilled water. The shoots were then placed on sterilized filter paper in a Petri dish to dry, and the remaining leaves and petioles were removed (**Figure 31d**). The node was trimmed to a size between 1 and 2 cm and used as explants for culture initiation (**Figure 31e**). Explants were randomly aseptically placed vertically on shoot induction media (**Figure 31f**). Cultures were maintained in the growth chamber at 26°C ± 1 under a 16/8 h light/dark photoperiod.

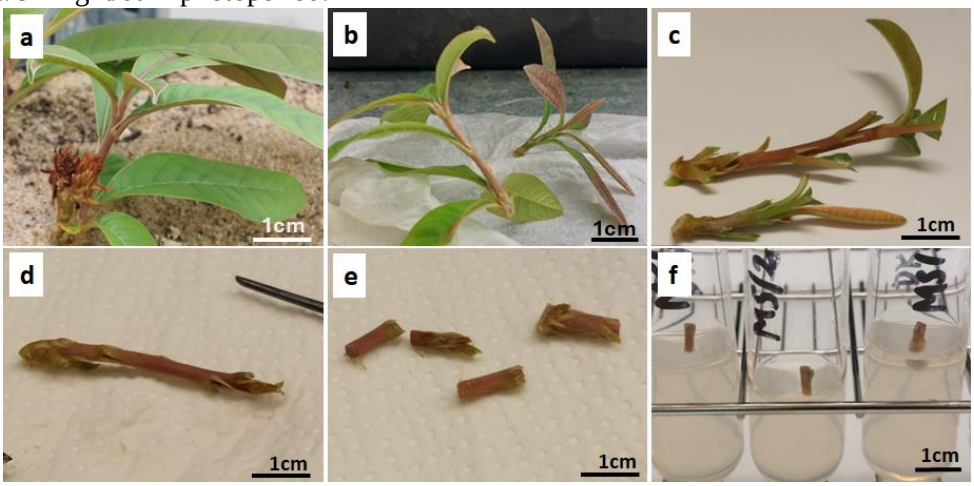


Figure 31. Preparation of explants for culture initiation. (a) location of a shoot on the donor tree, (b) Excised shoots; (c) shoot after removal of old leaves, (d) shoot after remaining petiole and leaf, (e) nodal explants, (f) explants on shooting medium.

6.3.1.2. Reference Standard protocols for Media Preparation

Standard procedures were used to prepare all media. Murashige and Skoog medium micro and macronutrients including vitamins and modified Murashige and Skoog medium micro and ½ macronutrients including vitamins were used in this study, as defined by Murashige and Skoog (279). The different base media, their composition and the strengths used in this study are shown in **Table 18**. An identification of the different strengths of each medium has also been defined in **Table 18** and will be used throughout this paper. Media were supplemented with 30 g sucrose, pH was adjusted to 5.8, gelled by 8 g/L Agar. All media and all instruments (scalpels, forceps, scissors, and paperwork) were sterilized by autoclaving at 121 °C and 15 psi for 15 min.

Table 18 Basal media composition and strength length used during in this study

Media	Composition (mg/L)			Strength	Code
	Macronutrients	Micronutrients	Vitamins		
Murashige and Skoog medium	$CaCl_2$:332.02 KH_2PO_4 :170 KNO_3:1900 $MgSO_4$:180.54 NH_4NO_3: 1650	CoCl2.6H2O: 0.025 $CuSO_4.5H_2O$: 0.025 FeNaEDTA: 36.7 H_3BO_3: 6.2 KI: 0.83 $MnSO_4.H_2O$: 16.9 $Na_2MoO_4.2H_2O$: 0.25 $ZnSO_4.7H_2O$: 8.6	Glycine : 2 myo-Inositol: 100 Nicotinic acid:0.5 Pyridoxine HCl:0.5 Thiamine HCl: 0.1	Full	MS
				Half	MS/2
				Quarter	MS/4
Murashige and Skoog medium mod. No. 1B	$CaCl_2$: 166 KH_2PO_4 : 85 KNO_3: 950 $MgSO_4$: 87.86 NH_4NO_3: 825	Idem	Idem	Full	MS1B
				Half	MS1B/ 2

Full (100%, w/v), Half (50%, w/v), Quarter (25%,w/v)

For the axillary shoot induction, only MS/2 was considered because it showed more promising results in shea axillary shoot induction than MS in previous studies (41,43,44). The MS/2 medium was supplemented with three plant growth regulators (PGRs) BAP, Kin and NAA. Three regimes of these plant growth regulators were applied: (1) use of cytokinin alone (BAP or Kin), (2) combination of cytokinin (BAP) and auxin (NAA) and (3) combination of two cytokinins (BAP and Kin) and one Auxin (NAA) (**Table 19**). For the first regime, the concentrations of BAP or Kin was: 0, 0.5, 1, 2.5 and 3.5 mg/L. The second regime concentrations of BAP:NAA are: 3.5:0.1 and 3:1 mg/L. Finally, the combination of BAP:NAA:Kin concentrations were: 3:1:1 and 3:1.2:1 mg/L. The selection of concentrations and combinations of growth regulators was done by consulting previous studies on Sapotaceae family or other recalcitrant plants *in vitro* culture (41,43,44,264–267).

Regenerated shoots from lateral bud breaks were transferred to rooting media consisting of MS, MS/2, MS1B and MS1B/2 with different PGRs. The different PGR combinations and concentrations are listed in **Table 19**. The plant growth regulators, indole-3-butyric acid (IBA), *meta*-topolin (*m*T), 1-Napthaleneacetic acid (NAA), Indole-3-acetic acid (IAA), and Putrescine were added to stimulate adventitious root induction. The cultures were maintained in a growth room under the same conditions as described above. The different concentrations of the combination (IBA:*m*T:Putrescine) in MS1B medium were: 3:0:0; 3:0.1:40 and 3:0.1:75 mg/L. Concentrations of plant growth regulators in MS medium were: 3 mg/L (IBA); 1.5:0.5:160 mg/L (IBA:NAA:Putrescine) or 1.5:0.5:1 mg/L (IBA:NAA:IAA). The choice of concentrations and combinations was the same as mentioned above.

For the rooting experiment, axillary shoots of different ages were used: young axillary shoots (4-6 weeks) and old axillary shoots (8 weeks and older).

Table 19. Plant growth regulators combinations and concentration used for shoot and root induction

PGR Comb.	C1(mg/L)	C2(mg/L)	C3(mg/L)	C4(mg/L)	C5(mg/L)	Med
Shoot induction						
BAP	0	0.5	1	2.5	3.5	MS/2
Kin		0.5	1	2.5	3.5	MS/2
BAP:NAA	3.5:0.1	3:1				MS/2
BAP:Kin:NAA	3:1:1	3:1.2:1				MS/2
Root induction						
IBA	0	3			MS/2, MS1B,	MS/4, MS1B/2
IBA:NAA:Putrescine	1.5:0.5:160				MS/2	
IBA:NAA:IAA	1.5:0.5:1				MS/2	
IBA:*m*T:Putrescine	3:0.1:40	3:0.1:75*			MS1B,	MS1B/2

PGR Comb.: plant growth regulators combinations; C1-C5: different concentrations; Med: Medium; concentration with asterisk was only used with MS1B.

6.3.1.3. Acclimatization

Rooted plantlets were removed from the rooting medium and the roots were carefully washed with distilled water to remove adherent agar. Plantlets were transferred to plastic pots (one per pot) containing a mixed substrate of autoclaved universal potting soil consisting of peat moss, fertilizer, lime, and perlite. They were then transferred to the greenhouse under the following conditions: 28±2 °C temperature, 85±5 % relative humidity and 16/8h light/dark photoperiod. The pots were covered with clear polyethylene plastic to maintain high humidity.

6.3.2. Genetic fidelity assessment using ISSR markers

6.3.2.1. DNA extraction and quantification

To assess the genetic fidelity of regenerated plantlets of *V. paradoxa* plants to the donor (plant from which explants are taken), fresh leaf samples of 100 mg were randomly collected from 5 acclimated plantlets and the donor plant. Leaf tissue was finely ground in liquid nitrogen using a pre-chilled mortar and pestle. Genomic DNA was isolated from the resultant tissue powder using the "Innovative Shea DNA extraction protocol" (207). The DNA pellet was resuspended in sterile double-distilled water and stored at -20 °C until further analysis. The concentration of the isolated DNA was determined using a Quantus Fluorometer (Promega, Madison, USA) in a total volume of 100 µl, containing 98.5 µl (1XTE Buffer), 0.5 µl (QuantiFluor dsDNA System; Madison, Promega) and 1 µl (DNA template).

6.3.2.2. ISSR-PCR analysis

Five (5) ISSR primers were used to assess the genetic fidelity of regenerated plantlets. ISSR markers have not been yet mentioned in *V. paradoxa* study. Therefore, the polymorphic level of the 5 ISSR primers was first assessed using 10 shea trees randomly selected from the shea seedling established in the tropical green-house as mentioned in the "Preparation of Explants and Surface Sterilization" section. The

seeds used to establish the seedlings were collected in Côte d'Ivoire in the regions of Poro and Bagoué. The five ISSR markers used were polymorphic and produced highly readable and reproducible bands. It should be noted that the effect of genotype on propagation efficiency was not considered in this study. However, one shea tree was identified as a donor prior to shoot induction to study the genetic fidelity of the regenerated shea trees from this donor using our method. The five ISSR primers were then used to assess the genetic fidelity of five regenerated shea trees. The sequence of the primers used and their characteristics are shown in **.**

Table 20. Polymerase chain amplification (PCR) was performed in a total volume of 20 µL containing: 10 µL of Go Taq Green Master Mix (2x), 1 µL of the primer (10 pmol/µL), 1 µL of DNA template (15 ng/µL) and 8 µL of sterile ddH_2O. The PCR run program was performed on a Bio-Rad T100 Thermal Cycler the PCR machine, with a denaturation step at 94 °C for 4 min, followed by 30 cycles of amplification at 94 °C for 30 s, 40 s at the primer's annealing temperature (**.**

Table 20), and a 72 °C for 2 min and a final extension at 72 °C for 5 min.

Table 20. Sequence and characteristics of ISSR primers

Locus code	Primer sequence	Tm (°C)
IS-1	5' AGA GAG AGA GAG AGAGG 3'	52.8
IS-2	5' GAG AGA GAG AGA GAGAT 3'	50.4
IS-3	5' GAG AGA GAG AGA GAGAC 3'	52.8
IS-4	5' TCT CTC TCT CTC TCTCC 3'	52.8
IS-5	5' AGA GAG AGA GAG AGA GYT 3'	52.6

Tm (°C): annealing temperature

ISSR amplification was performed three times and only reproducible PCR products were scored. Amplified fragments were separated on a 1% (w/v) agarose gel using 1× TAE buffer and stained with 0.1 µL/mL Gelred. Electrophoretic separation was performed at 180 volts for 25 minutes. Bands were visualized using a gel documentation system (Gel Doc **XR+** system; Bio-Rad model; Universal Hood II). The size of each amplicon was estimated by comparison with the Fast DNA Ladder (BioLabs, England). The banding patterns were compared to evaluate the genetic variability between samples. The fragments were scored as either present (1) or absent (0). The polymorphism information content (PIC) was determined following the general equation established by Serrote et al. for dominant markers (277):

$$PIC = 1 - (p^2 + q^2)$$

Where *p* is the frequency of present bands and *q*, the frequency of absent bands.

6.3.3. Collection of experimental data and statistical analysis

Cultures were observed weekly for signs of growth. For shoot induction, the data recorded were time for shoot induction (week), callus formation, percentage of explants that induced shoot, number of shoots per explant, percentage of elongated shoot, average shoot length (cm) and number of leaves. The experiment was repeated three times with 130 explants in each replicate (10 explants per treatment).

For root induction, the following data were recorded: callus formation, time for root induction (week), number of roots per shoot, root length and percentage of healthy shoots in the rooting media. Five new axillary shoots were used for each treatment with two replicates.

The collected data were analyzed using the R version 4.3.3. Analysis of variance (ANOVA) was used to determine the differences between treatment means. Multiple comparisons between means were performed using Tukey's pairwise comparison.

6.4. Results

Like many plant species of the Sapotaceae family, *Vitellaria paradoxa* is recalcitrant in tissue culture. In this section, we present the main findings of an efficient protocol for *in vitro* propagation of shea tree for axillary shoot proliferation, root induction and ex-vitro acclimatization. We further assess the genetic fidelity of the regenerated plants using ISSR markers.

6.4.1.1. Axillary shoot regeneration

Axillary shoots were induced when explants were cultured on MS/2 media supplemented with various combinations of the plant growth regulators BAP, Kin and NAA.

Axillary shoots were induced after 1 week on explants grown on media supplemented with Kin alone (1mg/L) and the combination BAP:Kin:NAA (3:1.2:1 mg/L). Media containing BAP alone and the combination of BAP:NAA (3:1 mg/L) induced axillary buds after 2 weeks **Table 21**). However, axillary shoots induced on media containing BAP alone showed no signs of elongation. Induced shoots were elongated on media supplemented with Kin alone, the combinations BAP:NAA, and BAP:Kin:NAA. The highest percentage of elongated axillary buds was obtained on MS/2 medium supplemented with the combination of 3:1.2:1 mg/L BAP:Kin:NAA (**Table 21**).

Media containing the hormone concentration of 3:1.2:1 mg/L BAP:Kin:NAA had the highest average length of axillary shoot per explant (5.2 cm). In contrast, media supplemented with Kin alone (1 mg/L) had the lowest average shoot length in the elongated axillary buds. A highly significant difference ($p<0.001$) in mean length was observed between treatments (**Table 21**). After 4 weeks of culture on medium containing 3:1.2:1 mg/L BAP:Kin:NAA, **Figure 32a** and **Figure 32c** show the elongation of the microshoots, while **Figure 32e** shows the development of the microshoot on medium containing 1 mg/L Kin.

Table 21. Effect of BAP, Kin and NAA concentrations on growth, callus formation and axillary shoot induction of *V. paradoxa* explants after 8 weeks culture on half-strength MS media.

Med.	PGRs (mg/L)			TSI	CFo	EIS (%)	PES (%)	NSE	ASL	NoL
	BAP	Kin	NAA							
MS/2	0	0	0	*NA*	N	0	0	0	0	0
	0.5	0	0	*NA*	N	0	0	0	0	0
	1	0	0	*NA*	N	0	0	0	0	0
	2.5	0	0	*NA*	N	0	0	0	0	0
	3.5	0	0	*NA*	N	0	0	0	0	0
	0	0.5	0	*NA*	N	0	0	0	0	0
	0	1	0	1	N	100	53.3	1	1.3 ± 0.2^{a}	0.84 ± 0.1^{a}
	0	2.5	0	*NA*	N	0	0	0	0	0
	0	3.5	0	*NA*	N	0	0	0	0	0
	3.5	0	0.1	2	L	46.67	13.3	1	1.5 ± 0.3^{a}	1.4 ± 0.2^{a}
	3	0	1	2	L	53.33	26.7	1	2.2±0.4a	1.7 ± 0.26^{a}
	3	1	1	1	M	100	86.7	1	3.5 ± 0.5^{b}	4.6 ± 0.55^{b}
	3.3	1.2	1	1	M	100	93.3	1	5.2 ± 0.36^{c}	6.7 ± 0.58^{c}

Med.: medium; PGRs: plant growth regulators, TSI: Time for shoot induction (NA: None), CFo: callus formation (N: None, L: Low, M: Moderate), EIS: percentage of explants that induced a shoot, PES: percentage of elongated shoots, NSE: number of induced shoots per explant, ASL: average shoot length, NoL: number of leaves

Globally, the first induced leaves were observed on the shoot tip explants after 2 weeks, while leaves were observed on the nodal explants after 4 weeks in culture (**Figure 32a** and **Figure 32c**). However, the number of leaves per explant was higher on nodal explants than on shoot tip explants after 6 weeks of culture (**Figure 32b and Figure 32d**). The number of leaves per explant was higher in the BAP:Kin:NAA treatment with an average of 6.7 leaves per explant after 8 weeks (**Table 21**), while the lowest average number of leaves (0.84 leaves/explant) was obtained from media supplemented with Kin alone. The number of leaves from the BAP:Kin:NAA combination was significantly different ($p < 0.001$) from the number of leaves from the BAP:NAA combination and from the treatment with Kin alone. Explants cultured on media containing BAP:Kin:NAA formed moderate callus at the basal side (**Figure 32b**), while media containing only Kin did not form callus (**Figure 32f**).

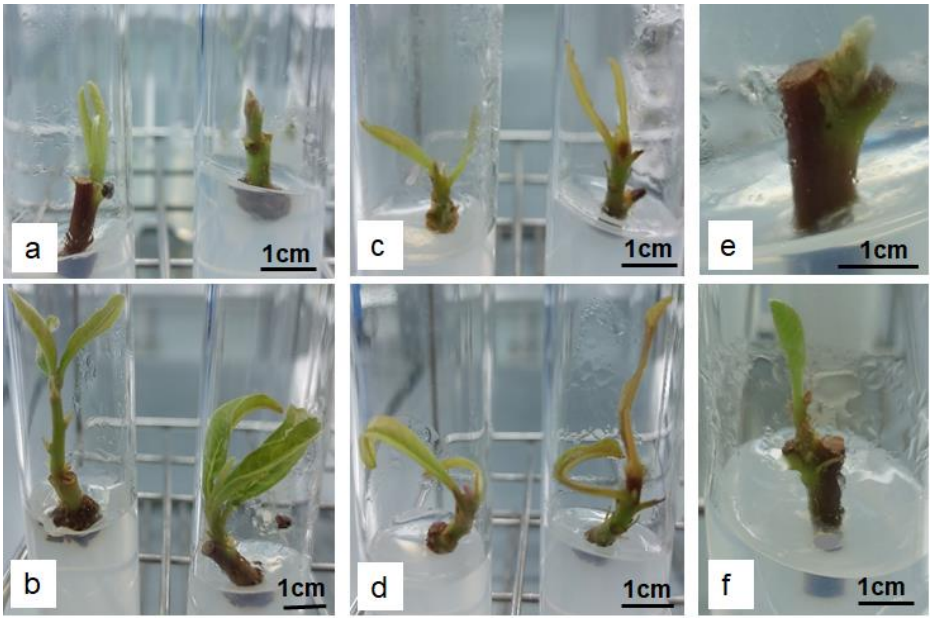


Figure 32. Growth and development of *V. paradoxa* microshoots on MS/2 medium supplemented with: *(1)* 3:1.2:1 mg/L BAP:Kin:NAA*:* **a** and ***c*** *(*4 weeks*), and* ***b*** *and* ***d*** *(6 weeks) on culture initiation media respectively*; *and (2)* 1mg/L Kin ***e*** and **f:** 4 and 6 weeks respectively *on initiation media*.

6.4.1.2. Root induction

Adventitious roots were induced on regenerated axillary buds of *Vitellaria paradoxa* after 12 weeks on rooting media consisting of MS1B/2 supplemented with a combination of IBA, *m*T, and putrescine (3:0.1:40 mg/L). All 4- or 5-week-old axillary buds were rooted and formed 1 or 2 roots on the MS1B/2 medium supplemented with 3 mg/L IBA, 0.1 mg/L *m*T, and 40 mg/L putrescine (**Figure 33a**). It should be noted that the color of the medium changes in rooted tubes due to callus degradation (**Figure 33**b and **Figure 33**c). The average root length measured was 2.2 cm (**Table 22**). However, no roots were observed on axillary buds of 8 weeks and older from the same treatment (**Figure 33d**). No evidence of rooting was observed when shoots were cultured on media containing MS1B (**Figure 33e**) and MS/2 media or MS/4.

Table 22. Effect of culture media, putrescine, indole-3-butyric acid (IBA), indole-3-acetic acid (IAA) and 1-naphthalene acetic acid (NAA) on rooting percentage, root length, callus formation for shea tree (*Vitellaria paradoxa* C.F Gaertn) after 12 weeks of culture.

Medium	**Plant growth regulators (mg/L)**					**CFo**	**Root induction**		**RAL (cm)**	**HAB**
	IBA	***m*T**	**IAA**	**NAA**	**Putrescine**		**4-6 ws**	**≥8 ws**		
MS1B	0	0	0	0	0	N	-	-	-	100
	3	0	0	0	0	L	-	-	-	100
	3	0.1	0	0	40	M	-	-	-	100
	3	0.1	0	0	75	N	-	-	-	100
MS1B/2	0	0	0	0	0	N	-	-	-	100
	3	0	0	0	0	L	-	-	-	100
	3	0.1	0	0	40	M	R+	-	2.2±0.3	100
MS/2	0	0	0	0	0	N	-	-	-	100
	3	0	0	0	0	L	-	-	-	100
	1.5	0	1	0.5	0	N	-	-	-	100
	1.5	0	0	0.5	160	N	-	-	-	100
MS/4	0	0	0	0	0	N	-	-	-	100
	3	0	0	0	0	L	-	-	-	100
	1.5	0	1	0.5	0	N	-	-	-	100
	1.5	0	0	0.5	160	N	-	-	-	100

CFo: Callus formation (N: None, L: Low; M: Moderate) ; 4-6 ws: 4 to 6 weeks axillary shoot; ≥ 8 ws: 8 weeks or older axillary buds; RAL: Root average length; HAB: healthy axillary bud; R+: root induction, (-): None.

These results show that *Vitellaria paradoxa* can be successfully rooted , although the axillary shoots must be very young. All microshoots that did not exhibit roots were still healthy at the end of the experiment (**Figure 33f**).

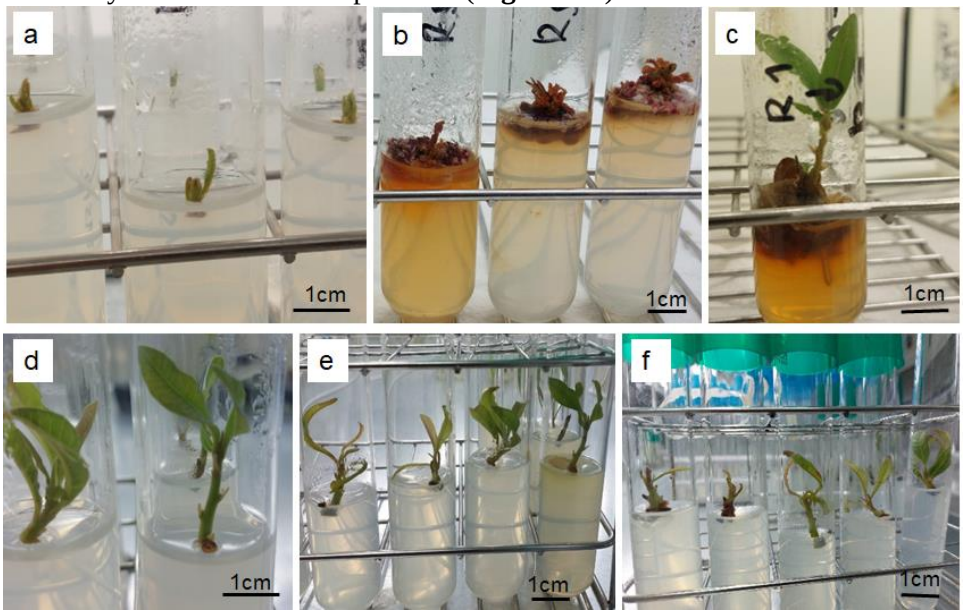


Figure 33. Microshoots onto rooting media MS1B/2 containing 3:0.1:40 mg/mL IBA:mT:putrescine. a: *Four* weeks *(young)* axillary buds on rooting media; b: young axillary buds after 6 weeks; c: in vitro rooted plantlet after 12 weeks, d: 8 weeks axillary buds (old) on rooting media; e: old axillary buds after a week; f: old axillary buds after a 12 weeks

6.4.1.3. *Ex vitro* acclimatization

The rooted plantlets transferred to a substrate containing universal compost and perlite acclimated successfully. The survival rate of 100% was observed on both plantlets with one or two roots. Some plantlets showed signs of growth after 1 week and exhibited more green leaves compare to the initial color (see **Figure 34b** and **Figure 34d**). The plantlets continued to grow, producing new leaves and elongating shoots.

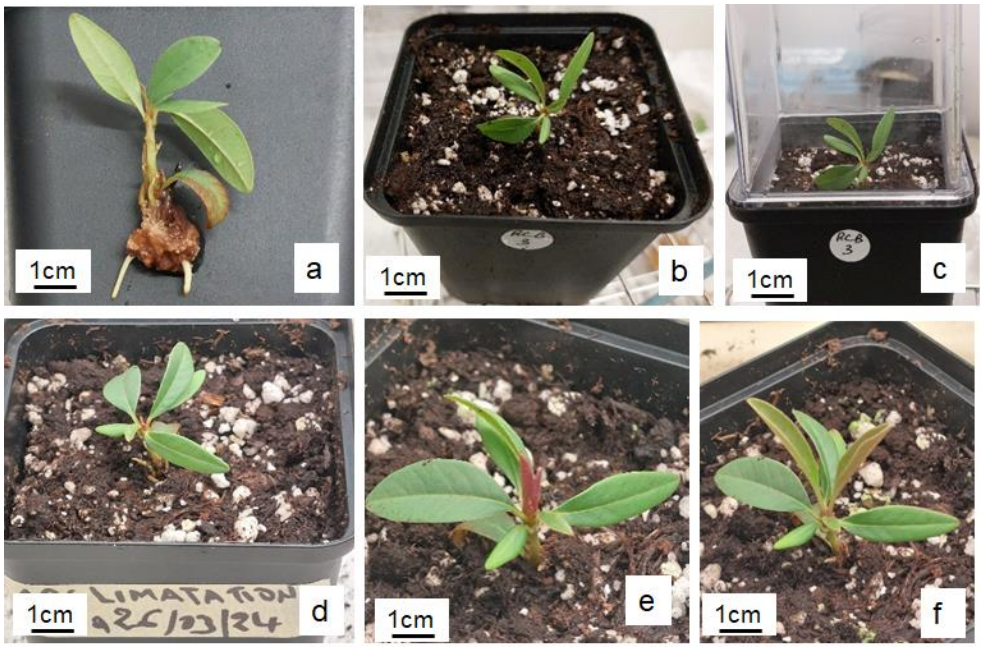


Figure 34. *Ex vitro* acclimatization in universal compost-perlite mixture substrate. a: washed plantlet to remove adherent agar; b: plantlet transferred to substrate; c: plantlet in pot covered to keep moisture; d: plantlet after one week in substrate; e: plantlet after 2 weeks in substrate; f: plantlet after 3 weeks in substrate.

6.4.2. Genetic Fidelity Studies of Acclimated Plantlets of *V. paradoxa*

6.4.2.1. Polymorphism of the 5 ISSR primers in *V. paradoxa*

To access the genetic fidelity of the regenerated shea trees, 5 ISSR primers were used. Due to the lack of knowledge of the level of polymorphism of these five markers in *V. paradoxa*, 10 shea trees were used to evaluate their variation. The total number of bands produced was 34, consisting of 23 polymorphic and 11 monomorphic bands (**Table 23**). The average polymorphic percentage was 65%. The highest polymorphic percentage was found in IS-4 (87.5%), while the lowest polymorphic percentage was found in IS-2 and IS-5 (50%). The polymorphism information content (PIC) ranged from 0.17 (IS-5) to 0.32 (IS-3) with an average of 0.22 (**Table 23**). The ISSR markers IS-1, IS-3 and IS-4 produced the same number of bands, which was 8, while IS-5 produced the lowest number of bands, which consisted of 4.

Table 23. Polymorphism assessment of the 5 ISSR primers within 10 shea tree genotypes.

Primers	Sequence	Tm (°C)	TB	PB	MB	%P	%M	PIC
IS-1	$(AG)_8G$	52.8	8	5	3	62.5	37.5	0.18
IS-2	$(GA)_8T$	50.4	6	3	3	50	50	0.14
IS-3	$(GA)_8C$	52.8	8	6	2	75	25	0.32
IS-4	$(TC)_8C$	52.8	8	7	1	87.5	12.5	0.3
IS-5	$(AG)_8YT$	52.6	4	2	2	50	50	0.17
Total/Mean			34	23	11	65	35	0.22

Tm (°C): annealing temperature, TB: total bands; PB: polymorphic bands, MB: monomorphic bands; %P: percentage of polymorphic bands; %M: percentage of monomorphic bands; PIC: polymorphism information content.

The **Figure 35** showed the polymorphism of the primer IS-4 within the 10 shea trees.

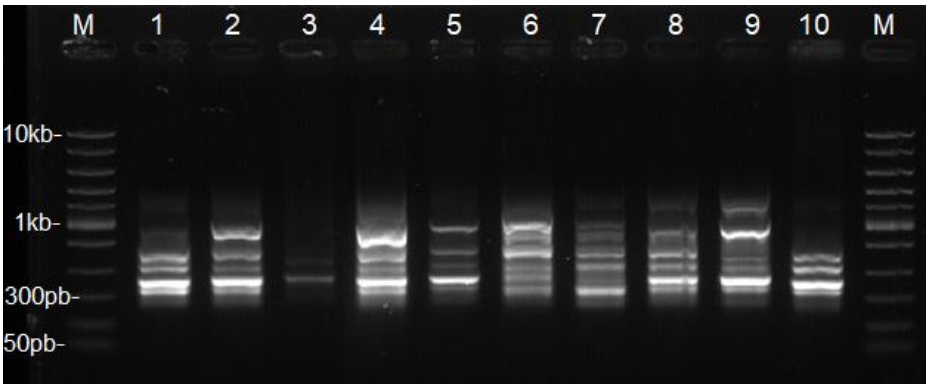


Figure 35. Polymorphic band patterns of the ISSR primer IS-4 within the shea trees; M: Fast DNA Ladder; numbers represent the genotypes

6.4.2.2. Genetic Fidelity Analysis

The five ISSR primers were then used to assess the genetic fidelity of five regenerated shea microshoots, since their polymorphic nature had been established. The screening process produced 34 scorable and reproducible bands, ranging from 200 to 2000 bp (**Table 24**). All ISSR primers successfully produced DNA bands with an average of 7 bands per primer.

Table 24. Five ISSR primers used in genetic fidelity assessment of *Vitellaria paradoxa*

Locus code	Primer sequence	Tm(°C)	No of amplified bands	Amplificated bands range (bp)
IS-1	$(AG)_8G$	52.8	9	200-1000
IS-2	(GA)8T	50.4	5	300-900
IS-3	(GA)8C	52.8	6	200-1200
IS-4	$(TC)_8C$	52.8	8	300-2000
IS-5	(AG)8YT	52.6	6	300-1000

The primer IS-1 generated the highest number of bands, with nine bands ranging from 200 to 1000 bp. In contrast, IS-2 produced the lowest number of bands, with 5 bands ranging from 300 to 900 bp. The banding patterns observed were monomorphic (**Figure 36a**). The **Figure 36b** and **Figure 36c** show the banding patterns between regenerated shea trees and the donor plant.

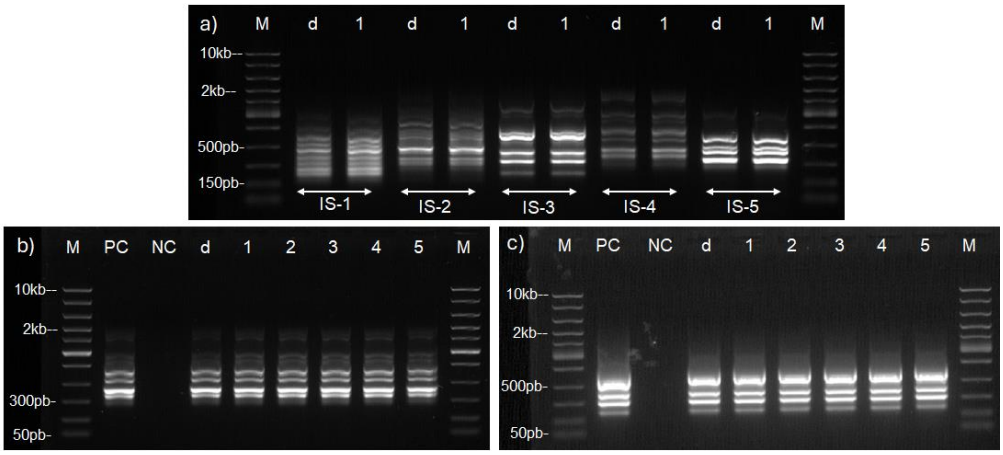


Figure 36. Genetic fidelity assessment. a) The five ISSR molecular markers profiling of *in vitro* regenerated plants with donor plants of *V. paradoxa*. M: Fast DNA Ladder; d: Donor plant DNA banding pattern; 1: in-vitro regenerated plants DNA banding patterns, b) banding pattern between regenerated plants with IS-4; c) banding pattern between regenerated plants with IS-3. PC and NC represent positive control and negative control respectively.

6.5. Discussion

In vitro regeneration of *Vitellaria paradoxa* is challenging. Providing an efficient *in vitro* protocol for shea tree regeneration is a prerequisite for its domestication and conservation regarding climate change.

Plant growth can be influenced by the appropriate use of heterocyclic compounds as a factor of plant growth regulators. The efficiency of plant regeneration depends on the composition of the medium: macro- and micronutrients, type and concentration of growth regulators and type of carbon sources (280).

In this study, half strength MS medium (MS/2) supplemented with Kin alone or combinations of BAP:Kin:NAA and BAP/NAA well induced new axillary shoots and leaves. Previous studies reporting the use of half-strength MS on *Vitellaria paradoxa* have demonstrated its efficacy in axillary shoot induction (41,43,44). Only half strength of MS medium was used in this study since full strength has been shown to have an inhibitory effect on *Vitellaria paradoxa* axillary shoot induction (41,43,44). Media containing full strength MS may have exhibited an inhibitory effect due to excessive ion concentration (281).

In this study, cytokines and auxin were used as plant growth hormones to induce axillary shoots. Cytokines are known to stimulate axillary shoot development (282,283). This is consistent with our investigation, in which media supplemented with Kin alone showed new axillary shoots on the explants. Several studies have used cytokines for *in vitro* shoot propagation of various woody plants and medicinal plants (44,283,284). However, while BAP alone induced axillary buds, the induced shoots were not elongated in shea tree. Different results were obtained in the micropropagation of *Achras sapota* (Sapotaceae) (267).

The combination of cytokines and auxin has often shown better growth and shoot proliferation in shea tree (41) and *Madhuca latifolia*, a Sapotaceae plant species (265). Our results are consistent with this, as we obtained better shoot proliferation and growth when cytokines and auxin were combined.

Combinations of the cytokines used in different proportions affected the growth in different patterns. In this study, explants showed variable responses to different concentrations and combinations of the plant growth regulators BAP, NAA and Kin. Numerous morphogenetic responses have been reported for *vitellaria paradoxa* in the presence of various plant growth regulators (41,43,44).

The concentration of cytokines used in different combinations promoted the growth of new axillary shoots and leaves as the experiment progressed. Similar results have been reported with increasing shoot length of *V. paradoxa* cultured on media supplemented with different ratios of BAP:NAA (41) and GA_3/BAP (43). The combination of two cytokines (BAP and Kin) with an auxin (NAA) produced the best results in terms of number of leaves and shoot length, compared to the combination of a cytokine and an auxin or the use of a single hormone. Authors reported that the combination of two different cytokines induced a higher number of shoots and higher shoot lengths compared to the singular effects of the individual hormones (43,285). The particular death rate (6.67%) of induced shoots on MS/2 media supplemented with BAP, Kin and NAA is attributed to the size of the initial explants. Indeed, explants shorter than 1.5 cm were converted to callus.

To our knowledge, the shea tree is a recalcitrant species in tissue culture. In fact, adventitious root formation is very difficult to achieve, while it is critical for successful vegetative propagation of this species (36). The development of healthy and vigorous roots is essential for successful acclimatization to ex vitro conditions.

The high percentage (100%) of rooting was obtained when young new axillary shoots (4-6 weeks) were transferred to media containing 3 mg/L IBA, 0.1 mg/L *m*T and 40 mg/L putrescine, while no signs of rooting were observed in the same media

with old axillary shoots (7 weeks and older), can be explained that *Vitellaria paradoxa* rooting is efficient when the explant is young. However, older axillary shoots could be used as explants for subculture because they had a high average shoot length. This result is consistent with that obtained in a previous study of V*itellaria paradoxa* cuttings rooting using juvenile and mature shea trees treated with IBA (36). The authors observed a rooting rate of 41% in juvenile shea trees, while only 28 % was observed in mature shea trees. The use of plant growth regulators such as putrescine has been shown to have a positive effect on adventitious root formation in *Argania spinosa* (Sapotaceae) (264) and other plant species (286). In our study, no signs of rooting were observed on media containing half and quarter-strength MS media. However, authors observed 8% rooting of *V. paradoxa* (41). *In vitro* rooting was reported when a combination of BAP/NAA was used for an axillary shoot induction study on shea tree (44).

Plantlet acclimatization is a critical step that determines the efficiency of the entire micropropagation process. During acclimation, the regenerants are exposed to new growth conditions that cause biotic and abiotic stresses. These stresses include water loss, tissue dehydration, and reduction in synthesis processes (287). The substrate used during acclimation has a significant effect on the survival and growth of plantlets regenerated by *in vitro* culture (264,288,289). Other authors noted the positive effect of *meta*-topolin on plantlet acclimatization when used in shoot induction media (290–292). The findings of this study show that *Vitellaria paradoxa* plantlets transferred in potting soil-perlite mixture showed efficient adaptation to ex vitro conditions, as well as good growth and development. It could be postulated that a low concentration of meta-topolin (0.1 mg/L) in the shea tree rooting medium could improve rooting and the acclimation of *in vitro* rooted plantlets. However, further studies are needed to confirm this claim. Mixture substrate, such as the combination of potting soil and perlite, has demonstrated its efficiency in ex vitro acclimation of plant species of the Sapotaceae family (39,264).

Few studies have been conducted on micropropagation systems of Sapotaceae family plant species, and none of them has evaluated the genetic homogeneity of plants grown *in vitro*. ISSR markers have been widely used to assess genetic fidelity in various plant species (271,273,275,276,293,294).

In this study, the polymorphic level of 5 ISSR primers was primarily evaluated in shea tree. The polymorphic and monomorphic percentages were 65% and 35%, respectively, while the polymorphism information content was 0.22. This indicated a moderate genetic diversity among the shea trees using the ISSR markers (277). Regarding the polymorphism information found in this study, the ISSR primers demonstrated the advantages of ISSR markers in detecting genetic variation in *V. paradoxa*. Therefore, the five ISSR primers were then used to evaluate the genetic fidelity of *V. paradoxa* plantlets regenerated by our *in vitro* regeneration protocol. The use of molecular markers to investigate genetic fidelity in plants with long juvenile periods (10-15 years) such as *V. paradoxa* is primordial. Regarding the genetic fidelity analysis, the five ISSR primers produced 34 bands ranging from 200 to 2000 bp. These results were consistent with previous genetic fidelity assessment studies in

Muehlenbeckia platyclada micropropagation (275), in *Crinum brachynema* (295), *Lilium* spp (296), *Curcuma angustifolia* (293), and *Acacia auriculiformis* (294). The plantlets produced *in vitro* did not show any polymorphisms as the DNA bands amplified with the ISSR primers were monomorphic to the donor plant (**Figure 36**).

6.6. Conclusion

The efficiency of micropropagation of *V. paradoxa* depends on the Murashige and Skoog medium composition, the type and concentration of plant growth regulators. This study showed differences in the response of *Vitellaria paradoxa* tissue cultures to different concentrations and different regimes (single or combined) of plant growth regulator application. The best shoot regeneration response was observed on half strength MS medium supplemented with 3:1.2:1 mg/L BAP:Kin:NAA. The presence of the cytokinin Kin in the hormone combinations resulted in better shoot growth and leaf development, although the cytokinin Kin alone did not induce vigorous shoot growth. Rooting was induced only on half strength of modified MS medium with 3:0.1:40 mg/L IBA:*m*T:Putrescine. Plantlets were successfully acclimated on universal compost-perlite mixture substrate. No genetic alterations were found in the regenerated plants based on ISSR analysis. The developed micropropagation protocol is the most advanced in *V. paradoxa* micropropagation and genetic fidelity analysis is helpful for mass multiplication, conservation, and *in vitro* culture-based biotechnological experiments. The micropropagation protocol presented in this study can be used for routine *in vitro* regeneration of *Vitellaria paradoxa*. However, further studies are required for the induction of multiple microshoots.

Chapter 7

General Discussion, conclusions and perspectives

General Discussion, conclusions and perspectives

7.1. General Discussion

The research presented in this thesis represents a concerted effort to establish a comprehensive shea tree improvement program in Côte d'Ivoire. By systematically addressing conservation, genetic characterization, and the identification of economically valuable traits, the work lays the foundation for a sustainable and resilient shea cultivation system that is both scientifically robust and locally grounded. The integration of participatory practices with scientific innovation holds the potential to revolutionize the shea industry, ensuring improved yields and product quality while maintaining biodiversity and environmental integrity.

The research is structured around an interconnected sequence of steps that lead to the identification and propagation of superior shea tree genotypes. The program began with a participatory approach to identifying superior shea trees (SSTs), leveraging the invaluable local knowledge of farmers. A recent study highlighted the importance of participatory tree selection approach in capturing wider variation and located extreme phenotypes in indigenous fruit tree domestication (297). As demonstrated in Chapter 3, farmers selected SSTs based on criteria such as fruit size, yield, pulp taste, and consistency in production. These traits reflect the local priorities of the shea-growing communities and set a foundation for the scientific validation of these selections. The combination of farmers' knowledge with morphological analysis offers a framework for breeding that is both locally relevant and genetically diverse (298,299). This participatory model is particularly critical in the context of sustainable agriculture, where the priorities of smallholder farmers must align with scientific innovations to ensure long-term success (300,301). Participatory plant breeding approaches have been apply to a wide range of plant species including potato, tomato, wheat, maize and Brassica (302).

One of the main challenges encountered during the morphological assessment of SSTs in Chapter 3 was the relatively high mortality rates and susceptibility to hemiparasitic plant infestations, which impede the sustainability of shea populations. The *in situ* conservation of 8 plant species highlighted that these species are threatened of declining (303). These challenges highlight the importance of genetic diversity as a fundamental driver of resilience. Chapters 3 and 4 underscore the rich morphological and genetic diversity found in Côte d'Ivoire's shea populations, which is critical for enhancing the species' adaptive potential, especially in response to environmental pressures such as climate change (304–307). This trend was observed in the morphological characterization of *Phoebe cooperiana* genotypes, an indigenous plant species in India, selected through participatory survey (297). The exploration of molecular diversity using SNP markers further enriched the genetic profile of the selected superior shea trees, establishing a valuable *ex situ* core collection to conserve

this diversity. These molecular techniques pave the way for more precise genetic interventions, supporting long-term conservation efforts and laying the groundwork for the quantitative trait nucleotide (QTN) identification (25,63). While *ex situ* core collections, established through molecular diversity analyses, offer a valuable conservation strategy, the research emphasizes that *in situ* conservation strategies are equally essential. This dual approach ensures that the dynamic and evolving nature of shea tree populations is maintained in their natural habitats, which is crucial for safeguarding the long-term health and productivity of the species (308–310).

The discovery of quantitative trait nucleotides (QTNs) linked to shea butter content and fatty acid composition in Chapter 5 represents a pivotal step toward precision breeding. Marker-assisted selection (MAS) provides an efficient tool for selecting trees with desirable traits early in the breeding process (311,312). By identifying QTNs, the research facilitates targeted breeding efforts that focus on economically valuable traits, such as improved oil quality and yield. However, translating molecular discoveries into practical applications poses several challenges. For instance, validating QTN associations across diverse environments is crucial to ensuring that the identified markers are reliable and consistent (313). Additionally, controlled breeding populations must be developed to facilitate crossbreeding that can lead to the creation of superior hybrids. The integration of genomic selection (GS) into breeding strategies would further enhance the precision of trait selection, improving the speed and accuracy of the breeding process (314–316).

Another significant challenge identified in this research is the propagation bottleneck that limits shea tree domestication. Shea trees are notoriously slow-growing, with a long juvenile phase and large yield variability (67,317,318). The research in Chapter 6 addresses this issue by developing in vitro propagation, which could enable the rapid multiplication of elite genotypes and more importantly for the research. For instance, micropropagation has played an important role in the production of healthy, disease-free plants and in the rapid multiplication of scions and rootstocks with desirable traits in apple (282). Grafting approaches have been also optimized for high successful rate in shea tree (29,31,32,318). These techniques offer great promise for overcoming the propagation constraints associated with shea tree domestication. However, scalability remains a key hurdle. For smallholder farmers to benefit from these innovations, the techniques must be cost-effective and suitable for widespread adoption. Issues such as graft compatibility and the reduction of propagation costs must be addressed to ensure that these advancements can be implemented at scale (297,319,320).

The integrated shea tree improvement program outlined in this study envisions the use of both *ex situ* and *in situ* conservation strategies, alongside advanced molecular techniques and traditional breeding methods. The establishment of a comprehensive germplasm collection, together with continued genetic characterization through molecular markers, will ensure that valuable genetic diversity is preserved and that

selection efforts are guided by robust genetic information. Expanding the scope of genome-wide association studies (GWAS) to include traits such as drought tolerance, disease resistance, and nutritional quality would further strengthen the breeding program, particularly in the face of emerging environmental challenges (321,322).

A key aspect of the improvement program is the development of breeding and selection strategies that utilize MAS and GS to accelerate breeding cycles. Controlled pollination and the creation of breeding populations will allow for the development of superior hybrids with traits that are economically desirable and adapted to local environments. Additionally, optimizing propagation methods for large-scale multiplication of these genotypes will facilitate the transition from wild harvesting to more sustainable orchard-based cultivation systems (323–325). As such, this integrated approach is likely to yield a system where shea cultivation can be expanded, more efficiently managed, and made economically viable for smallholder farmers.

Finally, the success of this program depends on capacity building and active farmer involvement. It is crucial that local farmers are trained in modern propagation techniques, breeding strategies, and orchard management to ensure that they are equipped to implement these innovations effectively. Engaging farmers in on-farm trials and demonstrations will help bridge the gap between scientific research and real-world agricultural practices, ensuring the adoption of improved genotypes and best management practices (326–328).

The research conducted in this thesis provides a solid foundation for the conservation, genetic improvement, and sustainable cultivation of shea trees in Côte d'Ivoire. By integrating traditional knowledge with modern scientific techniques, the program offers a pathway to enhancing the productivity and resilience of shea cultivation. Scaling up propagation techniques, addressing emerging environmental challenges, and fostering collaborations among researchers, farmers, and policymakers will be key to ensuring the long-term success of this shea tree improvement program. Ultimately, this integrated approach not only promises to secure the future of shea production in Côte d'Ivoire but also serves as a model for other underutilized tree crops in the region.

7.2. General Conclusions

This thesis investigated the sustainable conservation and genetic improvement of the shea tree (*Vitellaria paradoxa*) in Côte d'Ivoire through a multi-disciplinary approach encompassing morphological, molecular, and biotechnological studies. The work addressed critical issues surrounding the conservation of this vital species and sought innovative solutions to enhance its productivity and adaptability for local and international applications.

Key results are as follows:

1. Assessment of *In Situ* Conservation Efforts

The *in situ* conservation of 405 "superior" shea trees, established in 2017, revealed significant challenges:

- **Mortality and Threats**: After three years, 8% of the conserved trees had died, with Bagoué region exhibiting a higher mortality rate (12%) than Tchologo region (5%). Additionally, 84% of trees were infested with hemiparasitic plants (Loranthaceae), compromising sustainability.

- **Diversity and Morphology**: Morphological analyses of 220 trees showed significant diversity across 12 qualitative traits, such as crown shape, branch density, and leaf color. Shannon diversity indices ranged from 0.55 to 0.98, demonstrating high genetic variability. The most prevalent crown shape was broadly pyramidal (26.36%).

The *in situ* collection of shea trees in Côte d'Ivoire exhibits significant genetic diversity, which is essential to preserve for future generations. However, this study highlights the limitations inherent in *in situ* conservation methods. To address these challenges and ensure the long-term preservation and utilization of this valuable genetic heritage, the implementation of an *ex situ* conservation approach is strongly recommended. Therefore, a dual conservation approach, combining both *in situ* and *ex situ* methods, is necessary to ensure the sustainability of shea genetic resources. The study contributes to the scientific understanding of shea diversity and provides actionable recommendations for its preservation and improvement.

2. Molecular Diversity and Core Collection Development

The molecular study, leveraging 7,559 SNP markers, underscored the moderate genetic diversity within the shea population:

- **Diversity Metrics**: The average genetic diversity was 0.26, and AMOVA revealed an intra-individual variance of 100%, indicating low genetic differentiation.

- **Core Collection**: A core collection of 100 trees was established, capturing 30% of the population while preserving all allelic diversity. This collection forms the foundation for an *ex situ* conservation strategy to mitigate the limitations of *in situ* conservation.

Genomic studies have enabled the identification of a reduced core collection of individuals, encompassing the entire genetic diversity of the *in situ* collection while significantly reducing the number of specimens. This approach rationalizes the establishment of the collection in terms of cost, space, and long-term maintenance labor. The core collection will be established at Botanical garden of University Peleforo Gon Coulibaly of Korhogo and managed by the African Centre for Shea Research and Applications. It will serve as a valuable resource for shea tree varietal improvement in Côte d'Ivoire and other countries engaged scientifically and economically by the shea tree.

3. Genomic Insights into Shea Butter Traits

A genome-wide association study (GWAS) identified 25 significant quantitative trait nucleotides (QTNs) linked to shea butter production. Key findings include:

- **Fatty Acid Composition**: Oleic and stearic acids, constituting 85–90% of total fatty acids, were primary determinants of shea butter's unique properties.
- **Geographic Variability**: Region-level differences in fatty acid profiles highlighted the potential for region-specific cultivation strategies to optimize shea butter quality.

The assessment of the shea butter content of the superior shea trees enabled the identification of genotypes with high agronomic value in terms of shea butter production. These shea trees will serve as parents for varietal improvement. The high fat genotypes produce an interesting shea butter in terms of fatty acid composition. The shea butter of Côte d'Ivoire could be used in different application domains. However, a complete characterization of this shea butter is essential in order to determine specific domain of application.

4. Biotechnological Advancements in Propagation

The development of vegetative propagation techniques provided viable pathways for conservation and domestication:

- **Grafting**: Demonstrated high success rates but was limited by factors such as rootstock influence and seasonality.
- ***In Vitro* Culture**: A breakthrough axillary shoot regeneration protocol using MS/2 medium supplemented with specific growth regulators achieved 100% rooting success. Genetic fidelity of *in vitro*-propagated plants was confirmed using ISSR markers, ensuring they were true-to-type.
- ***Research***: Micropropagation is a powerful technique widely used in plant science for the rapid multiplication of plants under controlled laboratory conditions. Its significance in tree research lies in several key areas, including mass production of elite genotype, conservation of genetic resources, overcome propagation barriers, support for breeding programs, pathogen-free plant production and facilitating biotechnology research.

The micropropagation study has provided a suitable protocol for *in vitro* regeneration of the shea tree. This protocol will serve to multiply the genotypes of interest, thereby enhancing the agronomic potential and sustainability of shea tree cultivation.

7.3. Perspectives and Futures Directions

Building upon the findings of this research, several avenues can be pursued to enhance the conservation, genetic improvement, and productivity of the shea tree (*Vitellaria paradoxa*). These recommendations aim to ensure the species' long-term preservation and maximize its socio-economic and ecological benefits:

1. **Implementation and Management of the *Ex Situ* Collection:**

The 100 genotypes in the core collection should be propagated via grafting within six months and conserved in a secure *ex situ* site, such as the botanical garden at Peleforo Gon Coulibaly University. Establishing a centralized *ex situ* repository will ensure long-term preservation and enable controlled breeding programs for the genetic improvement of the shea tree.

2. Assess shea trees production

To assess the production of shea trees, key parameters such as fruit yield need to be evaluated. These indicators provide insights into the productivity and economic potential of individual trees, supporting selection for breeding and domestication programs.

3. **Comprehensive Characterization:**

Further characterization of superior shea trees should include triglyceride profiles, melting points, and free fatty acid content, which are critical parameters for industrial applications, particularly in food and cosmetics sectors.

4. **Molecular Breeding:**

Leverage quantitative trait nucleotides (QTNs) and simple sequence repeat (SSR) markers for marker-assisted selection (MAS) and genome-wide selection. These tools can expedite the development of high-fat-content shea varieties and improve breeding efficiency.

5. **Advanced Genomic Studies:**

Employ cutting-edge genome editing tools, such as CRISPR, to target and enhance essential traits, including oil yield, fatty acid composition, and resistance to biotic and abiotic stressors.

6. **Integrated Pest Management:**

Develop sustainable and environmentally friendly strategies to control Loranthaceae infestations, including biological control agents and agroecological management practices.

7. **Domestication Programs:**

Scale up vegetative propagation techniques and distribute genetically superior planting materials to farmers. These efforts will help reduce the long juvenile phase of shea trees and boost productivity.

8. **Climate Resilience:**

Investigate the adaptive capacity of various genetic groups to climate variability. Integrate these findings into reforestation and conservation strategies to enhance the resilience of shea populations to changing environmental conditions.

9. **Economic Empowerment:**

Empower local communities, particularly women, by promoting value-added processing of shea butter. This approach will improve economic benefits and incentivize conservation efforts.

10. **Environmental Genomics:**

Conduct genome-wide association studies (GWAS) that integrate genomic, biochemical, and climatic data to uncover the impact of environmental factors on fatty acid biosynthesis and tree adaptation.

11. **Optimization of Propagation Methods:**

Improve *in vitro* propagation protocols by increasing the number of shoots per explant and refining cutting techniques. These optimizations will provide cost-effective propagation methods suitable for widespread application, especially in developing countries.

12. **Community Engagement:**

Involve local communities in conservation and propagation initiatives, highlighting the economic and ecological advantages of sustainable shea tree management. This approach aims to reduce tree cutting and habitat destruction.

This thesis lays a strong foundation for an integrated approach to conserving and utilizing shea resources. By balancing ecological sustainability with socio-economic development, these strategies can ensure the long-term viability of the shea tree while contributing to the livelihoods of local communities and meeting the demands of global industries.

References

1. Coelho N, Gonçalves S, Romano A. Endemic Plant Species Conservation: Biotechnological Approaches. Plants. mars 2020;9(3):345.

2. Stalker HT, Warburton ML, Harlan JR. Harlan's Crops and Man: People, Plants and Their Domestication. John Wiley & Sons; 2021. 324 p.

3. Savadi S, Mangalassery S, Sandesh MS. Advances in genomics and genome editing for breeding next generation of fruit and nut crops. Genomics. 1 nov 2021;113(6):3718-34.

4. Neale DB, Kremer A. Forest tree genomics: growing resources and applications. Nat Rev Genet. févr 2011;12(2):111-22.

5. Gaisberger H, Fremout T, Kettle CJ, Vinceti B, Kemalasari D, Kanchanarak T, et al. Tropical and subtropical Asia's valued tree species under threat. Conserv Biol. juin 2022;36(3):e13873.

6. Plieninger T, Muñoz-Rojas J, Buck LE, Scherr SJ. Agroforestry for sustainable landscape management. Sustain Sci. 1 sept 2020;15(5):1255-66.

7. Sallé G, Boussim JI, Raynal-Roques A, Brunck F. Physiologie des arbres et arbustes en zones arides et semi-arides. John Libbey; 1993. Le karité : état de nos connaissances et perspectives de recherche.

8. Hall JB, Aebischer DP, Tomlinson HF, Osei-Amaning E, Hindle JR. Vitellaria paradoxa: a monograph. Publisher: School of Agricultural and Forest Sciences, University of Wales. 1996;

9. Glew D, Lovett PN. Life cycle analysis of shea butter use in cosmetics: from parklands to product, low carbon opportunities. Journal of Cleaner Production. 1 avr 2014;68:73-80.

10. Naughton CC, Lovett PN, Mihelcic JR. Land suitability modeling of shea (Vitellaria paradoxa) distribution across sub-Saharan Africa. Applied Geography. 1 mars 2015;58:217-27.

11. Boffa J, Yaméogo G, Nikiema P, Taoda JB. What future for the shea tree. Agroforestry today [Internet]. 1996 [cité 31 oct 2024]; Disponible sur: https://www.semanticscholar.org/paper/What-future-for-the-shea-tree-Boffa-Yam%C3%A9ogo/1abda64e7b06344fb03969306a787e0d3cb37bca

12. Maranz S, Wiesman Z. Evidence for indigenous selection and distribution of the shea tree, Vitellaria paradoxa, and its potential significance to prevailing parkland savanna tree patterns in sub-Saharan Africa north of the equator. Journal of Biogeography. 2003;30(10):1505-16.

13. Sanou H, Lamien N. Vitellaria paradoxa, shea butter tree. Conservation and Sustainable Use of Genetic Resources of Priority Food Tree Species in sub-Saharan Africa. Biovers Internat. 2011;

14. Allal F, Sanou H, Millet L, Vaillant A, Camus-Kulandaivelu L, Logossa ZA, et al. Past climate changes explain the phylogeography of Vitellaria paradoxa over Africa. Heredity. août 2011;107(2):174-86.

15. Sirima D, Konate SR, Lankoande A, Hema EE. Contribution du karité dans la réduction de dégâts de cultures annuelles par les éléphants autour de la forêt classée de Niangoloko, Burkina Faso. Sciences Naturelles et Appliquées. 31 oct 2024;43(1):107-22.

16. NEPC. shea-nut-profile [Internet]. NEPC. 2023 [cité 12 nov 2024]. Disponible sur: https://nepc.gov.ng/blog/market-report/shea-nut-profile/

17. USAID. Natural Resource Product Analysis - Shea Radmap [Internet]. 2018. Disponible sur: https://pdf.usaid.gov/pdf_docs/PA00TGBP.pdf

18. UNDP. Rapport de consultation nationale des femmes dans le commerce. 2023.

19. Lovett P. The Shea Butter Value Chain: Production, Transformation & Marketing in West Africa. USAID West Africa Trade Hub; 2004. Report No.: Technical Report No. 2.

20. Carette C, Malotaux M, Leeuwen MV, Tolkamp M, Kassaw A, Dordaa S, et al. Shea nut and butter in Ghana Opportunities and constraints for local processing. In 2009 [cité 19 oct 2023]. Disponible sur: https://www.semanticscholar.org/paper/Shea-nut-and-butter-in-Ghana-Opportunities-and-for-Carette-Malotaux/bd5a492a393875a3bd8701c61a0f47b9e2baa1bf?sort=relevance&citationIntent=methodology

21. Bup DN, Mohagir AM, Kapseu C, Mouloungui Z. Production zones and systems, markets, benefits and constraints of shea (Vitellaria paradoxa Gaertn) butter processing. OCL. 1 mars 2014;21(2):D206.

22. Rousseau K, Gautier D, Wardell DA. Coping with the upheavals of globalization in the shea value chain: the maintenance and relevance of upstream

shea nut supply chain organization in western Burkina Faso. World Development. 2015;66:413-27.

23. Cardi C, Vaillant A, Sanou H, Kelly BA, Bouvet JM. Characterization of microsatellite markers in the shea tree (Vitellaria paradoxa C. F Gaertn) in Mali. Molecular Ecology Notes. 2005;5(3):524-6.

24. Sanou H, Lovett PN, Bouvet JM. Comparison of quantitative and molecular variation in agroforestry populations of the shea tree (Vitellaria paradoxa C.F. Gaertn) in Mali. Molecular Ecology. 2005;14(8):2601-10.

25. Hale I, Ma X, Melo ATO, Padi FK, Hendre PS, Kingan SB, et al. Genomic Resources to Guide Improvement of the Shea Tree. Frontiers in Plant Science [Internet]. 2021;12. Disponible sur: https://www.frontiersin.org/articles/10.3389/fpls.2021.720670

26. IUCN. IUCN Red List of Threatened Species: Vitellaria paradoxa. IUCN Red List of Threatened Species [Internet]. 1 janv 1998 [cité 27 sept 2023]; Disponible sur: https://www.iucnredlist.org/en

27. Diarrassouba N. Caractérisation morphologique chez le karité : Vitellaria paradoxa C. F. Gaertn. (Sapotaceae) et analyse de certains facteurs anthropiques affectant la structure de la diversité génétique des peuplements en Côte d'Ivoire. [(Doctoral dissertation, Thèse de Doctorat Unique ès sciences, Génétique et Amélioration des plantes, Université d'Abidjan-Cocody (Côte d'Ivoire)).]. [Abidjan, Côte d'Ivoire]: Université de Cocody; 2008.

28. Lovett PN, Haq N. Diversity of the Sheanut tree (Vitellaria paradoxa C.F. Gaertn.) in Ghana. Genetic Resources and Crop Evolution. 1 juin 2000;47(3):293-304.

29. Sanou H, Kambou S, Teklehaimanot Z, Dembélé M, Yossi H, Sina S, et al. Vegetative propagation of Vitellaria paradoxa by grafting. Agroforestry Systems. 1 janv 2004;60(1):93-9.

30. Sawadogo BGB. Greffage In Situ de Vitellaria paradoxa C.F. Gaertn Dans La Province du Sanguie. Universite Nazi Boni (UNB), Dindéresso, Burkina Faso; 2018.

31. Yao SDM, Alui KA, Kouame NMT, Ble PA, Kone B, Diarrassouba N. Réussir le « greffage en fente simple » et le « greffage de côté dans l'aubier » du karité. Journal of Applied Biosciences. 30 juill 2019;137:13961-72.

32. Yao SDM, Diarrassouba N, Alui KA, Fofana VP, Ble PA, Diallo R, et al. Effect of the Grafting Method on the Recovery and Growth of Juvenile Shea (Vitellaria Paradoxa Gaertn C.F) Plants Grafted in Nursery. East African Scholars Journal of Agriculture and Life Sciences. 2020;3(12):406-14.

33. Yeboah J, Lowor S, Amoah F. The rooting performance of shea (Vitellaria paradoxa gaertn) stem cuttings as influenced by wood type, sucrose and rooting hormone. Scientific Research and Essays. 2009;4(5):521-5.

34. Yeboah J, Akrofi AY, Owusu-Ansah F. Influence of selected fungicides and hormone on the rooting success of Shea (Vitellaria paradoxa Gaernt) stem cuttings. Agric Biol JN Am. 2010;1(3):313-20.

35. Yeboah J, Segbefia MAD, Dadzie AM, Padi F, Lowor ST, Agene VN, et al. Cutting Propagation of Shea (Vitellaria paradoxa C.F. Gaertn) Tree Using Shoot Types and Application of Auxin. JAS. 15 nov 2020;12(12):213.

36. Amissah N, Akakpo B, Yeboah J, Blay E. Asexual Propagation of Sheanut Tree (Vitellaria paradoxa C.F. Gaertn.) Using a Container Layering Technique. American Journal of Plant Sciences. 21 août 2013;04(09):1758.

37. Akakpo DB, Amissah N, Yeboah J, Blay E. Effect of Indole 3-Butyric Acid and Media Type on Adventitious Root Formation in Sheanut Tree (Vitellaria paradoxa; C. F. Gaertn.) Stem Cuttings. AJPS. 2014;05(03):313-8.

38. Fotso, Sanonne, Omokolo Ndoumou D, Donfagsiteli Tchinda N. Comparaison des premières étapes de l'embryogenèse somatique chez Baillonella toxisperma et Vitellaria paradoxa (Sapotacées). Biotechnol Agron Soc Environ. 1 janv 2008;12(2):131-8.

39. Adu-Gyamfi PKK, Barnor MT, Dadzie AM, Lowor S, Opoku SY, Opoku-Ameyaw K, et al. Preliminary investigation on somatic embryogenesis from immature cotyledon explants of shea (Vitellaria paradoxa G.). Journal of Agricultural Science and Technology. 2012;2:1171-6.

40. Issali AE, Diarrassouba N, Nguessan EA, Traore A, Kouassi KM, Amoncho A, et al. Typology of seven Vitellaria paradoxa genotypes in relation to their response to callogenesis. IJAIEM. 2013;2(2):160-9.

41. Lovett PN, Haq N. Progress in developing in vitro systems for shea tree (Vitellaria paradoxa C.F. Gaertn.) propagation. Forests, Trees and Livelihoods. 1 mars 2013;22(1):60-9.

42. Aisagbonhi EP, Isalar CE, Odenore VD, Ogbebor JU, Eke CR, Asemota O, et al. The interplay between explant developmental stages and phytohormone type in callogenesis of shea tree (Vitellaria paradoxa C. F. Gaertn). EIJST. 2015;4(7):50-7.

43. Afolabi JO, Olorode EM, Olomola DB, Fasakin YO, Adekunle EA. Effects of different media strengths and hormone concentrations on in-vitro regeneration of *Vitellaria paradoxa* C.F. Gaertn. Nig J Biotechnol. 28 août 2020;37(1):150-8.

44. Afful NT, Abdulai I, Azu E, Elegba W, Annor C, Akama C, et al. In vitro regeneration of Vitellaria paradoxa from shoot tip explants. BioTechnologia (Pozn). 2022;103(1):71-9.

45. Soro D, N'Da-Adopo A, Da KP, Traore D. Lutte Contre Les Parasites Chez Le Karite. Agronomie Africaine. 2004;16(3):21-8.

46. Lamien N, Tigabu M, Dabiré R, Guinko S, Oden PC. Insect (Salebria sp.) infestation and impact on Vitellaria paradoxa CF Gaertn. fruit production in agroforestry parklands. Agroforestry Systems [Internet]. 2008 [cité 8 nov 2024];72(1). Disponible sur: https://res.slu.se/id/publ/78166

47. Boussim IJ, Yonli D, Medah NM, Guinko S, Salle G. Prospects for an integrated control of Loranthaceae species parasitizing Vitellaria paradoxa C. F. Gaertn in Burkina Faso. International Journal of Biological and Chemical Sciences. 2012;6(1):355-64.

48. Ahamidé IDY, Tossou MG, Yédomonhan H, Adomou AC, Houénon J, Akoègninou A. Diversité Des Loranthaceae Et Leur Impact Sur Vitellaria Paradoxa C.F.Gaertn.: Un Fruitier À Grande Valeur Socio-Économique Au Nord-Bénin. European Scientific Journal, ESJ. 31 août 2017;13(24):217-217.

49. Yao SDM, Diarrassouba N, Alui KA, N'Goran KSB, Koffi BZE, Fofana IJ. Loranthaceae Species Infesting Shea Trees (Vitellaria paradoxa Gaertn. C.F.) and Factors Involving Attacks in Northern Côte D'Ivoire. Ecologia. 2020;10(3):101-9.

50. Fontaine C, Lovett PN, Sanou H, Maley J, Bouvet JM. Genetic diversity of the shea tree (Vitellaria paradoxa C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers. Heredity (Edinb). déc 2004;93(6):639-48.

51. Sanou H, Picard N, Lovett PN, Dembélé M, Korbo A, Diarisso D, et al. Phenotypic Variation of Agromorphological Traits of the Shea Tree, Vitellaria paradoxa C.F. Gaertn., in Mali. Genet Resour Crop Evol. 1 févr 2006;53(1):145-61.

52. Gwali S, Vaillant A, Nakabonge G, Okullo JBL, Eilu G, Muchugi A, et al. Genetic diversity in shea tree (Vitellaria paradoxa subspecies nilotica) ethno-varieties in Uganda assessed with microsatellite markers. Forests, Trees and Livelihoods. 3 juill 2015;24(3):163-75.

53. Abdulai I, Krutovsky KV, Finkeldey R. Morphological and genetic diversity of shea tree (Vitellaria paradoxa) in the savannah regions of Ghana. Genet Resour Crop Evol. 1 août 2017;64(6):1253-68.

54. Odoi JB, Adjei EA, Hendre P, Nantongo JS, Ozimati AA, Badji A, et al. Genetic diversity and population structure among Ugandan shea tree (Vitellaria paradoxa subsp. nilotica) accessions based on DarTSeq markers. Crop Science. 1 juill 2023;63(4):2297-309.

55. Okullo JBL, Hall JB, Obua J. Leafing, flowering and fruiting of Vitellaria paradoxa subsp. nilotica in savanna parklands in Uganda. Agroforestry Systems. 1 janv 2004;60(1):77-91.

56. Bayala J, Ouédraogo SJ, Ong CK. Early growth performance and water use of planted West African provenances of Vitellaria paradoxa C.F. Gaertn (karité) in Gonsé, Burkina Faso. Agroforest Syst. 1 févr 2009;75(2):117-27.

57. Aleza K, Villamor GB, Nyarko BK, Wala K, Akpagana K. Shea (Vitellaria paradoxa Gaertn C. F.) fruit yield assessment and management by farm households in the Atacora district of Benin. PLOS ONE. 18 janv 2018;13(1):e0190234.

58. Kelly BA, Poudyal M, Bouvet JM. Impact of land use and land use history on fruits production of Vitellaria paradoxa (Shea tree) according to agroclimatic zones in Mali (West Africa). Current Botany. 2019;01-7.

59. Wei Y, Ji B, Siewers V, Xu D, Halkier BA, Nielsen J. Identification of genes involved in shea butter biosynthesis from Vitellaria paradoxa fruits through transcriptomics and functional heterologous expression. Appl Microbiol Biotechnol. 1 mai 2019;103(9):3727-36.

60. Diarrassouba N, Yao S.D.M., Traoré B. Identification participative et caractérisation des arbres élites de karité dans la zone de production en Côte d'Ivoire. 2017 p. 15 pages. (projet FIRCA/Karité). Report No.: N° 069/2016.

61. Sandwidi A, Diallo BO, Lamien N, Vinceti B, Sanon K, Coulibaly P, et al. Participatory identification and characterisation of shea butter tree (Vitellaria paradoxa C.F. Gaertn.) ethnovarieties in Burkina Faso. Fruits, The International Journal of Tropical and Subtropical Horticulture. 2018;73(3):141-52.

62. Odoi JB, Muchugi A, Okia CA, Gwali S, Odong TL. Local knowledge, identification and selection of shea tree (Vitellaria paradoxa) ethnovarieties for pre-breeding in Uganda. The Journal of Agriculture and Natural Resources Sciences. 5 août 2020;7(1):22-33.

63. Odoi JB, Adjei EA, Barnor MT, Edema R, Gwali S, Danquah A, et al. Genome-Wide Association Mapping of Oil Content and Seed-Related Traits in Shea Tree (Vitellaria paradoxa subsp. nilotica) Populations. Horticulturae. juill 2023;9(7):811.

64. Elias M. Transforming nature's subsidy: Global markets, Burkinabè women and African shea butter [Internet]. McGill University; 2010 [cité 31 janv 2025]. Disponible sur: https://escholarship.mcgill.ca/concern/theses/5x21tf758

65. Diarrassouba N. Connaissances locales et leur utilisation dans la gestion des parcs à karité en Côte d'Ivoire. Afrika Focus [Internet]. 1 janv 2008 [cité 31 janv 2025]; Disponible sur: https://www.academia.edu/85440702/Connaissances_locales_et_leur_utilisation_dans_la_gestion_des_parcs_%C3%A0_karit%C3%A9_en_C%C3%B4te_d_Ivoire

66. Carney J, Elias M. Revealing Gendered Landscapes: Indigenous Female Knowledge and Agroforestry of African Shea. Canadian Journal of African Studies / Revue canadienne des études africaines. 1 janv 2006;40(2):235-67.

67. Boffa JM. Opportunities and challenges in the improvement of the shea (Vitellaria paradoxa)" (2015). resource and its management. 2015;

68. Improving Land Access for Women [Internet]. LandLinks. [cité 31 janv 2025]. Disponible sur: https://www.land-links.org/project/improving-land-access-for-women/

69. Elias M. Gender, knowledge-sharing and management of shea (Vitellaria paradoxa) parklands in central-west Burkina Faso. Journal of Rural Studies. 1 avr 2015;38:27-38.

70. Kiptot E, Franzel S. Gender and agroforestry in Africa: a review of women's participation. Agroforest Syst. 1 janv 2012;84(1):35-58.

71. Ocheje PD. Women's land rights in sub-Saharan Africa: Between the law and cultural norms 1. In: Expanding Perspectives on Human Rights in Africa. Routledge; 2019.

72. Boffa JM, Knudson DM, Yameogo G, Nikiema P. Shea nut (Vitellaria paradoxa) production and collection in agroforestry parklands of Burkina Faso. Non-Wood Forest Products (FAO). 1996;

73. Gakou M, Force JE, McLaughlin WJ. Non-timber forest products in rural Mali: a study of villager use. Agroforest Syst. 1 déc 1994;28(3):213-26.

74. Elias M, Carney J. Shea Butter, Globalization, and Women of Burkina Faso. In: A Companion to Feminist Geography [Internet]. John Wiley & Sons, Ltd; 2005 [cité 31 janv 2025]. p. 93-108. Disponible sur: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470996898.ch7

75. Gwali S, Okullo JBL, Eilu G, Nakabonge G, Nyeko P, Vuzi P. Traditional management and conservation of shea trees (Vitellaria paradoxa subspecies nilotica) in Uganda. Environ Dev Sustain. 1 juin 2012;14(3):347-63.

76. Luedeling E, Neufeldt H. Carbon sequestration potential of parkland agroforestry in the Sahel. Climatic Change. 1 déc 2012;115(3):443-61.

77. Sanogo K, Gebrekirstos A, Bayala J, Villamor GB, Kalinganire A, Dodiomon S. Potential of dendrochronology in assessing carbon sequestration rates of *Vitellaria paradoxa* in southern Mali, West Africa. Dendrochronologia. 1 déc 2016;40:26-35.

78. Dimobe K, Kuyah S, Dabré Z, Ouédraogo A, Thiombiano A. Diversity-carbon stock relationship across vegetation types in W National park in Burkina Faso. Forest Ecology and Management. 15 avr 2019;438:243-54.

79. Alui KA, Diarrassouba N, Yao SDM. Carbon Sequestration Potential of Shea Trees (Vitellaria paradoxa C.F. Gaertn.) in Parklands under Two Soil Types (Ferralsol and Cambisol) in Northern Côte d'Ivoire. International Journal of Sciences. 2020;9(02):14-23.

80. Baziari F, Henquinet KB, Cavaleri MA. Understanding farmers' perceptions and the effects of shea (Vitellaria paradoxa) tree distribution in agroforestry parklands of Upper West Region, Ghana. Agroforest Syst. 1 avr 2019;93(2):557-70.

81. Nyamekye C, Thiel M, Schönbrodt-Stitt S, Zoungrana BJB, Amekudzi LK. Soil and Water Conservation in Burkina Faso, West Africa. Sustainability. sept 2018;10(9):3182.

82. Ogwok G, Alele PO, Kizza S. Influence of Shea tree (Vitellaria paradoxa) on maize and soybean production. PLOS ONE. 9 avr 2019;14(4):e0201329.

83. Teklehaimanot Z. Exploiting the potential of indigenous agroforestry trees: Parkia biglobosa and Vitellaria paradoxa in sub-Saharan Africa. In: Nair PKR, Rao MR, Buck LE, éditeurs. New Vistas in Agroforestry: A Compendium for 1st World Congress of Agroforestry, 2004 [Internet]. Dordrecht: Springer Netherlands; 2004 [cité 3 nov 2024]. p. 207-20. Disponible sur: https://doi.org/10.1007/978-94-017-2424-1_15

84. Bazié HR, Bayala J, Zombré G, Sanou J, Ilstedt U. Separating competition-related factors limiting crop performance in an agroforestry parkland system in Burkina Faso. Agroforest Syst. 1 mars 2012;84(3):377-88.

85. CABI. Vitellaria paradoxa (shea tree). CABI Compendium. 21 nov 2019;CABI Compendium:56510.

86. Chevalier A. Nouvelles recherches sur l'Arbre à beurre du Soudan Butyrospermum Parkii. 1948 [cité 31 janv 2025]; Disponible sur: https://www.persee.fr/doc/jatba_0370-5412_1948_num_28_307_6157

87. Bremer B, Bremer K, Chase MW, Fay MF, Reveal JL, Soltis DE, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society. 1 oct 2009;161(2):105-21.

88. Zhang J, Kurita M, Shinozaki T, Ukiya M, Yasukawa K, Shimizu N, et al. Triterpene glycosides and other polar constituents of shea (*Vitellaria paradoxa*) kernels and their bioactivities. Phytochemistry. 1 déc 2014;108:157-70.

89. Baky MH, Kamal AM, Elgindi MR, Haggag EG. A Review on Phenolic Compounds from Family Sapotaceae. J Pharmacogn Phytochem. 2016;5(2):280-7.

90. Choungo Nguekeng PB, Hendre P, Tchoundjeu Z, Kalousová M, Tchanou Tchapda AV, Kyereh D, et al. 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. déc 2021;12(12):1740.

91. Maranz S, Kpikpi W, Wiesman Z, De Saint Sauveur A, Chapagain B. Nutritional values and indigenous preferences for Shea Fruits(vitellaria paradoxa C.F. Gaertn. F.) in African Agroforestry Parklands. Econ Bot. 1 déc 2004;58(4):588-600.

92. Orwa C, Mutua A, Kindt R, Jamnadass R, Simons A. Agroforestree Database: a tree reference and selection guide version 4.0. World Agroforestry Centre, Kenya. 1 janv 2009;

93. Jimoh AA, Aina ST. Investigation of the physical and mechanical properties of Shea Tree timber (Vitellaria paradoxa) used for structural applications in Kwara State, Nigeria. Journal of Applied Sciences and Environmental Management. 29 nov 2017;21(5):961-5.

94. Yao SDM, Diarrassouba N, Attikora A, Fofana IJ, Dago DN, 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. International Journal of Genetics and Molecular Biology. 3 juin 2020;12:1-10.

95. Oyen LPA, Lemmens RHMJ. Ressources vegetales de l'Afrique tropicale : precurseur [Internet]. PROTA; 2002 [cité 31 oct 2024]. Disponible sur: https://research.wur.nl/en/publications/ressources-vegetales-de-lafrique-tropicale-precurseur

96. Delaney A, Dembele A, Nombré I, Gnane Lirasse F, Marshall E, Nana A, et al. Local-scale tree and shrub diversity improves pollination services to shea trees in tropical West African parklands. Journal of Applied Ecology. 2020;57(8):1504-13.

97. Delolme A. Etude du Karité à la station agricole de Ferkessedougou. Oléagineux, 4 : 186-200. Oléagineux. 4e éd. 1947;186-200.

98. Guira M. Contribution à l'étude de la Biologie florale du karité,Butyrospèremum paradoxum SUP SP.parkII (G.DON)hepper, dans la reregion de Ouagadougou,Burkina Faso. Université de Ouagadougou; 1989.

99. Diarrassouba N, Nde Bup D, Kapseu C, Kouame C, Sangaré A. Typology of shea trees (Vitellaria paradoxa C.F. Gaertn.) using morphological traits in Côte d'Ivoire. Geneconserve. 8 (33) : 752 – 780.; 2009.

100. Tom-Dery D, Eller F, Reisdorff C, Jensen K. Shea (Vitellaria paradoxa C. F. Gaertn.) at the crossroads: current knowledge and research gaps. Agroforest Syst. 1 oct 2018;92(5):1353-71.

101. Aubreville A. Flore Forestière Soudano-Guinéenne. A.O.F.- CAMEROUN A. E.F, . 523 p.; 1950.

102. Adam A, Acheampong A, Abdul-Mumeen I. Effect of soil variation on quality of shea butter in selected areas of the northern region of Ghana. JABSD. 18 juill 2015;7(5):43-50.

103. Salako1 G, Sawyerr1 H, Bashir2 A, Adebayo2 A, Adio3 A. Deductive and Multi-criteria Approach to Ecosystem Modeling and Habitat Mapping of Shea Butter Trees (Vitellaria Paradoxa) in the Tropical Savanna. International Journal of Environment, Agriculture and Biotechnology [Internet]. 16 déc 2017 [cité 3 févr 2025];2(6). Disponible sur: https://ijeab.com/detail/deductive-and-multi-criteria-approach-to-ecosystem-modeling-and-habitat-mapping-of-shea-butter-trees-vitellaria-paradoxa-in-the-tropical-savanna/

104. Attikora AJP. Amélioration génétique du karité (Vitellaria paradoxa C.F. Gaertn) en Côte d'Ivoire : Analyse de la diversité morphologique des arbres élites identifiés dans les régions de la Bagoué et du Tchologo. 23 août 2018 [cité 25 sept 2024]; Disponible sur: https://matheo.uliege.be/handle/2268.2/6056

105. Ouoba HY, Bastide B, Kaboré SA, Seghieri J, Boussim IJ. Structure des populations de Vitellaria paradoxa C.F. Gaertn. (karité) dans les parcs agroforestiers au Burkina Faso. Biotechnol Agron Soc Environ [Internet]. 1 janv 2023 [cité 1 nov 2024]; Disponible sur: https://popups.uliege.be/1780-4507/index.php?id=20329

106. Serpanité G. La production de karité (Butyrospermum paradoxum Gaertn. f. Hepper) des parcs arborés de l'Ouest Burkina Faso : effets de différents modes de gestion- fdi:010010093- Horizon [Internet]. 1996 [cité 3 févr 2025]. Disponible sur: https://www.documentation.ird.fr/hor/fdi:010010093

107. McKey D, Emperalre L, Élias M, Pinton F, Robert T, Desmoulière S, et al. Gestions locales et dynamiques régionales de la diversité variétale du manioc en Amazonie. Genetics, Selection, Evolution : GSE. 1 déc 2001;33(Suppl 1):S465.

108. Mkwezalamba I, Munthali CRY, Missanjo E. Phenotypic Variation in Fruit Morphology among Provenances of Sclerocarya birrea (A. Rich.) Hochst. International Journal of Forestry Research. 2015;2015(1):735418.

109. Sudrajat DJ. Genetic variation of fruit, seed, and seedling characteristics among 11 populations of white jabon in Indonesia. Forest Science and Technology. 2 janv 2016;12(1):9-15.

110. Oluoch P, Nyaboga EN, Bargul JL. Analysis of genetic diversity of passion fruit (*Passiflora edulis* Sims) genotypes grown in Kenya by sequence-related amplified polymorphism (SRAP) markers. Annals of Agrarian Science. 1 déc 2018;16(4):367-75.

111. Ramanatha Rao V, Hodgkin T. Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell, Tissue and Organ Culture. 1 janv 2002;68(1):1-19.

112. Lassois L, Denancé C, Ravon E, Guyader A, Guisnel R, Hibrand-Saint-Oyant L, et al. Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers. Plant Mol Biol Rep. 1 août 2016;34(4):827-44.

113. Diarrassouba N, N'Guessan A, Kouablan K, Abdourahamane S. Evaluation des performances de quelques descripteurs quantitatifs et leur utilisation dans la structuration de la population d'un parc naturel de karite en Côte d'Ivoire. Plant Genetics Ressources Newsletter. No. 152. 2007;65-72.

114. Yao SDM, Konan KL, Sie R, Diarrassouba N, Lekadou T, Koffi EBZ, et al. Fiabilité d'une liste minimale de descripteurs agromorphologiques recommandée par le COGENT dans l'étude de la diversité génétique du cocotier (Cocos nucifera L.). Journal of Animal and Plant Sciences. 31 août 2015;26:4006-22.

115. Karambiri M, Elias M, Vinceti B, Grosse A. Exploring local knowledge and preferences for shea (*Vitellaria paradoxa*) ethnovarieties in Southwest Burkina Faso through a gender and ethnic lens. Forests, Trees and Livelihoods. 2 janv 2017;26(1):13-28.

116. Gwali S, Okullo JBL, Eilu G, Nakabonge G, Nyeko P, Vuzi P. Folk Classification of Shea Butter Tree (Vitellaria paradoxa subsp. nilotica) Ethno-varieties in Uganda. Ethnobotany Research and Applications. 1 juill 2011;9:243-56.

117. Djékota CN. Etude de la variabilité morphologique et taxonomie des morphotypes chez le Karité (Vitellaria paradoxa C.F. Gaertn sous-espèce paradoxa, Sapotaceae) dans la région du Mandoul au Tchad. [Sénégal]: Cheikh Anta Diop; 2014.

118. Diarrassouba N, Nde Bup D, Kapseu C, Kouame C, Sangaré A. Phenotypic Diversity of Shea (Vitellaria ParadoxaC.F. Gaertn.) Populations across Four Agro-Ecological Zones of Cameroon. Journal of Crop Scienceand Biotechnolog. 1 janv 2008;10:211-8.

119. Soro D, Traore K, Kassi NJ. Variabilité des caractères morphologiques chez le karité (Vitellaria paradoxa), dans le Nord de la Côte d'Ivoire. International Journal of Biological and Chemical Sciences [Internet]. 2011 [cité 22 nov 2022];5(3). Disponible sur: https://www.ajol.info/index.php/ijbcs/article/view/72263

120. Najimi B, El Jaafari S, Jlibène M, Jacquemin JM. Applications des marqueurs moléculaires dans l'amélioration du blé tendre pour la résistance aux maladies

et aux insectes. Biotechnol Agron Soc Environ [Internet]. 1 janv 2003 [cité 5 nov 2024]; Disponible sur: https://popups.uliege.be/1780-4507/index.php?id=14450

121. Abdou R, Bakasso Y, Adam T, Saadou M, Baudoin JP. Biologie, diversité et outils pour l'analyse de la diversité génétique de l'oignon, Allium cepa L. (synthèse bibliographique). Biotechnol Agron Soc Environ [Internet]. 1 janv 2015 [cité 5 nov 2024]; Disponible sur: https://popups.uliege.be/1780-4507/index.php?id=12086

122. Attikora AJP, Yao SDM, Dago DN, Silué S, De Clerck C, Kwibuka Y, et al. 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 Biology. 1 oct 2024;24(1):913.

123. Kelly B, Hardy O, Bouvet JM. Temporal and spatial genetic structure in Vitellaria paradoxa (shea tree) in an agroforestry system in southern Mali. Molecular Ecology. 2004;13(5):1231-40.

124. Logossa ZA, Camus-Kulandaivelu L, Allal F, Vaillant A, Sanou H, Kokou K, et al. Molecular data reveal isolation by distance and past population expansion for the shea tree (Vitellaria paradoxa C.F. Gaertn) in West Africa. Molecular Ecology. 2011;20(19):4009-27.

125. Bouvet JM, Fontaine C, Sanou H, Cardi C. An analysis of the pattern of genetic variation in Vitellaria paradoxa using RAPD markers. Agroforestry Systems. 1 janv 2004;60(1):61-9.

126. Anyomi WE, Barnor MT, Eleblu JSY, Danquah A, Avicor SW, Ofori K, et al. Elucidation of the Genetic Diversity within Some In Situ Shea Germplasm in Ghana. Agronomy. sept 2023;13(9):2256.

127. Lovett PN. 5 - Shea butter: Properties and processing for use in food. In: Talbot G, éditeur. Specialty Oils and Fats in Food and Nutrition [Internet]. Woodhead Publishing; 2015 [cité 1 août 2024]. p. 125-58. (Woodhead Publishing Series in Food Science, Technology and Nutrition). Disponible sur: https://www.sciencedirect.com/science/article/pii/B9781782423768000053

128. Goumbri BWF, da Silva TLT, Marini RD, Semdé R, Somé TI, Danthine S. African Shea Butter Properties Related to Common Extraction Technologies: A Review. Food Bioprocess Technol. 1 févr 2022;15(2):231-48.

129. Abdul-Hammed M, Jaji AO, Adegboyega SA. Comparative studies of thermophysical and physicochemical properties of shea butter prepared from

cold press and solvent extraction methods. Journal of King Saud University - Science. 1 juin 2020;32(4):2343-8.

130. Honfo FG, Akissoe N, Linnemann AR, Soumanou M, Van Boekel MAJS. Nutritional Composition of Shea Products and Chemical Properties of Shea Butter: A Review. Critical Reviews in Food Science and Nutrition. 1 janv 2014;54(5):673-86.

131. Badu M, Awudza J, Budd PM, Yeates S. Determination of Physical Properties and Crystallization Kinetics of Oil From Allanblackia Seeds and Shea Nuts Under Different Thermal Conditions. European Journal of Lipid Science and Technology. 2018;120(3):1700156.

132. Akihisa T, Kojima N, Katoh N, Kikuchi T, Fukatsu M, Shimizu N, et al. Triacylglycerol and Triterpene Ester Composition of Shea Nuts from Seven African Countries. Journal of Oleo Science. 2011;60(8):385-91.

133. Allal F, Piombo G, Kelly BA, Okullo JBL, Thiam M, Diallo OB, et al. Fatty acid and tocopherol patterns of variation within the natural range of the shea tree (Vitellaria paradoxa). Agroforest Syst. 1 oct 2013;87(5):1065-82.

134. Maranz S, Wiesman Z, Bisgaard J, Bianchi G. Germplasm resources of Vitellaria paradoxa based on variations in fat composition across the species distribution range. Agroforestry Systems. 1 janv 2004;60(1):71-6.

135. Goumbri BWF, Kouassi AK, Djang'eing'a RM, Semdé R, Mouithys-Mickalad A, Sakira AK, et al. Quality Characteristics and Thermal Behavior Diversity of Traditional Crude Shea (Vitellaria paradoxa Gaertn) Butter from Burkina Faso. Food Biophysics. 1 sept 2024;19(3):609-26.

136. Abdul-Mumeen I, Beauty D, Adam A. Shea butter extraction technologies: Current status and future perspective. AJBR. 28 févr 2019;13(2):9-22.

137. Ofoegbu-Chibuzo NE, Chukwu UJ, Okoye IP. Physicochemical Analysis and Fatty Acid Content of Chemical and Traditional Extracts of Shea Kernel (Vitellaria paradoxa) from Kwara State Nigeria. Open Access Library Journal. 4 janv 2022;9(1):1-11.

138. Olaniyan AM, Oje K, Olaniyan AM, Oje K. Quality characteristics of shea butter recovered from shea kernel through dry extraction process. Journal of Food Science and Technology (Mysore). 1 janv 2007;(Mysore):404-7.

139. Kodji L, Soudy ID, Mbailao M, Togdjim T, Trebo SA, Beyamra C. Assessment of the Physicochemical Quality of Unrefined Shea Butter Produced in the

Localities of Sarh and Koumra (Chad). Food and Nutrition Sciences. 26 juin 2023;14(6):561-77.

140. Yao AL, Adingra KMD, N'gbo MLK, Tano K. Impact of Different Extraction Processes on the Quality of Shea Butter (Vitellaria paradoxa) Produced by Women in the Northern Regions of Côte d'Ivoire. Biotechnology Journal International. 30 juill 2024;28(4):142-51.

141. Ræbild A, Hansen UB, Kambou S. Regeneration of Vitellaria paradoxa and Parkia biglobosa in a parkland in Southern Burkina Faso. Agroforest Syst. 1 juill 2012;85(3):443-53.

142. Ouoba H, Bastide B, Coulibaly-Lingani P, Kaboré A, Paulin O, Sylvie Y, et al. Régénération assistée du karité (Vitellaria paradoxa C. F. Gaertn.) dans les parcs agroforestiers au Burkina Faso. European Scientific Journal. 31 déc 2020;16:23.

143. Salle G, Boussim J, Raynal-Roques A, Brunck F. Le karité une richesse potentielle. perspectives de recherche pour améliorer sa production. Bois et forêts des tropiques. 1991;(228):11-23.

144. Boussim JI, Sallé G, Guinko S. Tapinanthus parasite du Karité au Burkina Faso. 1ère partie identification et distribution. BOIS & FORETS DES TROPIQUES. 1 déc 1993;238:45-52.

145. Odebiyi JA, Bada SO, Awodoyin R, Oni PI, Omoloye AA. Population structure of Vitelaria paradoxa Gaertn. F. and Parkia biglobosa(Jacq.) Benth.in the Agroforestry Parklands of Nigerian humid Savanna. West African Journal of Applied Ecology [Internet]. 2004 [cité 29 janv 2025];5(1). Disponible sur: https://www.ajol.info/index.php/wajae/article/view/45597

146. Gaisberger H, Kindt R, Loo J, Schmidt M, Bognounou F, Da SS, et al. Spatially explicit multi-threat assessment of food tree species in Burkina Faso: A fine-scale approach. PLOS ONE. 7 sept 2017;12(9):e0184457.

147. Ouédraogo P, Bationo BA, Sanou J, Traoré S, Barry S, Dayamba SD, et al. Uses and vulnerability of ligneous species exploited by local population of northern Burkina Faso in their adaptation strategies to changing environments. Agriculture & Food Security. 1 mai 2017;6(1):1-16.

148. Lovett P, Phillips L. Agroforestry Shea Parklands of Sub-Saharan Africa: Threats and Solutions. 2018.

149. Kaboré SA, Bastide B, Traoré S, Boussim JI. Dynamique du karité, Vitellaria paradoxa, dans les systèmes agraires du Burkina Faso. Bois for trop. 1 sept 2012;313(313):47.

150. Coulibaly-Lingani P. Vulnérabilité des Produits Forestiers non Ligneux « Vitellaria Paradoxa » et « Parkia Biglobosa » au Burkina Faso - Une Évaluation des Perceptions de la Population Locale Basée sur le Genre et l'Ethnicité. European Scientific Journal, ESJ. 31 janv 2023;19(2):164-164.

151. Azongnide GG, Christine O, Valerien Z, Sourou B. Scientific Knowledge on Vitellaria paradoxa, A Vulnerable Species: State of Research Works and Perceptive. 1 janv 2021;9:39-51.

152. KONE M. Vitellaria paradoxa: aires de distribution, paysages et services écosystémiques dans la région du Poro, Côte d'Ivoire. Regardsuds [Internet]. 19 janv 2022 [cité 26 nov 2024];Nous publions un volume par an(2021 Numéro 2). Disponible sur: https://regardsuds.org/vitellaria-paradoxa-aires-de-distribution-paysages-et-services-ecosystemiques-dans-la-region-du-poro-cote-divoire/

153. Diarrassouba N, Yao SDM, Alui KA, Ble K. Comment restaurer les parcs à karité dégradés par la technique de Régénération Naturelle Assistée (RNA) ? 2020.

154. Lawali S, Diouf A, Morou B, Kona KA, Saidou L, Guero C, et al. Régénération Naturelle Assistée (RNA) : outil d'adaptation et résilience des ménages ruraux d'Aguié au Niger. International Journal of Biological and Chemical Sciences. 2018;12(1):75-89.

155. Dianda M, Ouédraogo S, Diop T. Variation de la réponse à l'endomycorhization en pépinière des plants issus de deux semenciers de karité au Burkina Faso. International Journal of Biological and Chemical Sciences. 15 avr 2011;4.

156. Boffa JMJ. Productivity and management of agroforestry parklands in the Sudan zone of Burkina Faso, west Africa. Theses and Dissertations Available from ProQuest. 1 janv 1995;1-101.

157. Boffa JM, Taonda SJB, Dickey JB, Knudson DM. Field-scale influence of karité (Vitellaria paradoxa) on sorghum production in the Sudan zone of Burkina Faso. Agroforestry Systems. 1 juill 2000;49(2):153-75.

158. Lovett PN, Haq N. Evidence for anthropic selection of the Sheanut tree (Vitellaria paradoxa). Agroforestry Systems. 1 mai 2000;48(3):273-88.

159. Maanikuu PMI, Peker K. Medicinal and Nutritional Benefits from the Shea Tree- (Vitellaria Paradoxa). Journal of Biology, Agriculture and Healthcare. 2017;7(22):51.

160. Moore S. The role of Vitellaria Paradoxa in poverty reduction and food security in the Upper East region of Ghana. 1 janv 2008;3.

161. Masters ET. Traditional food plants of the upper Aswa River catchment of northern Uganda—a cultural crossroads. Journal of Ethnobiology and Ethnomedicine. 6 avr 2021;17(1):24.

162. J. S. L. G, Dalziel JM. The Useful Plants of West Tropical Africa. The Geographical Journal. 1 mars 1938;91:279.

163. Talbot G, Slager H. Cocoa butter equivalents and improvers - Their use in chocolate and chocolate-coated confectionery. Agro Food Industry Hi-Tech. 1 mai 2008;19:28-9.

164. Jahurul MHA, Zaidul ISM, Norulaini NAN, Sahena F, Jinap S, Azmir J, et al. Cocoa butter fats and possibilities of substitution in food products concerning cocoa varieties, alternative sources, extraction methods, composition, and characteristics. Journal of Food Engineering. 1 août 2013;117(4):467-76.

165. Nahm HS, Juliani HR, Simon JE. Quality Characteristics of Shea Butter, Vitellaria paradoxa. In: African Natural Plant Products Volume II: Discoveries and Challenges in Chemistry, Health, and Nutrition [Internet]. American Chemical Society; 2013 [cité 14 nov 2024]. p. 167-84. (ACS Symposium Series; vol. 1127). Disponible sur: https://doi.org/10.1021/bk-2013-1127.ch012

166. Di Vincenzo D, Maranz S, Serraiocco A, Vito R, Wiesman Z, Bianchi G. Regional Variation in Shea Butter Lipid and Triterpene Composition in Four African Countries. J Agric Food Chem. 1 sept 2005;53(19):7473-9.

167. Glew D, Lovett PN. Life cycle analysis of shea butter use in cosmetics: from parklands to product, low carbon opportunities. Journal of Cleaner Production. 1 avr 2014;68:73-80.

168. Wardell DA, Elias M, Zida M, Tapsoba A, Rousseau K, Gautier D, et al. Shea (Vitellaria paradoxa C. F. Gaertn.) – a peripheral empire commodity in French West Africa, 1894–1960. International Forestry Review. 1 déc 2021;23(4):511-33.

169. Wardell DA, Tapsoba A, Lovett PN, Zida M, Rousseau K, Gautier D, et al. Shea (Vitellaria paradoxa C. F. Gaertn.) – the emergence of global production

networks in Burkina Faso, 1960–2021 1. International Forestry Review. 1 déc 2021;23(4):534-61.

170. Jasaw GS, Saito O, Gasparatos A, Shoyama K, Boafo YA, Takeuchi K. Ecosystem services trade-offs from high fuelwood use for traditional shea butter processing in semi-arid Ghana. Ecosystem Services. 1 oct 2017;27:127-38.

171. GEPA Exporters Portal [Internet]. 2014. GEPA - Ghana Export Promotion Authority.

172. Bello-Bravo J, Lovett PN, Pittendrigh BR. The Evolution of Shea Butter's « Paradox of paradoxa » and the Potential Opportunity for Information and Communication Technology (ICT) to Improve Quality, Market Access and Women's Livelihoods across Rural Africa. Sustainability. mai 2015;7(5):5752-72.

173. Schreckenberg K, Awono A, Degrande A, Mbosso C, Ndoye O, Tchoundjeu Z. Domesticating Indigenous Fruit Trees as a Contribution to Poverty Reduction. Forests, Trees and Livelihoods. 1 janv 2006;16(1):35-51.

174. Brou YT. Climat, mutations socio-économiques et paysages en Côte d'Ivoire. [Université des Sciences et Technologies de LILLE, (2005) 212.]: Université Lille1 - Sciences et Technologies; 2005.

175. N'Guessan KA, Diarrassouba N, Alui KA, Nangha KY, Fofana IJ, Yao-Kouame A. Indicateurs de dégradation physique des sols dans le Nord de la Côte d'Ivoire: Cas de Boundiali et Ferkessédougou. Afrique Science: Revue Internationale des Sciences et Technologie. 28 juill 2015;11(3):115-28.

176. IPGRI, INIA. Descriptors for Shea tree (Vitellaria paradoxa). International Plant Genetic Resources Institute. Rome, Italy; Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Madrid, Spain.ISBN-13: 978-92-9043-687-4 ISBN-10: 92-9043-687-5; 2006.

177. Saidou A. Effet d'un système agroforestier à karité (Vitellaria paradoxa c.f. gaertn) sur le sol et le potentiel de production du maïs (Zea maize) en zone Soudanienne du Bénin. International Journal of Biological and Chemical Sciences. 1 janv 2013;

178. Anyomi WE, Barnor MT, Danquah A, Ofori K, Padi FK, Avicor SW, et al. Heritability and Genetic Advance Estimates of Key Shea Fruit Traits. Agronomy. mars 2023;13(3):640.

179. Azo'O JRN, Tchatat M, Mony R, SD D. Parasitisme et ethnobotanique des Loranthaceae à Lokomo (Est-Cameroun). Journal of Animal and Plant Sciences. 2 oct 2013;19:29223-2932.

180. Cubero JI. Conservation through plant breeding and in-situ conservation. Bocconea 7: 207-2121997. 1997;

181. Nguyen NN, Kim M, Jung JK, Shim EJ, Chung SM, Park Y, et al. Genome-wide SNP discovery and core marker sets for assessment of genetic variations in cultivated pumpkin (Cucurbita spp.). Hortic Res. 1 août 2020;7(1):1-10.

182. Abbasi Kejani A, Hosseini Tafreshi SA, Khayyam Nekouei SM, Mofid MR. Efficient isolation of high quality nucleic acids from different tissues of Taxus baccata L. Mol Biol Rep. 1 févr 2010;37(2):797-800.

183. Nagori R, Sharma P, Habibi N, Purohit SD. An Efficient Genomic DNA Extraction Protocol for Molecular Analysis in Annona reticulata. Natl Acad Sci Lett. 1 avr 2014;37(2):137-40.

184. Aboul-Maaty NAF, Oraby HAS. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. Bull Natl Res Cent. 12 févr 2019;43(1):25.

185. Rezadoost MH, Kordrostami M, Kumleh HH. An efficient protocol for isolation of inhibitor-free nucleic acids even from recalcitrant plants. 3 Biotech. 13 févr 2016;6(1):61.

186. Ahmadi E, Kowsari M, Azadfar D, Salehi Jouzani G. Rapid and economical protocols for genomic and metagenomic DNA extraction from oak (Quercus brantii Lindl.). Annals of Forest Science. 28 mars 2018;75(2):43.

187. Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, et al. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (Coffea sp.). Theor Appl Genet. 1 sept 2000;101(4):669-76.

188. Varma A, Padh H, Shrivastava N. Plant genomic DNA isolation: An art or a science. Biotechnology Journal. 2007;2(3):386-92.

189. Allal F, Vaillant A, Sanou H, Kelly B, Bouvet JM. Isolation and characterization of new microsatellite markers in shea tree (Vitellaria paradoxa C. F. Gaertn). Molecular Ecology Resources. 2008;8(4):822-4.

190. Mohammed HI, Mohammed Z, Warra A, Abdulrahman Y, Sabo I, Ibrahim G, et al. Physicochemical and Genetic Diversity Studies of Vitellaria paradoxa in

Northern Nigeria. Journal of Current Biomedical Research. 28 févr 2022;2(1):19-37.

191. Doyle JJ, Doyle JL, éditeurs. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin. 1987;

192. Youssef M, Valdez-Ojeda R, Ku-Cauich J, EscobedoGraciaMedrano R. Enhanced Protocol for Isolation of Plant Genomic DNA. Journal of Agriculture and Environmental Sciences. 1 déc 2015;4:172-80.

193. Stulnig TM, Amberger A. Exposing contaminating phenol in nucleic acid preparations. Biotechniques. mars 1994;16(3):402-4.

194. Scobeyeva VA, Omelchenko DO, Dyakov LM, Konovalov AS, Speranskaya AS, Krinitsina AA. Comparison of Some Plant DNA Extraction Methods. Russ J Genet. 1 mai 2018;54(5):576-86.

195. Schmiderer C, Lukas B, Novak J. Effect of different DNA extraction methods and DNA dilutions on the amplification success in the PCR of different medicinal and aromatic plants. Zeitschrift für Arznei- & Gewürzpflanzen. 2013;18(2):65-72.

196. Henry RJ. Plant DNA extraction. Plant genotyping: the DNA fingerprinting of plants. janv 2001;239-49.

197. Kang TJ, Yang MS. Rapid and reliable extraction of genomic DNA from various wild-type and transgenic plants. BMC Biotechnol. 2 sept 2004;4(1):20.

198. Souza DC de, Teixeira TA. A simple and effective method to obtain high DNA quality and quantity from Cerrado plant species. Mol Biol Rep. août 2019;46(4):4611-5.

199. Pandey S, Alam A, Chakraborty D, Sharma V. An Improved Protocol for Genomic DNA Isolation from Bryophyte Species. Proc Natl Acad Sci, India, Sect B Biol Sci. 1 sept 2019;89(3):823-31.

200. Li H, Li J, Cong XH, Duan YB, Li L, Wei PC, et al. A high-throughput, high-quality plant genomic DNA extraction protocol. Genet Mol Res. 1 oct 2013;12(4):4526-39.

201. Sambrook J, Russell DW. The condensed protocols from molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2006. V, 800 S.

202. Lade BD, Patil AS, Paikrao HM. Efficient genomic DNA extraction protocol from medicinal rich Passiflora foetida containing high level of polysaccharide and polyphenol. SpringerPlus. 23 août 2014;3(1):457.

203. Neumann K, Kahlheber S, Uebel D. Remains of woody plants from Saouga, a medieval west African village. Veget Hist Archaebot. 1 juin 1998;7(2):57-77.

204. Masters ET, Yidana JA, Lovett PN. Reinforcing sound management through trade: Shea tree products in Africa. Unasylva. 2004;55(219):46-52.

205. Attikora AJP, Diarrassouba N, Yao SDM, Clerck CD, Silue S, Alabi T, et al. Morphological traits and sustainability of plus shea trees (Vitellaria paradoxa C.F.Gaertn.) in Côte d'Ivoire. Biotechnologie, Agronomie, Société et Environnement [Internet]. 25 sept 2023 [cité 19 oct 2023]; Disponible sur: https://orbi.uliege.be/handle/2268/307173

206. Luo Z, Brock J, Dyer JM, Kutchan T, Schachtman D, Augustin M, et al. Genetic Diversity and Population Structure of a Camelina sativa Spring Panel. Frontiers in Plant Science [Internet]. 2019 [cité 19 oct 2023];10. Disponible sur: https://www.frontiersin.org/articles/10.3389/fpls.2019.00184

207. Attikora AJP, Silué S, Yao SDM, De Clerck C, Shumbe L, Diarrassouba N, et al. An innovative optimized protocol for high-quality genomic DNA extraction from recalcitrant Shea tree (Vitellaria paradoxa, C.F. Gaertn) plant and its suitability for downstream applications. Mol Biol Rep. 22 janv 2024;51(1):171.

208. Brown AHD. Core collections: a practical approach to genetic resources management. Genome. 15 janv 1989;31(2):818-24.

209. Frankel OH. Plant genetic resources today: a critical appraisal. Crop genetic resources: conservation and evaluation. 1984;241-56.

210. Tchokponhoué DA, Achigan-Dako EG, N'Danikou S, Nyadanu D, Kahane R, Houéto J, et al. Phenotypic variation, functional traits repeatability and core collection inference in Synsepalum dulcificum (Schumach & Thonn.) Daniell reveals the Dahomey Gap as a centre of diversity. Sci Rep. 11 nov 2020;10(1):19538.

211. Dekoula CS, Kouame B, N'goran EK, Yao FG, Ehounou JN, Soro N. Impact De La Variabilité Pluviométrique Sur La Saison Culturale Dans La Zone De Production Cotonnière En Côte d'Ivoire. European Scientific Journal, ESJ. 30 avr 2018;14(12):143-143.

212. Kouamé B, Ehounou JN, Kassin KE, Dekoula CS, Yao GF, N'goran EK, et al. Caractérisation Des Paramètres Agroclimatiques Clés De La Saison Culturale En Zone De Contact ForêtSavane De Côte-d'Ivoire. European Scientific Journal, ESJ. 31 déc 2018;14(36):243-243.

213. Koffi AA, Kouassi FA, N'Goran SBK, Soro D. Les Loranthaceae, parasites des arbres et arbustes : cas du département de Katiola, au nord de la Côte d'Ivoire. International Journal of Biological and Chemical Sciences. 2014;8(6):2552-9.

214. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Mol Biol. 2012;888:67-89.

215. Ketema S, Tesfaye B, Keneni G, Fenta BA, Assefa E, Greliche N, et al. DArTSeq SNP-based markers revealed high genetic diversity and structured population in Ethiopian cowpea [Vigna unguiculata (L.) Walp] germplasms. PLOS ONE. 8 oct 2020;15(10):e0239122.

216. Perrier X, Jacquemoud-Collet JP. DARwin software http://DARwin. cirad. fr. 2006.

217. Mantello CC, Cardoso-Silva CB, Silva CC da, Souza LM de, Junior EJS, Gonçalves P de S, et al. De Novo Assembly and Transcriptome Analysis of the Rubber Tree (Hevea brasiliensis) and SNP Markers Development for Rubber Biosynthesis Pathways. PLOS ONE. 21 juill 2014;9(7):e102665.

218. Fufa TW, Abtew WG, Amadi CO, Oselebe HO. DArTSeq SNP-based genetic diversity and population structure studies among taro [(Colocasia esculenta (L.) Schott] accessions sourced from Nigeria and Vanuatu. PLOS ONE. 10 nov 2022;17(11):e0269302.

219. Peringottillam M, Kunhiraman Vasumathy S, Selvakumar HKK, Alagu M. Genetic diversity and population structure of rice (Oryza sativa L.) landraces from Kerala, India analyzed through genotyping-by-sequencing. Mol Genet Genomics. 1 janv 2022;297(1):169-82.

220. Guo C, McDowell IC, Nodzenski M, Scholtens DM, Allen AS, Lowe WL, et al. Transversions have larger regulatory effects than transitions. BMC Genomics. 19 mai 2017;18(1):394.

221. Chen W, Hou L, Zhang Z, Pang X, Li Y. Genetic Diversity, Population Structure, and Linkage Disequilibrium of a Core Collection of Ziziphus jujuba Assessed with Genome-wide SNPs Developed by Genotyping-by-sequencing and SSR Markers. Front Plant Sci. 2017;8:575.

222. Eltaher S, Sallam A, Belamkar V, Emara HA, Nower AA, Salem KFM, et al. Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. Frontiers in Genetics [Internet]. 2018 [cité 19 oct 2023];9. Disponible sur: https://www.frontiersin.org/articles/10.3389/fgene.2018.00076

223. Frankham R, Ballou JD, Briscoe DA. Higher Education from Cambridge University Press. Cambridge University Press; 2010 [cité 19 oct 2023]. Introduction to Conservation Genetics. Disponible sur: https://www.cambridge.org/highereducation/books/introduction-to-conservation-genetics/696B4E558C93F7FBF9C33D6358EA7425

224. Odoi JB, Adjei EA, Hendre P, Nantongo JS, Ozimati AA, Badji A, et al. Genetic diversity and population structure among Ugandan shea tree (Vitellaria paradoxa subsp. nilotica) accessions based on DarTSeq markers. Crop Science. 2023;63(4):2297-309.

225. Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, et al. PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics. 15 août 2007;23(16):2155-62.

226. De Beukelaer H, Davenport GF, Fack V. Core Hunter 3: flexible core subset selection. BMC Bioinformatics. 31 mai 2018;19(1):203.

227. Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA. Statistical Techniques for Defining Reference Sets of Accessions and Microsatellite Markers. Crop Science. 2011;51(6):2401-11.

228. Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet. 1 févr 2013;126(2):289-305.

229. Bai Q, Cai Y, He B, Liu W, Pan Q, Zhang Q. Core set construction and association analysis of Pinus massoniana from Guangdong province in southern China using SLAF-seq. Sci Rep. 11 sept 2019;9(1):13157.

230. Gan ST, Teo CJ, Manirasa S, Wong WC, Wong CK. Assessment of genetic diversity and population structure of oil palm (Elaeis guineensis Jacq.) field genebank: A step towards molecular-assisted germplasm conservation. PLOS ONE. 29 juill 2021;16(7):e0255418.

231. Ithnin M, Teh CK, Ratnam W. Genetic diversity of Elaeis oleifera (HBK) Cortes populations using cross species SSRs: implication's for germplasm utilization and conservation. BMC Genetics. 19 avr 2017;18(1):37.

232. Bai Q, He B, Cai Y, Lian H, Zhang Q, Liang D, et al. Genetic Diversity and Population Structure of Schima superba From Southern China. Frontiers in Ecology and Evolution [Internet]. 2022 [cité 19 oct 2023];10. Disponible sur: https://www.frontiersin.org/articles/10.3389/fevo.2022.879512

233. Yonezawa K, Nomura T, Morishima H, Nonura T. Sampling strategies for use in stratified germplasm collections. In 1995 [cité 19 oct 2023]. Disponible sur: https://www.semanticscholar.org/paper/Sampling-strategies-for-use-in-stratified-germplasm-Yonezawa-Nomura/d24d31793c09f5c9044085d6ebab1bc924a36d1a

234. Hintum TJL van, Brown AHD, Spillane C. Core Collections of Plant Genetic Resources. Bioversity International; 2000. 51 p.

235. Agarwal M, Shrivastava N, Padh H. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. avr 2008;27(4):617-31.

236. Schlottfeldt S, Walter MEMT, de Carvalho ACPLF, Soares TN, Telles MPC, Loyola RD, et al. Multi-objective optimization for plant germplasm collection conservation of genetic resources based on molecular variability. Tree Genetics & Genomes. 5 févr 2015;11(2):16.

237. Bondé L, Ouédraogo O, Ouédraogo I, Thiombiano A, Boussim JI. Variability and estimating in fruiting of shea tree (Vitellaria paradoxa C.F. Gaertn) associated to climatic conditions in West Africa: implications for sustainable management and development. Plant Production Science. 3 avr 2019;22(2):143-58.

238. Ray J, Smith KW, Bhaggan K, Nagy ZK, Stapley AGF. Characterisation of high 1,3-distearoyl-2-oleoyl-sn-glycerol content stearins produced by acidolysis of high oleic sunflower oil with stearic and palmitic acids. European Journal of Lipid Science and Technology. 2014;116(5):532-47.

239. Attikora AJP, Silué S, Yao SDM, De Clerck C, Shumbe L, Diarrassouba N, et al. An innovative optimized protocol for high-quality genomic DNA extraction from recalcitrant Shea tree (Vitellaria paradoxa, C.F. Gaertn) plant and its suitability for downstream applications. Molecular Biology Reports. 22 janv 2024;51(1):171.

240. Kouassi AK, Alabi T, Cissé M, Purcaro G, Moret S, Moret E, et al. Assessment of composition, color, and oxidative stability of mango (Mangifera indica L.) kernel fats from various Ivorian varieties. Journal of the American Oil Chemists' Society. 2023;101(3):283-95.

241. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 1 juin 2000;155(2):945-59.

242. Perrier X, Bonnot AF and F. Methods of data analysis. In: Genetic Diversity of Cultivated Tropical Plants. CRC Press; 2003.

243. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 20 janv 2016;6(1):19444.

244. Tamba CL, Zhang YM. A fast mrMLM algorithm for multi-locus genome-wide association studies [Internet]. bioRxiv; 2018 [cité 26 juill 2024]. p. 341784. Disponible sur: https://www.biorxiv.org/content/10.1101/341784v1

245. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Brief Bioinform. 20 juill 2018;19(4):700-12.

246. Zhang J, Feng JY, Ni YL, Wen YJ, Niu Y, Tamba CL, et al. pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. Heredity. juin 2017;118(6):517-24.

247. Ren WL, Wen YJ, Dunwell JM, Zhang YM. pKWmEB: integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. Heredity. mars 2018;120(3):208-18.

248. Tamba CL, Ni YL, Zhang YM. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. PLOS Computational Biology. 31 janv 2017;13(1):e1005357.

249. Allal F, Sanou H, Millet L, Vaillant A, Camus-Kulandaivelu L, Logossa ZA, et al. Past climate changes explain the phylogeography of Vitellaria paradoxa over Africa. Heredity. août 2011;107(2):174-86.

250. Yang Y, Kong Q, Lim ARQ, Lu S, Zhao H, Guo L, et al. Transcriptional regulation of oil biosynthesis in seed plants: Current understanding, applications, and perspectives. Plant Communications. 12 sept 2022;3(5):100328.

251. Coban F, Ozer H, Lan Y. Genetic and environmental influences on fatty acid composition in different fenugreek genotypes. Industrial Crops and Products. 15 déc 2024;222:119774.

252. Davrieux F, Allal F, Piombo G, Kelly B, Okulo JB, Thiam M, et al. Near Infrared Spectroscopy for High-Throughput Characterization of Shea Tree (Vitellaria paradoxa) Nut Fat Profiles. J Agric Food Chem. 14 juill 2010;58(13):7811-9.

253. Ugese F, Baiyeri P, Mbah B. Agroecological variation in the fruits and nuts of shea butter tree (Vitellaria paradoxa C. F. Gaertn.) in Nigeria. Agroforestry Systems. 1 juin 2010;79:201-11.

254. Zhang JL, Zhang SB, Zhang YP, Kitajima K. Effects of phylogeny and climate on seed oil fatty acid composition across 747 plant species in China. Industrial Crops and Products. 1 janv 2015;63:1-8.

255. Zhang YM, Jia Z, Dunwell JM. Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. Front Plant Sci [Internet]. 11 févr 2019 [cité 2 août 2024];10. Disponible sur: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2019.00100/full

256. Peng Y, Liu H, Chen J, Shi T, Zhang C, Sun D, et al. Genome-Wide Association Studies of Free Amino Acid Levels by Six Multi-Locus Models in Bread Wheat. Front Plant Sci [Internet]. 14 août 2018 [cité 2 août 2024];9. Disponible sur: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.01196/full

257. Zhong H, Liu S, Sun T, Kong W, Deng X, Peng Z, et al. Multi-locus genome-wide association studies for five yield-related traits in rice. BMC Plant Biology. 10 août 2021;21(1):364.

258. Kouassi AK, Alabi T, Purcaro G, Blecker C, Danthine S. Assessment of the Impact of Annual Growing Conditions on the Physicochemical Properties of Mango Kernel Fat. Horticulturae. août 2024;10(8):814.

259. Elias M. Influence of agroforestry practices on the structure and spatiality of shea trees (Vitellaria paradoxa C.F. Gaertn.) in central-west Burkina Faso. Agroforest Syst. 1 févr 2013;87(1):203-16.

260. Ræbild A, Larsen AS, Jensen JS, Ouedraogo M, De Groote S, Van Damme P, et al. Advances in domestication of indigenous fruit trees in the West African Sahel. New Forests. 1 mai 2011;41(3):297-315.

261. Frimpong EB, Kpogoh PK, Akuoko S. Vegetation propagation of shea, kola and Allanblackia. Cocoa Research Institute, Ghana Annual Report. 1988;89:127-30.

262. Lovett PN, Azad AK, Paudyal K, Haq N. Genetic diversity and development of propagation techniques for tropical fruit trees. Domestication, Production and Utilisation of New Crops. 1996;286-7.

263. Singh M, Sonkusale S, Niratker C, Shukla P. Micropropagation of Shorea robusta: an economically important woody plant. Journal of Forest Science. 2014;60(2):70-4.

264. Amghar I, Ibriz M, Ibrahimi M, Boudra A, Gaboun F, Meziani R, et al. In Vitro Root Induction from Argan (Argania spinosa (L.) Skeels) Adventitious Shoots: Influence of Ammonium Nitrate, Auxins, Silver Nitrate and Putrescine, and Evaluation of Plantlet Acclimatization. Plants (Basel). 26 mai 2021;10(6):1062.

265. Bansal YK, Chibbar T. Micropropagation of Madhuca latifolia Macb. through nodal culture. Plant Biotechnology. 2000;17(1):17-20.

266. Jordan M, Oyanedel E. Regeneration of Pouteria lucuma (Sapotaceae) plants in vitro. Plant Cell Tiss Organ Cult. 1 déc 1992;31(3):249-52.

267. Purohit SD, Singhvi A. Micropropagation of *Achras sapota* through enhanced axillary branching. Scientia Horticulturae. 31 août 1998;76(3):219-29.

268. Pijut PM, Beasley RR, Lawson SS, Palla KJ, Stevens ME, Wang Y. In vitro propagation of tropical hardwood tree species—a review (2001–2011). Propagation of Ornamental Plants. 2012;12(1):25-51.

269. Aguwa I, Gana AS, Salaudeen MT, Tolorunse KD, Nmadu JN, Baba KM, et al. Effect of different concentrations of plant growth hormones on callus induction and regeneration of Shea tree (Vitellaria paradoxa). AJB. 31 juill 2022;21(7):334-41.

270. Manokari M, Mehta SR, S. P, Badhepuri MK, Dulam S, K. J, et al. Meta-Topolin mediated improved micropropagation, foliar micro-morphological traits, biochemical profiling, and assessment of genetic fidelity in *Santalum album* L. Industrial Crops and Products. 1 nov 2021;171:113931.

271. El-Sayed M, Aly UI, Mohamed MS, Rady MR. In vitro regeneration and molecular characterization of Jatropha curcas plant. Bulletin of the National Research Centre. 6 mai 2020;44(1):70.

272. Rohela GK, Jogam P, Saini P, Sandhya D, Peddaboina V, Shekhawat MS. Assessing the Genetic Stability of In Vitro Raised Plants. In: Gupta S, Chaturvedi P, éditeurs. Commercial Scale Tissue Culture for Horticulture and Plantation Crops [Internet]. Singapore: Springer Nature Singapore; 2022 [cité

20 juin 2024]. p. 245-76. Disponible sur: https://link.springer.com/10.1007/978-981-19-0055-6_11

273. Wu B, Zhang NS, Dixon B, Sierra I, Kan S, Layton A, et al. Reliable callus-induced plantlet regeneration from leaf explants of Lagerstroemia speciosa and genetic fidelity assessment through ISSR markers. Plant Cell Tiss Organ Cult. 14 juin 2024;157(3):76.

274. Savitikadi P, Jogam P, Rohela GK, Ellendula R, Sandhya D, Allini VR, et al. Direct regeneration and genetic fidelity analysis of regenerated plants of *Andrographis echioides* (L.) - An important medicinal plant. Industrial Crops and Products. 1 nov 2020;155:112766.

275. Badhepuri MK, Beeravelli PR, Arolla RG, Jogam P, Rohela GK, Singisala NR. Micropropagation and genetic fidelity analysis using SCoT and ISSR markers in Muehlenbeckia platyclada (F.Muell.) meisn. Plant Cell Tiss Organ Cult. 17 mai 2024;157(3):51.

276. Kudikala H, Jogam P, Sirikonda A, Mood K, Allini VR. In vitro micropropagation and genetic fidelity studies using SCoT and ISSR primers in Annona reticulata L.: an important medicinal plant. Vegetos. 1 sept 2020;33(3):446-57.

277. Serrote CML, Reiniger LRS, Silva KB, Rabaiolli SM dos S, Stefanel CM. Determining the Polymorphism Information Content of a molecular marker. Gene. 5 févr 2020;726:144175.

278. Kaur S, Panesar PS, Bera MB, Kaur V. Simple Sequence Repeat Markers in Genetic Divergence and Marker-Assisted Selection of Rice Cultivars: A Review. Critical Reviews in Food Science and Nutrition. 2 janv 2015;55(1):41-9.

279. Murashige T, Skoog F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia Plantarum. 1962;15(3):473-97.

280. Ilczuk A, Jacygrad E. In vitro propagation and assessment of genetic stability of acclimated plantlets of Cornus alba L. using RAPD and ISSR markers. In Vitro CellDevBiol-Plant. 1 sept 2016;52(4):379-90.

281. Bell RL, Srinivasan C, Lomberk D. Effect of nutrient media on axillary shoot proliferation and preconditioning for adventitious shoot regeneration of pears. In Vitro CellDevBiol-Plant. 1 déc 2009;45(6):708-14.

282. Dobránszki J, da Silva JAT. Micropropagation of apple--a review. Biotechnol Adv. 2010;28(4):462-88.

283. Shekhawat MS, Kannan N, Manokari M, Ravindran CP. In vitro regeneration of shoots and ex vitro rooting of an important medicinal plant Passiflora foetida L. through nodal segment cultures. J Genet Eng Biotechnol. déc 2015;13(2):209-14.

284. Arab MM, Yadollahi A, Shojaeiyan A, Shokri S, Ghojah SM. Effects of nutrient media, different cytokinin types and their concentrations on in vitro multiplication of G\times N15 (hybrid of almond\times peach) vegetative rootstock. Journal of genetic engineering and biotechnology. 2014;12(2):81-7.

285. Ashraf MF, Aziz MA, Kemat N, Ismail I. Effect of cytokinin types, concentrations and their interactions on in vitro shoot regeneration of Chlorophytum borivilianum Sant. & Fernandez. Electronic Journal of Biotechnology. nov 2014;17(6):275-9.

286. Hausman JF, Gevers C, Gaspar T. Involvement of putrescine in the inductive rooting phase of poplar shoots raised in vitro. Physiologia Plantarum. 1994;92(2):201-6.

287. Silva KB da, Reiniger LRS, Stefanel CM, Rabaiolli SM dos S. Sucrose and substrates on the acclimatization of micropropagated *Luehea divaricata plants*. Floresta Ambient. 27 mars 2020;27:e20171170.

288. Lone AB, Barbosa CM, Takahashi LSA, Faria RT de. Acclimatization of the Cattleya (Orchidaceae) in alternative substrates to tree fern fiber and sphagno. Acta Sci, Agron. déc 2008;30:465-9.

289. Stefanello S, Silveira EV, Oliveira LK, Besson JCF, Dutra GMN. Efficiency of substrates on acclimatization of in vitro propagated Miltonia flavescens Lindl. Rev Em Agronegócios E Meio Ambiente. 2009;2:467-76.

290. Zaytseva YG, Ambros EV, Novikova TI. Meta-topolin: Advantages and Disadvantages for In Vitro Propagation. In: Ahmad N, Strnad M, éditeurs. Meta-topolin: A Growth Regulator for Plant Biotechnology and Agriculture [Internet]. Singapore: Springer; 2021 [cité 8 avr 2024]. p. 119-41. Disponible sur: https://doi.org/10.1007/978-981-15-9046-7_11

291. Naaz A, Hussain SA, Anis M, Alatar AA. Meta-topolin improved micropropagation in Syzygium cumini and acclimatization to ex vitro conditions. Biologia plantarum. 19 janv 2019;63(1):174-82.

292. Ahmad A, Ahmad N, Anis M. Preconditioning of Nodal Explants in Thidiazuron-Supplemented Liquid Media Improves Shoot Multiplication in Pterocarpus marsupium (Roxb.). In: Ahmad N, Faisal M, éditeurs. Thidiazuron: From Urea Derivative to Plant Growth Regulator [Internet]. Singapore: Springer; 2018 [cité 10 avr 2024]. p. 175-87. Disponible sur: https://doi.org/10.1007/978-981-10-8004-3_8

293. Jena S, Ray A, Sahoo A, Sahoo S, Kar B, Panda PC, et al. High-frequency clonal propagation of Curcuma angustifolia ensuring genetic fidelity of micropropagated plants. Plant Cell Tiss Organ Cult. 1 déc 2018;135(3):473-86.

294. Yadav R, Yadav N, Kumar S. An Improved Micropropagation and Assessment of Genetic Fidelity in Multipurpose Medicinal Tree, Acacia auriculiformis. Proc Natl Acad Sci, India, Sect B Biol Sci. 1 déc 2016;86(4):921-9.

295. Kaur H, Lekhak MM, Ochatt SJ, Kumar V. Somatic embryogenesis and genetic homogeneity assessment of regenerated plants of Crinum brachynema (Amaryllidaceae): an endemic critically endangered medicinal plant. Plant Cell Tiss Organ Cult. 18 déc 2023;156(2):33.

296. Yadav R, Yadav N, Pal M, Goutam U. Multiple shoot proliferation, bulblet induction and evaluation of genetic stability in Asiatic hybrid lily (Lilium sp.). Ind J Plant Physiol. 1 déc 2013;18(4):354-9.

297. Pabin J, Lyngdoh N, Dobhal S, Upadhyaya K, Devi MB. Participatory tree selection and fruit morphological characterization of Phoebe cooperiana (U.N Kanjilal ex A. Das) in the Eastern Himalayas of India. Plant Genetic Resources. déc 2023;21(6):481-9.

298. Gesesse CA, Nigir B, de Sousa K, Gianfranceschi L, Gallo GR, Poland J, et al. Genomics-driven breeding for local adaptation of durum wheat is enhanced by farmers' traditional knowledge. Proc Natl Acad Sci U S A. 120(14):e2205774119.

299. Marketing and Societal Issues in Breeding. In: Principles of Plant Genetics and Breeding [Internet]. John Wiley & Sons, Ltd; 2012 [cité 2 févr 2025]. p. 489-489. Disponible sur: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118313718.part8

300. Pawera L, Manickam R, Wangungu C, Bonnarith U, Schreinemachers P, Ramasamy S. Guidance on farmer participation in the design, testing and scaling of agricultural innovations. Agricultural Systems. 1 juin 2024;218:104006.

301. Dossou-Yovo ER, Arouna A, Benfica R, Mujawamariya G, Yossa R. A participatory framework for prioritizing climate-smart agriculture innovations in rice-based systems: A case study of Mali. Smart Agricultural Technology. 1 mars 2024;7:100392.

302. Colley MR, Dawson JC, McCluskey C, Myers JR, Tracy WF, Bueren ETL van. Exploring the emergence of participatory plant breeding in countries of the Global North – a review. The Journal of Agricultural Science. juill 2021;159(5-6):320-38.

303. Whitlock R, Hipperson H, Thompson DBA, Butlin RK, Burke T. Consequences of *in-situ* strategies for the conservation of plant genetic diversity. Biological Conservation. 1 nov 2016;203:134-42.

304. Weiskopf SR, Rubenstein MA, Crozier LG, Gaichas S, Griffis R, Halofsky JE, et al. Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource management in the United States. Science of The Total Environment. 1 sept 2020;733:137782.

305. Kelly M. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. Philosophical Transactions of the Royal Society B: Biological Sciences. 28 janv 2019;374(1768):20180176.

306. Aurelle D, Thomas S, Albert C, Bally M, Bondeau A, Boudouresque CF, et al. Biodiversity, climate change, and adaptation in the Mediterranean. Ecosphere. 2022;13(4):e3915.

307. FAO. The role of genetic resources for food and agriculture in adaptation to and mitigation of climate change [Internet]. 2022 [cité 2 févr 2025]. Disponible sur: https://openknowledge.fao.org/items/c43bd34c-481d-4e38-832e-d07ff655f1a6

308. Rouichi S, Ghanem ME, Amri M. In-situ and ex-situ conservation priorities and distribution of lentil wild relatives under climate change: A modelling approach. Journal of Applied Ecology [Internet]. [cité 2 févr 2025];n/a(n/a). Disponible sur: https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.14842

309. Priyanka V, Kumar R, Dhaliwal I, Kaushik P. Germplasm Conservation: Instrumental in Agricultural Biodiversity—A Review. Sustainability. janv 2021;13(12):6743.

310. Migicovsky Z, Warschefsky E, Klein LL, Miller AJ. Using Living Germplasm Collections to Characterize, Improve, and Conserve Woody Perennials. Crop Science. 2019;59(6):2365-80.

311. De Mori G, Cipriani G. Marker-Assisted Selection in Breeding for Fruit Trait Improvement: A Review. International Journal of Molecular Sciences. janv 2023;24(10):8984.

312. Nayak SN, Singh VK, Varshney RK. Marker-Assisted Selection. In: Thomas B, Murray BG, Murphy DJ, éditeurs. Encyclopedia of Applied Plant Sciences (Second Edition) [Internet]. Oxford: Academic Press; 2017 [cité 2 févr 2025]. p. 183-97. Disponible sur: https://www.sciencedirect.com/science/article/pii/B9780123948076001921

313. Li L, Bian Y, Dong Y, Song J, Liu D, Zeng J, et al. Identification and validation of stable quantitative trait loci for yield component traits in wheat. The Crop Journal. 1 avr 2023;11(2):558-63.

314. Sinha D, Maurya AK, Abdi G, Majeed M, Agarwal R, Mukherjee R, et al. Integrated Genomic Selection for Accelerating Breeding Programs of Climate-Smart Cereals. Genes. juill 2023;14(7):1484.

315. Cobb JN, Biswas PS, Platten JD. Back to the future: revisiting MAS as a tool for modern plant breeding. Theor Appl Genet. 1 mars 2019;132(3):647-67.

316. Kumar R, Das SP, Choudhury BU, Kumar A, Prakash NR, Verma R, et al. Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing. Biological Research. 7 nov 2024;57(1):80.

317. Koloche IM, Ikuenobe CE, Solomon S, Enaberue LO, Garba H, Mohammed A, et al. Studies on Shea (Vitellaria paradoxa) fruits storage and different preservation technologies for domestication. IJBC. 30 avr 2021;13(2):42-8.

318. Yidana JA. Progress in developing technologies to domesticate the cultivation of shea tree (Vitellaria paradoxa L.) in Ghana Progres realises dans le developpement des technologies pour domestiquel la cultivation du karitier (Vitelleria paradoxa L.) au Ghana. Agricultural and Food Science Journal of Ghana. 2004;3:249-68.

319. Leakey RRB. Chapter 27 - The Participatory Domestication of West African Indigenous Fruits: This chapter was previously published in Leakey, R.R.B., Schreckenberg, K., Tchoundjeu, Z., 2003. International Forestry Review, 5, 338–347, with permission from Commonwealth Forestry Association. In: Leakey RRB, éditeur. Multifunctional Agriculture [Internet]. Academic Press; 2017 [cité 2 févr 2025]. p. 287-95. Disponible sur: https://www.sciencedirect.com/science/article/pii/B9780128053560000271

320. LEAKEY RRB, PAGE T. The 'Ideotype Concept' and Its Application to the Selection of Cultivars of Trees Providing Agroforestry Tree Products. Forests, Trees and Livelihoods. 1 janv 2006;16(1):5-16.

321. Sahito JH, Zhang H, Gishkori ZGN, Ma C, Wang Z, Ding D, et al. Advancements and Prospects of Genome-Wide Association Studies (GWAS) in Maize. International Journal of Molecular Sciences. janv 2024;25(3):1918.

322. Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genetics & Genomes. 1 avr 2011;7(2):241-55.

323. Haile B, Tesfaye B, Olango TM. Methods for vegetative propagation of wild enset (*Ensete ventricosum* (Welw.) Cheesman) that make genotype conservation possible. Heliyon. 1 nov 2021;7(11):e08416.

324. Şimşek Ö, Dalda Şekerci A, Isak MA, Bulut F, İzgü T, Tütüncü M, et al. Optimizing Micropropagation and Rooting Protocols for Diverse Lavender Genotypes: A Synergistic Approach Integrating Machine Learning Techniques. Horticulturae. janv 2024;10(1):52.

325. Karsen PA, Lötze E, Valentine AJ, Hoffman EW. Propagation and cultivation practices of honeybush (Cyclopia spp.) for the sustainable production of an export quality indigenous South African tea. Crop Science. 2022;62(5):1702-33.

326. Roo N de, Andersson JA, Krupnik TJ. On-farm trials for development impact? the organisation of research and the scaling of agricultural technologies. Experimental Agriculture. avr 2019;55(2):163-84.

327. Lacoste M, Cook S, McNee M, Gale D, Ingram J, Bellon-Maurel V, et al. On-Farm Experimentation to transform global agriculture. Nat Food. janv 2022;3(1):11-8.

328. Adefila AO, Ajayi OO, Toromade AS, Sam-Bulya NJ, Adefila AO, Ajayi OO, et al. Integrating traditional knowledge with modern agricultural practices: A sociocultural framework for sustainable development. World Journal of Biology Pharmacy and Health Sciences. 2024;20(2):125-35.

Appendices

Peer review publications and Scientific communication

Appendices. Peer review publications and scientific communication

Peer review publications

Publications and manuscripts included in the thesis

This thesis is a sum of five manuscript articles including four published articles (Articles 1, 2, 3 and 4), and one manuscript accepted (Article 5).

Article 1. **Attikora AJP**, Diarrassouba N, Yao SMD, De Clerck C, Silué S, Alabi T, Lassois L,. *Morphological traits and sustainability of plus shea trees (Vitellaria paradoxa C.F. Gaertn.) in Côte d'Ivoire. Biotechnol Agron Soc. Environ.* 25 sept 2023 DOI: 10.25518/1780-4507.20462

Article 2. **Attikora AJP**, Silué S, Yao SMD, De Clerck C, Shumbe L, Diarrassouba N, Fofana IJ, Alabi T, Lassois L. *An innovative optimized protocol for high-quality genomic DNA extraction from recalcitrant Shea tree (Vitellaria paradoxa, C.F. Gaertn) plant and its suitability for downstream applications*. *Mol Biol Rep* 51, 171 (2024). https://doi.org/10.1007/s11033-023-09098-6

Article 3. **Attikora AJP**, Yao SMD, Dago DN, Silué S, De Clerck C, Kwibuka Y, Diarrassouba N, Alabi T, Achigan-Dako EG, 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* 24, 913 (2024). https://doi.org/10.1186/s12870-024-05617-0

Article 4. Attikora AJP, Kouassi KA, Yoa SDM, Dago DN, Silué S, De Clerck C, Diarrassouba N, Alabi T, Achigan-Dako EG, Fauconnier ML, Danthine S, Lassois L. *Genome-wide association study of fat content and fatty acid composition of shea tree* (*Vitellaria paradoxa* C.F. Gaertn subsp. *paradoxa*). *BMC Genomics* **26**, 164 (2025). https://doi.org/10.1186/s12864-025-11344-z

Article 5. **Attikora AJP**, Silué S, Koné M, Silué N, Kwibuka Y, Yao SDM, De Clerck C, Him SL, Diarrassouba N, Alabi T, Lassois L. *Efficient in vitro Regeneration and Genetic Fidelity Analysis of Shea Tree (Vitellaria paradoxa Gaertn) Using ISSR Markers.* **Accepted in *Electronical Journal of Biotechnology***

Other peer review publications

Yao SDM, Diarrassouba N, **Attikora A**, Fofana IJ, Dago DN, 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.

Scientific communications

1. **Affi Jean Paul ATTIKORA** (2020). *Analysis of the morphological diversity of elite Shea trees (Vitellaria paradoxa, C. Gaertn.) identified in the Bagoué and Tchologo counties, Côte d'Ivoire.* The 25th National Symposium for Applied Biological Sciences (NSABS) conference at University of Liege, Gembloux Agro Bio-Tech, Espace Senghor, on 31st January 2020.

2. **Affi Jean Paul ATTIKORA** (2022). « *Caractérisation de la diversité des arbres à karité de Côte d'Ivoire et conseils de gestion pour une conservation durable du patrimoine génétique* ». 8th edition of the « Conférence Nationale du Karité du Bénin », Université d'Abomey-Calavi. December 8-10, 2022

3. **Affi Jean Paul ATTIKORA** (2024). *Genetic Diversity and Population Structure of Shea Tree for Establishing a Core Collection in Côte d'Ivoire*. AGROSYM2024, University of East-Sarajevo, Jahorina, Bosnia & Herzegovina. October 10-13, 2024

4. **Affi Jean Paul ATTIKORA** (2025). Genetic Diversity and Population Structure of Shea Trees for Sustainable Conservation and Breeding. PAG 32, San Diego, USA, January 10-15, 2025